Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

GENOMES IN SPACE AND TIME:

INSIGHTS INTO THE FUNCTIONAL THREE-DIMENSIONAL ORGANIZATION OF PROKARYOTIC AND EUKARYOTIC GENOMES IN RESPONSE TO ENVIRONMENTAL STIMULI AND CELL CYCLE PROGRESSION

A thesis presented in partial fulfilment of the requirements for the degree of

Doctorate in Philosophy

In

Genetics

At Massey University, Albany, New Zealand

Ralph Stefan Grand

BSc (Hons)

2014

ABSTRACT

The specific three-dimensional organization of prokaryotic and eukaryotic genomes and its contribution to cellular functions is increasingly being recognized as critical.

Bacterial chromosomes are highly condensed into a structure called the nucleoid. Despite the high degree of compaction in the nucleoid, the genome remains accessible to essential biological processes such as replication and transcription. Here I present the first high-resolution Chromosome Conformation Capture based molecular analysis of the spatial organization of the *Escherichia coli* nucleoid during rapid growth in rich medium and following an induced amino-acid starvation that promotes the stringent response. My analyses identified the presence of origin and terminus domains in exponentially growing cells. Moreover, I observe an increased number of interactions within the origin domain and significant clustering of SeqA binding sequences, suggesting a role for SeqA in clustering of newly replicated chromosomes. By contrast, “Histone-like” protein (*i.e.* Fis, IHF, H-NS) binding sites did not cluster suggesting that their role in global nucleoid organization does not manifest through the mediation of chromosomal contacts. Finally, genes that were down-regulated after induction of the stringent response were spatially clustered indicating that transcription in *E. coli* occurs at transcription foci.

The successful progression of a cell through the cell cycle requires the temporal regulation of gene expression, the number and condensation levels of chromosomes and numerous other processes. Despite this, detailed investigations into how the genome structure changes through the cell cycle and how these changes correlate with functional changes have yet to be performed. Here I present the results of a high resolution study in which we used synchronized Fission yeast (*Schizosaccharomyces pombe*) cells to investigate changes in genome organization and transcription patterns during the cell cycle. The small size of the Fission yeast genome makes this organism particularly amenable to studies of the spatial organization of its chromosomes. I detected cell cycle dependent changes in connections within and between chromosomes. My results show that chromosomes are effectively circular throughout the cell cycle and that they remain connected even during the M phase, in part by the co-localization of

repeat elements. Furthermore, I identified the formation and disruption of chromosomal interactions with specific groups of genes in a cell cycle dependent manner, linking genome organization and cell cycle stage specific transcription patterns. Determining the structure and transcript levels for matched synchronized cells revealed: 1) that telomeres of the same chromosome co-localization throughout the cell cycle, effectively circularizing the chromosomes; 2) that genes with high transcript levels are highly connected with other genomic loci and highly expressed genes at specific stages of the cell cycle; 3) that interactions have positive and negative effects on transcript levels depending on the gene in question; and 4) that metaphase chromosomes assume a 'polymer melt' like structure and remain interconnected with each other. I hypothesize that the observed correlations between transcript levels and the formation and disruption of cell cycle specific chromosomal interactions, implicate genome organization in epigenetic inheritance and bookmarking.

Over the course of mitochondrial evolution, the majority of genes required for its function have been transferred and integrated into nuclear chromosomes of eukaryotic cells. The ongoing transfer of mitochondrial DNA to the nucleus has been detected, but its functional significance has not been fully elucidated. To determine whether the recently detected interactions between the mitochondrial and nuclear genomes (mt-nDNA interactions) in *S. cerevisiae* are part of a DNA-based communication system I investigated how the reduction in interaction frequency of two mt-nDNA interactions (COX1-MSY1 and Q0182-RSM7) affected the transcript level of the nuclear genes (MSY1 and RSM7). I found that the reduction in interaction frequency correlated with increases in MSY1 and RSM7 transcript levels. To further investigate whether mt-nDNA interaction could be detected in other organisms and characterize their possible functional roles, I performed Genome Conformation Capture (GCC) on Fission yeast cell cycle synchronized in the G1, G2 and M phases of the cell cycle. I detected mt-nDNA interactions that vary in strength and number between the G1, G2 and M phases of the Fission yeast cell cycle. Mt-nDNA interactions formed during metaphase were associated with nuclear genes required for the regulation of cell growth and energy availability. Furthermore, mt-nDNA interactions formed during the G1 phase involved high efficiency, early firing replicating origins of DNA replication. Collectively, these results implicate the ongoing transfer of regions of the mitochondrial genome to the nucleus in the regulation of nuclear gene transcription and cell cycle progression following exit from metaphase. I propose that these

interactions represent an inter-organelle DNA-mediated communication mechanism.

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr Justin O'Sullivan for taking me on to do my PhD and for his guidance throughout. Having someone to discuss ideas with has been extremely beneficial to my development, thank you Justin. I am also grateful for the opportunities you gave me, your wisdom, continued support and constructive criticism.

Thank you to my co-supervisors Professor Rob Martienssen, Professor Paul Rainey, and Dr Austen Ganley for being there to answer my questions, discuss ideas, and for your support. Thank you also to Dr Beatrix Jones for helping me with my statistical questions and Professor Philippe Collas for accommodating me in his lab.

Thank you to all the staff and students in building 11, 12 and 14 that have come and gone over the years that I have been at Massey, in particular past and present members of the O'Sullivan group, as well as members of Austen Ganley's and Evelyn Sattlegger's group, for all your help, time, use of equipment, consultation and friendship. I could not have asked for a better bunch of encouraging and understanding people to work with.

During my PhD (and entire time at Massey University) I have had the privilege of meeting some amazing people. My laboratory buddy and partner in crime Robbie Wilson, the best lab demonstrator ever Jarod Young, Chris Rodley, Martina Dautel, Jyothsna Visweswaraiah, Saumya Agrawal, Mack Saraswat, Lutz Gehlen, Matthew Woods, and Andrew Cridge to name a few, thank you all very much for your time, knowledge and friendship. A special thanks to Lutz for teaching me how to program, it was great fun to learn and helped me extensively during my PhD. To all of those whose name is not mentioned that have helped me in some way, my apologies; thank you for your friendship and I am very appreciative of the time you took to help me out.

I would like to thank my parents and brother for their ongoing support and belief in me through the good and the bad times. Without them none of this would have been possible. Thank you.

To all my friends, thank you for being understanding on all the occasions when I could not do things because of work and being there when I had the time. Whether it was a fishing trip away for a few days, a gaze at the night sky, a beer or glass of wine, thanks you all for being my friends and supporting me throughout.

TABLE OF CONTENT

ABSTRACT I

ACKNOWLEDGMENTS V

TABLE OF CONTENT VII

FIGURES XIII

TABLES XVI

ABBREVIATIONS XVII

1 INTRODUCTION 2

1.1 THE THREE-DIMENSIONAL ORGANIZATION OF GENOMES 3
1.2 THE THREE HIERARCHICAL LEVELS OF GENOME ORGANIZATION 4
1.2.1 THE PRIMARY LEVEL OF GENOME ORGANIZATION: THE LINEAR ARRANGEMENT OF CHROMOSOMES 4
1.2.1.1 DNA replication of circular and linear chromosomes 7
1.2.2 THE SECOND LEVEL OF GENOME ORGANIZATION: SUPERCOILING AND PROTEIN BINDING ARE REQUIRED FOR THE COMPACTION OF CHROMOSOMES 9
1.2.2.1 Supercoiling is the primary mechanism for the secondary level organization of bacterial genomes 9
1.2.2.2 Efficient packaging of eukaryotic genomes is primarily accomplished by histone proteins 11
1.2.1 THE TERTIARY LEVEL OF GENOME ORGANIZATION: GENOMES IN SPACE 12
1.2.1.2 The spatial organization of the bacterial nucleoid 14
1.2.1.3 The spatial organization of eukaryotic genomes 18
1.2.1.3.1 Metaphase chromosomes 23
1.3 GENOMES IN TIME 24
1.3.1 THE RESPONSE TO ENVIRONMENTAL STIMULI 24
1.3.2 THE RESPONSE TO CELL CYCLE PROGRESSION 24
1.4 DNA TRANSFER BETWEEN ORGANELLES 26
1.5 CENTRAL QUESTIONS 29
1.6 METHODS FOR THE STUDY OF SPATIAL ORGANIZATION 30
1.6.1 MOLECULAR BIOLOGY TECHNIQUES FOR THE IDENTIFICATION OF CHROMOSOME ORGANIZATION 32
1.6.1.1 Cross-Linking 32
1.6.1.2 Chromosome Conformation Capture (3C) 35
1.6.1.3 Circular Chromosome Conformation Capture (4C) 36
1.6.1.4 Chromosome Conformation Capture Carbon-Copy (5C) 37
1.6.2 GLOBAL IDENTIFICATION OF DNA INTERACTIONS 37
1.6.2.1 Genome Conformation Capture (GCC) 38

1.6.2.2 Hi-C 39
1.6.3 METHODS TO LOCALIZE PROTEINS AT CHROMOSOME INTERACTION SITES 40
1.6.3.1 Combined 3C-ChIP-Cloning (6C) 40
1.6.3.2 Chromatin Interaction Analysis Using Paired-End Tag Sequencing (Chia-PET) 42
1.6.4 WHAT DO THE INTERACTIONS CAPTURED REPRESENT 42
1.7 GOALS OF THE THESIS 43

2 GENOME CONFORMATION CAPTURE REVEALS THAT THE *ESCHERICHIA COLI* CHROMOSOME IS ORGANIZED BY REPLICATION AND TRANSCRIPTION 46

2.1 INTRODUCTION 47
2.2 RESULTS 49
2.2.1 ORIGIN AND TERMINUS DOMAINS EXIST WITHIN THE *E. COLI* NUCLEOID 51
2.2.2 INTERACTIONS WITHIN THE ORI AND TER REGIONS ARE LINKED TO REPLICATION 52
2.2.3 CLUSTERING OF MATP AND SEQA BINDING SITES LINK NUCLEOID STRUCTURE AND REPLICATION 57
2.2.4 INTRA- OR INTER-NAP BINDING SITE CLUSTERING DOES NOT CONTRIBUTE TO THE GLOBAL ORGANIZATION OF THE *E. COLI* NUCLEOID. 60
2.2.5 GENES UP OR DOWN REGULATED FOLLOWING SHX TREATMENT EXIST IN DIFFERENT SPATIAL ENVIRONMENTS CONFIRMING FUNCTIONAL COMPARTMENTALIZATION OF THE NUCLEOID. 62
2.3 DISCUSSION 66
2.3.1 IS THE *E. COLI* NUCLEOID SHAPED AS A SAUSAGE OR ROSETTE? 67
2.3.2 REPLICATION CONTRIBUTES TO NUCLEOID ORGANIZATION THROUGH SEQA. 68
2.3.3 WHAT ROLE DOES THE MATS5-10 LOOP PLAY IN NUCLEOID ORGANIZATION? 70
2.3.4 DO "HISTONE-LIKE" NAPS PLAY A ROLE IN GLOBAL NUCLEOID STRUCTURE? 70
2.3.5 DO TRANSCRIPTION FOCI HAVE A ROLE IN NUCLEOID ORGANIZATION? 71
2.3.6 DOES A NUCLEOLUS-LIKE STRUCTURE FORM WITHIN THE *E. COLI* NUCLEOID? 72
2.3.7 EPISTATIC INTERACTIONS AND THE CHROMOSOME INTERACTION NETWORK 73
2.3.8 CONCLUSION 73
2.4 MATERIALS AND METHODS 74
2.4.1 STRAINS AND GROWTH CONDITIONS 74
2.4.2 GENOME CONFORMATION CAPTURE (GCC) 74
2.4.3 PRODUCTION OF EXTERNAL LIGATION CONTROLS FOR GCC LIBRARY PREPARATION 75
2.4.5 GENOME CONFORMATION CAPTURE NETWORK ASSEMBLY, EFFECTS OF SAMPLE PRODUCTION AND PROCESSING, AND BIOINFORMATICS ANALYSIS 76
2.4.6 QUANTIFICATION OF THE EFFECTS OF BAR-CODING, SEQUENCING AND BIOLOGICAL REPLICATES ON NETWORK PATTERNS 77
2.4.7 FILTERING BASED ON LIGATION CONTROLS 78
2.4.8 COLLECTOR'S CURVE 79
2.4.9 ANALYSIS OF LOOP SIZE AND INTERACTING FRAGMENT DISTRIBUTIONS 80
2.4.10 INTERACTION FREQUENCY VERSUS DISTANCE FROM THE ORIGIN OF REPLICATION 81
2.4.11 GENOME COPY NUMBER 81
2.4.12 TRANSCRIPTION MICROARRAY 81

2.4.13 GENE ONTOLOGY (GO) TERM ENRICHMENT ANALYSIS OF THE SIGNIFICANTLY UP AND DOWN REGULATED GENES 82
2.4.14 CORRELATING TRANSCRIPTION LEVEL WITH INTERACTION FREQUENCY 82
2.4.15 TRANSCRIPTION REGULONS 83
2.4.16 GC CONTENT AND TRANSCRIPTION LEVELS 83
2.4.17 MATS, SEQA, SLMA AND NAP CLUSTERING ANALYSES 83
2.4.18 INTERACTIONS AND CLUSTERING OF GENES THAT SIGNIFICANTLY CHANGE THEIR EXPRESSION LEVEL UPON SHX TREATMENT 84
2.5 DATA ACCESS 84
2.6 FUNDING 84

3 GLOBAL ANALYSES OF CELL CYCLE DEPENDENT CHANGES IN FISSION YEAST GENOME ORGANIZATION REVEAL CORRELATIONS WITH ALTERATIONS IN TRANSCRIPT LEVELS 85

3.1 INTRODUCTION 86
3.2 RESULTS 88
3.2.1 *S. POMBE* GENOME ORGANIZATION CHANGES THROUGHOUT THE CELL CYCLE 88
3.2.2 CHANGES IN INTERACTIONS BETWEEN REPEAT CONTAINING GENOMIC REGIONS CONTRIBUTE TO CELL CYCLE SPECIFIC GENOME ORGANIZATION AND GENE EXPRESSION 96
3.2.3 CELL CYCLE DEPENDENT INTERACTIONS ARE NOT JUST ABOUT GENE REGULATION 100
3.2.4 GENES WITH HIGH TRANSCRIPT LEVELS THROUGHOUT THE CELL CYCLE ARE ASSOCIATED WITH A HIGH PROPORTION OF INTER-CHROMOSOMAL CO-LOCALIZATION. 105
3.2.5 GENES DIFFERENTIALLY REGULATED DURING THE G1 – G2 – M CELL CYCLE PHASE TRANSITIONS WERE LINEARLY AND SPATIALLY CO-LOCALIZED 110
3.3 DISCUSSION 114
3.4 CONCLUSION 117
3.5 MATERIALS AND METHODS 118
3.5.1 STRAINS, GROWTH CONDITIONS AND SYNCHRONIZATION 118
3.5.2 SYNCHRONIZATION EFFICIENCY 118
3.5.3 CHROMATIN ISOLATION FOR GENOME CONFORMATION CAPTURE (GCC) 119
3.5.4 PRODUCTION OF EXTERNAL LIGATION CONTROLS FOR GCC LIBRARY PREPARATION 120
3.5.5 NETWORK ASSEMBLY 121
3.5.6 SIGNIFICANCE CUT-OFF CALCULATIONS 121
3.5.7 COLLECTOR'S CURVE 122
3.5.8 HEAT MAPS 122
3.5.10 RNA EXTRACTION 123
3.5.11 TRANSCRIPTOME ANALYSIS 123
3.5.12 THE CHROMOSOME DISTRIBUTION OF GENES WITH HIGH, LOW, AND DIFFERENTIAL TRANSCRIPT LEVELS 124
3.5.13 LOOP LENGTHS OF INTERACTIONS WITHIN CHROMOSOMES 124
3.5.14 THE DETERMINATION OF GENOME CONNECTIVITY AND CO-LOCALIZATION (CLUSTERING) LEVELS 125
3.5.15 GENE ONTOLOGY (GO) ANALYSIS 125

3.5.16 EXTRACTING GENES ASSOCIATED WITH DNA INTERACTIONS FOR GO AND TRANSCRIPTION OVERLAP ANALYSES 125
3.5.17 DETERMINING THE LIKELIHOOD OF ASSOCIATION WITH SPECIFIC INTERACTIONS 126

4 MITOCHONDRIAL-NUCLEAR DNA INTERACTIONS CONTRIBUTE TO THE REGULATION OF NUCLEAR TRANSCRIPT LEVELS AND THE REGULATION OF CELL CYCLE PROGRESSION 127

4.1 INTRODUCTION 129
4.2 RESULTS 132
4.2.1 MT-NDNA INTERACTIONS CONTRIBUTE TO THE REGULATION OF NUCLEAR TRANSCRIPT LEVELS AS PART OF THE INTER-ORGANELLE COMMUNICATION SYSTEM 132
4.2.2 MT-NDNA INTERACTIONS VARY THROUGHOUT THE *S. POMBE* CELL CYCLE 136
4.2.3 MT-NDNA INTERACTIONS ARE NOT ENRICHED FOR NUCLEAR ENCODED MITOCHONDRIAL GENES. 142
4.2.4 MT-NDNA INTERACTIONS SPECIFICALLY FORMED DURING METAPHASE OCCUR WITH GENES REQUIRED FOR CELL GROWTH AND DNA SYNTHESIS 144
4.2.5 G1 PHASE MT-NDNA INTERACTIONS ARE ASSOCIATED WITH HIGH EFFICIENCY, EARLY REPLICATING ORIGINS OF DNA REPLICATION 149
4.3 DISCUSSION 152
4.3.1 MT-NDNA INTERACTIONS REGULATE NUCLEAR ENCODED GENES IN *S. CEREVISIAE* 152
4.3.2 MT-NDNA INTERACTIONS DETECTED AT SPECIFIC PHASES OF THE *S. POMBE* CELL CYCLE ARE IMPLICATED IN THE ESTABLISHMENT OF TRANSCRIPTION AND INITIATING DNA REPLICATION 153
4.4 MATERIALS AND METHODS 156
4.4.1 STRAINS AND GROWTH CONDITIONS 156
4.4.1.1 Saccharomyces cerevisiae 156
4.4.1.3 Schizosaccharomyces pombe 157
4.4.2 HARVESTING CELLS AND CHROMATIN PREPARATION 157
4.4.3 CHROMOSOME CONFORMATION CAPTURE (3C) SAMPLE PREPARATION 157
4.4.4 QUANTITATIVE 3C ANALYSES 158
4.4.5 RNA EXTRACTION 158
4.4.6 QUANTITATIVE REVERSE TRANSCRIPTION-PCR 159
4.4.7 THE GENERATION OF *S. POMBE* DATA 159
4.4.8 SIGNIFICANCE CUT-OFF CALCULATIONS 159
4.4.9 EXTRACTING NUCLEAR GENES ASSOCIATED WITH MT-NDNA INTERACTIONS FOR GO AND TRANSCRIPTION ANALYSES 160
4.4.10 DETERMINING IF NUCLEAR ENCODED GENE SETS ASSOCIATED WITH MT-NDNA INTERACTIONS WERE DETECTED IN SPECIFIC TRANSCRIPTION DATA SETS 160
4.4.11 DETERMINING WHETHER NUCLEAR RESTRICTION FRAGMENTS INVOLVED IN MT-NDNA INTERACTIONS OVERLAP WITH NUCLEAR ENCODED MITOCHONDRIAL GENES AND ARS SITES 161

5 DISCUSSION AND FUTURE DIRECTIONS 163

5.1 CONSERVED AND STRUCTURAL VS DYNAMIC AND FUNCTIONAL 165

5.2 RESPONSE TO THE ENVIRONMENT AND CELL CYCLE PROGRESSION 170
5.2.1 THE LINEAR AND SPATIAL CLUSTERING OF CO-REGULATED GENES 171
5.2.3 SPATIAL BOOKMARKING? 175
5.3 REPEAT ELEMENTS MAY BE INVOLVED IN THE ESTABLISHMENT OF SPECIFIC TRANSCRIPTIONAL PROFILES 176
5.4 SHARING INFORMATION FOR A MUTUAL BENEFIT 178
5.4.1 MT-NDNA INTERACTIONS PARTICIPATE IN THE COORDINATION OF MITOCHONDRIAL AND NUCLEAR PROCESSES 178
5.4.2 MITOCHONDRIA IN NEURONS 179
5.4.3 IMPLICATIONS FOR THE CO-EVOLUTION OF OTHER HOST-PARASITE INTERACTIONS 180
5.5 FINAL REMARKS 181

6 BIBLIOGRAPHY 183

7 APPENDIX I 212

8 APPENDIX II 239

9 APPENDIX III 253

10 APPENDIX IV 264

FIGURES

Figure 1.1. Nuclear architecture is functionally linked to the organization and sorting of regulatory information. 5
Figure 1.2. DNA in eukaryotic cells is wound around nucleosomes that interact to form a chromatin fibre. 13
Figure 1.3. Molecular methods for the detection of chromosomal interactions. Refer to footnote for the figure legend. 34
Figure 2.1. Ori and Ter domains are present within the *E. coli* nucleoid. 50
Figure 2.2. There is a reduction in the number of long distance interactions in the SHX treated nucleoids. 53
Figure 2.3. Origin proximal interactions are more frequently detected. 54
Figure 2.4. There is an origin preferential reduction in the number of partners each restriction fragment interacts with in the SHX treated cells. 56
Figure 2.5. Binding sites for nucleoid associated proteins MatP and SeqA exhibit differing degrees of spatial clustering within the exponential and SHX treated E. coli nucleoids. 58
Figure 2.6. Annotated genes with transcripts that were up (644 genes) or down (687 genes) regulated following SHX treatment existed in different spatial environments. Refer to footnote for figure legend. 64
Figure 2.7. Genes whose transcript level changes upon SHX treatment did not correlate with genomic GC content and the raw transcript levels did not correlate with interaction frequency. 65
Figure 2.8. Spatial model of exponential phase nucleoid organization in *E. coli*. 68
Figure 2.9. Interactions attributed to H-NS clustering identified by Wang et al. 2011 were confirmed as occurring between A) gadA:gadB and B) aceA:yddB (indicated by the grey bars linking the ORFs (black boxes)). 71
Figure 2.10. Biological replicates of the E. coli interaction networks were highly correlated at the HhaI restriction fragment level. 77
Figure 2.11. A collectors curve was generated to determine the level of saturation of interaction detection. 80
Figure 3.1. The spatial organization of the *S. pombe* genome changes throughout the cell cycle. Refer to footnote for figure legend. 90
Figure 3.2. Correlation plots of *S. pombe* interaction networks for individual biological replicates and for each cell cycle phase. 91
Figure 3.3. The number of unique interactions detected was not fully saturated. 92
Figure 3.4. There is a marked increase in the number of within chromosome interactions with a loop length of up to ~5Kb in M phase chromosomes. 93
Figure 3.5. Interactions within chromosomes were predominantly shared by the three cell cycle phases, while interactions between chromosomes were largely cell cycle specific. 94
Figure 3.6. All three chromosomes were circularized throughout the cell cycle. 95
Figure 3.7. Centromere clustering was observed in heat maps of interactions involving repetitive loci. Moreover, repeat regions appear to demarcate the genome. 97
Figure 3.8. Repeat elements contribute to the formation of cell cycle specific genome organization. 98

Figure 3.9. Co-localisation between LTR elements occurred predomenantly between chromosomes or LTR elements that were > 50 Kb apart on the same chromosome. .. 99
Figure 3.10. Genes that are differentially regulated during the G2 – M – G1 cell cycle transitions are overrepresented within interactions that form during this period of the cell cycle. Refer to footnote for figure legend. 101
Figure 3.11. Description of gene sets used for GO and overlap analysis. 102
Figure 3.12. Genes with high transcript levels are highly conserved throughtout the cell cycle while low transcript level genes are not. 105
Figure 3.13. Genes with high transcript levels are more highly connected to the genome and co-localize more frequently than low transcript level genes. 106
Figure 3.14. The percentage of inter- compared to intra-chromosomal co-localization for genes with high, low and differential transcript levels. 108
Figure 3.15. Genes with high transcript levels were non-randomly distributed across one or more chromosomes at each stage of the cell cycle. 109
Figure 3.16. Genes that undergo differential expression during cell cycle phase transitions are found in specific interaction environments and co-localize in a cell cycle specific way. 112
Figure 3.17. Cell cycle specific differentially regulated genes are non-randomly distributed across the linear chromosomal sequences. 113
Figure 3.18. The level of cell cycle phase synchronization was calculated using the septation index of *S. pombe* cells before and after synchronization. 119
Figure 4.1. Mt-nDNA interactions require active mitochondrial reverse transcriptase machinery. 134
Figure 4.2. Knocking out mitochondrial encoded reverse transcriptase activity results in increased transcript levels of nuclear genes that are involved in mt-nDNA interactions. 135
Figure 4.3. Deletion of *MRS1* (BY4741 *⊿mrs1*), a nuclear gene involved in splicing mitochondrial type-I introns, has no significant effect on the frequency of the *COX1-MSY1* interaction in glucose grown yeast cells. 136
Figure 4.4. Mitochondrial interaction networks for the biological replicates were highly correlated at the AseI restriction fragment level. 137
Figure 4.5. Mt-nDNA interactions are cell cycle dependent. 138
Figure 4.6. Mito-nuclear DNA interactions are largely specific to each cell cycle phase. By contrast, a large number of within mitochondrial interactions were unique to the M phase of the cell cycle. 140
Figure 4.7. Mitochondrial fragments that interact with nuclear chromosomes also interact with other mitochondrial fragments and M phase mitochondrial genomes are the most highly connected. 141
Figure 4.8. Mt-nDNA interactions do not preferentially involve nuclear encoded mitochondrial genes. 143
Figure 4.9. Nuclear genes associated with mt-nDNA interactions formed specifically during metaphase are enriched in biological functions related to genomic DNA synthesis and energy regulation. 145
Figure 4.10. Mt-nDNA interactions that formed specifically during metaphase were associated with highly transcribed nuclear genes required for ribosome function. 147
Figure 4.11. Mt-nDNA interactions do not globally correlate with the differential regulation of nuclear genes. 148

Figure 4.12. G1 phase mt-nDNA interactions are enriched for high efficiency, early replicating autonomously replicating sequences (origins of replication)..... 150
Figure 4.13. Three mitochondrial fragments are involved in the majority of the interactions with ARS sites during the G1 phase of the cell cycle. 151

TABLES

Table 1.1 Advantages and disadvantages of the different molecular methods for the detection of chromosomal interactions 41
Table 2.1. A summary of *E. coli* chromosomal interactions. 52
Table 2.2. MatS loci were highly clustered in exponentially growing cells. 58
Table 2.3. SeqA sites cluster within the *E. coli* nucleoid. Refer to footnote for the table legend. 59
Table 2.4. SlmA is highly connected with the genome but not with other SlmA sites. 61
Table 2.5. H-NS, IHF and Fis sites do not exhibit spatial clustering. 61
Table 2.6. Genes up and down regulated in response to SHX treatment were enriched in specific Gene Ontology terms. 63
Table 2.7. Genes that were highly expressed in exponential (exp) cells became weakly expressed in SHX treated cells and vice versa. 67
Table 2.8 *Escherichia coli* strain used in this study. 74
Table 2.9. PCR primers used to generate the ligation controls used in this study. 75
Table 2.10. Barcoding, sequencing lane and biological replicates did not affect the correlation between samples. 78
Table 2.11. The number of annotated *E. coli* genes that were significantly up or down regulated in response to SHX treatment. 82
Table 3.1. A large proportion of interactions detected at each phase of the cell cycle were within chromosomes. 92
Table 3.2. The restriction fragments involved in a high frequency interaction detected within chromosome II both contained LTR elements. 99
Table 3.3. The absence of the interaction between two LTR elements in chromosome II in G2 phase was associated with the up regulation of an overlapping gene. 99
Table 3.4. The number of genes that were significantly differentially regulated during each *S. pombe* cell cycle transition. 104
Table 3.5. Table highlighting the number of genes that had the highest (top 5%) and lowest (bottom 5%) transcript levels at each cell cycle phase and whether they were differentially regulated during cell cycle transitions. 104
Table 3.6. *Schizosaccharomyces pombe* strains used in this study. 118
Table 3.7. An example calculation of the cell culture synchronization efficiency. 119
Table 3.8. Ligation controls used in this study. 121
Table 4.1. The number of interactions within the mitochondrial genome and between the mitochondrial and nuclear genomes varies at different phases of the *S. pombe* cell cycle. 140
Table 4.2. The numbers of *S. pombe* genes within the highest, lowest, and differential transcript levels at each stage of the cell cycle. 144
Table 4.3. *Saccharomyces cerevisiae* strains used in this study. 156
Table 4.4. Primers and probes used in this study. 158

ABBREVIATIONS

h	hour
min	minute
s	second
mmol	millimole
temp	temperature
Rm	room
O/N	overnight
OD	optical density
nm	nano-meter
MQ	milliQ
H_2O	water
µl	microliters
rpm	revolutions per minute
ºC	degrees celsius
µg	micrograms
U	units
v/v	volume per volume
w/v	weight per volume
ml	milliliters
g	g-force
Amp	ampicillin
PCR	polymerase chain reaction
V	volts
bp	base pair
Kb	kilo-base pair
Mb	mega-base pair
NTC	no template control
EM	electron microscopy
µm	micro-meter
GCC	genome conformation capture
LTR	long terminal repeat
DNA	Deoxyribonucleic Acid
mtDNA	mitochondrial DNA
nDNA	nuclear DNA
mt-nDNA	mitochondrial – nuclear DNA interactions

3C	chromosome conformation capture
4C	circular chromosome conformation capture
5C	chromosome conformation capture carbon copy
6C	combined 3C-ChIP-cloning
ChIA-PET	chromatin interaction analysis with paired-end tag sequencing
G1 phase	growth one phase of the cell cycle
S phase	DNA synthesis phase of the cell cycle
G2 phase	growth two phase of the cell cycle
M phase	metaphase of the cell cycle
PE	paired-end

Chapter 1

Section 1.6 has been published see Appendix I:

Grand, R.S., Gehlen, L.R., and O'Sullivan, J.M. (2011). Methods for the Investigation of Chromosome Organization. Advances in Genetics Research, 5: 111-129.

Institute of Natural Sciences, Massey University Auckland, New Zealand.

Acknowledgements: The authors would like to apologize to colleagues whose work was not cited. Thanks to Dr Sue Mei Tan Wong for reviewing this chapter. The authors would also like to thank A. Ganley, R. McNab, and members of the O'Sullivan laboratory for comments on this chapter. LRG is funded by an SNSF fellowship (PBBSP3-130910). RSG is funded by a Massey University Ph.D. scholarship. Work in JMO's laboratory is funded by the Maurice & Phyllis Paykel trust and the Auckland Medical Research Foundation.

1 INTRODUCTION

The cell is the basic unit of all known life. It possesses characteristics shared by all living organisms, including the ability to sense and respond to the environment, and grow and reproduce. Cells are divided into two main classes, prokaryotes and eukaryotes. Prokaryotic cells are small (0.2-2 μm in diameter) and commonly contain genomes that range in size from 1-10 Mb. The genomes of prokaryotes can be composed of single or multiple circular and/or linear chromosomes and are not enclosed within a nuclear membrane. By contrast, eukaryotic cells are generally larger than prokaryotes (10-100 μm in diameter) and contain genomes >10 Mb in length. Eukaryotic genomes predominantly consist of multiple linear chromosomes that are contained within a nucleus. In addition, eukaryotic cells contain extra-nuclear DNA that exists within organelles, for example the mitochondria.

With the rapid development of DNA sequencing technology over recent years, came the hope that knowing the composition and distribution of genetic elements *e.g.* genes and regulatory sequences, in the DNA sequence would result in the understanding of an organism. However, it has become increasingly evident that the linear arrangement of elements in the DNA sequence cannot fully account for the genotype to phenotype translation observed (Weatherall, 2001). The spatial (three-dimensional (3D)) organization of chromosome(s) has increasingly been implicated as a key contributor to the genotype – phenotype translation. To gain an integrated understanding of the genotype – phenotype relationship, the role that the spatial organization of chromosomes, and changes therein, has in cellular processes, such as DNA replication and transcription, needs to be unravelled.

The recent development of Genome Conformation Capture (GCC) (Rodley, Bertels, Jones, & O'Sullivan, 2009) enabled, for the first time, the investigation of the *in vivo* 3D organization of chromosomes at the molecular level. GCC combines 'proximity-based ligation' with 'genome-wide sequencing' and was made possible due to the rapid advancement of next generation sequencing technology. Further combining GCC with gene expression data provides a unique opportunity to investigate how specific changes in spatial genome organization, in response to stimuli and through time (the fourth dimension), relates to changes in cellular processes that enable an organism to adapt.

1.1 THE THREE-DIMENSIONAL ORGANIZATION OF GENOMES

Ever since chromosomes were first described in the 1900s, the use of staining methods followed by visualization with light or electron microscopy, has clearly shown that chromosome(s) in prokaryotic and eukaryotic cells are organized in three-dimensions (Aula & Saksela, 1972; Fukui, 2009; Kite & Chambers, 1912; Marshak, 1951; Van Winkle, Renoll, Garvey, Palik, & Prebus, 1953). Advances in microscopic and proximity-based ligation methodologies have resulted in the accumulation of knowledge about the role that genome organization has in cellular processes (T. Cremer & Cremer, 2001; Grand, Gehlen, & O'Sullivan, 2011). These studies have led to the current view, where it is generally accepted that genomes are highly organized in space in a way that participates in the regulation of cellular processes (Berezney, 2002; Nicodemi & Prisco, 2009; Rippe, 2007).

Prokaryotic chromosomal DNA, which can consist of single or multiple circular and/or linear chromosomes, is not contained within a nuclear membrane (Casjens, 1998; Casjens et al., 2000; Suzuki, Iwata, & Yoshida, 2001). Despite this, prokaryotic chromosomes occupy a region at the centre of the cell that together with proteins and RNA forms a structure called the nucleoid. The nucleoid appears to be highly organized in space to facilitate the regulation of essential processes, including ribosomal DNA transcription, chromosome replication and segregation (Cabrera & Jin, 2006; Jin, Cagliero, & Zhou, 2012; Marshak, 1951; Mercier et al., 2008; Robinow & Kellenberger, 1994). Collectively, these observations support an important role for the spatial organization of the nucleoid in bacterial cell growth and survival.

The presence of a nucleus in eukaryotic cells, which separates the chromosomes and other cellular components from the cytoplasm, enables the establishment of specialized nuclear compartments. Chromosomes and many nuclear components have been visualized to occupy distinct subcompartments, formed by the packaging of chromatin fibres and aggregation of specific nuclear processing factors (Figure 1.1). For example, chromosomes occupy distinct regions called chromosomes territories (Bolzer et al., 2005; Boyle et al., 2001; Heard & Bickmore, 2007; Rinke et al., 1995; Schardin, Cremer, Hager, & Lang, 1985); actively transcribing RNA polymerase molecules aggregate into highly concentrated foci called transcription factories (Carter, Eskiw, & Cook, 2008; S.

Martin & Pombo, 2003); and pre-mRNA processing factors aggregate into nuclear speckles and Cajal bodies (Mintz & Spector, 2000; Morris, 2008; Nesic, Tanackovic, & Krämer, 2004). This organization allows cellular processes to occur simultaneously in separate subcompartments without interfering with each other.

The organization of eukaryotic and presumably also bacterial genomes has been proposed to manifest through three hierarchical levels of organization: 1) The organization of chromatin - a DNA and protein complex - into domains; 2) The spatial organization of nuclear processes such as transcription and replication (Figure 1.1); and 3) The interplay between the two, to arrange the genes and chromosomes in the nuclear (or cellular) space (Misteli, 2007). Each of these levels of organization has the potential to be regulated; therefore, it is not difficult to envisage that the regulation of nuclear processes in space and time is an extraordinarily complex process. In addition, describing the spatial organization of a single genome is a formidable problem, for two main reasons: 1) a genome's structure cannot be defined by a single spatial structure because the DNA/chromatin fibre is highly dynamic, with changes in condensation, positioning of particular loci and DNA content between individual cells; and 2) as a consequence of (1), even if the genome structure was relatively well defined in any one cell, there may be large variations between individual cells (Langowski, 2010). Together, the interplay between chromatin organization and nuclear processes, and the probabilistic nature of chromatin behaviour, illustrate why understanding genome structure is one of the most challenging problems in structural biology. Despite this challenge, it is a field of great interest because of the emerging role that genome structure has in the manifestation of disease and the potential for insights into the fundamentals of life itself (Göndör & Ohlsson, 2009; Kauffman, 1993; Marella, Bhattacharya, Mukherjee, Xu, & Berezney, 2009; Rajapakse & Groudine, 2011; Sanyal, Lajoie, Jain, & Dekker, 2012).

1.2 THE THREE HIERARCHICAL LEVELS OF GENOME ORGANIZATION

1.2.1 THE PRIMARY LEVEL OF GENOME ORGANIZATION: THE LINEAR ARRANGEMENT OF CHROMOSOMES

DNA is the predominant molecule used by all known organisms to store and transmit their genetic blueprint from generation to generation. The primary storage unit for DNA is the chromosome, which exist in an extraordinary diversity of forms varying in size (number of base pairs), topology (circular or linear), and the

arrangement of functional elements they encode. Perhaps the clearest example of the variation within chromosomes is illustrated by the differences between bacterial and eukaryotic genomes.

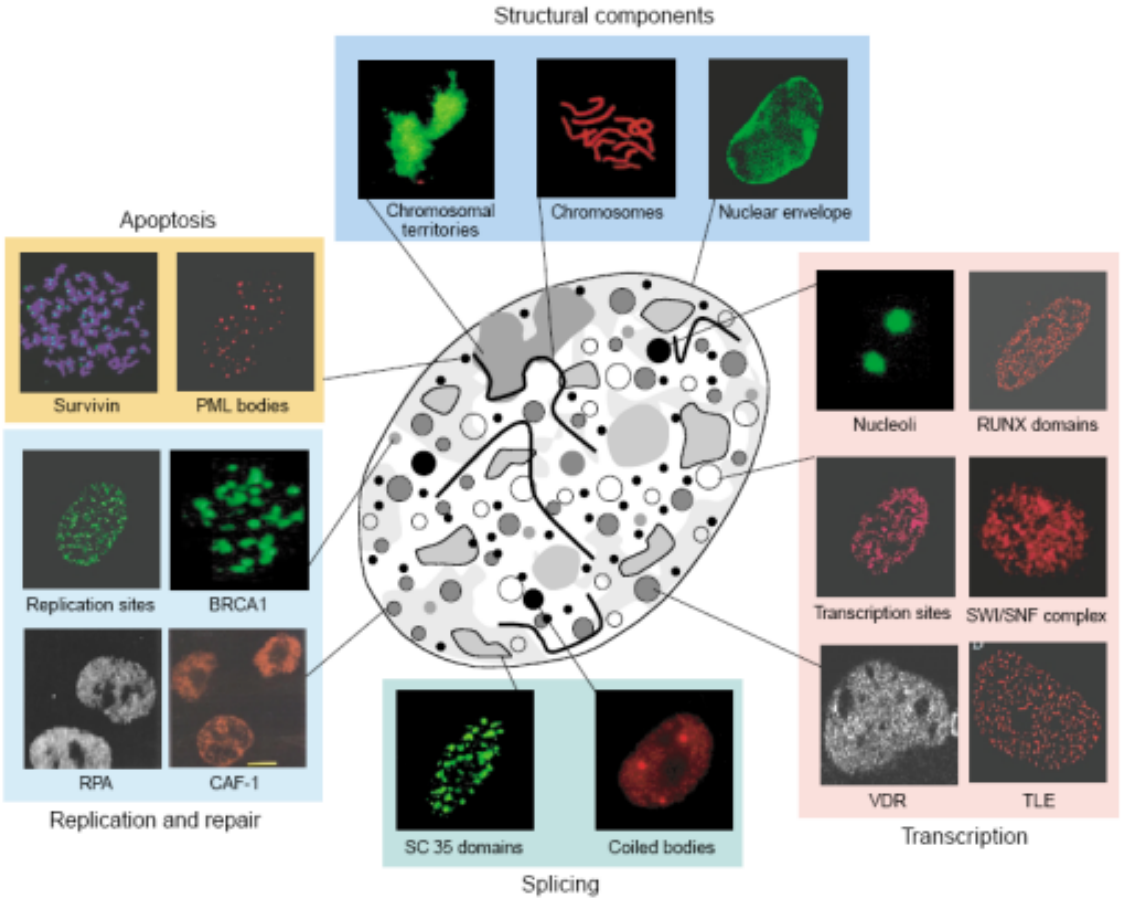


Figure 1.1. Nuclear architecture is functionally linked to the organization and sorting of regulatory information.

Immunofluorescence microscopy of the nucleus *in situ* has revealed the distinct non-overlapping subnuclear distribution of vital nuclear processes, including: DNA replication sites and proteins involved in replication, such as chromatin assembly factor-1 (CAF-1) and replication protein A (RPA); DNA damage as shown by BRCA1; chromatin remodeling (*e.g.* mediated by the SWI/SNF complex); structural parameters of the nucleus (*e.g.* the nuclear envelope, chromosomes and chromosomal territories); RUNX, transducin-like enhancer (TLE) and vitamin D3 receptor (VDR) domains for chromatin organization and transcriptional control of tissue-specific genes; RNA synthesis and processing, involving, for example, transcription sites; SC35 domains, coiled bodies and nucleoli; Subnuclear promyelocytic leukemia (PML) bodies; as well as proteins involved in cell survival (*e.g.* survivin). Adapted from (G. S. Stein et al., 2003).

The classical view, which arose from pioneering studies in *Escherichia coli*, was and still generally is that bacterial (prokaryotic) genomes consist of single, small (1-10Mb), circular DNA molecules, (Casjens, 1998; Hinnebusch & Tilly, 1993; Ishikawa & Naito, 1999). However, prokaryotes have since been found to contain genomes with a variety of forms, including multiple circular or linear chromosomes

and even combinations of the two. For example, *Borrelia burgdorferi*, the causative agent of Lyme disease, has 17 linear chromosomes and numerous plasmids (Hinnebusch & Tilly, 1993). The small size of bacterial genomes cannot be solely attributed to them containing fewer genes compared to eukaryotes as they can contain similar numbers of protein coding genes (*e.g. E. coli* has 4,288 genes and *S. pombe* has 4,970) (Blattner et al., 1997; Wood et al., 2002). Instead the reduced size of bacterial genomes is largely considered to result from the compact linear arrangement of genes and regulatory sequences, with little intergenic, non-coding regions (Casjens, 1998). Moreover, coding regions (genes) are typically not disrupted by introns and co-regulated genes are often clustered into operons within the linear sequence (Lawrence, 2002). The compact linear arrangement of bacterial genomes makes them highly proficient for the rapid proliferation and adaptation required for the survival in the niches they occupy.

By contrast, eukaryotic genomes are generally much larger (>10Mb) and have a predominantly linear arrangement (Ishikawa & Naito, 1999; Nosek, Kosa, & Tomaska, 2006). Although circular chromosomes have been reported in the Budding and Fission yeasts, they are mitotically and meiotically unstable (Fan, Rochet, Gaillardin, & Smith, 1992; Greenfeder & Newlon, 1992; Haber, Thorburn, & Rogers, 1984; Niwa & Yanagida, 1985). The large size of eukaryotic genomes is generally attributed to the extensive regions of non-coding and repetitive DNA found in their chromosomes (Jurka, Kapitonov, Kohany, & Jurka, 2007; Schueler & Sullivan, 2006; Wood et al., 2002). Initially these non-coding regions were thought to be non-functional and a type of ‘junk DNA’, however, they are becoming increasingly recognized as possessing regulatory functions (Birney et al., 2007; Doolittle, 2013; Weinstock, 2007). In addition, unlike bacterial genes, many eukaryotic genes are interrupted by introns and are not arranged into operons (Jurka et al., 2007; Schueler & Sullivan, 2006; Wood et al., 2002). However, the linear clustering of co-ordinately transcribed genes into regions of increased gene expression (RIDGES), has been observed (Caron et al., 2001; Lercher, Urrutia, & Hurst, 2002; Versteeg et al., 2003). Thus it is clear that prokaryotic and eukaryotic cells have adopted different ways to arrange the linear sequences of their chromosome(s).

It should be noted that not all the genetic material in a eukaryotic cell is contained within the nucleus. Ancient endosymbiosis events, whereby a primitive eukaryotic cell is thought to have engulfed a bacterium, have given rise to

intracellular organelles, the most prominent of which are the mitochondrion and the chloroplast (Kutschera & Niklas, 2005). These organelles possess their own genomes that encode a number of proteins essential for their function. Interestingly, these genomes are generally considered to be circular and have a compact linear gene arrangement, akin to that of bacteria. This is widely considered to be critical evidence in the argument that they are of bacterial origin (Anderson et al., 1981; Bullerwell, Leigh, Forget, & Lang, 2003; Chiron et al., 2007).

1.2.1.1 DNA REPLICATION OF CIRCULAR AND LINEAR CHROMOSOMES

DNA replication, the duplication of the genome prior to cell division, is central to cell proliferation. Chromosome replication tends to occur at set times during the cell cycle requiring the temporal regulation of the initiation of DNA replication and chromosome segregation (Edenberg & Huberman, 1975; Huberman & Riggs, 1966; D. Jackson, Wang, & Rudner, 2012; Nosek et al., 2006). The organization of eukaryotic genomes into multiple linear chromosomes results in four major challenges related to the regulation of chromosome replication and segregation, not encountered in circular bacterial genomes, which must be resolved. First, the fragmentation of genomes into separate chromosomes requires that they are replicated together during the Synthesis (S) phase of the cell cycle. Second, the rate of replication fork movement is impeded by the higher-order folding of chromatin that exists in eukaryotic cells (*e.g.* ~1.5 kb/min compared to ~40 kb/min in bacteria). Third, many eukaryotes contain epigenetic information that must be selectively duplicated along with replication in order to be transmitted to daughter cells. Finally, chromosome ends appear like double stranded breaks and, therefore, they must be protected, and in some cases extended, during DNA replication to insure chromosome stability and faithful segregation (Djupedal & Ekwall, 2009; Ishii et al., 2008; D. Jackson et al., 2012; Provost et al., 2002; T. Volpe et al., 2003). The spatial and temporal organization of DNA replication is essential for cell survival, with the latter especially important in eukaryotes, due to the required coordination of replication among the myriad of chromosomal origins.

The replication of circular bacterial chromosomes has been characterized in the greatest detail in *E. coli*. DNA replication initiates at a single unique origin of replication (*oriC*) and proceeds bidirectionally around the chromosome to a site opposite *oriC* called the replication terminus region (*ter*). The *E. coli oriC* contains

a conserved AT-rich region and multiple DnaA boxes, which are 9-nucleotide long sequences specifically bound by DnaA (Katayama, Ozaki, Keyamura, & Fujimitsu, 2010; Messer, 2002). The control of replication initiation is linked to the energy status of the cell through the replication initiator protein DnaA. This is achieved by the accumulation of DnaA bound to ATP (ATP-DnaA) above a critical threshold triggering the initiation of DNA replication (Kaguni, 2006; Katayama et al., 2010; Leonard & Grimwade, 2011). In *E. coli* the chromosome is thought to segregate concurrently with replication (Elmore, Müller, Vischer, Odijk, & Woldringh, 2005; Y. Li, Sergueev, & Austin, 2002; X. Wang, Possoz, & Sherratt, 2005). However, evidence also exists for a cohesion dependent mechanism whereby replicated chromosomes remain associated for a short period of time before segregation, similar to eukaryotes (Bates & Kleckner, 2005; Sunako, Onogi, & Hiraga, 2001). The consecutive nature of replication and segregation of bacterial chromosomes requires their simultaneous and coordinated regulation in space but not so much through time.

Unlike bacteria, in eukaryotes the events of DNA replication and segregation occur in distinct cell cycle phases, *i.e.* synthesis (S) phase and metaphase (M phase), respectively, separated by growth phases (Gap 1 (G1) and Gap 2 (G2) phases). The reduced DNA replication rate and presence of multiple chromosomes in eukaryotes necessitates the existence of many origins of replication. Aside from origins in the unicellular eukaryote *Saccharomyces cerevisiae*, in which a number of 10-15 bp sequences spread across ~150 bp appears sufficient to produce an active origin, eukaryotic origins tend to be more complex than those of bacteria (Bell & Dutta, 2002; Bell, 1995). For example, in the Fission yeast (*Schizosaccharomyces pombe*) origins consist of more degenerate 20-50 bp AT-rich sequences spread over at least 800 bp of sequence, while metazoan origins are even less well defined and can spread over thousands of base pairs of sequence (Bell & Dutta, 2002; Bell, 1995; Bielinsky & Gerbi, 2001; Clyne & Kelly, 1995; Dubey, Kim, Todorov, & Huberman, 1996; Dubey, Zhu, Carlson, Sharma, & Huberman, 1994). Furthermore, the initiation of DNA replication must be coordinated across many origins during S phase of the cell cycle. The initiation of DNA replication across the myriads of origins in eukaryotes is precisely controlled, only occurring at a small number of potential origins, with clusters of adjacent replicores (typically 4-10) firing at very similar times (Edenberg & Huberman, 1975; Huberman & Riggs, 1966; D. Jackson et al., 2012). Following duplication, sister chromatids are held together by the cohesion complex through the G2 and M

phases of the cell cycle until separation late in M phase (Farcas, Uluocak, Helmhart, & Nasmyth, 2011; Hakimi et al., 2002; Lengronne et al., 2004). The separation of chromosome replication and segregation into discreet phases of the cell cycle in eukaryotes and the required coordinated firing across many origins relies on the coordinated regulation of these processes in space and through time.

One of the consequences of the differences between bacterial and eukaryotic DNA replication is that for bacterial chromosomes the time at which a particular locus is replicated depends predominantly on its distance from the origin of replication. By contrast, the precise and dynamic regulation of DNA replication firing in eukaryotic chromosomes means that different regions of the genome will be replicated at different times according to their proximity to, and the firing efficiency of, the replication origin.

1.2.2 The second level of genome organization: Supercoiling and protein binding are required for the compaction of chromosomes

1.2.2.1 Supercoiling is the primary mechanism for the secondary level organization of bacterial genomes

Bacteria, like all organisms, must package their genome into a small cellular volume while enabling factors to access the DNA to perform vital cell processes such as DNA replication and gene expression (Ishihama, 2009). For example, the *E. coli* chromosome has a contour length, the length at maximum physically possible extension, of ~1,600 μm and is compacted ~1000-fold within a rod-shaped cell only ~1 μm in diameter and ~2-4 μm in length (Reyes-Lamothe, Wang, & Sherratt, 2008; Zimmerman, 2006). A substantial level of bacterial genome compaction is achieved by the introduction of negative supercoiling and binding of proteins that can isolate topological domains and bend DNA (Azam & Ishihama, 1999; Dame, 2005; D. Jackson et al., 2012; Reyes-Lamothe et al., 2008; Thanbichler, Wang, & Shapiro, 2005).

The principle mechanism by which bacterial chromosomes are condensed is through DNA supercoiling (D. Jackson et al., 2012; Reyes-Lamothe et al., 2008). The chromosomes of most bacteria are negatively supercoiled, with the mechanisms of regulation and influences on cellular processes best understood in *E. coli* and *Salmonella enterica*. Negative supercoils are introduced into the bacterial chromosomes by the ATP dependent gyrase enzymes (*gyrA* and *gyrB* genes) and relaxed by topoisomerase I (*topA* gene), IV (*parC* and *parE* genes),

and III (*topB* gens) (Rovinskiy, Agbleke, Chesnokova, Pang, & Higgins, 2012; Zechiedrich et al., 2000). Thus, the global supercoiled state of DNA within a bacterial cell is set by the relative abundance and activities of these counteracting enzymes (Snoep, Van Der Weijden, Andersen, Westerhoff, & Jensen, 2002). Evidence suggests that the introduction of negative supercoiling gives rise to the smallest unit of bacterial chromosome organization through the formation of non-constrained, independent supercoiled domains (Postow, Hardy, Arsuaga, & Cozzarelli, 2004). In *E. coli*, these supercoiled domains have been estimated to have an average size of ~10Kb (Postow et al., 2004; R. A. Stein, Deng, & Higgins, 2005).

The negative supercoiling of bacterial chromosomes makes a significant contribution to the compaction of bacterial nucleoids, but it does not account for the total level of compaction observed (Dame, 2005; Dillon & Dorman, 2010; Postow et al., 2004; Thanbichler et al., 2005; Zimmerman, 2006). Unlike eukaryotes, bacteria do not contain histone proteins. However, a number of bacterial Nucleoid Associated Proteins (NAPs) have been identified that are thought to behave as histone homologues. These NAPs exhibit varying degrees of DNA binding, bending, looping and dimerization properties *in vitro* (Azam & Ishihama, 1999; Dame, 2005; Dorman, 2013; Thanbichler et al., 2005). Four of the classical NAPs have been investigated in great detail: the heat-stable nucleoid-structuring protein (H-NS), the heat-stable protein (HU), the factor for inversion stimulation (Fis), and the integration host factor (IHF). The DNA binding and bending properties of these proteins are thought to function in place of eukaryotic histones playing a role in the isolation of topological domains and compaction of the nucleoid (Dorman, 2013). However, studies also indicate that *in vivo* the role of the NAPs could be more in the regulation of cellular processes, such as gene expression, rather than architectural (Dame, 2005; Grainger, Hurd, Goldberg, & Busby, 2006).

In addition, the recently characterized non-classical NAPs (*i.e.* SeqA, SlmA, and MatP) that exhibit macrodomain-specific DNA binding properties (reviewed in (Dame, Kalmykowa, & Grainger, 2011)), may represent alternative candidates that facilitate the secondary and tertiary level organization of bacterial nucleoids (see section 1.2.1.1 for further discussion).

1.2.2.2 EFFICIENT PACKAGING OF EUKARYOTIC GENOMES IS PRIMARILY ACCOMPLISHED BY HISTONE PROTEINS

In eukaryotes, supercoiling makes a smaller contribution to chromosome compaction, instead nuclear DNA is associated with numerous proteins in a complex called chromatin. The basic building block of chromatin is the nucleosome (Olins & Olins, 1974; C. L. F. Woodcock, Safer, & Stanchfield, 1976; C. L. F. Woodcock, Sweetman, & Frado, 1976). Nucleosomes are composed of two copies each of the four core histone proteins (H2A, H2B, H3 and H4). The C-terminal two thirds of each histone protein come together to form a hydrophobic protein octamer core, around which approximately 147 bp of DNA is wound (Luger, Mäder, Richmond, Sargent, & Richmond, 1997) (Figure 1.2A). In most eukaryotes a fifth histone, histone H1, associates with the linker DNA providing partial protection from nuclease digestion for ~20 bp of DNA (Allan, Cowling, Harborne, Cattini, & Gould, 1981; Happel & Doenecke, 2009; Christopher L Woodcock & Ghosh, 2010). Together, the wrapping of DNA around nucleosomes and association of the linker histone H1 contribute significantly to the compaction of eukaryotic chromosomes.

The reversible, post-translational modification of histone proteins further influences the level of chromosome compaction and contributes to the epigenetic code. The N-terminal tails of the core histone proteins protrude out from the octamer making them the most accessible to post-translational modification (Figure 1.2A). Amino acid residues in the N-terminal histone tails and the nucleosome core can undergo specific and variable post-translational modification, including methylation, acetylation and phosphorylation (Lennartsson & Ekwall, 2009; Strahl & Allis, 2000). These post-translational modifications form part of a ‘histone code’ that can be interpreted by other proteins to bring about specific downstream events. This histone code can be transient or stable. In the latter case, if the modifications are heritable, they constitute a true ‘epigenetic code’ (Lennartsson & Ekwall, 2009; Turner, 2000). Not only does the post-translational modification of nucleosomes influence the accessibility and interpretation of the DNA, the dynamic positioning of nucleosomes along chromosomes can also affect cellular possesses such as transcription (Bai & Morozov, 2010; Dai et al., 2009; Schones et al., 2008; Wu et al., 2013).

The contribution that nucleosomes make to eukaryotic genome compaction is two-fold. In addition to the winding of DNA around the nucleosome, interactions between neighbouring nucleosomes facilitate the further compaction of chromatin,

giving rise to the debated 30-nm fibre (Dehghani, Dellaire, & Bazett-Jones, 2005; Maeshima, Hihara, & Eltsov, 2010; Tremethick, 2007; van Holde & Zlatanova, 2007). Early studies of native chromatin using electron microscopy (EM) led to the proposal of two models for the 30-nm chromatin fibre: 1) the one-start helix/solenoid model (Figure 1.2B and D), and 2) the two-start helix model (Figure 1.2C and E). The former involves interactions between adjacent nucleosomes that are connected by linker DNA that is bent to follow a superhelical path. This results in 6-8 nucleosomes, encompassing approximately 1,200 bp of DNA, per turn of the solenoid (G. Li & Reinberg, 2011; Widom & Klug, 1985; C L Woodcock & Dimitrov, 2001). By contrast, in the two-start helix model, adjacent nucleosomes are connected by straight linker DNA. The two-start model was proposed based on the interpretation of EM experiments of chromatin in low ionic strength buffers that appeared to have a Zig-Zag like nucleosome arrangement (G. Li & Reinberg, 2011; S. P. Williams et al., 1986; C L Woodcock & Dimitrov, 2001). The different structures observed for the 30-nm fibre may in part be due to the majority of studies on chromatin structure having been done *in vitro* using isolated chromosomes. Conducting molecular level studies on chromatin structure *in vivo* may shed light on this debate.

1.2.1 The tertiary level of genome organization: Genomes in space

The primary and secondary level organization of bacterial and eukaryotic genomes is relatively well understood in comparison to their spatial organization *in vivo* and the molecular mechanisms that underlie the formation of spatial genome organization. This is particularly true in bacteria, where our understanding of the spatial organization of the nucleoid and its relation to cellular functions is limited. The limited ability to visualize the spatial organization of the nucleoid is partly owing to the small cell size of bacterial cells and the lack of appropriate tools for visualizing the chromosome *in vivo* (Ishihama, 2009; Sherratt, 2003; Woldringh & Nanninga, 2006).

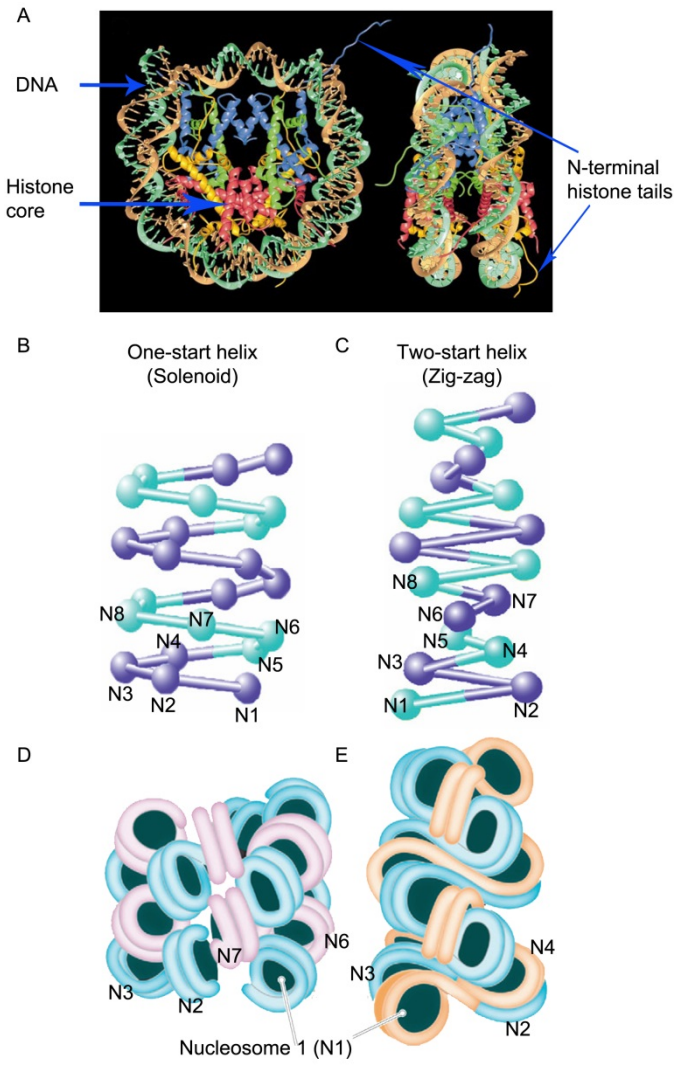


Figure 1.2. DNA in eukaryotic cells is wound around nucleosomes that interact to form a chromatin fibre.

A) Side and front views of the nucleosome octamer with associated DNA and protruding N-terminal tails modelled from crystallography data at a resolution of 2.8 angstrom. B-E) Display two current models of the 30-nm chromatin fibre. Two well-known structural models for 30-nm chromatin fibres are proposed: the one-start helix (solenoid) (B and D) and two-start helix (zig-zag) (C and E). At the top, a schematic representation is shown for the two different topologies of the 30-nm chromatin fibres (B and C). The substitute nucleosomes are numbered from N1 to N8. In the solenoid model proposed by Rhodes and colleagues, the 30-nm chromatin fibre is an inter-digitated one-start helix in which a nucleosome in the fibre interacts with its fifth and sixth neighbour nucleosomes (Robinson & Rhodes, 2006). Alternative helical gyres are coloured blue and magenta (D). In the zig-zag model suggested by Richmond and colleagues, the chromatin fibre is a two-start helix in which nucleosomes are arranged in a zig-zag manner such that a nucleosome in the fibre binds to the second neighbour nucleosome (Dorigo et al., 2004; Schalch, Duda, Sargent, & Richmond, 2005). Alternative nucleosomes pairs are colored blue and orange. Adapted from (G. Li & Reinberg, 2011; Luger et al., 1997).

1.2.1.2 THE SPATIAL ORGANIZATION OF THE BACTERIAL NUCLEOID

Despite the difficulties with investigating bacterial chromosomes, their gross level spatial organization has been recognized for some time. For example, bacterial chromosomes generally occupy a region at the centre of the cell that together with proteins and RNA forms a structure called the nucleoid. Further, the nucleoid appears to be highly organized to facilitate the control of essential cellular processes such as transcription and DNA replication (Cabrera & Jin, 2006; Jin et al., 2012; Marshak, 1951; Mercier et al., 2008; Robinow & Kellenberger, 1994).

Early microscopy studies revealed that bacterial chromosomes typically occupy a region at the centre of the cell ~1/4th its volume and appeared to be organized into a rosette structure formed of topological domains (Marshak, 1951; Robinow & Kellenberger, 1994). More recently, the use of fluorescence *in situ* hybridization and fluorescent repressor-operator systems to visualize individual genomic loci has allowed the spatial position of specific genomic loci to be observed in fixed and living bacterial cells during DNA replication and segregation (Espéli & Boccard, 2006; Gitai, Thanbichler, & Shapiro, 2005; X. Liu, Wang, Reyes-Lamothe, & Sherratt, 2010; Reyes-Lamothe et al., 2008). These studies have revealed that DNA replication and segregation occur concurrently in a highly organized manner that results in a Left-Right-Left-Right replicore arrangement of the two segregated chromosomes (X. Liu et al., 2010).

The use of molecular and recombination based methodologies has identified the existence of macrodomains within the *Escherichia coli* nucleoid. Four structured macrodomains ranging in size from ~0.5 to 1 Mb have been identified; Origin (Ori), Terminus (Ter), Left (L), and Right (R). These domains are defined by the occurrence of preferential intra-domain recombination between lambda att sites when compared to inter-domain recombination. Two non-structured (NS) domains that flank the Ori domain have also been characterized (Boccard, Esnault, & Valens, 2005; Espeli, Mercier, & Boccard, 2008; Ishihama, 2009; Mercier et al., 2008; Niki, Yamaichi, & Hiraga, 2000; Thiel, Valens, Vallet-Gely, Espéli, & Boccard, 2012; Valens, Penaud, Rossignol, Cornet, & Boccard, 2004). Together, these results reveal that bacterial nucleoids are highly organized in space; however, they do not provide insight into the mechanisms whereby this organization is established and regulated.

NAPs, along with their role in nucleoid compaction (see section 1.2.2.1), may represent additional candidates for the spatial organization of bacterial genomes. The classical NAPs: H-NS, HU, Fis, and IHF, are able to bind, bend and form loops in DNA. In fact, early evidence that looped DNA structures might be formed as part of a regulatory mechanism for the control of transcriptional activation, came from studies in bacteria on the ara operon (Dunn, Hahn, Ogden, & Schleif, 1984; Hahn, Dunn, & Schleif, 1984; Hahn, Hendrickson, & Schleif, 1986; K. S. Matthews, 1992). In addition to their role in isolating topological domains and regulating gene expression (mentioned above, see section 1.2.2.1), the classical NAPs (H-NS, HU, Fis, and IHF) have been implicated in loop formation and higher order nucleoid organization (Dame, 2005; Dorman, 2013; Grainger et al., 2006; McGovern, Higgins, Chiz, & Jaworski, 1994; W. Wang, Li, Chen, Xie, & Zhuang, 2011).

H-NS has been extensively show to form a bridging structure following binding to DNA *i.e.* DNA-H-NS-DNA bridges (Dorman & Kane, 2009; Wiggins, Dame, Noom, & Wuite, 2009). H-NS binds to genes to repress their transcription and its ability to dimerize with other H-NS proteins, results in the spatial co-localization of linearly distant H-NS repressed genes. Consistent with this idea, H-NS has been visualized to exist in a small number of highly concentrated foci within the nucleoid (W. Wang et al., 2011). Moreover, genes that are repressed by H-NS tend to co-localize with H-NS focal points (W. Wang et al., 2011). The overexpression of H-NS has also been shown to result in nucleoid condensation and the global reduction in transcription (Dame, 2005; McGovern et al., 1994; Smyth et al., 2000; Spurio et al., 1992; W. Wang et al., 2011).Collectively, these results support a role for H-NS in the compaction on the nucleoid and the coordination of the DNA in each of the two replichores via the formation of specific H-NS foci (W. Wang et al., 2011).

Fis and IHF also exhibit DNA-bending properties and are able to bind many DNA targets within the *E. coli* chromosome. However, H-NS, Fis, and IHF are found to bind to DNA sites within all macrodomains *i.e.* they do not possess any macrodomain specific binding properties (Grainger et al., 2006). This suggests that NAPs might contribute to nucleoid organization at a different level to that of the macrodomain specific proteins, perhaps at the more localized level of plectonemic supercoil microdomains.

NAPs are expressed at high abundance during specific growth phases (Ohniwa, Ushijima, Saito, & Morikawa, 2011). Due to the high abundance of NAPs and their wide-spread DNA binding pattern, it is not unexpected that they also influence transcription levels of target genes. The majority of binding targets for H-NS, IHF and Fis were located in intergenic, potentially regulatory regions, of the *E. coli* chromosome (Grainger et al., 2006). Moreover, they are known to act like transcription factors at specific growth phase or sets of genes (Dorman, 2013; Luijsterburg, Noom, Wuite, & Dame, 2006). Overall, *in vivo* studies suggest the role of NAPs is to influence patterns of global genome expression, thus these proteins may function in more of a regulatory role within the nucleoid structure (Dame, 2005; Grainger et al., 2006; and see chapter 2).

The recently characterized non-classical NAPs (*i.e.* SeqA, synthetic lethal with a defective Min system (SlmA), and MatP) exhibit macrodomain-specific DNA binding properties across the *E. coli* chromosome. These non-classical NAPs may represent additional candidates for the spatial organization of bacterial genomes (Dame et al., 2011).

MatP is a small DNA-binding protein that is associated exclusively with the Ter domain (Espéli et al., 2012; Mercier et al., 2008). It binds to a signature motif *matS* repeated 23 times within the Ter region (Mercier et al., 2008). Binding of MatP to the Ter region is an essential component in coordinating chromosome segregation and cell division; cells deficient in MatP display an anucleate phenotype and there is no extended colocalization of the replicated Ter domains (Dame et al., 2011). Moreover, in the absence of MatP, the Ter domain becomes more mobile and displays a lower degree of compaction. The exclusive relationship that MatP has with the Ter domain reflects the importance of MatP in processes such as preventing premature daughter chromosome segregation during cell division by linking the Ter macrodomains from the replicated chromosomes (Thiel et al., 2012). Collectively this evidence indicates that MatP plays an important role in Ter domain organization; however the precise molecular mechanism of MatP action is yet to be elucidated.

SeqA was originally identified as the factor responsible for sequestration of chromosome replication origins (Lu, Campbell, Boye, & Kleckner, 1994). Subsequent evidence revealed that SeqA is also involved in the negative regulation of chromosome replication initiation and delays premature separation of

newly replicated chromosomes (Bach, Krekling, & Skarstad, 2003; Sánchez-Romero et al., 2010). SeqA binds to specific hemimethylated GATC sequences which are distributed throughout all newly replicated chromosomal domains excluding the Ter macrodomain, but are most densely concentrated within the Ori macrodomain, specifically at *oriC* where replication is initiated (Bach et al., 2003). It was recently found that SeqA binds to the *E. coli* chromosome as a dimer which can multimerize in a reversible, concentration-dependent manner to form a left handed filament (Sánchez-Romero et al., 2010; Waldminghaus & Skarstad, 2009). This suggests that in addition to playing a regulatory role in replication, SeqA may also play a structural role in linking spatially separated binding sites (see chapter 2).

The SlmA protein was initially identified as a “nucleoid occlusion factor”, it is involved in the positioning and proper assembly of the tubulin-like FtsZ protein and subsequent recruitment of additional septal ring components into a ring structure (the “Z-ring”) at mid-cell prior to cell division (Bernhardt & De Boer, 2005; Cho, McManus, Dove, & Bernhardt, 2011). SlmA binds at 24 defined sites within the genome; the specific DNA sequence SlmA binds to is mainly found in the Ori macrodomain and the flanking unstructured regions (NS-left and NS-right) (Dame et al., 2011; Tonthat et al., 2011). Similar to the distribution of SeqA binding sites, SlmA-binding sites are absent from the Ter macrodomain (Tonthat et al., 2011). The SlmA binding sites are found mostly within coding regions, consistent with observations that SlmA seems to function as a regulator of gene expression (Tonthat et al., 2011). On the other hand, MatP and SeqA have not yet been shown to possess any gene expression regulation activities, only structural properties (Dorman, 2013).

Active cellular processes such as transcription have also been implicated in the spatial organization of bacterial nucleoids. For example, in *E. coli* the spatial co-localization of the rDNA genes into a single or small number of foci in the nucleoid appears to contribute to their high level of transcription during vegetative growth. The seven rDNA genes in *E. coli* are spread out over approximately one-third of the genome surrounding the origin of replication; however, they have been identified to co-localize with high concentrations of actively transcribing RNA polymerase (Cabrera & Jin, 2003; Cagliero & Jin, 2013). The active transcription of the rDNA gene operons condenses the *E. coli* nucleoid and accounts for approximately ~85% of a cells total transcription (Bremer & Dennis, 1996; Cabrera,

Cagliero, Quan, Squires, & Ding, 2009). Further, support that transcription influences nucleoid condensation comes from the observation that the transcriptional response to chemical treatment or entry into stationary phase, correlate with the observed condensation of the nucleoid (Cabrera & Jin, 2003; Cagliero & Jin, 2013; Frenkiel-Krispin et al., 2004). Therefore, the regulation of transcription and other active cellular processes may also contribute to the spatial organization and compaction of bacterial nucleoids.

The spatial organization of the bacterial nucleoids has been proposed to be determined by a balance between expansion and compaction forces (Woldringh, Jensen, & Westerhoff, 1995). The compaction forces include DNA supercoiling (Sawitzke & Austin, 2000; Stuger et al., 2002; Travers & Muskhelishvili, 2005), DNA binding proteins (Dame, 2005; Luijsterburg et al., 2006), macromolecular crowding (Murphy & Zimmerman, 2001; Odijk, 1998; Zimmerman & Murphy, 1996), and entropy-driven depletion attraction (Marenduzzo, Micheletti, & Cook, 2006). These compaction forces are opposed by expansion forces such as transertion (the simultaneous transcription, translation and insertion of proteins into the membrane) and transcription (Binenbaum, Parola, Zaritsky, & Fishov, 1999; Dworkin & Losick, 2002; Kruse et al., 2006; Woldringh, 2002). According to this proposition, inhibition of both translation and transcription would disrupt transertion and lead to nucleoid compaction. In support of this concept, it is consistently reported that chloramphenicol, a translation inhibitor, induces nucleoid compaction (Van Helvoort, Kool, & Woldringh, 1996; Zimmerman, 2002). However, the inhibition of transcription by rifampin has been reported to both induce nucleoid expansion (Cabrera & Jin, 2003; Dworsky & Schaechter, 1973; Pettijohn & Hecht, 1974; Q. Sun & Margolin, 2004) and compaction (Binenbaum et al., 1999; Van Helvoort, Huls, Vischer, & Woldringh, 1998; Zimmerman & Murphy, 2001; Zusman, Carbonell, & Haga, 1973). Thus, the roles that transcription and translation have in the organization of bacterial nucleoids remains poorly understood (Saier Jr., 2008).

1.2.1.3 THE SPATIAL ORGANIZATION OF EUKARYOTIC GENOMES

The larger size of eukaryotic cells and their genomes compared to bacteria has made it easier to visualize the spatial organization of chromatin microscopically. This has led to considerable insight into the gross spatial organization of eukaryotic genomes having been gained by the use of microscopy, together with

various tagging methods (*e.g.* FISH and Immunofluorescence), to visualize the organization of chromosomes and the localization of protein factors (Figure 1.1).

At the level of the nucleosome, the regulated post-translational modification of histone proteins can vary the condensation state of the chromatin fibre, affecting the accessibility to the genetic code. Together with the association of additional proteins (*e.g.* the family of Heterochromatin Protein 1 (HP1) proteins, except for in *S. cerevisiae*), post-translational histone modifications change the condensation state of the chromatin fibre (Gelato & Fischle, 2008; Hall et al., 2002; Motamedi et al., 2008). This variation in chromatin condensation state is understood to alter the accessibility of DNA to regulatory proteins and consequently further affect cellular processes including gene transcription (Gelato & Fischle, 2008) and recombination (Bisht, Arora, Ahmed, & Singh, 2008).

Chromatin generally exists in two broadly distinguishable forms, heterochromatin and euchromatin. These two forms of chromatin were initially defined in the early 20th century following the observation of chromosomal regions that remained condensed throughout interphase in liverworts cells in comparison to de-condensed regions (Gelato & Fischle, 2008; Heitz, 1929). Heterochromatic and euchromatic regions can now be further distinguished from each other by the different chromatin markers that they contain (*e.g.* histone variants and post-translational modifications) (Campos & Reinberg, 2009; Gelato & Fischle, 2008; Kamakaka & Biggins, 2005). For example, euchromatin is associated with chromatin markers that indicate loosely packed histones (*e.g.* Histone H3 Lysine 4 methylation (H3K4me)) and normally contains a high density of actively transcribed genes. By contrast, heterochromatin is associated with markers that are indicative of tightly condensed histones (*e.g.* Histone H3 Lysine 9 methylation (H3K9me)) and a low frequency of active genes (Gelato & Fischle, 2008).

Eukaryotic genomes are often riddled with repeat sequences that are frequently maintained in a heterochromatic state throughout the cell cycle (Jurka et al., 2007; Schueler & Sullivan, 2006; Wood et al., 2002). The formation of heterochromatin at repetitive regions, such as the telomeres and centromeres, is important to maintain genome integrity and the fidelity of chromosome segregation (Djupedal et al., 2009; Ishii et al., 2008; Provost et al., 2002; T. Volpe et al., 2003). A commonly observed feature of constitutively heterochromatic regions is that they cluster into a number of distinct foci. For example, in *S. pombe* the three centromeres can be

visualized as a single focus that co-localizes with the spindle pole body (microtubule organizing centre) and the telomeres tend to cluster into a number of foci at the nuclear periphery (Alfredsson-Timmins, Henningson, & Bjerling, 2007; Alfredsson-Timmins, Kristell, Henningson, Lyckman, & Bjerling, 2009). In addition to the observed spatial clustering of heterochromatic regions, the co-localization of genes with these regions and clustering of repeat elements have also been detected. In *S. pombe* the pol III transcribed genes; the tRNA and 5S rRNA genes, frequently co-localize with heterochromatin and *Tf2* retrotransposons cluster together in a CENP-B dependent manner. Moreover, the co-localization of genes with heterochromatin and clustering of repeat elements (*e.g. Tf2* retrotransposons) have likely roles in the regulation of gene expression by supressing the expression of nearby genes (Alfredsson-Timmins et al., 2007; Cam, Noma, Ebina, Levin, & Grewal, 2008; Chikashige et al., 1997; Iwasaki, Tanaka, Tanizawa, Grewal, & Noma, 2010).

At a gross level, one of the most generally accepted examples of chromosome organization is that individual chromosomes occupy distinct spatial positions relative to other chromosomes within the nucleus (*i.e.* chromosome territories) (Bolzer et al., 2005; T. Cremer & Cremer, 2001; Gehlen et al., 2012; Lieberman-Aiden et al., 2009; Tanizawa et al., 2010). However, how the chromatin fibre is organized at the molecular level to give rise to chromosome territories while participating in the regulation of cellular processes has not been well established. In part due to the limited resolution of microscopy. Molecular based methods have begun to give insights into the detailed organization of eukaryotic chromosomes, such as the formation of DNA loops between linearly distal regions on the same chromosome (intra-chromosomal) and between different chromosomes (inter-chromosomal).

Chromatin loops between gene promoters and distal regulatory elements (including but not limited to: enhancers, silencers, imprinting control regions and locus control regions), have been established as participating in the regulation of gene transcription (Q. Li, Barkess, & Qian, 2006; Palstra et al., 2003; Sproul, Gilbert, & Bickmore, 2005; Tolhuis, Palstra, Splinter, Grosveld, & De Laat, 2002). For example, the developmentally dependent formation of inter- and intra-chromosomal interactions in naïve CD4+ T-cells between the promoter regions of both the IFN-γ (5 genes) and IL-4 (13 genes) genes and the regulatory T_H2 locus control region on chromosome 11, dictate cell fate (Lee, Spilianakis, & Flavell,

2005; Spilianakis & Flavell, 2004; Spilianakis, Lalioti, Town, Lee, & Flavell, 2005; A. Williams, Spilianakis, & Flavell, 2010). These interactions facilitate the establishment of transcription profiles that promote the differentiation of naïve CD4+ T-cells into either T-helper-cells 1 (T_H1) or T-helper-cells 2 (T_H2). The formation of T_H1 cells requiring the expression of the interferon-γ gene while T_H2 cells require expression of interleukins 4, 5, and 13, which are on mouse chromosomes 10 and 11, respectively (Spilianakis & Flavell, 2004; Spilianakis et al., 2005). Chromosomal interactions, however, do not always stimulate transcription, the resulting effect of the formation of an interaction depends on a number of factors, including the identity of the effector element(s) (*e.g.* enhancer or repressor) and/or the chromatin state of the interacting loci (Simonis et al., 2006; Smallwood & Ren, 2013).

Investigations into how these chromatin loops and interactions are formed have led to the identification of a number of factors that are involved in the establishment and maintenance of specific higher order chromatin structures. Among these is cohesin, a protein complex that forms a ring structure around replicated chromatids during interphase, which has also been implicated in the formation of large chromosomal loops (Cipak, Spirek, & Gregan, 2008; Farcas et al., 2011; Hakimi et al., 2002; Lengronne et al., 2004). These cohesin mediated loops have been postulated to function as barriers between functional and non-functional chromatin domains (Kim, Cecchini, & Kim, 2011). In addition, the insulation factor CTCF, found in higher eukaryotes, binds at sites that overlap cohesin binding sites and help to demarcate between hetero- and eu-chromatin (Handoko et al., 2011). Furthermore, CTCF bound at specific sites in the DNA can bring together strands of DNA, thus forming chromatin loops, and anchors DNA to cellular structures like the nuclear lamina (Guelen et al., 2008). Therefore, it has been proposed that the primary role of CTCF is to regulate the 3D structure of chromatin (Phillips & Corces, 2009). For example, the involvement of CTCF in the formation of specific inter- and intra-chromosomal interactions including those at the β-globin, IFN-γ, and Igf2/H19 loci, indicates that CTCF may be a 'master genome weaver' in mammalian cells (Botta, Haider, Leung, Lio, & Mozziconacci, 2010; Handoko et al., 2011; Kurukuti et al., 2006; Phillips & Corces, 2009; Sekimata et al., 2009; Splinter et al., 2006). Even though CTCF is not found in lower eukaryotes, other potential insulator factors involved in the demarcation of active from inactive chromatin domains have been identified, such as transcription factor III (TFIIIC) in the Budding yeast (Valenzuela, Dhillon, & Kamakaka, 2009).

An emerging theme is that the regulation of transcription plays a significant role in the spatial organization of eukaryotic genomes (Bartlett et al., 2006; Carter et al., 2008; Eskiw et al., 2010; S. Martin & Pombo, 2003; Osborne et al., 2004; Papantonis & Cook, 2013; Xu & Cook, 2008). A clear illustration of this is the observed accumulation of actively transcribing versions of RNA-polymerase II (pol II) into highly concentrated foci; termed transcription factories (Bartlett et al., 2006; Carter et al., 2008; Eskiw et al., 2010; Mitchell & Fraser, 2008; Papantonis & Cook, 2013). The number of observable transcription factories is far fewer than that of actively transcribed genes. This has led to the finding that many actively transcribed loci co-localize in the vicinity of transcription factories (S. Martin & Pombo, 2003; Mitchell & Fraser, 2008). As a result, the regulated co-localization of genes with transcription factories, and maybe even the length of time that they remain associated, has been postulated to facilitate the regulation of transcription. Consistent with this idea, genes that have high transcript levels have been found to preferentially co-localize with each other, perhaps reflecting the co-occupancy of transcription factories (Papantonis & Cook, 2013; Tanizawa et al., 2010). The enrichment of certain regulatory factors at transcription factories has also been observed and suggests that different factories maybe specialize in the transcription of specific groups of genes (Bartlett et al., 2006; S. Martin & Pombo, 2003; Xu & Cook, 2008).

Increasing evidence points towards transcription factors themselves having a role in the establishment of three-dimensional genome organization to facilitate proficient, coordinated, tissue specific transcription patterns (de Wit et al., 2013; Jing et al., 2008; Mastrangelo, Courey, Wall, Jackson, & Hough, 1991; Nolis et al., 2009; W. Zhao et al., 2011). For example, genomic clusters of sites bound by the Oct4 and Nanog pluripotency factors were found to have a high propensity to associate with each other in the nuclear space. The association of the Oct4 and Nanog binding sites was specific to the pluripotency state and dependent on the presence of these factors (de Wit et al., 2013). At a more local level, the GATA transcription factor in mammals has been shown to establish loops between enhancer elements and the promoter of the proto-oncogene *Kit* locus in a development, transcription, and DNA methylation dependent manner (Jing et al., 2008). Collectively, these results support a role for transcription, and the regulation thereof, in shaping the spatial organization of eukaryotic and likely also bacterial genomes.

1.2.1.3.1 METAPHASE CHROMOSOMES

One of the classical depictions of a chromosome is that of the highly condensed, X shaped metaphase chromosome. The structure of isolated metaphase chromosomes has been studied extensively using microscopy and enzymatic digestion experiments (Matsuda et al., 2010; Ohta, Wood, Bukowski-Wills, Rappsilber, & Earnshaw, 2011; Poirier & Marko, 2002). Early hypotheses that postulated how the condensation of metaphase chromosomes was achieved led to the idea that chromatin loops were anchored to a central protein scaffold (Haapala & Nokkala, 1982; Harrison, Allen, Britch, & Harris, 1982). Key studies utilizing protease, RNAse and DNAse digestion together with micromechanical force measurements have since revealed that the integrity of metaphase chromosome structure and their inherent elasticity is largely due to the chromatin fibre itself. It has now been hypothesized that metaphase chromosomes are made up of a continuous chromatin network rather than a protein scaffold, potentially constrained by chromatin-crosslinking elements spaced ~15 Kb apart (Poirier & Marko, 2002). Recently, topoisomerase II activity has been shown to regulate the level of DNA catenation and thus has been implicated in the establishment of condensed metaphase chromosomes and sister chromatid cohesion (SCC) (Bauer, Marie, Rasmussen, Kristensen, & Mir, 2012; Farcas et al., 2011). However, the *in vivo* structure of metaphase chromosomes at the molecular level has not yet been visualized and it therefore remain unknown whether the chromatin fibre is folded in a regular fashion or exists more as a dynamic ‘polymer melt’ (Maeshima et al., 2010).

Despite the apparent exclusion of condensed metaphase chromosomes from interacting with each other at the gross level they have been shown to remain interconnected. However, the level of interconnectedness remains unclear. Early observations of isolated metaphase chromosomes provided evidence that there are connections between metaphase chromosomes, but these connections were initially presumed to be an artefact of chromosome isolation (Hoskins, 1968; Korf & Diacumakos, 1978, 1980). More recently, the presence of connections between metaphase chromosomes have been confirmed and identified to contain DNA (Kuznetsova et al., 2007; Maniotis, Bojanowski, & Ingber, 1997). For example, threads observed to form between mitotic chromosome in mouse cell lines have been found to include satellite DNA and the CENP-B protein (Kuznetsova et al., 2007).

1.3 GENOMES IN TIME

It is clear that bacterial and eukaryotic genomes are organized at many different hierarchical levels, each of which has the potential to be regulated, resulting in the manifestation of a highly complex, three-dimensional genome organization. However, due to the highly dynamic nature of all these levels of organization as a result of active and passive processes, and due to the continually changing environment, the spatial organization of chromosomes is not static. Quite the contrary, many cellular processes, including the three-dimensional organization of chromosomes, are known to change in response to environmental stimuli and cell cycle progression.

1.3.1 THE RESPONSE TO ENVIRONMENTAL STIMULI

The change in shape of bacterial nucleoids in response to the growth environment was observed early on (Cabrera et al., 2009; Cabrera & Jin, 2003; Cagliero & Jin, 2013; Dworsky & Schaechter, 1973; Jin et al., 2012; Margolin, 2010; Van Helvoort et al., 1996). For example, nucleoids in rapidly growing cells were large in comparison to those of slow growing cells, and were seen to take on other shapes in response to certain antibiotics (Zimmerman, 2006). The most well studied change in nucleoid organization in response to an antibiotic occurs within minutes of treatment with the bacteriostat chloramphenicol, with the nucleoid being converted into very compact, axially localized doublet shapes (Zimmerman, 2003). A more naturally occurring response to an environmental stimulus is the observable, serine hydroxamate (SHX) induced, expansion of the *E. coli* nucleoid (Cashel, Gentry, Hernandez, & Vinella, 1996). SHX artificially induces the amino acid starvation response, called the stringent response, which, in addition to the observed expansion of the nucleoid, results in transcriptional changes and inhibits the initiation of DNA replication (Durfee, Hansen, Zhi, Blattner, & Ding, 2008; Traxler et al., 2008). Even though the mechanism(s) behind the re-structuring of the nucleoid in response to growth and stress is still largely unknown, these observations suggests a relationship between transcription and the organization of the nucleoid (Jin & Cabrera, 2006).

1.3.2 THE RESPONSE TO CELL CYCLE PROGRESSION

In eukaryotes the organization of chromosomes and nuclear processes in space is known to change over time in response to, for example, environmental stimuli and cell cycle progression. The cell cycle is a temporal process that has been

under intensive investigation for many years. As a result, many cellular processes have been found to fluctuate throughout the cell cycle, including gene transcription and chromosome organization. One of the most dramatic changes in chromosome condensation state is observed during metaphase, which coincides with a reduction in gene transcription (Funabiki, Hagan, Uzawa, & Yanagida, 1993; McInerny, 2011; Müller, 1995; Oliva et al., 2005; Orlando et al., 2008; Rustici et al., 2004; Zwicker & Müller, 1995). It remains unresolved whether the spatial organization of chromosomes is transmitted through metaphase or re-established *de novo* upon cells entering the G1 phase (Essers et al., 2005; Gerlich et al., 2003; Walter, Schermelleh, Cremer, Tashiro, & Cremer, 2003). In addition, the role that cell cycle specific changes in genome organization, particularly during phases where large-scale chromatin re-organization occurs (*e.g.* metaphase), have in the establishment of cell cycle specific transcription, is unknown (Thomson, Gilchrist, Bickmore, & Chubb, 2004; Walter et al., 2003). Chromatin based epigenetic mechanisms that influence the condensation level of mitotic chromosomes at specific loci and nuclear factors that remain bound to mitotic chromosomes have been identified. The formation of specific chromatin structures and binding of nuclear factors have been shown to transmit gene regulatory information through the transcriptionally silent mitotic chromosomes. For example, the binding of transcription factors (TFs) at specific genomic sites facilities reestablishment of transcriptional activity during exit from mitosis and entry into the following growths phase, the Gap 1 (G1) phase (Kadauke & Blobel, 2013). The transmission of regulatory information (*e.g.* chromatin state and TF binding) through metaphase that facilitates the post-mitotic re-activation of gene expression, is a concept that has been termed ‘mitotic bookmarking’ (Kadauke & Blobel, 2013; Sarge & Park-Sarge, 2005; R. Zhao, Nakamura, Fu, Lazar, & Spector, 2011).

In *S. cerevisiae*, the inheritance of the spatial position of genes within the nucleus through the cell cycle enhances the rate at which transcription can be induced following an initial burst of transcription (D. G. Brickner et al., 2007; J. H. Brickner, 2009; Tan-Wong, Wijayatilake, & Proudfoot, 2009). For example, the exposure of *S. cerevisiae* cells to galactose induced the transcription of the *GAL1* locus and its re-localization to the nuclear periphery. The previously expressed, peripherally localized forms of *GAL1* are activated much faster in response to galactose re-exposure than the long-term repressed forms of *GAL1* found in the nucleoplasm. Furthermore, the peripheral localization of the *GAL1* locus has been found to be mediated by the presence of specific histone variants *i.e.* H2A.Z.

Interestingly, the peripheral localization of the *GAL1* locus can be inherited through multiple cell cycle generations in the absence of transcription, serving as a form of memory of recent transcriptional activity (D. G. Brickner et al., 2007; J. H. Brickner, 2009; Tan-Wong et al., 2009).

Due to its complexity, it is difficult to define the details of the mammalian cell cycle. As a result, more tractable model organisms have been used to gain a detailed understanding of cell cycle regulation. One organism that has contributed significantly to our understanding of the cell cycle is the Fission yeast, *Schizosaccharomyces pombe*. The Fission yeast has a genome structure and cell cycle similar to that of higher eukaryotes. For example, the linear chromosome arrangement and conserved genomic features, such as constitutive pericentromeric and telomeric heterochromatin, are shared with mammals. The genetic malleability of the Fission yeast genome has enabled the in-depth investigation of how the cell cycle is regulated, which has resulted in the near complete description of cell cycle control in this organism (Coudreuse & Nurse, 2010; Nurse, Thuriaux, & Nasmyth, 1976). Our ability to control the cell cycle has led to the identification of many cellular processes that fluctuate throughout, such as gene transcript levels (McInerny, 2004; Rustici et al., 2004), heterochromatin formation (Kloc, Zaratiegui, Nora, & Martienssen, 2008) and reorganization of telomere, centromere, and mating type loci clustering (Alfredsson-Timmins et al., 2009; Funabiki et al., 1993).

1.4 DNA TRANSFER BETWEEN ORGANELLES

In addition to the requirement for eukaryotic cells to spatially and temporally regulate the organization of their nuclear genomes, they must also coordinate nuclear processes with intra-cellular organelles that contain their own genomes. One of these organelles, the mitochondrion, has a central role in the metabolic systems of most eukaryotic cells (Vellai & Vida, 1999). The mitochondrial organelle is thought to have arisen by way of an ancient endosymbiosis event (Kutschera & Niklas, 2005; Timmis, Ayliff, Huang, & Martin, 2004). As the proposed engulfed bacterial cell evolved into the mitochondrial organelle, most of the genes on its genome were transferred and integrated into the host cell's nuclear chromosomes (Timmis et al., 2004). The transfer of mitochondrial DNA to the nucleus has resulted in a drastic reduction in mitochondrial genome size. As a result, in the Budding and Fission yeasts, as in other organisms, only a small number of the genes essential to the function of the electron transport chain, which is necessary

for respiratory growth, are now encoded by the mitochondrial genome (Foury, Roganti, Lecrenier, & Purnelle, 1998; Lagunas, 1976; Schäfer, 2003; Wood et al., 2002).

The transfer of mitochondrial genomic fragments to the nucleus is not only a historical event. Rather, the ongoing transfer of mitochondrial DNA to the nucleus in various organisms including the Budding yeast, plants, and humans has been detected (Brennicke, Grohmann, Hiesel, Knoop, & Schuster, 1993; C. Y. Huang, Ayliffe, & Timmis, 2003; Liang, 1996; W. Martin, 2003; Ricchetti, Fairhead, & Dujon, 1999; Ricchetti, Tekaia, & Dujon, 2004; Rodley et al., 2012, 2009; Stegemann, Hartmann, Ruf, & Bock, 2003; Thorsness & Fox, 1990, 1993; Yu et al., 2003). Studies have demonstrated that mitochondrial DNA in *S. cerevisiae* is transferred to the nucleus at a rate of $2x10^{-5}$ per cell per generation (Thorsness & Fox, 1990). However, the mechanism of transfer remains unknown, despite having been shown to occur in a nuclear gene dependent manner (Campbell & Thorsness, 1998; Thorsness & Fox, 1993). Furthermore, the functional roles of the mitochondrial regions once in the nuclear compartment have not been characterized.

Additional evidence for the continued transfer of regions of the mitochondrial genome to the nucleus is provided by the identification of NUclear MiTochondrial sequences (*NUMTs*) (Lenglez, Hermand, & Decottignies, 2010; Ricchetti et al., 2004). NUMTs are current mitochondrial genome sequences that are found integrated into nuclear chromosomes. These NUMTs are proposed to have nuclear functions (Blank et al., 2008; Chatre & Ricchetti, 2011; Lenglez et al., 2010; Ricchetti et al., 1999). For example, the identification of NUMT insertions at double-stranded DNA break repair sites in the Budding yeast may implicate mitochondrial DNA fragments in the repair of double-stranded breaks (Ricchetti et al., 1999). NUMTs have also been found to contain autonomously replicating sequence (ARS) motifs sequences and in the Fission yeast preferentially insert near to origins of DNA replication (Blank et al., 2008; Chatre & Ricchetti, 2011; Lenglez et al., 2010). This provides further support that NUMTs have a functional role, implicating them in the initiation of DNA replication. Alternatively, origins of DNA replication may be prone to double-stranded breaks that were repaired by integration of a mitochondrial DNA fragment. Further evidence that mitochondrial DNA promotes DNA replication in *S. cerevisiae* came from experiments overexpressing the mitochondrial maintenance protein Abf2p. The increase in the

amount of mitochondrial DNA due to the stabilization of the mitochondrial genome promoted nuclear DNA replication through reducing the Sir2 mediated deacetylation of specific ARS sites (Blank et al., 2008). These results suggest that both the presence of NUMTs and the ongoing transfer of mitochondrial genome fragments to the nucleus participate in the regulation of nuclear processes.

Loops between distal regulatory regions (*e.g.* enhancers) and contacts between chromosomes are known to occur with promoter regions of genes that the contacts regulate (Spilianakis & Flavell, 2004; Spilianakis et al., 2005; Tolhuis et al., 2002; A. Williams et al., 2010). Therefore, given that regions of the mitochondrial genome transferred to the nucleus were found to be enriched in specific sequences that influence nuclear processes (*i.e.* ARS elements and DNA replication); perhaps the presence of regulatory sequences could be involved in the manipulation of nuclear transcription profiles. It is well known that the successful progression through the cell cycle requires the coordination between the nuclear and mitochondrial genomes (Chatre & Ricchetti, 2013; Chu et al., 2007; Crider et al., 2012; Lebedeva & Gerald S. Shadel, 2007; Mandal, Guptan, Owusu-Ansah, & Banerjee, 2005; Mitra, Wunder, Roysam, Lin, & Lippincott-Schwartz, 2009). For example, studies in *S. cerevisiae* have revealed that the G1 to S phase cell cycle checkpoint is regulated by the nuclear abundance of mitochondrial DNA (Mandal et al., 2005; Mitra et al., 2009). Specifically, the absence of mtDNA in the nucleus appears to trigger the Rad53 DNA damage response checkpoint inhibiting progression from G1 to S phase of the cell cycle (Crider et al., 2012). However, how the mitochondrial fragments once in the nucleus influence cellular processes is not well understood.

Rodley et al., (2009) developed a new method called GCC for the detection of chromosomal contacts (DNA-DNA interactions) on a genome-wide scale. The application of GCC to the study of genome organization in *S. cerevisiae* resulted in the unexpected detection of nucleic acids of mitochondrial origin interacting with nuclear loci (hereinafter referred to as mt-nDNA interactions) (Rodley et al., 2009). These mt-nDNA interactions were found to occur at statistically significant frequencies and change depending on the energetic state of the cells (Rodley et al., 2009). These results were confirmed by the specific analysis of one of these interactions that demonstrated carbon source dependent fluctuations in interaction frequency (Rodley et al., 2009).

In addition to an apparent role in regulating the cell cycle in yeast, the quality and quantity of mitochondrial DNA has been shown to affect patterns of nuclear transcription and replication (Blank et al., 2008; V. Parikh, Morgan, Scott, Clements, & Butow, 1987; V. S. Parikh, Conrad-webb, Docherty, & Butow, 1989). These observations suggest that the formation of specific contacts between regions of the mitochondrial genome that have been transferred to the nucleus and nuclear loci may contribute to the regulation of transcriptional level from nuclear encoded genes. Future work looking at the specific formation of mt-nDNA interactions and their role in regulating gene transcript levels or other nuclear processes such as the initiation of DNA replication should be performed. Furthermore, in addition to the use of *S. cerevisiae* as a model for the investigation of mt-nDNA interactions, the Fission yeast would provide a useful counterpart because it shares many features with higher eukaryotes, including the dependence on mitochondria for survival (Chiron et al., 2007; Coudreuse & Nurse, 2010; Fantes & Nurse, 1978; Nurse et al., 1976; Schäfer, 2003; Weir & Yaffe, 2004).

1.5 Central Questions

The accumulating wealth of knowledge about the packaging and spatial organization of genomes over the past century has extensively expanded our understanding of genome organization. It has become clear that prokaryotic and eukaryotic genomes are organized in three-dimensional space and disrupting it can alter the phenotype. However, there still remain many unanswered questions. Central of these is the link between a particular spatial organization and cellular function. Furthermore, do changes in genome organization through time facilitate the adaptation of a cell to particular environments or situation?

Underpinning the ability to investigate such questions is the development of methodologies that enable the precise interrogation of organism's genomes. In the past decade there has been the advancement in a number of techniques that begin to allow us to investigate these questions in depth. These include the development of super-high resolution microscopy and Chromosome Conformation Capture (3C) technologies (Bystricky, Heun, Gehlen, Langowski, & Gasser, 2004; Dekker, Rippe, Dekker, & Kleckner, 2002; Dostie et al., 2006; Dostie, Zhan, & Dekker, 2007; Gondor, Rougier, & Ohlsson, 2008; Kritikou, 2005; Lieberman-Aiden et al., 2009; Rodley et al., 2009; Sexton et al., 2012). However, each of these methods has its limitations and, therefore, to unravel the spatial and temporal organization of genomes and its role in cellular function necessitates a multifaceted approach.

In the following section I will review the past and present methods that have led to our current understanding of genome organization and that show great promise for advancing our knowledge about the structure – function relationship.

1.6 METHODS FOR THE STUDY OF SPATIAL ORGANIZATION

Early research into the spatial organization of chromosomes using staining and microscopy methods led to the first insights into prokaryotic and eukaryotic three-dimensional genome organization and have since revealed many features of chromosome organization (Aula & Saksela, 1972; T. Cremer & Cremer, 2001; Grond, Derksen, & Brakenhoff, 1982; Kite & Chambers, 1912; Marshak, 1951; Squarzoni, Cinti, Santi, Valmori, & Maraldi, 1994; Van Winkle et al., 1953). Microscopic measurements can be performed on fixed or living cells with both approaches having distinct advantages and disadvantages with respect to experimental limitations, image quality, and range of questions that can be addressed. For example, the use of live cells allows the experimenter to follow the actual movement of a locus in real-time (Berger et al., 2008; Heun, Laroche, Shimada, Furrer, & Gasser, 2001). Of particular importance to the microscopic study of genome organization has been the ability to fluorescently stain specific genomic loci, as it allows the measurement of a locus' position with respect to other loci or to nuclear structures on a single cell level.

The classical techniques of immunofluorescence (IF) and fluorescence *in situ* hybridization (FISH), which are performed on fixed samples, still play an important role in the investigation of nuclear architecture despite the recent development of *in vivo* live cell fluorescence imaging. IF uses fluorescently labelled antibodies against the target molecule or, more commonly, a primary antibody against the target and a labeled secondary antibody which then recognizes the primary antibody. FISH was pioneered by Cremer (C. Cremer, Rappold, Gray, Müller, & Ropers, 1984; T. Cremer, Lichter, Borden, Ward, & Manuelidis, 1988), Pinkel (Pinkel et al., 1988) and Lichter (Lichter et al., 1988) in the 1980's, for the detection of chromosome abnormalities and developmental defects, including trisomies (T. Cremer et al., 1988; Pinkel et al., 1988). It relies upon the hybridization and microscopic detection of fluorescently labelled nucleic acid probes to locate specific DNA sequences or even entire chromosomes within cells or nuclei (Volpi & Bridger, 2008). Since IF and FISH work on fixed samples they generally tolerate higher light intensities and longer exposure times than live cell imaging thus they typically produce higher quality images. The requirement for

long exposure times is also shared by the recently developed sub-diffraction techniques, which include variations of spatially modulated illumination (SMI and SPDM) (Baddeley, Batram, Weiland, Cremer, & Birk, 2007; Hildenbrand et al., 2005; Lemmer et al., 2008), photo-activated localization microscopy (PALM) (Ribeiro et al., 2010), stochastic optical reconstruction microscopy (STORM) (B. Huang, Wang, Bates, & Zhuang, 2008), and coherent diffraction microscopy (Jiang et al., 2010). Therefore, these sub-diffraction techniques also only work on fixed samples. While fixed cells enable the use of higher light intensities and exposure times, the preparation of cells for these techniques involves harsh treatments to allow the access of macromolecules that are used to detect the loci (Meister, Gehlen, Varela, Kalck, & Gasser, 2010). Therefore, great care has to be taken to prevent the appearance of artefacts while preserving as much of the nuclear 3D structure as possible during these treatments.

Live cell imaging techniques using fluorescently tagged repressor/operator combinations were developed to enable *in vivo* observations of loci (Robinett et al., 1996; Straight, Belmont, Robinett, & Murray, 1996). These initial studies have been extended to include detailed investigations into, for example, telomere positioning (*e.g.* (Heun et al., 2001; Schober, Ferreira, Kalck, Gehlen, & Gasser, 2009; X. Wang et al., 2008)), artificial chromosomes (*e.g.* (Levi, Ruan, Plutz, Belmont, & Gratton, 2005; Schober et al., 2008)), and low and high-throughput studies of locus positioning (*e.g.* (Berger et al., 2008; Cabal et al., 2006; Therizols, Duong, Dujon, Zimmer, & Fabre, 2010)). In addition to the obvious advantage that the cells are alive, *in vivo* imaging allows the observation of single cells over time. However, these advantages come at the cost of technical limitations. For example, it is essential to minimize the cellular damage caused by the high intensity illumination. Therefore, light intensity and exposure times are usually kept as low as possible, which limits the quality of the images and the number of images that can be taken. For each experiment, the optimum balance between spatial resolution, time resolution, and image quality has to be identified. In addition, image acquisition of live cells is usually more laborious than that for fixed cells which makes it more difficult to image sufficient numbers of cells.

A distinct advantage of fluorescence microscopy techniques compared to most other methods is their ability to monitor cellular and nuclear architecture at the single cell level. This makes it possible to identify positive or negative correlations between internally and externally driven processes (*e.g.* the cell cycle) (Heun et

al., 2001). These relationships are likely to be averaged out in bulk ensemble measurements. However, the corollary is that even supposed high-throughput microscopic techniques are limited to the observation of a small number of preselected genomic loci in a relatively small number of cells.

1.6.1 Molecular Biology Techniques for the Identification of Chromosome Organization

Reductionist approaches to studying genome organization drove the development of molecular techniques that rely upon proximity-based ligation events between spatially adjacent loci (Cullen, Kladde, & Seyfred, 1993; Mukherjee, Erickson, & Bastia, 1988). Chromosome Conformation Capture (3C; see section 1.6.1.2 and Figure 1.3c) was the first widely applied molecular technique developed specifically for this experimental niche (Dekker et al., 2002). Unfortunately, the a priori requirement to know, or suspect, the identity of the spatially proximal loci is a serious limitation to the 3C technique. Circular Chromosome Conformation Capture (4C; see section 1.6.1.3 and Figure 1.3d) (Simonis et al., 2006; Z. Zhao et al., 2006) and Chromosome Conformation Capture Carbon Copy (5C; see section 1.6.1.4 and Figure 1.3e) (Dostie et al., 2006) were subsequently developed to enable in-depth interrogations of the role of individual regions within the global genome organization. Unlike microscope based technologies, these methods can be applied following standard molecular training and using inexpensive equipment. Therefore, it has not taken long for these proximity-based ligation techniques to gain wide-spread acceptance as rapid and adaptable methodologies for the study of genomic organization at molecular level resolution. The fundamental basis of all of the following techniques is relatively simple. Genomic organization is captured by a combination of fragmentation and ligation prior to the identification of the interacting loci (Figure 1.3a).

1.6.1.1 Cross-Linking

The internal, most commonly protein-protein, linkages that hold DNA fragments together may (Dekker et al., 2002) or may not (Cullen et al., 1993) be stabilized by cross-linking. Cross-links are typically short range and reversible forming between stably bound protein molecules. The most commonly used cross-linking agent is formaldehyde as it joins primary amines through a Schiff's intermediate (Fujita & Wade, 2004; V. Jackson, 1999; Schmiedeberg, Skene, Deaton, & Bird, 2009), resulting in the formation of covalent linkages that are approximately 2 Ångstrom (Å) in length. Alternative cross-linking systems exist (for a brief review and

protocols see (Fujita & Wade, 2004)), but they are not widely employed in the study of genome organization.

There are several attributes of cross-linking that must be taken into account when designing experiments and subsequently interpreting the results. Firstly, it is important to recognize that the size of the cross-link (2 Å, the linear distance covered by the formaldehyde link) does not mean that the interaction is that close to begin with, particularly as hydrogen bonds involving amines occur at distances ranging from 1.2 Å to 4 Å (Gilli & Gilli, 2009). Secondly, cross-linking agents can covalently stabilize large complexes which physically link loci that remain separated by relatively large distances and hence are not clustered within a small volume. Thirdly, it has been demonstrated that interactions captured by formaldehyde cross-linking must be stable for periods of approximately 5 seconds or more (Schmiedeberg et al., 2009). This may explain observations that while histones are readily cross-linked to DNA (V. Jackson, 1978), other proteins, including regulatory factors, are not (Solomon & Varshavsky, 1985). Even increased times of cross-linking will not necessarily result in a comprehensive and fully representative set of the interactions that were present within the sample.

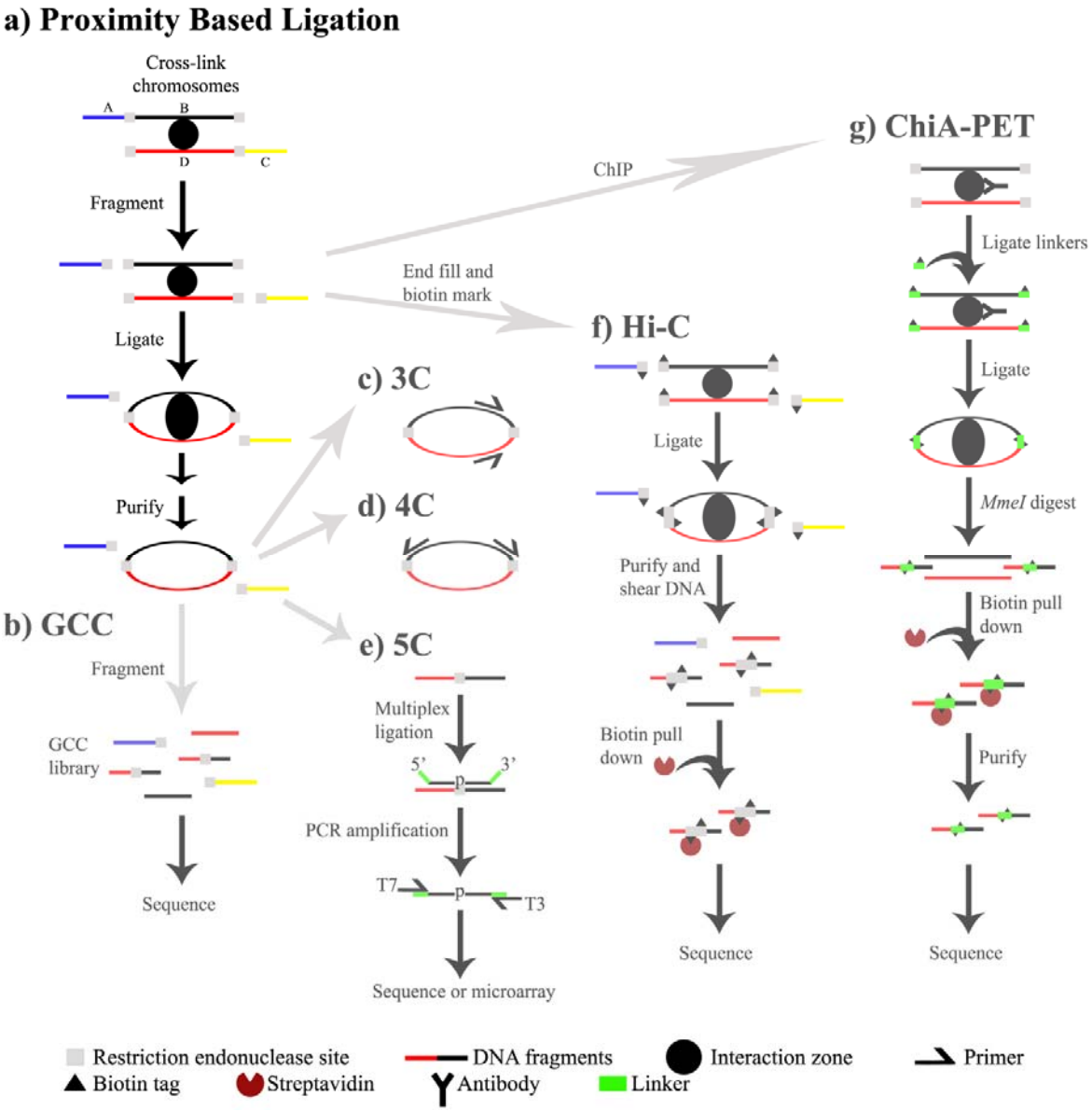


Figure 1.3. Molecular methods for the detection of chromosomal interactions. Refer to footnote [1] for the figure legend.

[1] a) The chromatin preparation module that forms the core of the proximity-based ligation methodologies. Chromatin is cross-linked, fragmented, and diluted prior to intra-molecular ligation, Cross-links are reversed and the library purified prior to interaction detection. Fragmentation is achieved by restriction enzyme digestion (*e.g.* (Dekker et al., 2002)) or sonication (*e.g.* (Fullwood, Liu, et al., 2009)). Fragmentation by sonication requires end repair. Hi-C (Lieberman-Aiden et al., 2009) and ChiA-PET (Fullwood, Liu, et al., 2009) deviate from this general scheme through the filling of the cohesive ends f) and incorporation of linkers and g), respectively. b) GCC is the simplest high-throughput methodology for the study of genome organization. GCC combines proximity-based ligation with random fragmentation and high-throughput sequencing in order to determine the genome organization. 3C and 4C incorporate c) direct and d) inverse PCR amplification to identify specific (*e.g.* (Dekker et al., 2002)) and unknown (*e.g.* (Simonis et al., 2006; Z. Zhao et al., 2006)) interactions, respectively. Amplified products are sequenced or hybridized to a microarray to confirm their identity. e) 5C (Dostie et al., 2006) utilizes primer ligation and amplification with universal T7 forward and reverse primers to amplify interacting segments. Interactions are subsequently identified by micro-array analyses or high-throughput sequencing. f) Hi-C (Lieberman-Aiden et al., 2009): Sticky ends produced by restriction enzyme digestion are end-filled with nucleotides, one of which is

1.6.1.2 CHROMOSOME CONFORMATION CAPTURE (3C)

Cross-linking is the first step in 3C and it is typically performed for 10 minutes (Dekker et al., 2002; Duan et al., 2010; Fullwood, Liu, et al., 2009; Lieberman-Aiden et al., 2009; O'Sullivan, Sontam, Grierson, & Jones, 2009; Rodley et al., 2009; Splinter, Grosveld, & de Laat, 2004; Z. Zhao et al., 2006). Once interactions have been captured by cross-linking, cells or nuclei are lysed and the cross-linked genomes are fragmented, usually by restriction enzyme digestion. The choice of restriction enzyme is dependent upon empirical and hypothetical considerations. Firstly, the sequence of the genomic region(s) under investigation automatically limits the choice of restriction enzymes that can be used, as certain restriction sites will or will not be present. Moreover, the desired resolution also affects the choice, as restriction enzymes that have a four nucleotide recognition site typically, but not uniformly, achieve a higher resolution than enzymes that have a six nucleotide recognition site. Secondly, while the use of restriction enzymes that leave overhangs or blunt ends is in many respects a personal preference when performing 3C, cohesive ends result in more efficient ligation (Sambrook & Russell, 2001; Sugino, Goodman, & Heyneker, 1977). Thirdly, the choice of enzyme is affected by the presence of DNA modification systems within the organism and the chances the modified sites overlap restriction sites. Finally, it is critical that the enzyme is easily inactivated following the digestion as carry-over of active restriction enzyme into the ligation step is counter-productive. Inactivation is generally achieved by the addition of SDS and incubation at moderate temperatures (≤65°C) in order to minimize the temperature dependent reversal of the formaldehyde mediated cross-links (V. Jackson, 1978; Solomon & Varshavsky, 1985).

Following digestion, the cross-linked genomic restriction fragments are diluted into a large volume prior to ligation of the free ends. In simplest terms, ligation requires that free DNA ends and ligase come together in the same place at one time. Diluting DNA samples reduces the chances of the tri-partite association,

---

biotinylated (black triangle). Intra-molecular ligation is performed prior to DNA purification and shearing. Biotinylated junctions are subsequently isolated and identified by paired-end sequencing. g) ChiA-PET (Fullwood, Liu, et al., 2009) deviates from the general chromatin preparation scheme through the integration of chromatin immunoprecipitation to fractionate the restriction enzyme digested chromatin. Subsequent steps include the incorporation of a biotinylated linker, prior to a final purification and high-throughput sequencing. Enhanced 4C (Schoenfelder et al., 2010), 6C, (Tiwari et al., 2008) and Associated Chromosome Trap (ACT) (Ling et al., 2006) are further variations on the proximity ligation method and were not included in this figure for simplicity. Adapted from (Grand et al., 2011)

unless the free DNA ends are from the same molecule. However, ligatable free ends can come from two or more linear DNA fragments that are physically held together in one complex by internal linkages. In such a situation, dilution will separate the non-linked DNA fragments but not those within a single complex. Importantly, dilution promotes the formation of intra-molecular ligation products by reducing the chances of inter-molecular ligations within the fixed time period. Moreover, the frequency of free DNA end association depends on factors that include the chromatin state and as a consequence the real and apparent length of the DNA fragment(s) (Ringrose, Chabanis, Angrand, Woodroofe, & Stewart, 1999; Rippe, 2001). Internal controls are often used to determine the inter-molecular ligation frequency and to standardize between samples. However, the use of internal controls requires evidence that the chosen loci are unlinked, spatially separate, and have similar interaction frequencies in the different samples (Dekker, 2006; Splinter et al., 2004). The use of external controls is more advisable as a means to determine the frequency of inter-fragment ligation events (work since published herein: (Cagliero, Grand, Jones, Jin, & O'Sullivan, 2013)).

Finally, cross-links are reversed and the DNA fragments are purified. Following purification, the mixture of linear and circular DNA molecules constitutes a 3C library. PCR is subsequently used to obtain preliminary confirmation of the presence of an interaction within the 3C library. PCR primers are designed to amplify across the reconstituted restriction site (Figure 1.3c) formed by the ligation of two interacting restriction fragments. If an interaction occurred a PCR product will be produced. In contrast, an absence of ligation results in the primer annealing sites (Figure 1.3c) being in a non-convergent orientation that prevents PCR amplification (Dekker et al., 2002). The identity of the PCR product is subsequently confirmed by restriction digestion and/or sequencing.

1.6.1.3 Circular Chromosome Conformation Capture (4C)

4C is a global method suitable for the detection of DNA regions that interact with a predetermined region of interest (the bait). 4C is similar to 3C, in that its reliance on PCR for the identification of chromosomal interactions, albeit inverse PCR directed outwards from the bait fragment (Figure 1.3d). There are two versions of 4C, both of which utilize inverse PCR (Simonis et al., 2006; Z. Zhao et al., 2006). As such, 4C enables the identification of 'unknown' interacting DNA fragments and thus removes the 3C requirement of *a priori* suspecting the identity of both interacting loci.

1.6.1.4 CHROMOSOME CONFORMATION CAPTURE CARBON-COPY (5C)

5C ((Dostie et al., 2006), reviewed in (van Berkum & Dekker, 2009)) employs a different approach to identify interacting DNA fragments as it is not anchored to a single locus and therefore maps the global interaction networks. As in 3C, 5C requires oligonucleotides that anneal either side of the restriction enzyme site(s) of interest. These oligonucleotides contain universal T7 primer binding sites and are designed to anneal adjacent to each other on the same strand of the ligated template, provided the two DNA fragments interact. This arrangement enables the oligonucleotides to be ligated together and amplified, using the T7 forward and reverse primers which are complementary to the ends of the oligonucleotides. The primer ligation step utilized in 5C is an extension of the ligase chain reaction developed in the early 1990's (reviewed in (Wiedmann et al., 1994)). Amplification products, and hence chromosomal interactions, are subsequently identified by sequencing or microarray detection (Figure 1.3e) (Dostie et al., 2006). 5C provided a step towards the elucidation of global genome organization by enabling multiplexing within a single 3C sample, using forward and reverse oligonucleotides that span a single region or an entire genome.

1.6.2 GLOBAL IDENTIFICATION OF DNA INTERACTIONS

The development of 3C enabled genome organization to be readily studied at molecular level resolution. However, 3C remains limited to the analysis of small numbers of chromosomal loci. Moreover, both 3C and 4C require the *a priori* selection of at least one of the two interacting fragments. This bias represents a serious hurdle between these techniques and their application to the study of global genome organization. While the requirement for *a priori* knowledge can be circumvented by the 5C method, the global application of this method requires blanketing the genome with forward and reverse oligonucleotides. Therefore, the genome-wide application of these methods is impractical.

GCC (Rodley et al., 2009), Hi-C (Duan et al., 2010; Lieberman-Aiden et al., 2009) and Chromatin Interaction Analysis using Paired-End Tag sequencing (ChiA-PET) (Fullwood, Liu, et al., 2009) overcome the requirements for the *a priori* identification of regions of interest. These techniques rely on cross-linking and intra-molecular ligation (Dekker et al., 2002). However, rather than designing primers for the detection of the interacting fragments, the purified ligated DNA samples are sequenced using high-throughput genome sequencing technology.

The large numbers of sequence tags that are obtained are analyzed to construct genome-wide maps of the interactions that were captured at the time of cross-linking. As such, the high-throughput techniques facilitate the unbiased and detailed interrogation of global genome organization. ChiA-PET (Fullwood, Liu, et al., 2009) incorporates chromatin immunoprecipitation (ChIP) and is designed to implicate proteinaceous factors in the maintenance of chromosomal interactions. ChiA-PET will be discussed in more detail in section 1.6.3.2.

ChIA-PET, GCC and Hi-C have been used to produce the first generation of experimentally defined models of global eukaryotic genome organization (Duan et al., 2010; Fullwood, Liu, et al., 2009; Lieberman-Aiden et al., 2009; Rodley et al., 2009). As a result, proximity ligation and the proximity-based ligation technologies have increased our ability to perform in-depth and rapid hypothesis driven investigations of genome organization (*e.g.* (Z. Liu & Garrard, 2005; Lomvardas et al., 2006; Murrell, Heeson, & Reik, 2004; Shoshani, Benvenisty, Trus, & Reshef, 1991; Spilianakis et al., 2005; Tolhuis et al., 2002)) and its roles in the regulation of gene transcription (Bartkuhn & Renkawitz, 2008; Cook, 1999; Goetze et al., 2007; Simonis & de Laat, 2008; G. S. Stein et al., 2003) and epigenetic memory (D. G. Brickner et al., 2007; J. H. Brickner, 2009; Laine, Singh, Krishnamurthy, & Hampsey, 2009; Tan-Wong et al., 2009).

1.6.2.1 Genome Conformation Capture (GCC)

The first eukaryotic global chromosome interaction map was produced from exponentially growing *Saccharomyces cerevisiae* cells using GCC (Rodley et al., 2009). GCC is the simplest of the high-throughput techniques. GCC effectively extends the 3C methodology by identifying spatial proximity by direct sequencing (Figure 1.3b) (Rodley et al., 2009). Moreover, GCC does not require the specific amplification of the captured interactions, or the incorporation of linkers, or labelled nucleotides (Table 1.1). As such, it can be used to produce high resolution, unbiased chromosome interaction maps. However, the lack of enrichment means that large numbers of sequence tags are required, even for relatively small genomes. Enrichment strategies that rely upon immunoprecipitation of the restriction enzyme site were proposed (Rodley et al., 2009) and have subsequently been developed (unpublished). The incorporation of these enrichment strategies will enable the investigation of larger genomes. However, the existing GCC methodology can be easily applied to the study of genomes under 20 Mb in size.

1.6.2.2 Hi-C

Hi-C (Lieberman-Aiden et al., 2009) is similar to GCC (Rodley et al., 2009) except that it incorporates an enrichment step that is specific for the ligated products. Enrichment is achieved by filling the restriction site 5′ overhang with a biotinylated residue prior to blunt-end proximity ligation. The resulting ligation products consist of fragments that were originally in close spatial proximity in the nucleus, marked with biotin at the junction. The ligation product library is subsequently sheared and fragments containing the interaction junction are enriched using streptavidin beads. The library is then analyzed using high-throughput DNA sequencing, producing a catalogue of interacting fragments (Figure 1.3f).

Hi-C provides a crucial step towards the identification of DNA interactions on a genome-wide scale in large genomes albeit with some limitations. Firstly, blunt end ligation is less efficient than sticky ended ligation (Sambrook & Russell, 2001; Sugino et al., 1977). Secondly, combining phenol:chloroform extractions with biotinylated DNA is problematic, due to purification biases introduced by the hydrophobic nature of the biotin moiety (Laman, Kurjukov, Bulgakova, Anikeeva, & Brovko, 2001; Langer, Waldrop, & Ward, 1981; Lucas, Wu, Guo, Chi, & Chen, 2006). However, this is likely to only be an issue for smaller DNA fragments, and can be overcome by replacing the phenol:chloroform purification steps with column purifications.

Since the publication of the above book chapter (see section 1.6) a number of additional methods based on the proximity-based ligation technique to identify DNA-DNA contacts on a genome-wide scale have been published (Duan et al., 2010; Sexton et al., 2012; Tanizawa et al., 2010). Two of these variants involve the additional restriction enzyme digestion of a 3C library followed by either: 1) the re-ligation and digestion with the initial enzyme to linearize ligation events between non-adjacent restriction fragments (Tanizawa et al., 2010), or 2) the sequential insertion of a type IIS or type III restriction enzyme site and biotinylated adaptor. Both of these methods enable the enrichment of interacting fragments that are subsequently sequenced and used to construct probability interacting networks (Duan et al., 2010). Most recently, a method claimed to be a simplified Hi-C variant, where 3C libraries are just fragmented and sequenced, is in fact essentially GCC (Sexton et al., 2012).

1.6.3 Methods to Localize Proteins at Chromosome Interaction Sites

3C, 4C and 5C were developed to detect chromosomal interactions and do not address the issue of whether the interactions are direct or mediated by a protein or nucleic acid. Enhanced-ChIP-4C (e4C) (Schoenfelder et al., 2010), 3C-ChIP-cloning (6C) (Tiwari, Cope, McGarvey, Ohm, & Baylin, 2008), ChiA-PET (Fullwood, Liu, et al., 2009) and variations thereof (Cai, Lee, & Kohwi-Shigematsu, 2006; Carroll & Brown, 2005; Horike, Cai, Miyano, Cheng, & Kohwi-Shigematsu, 2005) were developed in an attempt to directly address this issue.

1.6.3.1 Combined 3C-ChIP-Cloning (6C)

6C combines the cross-linking, digestion, and intra-molecular ligation of chromatin with ChIP and cloning (Tiwari et al., 2008). The incorporation of ChIP requires that *a priori* assumptions are made about the identity of the protein(s) that mediate the chromosomal interactions of interest. The detection of chromosomal interactions places the antigen targeted by the ChIP within the localized neighbourhood of the interaction. The size of the neighbourhood is dictated by the fragmentation method (*e.g.* restriction enzyme) and the size of the resulting fragment(s).

A positive result in 6C stops short of proving that the antigen is directly involved in the establishment and/or maintenance of the specified interaction. Proof of involvement in the actual interaction requires the deletion, knock-down or inactivation of the protein of interest and subsequent testing for the presence of the interaction by 3C or another technique.

Table 1.1 Advantages and disadvantages of the different molecular methods for the detection of chromosomal interactions

Method	Reference	Advantage(s)	Disadvantage(s)
Chromosome Conformation Capture (3C)	(Dekker et al., 2002)	1) Detects inter- and intra-chromosomal interactions. 2) Uses both sticky and blunt end ligation.	1) Requires *a priori* predictions of DNA interactions. 2) Restricted to the detection of a single interaction per PCR reaction.
Circular Chromosome Conformation Capture (4C)	(Z. Zhao et al., 2006)	Detects all DNA fragments interacting with a particular region.	1) Requires *a priori* selection of one interaction partner – the bait. 2) Limited to the identification of interactions with a single locus at a time.
Chromosome Conformation Capture Carbon Copy (5C)	(Dostie et al., 2006)	Allows the global identification of DNA interactions through multiplexing and scaling.	Does not scale well due to the requirement to design oligonucleotides that anneal adjacent to every restriction site in the genome.
Combined 3C-ChIP-Cloning (6C)	(Tiwari et al., 2008)	Determines if specific proteins bind in the vicinity of chromosomal interactions.	Requires *a priori* assumptions as to the identity of the proteins in the vicinity of the chromosomal interactions.
Chromatin interaction Analysis with Paired-End Tag (ChiA-PET) sequencing	(Fullwood, Liu, et al., 2009)	Global method to identify chromosomal interactions in the vicinity of specific proteins.	1) Requires *a priori* assumptions as to the identity of the proteins in the vicinity of the chromosomal interactions. 2) Requires the ligations of linker sequences before intra-molecular ligation.
Genome Conformation Capture (GCC)	(Rodley et al., 2009)	A global un-biased method for the detection of chromosomal interactions.	Not suitable for the analysis of large genomes as it lacks an enrichment step.
Hi-C	(Lieberman-Aiden et al., 2009)	1) A global un-biased method for the detection of chromosomal interactions. 2) Enriches for interacting DNA fragments.	1) Ligation efficiency is reduced due to the use of blunt end ligation. 2) Potential bias due to use of biotin moiety during DNA purification.

1.6.3.2 CHROMATIN INTERACTION ANALYSIS USING PAIRED-END TAG SEQUENCING (CHIA-PET)

ChiA-PET ((Fullwood & Ruan, 2009; Fullwood, Wei, Liu, & Ruan, 2009; Fullwood, Liu, et al., 2009), reviewed in (Fullwood, Han, Wei, Ruan, & Ruan, 2010)) was specifically developed to enable the large-scale, *de novo* analysis of higher-order chromatin structure (Fullwood, Liu, et al., 2009). ChiA-PET (Fullwood, Liu, et al., 2009) uses a similar general strategy to 6C in order to place antigens within the vicinity of chromosomal interactions on a global scale (Figure 1.3g). As such, it differs from the other global 3C based technologies through the incorporation of a ChIP step. ChIP is introduced following chromatin fragmentation, to fractionate and enrich for cross-linked chromatin DNA segments bound to the protein of interest. The incorporation of this ChIP step reduces the level of complexity of the fragment library prior to the proximity-based ligation step. The proximity-based ligation step was modified to incorporate linkers containing restriction sites that enable the biotin mediated enrichment of the ~27 bp of sequence at the ends of the interacting DNA sequences (Figure 1.3g) (Fullwood et al., 2010). Enrichment occurs prior to sequencing and hence reduces sample complexity. Thus, ChiA-PET enables the proximity of protein and chromosomal interactions to be determined, within the resolution of the restriction fragment size.

ChIA-PET and 6C are powerful techniques, yet while it is reasonable to assume that mechanisms exist to promote the correct association of interacting loci and prevent accidental interactions (Dobi & Winston, 2007), there is no *a priori* reason to think that the formation of chromosomal interactions at different sites is mediated by the same signal, nucleic acid, or protein moieties (Ansari & Hampsey, 2005; Bartlett et al., 2006; Laine et al., 2009; Singh & Hampsey, 2007; Tan-Wong et al., 2009). Moreover, it is possible that 'accidental' interactions have functional roles in certain nuclear processes (*e.g.* DNA repair).

1.6.4 WHAT DO THE INTERACTIONS CAPTURED REPRESENT

Microscopic observations that whole chromosomes and individual loci occupy specific regions within a cell (nucleus) imply that certain DNA sequences are spatially more proximal to each other than others. With the development of 3C methods, this has been shown to transcend to the molecular level with specific regions of DNA detected colocalizing more frequently than others. The 3C methods have enabled unprecedented insight into the local and global organization of genomes based on the frequency of contacts detected between DNA segments.

This has put great emphasis on the ‘interactions’ detected between distant regions of DNA (both inter- and intra-chromosomal) and what the fluctuation of the contact frequency may imply.

The detection of an interaction between two genomic sequences can be a result of three main processes: 1) the thermodynamic movement of the DNA polymer and molecular crowding within the cell or nucleus results in the random contact between two chromosome regions; 2) two chromosome regions come into close spatial proximity as part of defining structural features such as centromere and telomere clustering, and the formation of the nucleolus; and 3) two chromosome regions come into close spatial proximity through an active (*e.g.* ATP driven) cellular process that influences functional output of the genome (*i.e.* transcription, replication and DNA repair). Experimental controls and statistical methods are used to distinguish between the different types of interactions with the greatest interest in the interactions that contribute to cellular function, the genotype to phenotype translation. Clear examples of functional interactions detected as a result of DNA loop formation between regulatory sequences and promoters of the genes they regulate are beginning to emerge and the role of the interactions is being unravelled. However, detailed examples are few and the understanding of the organization of the genome at the global level and its involvement in transcription remain vague. Furthermore, the extent and importance of changes in genome organization over time in response to the environment remains largely unexplored.

1.7 GOALS OF THE THESIS

The extensive investigation of the spatial organization of bacterial nucleoids and eukaryotic nuclear chromosomes has clearly demonstrated that they are highly organized. The establishment of the spatial organization of genomes appears to be a balance between the forces and processes that compact and compartmentalize the chromosomes (*e.g.* DNA supercoiling, DNA binding proteins, macromolecular crowding, and entropy-driven depletion attraction) and the requirement for the DNA to be accessible to essential cellular processes such as DNA replication and transcription. Despite the significant advances in techniques for the investigation of spatial genome organization our understanding of the functional significance of specific three-dimensional genome conformations is limited. Furthermore, how changes in the spatial organization of genomes through time participate in the

regulation of cellular processes such as transcription, and consequently in a cell's ability to adapt, have been minimally characterized.

To gain further insight into the role that the spatial organization of genomes has on the genotype - phenotype translation will require the integrated study of chromosome organization at the molecular level and cellular processes. Further, most studies of the spatial genome organization using proximity-based ligation methods have been performed on asynchronous cells and multicellular organisms. Consequently, due to the dynamic nature of the chromatin fibre, this may cloud our interpretation of the relationship between genome structure and function. Therefore, the utilization of synchronous cell cultures will likely provide a more detailed understanding of the role spatial genome organization plays in functional cellular processes. In an attempt to further our understanding of the role that the spatial organization of genomes and changes therein have on the genotype - phenotype translation, I have addressed three main questions relating to genome structure and function in space and through time.

1) Do changes in the spatial organization of the *Escherichia coli* nucleoid in response to nutrient deprivation have a functional role in adapting to the stress?

Following an induced amino acid starvation by the treatment with serine hydroxamate there is an observable expansion of the *E. coli* nucleoid and changes in gene transcription. Combining GCC with microarray data on gene transcript levels, I investigated the potential involvement for specific DNA-DNA contacts in the modulation of specific gene transcript levels enabling the cell to cope with the changing environment. I further investigated the role of DNA replication and the classical NAPs in the spatial organization of the nucleoid.

2) Are there changes in spatial genome organization throughout the cell cycle of *Schizosaccharomyces pombe* and do these changes facilitate the establishment of cell cycle specific transcription profiles?

S. pombe cells were synchronized at three phases of the cell cycle: G1, G2 and M phases and GCC and RNA-seq were performed on these cells. I determined whether there were detectable differences in genome organization between the three phases of the cell cycle, which included an *in vivo* molecular level analysis of metaphase chromosomes. I investigated the functional role of the changes in

genome organization observed in the establishment and maintenance of cell cycle specific transcription profiles.

3) Can mt-nDNA interactions that were detected in the Budding yeast also be found in the Fission yeast and do they serve functional roles?

It had been demonstrated in *S. cerevisiae* that Interactions between mitochondrial genes (*i.e. COX1* and *Q0182,* a dubious mitochondrial ORF) and nuclear encoded loci (*i.e. MSY1* and, *RSM7*, respectively), are dependent upon a functional electron transport chain and mitochondrial encoded reverse transcriptase machinery (Rodley et al., 2012). I investigated whether the levels of the nuclear encoded *MSY1* and *RSM7* gene transcripts changed when the interaction frequency of the respective interactions was reduced by the knockout of mitochondrial reverse transcriptase activity.

I characterized the relationship between mt-nDNA interactions and cellular function throughout the *S. pombe* cell cycle. I mapped mt-nDNA interactions using GCC for *S. pombe* cells synchronized at the G1, G2, and M phases of the cell cycle. The enrichment for particular protein products and transcriptional levels of nuclear encoded genes that were associated with the detected mt-nDNA interactions were assessed. In addition, the role for mt-nDNA interactions in the regulation of nuclear DNA replication was also investigated.

Chapter 2

2 GENOME CONFORMATION CAPTURE REVEALS THAT THE *ESCHERICHIA COLI* CHROMOSOME IS ORGANIZED BY REPLICATION AND TRANSCRIPTION

This chapter has been published see Appendix II:

Cagliero, C.*, **Grand, R.S.***, Jones, M.B., Jin, D.J. and O'Sullivan, J.M. (2013). Genome conformation capture reveals that the *Escherichia coli* chromosome is organized by replication and transcription. Nucleic Acids Research, 41(12): 6058-6071.

[1]Gene Regulation and Chromosome Biology Laboratory, Frederick National Laboratory for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD 21702, USA, [2]Institute of Natural and Mathematical Sciences, Massey University, Auckland 0745, New Zealand and [3]Liggins Institute, University of Auckland, Auckland 1023, New Zealand.

* The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Acknowledgements: The authors would like to thank Philippe Collas, Austen Ganley, Gary Greyling, Lutz R. Gehlen, Heather Hendrickson, Julia Horsfield, Rod McNab, Ana Pombo, Tom Schneider and Yan Ning Zhou for helpful discussions.

Author's contributions: CC, RSG, JDJ, and JMO conceived and directed the project, and wrote the manuscript. CC and RSG performed experiments and analysis. MBJ advised and helped with statistical analysis.

2.1 INTRODUCTION

Our understanding of the spatial organization of bacterial genomes and its relationship to cellular function is limited (for reviews see (Ishihama, 2009; Sherratt, 2003; Woldringh & Nanninga, 2006)). Yet it is clear that despite not being enclosed in a nuclear membrane, bacterial nucleoids are spatially organized within a defined sub-fraction of the cell volume (Boccard et al., 2005; X. Liu et al., 2010; Mercier et al., 2008; Postow et al., 2004; Thiel et al., 2012; Umbarger et al., 2011; X. Wang & Sherratt, 2010; Wiggins, Cheveralls, Martin, Lintner, & Kondev, 2010). Various molecular (reviewed in (Ishihama, 2009)) and recombination based methodologies have been used to identify the existence of micro- and macrodomains within the *Escherichia coli* nucleoid (*e.g.* (Espeli et al., 2008; Ishihama, 2009; Mercier et al., 2008; Thiel et al., 2012; Valens et al., 2004)). The four structured macrodomains (~ 0.5 to 1 Mb) that have been identified exhibit preferential intra-domain recombination between lambda att sites while inter-domain recombination is reduced (Boccard et al., 2005; Espeli et al., 2008; Mercier et al., 2008; Thiel et al., 2012; Valens et al., 2004). By contrast, microdomains are much smaller (average ~10 kb) and have been linked to the topological isolation of supercoils (Ishihama, 2009; Postow et al., 2004). Collectively, micro- and macrodomains are hypothesized to be critical for maintaining global organization whilst enabling the local levels of compaction required to fit a circular chromosome with an extended diameter of ~490 nm within a cell with a length as small as 1000 nm (Ishihama, 2009).

Unlike eukaryote chromatin, the bacterial nucleoid does not contain histones. However, Nucleoid Associated Proteins (NAPs) particularly histone-like proteins such as H-NS, HU, Fis and IHF are believed to act like histones and play a significant role in the organization of the nucleoid (Dame et al., 2011; Margolin, 2010; Tanaka et al., 1995; W. Wang et al., 2011). These NAPs exhibit DNA bending, looping, and bridging properties *in vitro.* However, studies also indicate that *in vivo* the role of the NAPs could be more regulatory than architectural (*e.g.* (Dame, 2005; Grainger et al., 2006)). Non-classical NAPs (*i.e.* SeqA, SlmA, and MatP) have been recently characterized as exhibiting macrodomain-specific DNA binding properties (reviewed in (Dame et al., 2011)) and may represent alternative candidates for organizational roles within the nucleoid.

The structure of the bacterial nucleoid is dynamic and affected by growth conditions and stress (Cabrera et al., 2009; Cabrera & Jin, 2003; Cagliero & Jin, 2013; Jin et al., 2012; Margolin, 2010). For example, serine hydroxamate (SHX) induces the stringent response (Cashel et al., 1996) and inhibits replication initiation through artificial amino-acid starvation. In terms of the biology of the *E. coli* nucleoid, the overall effect of the SHX-induced amino-acid starvation is an expansion of the nucleoid and a change in transcription patterns (Durfee et al., 2008; Traxler et al., 2008). This suggests a relationship between transcription and the organization of the nucleoid (Jin & Cabrera, 2006). However, the mechanism(s) behind the re-structuring of the nucleoid in response to growth and stress is still largely unknown.

Another long standing question is when and how the nascent nucleoid that arises from DNA replication segregates during bacterial cell growth (reviewed in (Woldringh & Nanninga, 2006)). In *E. coli*, the time required for the replication of the nucleoid is fixed at ~ 40 minutes (Bremer & Dennis, 1996). To maintain a fast growth rate, cells growing in rich media must initiate multiple rounds of replication before each division. Consequently, a typical cell growing in rich media contains up to 16 origins of replication (Nielsen, Youngren, Hansen, & Austin, 2007). Whether the nascent nucleoids segregate rapidly (Elmore et al., 2005; Y. Li et al., 2002; X. Wang et al., 2005) or remain associated following replication, by a cohesion dependent mechanism (*i.e.* the cohesion model) as seen in eukaryotes (Bates & Kleckner, 2005; Sunako et al., 2001), remains unresolved.

Advances in Chromosome Conformation Capture (3C) related methodologies (Dekker et al., 2002) enable the direct, high-resolution, detection of chromosome organization (*e.g.* (Dostie et al., 2006; Fullwood, Liu, et al., 2009; Lieberman-Aiden et al., 2009; Rodley et al., 2009; Tanizawa et al., 2010)). Recently, Chromosome Conformation Capture Carbon-Copy (5C) was used to generate a global DNA:DNA contact map for *Caulobacter crescentus* synchronized swarmer cells (Umbarger et al., 2011). Here we present a high resolution analysis of the DNA:DNA interactions within *E. coli* nucleoids in rapidly growing and starved cell populations. Using Genome Conformation Capture (GCC), we observe a clear relationship between DNA:DNA interactions, copy number and DNA replication. This suggests that nucleoids remain associated after replication, consistent with the cohesion model. Furthermore, SeqA binding sites exhibit replication-dependent clustering while

binding sites for the major histone-like proteins (Fis, H-NS and IHF) did not. Finally, we observe a correlation between gene regulation and spatial clustering.

2.2 RESULTS

In GCC, the spatial organization of the nucleoid is captured by formaldehyde cross-linking within intact cells prior to cell lysis and the isolation of the nucleoid (Figure 2.1A). Once isolated the nucleoid is digested, diluted and incubated with DNA ligase to enable the capture of spatially proximate, but linearly separated, loci (Figure 2.1A) (Rodley et al., 2009). This produces an interaction library that can be sequenced to identify the network of chromosomal interactions occurring at the moment of cross-linking. GCC differs from current competing unbiased 3C technologies in that all DNA material is sequenced without the prior selection of DNA fragments containing ligation products. Therefore, there are no enrichment introduced biases and DNA copy variation can be determined.

GCC relies upon the intra-molecular ligation of cross-linked loci. However, inter-molecular ligation events resulting from random associations during the procedure can also occur, leading to false positives. To reduce the chances of isolating false positives we: a) induce expansion of the nucleoid by isolation in a high salt environment (a “high-salt nucleoid” (Ishihama, 2009)), following cross-linking of the interacting loci; b) added external ligation controls during GCC library preparations to empirically measure the background level of random inter-molecular ligation events. Thus, we determined a cut-off, for the minimum number of sequences representing any one interaction, above which interactions were deemed significant (Materials and Methods section 2.4.7). The following analyses were only performed on interactions that were above this significance threshold.

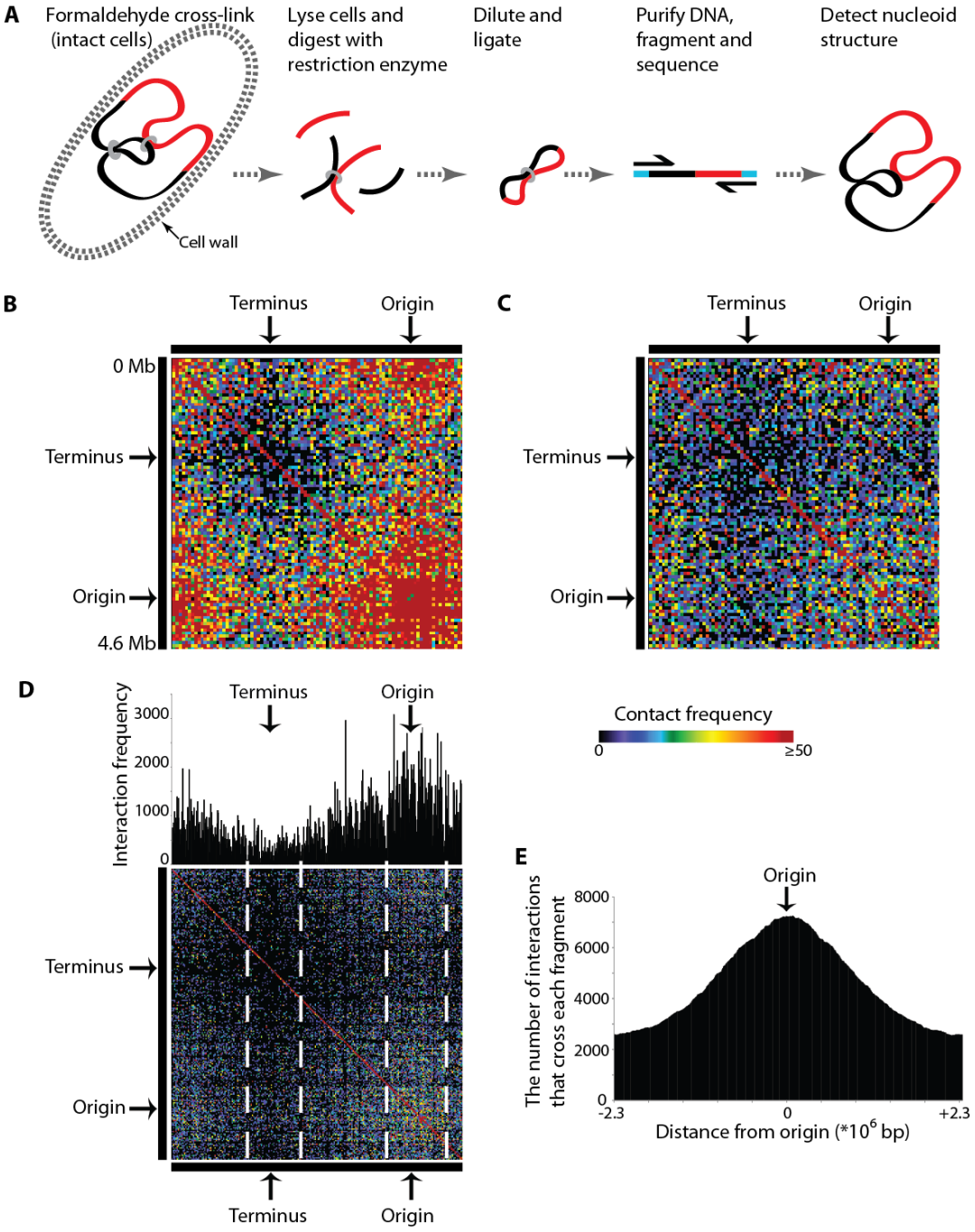


Figure 2.1. Ori and Ter domains are present within the *E. coli* nucleoid.

A) Schematic of the GCC procedure (Rodley et al., 2009). Intact cells are cross-linked with formaldehyde prior to lysis and the cross-linked nucleoids are isolated. The nucleoids are restriction digested, diluted and ligated to generate an interaction library. The interaction library is sequenced, following the addition of sequencing adapters (blue bars), and the network of interactions that define the nucleoid organization is determined. B) Genome-wide contact matrix (50 Kb bins) for exponentially growing *E. coli* nucleoids. The matrix highlights the Ori (high contact region) and Ter domains (low contact region). C) Genome-wide contact matrix (50 Kb bins) for nucleoids isolated from SHX treated *E. coli*. The Ori and Ter domains remain visible. D) Genome-wide contact matrix (20 Kb bins) and bar graph for exponentially growing nucleoids high-lighting regions of low interaction frequency ('domain boundaries') surrounding the Ori and Ter regions. E) Frequency of exponential phase interactions that cross each restriction fragment plotted as a function of distance from the Ori (0). Fixed boundaries are not observed. The profile for the SHX treated cells are no different (data not shown).

2.2.1 ORIGIN AND TERMINUS DOMAINS EXIST WITHIN THE *E. COLI* NUCLEOID

Chromosome interaction networks were determined for rapidly growing cells in rich medium harvested at early exponential phase and exponential cells treated with serine hydroxamate (SHX) (Figure 2.1B and C). The exponential phase chromosome interaction network (Figure 2.1B) is dominant in two regions: 1) a high frequency interaction domain surrounding the origin (Ori); and 2) a low frequency interaction domain surrounding the terminus (Ter). These Ori and Ter domains are also present in the interaction network for the SHX treated samples, although they are less pronounced (Figure 2.1C). Higher resolution (*i.e.* 20 Kb) emphasizes that the exponential phase interaction network contains regions that have a demonstrably lower average interaction frequency than the adjacent Ori and Ter domains (Figure 2.1D). We attribute these reductions to the presence of non-fixed domain boundaries within the population. We predicted that these boundaries would reduce interactions between domains and that this would be manifested as a reduction in the interactions that cross the boundary regions. However, despite the obvious Ori preference, there is no sharp reduction in the numbers of interactions that cross our apparent domain boundaries (Figure 2.1E). Despite the diffuse boundaries for the Ori and Ter domains, we observe several noticeable reductions in the interaction frequency at various locations in the chromosome that could represent additional domain boundaries.

Table 2.1. A summary of *E. coli* chromosomal interactions.

A) Summary of interactions for *E. coli* cells growing in the exponential phase (wt (LB)) or following SHX treatment. B) Breakdown of the interactions in A) according to whether they were shared or unique for each condition. C) A further breakdown of the interactions specific to the exponential phase data set, shared by the two conditions or specific to SHX treated data set. In the condition specific interactions a large proportion of the non-adjacent interactions were long distance (>800bp), in contrast only a very small fraction of the shared interactions were long distance.

A

	Total number of unique interactions	Number of self-interactions	Number of adjacent interactions	Number of non-adjacent interactions
wt (LB)	148,036	21,369 (14.4%)	31,142 (21.1%)	95,525 (64.5%)
SHX treated	127,652	21,258 (16.6%)	30,972 (24.3%)	75,422 (59.1%)

B

	Break down of interactions shared by both condition				Break down of interactions specific to each condition			
	Total number of interactions	Number of self-interactions	Number of adjacent interactions	Number of non-adjacent interactions	Total number of interactions	Number of self-interactions	Number of adjacent interactions	Number of non-adjacent interactions
wt (LB)	117,174 (79.2%)	20,749 (14.0%)	30,710 (20.8%)	65,715 (44.4%)	30,862 (20.8%)	620 (2.0%)	432 (1.4%)	29,810 (96.6%)
SHX treated	117,174 (91.8%)	20,749 (16.2%)	30,710 (24.1%)	65,715 (51.5%)	10,478 (8.2%)	509 (4.9%)	262 (2.5%)	9,707 (92.6%)

C

	Exponential specific	Shared	SHX specific
Total (T)	30,862	117,174	10,478
Non-adjacent (NA)	29,810 (96.6% T)	65,715 (56% T)	9,707 (92.6% T)
>800 bp	17,512 (56.7% T) (58.7% NA)	1,838 (1.6% T) (2.8% NA)	6,275 (59.9% T) (64.6% NA)

2.2.2 Interactions within the Ori and Ter regions are linked to replication

Comparisons of the chromosome networks from the exponential and SHX treated cells identified similar levels of self and adjacent interactions (Table 2.1A and B). However, SHX treatment results in fewer long distance interactions (between 800 bp and half the length of the genome, respectively; Figure 2.2A), shorter loop lengths (Figure 2.2B), and reduced numbers of partners per fragment (Figure 2.3A and Figure 2.4) when compared with the exponential network. These observations are consistent with SHX decreasing the overall compaction of the nucleoid (Cabrera et al., 2009; Cabrera & Jin, 2003; Jin et al., 2012).

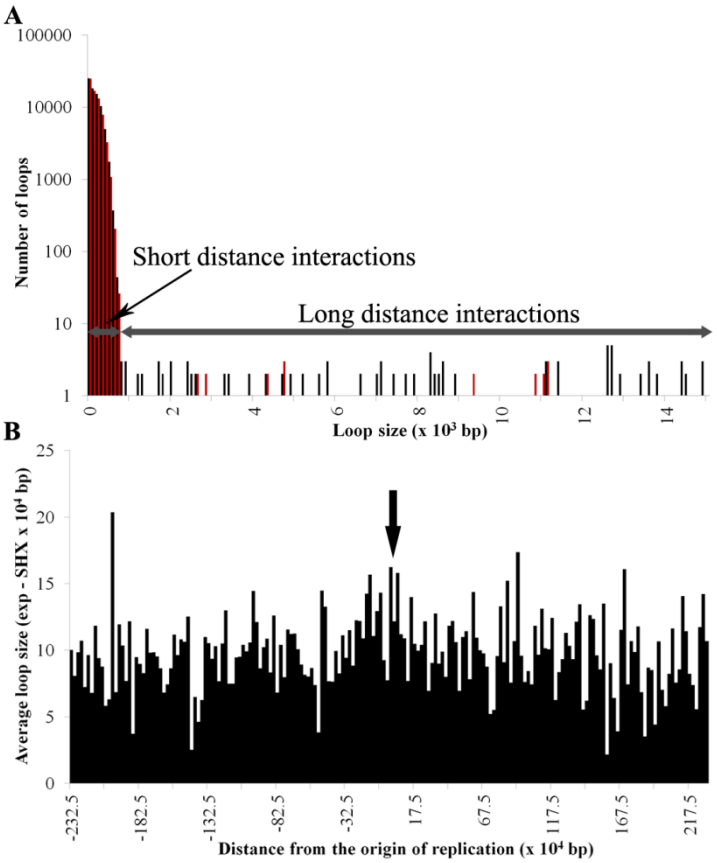


Figure 2.2. There is a reduction in the number of long distance interactions in the SHX treated nucleoids.

A) Interaction loop lengths within the *E. coli* genome can be divided into two distinct populations based on frequency: 1) Short (<800 bp) and 2) long distance (≥800 bp) interactions. The histogram of loop lengths (black, exponential phase; red, SHX treated) was calculated using 100 bp bins. For visualization, only loops less than 15 kb in length have been shown. B) Loop lengths are longer in exponential phase cells than following SHX treatment (~98 kb ± SE 2064 bp). The difference between the average loop length for each pair of interacting fragments in exponential phase compared to the SHX treated cells was calculated and a histogram plotted using 25,000 bp bins centred at the origin of replication (3,923,883 bp). Since the *E. coli* chromosome is circular the loop size will always be less than half the length of the genome, therefore whichever of the calculated length (L_c) or L_r=4,639,675-L_c was shorter than 4,639,675/2 was counted for the frequency distribution. exp, exponential; SHX, serine hydroxamate. Origin position is indicated by the arrow.

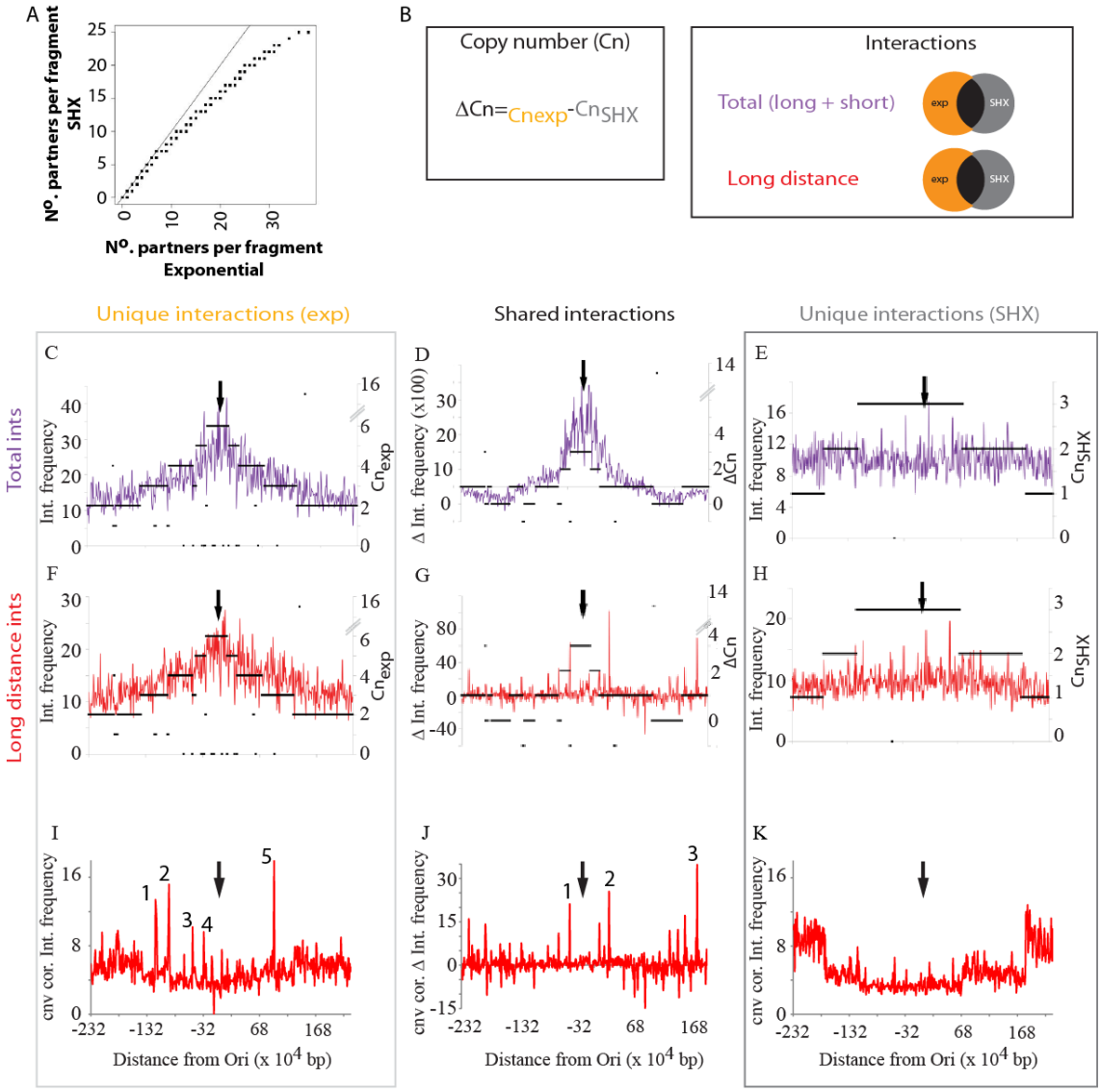


Figure 2.3. Origin proximal interactions are more frequently detected.

A) Fragments that interact have more partners in the exponential nucleoids as opposed to SHX treated nucleoids. The 45o line shows the expected pattern if the number of partners for each fragment is equal in both conditions. B) Schematic of the copy number and interaction comparisons that were performed. Comparisons between interaction frequency and copy number: C-E), total observed interactions; F-H), long distance (>800 bp) interactions. C) Interactions that are specific to exponential phase growth correlated with copy number. D) Differences in frequency for shared interactions between exponentially growing and SHX treated E. coli cells indicate a correlation with copy number. E) Interactions that were specific to SHX treated cells are copy number independent. F) Exponential phase specific long distance interactions correlated with copy number. G) Removal of short distance interactions (≤ 800 bp) removed the copy number dependence of the shared interactions. H) SHX specific interactions were independent of copy number. I) Correction of exponential specific long distance interactions identifies five peaks (I1-5) of increased interactions at positions I1) 2,753,883-2,773,883 bp; I2) 2,983,883-3,003,883 bp; I3) 3,413,883-3,423,883 bp; I4) 3,613,883-3,623,883 bp; I5) 224,208-234,208 bp J) Correction of shared long distance interactions identifies three peaks (J1-3) of increased interactions at positions 1) 3,643,883-3,653,883 bp; 2) 4,383,883-4,393,883 bp; 3) 1,404,208-1,414,208 bp. K) Correction of SHX specific long distance Interactions for copy number identifies a decrease in the relative frequency of interactions at the origin compared to the terminus. Interactions were tallied for 10,000 bp bins and corrected for the number of fragments per bin. Vertical, grey broken lines denote the position of the origin of replication. Copy number is depicted by black horizontal bars.

The high frequency of replication initiation in rapidly growing cells leads to an enrichment of origin-proximal loci which could explain the pronounced increase in the number of partners observed in this region in exponentially growing cells (Figure 2.4A). By contrast, treatment with SHX reduces this bias (Figure 2.4B). These results are consistent with the inhibition of replication initiation after SHX treatment leading to a reduction in the Ori:Ter copy number ratio (Ferullo & Lovett, 2008) or structural alterations within the origin domain.

To investigate whether interaction frequencies are affected by differences in copy number across the bacterial chromosome due to DNA replication, we compared interaction patterns and copy number before and after SHX treatment. Interactions were grouped according to the linear distance between the interacting loci and occurrence in the different environmental conditions (Figure 2.3B and C). The distribution of interaction strength and copy number relative to the origin was determined (Figure 2.3C-K). Exponential phase-specific and shared short distance interactions correlate with copy number (Figure 2.3C, D and F). By contrast, SHX-specific or shared long distance interactions do not correlate with copy number (Figure 2.3E, G and H). Critically, the ratio of Ori to Ter regions within both the exponential and SHX conditions remains at 3:1 (compare copy number Figure 2.3C and E). Thus, the observed decrease in the frequency of the interactions within the origin domain (compare Figure 2.1B and C) is either due to a decrease in the absolute number of origin sequences or to a structural alteration (*e.g.* expansion) of the Ori domain.

Correcting the frequency of long distance interactions by copy number, a feature of GCC, indicates that most genomic regions interact with similar frequencies within the exponential-specific and shared interaction sets (*i.e.* interactions that occur in both the exponential and SHX conditions; Figure 2.3I and J). However, there are several notable deviations from this trend (labeled peaks within Figure 2.3I and J). The observed deviations are due to interactions involving multiple fragments within each of the 10,000 bp segments that are plotted (Figure 2.3I and J). By contrast, copy number correction of the long distance SHX specific interactions identifies an increase in the interaction frequency within the Ter domain. The remainder of the genome shows relatively even and low interaction frequencies within the SHX specific interaction set (Figure 2.3K).

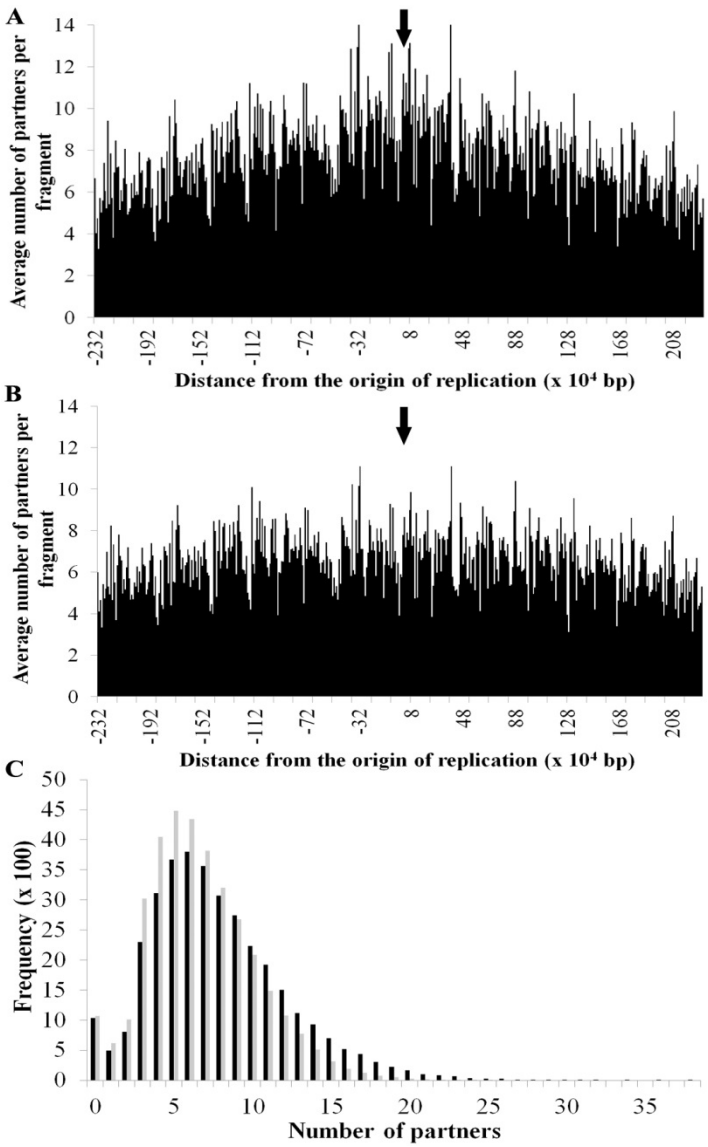


Figure 2.4. There is an origin preferential reduction in the number of partners each restriction fragment interacts with in the SHX treated cells.

The distribution of the number of partners per restriction fragment across the *E. coli* genome averaged for 10,000 bp bins for exponentially grown A) and SHX treated B) cells. There is a visible increase in partner number towards the origin of replication (black arrow). C) There is no inter-condition difference in the mean number of unique partners per restriction fragment (exponential, black bars; SHX, grey bars; T-Test). Adjacent and self-interactions were included in these calculations.

2.2.3 CLUSTERING OF MATP AND SEQA BINDING SITES LINK NUCLEOID STRUCTURE AND REPLICATION

To further investigate the link between replication and nucleoid organization we determined the clustering and interaction properties of loci containing characterized protein binding sites for the MatP, SlmA and SeqA proteins.

MatP is a protein that binds to matS sites and organizes the Ter macrodomain (Mercier et al., 2008). Analyses of matS loci identify significantly ($p < 0.008$) high clustering (*i.e.* inter-matS loci interactions) within the exponentially growing cells (Table 2.2). In contrast clustering of matS sites was not detected in the SHX treated cells. The clustering in the exponentially growing condition was attributed to a single specific interaction between matS10 and matS5 (Figure 2.5A). This interaction must result from intra- or inter-Ter associations of these matS sites (Figure 2.5A i-iv).

The finding that SeqA binds as a dimer, which multimerizes to form a left handed filament (reviewed in (Waldminghaus & Skarstad, 2009)), suggests that this protein may link spatially separated binding sites. Clustering of the 135 strongest confirmed SeqA binding sites present within exponentially growing *E. coli* (Sánchez-Romero et al., 2010) was significantly higher than the random set ($p<0.05$) (Table 2.3A). Moreover, these sites are significantly more prone to interact with other loci than random sites ($p<0.05$; Table 2.3A). Visualizing the positions of the SeqA-SeqA interactions that formed within the *E. coli* genome showed that they tend to occur towards, and involve, the Ori domain in exponential cells (Figure 2.5B and C). SeqA interactions that are shared between exponential and SHX treated nuclei predominantly link the left and right replichores (Figure 2.5C). By contrast, cells treated with SHX have a reduction in clusters involving SeqA sites surrounding the Ori domain and more inter-replichore interactions towards the terminal domain (Figure 2.5C and D). This is consistent with the progression of active replication forks that were initiated prior to SHX treatment.

Table 2.2. MatS loci were highly clustered in exponentially growing cells.

The total number of interactions and clustering for matS sites (+/- 50 bp) was determined in the exponential and SHX treated interaction data sets. Significant clustering of the matS sites was detected in the exponential data set but no clustering was detected in the SHX data set. The total number of interactions was Not Changed (NC) relative to the random data sets. nd, not detected, **p<0.001.

		Clustering		Interactions	
NAP	Interaction set	RS	CLS	RS	CLS
matP	exp	High**	High**	NC	NC
	SHX	nd	nd	NC	NC

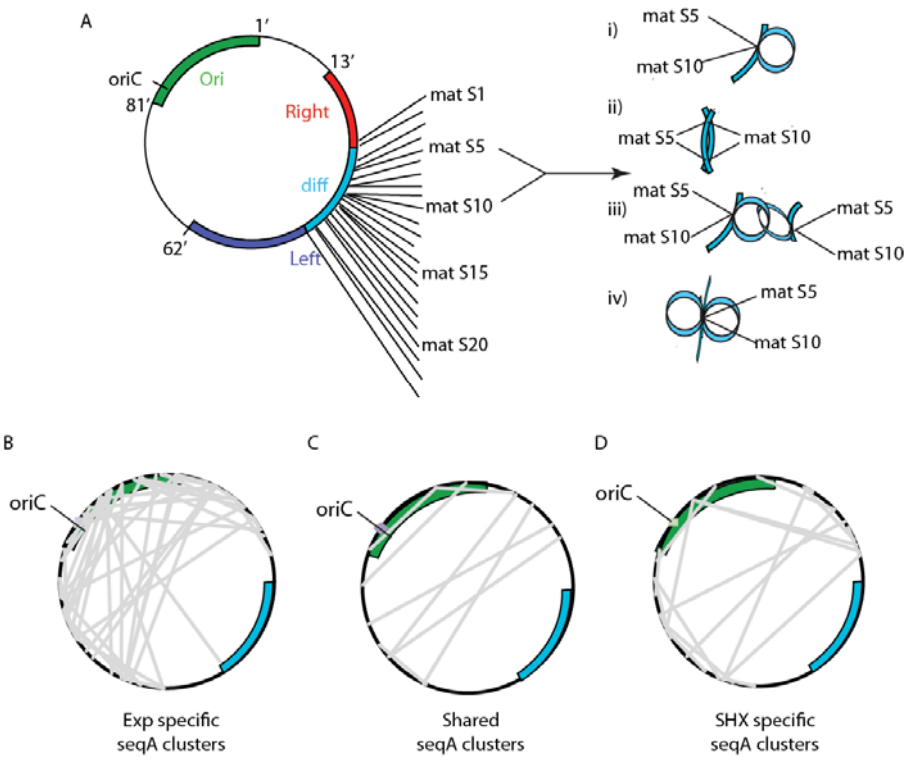


Figure 2.5. Binding sites for nucleoid associated proteins MatP and SeqA exhibit differing degrees of spatial clustering within the exponential and SHX treated E. coli nucleoids.

A) Regions that centred on matS binding sites (+/-50bp; (5)) show significantly increased clustering in the exponential condition, despite having interaction levels that were no different from random (Table 2.2). MatS site clustering is confined to two matS sites: matS5 and matS10 and may result from: A i) intra-chromosome interactions, or A ii-Iv) inter-chromosomal interactions. Critically, this clustering is not observed in the SHX treated nucleoid. B) Exponential specific spatial clustering of SeqA binding sites was concentrated around the origin. C) Spatial clusters of SeqA binding sites that were shared between conditions tended to occur between the left and right replichores. D) SHX specific interactions involved fewer SeqA binding sites and tended to be towards the terminus (Table 2.3).

.

A

		Clustering		Interactions	
NAP	Interaction set	RS	CLS	RS	CLS
SeqA	exp	High*	High	High**	High**
	SHX	High	High	High*	High*

B

SeqA locus			Interacting SeqA partner		
Gene	Start position	End position	Gene	Start position	End position
selB	3756040	3757884	ulaG	4416584	4416584
uraA	2616893	2618182	scpA	3058872	3058872
frlB	3499290	3500312	slt	4628756	4628756
kefA	485760	489122	ispG	2638708	2638708
gltL	683753	684478	ebgA	3220655	3220655
betA	324801	326471	yjfC	4411051	4411051
prpE	351930	353816	recB	2950483	2950483
galE	790262	791278	recB	2950483	2950483

C

SeqA locus			Interacting SeqA partners								
			First partner			Second partner			Third partner		
Gene	Start position	End position	Gene	Start position	End position	Gene	Start position	End position	Gene	Start position	End position
yhhJ	3623702	3624826	phnD	4321359	4321359						
hyfR	2609922	2611934	frlB	3499290	3499290						
sucA	757929	760730	smtA	972760	972760						
dmsA	940182	942626	ptsA	4137743	4137743						
potI	896307	897152	ptsA	4137743	4137743						
yghJ	3112572	3117134	frlB	3499290	3499290						
scpA	3058872	3061016	uxaA	3239849	3239849						
aaeB	3384243	3386210	dppA	3704121	3704121						
ypdE	2502532	2503569	parE	3171526	3171526						
betI	327971	328558	ybiW	859397	859397						
yghO	3127065	3128237	katG	4131858	4131858						
glpA	2350669	2352297	hda	2616097	2616097						
ybfF	711261	712025	uxaA	3239849	3239849						
prpE	351930	353816	maeB	2574120	2574120	ddlA	399053	399053	mukF	973542	973542
yfiQ	2717975	2720635	yghJ	3112572	3112572						
cusA	597937	601080	cysI	2886409	2886409	slt	4628756	4628756			

D

SeqA locus			Interacting SeqA partners											
			First partner			Second partner			Third partner			Fourth partner		
Gene	Start position	End position	Gene	Start position	End position	Gene	Start position	End position	Gene	Start position	End position	Gene	Start position	End position
yhhJ	3623702	3624826	psiE	4238348	4238348	hemN	4050068	4050068						
selB	3756040	3757884	psiE	4238348	4238348									
malZ	421739	423556	cusA	597937	597937									
araA	66835	68337	yahF	336002	336002									
hyfR	2609922	2611934	fadH	3229687	3229687									
nfrA	587205	590177	yjfC	4411051	4411051	nanA	3370705	3370705						
ebgA	3220655	3223747	dppA	3704121	3704121	pyrI	4469009	4469009						
ushA	504138	505790	smtA	972760	972760	mukF	973542	973542						
yahF	336002	337549	recB	2950483	2950483									
yfaL	2338439	2342191	scpC	3062824	3062824									
afuC	276980	278026	maeB	2574120	2574120									
scpA	3058872	3061016	ebgA	3220655	3220655	ybjH	4235657	4235657						
fryA	2500012	2502507	degS	3380222	3380222	ygiQ	3156949	3156949						
agaR	3275878	3276687	aceB	4213501	4213501									
hepA	60358	63264	sgrR	75644	75644	aaeB	3384243	3384243	scpC	3062824	3062824	hda	2616097	2616097
kefA	485760	489122	polA	4044989	4044989									
sgrR	75644	77299	afuC	276980	276980	uraA	2616893	2616893	hda	2616097	2616097			
gltL	683753	684478	mngB	767201	767201	slt	4628756	4628756						
md	1884888	1886015	ptsA	4137743	4137743									
betI	327971	328558	dnaA	3880349	3880349									
uxaA	3239849	3241336	phnD	4321359	4321359	kbaY	3281165	3281165						
yghO	3127065	3128237	sthA	4157413	4157413	frlB	3499290	3499290						
ybbO	517564	518373	ytfR	4449081	4449081									
hda	2616097	2616798	yjiN	4563989	4563989									
nanA	3370705	3371598	frlB	3499290	3499290	ptsA	4137743	4137743						
aroG	784856	785908	aslA	3982375	3982375									
glpA	2350669	2352297	pbpC	2643035	2643035									
betA	324801	326471	ybbO	517564	517564									
dinG	832293	834443	yjfC	4411051	4411051									
nrdB	2345406	2346536	ptsA	4137743	4137743									
prpE	351930	353816	cysI	2886409	2886409	ptsA	4137743	4137743	dnaC	4598261	4598261			
yfiQ	2717975	2720635	nfi	4196813	4196813									
pbpC	2643035	2645347	phnD	4321359	4321359									
miaB	692754	694178	scpC	3062824	3062824									

Table 2.3. SeqA sites cluster within the *E. coli* nucleoid. Refer to footnote [2] for the table legend.

[2] A) The total number of interactions and clustering for the strongest 135 confirmed SeqA binding sites was determined in the exponential and SHX treated interaction data sets. Significant clustering of the SeqA sites was detected in the exponential data set for the RS random sampling but not for the CLS random sampling and not in the SHX treated data set. The total number of interactions was significantly higher in both the exponential and SHX treated data sets. High $p<0.1$, *$p<0.05$, **$p<0.01$. B) Identities of the SeqA sites involved in clustering in

SlmA binds at 24 defined sites within the genome (Cho et al., 2011) and acts to prevent FtsZ polymerization and premature cell division prior to complete chromosome replication. Analyses of the clustering and interaction profiles of *E. coli* SlmA sites demonstrated that clustering of these sites was not different from that observed for randomly selected sites (Table 2.4). However, SlmA sites did exhibit a significantly increased propensity to interact with other genomic loci ($p<0.05$) compared to randomly spaced elements for both exponential and SHX treated cells (Table 2.4). The significant increase in interaction frequency was lost when comparisons were made with random sets that have conserved linear spacing (Table 2.4). Note that the differences observed in significance when the test data set was compared to randomly generated data sets (*i.e.* random spacing (RS) or conserved linear spacing (CLS)), confirms that the linear spacing of *E. coli* loci is important. Whether this is an effect or cause of spatial organization remains to be determined.

2.2.4 Intra- or inter-NAP binding site clustering does not contribute to the global organization of the *E. coli* nucleoid.

We investigated the clustering and interaction properties of H-NS, IHF and Fis binding sites which are not enriched in any particular macrodomain. There is no detectable clustering for the 200 bp regions surrounding the Fis, H-NS, and IHF binding sites in either the exponential or SHX treated nucleoids (Table 2.5A). Moreover, the classical NAP binding sites have depleted levels of interactions in exponentially growing *E. coli* cells (Table 2.5A). These results can be explained by restrictions in the flexibility of the DNA (and hence reduced ligation efficiencies) due to the binding of the NAP. However, increasing the length of the region surrounding the binding site has no effect on the clustering (data not shown). Additionally, we do not observe intra-NAP binding site clustering (Table 2.5B), consistent with the temporal isolation of the expression of these NAPs (Azam, Iwata, Nishimura, Ueda, & Ishihama, 1999).

---

the shared interactions (plotted in Figure 2.3C). C) Identities of the SeqA sites involved in clustering in the SHX specific interactions (plotted in Figure 2.3D). D) Identities of the SeqA sites involved in clustering in the exp specific interactions (plotted in Figure 2.3B). Clustering levels for B-D) were calculated using non-adjacent interactions that occurred at levels higher than the false detection rate. Several SeqA loci were identified as being involved in interactions with multiple other SeqA loci (*i.e.* First-Fourth partners). The GCC assay does not allow us to determine if interactions with multiple other SeqA loci occurred at the same time within a single complex. Interactions are only shown in one direction (*i.e.* SeqA locus to partner 1 and not partner 1 to Seq A).

Table 2.4. SlmA is highly connected with the genome but not with other SlmA sites.

The total number of interactions and clustering for the 24 defined SlmA binding sites was determined in the exponential and SHX treated interaction data sets. Clustering of the SlmA sites was Not Changed (NC) compared to random data sets in both the exponential and SHX treated data sets. The total number of interactions was significantly higher in both the exponential and SHX treated data sets apart from the CLS random sampling comparison for the SHX treated data set. High p<0.06, *p<0.05.

		Clustering		Interactions	
NAP	Interaction set	RS	CLS	RS	CLS
SlmA	exp	NC	NC	High*	High*
	SHX	NC	NC	High*	High

Table 2.5. H-NS, IHF and Fis sites do not exhibit spatial clustering.

A) The summed interaction or clustering strength for each condition was compared to 1,000 random data sets with random spacing (RS) or conserved linear spacing (CLS) between each element (see Methods). Clustering of characterized H-NS, IHF and Fis binding sites (+/-100bp; (18)) located within coding and non-coding sequences was not detected. H-NS, IHF and Fis binding sites had significantly lower interaction frequencies than random. However, treatment with SHX altered the interaction frequencies of genic H-NS and Fis sites such that they were no longer different from random. Short distance (<200 bp) and self-interactions were excluded from these analyses. B) Clustering of loci that contained one or more characterized H-NS, IHF and Fis sites (+/- 500 bp) was no different to random (coding loci) or lower than random (non-coding loci) confirming that these elements do not cluster either individually or collectively.

A

		Genic regions				Non-genic regions			
		Clustering		Interactions		Clustering		Interactions	
NAP	Interaction set	RS	CLS	RS	CLS	RS	CLS	RS	CLS
H-NS	exp	nd	nd	Low**	Low**	nd	nd	Low**	Low**
	SHX	nd	nd	NC	NC	nd	nd	Low**	Low**
IHF	exp	nd	nd	Low**	Low**	nd	nd	Low**	Low**
	SHX	nd	nd	Low	Low	nd	nd	Low**	Low**
FIS	exp	nd	nd	Low	Low	nd	nd	Low**	Low**
	SHX	nd	nd	NC	NC	nd	nd	Low**	Low**

B

		Clustering		Interactions	
NAP (500 bp)	Interaction set	RS	CLS	RS	CLS
FIS, H-NS, IHF coding	exp	NC	NC	Low**	Low**
	SHX	Low	NC	Low*	Low*
FIS, H-NS, IHF non-coding	exp	Low**	Low	Low**	Low**
	SHX	Low**	NC	Low**	Low**

Key
** p<0.001
* p<0.05
Low p<0.085
NC no change
nd not detected

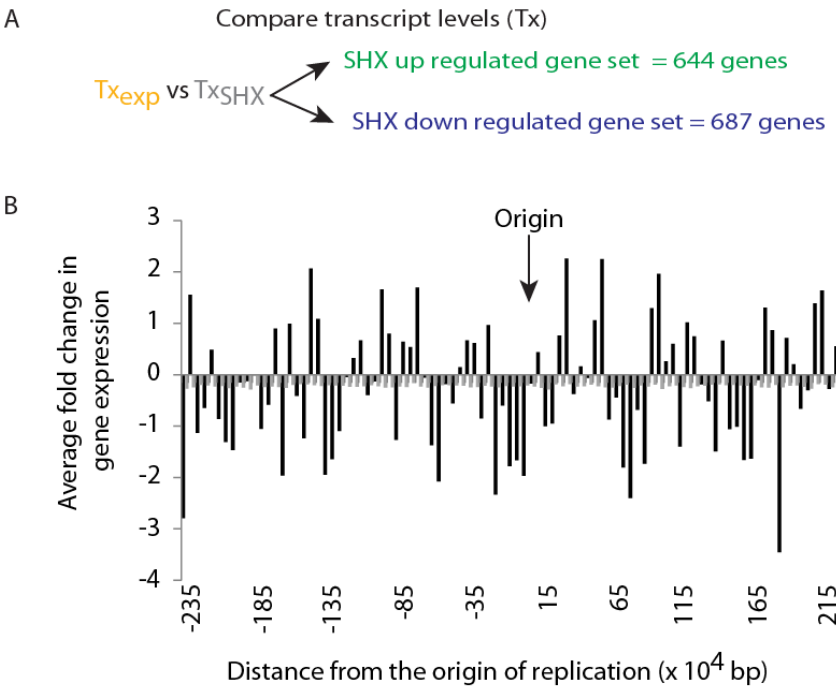
2.2.5 GENES UP OR DOWN REGULATED FOLLOWING SHX TREATMENT EXIST IN DIFFERENT SPATIAL ENVIRONMENTS CONFIRMING FUNCTIONAL COMPARTMENTALIZATION OF THE NUCLEOID.

Eukaryotic studies have identified a non-random distribution of gene expression associated with the presence of spatially distinct environments that promote or inhibit nuclear functions (*e.g.* (Cook, 1999; Iborra, Pombo, Jackson, & Cook, 1996; Versteeg et al., 2003)). Similarly, we observe that *E. coli* genes whose transcript levels increased or decreased in response to SHX treatment are over represented in some gene ontology terms (Table 2.6) and are non-randomly distributed across the linear genome (Figure 2.6A & B) in a manner that does not correlate with GC content (Figure 2.7A). There is no correlation between transcript level and interaction frequency at the level of specific restriction fragments (Figure 2.7B and C). However, the SHX down-regulated genes have high average transcript ($p<0.001$; Table 2.7), clustering and interaction (Figure 2.6C) levels in exponential phase cells. These results suggest that genes that are highly expressed in exponential phase and down-regulated following SHX treatment are not only linearly but also highly spatially clustered. In conjunction with microscopic observations of large RNA polymerase (RNAP) clusters (foci) within exponentially growing *E. coli* cells (Cabrera & Jin, 2003), our results support the hypothesis that the highly expressed exponential phase genes are associated with transcription foci. Despite this, genes down-regulated in response to SHX treatment ($p<0.001$; Table 2.7) remained highly clustered (Figure 2.6C). Similarly, up-regulated genes within lowly clustered regions do not increase their clustering upon activation (Figure 2.6C). As such, the maintenance of the clustering is independent of transcript levels and *ipso facto* transcription.

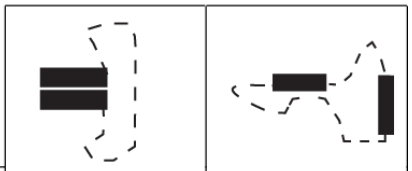
Table 2.6. Genes up and down regulated in response to SHX treatment were enriched in specific Gene Ontology terms.

An R package for the analysis of Gene Ontology (GO) term associations called GOStats was used to determine enriched GO terms in the up and down regulated gene sets. The table shows the significantly enriched GO terms in the up (top) and down (bottom) regulated gene sets.

GOBPID	P-value	OddsRatio	Expected Count	Count	Size	Term
Genes up regulated in response to SHX treatment						
GO:0016052	1.17e-06	2.3872	31.0473	56	274	carbohydrate catabolic process
GO:0051716	0.0004	2.6589	9.8580	21	87	cellular response to stimulus
GO:0006950	0.0019	2.2030	12.4642	23	110	response to stress
GO:0007154	0.0025	4.3016	2.6061	8	23	cell communication
GO:0009432	0.0025	4.3016	2.6061	8	23	SOS response
GO:0071496	0.0025	4.3016	2.6061	8	23	cellular response to external stimulus
GO:0006281	0.0031	2.9597	5.0990	12	45	DNA repair
Genes down regulated in response to SHX treatment						
GO:0034641	2.69e-10	2.3881	76.9565	121	402	cellular nitrogen compound metabolic process
GO:0009987	2.07e-08	2.3904	299.9339	338	1420	cellular process
GO:0009451	5.84e-08	5.4087	9.5049	26	45	RNA modification
GO:0009228	1.71e-06	21.1152	2.7458	11	13	thiamine biosynthetic process
GO:0042723	1.71e-06	21.1152	2.7458	11	13	thiamine-containing compound metabolic process
GO:0006259	9.18e-06	2.6646	20.4960	39	99	DNA metabolic process
GO:0009058	1.20e-05	1.6473	145.5313	182	689	biosynthetic process
GO:0006766	4.02e-05	3.7129	9.0825	21	43	vitamin metabolic process
GO:0044260	8.67e-05	1.6409	98.0507	127	500	cellular macromolecule metabolic process
GO:0044238	9.59e-05	1.6612	273.7425	303	1296	primary metabolic process
GO:0042559	0.0001	4.8948	5.2805	14	25	pteridine-containing compound biosynthetic process
GO:0044271	0.0001	1.7897	52.0022	75	252	cellular nitrogen compound biosynthetic process
GO:0006261	0.0001	3.0516	10.5610	22	50	DNA-dependent DNA replication
GO:0009108	0.0006	2.6234	12.0396	23	57	coenzyme biosynthetic process
GO:0019438	0.0007	3.4946	6.6034	15	32	aromatic compound biosynthetic process
GO:0000270	0.0007	2.7074	10.7722	21	51	peptidoglycan metabolic process
GO:0006022	0.0007	2.7074	10.7722	21	51	aminoglycan metabolic process
GO:0006308	0.0009	3.8301	5.4917	13	26	DNA catabolic process
GO:0018130	0.0013	1.9268	24.9693	39	121	heterocycle biosynthetic process
GO:0055086	0.0015	1.8559	27.6699	42	131	nucleobase-containing small molecule metabolic process
GO:0006807	0.0017	2.3493	11.9157	22	78	nitrogen compound metabolic process
GO:0006596	0.0023	4.2335	4.0132	10	19	polyamine biosynthetic process
GO:0006760	0.0027	6.6379	2.3234	7	11	folic acid-containing compound metabolic process
GO:0009246	0.0027	6.6379	2.3234	7	11	enterobacterial common antigen biosynthetic process
GO:0046656	0.0027	6.6379	2.3234	7	11	folic acid biosynthetic process
GO:0042364	0.0027	3.0254	6.6762	14	32	water-soluble vitamin biosynthetic process
GO:0009057	0.0032	2.4361	10.3498	19	49	macromolecule catabolic process
GO:0006575	0.0037	4.278	3.5907	9	17	cellular modified amino acid metabolic process
GO:0090304	0.0046	1.762	25.6893	38	131	nucleic acid metabolic process
GO:0009165	0.0057	2.1179	13.0957	22	62	nucleotide biosynthetic process



C



Gene set	Interaction set	Clustering RS	Clustering CLS	Interactions RS	Interactions CLS
SHX up regulated	exp specific	NC	NC	NC	NC
	SHX specific	NC	Low	Low	Low
	exp shared	Low*	Low*	NC	NC
	SHX shared	Low*	Low*	NC	NC
SHX down regulated	exp specific	High**	High**	High**	High**
	SHX specific	High**	High**	High**	High**
	exp shared	High*	High*	High**	High**
	SHX shared	High	High*	High*	High**

Key High = increased clustering or interactions (p<0.1)
Low = reduced clustering or interactions (p<0.1)
** p<0.001 * p<0.05 NC no change
Locus

Figure 2.6. Annotated genes with transcripts that were up (644 genes) or down (687 genes) regulated following SHX treatment existed in different spatial environments. Refer to footnote [3] for figure legend.

[3] A) Genes that changed transcript level (Tx) following treatment with SHX were identified. B) Analyses of positions of the up and down regulated genes across the *E. coli* genome identify non-random clustering within the linear sequence. Average expression levels were calculated for 50 Kb bins. Grey bars indicate the average expression across 50 Kb bins within a thousand randomized genomes. Autocorrelation analyses on the distribution of gene expression data across the genome demonstrated a strong predictive relationship up to 32 genes away (ACF

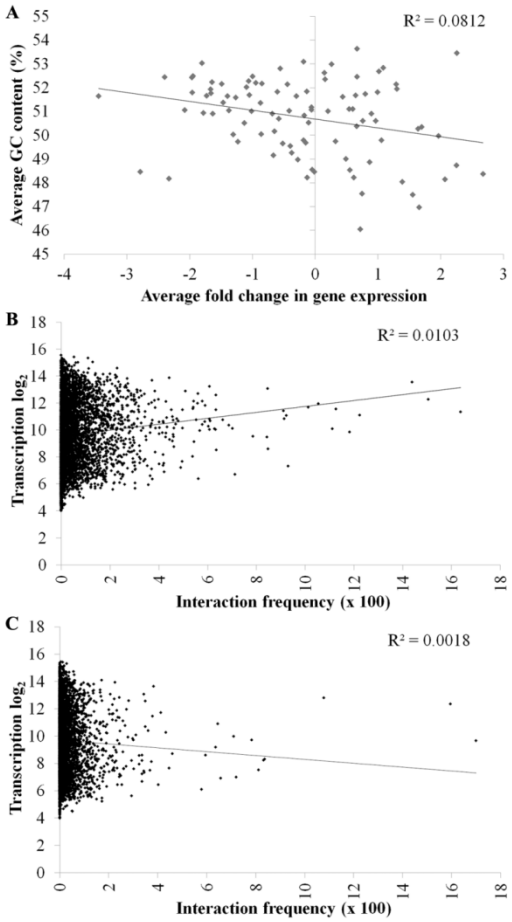


Figure 2.7. Genes whose transcript level changes upon SHX treatment did not correlate with genomic GC content and the raw transcript levels did not correlate with interaction frequency.

A) Change in gene expression does not correlate with GC content. The genome was divided into 50 Kb bins and the average change in gene expression and GC content calculated for each bin. The transcript levels (log2) of all annotated genes did not show any clear linear correlation with interaction frequency in both B) exponential phase and C) SHX interaction datasets (plots were generated from long distance interaction data [>800 bp]).

>0.83). C) Clustering and interaction patterns for up or down regulated genes demonstrate that up and down regulated genes occupy specific spatial environments. The amount of clustering within the up or down regulated gene sets, and between the up or down regulated genes and other loci, was compared to 1,000 randomly generated sets. 1,000 random sets of equivalent size (number and length) to the up or down-regulated sets were generated such that they i) randomized the spacing between elements (RS) or ii) conserved the linear spacing between the elements (CLS) involved in the interactions. Clustering and interaction counts were determined individually for the condition specific and shared data sets. Clustering and interaction data are shown for both exponential (exp) and SHX shared interaction sets because despite the interaction being shared the clustering or interaction frequency was specific for each condition. These analyses were performed on long distance interactions only.

2.3 DISCUSSION

The *E. coli* nucleoid has a complex structure that emerges from the sum of the cellular processes that occur within the bacterial cell. We identified two macrodomains within the *E. coli* chromosome interaction networks corresponding to the Ori and Ter domains that have been previously identified (Boccard et al., 2005; Espeli et al., 2008; Mercier et al., 2008; Niki et al., 2000; Thiel et al., 2012; Valens et al., 2004). However, the two remaining macrodomains (Left (L), Right (R)) and the two non-structured domains (NS) are not obvious within our data. Moreover, we did not identify hard boundaries surrounding either the Ori or Ter domain, consistent with earlier predictions (Boccard et al., 2005; Valens et al., 2004). It remains possible that the L, R and NS domains, and the domain boundaries were obscured due to the use of an unsynchronized population of cells. Alternatively the formation of the macrodomains and the previously observed reductions in inter-domain recombination rates (Valens et al., 2004) could be achieved by a combination of mechanisms of which physical segregation is only one component. This explanation is supported by the observation that a low level of connectivity remains between the Ter and Ori domains. Critically, this connectivity occurs at levels above those observed for random inter-molecular ligation under our experimental conditions and indicates that while these domains are largely separated there is some inter-domain mixing during the cell cycle. This is consistent with the observation that recombination rates between lambda att sites are reduced but not completely abolished between these domains (Valens et al., 2004).

The chromosome interaction networks we identified within both exponential and SHX treated *E. coli* cells contain variable numbers of short and long distance loops. The observation that the number of long distance interactions (long distance loops) reduced following treatment with SHX can be interpreted as indicating that the nucleoid expands under this condition, consistent with microscopic observations (Cabrera et al., 2009; Cabrera & Jin, 2003; Kuhlman & Cox, 2012). Either the observed expansion is specific and directed as part of the stress response or it is a non-specific consequence of SHX acting on the factors that mediate the interactions (*e.g.* rapid protein turn-over with no replacement). The exact reasons for the loss of interactions remain to be determined. However, the fact that SHX specific interactions form indicates a directed alteration in nucleoid organization.

Table 2.7. Genes that were highly expressed in exponential (exp) cells became weakly expressed in SHX treated cells and vice versa.

The summed raw transcript level for the exp or SHX transcription microarray data for genes that were significantly up (SHX/exp log2 ratio >1.5) or down (SHX/exp log2 ratio <-1.5) regulated after SHX treatment was compared to that of 1,000 randomly generated data sets containing the same number and length (bp) elements as the test set in question. Annotated *E. coli* gene transcripts which showed log2 changes of >1.5 upon treatment with SHX were included in the analysis. SHX, serine hydroxamate.

		Mean expression (per 20 kbp bin)		Expression variance (sum all bins)	
E. coli condition	Transcript N°. change on SHX shift	Test sample	Random sample (range)	Test sample	Random sample (range)
Exponential	Down	13,129 (High) p = < 0.001	6,910 – 10,337	225,684,597 p = 0.113	72,698,651 – 422,279,592
	Up	2,447 (Low) p = < 0.001	6,184 – 9,426	6,547,129 (Low) p = < 0.001	57,500,836 – 441,895,413
SHX	Down	2,271 (Low) p = < 0.001	6,464 – 9,992	9,293,946 (Low) p = < 0.001	58,259,459 – 372,181,089
	Up	12,733 (High) p = < 0.001	5,803 – 9,185	174,870,395 p = 0.154	41,320,084 – 388,277,906

2.3.1 IS THE *E. COLI* NUCLEOID SHAPED AS A SAUSAGE OR ROSETTE?

The presence of short and long distance loops within both networks points to the *E. coli* genome folding into a series of DNA loops connected to a central node (*i.e.* a rosette). This interpretation agrees with electron microscope observations of isolated nucleoids (reviewed in (Ishihama, 2009)). However, our observation that the Ter region has few contacts with itself (*i.e.* is extended in nature) and is less well connected to the remainder of the genome is consistent with previous observations made by David Sherratt's group (X. Liu et al., 2010; X. Wang, Liu, Possoz, & Sherratt, 2006). Therefore, despite differences in growth rate between the studies (X. Liu et al., 2010), our data also supports the hypothesis that the *E. coli* chromosome is organized as a sausage in which the bulk of the chromosome is organized into a compacted rod that is circularized by the Ter domain (Figure 2.8A; (X. Liu et al., 2010; X. Wang et al., 2006)). The apparent dichotomy of these interpretations is reconcilable through the realization that the isolation of a sausage shaped genome during preparation for electron microscopy would result in the appearance of a rosette. Thus, the sausage model is a variation of the

rosette model where the rosette is flattened through confinement or as a result of the biological processes within the live cell.

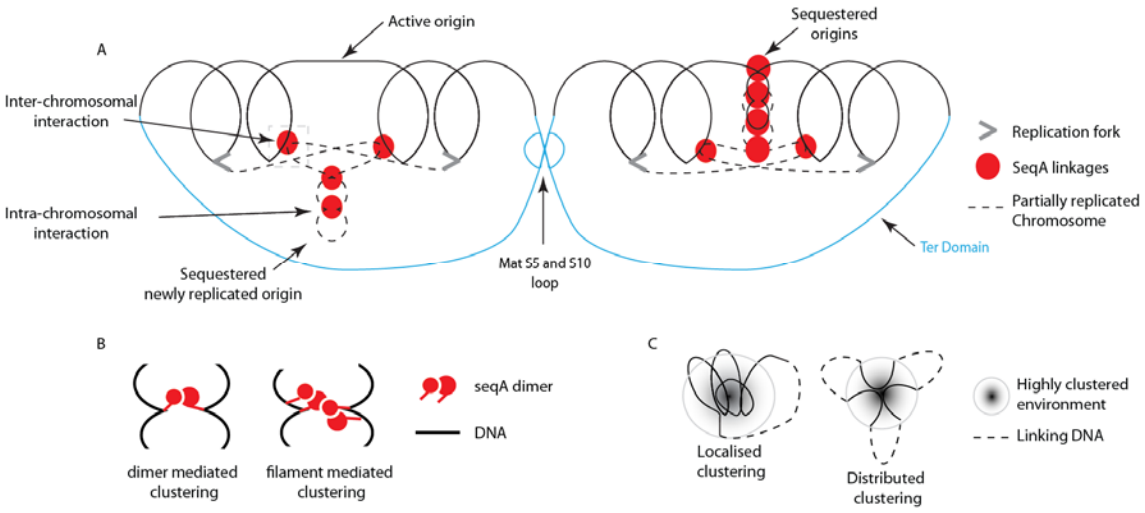


Figure 2.8. Spatial model of exponential phase nucleoid organization in *E. coli*.

A) The exponential phase *E. coli* nucleoid is organized into high interacting domains by nucleoid associated factors including, but not limited to, SeqA and MatP. SeqA promotes the intra- and inter-chromosomal clustering of hemi-methylated GATC sites in order to sequester recently replicated origins and contribute to chromosome segregation. Newly replicated origins can be sequestered individually (left) or through interactions between the recently replicated origins (right). The matS5-10 loop is hypothesized to form between chromosomes that have almost completed replication. The model illustrates some of the major findings in the study but for simplicity overlap between replichores has not been included in this cartoon. Similarly, only one replication process has been illustrated on each chromosome. Moreover, as our data is drawn from an unsynchronized population and we do not have data on the relative positions of the different elements through the cell cycle, we have not attempted to represent the dynamic nature of the positioning of the different elements through-out the cell cycle. B) SeqA can mediate interactions within or between chromosomes as either a dimer or filament. C) Highly clustered regions form as a result of localized and distributed clustering within and/or between the replicated chromosomes.

2.3.2 Replication contributes to nucleoid organization through SeqA.

The SeqA and SlmA proteins are implicated in the regulation of replication and chromosome separation (reviewed in (Dame et al., 2011)). Our results indicate that SlmA binding sites do not cluster as part of nucleoid occlusion during replication initiation or extension. Therefore, the dimerization necessary to activate SlmA occurs at a single or linearly-adjacent binding site(s) but does not result from spatial associations of distant SlmA sites. Consistent with the supposition by Dame et al. (Dame et al., 2011), the low levels of SlmA clustering observed indicate that any contribution that SlmA-FtsZ makes to nucleoid structure must be facilitated by tethering to an external framework (*e.g.* shortened preformed FtsZ polymers (Cho

et al., 2011), or non-functional protofilaments (Tonthat et al., 2011)) or the cell membrane.

By contrast, the replication dependent nature and distribution of the exponential phase SeqA mediated long distance interactions provides support for a role for SeqA clustering in the formation of an intra- and/or inter-chromosomal structure (Figure 2.8A and B). This is particularly true for SeqA interactions that form over the origin of replication and could function to sequester newly replicated origins and delay chromosome separation ((Bach et al., 2003; Lu et al., 1994; von Freiesleben, Rasmussen, & Schaechter, 1994), reviewed in (Dame et al., 2011; Waldminghaus & Skarstad, 2009)). As such, the SHX dependent loss of the long distance interactions is predicted if replication and segregation occur consecutively (Nielsen et al., 2007). Thus the loss of SeqA mediated interactions within the SHX treated nucleoid reflects an underlying spatial segregation of the replicated chromosome regions (Ferullo & Lovett, 2008). The predominance of SeqA clusters between loci that are approximately equidistant from the Ori within the SHX specific and shared interaction datasets represent links between the hemimethylated GATC sites trailing the replisome. We interpret the distinct subset of inter-replichore SeqA clusters as indicating that the DNA polymerases are pausing at specific genomic sites within the cell populations. Finally, there is no correlation between alterations to transcript levels and SeqA clustering (data not shown) therefore, SeqA clustering is independent of transcription. Collectively, these results support a strong linkage between replication and nucleoid organization (X. Liu et al., 2010).

For ease of visualization, the chromosomal interactions that we identified are presented as intra-chromosomal connections (Figure 2.1). This form of presentation is problematic as the proximity-based ligation data is probabilistic and represents a population average from unsynchronized cells (Grand et al., 2011). As such, it is impossible to determine which combinations of interactions occur within a single nucleoid. Secondly, while the sequences we obtain as part of the GCC protocol identify the interacting loci they do not provide information on whether the interactions occur within or between the chromosome(s). This is an important consideration when investigating nucleoid structure in exponential phase bacterial cells that contain and segregate partially replicated chromosomes (Sherratt, 2003). Therefore, it is possible that the formation of long distance SeqA dependent and independent interactions can be facilitated by overlaps between the

replichore arms that result from the chromosome alignment (*i.e.* inter-chromosomal (Figure 2.8A right)). Interestingly, such a system may contribute to gene dosage control as well as the control of chromosome segregation. However, it remains possible that interactions also occur within a chromosome (*i.e.* intra-chromosomal (Figure 2.8A left)). Future work should determine the contribution of inter- and intra-chromosomal interactions to the structure of the nucleoid in exponentially growing *E. coli* with a view to understanding how structure contributes to gene dosage control in this organism.

2.3.3 What role does the matS5-10 loop play in nucleoid organization?

MatS sites have a role in defining the Ter domain (Mercier et al., 2008; Thiel et al., 2012). *In vivo* experiments indicate that the definition of the Ter domain and condensation of this region are separable events with the condensation dependent on the presence of the MatP C-terminal coiled-coil domain which is responsible for tetramerization and looping (Dupaigne et al., 2012). We found that the matS5 and matS10 sites form a specific loop that surrounds the TerA site (1,339,796-1,339,791 bp) and is located away from the dif site (1,589,000 bp) towards the Ori on the right replichore. Note that matS5 is one of two matS sites (the other being matS21) that do not show *in vivo* MatP binding in an *E. coli* K12 derivative of MG1655 (Mercier et al., 2008). The question thus arises as to what contribution the matS5-10 interaction makes to the Ter domain structure and function. It is possible that the matS5-10 loop explains observations of a spatially separable, condensed region within the centre of the Ter linker domain (X. Liu et al., 2010). Furthermore, the absence of detectable matS clustering between the other matS loci raises the possibility of differentiation in the functions of the matS sites. However, further experiments are required to confirm these hypotheses and identify how or if MatP contributes to the formation of the matS5-10 loop.

2.3.4 Do "histone-like" NAPs play a role in global nucleoid structure?

The spatial clustering of NAP (*i.e.* H-NS, Fis and IHF) DNA binding sites is not significant within the gross spatial organization of the *E. coli* nucleoid we identified. Rather our results are consistent with the hypothesis that H-NS, IHF and Fis contribute to compaction through localized structuring (reviewed in (Luijsterburg et al., 2006)), gene regulation, or the formation of large protein heterocomplexes (reviewed in (Fang & Rimsky, 2008)). These results are in contrast to those of

Wang et al 2012. who identified H-NS clustering within the *E. coli* nucleoid using microscopic and proximity-ligation based measurements in slow growing early log phase cells (W. Wang et al., 2011). This apparent discrepancy may be due to the significant increase in resolution afforded by the use of the HhaI enzyme in our study. This conclusion is supported by our identification of interactions linking HhaI restriction fragments from within the larger EcoRI restriction fragments that were previously characterized as demonstrating an H-NS dependent association (Figure 2.9 (W. Wang et al., 2011)). Therefore, we propose that the previously recognized relationship between ligation efficiency and the presence/absence of *h-ns* mutants (W. Wang et al., 2011) was likely due to a combination of a global reorganization of localized genome structure (Bouffartigues, Buckle, Badaut, Travers, & Rimsky, 2007) and epistatic effects resulting from H-NS dependent transcriptional changes.

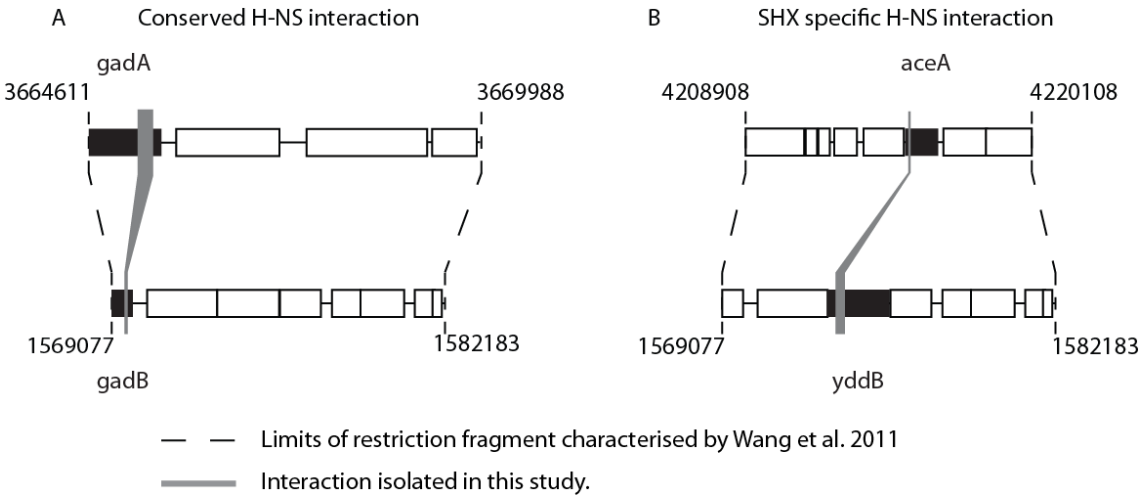


Figure 2.9. Interactions attributed to H-NS clustering identified by Wang et al. 2011 were confirmed as occurring between A) gadA:gadB and B) aceA:yddB (indicated by the grey bars linking the ORFs (black boxes)).

The previously documented connections between the H-NS binding site containing gadA:ydeO and ydeO:arpA gene pairs occurred on large restriction fragments (illustrated as all of the DNA containing multiple ORFs occurring between the broken lines) that contained some H-NS binding sites and a mixture of H-NS regulated and non-regulated genes. Our interactions were fully contained within these large restriction fragments. The widths of the grey bars show the length of the restriction fragments that were characterized as interacting in this study.

2.3.5 DO TRANSCRIPTION FOCI HAVE A ROLE IN NUCLEOID ORGANIZATION?

The observed organization of highly transcribed genes into clustered spatial environments is consistent with the hypothesis that some clustering is occurring around transcription foci (*e.g.* (Carter et al., 2008)). Similarly, the copy-number independent long distance interactions may reflect sequence driven intra-chromosomal nucleoid folding for the coordination of transcription through

enhancer-like interactions consistent with previous observations in bacteria (Dandanell, Valentin-Hansen, Larsen, & Hammer, 1987; Reitzer & Magasanik, 1986; W. Wang et al., 2011) and eukaryotes (*e.g.* (Palstra et al., 2003; Sanyal et al., 2012; Tolhuis et al., 2002)). The existence of these prokaryotic transcription foci is supported by microscopic observations of RNA polymerase foci within *E. coli* cells (Cabrera & Jin, 2003; Cagliero & Jin, 2013; W. Wang et al., 2011). The fact that similar clustering was observed in *P. aeruginosa* (data not shown) and among highly transcribed genes in *Schizosaccharomyces pombe* (Tanizawa et al., 2010) implies that the clustering of highly transcribed genes may be a ubiquitous feature of the control of gene expression.

It is likely that the linear gene clusters (Figure 2.6A) form into combinations of localized and distributed spatial clusters (Figure 2.8C). Given that RNA polymerase is redistributed following SHX treatment (Cabrera et al., 2009; Cabrera & Jin, 2003), decreases in the number of long distance interactions (*i.e.* reductions in extent of distributed clustering) we observed following stress induction could be interpreted as indicating that RNA polymerase mediates some interactions. However, the identification of a core interaction pattern that is conserved within the *E. coli* nucleoid following SHX treatment indicates that at least some of these interactions are stable to a significant redistribution of RNA polymerase. This result agrees with eukaryotic studies that demonstrate long distance interactions are insensitive to inhibition of ongoing RNA polymerase transcription (Palstra et al., 2008). Furthermore, the high levels of clustering and interactions observed at genes that were highly expressed in the exponential phase and subsequently down-regulated by SHX treatment indicates that the localized clustering – but not necessarily the identity of the partners – is stable. However, it remains possible that transcription associated interactions respond slowly to environmental change, allowing for short term fluctuations in environmental conditions without the requirement for major rearrangement of genome organization. This forms an epigenetic memory that is capable of being inherited (Veening, Smits, & Kuipers, 2008) similar to that observed in yeast (D. G. Brickner et al., 2007; J. H. Brickner, 2009, 2010; Laine et al., 2009; Tan-Wong et al., 2009).

2.3.6 Does a nucleolus-like structure form within the *E. coli* nucleoid?

It has been proposed that the formation of transcription factories that include the ribosomal RNA genes and ribosomal protein encoding loci could induce the

compaction of the nucleoid through the formation of a nucleolus-like structure (Cabrera & Jin, 2006; Jin et al., 2012; O'Sullivan, 2011). However, we found no evidence that the nucleoid structure promotes the clustering of ribosomal RNA genes and ribosomal protein encoding loci (data not shown). This may be due to technical limitations in the analysis of repetitive loci that cannot be unambiguously positioned onto the reference genome. Alternatively, it may be due to the very high levels of transcriptional activity at these loci interfering with the cross-linking and ligation steps during the preparation of our chromosome interaction libraries. *In silico* modeling of the nucleoid that incorporates biophysical parameters and interaction frequencies (similar to (Gehlen et al., 2012; Umbarger et al., 2011)) may resolve this issue.

2.3.7 Epistatic interactions and the chromosome interaction network

The bacterial cell is a complex structured entity in which each part exists "for and by means of the whole" (Kauffman, 1995). As such nucleoid structure is an integral – inseparable – part of the cells response to environmental challenge. Moreover, the contribution of any one gene to the bacterial phenotype relies upon its relationship with other genes on levels that include: regulation; transcription; translation; complex formation; and function. Therefore, it is likely that the interaction network we have determined contains information on epistatic relationships between multiple genes that occur at the regulatory, transcriptional and translational levels due to the co-dependence of these processes in *E. coli*. Future work should interrogate prokaryotic interaction networks for evidence of epistatic relationships and must address the mechanism(s) governing the organization of global structure.

2.3.8 Conclusion

The detection of both long and short distance interactions within the *E. coli* nucleoid is consistent with empirical measures and modeling, which indicated that intra-nucleoid interactions play a dominant role in shaping the *E. coli* nucleoid (Wiggins et al., 2010). However, the long distance interactions did not consistently involve loci located equidistant from the Ori on opposite replichores and, therefore, it is unlikely that the *E. coli* nucleoid is preferentially structured as ellipsoids as observed in *C. crescentus* (Umbarger et al., 2011). Rather our study indicates that the chromosome(s) within exponentially fast growing *E. coli* cells are structured by interactions that are linked to the ongoing replication and transcription processes

within the cell. The specificity of the observed interactions identifies spatial organization as a significant factor in bacterial gene regulation and indicates that the spatial clustering of highly regulated genes is a ubiquitous feature of gene regulation.

2.4 MATERIALS AND METHODS

2.4.1 STRAINS AND GROWTH CONDITIONS

For GCC analyses (36), *Escherichia coli* strains (Table 2.8) were recovered from -80°C on Luria Bertani (LB) agar (2%) plates (24 hours, 37°C). LB medium (3 ml, Gibco) starter cultures were inoculated and grown (37°C, 220 rpm, 16 h). The Optical Density (OD_{600}) of cultures was measured and used to inoculate LB test cultures to an OD_{600} of ~0.02. The test cultures were grown (37°C, 220 rpm) until the OD_{600} reached ~0.25 and the cells were harvested. For the serine hydroxamate (SHX) treated samples the cultures were treated with SHX (500 µg/ml, 30 min) before harvesting.

Table 2.8 *Escherichia coli* strain used in this study.

Strain	Genotype	Reference
E. coli K12 CC72	F- lambda- *ilvG- rfb*-50 *rph*-1	(Cagliero & Jin, 2013)

2.4.2 GENOME CONFORMATION CAPTURE (GCC)

E. coli chromatin was prepared according to (Rodley et al., 2009) with minor modifications. A total of $5*10^9$ cells were cross-linked with formaldehyde (1% final v/v, 20 min, RT) and then quenched with glycine (125mM final, 10 min). Cells were collected by centrifugation (4000 rpm, 15 min, 4°C), washed twice (1% PBS, 1% TritonX-100, 5ml/50ml culture) and pelleted (4000 rpm, 15 min, 4°C). Cell pellets were suspended in 800µl of B1 lysis buffer (10mM Tris pH 8.0, 50mM NaCl, 10mM EDTA, 20% (w/v) sucrose, 1mg/ml lysozyme) and incubated (37°C, 30 min). 800 µl of B2 lysis buffer (200mM Tris pH 8.0, 600mM NaCl, 4% TritonX-100, 1x Complete protease inhibitor cocktail tablet (Roche) per 10ml of buffer was added just before use) was gently added, mixed by inversion 3-4 times and incubated (37°C, 10 min). The cell lysate was centrifuged (21,500g, 20 min, 4°C) and the supernatant decanted. The chromatin was washed once with 1ml of chromatin digestion buffer (10mM Tris-HCl pH 8.0, 5mM MgCl2, 0.1% TritonX-100) by inverting the tube 3-4 times and centrifuged (21,500g, 20 min, 4°C). The supernatant was decanted and the chromatin pellet was suspended in 500µl chromatin digestion buffer. Chromatin samples were aliquoted into 10 sets of $5*10^8$ cells. Samples were digested with

HhaI (100U, New England Biolabs). A ligation control was added to the digestion chromatin (see section 2.4.3 and Table 2.9), the samples were diluted (~20-fold) and ligated with T4 DNA ligase (20U, Invitrogen). Following ligation, cross-links were removed in the presence of proteinase K (0.45U, Fermentas). RNA was removed and pUC19 plasmid (27.4pg/2ml) was added as a sequencing control prior to three extractions with 1:1 Phenol:Chloroform. DNA was column purified (Zymo, DNA clean and concentratorTM-5 kit) according to the manufacturer's instructions and eluted in milliQ H2O before combining for sequencing. 3µg of purified DNA was sent for paired-end (PE) sequencing (100 bp) at the ATC sequencing facility (Rockville, MD, USA) on an Illumina Hi-Seq.

Table 2.9. PCR primers used to generate the ligation controls used in this study.

PCR primers were designed to amplify a region form the Lambda phage genome and pRS426 that contained a HhaI site near to one end that once cut would produce a product 180-200 bp in length.

Primer name	Sequence	Product size
Lambda GCC F2	TGAAGAATGCCAGAGACTCC	198bp
Lambda GCC R2	ACCCCGGTATCAGTTCATCC	
pRS426 GCC F2	AGTCACTGGCGCTTGGTCTGACAGTTACCAATGC	187bp
pRS426 GCC R2	GATAAATCTGGAGCCGGTGA	

2.4.3 Production of external ligation controls for GCC library preparation

External ligation controls were produced by PCR amplification of short regions from the Lambda phage genome and the pRS426 plasmid. Primers (Table 2.9) were designed to include a HhaI site at one end of the final product. PCR products were purified using a PCR purification kit (Qiagen), digested with HhaI (4U, 37°C, 2h) and purified again. Purified, digested PCR products were introduced into the GCC samples at a 1:1 ratio with the number of genomes prior to the ligation step during GCC preparation. The pRS426 fragment was introduced into the exponential phase (LB grown) samples and resulted in 220 separate ligation events with HhaI restriction fragments on the genome. The Lambda phage fragment was introduced into the SHX treated samples and resulted in 2 ligation events with HhaI fragments on the genome.

2.4.5 Genome Conformation Capture network assembly, effects of sample production and processing, and bioinformatics analysis

To identify interacting DNA fragments from the PE sequence reads, network assembly was performed using the Topography suite v1.19 (Rodley et al., 2009). GCC networks were constructed from 100 bp PE Illumina Genome Analyser sequence reads. Topography uses the SOAP algorithm (R Li, Li, Kristiansen, & Wang, 2008) to position PE tags and single ends which contain a HhaI restriction enzyme site onto the *E. coli* (NC_000913) reference genome. The reference genome also contained the pUC19 (SYNPUC19CV) sequence and the sequences of the pRS426 plasmid and Lambda phage ligation controls (Table 2.9). No mismatches or unassigned bases (N) were allowed during positioning.

Except where indicated, bioinformatics and statistical analyses were performed on interactions identified by sequence reads that were uniquely mapped onto the reference genome and were above the cut-off value derived from the ligation control interactions (see section 2.4.7). A breakdown of the interactions present in the *E. coli* samples is provided in Table 2.1. All bioinformatics analysis was performed using in house Perl and Python scripts. Except where indicated, statistical analyses were performed in R (R Development Core Team, 2008).

The use of PE sequencing reads means that the HhaI restriction enzyme site does not have to be present in either of the sequences to detect an interaction. This, however, results in all PE reads effectively representing and interaction. Therefore, to insure that we could detect the same interactions using single ended reads, which necessitates that there is a HhaI restriction enzyme site detected in the sequence, we reconstructed part of the network using single-end (SE) reads and compared the interactions we identified to those detected in the PE analysis. Sequences from one of the two *E. coli* exponential phase biological replicates were used for SE network assembly as above. The number of interactions detected in this data set was compared to the interactions detected in the PE networks for the combined biological replicates. This analysis demonstrated that 92% of the interactions detected in the PE analysis were also detected in the SE analysis of one of the biological replicates for the exponentially growing cells. Therefore, the interactions detected by PE sequencing are real and not just an artefact of the PE analysis.

2.4.6 Quantification of the effects of bar-coding, sequencing and biological replicates on network patterns

The interactions within the biological repeats were highly correlated (Figure 2.10). To further investigate the level of agreement between these samples, and the effects of sequencing on this agreement, we used Cohen's Kappa is a statistic that summarizes agreement in categorical variables. Binning the interaction counts into categories allows us to reflect the greater importance of concordance among interaction pairs with high counts. Cohen's Kappa ranges between 0 and 1 with 1 being perfect agreement; however, it is a conservative measure that controls for the level of chance matching expected. (See, *e.g.*, (Landis & Koch, 1977)). Thus, the expected level of chance matching, controlled by the marginal probability of landing in each category, influences Kappa; we will study a case where these marginal probabilities are identical across the different situations of interest, so the Kappa measures can be directly compared.

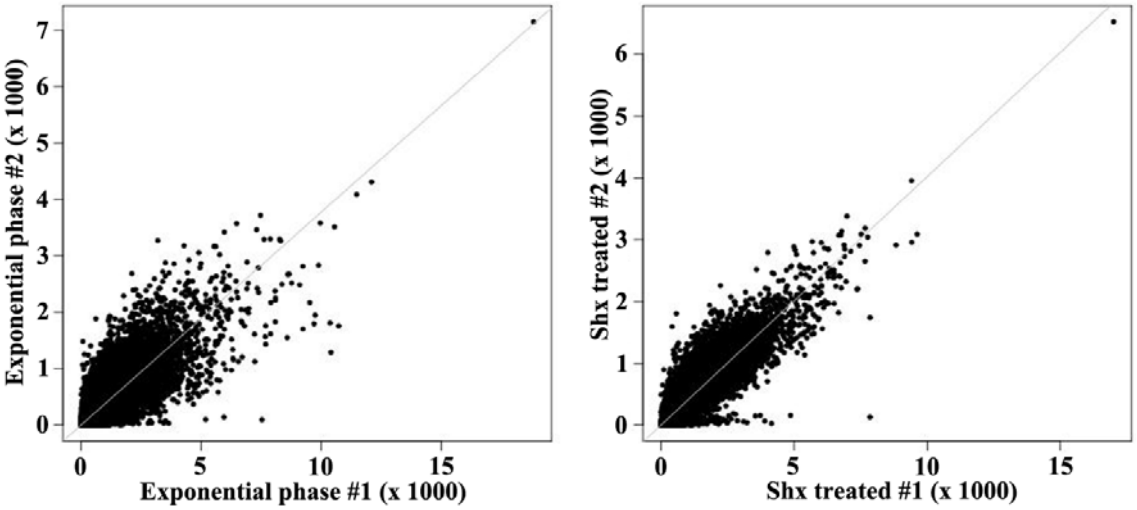


Figure 2.10. Biological replicates of the E. coli interaction networks were highly correlated at the HhaI restriction fragment level.

Exponential phase *E. coli* cells (left, R^2=0.773) and SHX treated cells (right, R^2=0.8741). Scatter plots were constructed from all interactions in each dataset involving only HhaI fragments which could be uniquely positioned on the reference genome.

To study the level of agreement between different sequencing results from different lanes, barcodes, and biological replicates, we have considered all interactions observed for a particular condition. The data have been modified to reflect how often each interaction occurs in each scenario, *i.e.*, many counts of "0" have been added to show when an interaction is possible under a condition, but was not observed for a particular lane/barcode/replicate combination. For each combination of factors, we have then labelled each interaction as occurring in the

bottom 95%, top 5% but not top 1%, top 1% but not top 0.5%, and top 0.5%. The top 5% typically starts at a count of around 5. The choice of these quantiles was based on the fact that lower counts have a large number of interactions with that count, thus the nominal proportion associated with the quantile can be misleading; *e.g.* if the median is 2, there may actually be 70% of the interactions with count ≤2. For the values chosen, the nominal proportions correspond well to the actual proportions, which are the marginal probabilities referred to above, enabling proper comparison of Kappa values. Kappa is defined for agreement between two variables, so all pairs of factor combinations are considered.

Barcoding, sequencing lane and biological replicates did not strongly affect the correlation between samples (Table 2.10). The differences introduced by different lanes and barcodes are indistinguishable from different barcodes alone; however, as expected, different biological replicates have lower agreement than those that differ only in barcode/lane (Table 2.10). The technical replicates (L1a, L1b, L1c and S1a, S1b, S1c) were highly correlated and were combined into L1 and S1 respectively. Following this pooling, Pearson's correlation analysis demonstrated that the biological replicates of the *E. coli* interaction networks were highly correlated at the HhaI restriction fragment level (exponential phase *E. coli* cells, r=0.879; and SHX treated cells r=0.935; Table 2.10). Thus the biological replicates (L1, L2 and S1, S2) were combined into exponential and SHX samples for the remainder of the analysis, respectively.

Table 2.10. Barcoding, sequencing lane and biological replicates did not affect the correlation between samples.

We calculated Cohen's Kappa for agreement between each pair of datasets; the values are listed by the differing factors for that pair (*i.e.* differing sequencing lane only, differing barcode and lane, and differing biological replicate, barcode and lane; see section 2.4.6).

	Condition	
Difference	exponential	SHX
Barcode only	0.73	0.70
Barcode and Lane	0.74, 0.74	0.69, 0.71
Barcode, Lane, Biological replicate	0.62, 0.62, 0.64	0.65, 0.68, 0.67

2.4.7 Filtering based on ligation controls

We wanted to know which of our individual interactions were above experimental noise. During the preparation of the GCC samples, random ligation events can occur during: 1) the GCC ligation step, and 2) the sequencing preparation step (*i.e.*, the addition of linkers during sequencing library preparation).

In an attempt to control for this, external controls were added during the GCC library preparation to obtain estimates of the rates of inter-molecular ligation events. The three pUC19-gDNA ligation events with the highest interaction frequencies in both exponential and SHX treated samples were further assessed. The sequences of the *E. coli* fragments found to interact with pUC19 were extracted and fragmented using a 13 bp sliding window, shifted 1 bp at a time. These fragments were subsequently aligned to the pUC19 fragments they interacted with. If the pUC19 fragments and *E. coli* sequences aligned multiple times then the high level of interactions was attributed to a miss-alignment and not to a random ligation event. There was one *E. coli*-pUC19 interaction (frequency of 4) in the SHX treated sample that did not occur due to miss-alignment. Therefore, the cut-off for filtering out random ligations was set to 4 interactions; only fragment pairs that occurred 5 or more times were considered for analysis.

2.4.8 Collector's curve

Collector's curves were generated using the total interaction data set from exponential growing and SHX treated *E. coli* cells, including the internal ligation controls. Specified fractions of the total interactions (*e.g.* 10%, 20%,…) were randomly sampled 100 times. For each random data set, the fragments that had an interaction frequency strictly greater than the number of interactions detected between *E. coli* genomic loci and the internal ligation controls were considered significant. The significant interactions were compared to the significant interactions present in the original data file containing only significant non-adjacent interactions. The number of interactions shared between the files was averaged for the 100 random data sets at each specified fraction of the total interactions sampled. The percentage to total was calculated from these averages and plotted as a collector's curve (Figure 2.11). Over 75% of the interactions present in each condition were detected after sampling just 10% of the total interactions. Therefore, we concluded that we have sampled a significant proportion of the interactions that were present in the *E. coli* cells at the time of cross-linking.

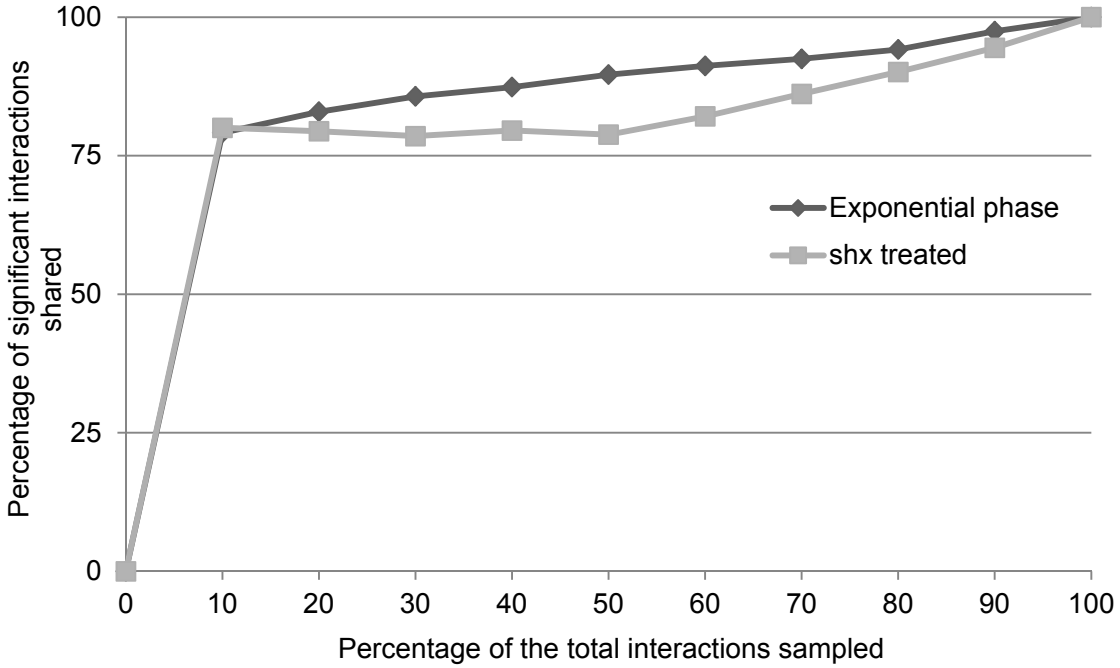


Figure 2.11. A collectors curve was generated to determine the level of saturation of interaction detection.

The average percentage of significant interactions in the random data sets that were also detected in the original data file was graphed. Over 75% of the interactions in the original data file were detected after sampling just 10% of the total interactions. We conclude that we have sampled a significant proportion of the interactions that were present in the populations of *E. coli* cells at the time of cross-linking.

2.4.9 Analysis of loop size and interacting fragment distributions

Loop size calculations were performed on GCC interaction networks that had repetitive sequences and adjacent interactions removed. The size of the loop between two interacting DNA fragments was determined by taking the absolute value after subtracting the end position of the first interacting partner (having the smaller end coordinate of the two fragments) from the start position of the second interacting partner (having the larger start coordinate of the two fragments). Because the bacterial genomes used in this study are circular, if the loop size was greater than half the size of the genome than the loop size was subtracted from the total genome size to give the actual loop size. Size specified bins were used to count the number of loops that were of a particular size and the data was plotted as the number of loops versus the bin size. To determine the loop size as a function of the distance from the origin of replication, the distance of each interacting partner from the origin was calculated and the loop size was associated with the calculated distance. The loop sizes were binned and corrected for the number of loops per bin. The difference in the average loop size per bin between exponential and SHX samples was plotted.

2.4.10 Interaction frequency versus distance from the origin of replication

All GCC interaction network data or only long distance interaction network data (>800 bp) was used to calculate the interaction frequency as a function of the distance from the origin of replication. The distance from the origin of replication was calculated for each interacting partner and then the total interaction frequency for each interacting fragment was assigned to this particular distance from the origin. The data was binned (10,000 bp) and plotted as total interaction frequency per bin, corrected for the number of fragments per bin, versus the distance from the origin of replication.

2.4.11 Genome copy number

Copy number was determined across the *E. coli* genome using Control-Free Copy number and genotype caller (Control-FREEC) (Boeva et al., 2012). The *E. coli* input sequences were in the SAM format, genome length was set at 4,639,675 bp, window size = 1,000, and telocentromeric = 0. The GC profile was calculated and included.

2.4.12 Transcription microarray

E. coli strain CC72 (Table 2.8) (Cagliero & Jin, 2013) was grown in LB (Gibco, lot # 817849) at 37°C until the Optical Density (OD_{600}) reached 0.2. The SHX sample was treated at OD_{600} 0.2 with SHX (500µg/ml, 30 min) prior to RNA isolation. RNA was isolated using the hot phenol procedure. Briefly, 800µl of cells were mixed with 700 µl of 65°C phenol (pH 5.0) and 100µl of 16 x lysis buffer (320mM Na Acetate, 8% SDS, 16mM EDTA) and incubated (65°C, 5 min). The RNA was extracted twice with Phenol:Chloroform:Isoamylalcohol (25:24:1, pH 8.0), precipitated with isopropanol and suspended in DEPC water. To remove the DNA, 40µg of RNA was treated with Turbo DNaseI (Ambion). The RNA was extracted twice with Phenol:Chloroform:Isoamylalcohol (25:24:1, pH 8.0), precipitated with isopropanol and suspended in DEPC treated water (Invitrogen). The cDNA library was constructed using the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) according to manufacturer's instructions and the cDNA was sent to Roche-Nimblegen for microarray hybridization. Each experiment (exponential or SHX) is a pool of three biological replicates. A total of two technical replicates were performed per condition (exponential and SHX).

The Expression values were generated using quantile normalization (Bolstad, Irizarry, Åstrand, & Speed, 2003; Irizarry, Bolstad, et al., 2003; Irizarry, Hobbs, et al., 2003). The average expression level of each gene from the two biological replicates for exponential and SHX samples were calculated. To determine which genes were significantly up and down regulated in SHX treated compared to exponential samples the log2 of the SHX/exponential ratio was calculated. Genes were considered significantly up regulated if the log2 ratio was greater than +1.5 and the log2 of the raw expression level in SHX was greater than 9. In contrast genes were considered to be significantly down regulated if the log2 ratio was smaller than -1.5 and the log2 of the raw expression level in exponential phase was greater than 9 (Table 2.11).

Table 2.11. The number of annotated *E. coli* genes that were significantly up or down regulated in response to SHX treatment.

Up regulated	644
Down regulated	687

2.4.13 Gene Ontology (GO) term enrichment analysis of the significantly up and down regulated genes

To determine whether the significantly up and down regulated genes were enriched for particular GO terms an R package called GOstats was used (Falcon & Gentleman, 2007). The 'gene universe' contained all annotate genes from the *E. coli* gene products list (http://regulondb.ccg.unam.mx/data/GeneProductSet.txt) and was compared to the significantly up or down regulated gene sets (Table 2.11). A standard hypergeometric test was used with a p-value cut-off of <0.01 (Table 2.6).

2.4.14 Correlating transcription level with interaction frequency

The raw expression level of all genes as well as just genes with a high expression level (log2 of the raw expression level >9) was correlated with the total interaction frequency with these genes. Using the genomic coordinates of the genes obtained from a gene products list (http://regulondb.ccg.unam.mx/data/GeneProductSet.txt), the interaction frequency with each of these regions was determined by summing up the total interaction frequency within each region. Where the region overlapped with a restriction fragment the interaction frequency was proportionally assigned to the region

depending on the percentage of overlap. The interaction frequency of each gene was then plotted against the raw expression level for each gene.

2.4.15 Transcription regulons

We wanted to determine whether the distribution of significantly up and down regulated genes was non-random across the genome relative to the origin of replication. The distance of each gene from the origin was calculated. The fold change in gene expression was binned (50,000 bp bins) according to the distance of the gene from the origin. The gene expression values were shuffled randomly 1,000 times. After each round of shuffling, a distance from the origin was assigned to each value and the values were binned as above. The expression level in each bin for the real data and the randomly generated data was corrected for the number of genes per bin. The average fold change in gene expression for the real data and randomly generated data was then plotted against the distance from the origin of replication. Additionally, an auto correlation analysis was performed in R using the non-binned expression data to determine whether the pattern of up and down regulated genes seen across the genome is random or not.

2.4.16 GC content and transcription levels

The *E. coli* genome was fragmented using a 1,000 bp sliding window shifted 1 bp at a time and the percentage GC content of each fragment was calculated. The distance of each fragment from the origin of replication was calculated and the percentage GC content for each fragment was placed into the appropriate 50,000 bp bin. The average GC content per bin was calculated and plotted against the average fold change in gene expression per bin.

2.4.17 MatS, SeqA, SlmA and NAP clustering analyses

Nucleoid Associated Protein (NAP) binding sites were obtained from (Grainger et al., 2006). MatP binding sites (MatS) were obtained from (Mercier et al., 2008). Regions for analysis were defined by taking a specified number of bases (50, 100, or 250 bp) either side of the peak binding position for NAPs or centre of the MatP binding site for MatS. For SeqA the strongest 135 confirmed SeqA binding sites where obtained from (Sánchez-Romero et al., 2010) and the 24 defined SlmA binding sites were obtained from (Cho et al., 2011). To determine whether these regions could be found in a different interacting environment compared to what would be expected by random chance, the total number of interactions with each of

the individual regions and the number of interactions that occurred between the regions of interest (*i.e.* clustering) was determined from our GCC interaction network. We then generated 1,000 random data sets of the same number and length (bp) as the actual region data set using two methods: 1) randomly selecting a start position for each region and then making it the same length as the region for which the random coordinate was being generated (*i.e.* random spacing [RS]); or 2) randomly select the start position for the first region and then sequentially determining the start and end position of all the other regions in the set such that the linear distances between regions were maintained (*i.e.* conserved linear spacing [CLS]). This ensured that the particular interaction frequencies we observed were not due to the linear arrangement of the regions around the circular genome. 1,000 random data sets were generated for the RS and CLS methods and the total interaction and clustering frequencies were calculated from our GCC interaction network. The frequency with which the total interaction and clustering frequency of the actual data was higher or lower than the random data sets was used to estimate significance.

2.4.18 Interactions and clustering of genes that significantly change their expression level upon SHX treatment

Genomic coordinates of genes that significantly change their expression level upon treatment with SHX were obtained from http://regulondb.ccg.unam.mx/data/GeneProductSet.txt. The total number of interactions with each of the individual genes and the number of interactions that occurred between the genes of interest was determined as for MatS, SeqA, SlmA and NAP clustering, above.

2.5 DATA ACCESS

The GCC data has been banked with Gene expression omnibus (GSE40603). Expression data has been deposited GSE40304.

2.6 FUNDING

This work was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research [CC and JD]; the Marsden Fund [JOS]; and Massey University research fund [JOS and BHAR]; Massey University scholarship [RSG]; Ministry for Science Innovation [BHAR]; Canadian Institutes for Health Research [REWH].

Chapter 3

3 GLOBAL ANALYSES OF CELL CYCLE DEPENDENT CHANGES IN FISSION YEAST GENOME ORGANIZATION REVEAL CORRELATIONS WITH ALTERATIONS IN TRANSCRIPT LEVELS

Since the submission of this thesis a modified version of this manuscript has been accepted as a 'Breakthrough article' in the journal Nucleic Acid Research:

Grand, R.S.[1,2], Jones, B.M.[2], Greyling, G.[1], Tsai, P.[3], Martienssen, R.[4], O'Sullivan, J.M.[1] (submitted to Genome Biology). Global analysis of cell cycle dependent changes in Fission yeast genome organization reveal correlations with alterations in transcript levels.

[1]Liggins institute, University of Auckland, Grafton Auckland 1032, NZ, [2]Institute of Natural and Mathematical Sciences, Massey University, Albany Auckland 0745, NZ, [3]School of Biological Sciences, University of Auckland, Auckland, NZ, [4]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

Acknowledgements: The authors would like to thank Julia Horsfield for comments on the manuscript. This work was supported by The Marsden fund (JMOS) and a Massey University Doctoral Scholarship (RSG).

Author's contributions: RSG, RM, and JMO conceived and directed the project, and wrote the manuscript. RSG performed experiments and analysis. MBJ advised and helped with statistical analysis. GG performed some computer analysis. PT advised on transcriptome analysis.

3.1 INTRODUCTION

The spatial and temporal organization of genomes are increasingly recognized as key contributors to genome maintenance and gene regulation in both prokaryotes and eukaryotes (Cagliero et al., 2013; Cavalli & Misteli, 2013; de Wit et al., 2013; Denholtz & Plath, 2012; Peric-Hupkes et al., 2010). The spatial arrangement of chromosomes at a given moment in time is the sum of hierarchical levels of organization. High resolution microscopy and proximity-based ligation techniques are beginning to reveal the dynamics of this hierarchical spatial organization and its effects on genome function (de Wit et al., 2013; Grand et al., 2011; Lieberman-Aiden et al., 2009; Rodley et al., 2009; Sanyal et al., 2012; Sexton et al., 2012; Tanizawa et al., 2010).

Cell growth proceeds in an ordered manner through a regulated cycle consisting of the gap 1 (G1), synthesis (S), gap 2 (G2) and mitotic (M) phases. The primary cell cycle checkpoint where cells determine whether to progress or pause varies in different species. For example, in the Bakers yeast, *Saccharomyces cerevisiae*, and mammals, the primary checkpoint is during the G1/S phase transition. Conversely, in the evolutionarily distal Fission yeast, *Schizosaccharomyces pombe*, primary control occurs at the G2/M transition (Forsburg & Nurse, 1991; Lukas, Lukas, & Bartek, 2004). *S. pombe* spends ~75% of its time in the G2 phase of the cell cycle, where there are two copies of each chromosome. By contrast, relatively little time is spent in M phase, where chromosomes are highly condensed and are actively segregated into the daughter cells. Similarly, little time is spent in the G1 and S phases of the cell cycle, where the cells commit to the cell cycle and synthesize DNA, respectively.

Gene transcript levels (McInerny, 2004; Rustici et al., 2004), heterochromatin formation (Kloc et al., 2008), telomere, centromere, and mating type loci clustering (Alfredsson-Timmins et al., 2009; Funabiki et al., 1993) fluctuate throughout the *S. pombe* cell cycle. Metaphase chromosome condensation coincides with a marked reduction in transcription and remains arguably the most obvious change that occurs during the cell cycle of all cells. Cell type specific arrangements of chromosomes (Parada, McQueen, & Misteli, 2004) and looping between genes and regulatory elements, within and between chromosomes, contribute to cell type specific phenotypes and temporal processes (Cagliero et al., 2013; Cavalli & Misteli, 2013; de Wit et al., 2013; Denholtz & Plath, 2012; Peric-Hupkes et al.,

2010). This raises the question: are specific genome organizations propagated through the cell cycle as a form of epigenetic inheritance (Blomen & Boonstra, 2011; Essers et al., 2005; Gerlich et al., 2003)?

The complexity of cell cycle regulation and size of metazoan genomes make it difficult to interrogate the relationship between genome structure and function over the course of the cell cycle. The Fission yeast, *S. pombe*, shares many mammalian features including linear chromosomes, constitutive pericentromeric and telomeric heterochromatin. Furthermore, *S. pombe* cell cycle regulation is well characterised and controllable enabling the in depth investigation of cell cycle dependent processes (Coudreuse & Nurse, 2010; Nurse et al., 1976).

Mixed tissues and asynchronicity, together with the dynamic nature of chromatin, potentially preclude the identification of cell cycle specific genome organization and cloud our understanding of the genome structure-function relationship. To date there is no high resolution comparative molecular study that interrogates genome organization and nuclear function as cells progress through the cell cycle. Rather, with three studies where cells were in a single cell cycle phase (Naumova et al., 2013; Umbarger et al., 2011; Zhang et al., 2012), proximity-based ligation studies have been performed on multicellular organisms or asynchronous cell populations (de Wit et al., 2013; Lieberman-Aiden et al., 2009; Nagano et al., 2013; Rodley et al., 2009; Sexton et al., 2012; Tanizawa et al., 2010). Here we present high resolution structures for G1, G2 and M phase chromosomes produced by Genome Conformation Capture (GCC) in synchronized populations of Fission yeast cells. We identify dynamic connections between and within chromosomes through all phases of the cell cycle. Moreover, specific subsets of these interactions appeared to have positive and negative effects on transcript levels and correlated with waves of transcriptional activity between the cell cycle phases. The correlations between transcript levels and the formation and disruption of cell cycle specific chromosomal contacts that we observed implicate genome organization in epigenetic inheritance and bookmarking.

3.2 RESULTS

3.2.1 *S. POMBE* GENOME ORGANIZATION CHANGES THROUGHOUT THE CELL CYCLE

Microscopic studies show dramatic shifts in chromatin organization during the cell cycle, from a highly condensed structure in metaphase to open, decondensed fibres in interphase (Funabiki et al., 1993). We therefore wanted to investigate whether studying synchronized populations of cells would reveal further details about the organization of genomes and if the formation of cell cycle specific organization is important, for example in the regulation of gene transcription. We considered a number of methods for obtaining populations of *S. pombe* cells synchronized at specific phases of the cell cycle, including: Centrifugal elutriation, Lactose gradient, nitrogen deprivation, chemical treatment, and temperature sensitive cell division cycle mutants (Fantes & Nurse, 1978; Faraday et al., 1994; Gómez & Forsburg, 2004; Hirano, Hiraoka, & Yanagida, 1988; Walker, 1999).

Centrifugal elutriation and lactose gradient methods separate cells in a culture based on their size allowing for the isolation of cell size fractions. In the case of *S. pombe* the smallest cells are in the G2 phase (Walker, 1999). These G2 phase cells can be isolated and grown further to obtain subsequent cell cycle phases. These methods do not perturb the cells and produce reasonably high levels of G2 phase synchronized cells but the level of synchronization diminishes rapidly upon continued culturing. The removal of nitrogen from the growth medium or addition of chemicals (Thymidine treatment) can arrest *S. pombe* cells at specific stages of the cell cycle (Faraday et al., 1994; Gómez & Forsburg, 2004). These methods produce high levels of synchronized cells; however, the effect of the drastic treatment on the cell may compound the results. Furthermore, as for the cell size selection methods, the cells must be further cultures to obtain subsequence cell cycle phases. Alternatively, different chemical treatments could be used to isolate different cell cycle phases, but such chemicals for the use in *S. pombe* are limited and this would likely further complicate the resulting data. As a result we decided to utilize temperature sensitive *S. pombe* mutant cells that when shifted from a permissive to restrictive temperature for a defined amount of time, become synchronized at specific stages of the cell cycle. These mutants have been shown to have little influence on normal cell growth and produce cell populations with a high proportion of synchronized cells. To further minimize the effect that the shift in temperature might have on genome organization and cellular processes that would

confound the results, mutants that undergo the same temperature shift where selected (Aves, Durkacz, Carr, & Nurse, 1985; Fantes & Nurse, 1978; Hirano et al., 1988; Nurse, 1975).

We used Genome Conformation Capture (GCC; Rodley et al. 2009) to determine the high resolution spatial genome organization of *S. pombe* cells synchronized in the G1, G2 and M phases (>95%, >95% and >80% synchronized, respectively) of the cell cycle (Figure 3.1; A(i), B(i) and C(i)). The biological replicates were highly correlative (Figure 3.2A-C) indicating that the chromosomal interactions we sampled were reproducibly detected within the cell populations, despite not saturating the total number of interactions (Figure 3.3).

Inter- and intra-chromosomal interactions between uniquely positioned loci varied in number and strength at each of the cell cycle phases (Figure 3.1: A(ii), B(ii) and C(ii); Figure 3.2D-F; Table 3.1. All analyses were performed on uniquely positioned loci unless otherwise stated). The majority (~80% in G1 and ~90% in G2 and M phases) of the interactions occurred within chromosomes (Table 3.1), supporting the existence of chromosome territories in *S. pombe* (Scherthan, Bahler, & Kohli, 1994; Tanizawa et al., 2010), and were predominantly shared by all three cell cycle phases: G1, G2 and M phase (Figure 3.5). M phase chromosomes had the largest number of phase specific, intra-chromosome interactions, with a clear increase in the number of interactions with loop lengths up to 5 Kb (Figure 3.4A and Figure 3.5C). By contrast, interactions between chromosomes were predominantly cell cycle specific, with the largest number forming in G1 phase (Figure 3.5B), likely reflecting the interaction between chromatids in the G2 and M phases.

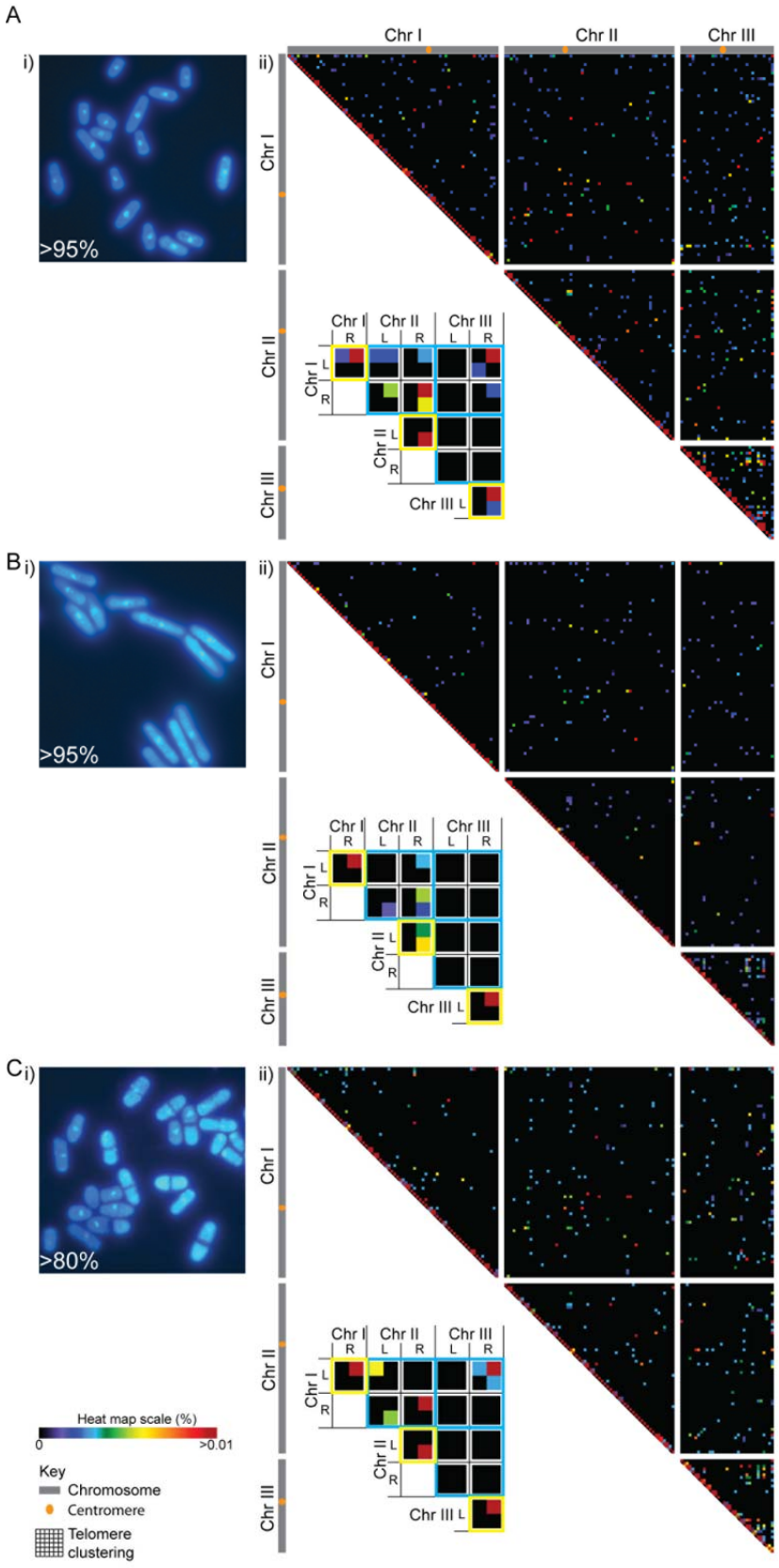


Figure 3.1. The spatial organization of the *S. pombe* genome changes throughout the cell cycle. Refer to footnote [4] for figure legend.

[4] CDC mutants were used to generate synchronized *S. pombe* cells for the G1, G2, and M phases of the cell cycle. Microscopy of DAPI and calcofluor white stained cells confirmed cell

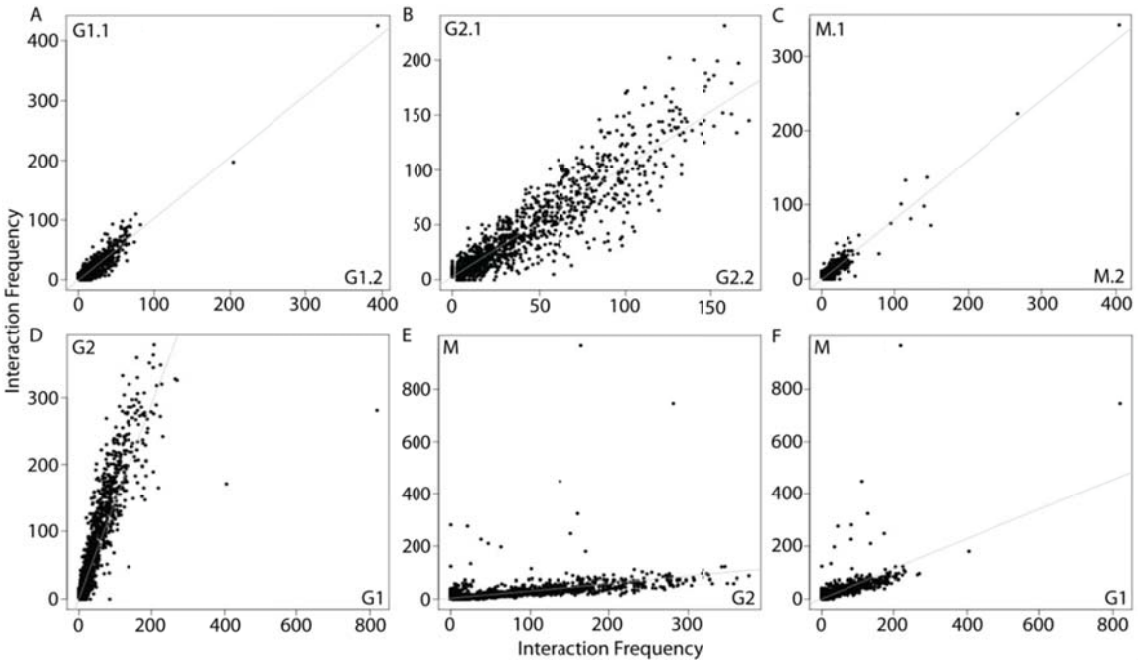


Figure 3.2. Correlation plots of *S. pombe* interaction networks for individual biological replicates and for each cell cycle phase.

The interaction networks of the biological replicates were highly correlated at the AseI restriction fragment level. A) G1 phase biological replicates (G1.1, replicate one and G1.2, replicate two, R2=0.8643), B) G2 phase biological replicates (G2.1, replicate one and G2.2, replicate two, R2=0.8783), B) M phase biological replicates (M.1, replicate one and M.2, replicate two, R2=0.8692). D) G1 and G2 phase interaction networks were also highly correlated (G1, G1 phase and G2, G2 phase, R2=0.8058), apart from two outliers in G1 phase that were attributed to interactions with LTR elements. In contrast, there was a low correlation between the G1 and M phase (G1, G1 phase and M, M phase, R2=0.5335), and G2 and M phase (G2, G2 phase and M, M phase, R2=0.3239) interaction networks. Uniquely mapped data files containing significant interactions that have had duplicates, self and adjacent interactions removed were used to generate these plots in R.

synchronization; G1 (>95%), G2 (>95%) and M (>80%) phase (A(i), B(i) and C(i)). Increased septation index (from ~16% to ~50%), highly condensed chromosomes, and the presence of enucleate cells, following DAPI staining are characteristic traits described for synchronous M phases cells (Hirano et al., 1988). Differences in the number and strength of interactions detected at each phase of the cell cycle are observed (A(ii) G1, B(ii) G2 and C(ii) M phase). Enlargement of the sub-telomeric (150 Kb on chromosomes I and II) and regions adjacent to the ribosomal repeats (150 Kb on each arm of chromosome III; A(ii), B(ii) and C(ii), insets) reveals differential co-localization at different stages of the cell cycle. Significant non-adjacent interactions were converted into an interaction matrix for 75 Kb segments expressed as the percentage total interactions and plotted as heat maps.

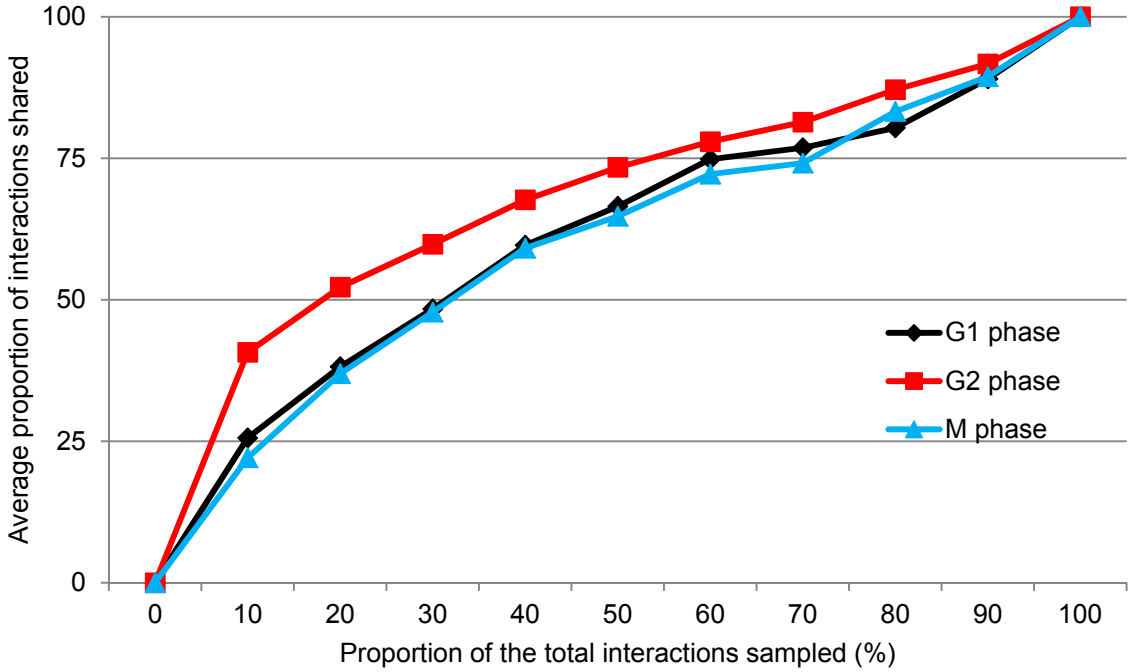


Figure 3.3. The number of unique interactions detected was not fully saturated.

A collector's curve was generated to determine whether the number of unique interactions detected at each phase of the cell cycle reached a level of saturation with the depth of sequencing that was performed. The average percentage of interactions in the significant, non-adjacent data sets that were also detected in randomly generated data sets was graphed. The collector's curves indicated that, despite the high correlation between biological replicates, the interaction network was not sampled to saturation. However, the high degree of correlation between the interaction networks detected for the biological replicates and reduced correlations between different cell cycle phases, indicates that the most frequently occurring interactions were sampled. The highest level of interaction detection was obtained for G2 phase.

Table 3.1. A large proportion of interactions detected at each phase of the cell cycle were within chromosomes.

Different total numbers of significant non-adjacent interactions were detected at each phase of the cell cycle, the majority of which were intra-chromosomal. There is a ~10% increase in the proportion of intra-chromosomal interactions in the two cell cycle phases that contain two copies of each chromosome (*i.e.* G2 and M). Data files containing only uniquely aligned significant non-adjacent interactions were used for this analysis.

Phase of the cell cycle	Total number of interactions	Intra-chromosomal interactions	Inter-chromosomal interactions
G1 phase	1,896	1,534 (80.91%)	362 (19.09%)
G2 phase	1,646	1,479 (89.85%)	167 (10.15%)
M phase	2,211	1,973 (89.24%)	238 (10.76%)

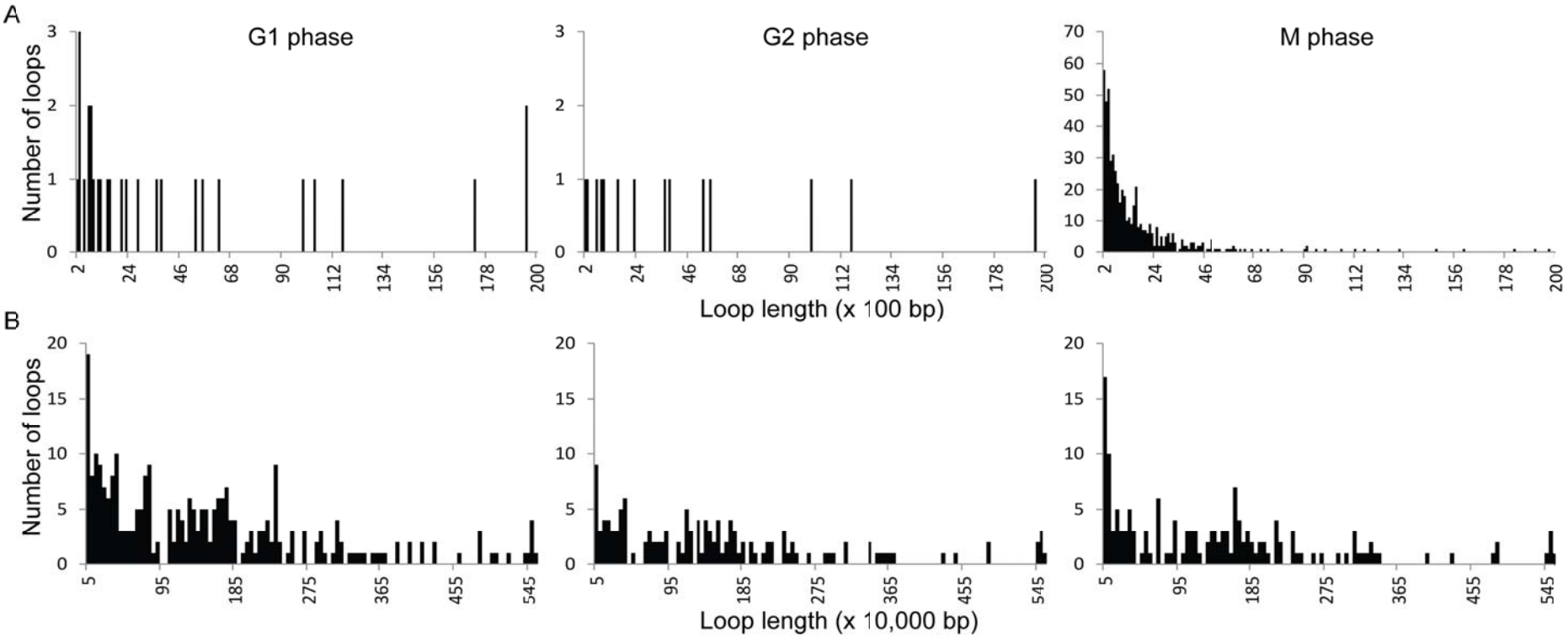


Figure 3.4. There is a marked increase in the number of within chromosome interactions with a loop length of up to ~5Kb in M phase chromosomes.

The length of the loop (bp) between interacting genomic fragments was calculated and binned (A, 100 bp bins; B, 50 Kb bins) for all intra-chromosomal interactions detected at each cell cycle phase; G1, G2 and M. Only loops with a length of ≤20 Kb are shown in A). The majority of loops were between fragments less than 200 bp apart (G1 phase = 83.6%, G2 phase = 91.0%, and M phase = 68.1%, not shown for clarity) with the lowest proportion in M phases chromosomes. There was an increase in the number of interactions with loop lengths of ≤~5 Kb in M phase chromosomes. There was little difference in the proportion of distal (>50 Kb loop) intra-chromosomal interactions between the different phases of the cell cycle (G1 phase = 85.4%, G2 phase = 92.0%, and M phase = 93.1%). Data files containing only uniquely aligned significant non-adjacent interactions were used in these analyses.

The co-localization of telomeres and centromeres has been observed microscopically in *S. pombe* (Hall, Noma, & Grewal, 2003), but the identity of the clustering regions and the specific effects of clustering on nuclear processes are not known. We detected cell cycle specific co-localization between sub-telomeric domains (terminal 150 kb on Chromosomes I and II) and the 150 kb adjacent to rDNA repeats on the left and right arms of Chromosome III (blue highlight in Figure 3.1; A(ii), B(ii) and C(ii) insets). Furthermore, all chromosomes "circularize" and centromeres co-localize throughout the cell cycle (yellow highlight in Figure 3.1; A(ii), B(ii) and C(ii) insets, Figure 3.6 and Figure 3.1). These results are consistent with previously observed characteristics of Fission yeast genome organization and implies that chromosomes assuming a Rabl conformation (Alfredsson-Timmins et al., 2007, 2009; Tanizawa et al., 2010).

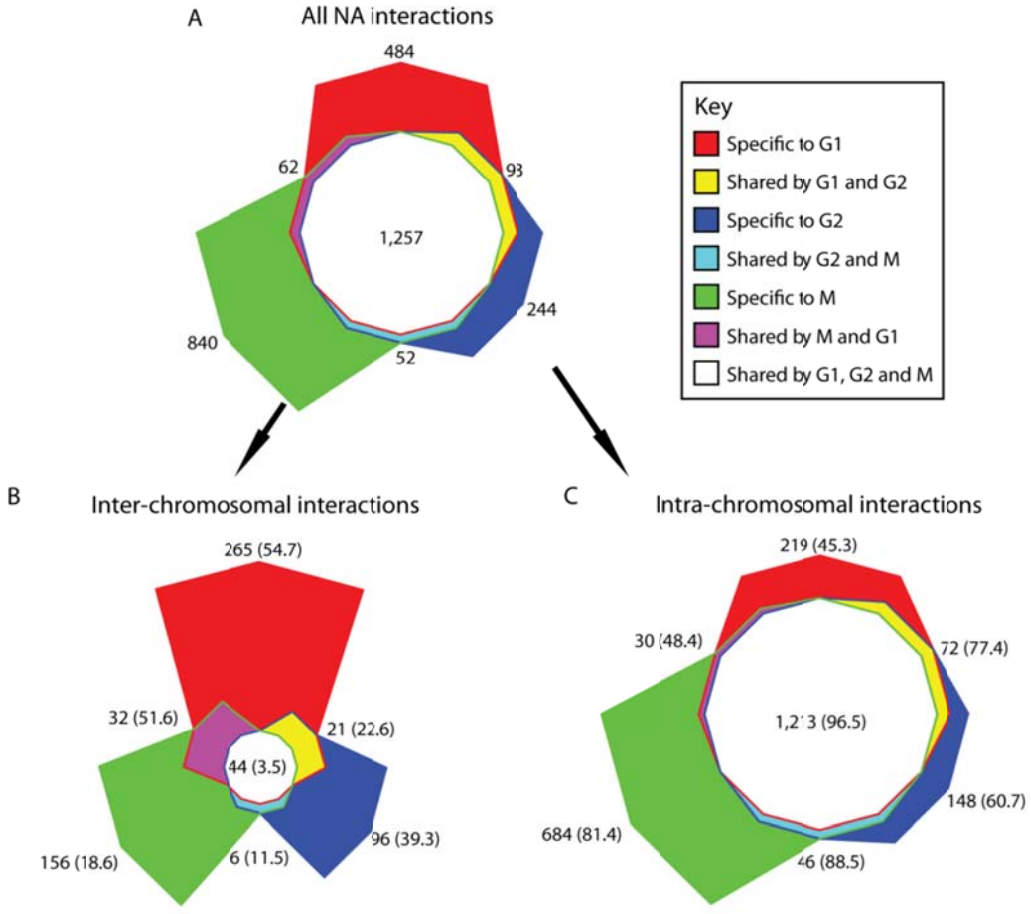


Figure 3.5. Interactions within chromosomes were predominantly shared by the three cell cycle phases, while interactions between chromosomes were largely cell cycle specific.

Three way proportional Venn-diagrams illustrating the number of interactions shared and specific to each cell cycle phase. A) All uniquely aligned significant non-adjacent interactions and subsets thereof; B) intra-chromosomal interactions, and C) inter-chromosomal interactions. The numbers in parentheses in B) and C) are the percentages of the total number of non-adjacent interactions in each category. NA, non-adjacent.

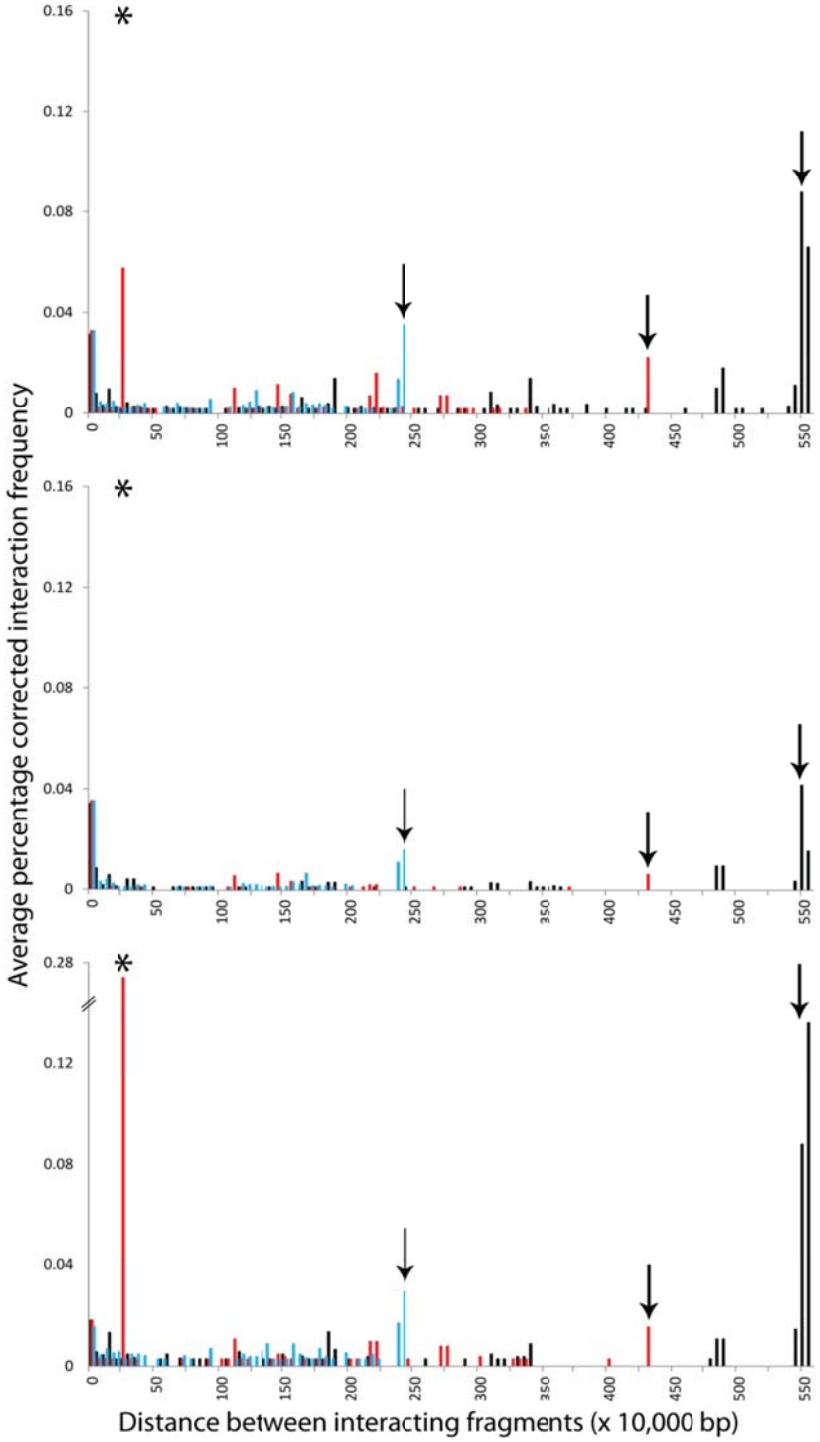


Figure 3.6. All three chromosomes were circularized throughout the cell cycle.

The size of the loop between two interacting fragments, for nuclear intra-chromosomal interactions, was calculated and the interaction frequency was binned (50 Kb bins) based on the loop length. The average interaction frequency for each 50 Kb bin was then calculated and plotted. G1 (top), G2 (middle), and M (Bottom). Intra-chromosomal telomere co-localization (chromosome circularization) of chromosomes I (black) and II (red) and rDNA co-localization on chromosome III (blue) occur at high frequency (arrows), consistent with microscopic observations. An interaction between two fragments on chromosome II (red; Table 3.2 and Table 3.3) fluctuated in frequency throughout the cell cycle (asterisk). Data files containing only uniquely aligned significant non-adjacent interactions were used in these analyses.

3.2.2 CHANGES IN INTERACTIONS BETWEEN REPEAT CONTAINING GENOMIC REGIONS CONTRIBUTE TO CELL CYCLE SPECIFIC GENOME ORGANIZATION AND GENE EXPRESSION

We observed a large number of interactions within and between two regions on either arm of the repeat rich chromosome III (positions: 150-675Kb and 1.725-2.25Mb; Figure 3.8A). The uniquely mapping interactions overlapped extensively with interactions involving repetitive sequences within these regions of chromosome III (Figure 3.8A and Figure 3.7A and B; unique (red), repeat (green) and overlap (yellow)). We observed that interactions between repetitive regions remained relatively unchanged throughout the cell cycle. Yet, unique interactions occurring within the bounds of these repetitive regions fluctuated (Figure 3.8B and C), suggesting that interactions between repeat regions contribute to genome organization by demarcating sub-genomic regions.

In *S. pombe*, LTR elements are bound by the CENP-B protein Abp1 and co-localize into *Tf* bodies (H P Cam et al. 2008; Lorenz et al. 2012). However, it is unknown if the spatial environment (connectivity with the genome) in which LTR elements reside changes throughout the cell cycle. We determined the frequency with which LTR elements interacted with any other genomic region (Figure 3.8D(i, ii, iii)), and co-localized with each other (Figure 3.8E(i, ii, iii)). LTRs co-localized with each other at a significantly high frequency relative to randomly selected loci in all inter-chromosomal interaction data sets (Figure 3.8E(ii)). Similarly, they were connected with other genomic loci at a significantly high frequency but these interactions varied between cell cycle phases (Figure 3.8D(ii)). LTR elements on the same chromosome were more likely to co-localize with each other in G2 and M phase (Figure 3.8D(iii) and E(iii)), perhaps because of interactions between homologs (which could not occur in G1 phase). Such interactions were predominantly between distal LTR elements (>50 Kb apart) (Figure 3.9).

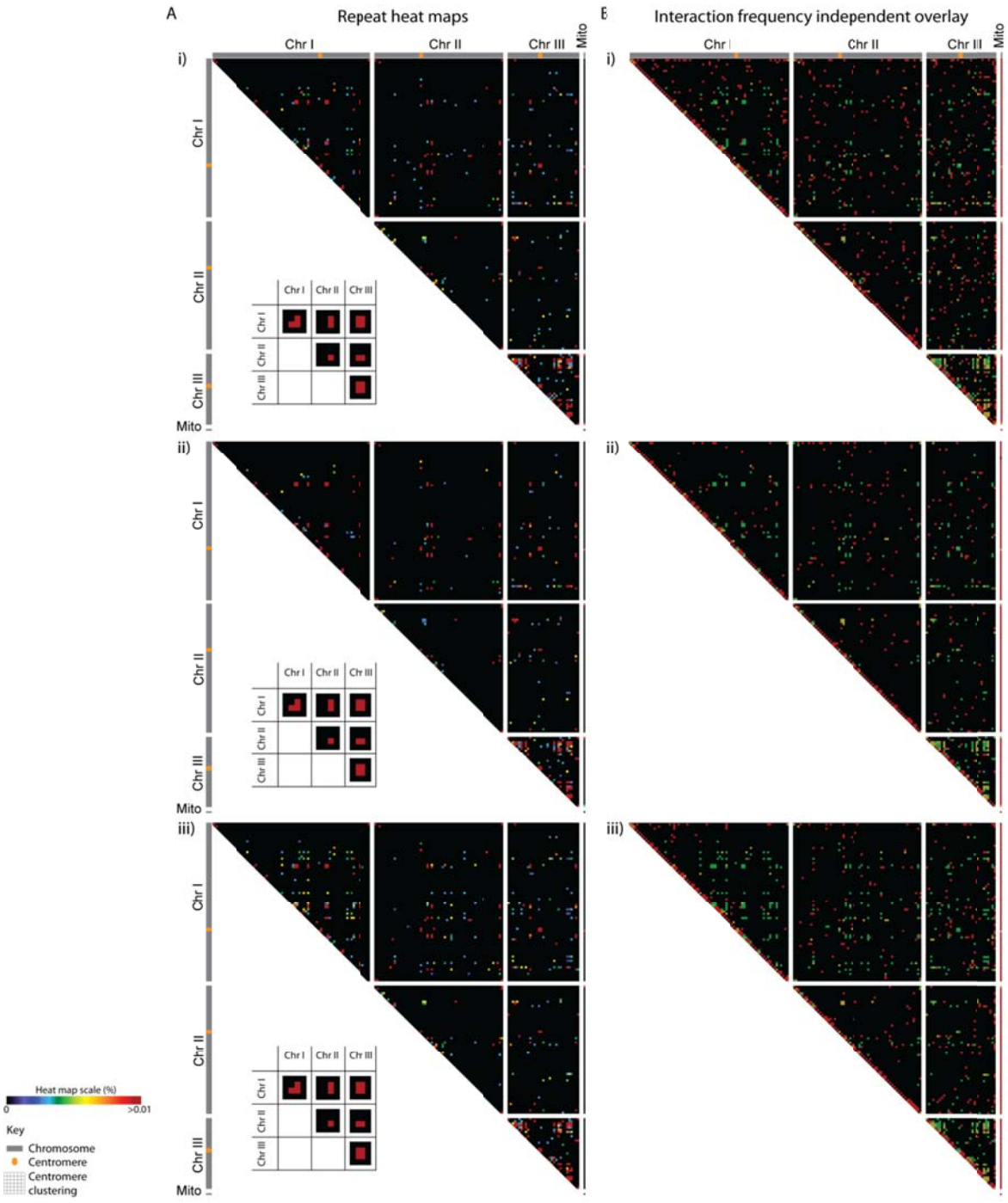


Figure 3.7. Centromere clustering was observed in heat maps of interactions involving repetitive loci. Moreover, repeat regions appear to demarcate the genome.

The inter-chromosomal co-localization of all three chromosome centromeres is observed at all three phases of the cell cycle in heat maps of interactions involving repetitive loci (A, insets). Overlaying the unique (red) and repetitive (green) heat maps revealed regional connections that overlapped (yellow) and occurred separately (B). The cut-off for significant interactions was set 30X higher for the repeat heat maps than the unique heat maps.

A strong intra-chromosomal interaction was detected between two restriction fragments on chromosome II (259,508 bp apart) that contained LTR elements (Figure 3.6 and Table 3.2). This interaction occurred at a very high frequency in G1 phase, disappeared in G2 phase and then returned with the highest frequency in M phase (Figure 3.6). Interestingly, the disappearance of the interaction in G2 phase correlated with the transcriptional up regulation of an ubiquitin ligase gene that overlapped the fragment (Table 3.2 and Table 3.3) consistent with earlier observations that the disruption of LTR co-localization results in the up regulation of nearby genes (Cam et al., 2008; Lorenz et al., 2012).

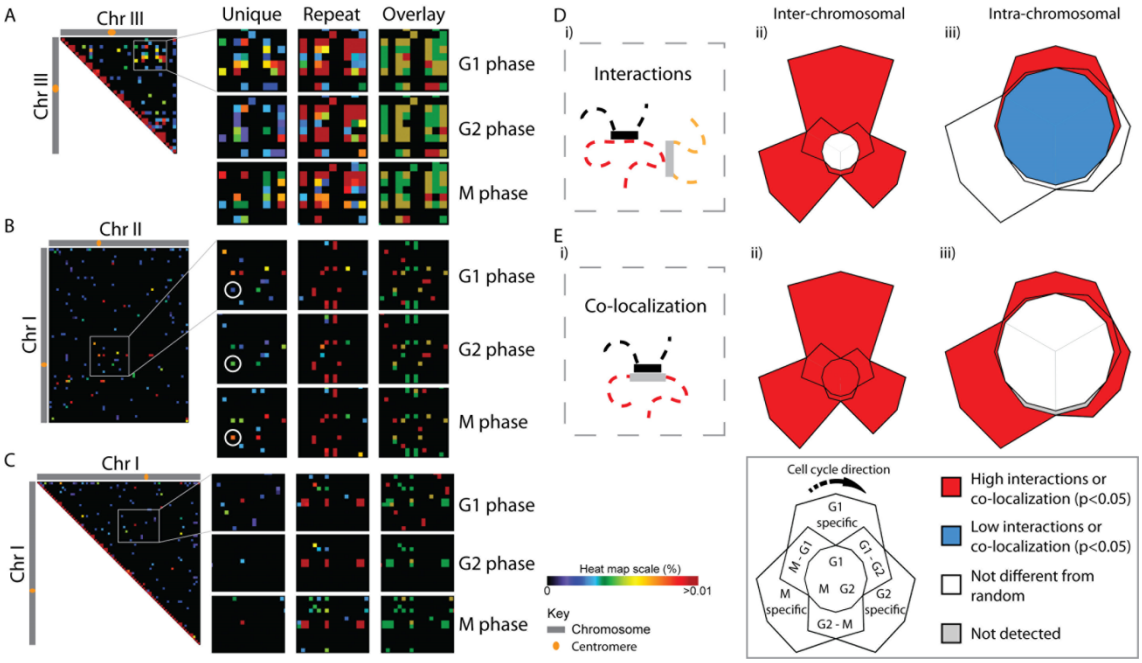


Figure 3.8. Repeat elements contribute to the formation of cell cycle specific genome organization.

Unique interactions that occurred within and between two regions on chromosome III (coordinates: 150-675 Kb and 1.725-2.25 Mb) overlapped extensively with repeat interactions and were consistent throughout the cell cycle (A). There were also changes in the number and strength of unique interactions between (B) and within (C) chromosomes that formed predominately between repeat interactions. Unique and repeat heat maps are scaled as in Fig. 1. The overlay heat maps are interaction frequency independent (unique; red, repeats; green, overlap; yellow) and therefore the heat map scale is not applicable. White circles display unique inter-chromosomal centromere clustering of Chr I and Chr II centromeres. (D and E) LTR elements interacted with the rest of the genome (D(i)) at a significantly high level in most inter- and some intra-chromosomal interaction data sets (D(ii, iii)). Similarly, LTR co-localization (E(i)) was observed at a high level in all inter- E(ii) and most intra-chromosomal E(iii) interaction data sets.

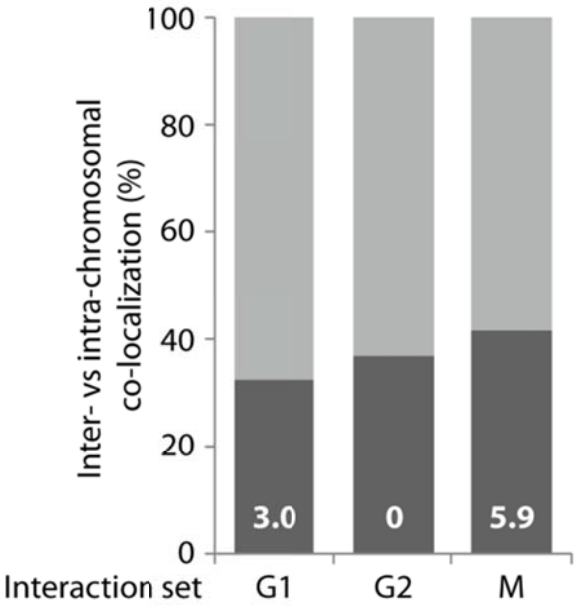


Figure 3.9. Co-localisation between LTR elements occurred predomenantly between chromosomes or LTR elements that were > 50 Kb apart on the same chromosome.

The percentage of detected LTR element co-localization that was inter- (Light grey) and intra-chromosomal (Dark grey) at each phase of the cell cycle was plotted. The percentage of intra-chromosomal co-localization that was between LTR elements <50 Kb apart in the linear sequence is annotated on the dark grey bars.

Table 3.2. The restriction fragments involved in a high frequency interaction detected within chromosome II both contained LTR elements.

Chromosomal coordinates and lengths of the restriction fragments involved in a high frequency intra-chromosomal interaction detected within chromosome II (Figure 3.6). Both fragments contain an LTR element.

Chromosme (Fragment ID)	Coordinates	Fragment length (bp)
Chr II (F1)	2163364 - 2164371	1,007
Chr II (F2)	2423879 - 2432127	9,000

Table 3.3. The absence of the interaction between two LTR elements in chromosome II in G2 phase was associated with the up regulation of an overlapping gene.

An ubiquitin ligase gene overlapped (~70%) one of the fragments involved in the high frequency intra-chromosomal interaction on chromosome II (Table 3.2). The absence of the interaction in the G2 phase was associated with a significant increase in the transcript level of the ubiquitin ligase gene.

	Fold change in transcript level		
Fragment ID (see Table 3.2) and overlapping gene(s)	G1 – G2 phase (significant)	G2 – M phase (significant)	M – G1 phase (significant)
F2: Ubiquitin-protein ligase E3 (SPBC21D10.09c).	-0.3075 (no)	0.9183 (yes)	-0.6108 (no)

3.2.3 CELL CYCLE DEPENDENT INTERACTIONS ARE NOT JUST ABOUT GENE REGULATION

In addition to repeat interactions, the formation of chromosomal contacts can be due to regulatory interactions associated with transcription. To identify chromosomal connections that associated with transcriptional activity in *S. pombe*, we compared interaction sets with matched transcriptomes. The number of genes from specific interaction associated gene sets (Figure 3.10; Venn-diagram keys on left, Figure 3.11) that were shared with transcriptome gene sets (Table 3.4 and Table 3.5) was compared to random gene sets (Figure 3.11). If the transcriptome gene sets were overrepresented in the interaction associated gene sets, the shared genes (see Appendix CD: Supplementary Spread sheet S2) were used for Gene Ontology analysis (see Appendix CD: Supplementary Spread sheet S3).

Genes in the high transcript level gene set (top 5%) were largely conserved between the cell cycle phases (Figure 3.12A). These genes were overrepresented in inter-chromosomal interactions that were shared with the G1 phase of the cell cycle (Figure 3.10A) and the shared genes were enriched in gene ontology (GO) terms related to cell growth, in particular ribosome biogenesis and function (see Appendix CD: Supplementary Spread sheet S3). Similarly, the high transcript level gene set was more likely to be associated with intra-chromosomal interactions that were specific to G1 phase, or shared by the G2 and M phases (Figure 3.10B), with the shared genes again being ribosome genes (see Appendix CD: Supplementary Spread sheet S3). This gene set was also less likely to be associated with intra-chromosomal interactions that formed specifically in the M phase (Figure 3.10B). These results are consistent with a regulatory role for these interactions in the control of the highly transcribed genes.

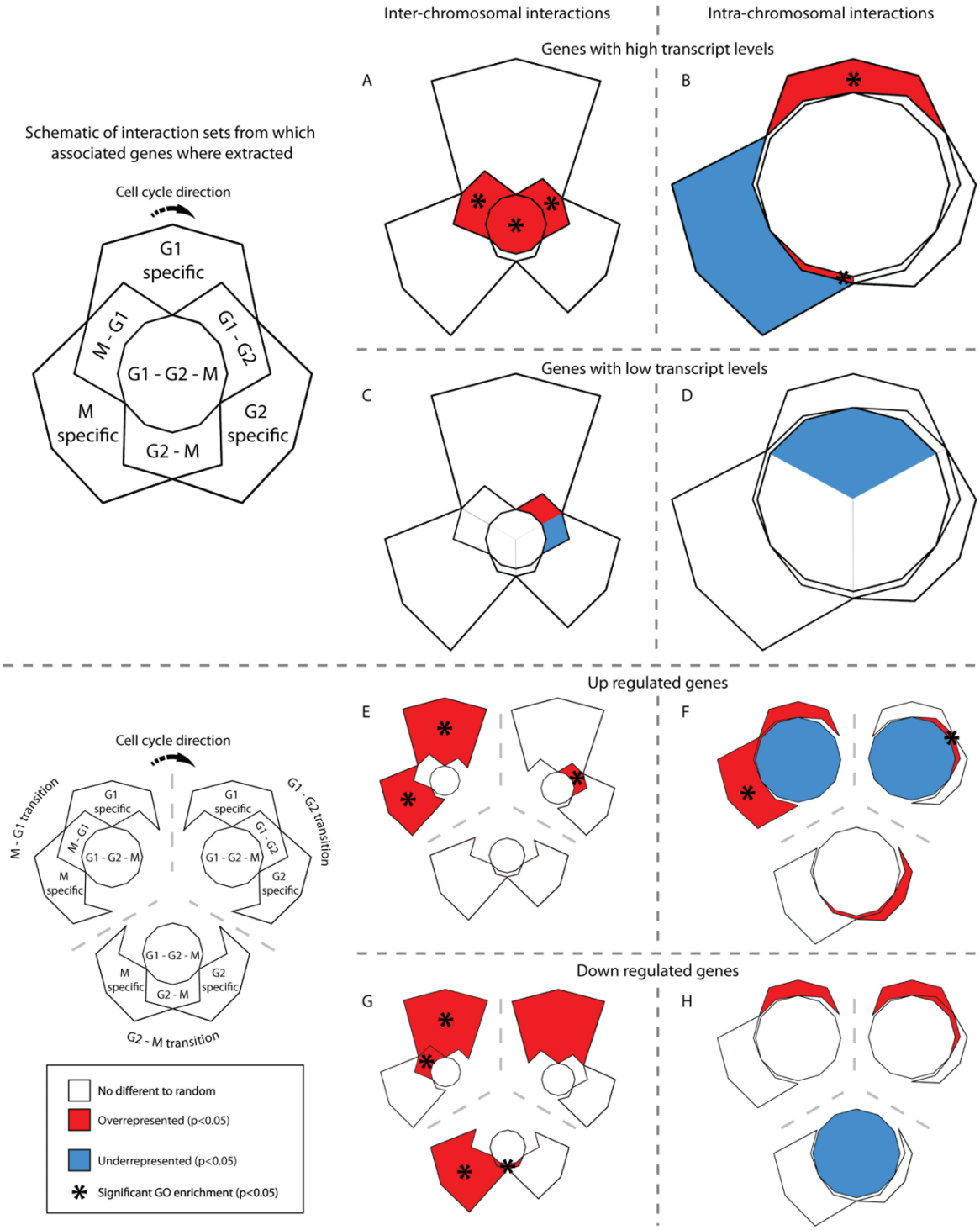


Figure 3.10. Genes that are differentially regulated during the G2 – M – G1 cell cycle transitions are overrepresented within interactions that form during this period of the cell cycle. Refer to footnote [5] for figure legend.

[5] Interaction data was divided into inter- or intra-chromosomal (Table 3.1 and Figure 3.5) and then broken down into seven categories: specific to G1 phase, shared by G1 and G2 phase, specific to G2 phase, shared by G2 and M phase, specific to M phase, shared by M and G1 phase and shared by all phases G1, G2 and M (Keys on left Top and Bottom; Figure 3.5). The genes associated with each of the interactions in the data sets were extracted (Figure 3.11) and compared to cell cycle transcription data (Table 3.4 and Table 3.5) to determine if there was a significant ($p<0.05$) overlap between the two gene sets. If genes were overrepresented (red box) within an interaction set, the genes present in both sets were analysed for GO term

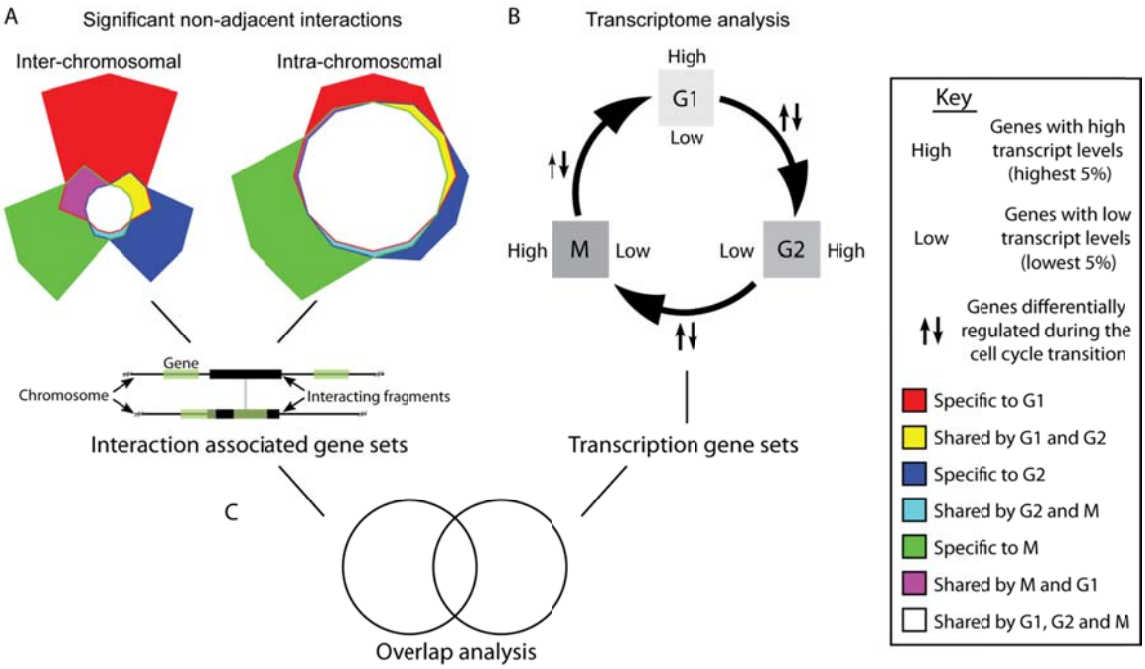


Figure 3.11. Description of gene sets used for GO and overlap analysis.

Genes associated with cell cycle specific inter- and intra-chromosomal interactions were extracted (A) if they overlapped, were internal too, or were up or down stream of the interacting fragments. Transcriptome analysis of the cell cycle identified sets of genes with the highest (top 5%), lowest (bottom 5%) and differentially regulated transcript levels (B). Comparisons between the interaction associated and transcription gene sets were used to determine whether genes in the transcription gene sets were overrepresented, underrepresented, not different from random, or did not overlap the interaction associated gene sets (C).

---

enrichment (see Appendix CD: Supplementary Spread sheet S3). High transcript level genes were more likely to be associated with an interaction in specific inter- and intra-chromosomal data sets (A and B). This is particularly obvious in the cell cycle phases that require the regulation of cell growth; G2 – M – G1, and was reflected in the overlapping genes being enriched in GO terms associated with cell growth, particularly ribosome biogenesis and function (A, B and see Appendix CD: Supplementary Spread sheet S3). The high transcript level genes were also underrepresented in M phase specific intra-chromosomal interactions (B). In contrast, low transcript level genes were predominantly not more likely to be associated with an interaction than random and overrepresented gene sets did not have GO term enrichments (C and D). The differentially regulated genes (up or down regulated during cell cycle transitions) were overrepresented in many of the interaction data sets again associated with phases of the cell cycle where cell growth is regulated (G2 – M – G1) and were enrichment in GO terms related to cell growth (see Appendix CD: Supplementary Spread sheet S3). These genes were never overrepresented in interaction sets shared by all three cell cycle phases.

The genes that had the lowest (5%) associated transcript levels were largely specific to each phase of the cell cycle (Figure 3.12B). These genes were only overrepresented in inter-chromosomal interactions that were shared during the G1-G2 phases (Figure 3.10C). However, there was no GO term enrichment meaning they did not share any common biological function, component or compartment (Figure 3.10C).

The gene sets that contained genes whose transcript levels were differentially regulated during the cell cycle transitions were over and underrepresented in subsets of the inter- and intra-chromosomal interactions (Figure 3.10E-H). Moreover, this overrepresentation was frequently associated with GO term enrichment (Figure 3.10E-H; see Appendix CD: Supplementary Spread sheet S3). Notably, the level of representation of the differentially regulated genes within the interaction sets did not directly correlate with the direction of transcript level change (*i.e.* up regulated genes were not simply overrepresented in the interaction set). Rather, the relationship between interaction formation and changes in transcript levels is complex. For example, genes whose transcript levels increase during the M-G1 phase transition are overrepresented in M and G1 phase specific interactions (Figure 3.10E and F). Similarly, genes that are up regulated during the G1-G2 phase transition are overrepresented in interactions that form in the G1 phase and are maintained through S phase into the G2 phase of the cell cycle (*i.e.* shared by the two consecutive cell cycle phases; Figure 3.10E and F).

Genes whose transcript levels were down regulated during the M-G1 and G2-M phase transitions were overrepresented in interactions that were shared between the two, or specific to the phase in which they were down regulated (Figure 3.10G and H). By contrast, genes that were down regulated during the G1-G2 phase transition were overrepresented in interactions that formed in the G1 phase (Figure 3.10G). These results are consistent with the idea that the association of specific genes with the formation and disruption of interactions facilitates their correct cell cycle specific regulation.

Table 3.4. The number of genes that were significantly differentially regulated during each *S. pombe* cell cycle transition.

RNA-seq data was analysed using cufflinks to determine genes that were significantly up and down regulated during each *S. pombe* cell cycle transition: G1 – G2 phase, G2 – M phase, and M – G1 phase. The number of genes that had a >2-fold change in transcript level is also displayed.

	G1 - G2 phase	G2 - M phase	M - G1 phase
Total number of genes differentially expressed	**198**	**346**	**239**
Number of all significantly up regulated genes	102 (51.51%)	138 (39.88%)	150 (62.76%)
Number of all significantly down regulated genes	96 (48.49%)	208 (60.12%)	89 (37.24%)
Genes with a differential expression >=2 (percentage of total)	**91 (45.96%)**	**142 (41.04%)**	**70 (29.29%)**
Number of genes up regulated (cut-off >= 2)	77 (84.62%)	46 (32.39%)	26 (37.14%)
Number of genes down regulated (cut-off <= -2)	14 (15.38%)	96 (67.61%)	44 (62.86%)

Table 3.5. Table highlighting the number of genes that had the highest (top 5%) and lowest (bottom 5%) transcript levels at each cell cycle phase and whether they were differentially regulated during cell cycle transitions.

The output file generated by RNA-seq analysis was used to determine genes that had the highest (top 5%) and lowest (bottom 5%, excluding genes that were not expressed) transcript levels at each phase of the cell cycle. The number of genes within these genes sets that was significantly up or down regulated entering each cell cycle phase was determined.

	G1 phase	G2 phase	M phase
Number of genes in the highest or lowest expressed 5%	195	197	196
Number of highly expressed genes that significantly change their expression	6 (3.08%)	17 (8.63%)	19 (9.69%)
High and up regulated	1	4	7
High and down regulated	5	13	12
Number of lowly expressed genes that significantly change their expression	41 (21.03%)	12 (6.09%)	9 (4.59%)
Low and up regulated	41	12	8
Low and down regulated	0	0	1

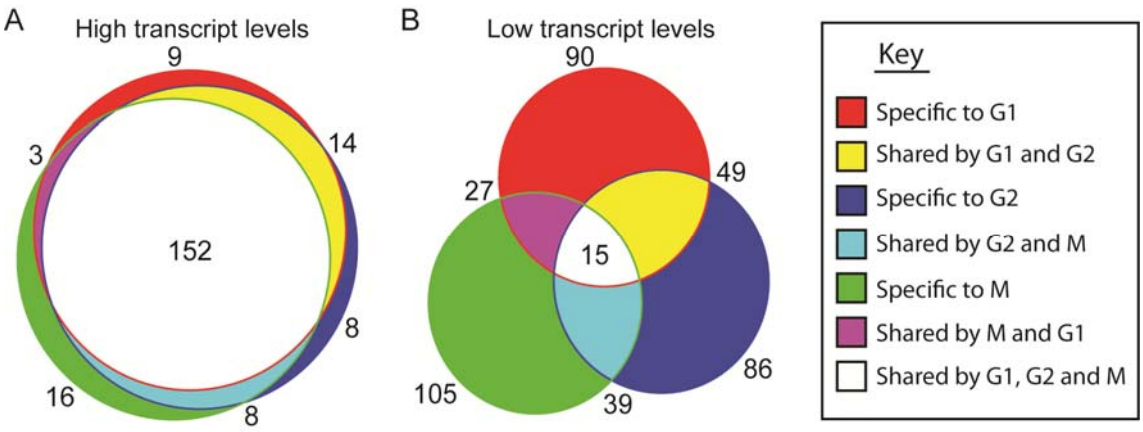


Figure 3.12. Genes with high transcript levels are highly conserved throughtout the cell cycle while low transcript level genes are not.

The majority of genes with high transcript levels were conserved at each cell cycle phase A). By contrast, genes with low transcript levels were largely specific to each cell cycle phase B). Proportional Venn-diagrams of the overlap between genes with the highest and lowest transcript levels at each phase of the cell cycle are displayed.

3.2.4 Genes with high transcript levels throughout the cell cycle are associated with a high proportion of inter-chromosomal co-localization.

We investigated whether genes with high or low transcript levels occupied different spatial environments. We determined if the frequency with which the genes that had high or low transcript levels interacted with any other genomic region was different from random (Figure 3.13A). Genes with high transcript levels interacted with other genomic loci at a high frequency in the inter- (Figure 3.13B(i)) and intra-chromosomal (Figure 3.13B(ii)) G1 phase specific and shared interaction data sets. Genes with low transcript levels only interacted with other genomic loci at a high frequency in the G1 phase specific intra-chromosomal interactions (Figure 3.13C(i) and (ii)). Therefore, genes with high transcript levels are highly connected to other chromosomal loci while low transcript level genes are not.

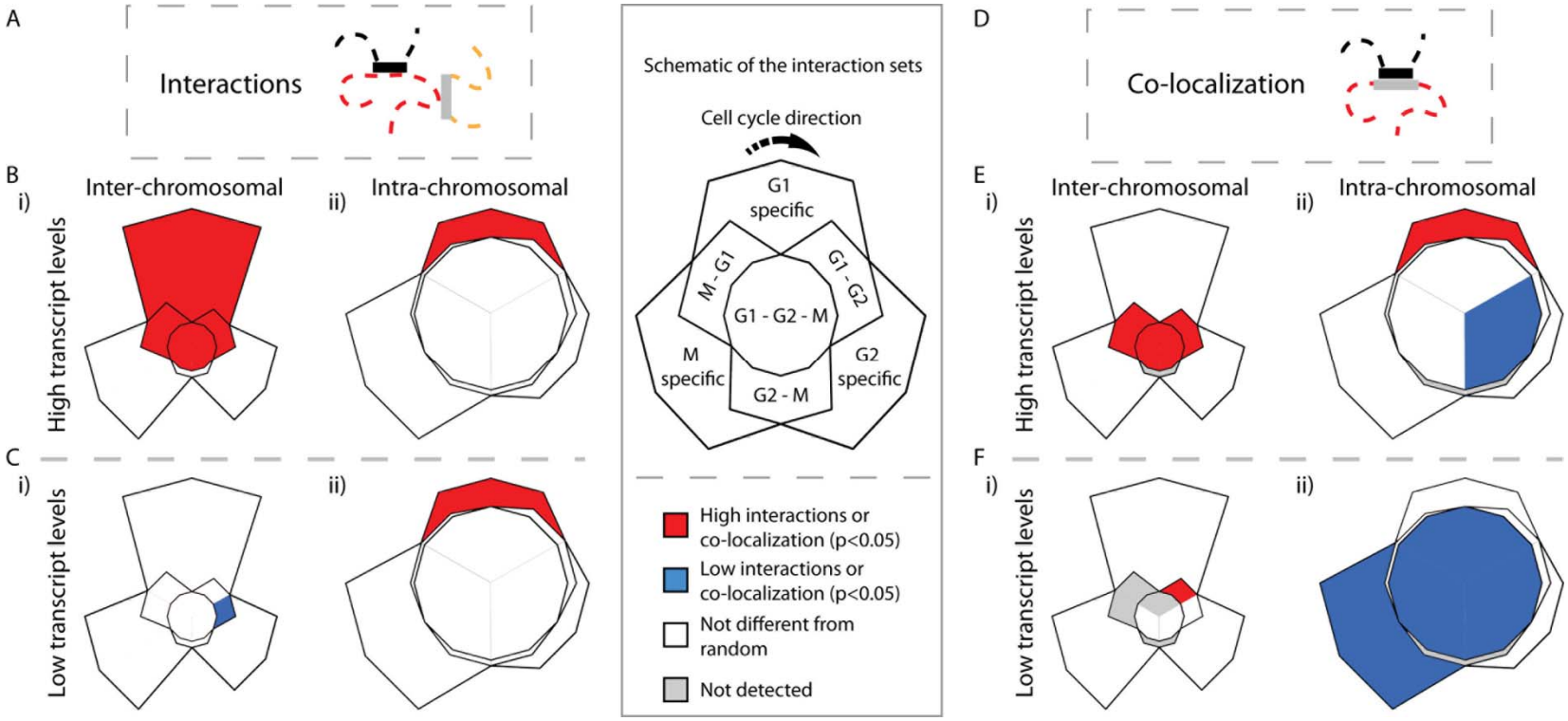


Figure 3.13. Genes with high transcript levels are more highly connected to the genome and co-localize more frequently than low transcript level genes.

To investigate if genes with high and low transcript levels occupied specific spatial environments the frequency that these genes interacted with the genome (A) or co-localized with each other (D) was compared to random genomic regions in specific interaction data sets (Key, middle). Proportional Venn-diagrams of the inter- and intra-chromosomal interaction data sets are coloured to represent the level at which the genes interact with the genome (B and C) or co-localize with each other (E and F). Genes with consistently high transcript levels interacted with the genome at a high frequency in specific inter- and intra-chromosomal interaction sets (B(i) and (ii)). Genes with low transcript levels did not interact with the genome at a higher than random frequency in the inter-chromosomal interactions (C(i)) but did in the G1 phases specific intra-chromosomal interactions (C(ii)). High transcript level genes co-localized with each other at a high frequency in similar inter- and intra-chromosomal interaction sets in which they interacted with the genome at a high frequency (E(i) and (ii)). In contrast, there was only one high level of co-localization of genes with low transcript levels in the inter-chromosomal interactions (F(i)) and low levels in intra-chromosomal interactions (F(ii)).

We then looked specifically at whether genes with high or low associated transcript levels co-localize with each other. The high transcript level genes co-localized with each other at a significantly high frequency in interaction data sets shared with the G1 phase of the cell cycle (Figure 3.13E(i) and (ii)) that overlap those in which they were highly connected with the genome (Figure 3.13B(i) and (ii)). Co-localization between these highly transcribed genes occurs predominantly between chromosomes (>70% inter-chromosomal co-localization; Figure 3.14) and those interactions that do occur within chromosomes (intra-chromosomal co-localization) are between distal genes (>50 Kb apart; Figure 3.14). High transcript level genes also had a non-random linear chromosomal distribution across one or more chromosomes at each cell cycle phase (Figure 3.15). In contrast, genes with low transcript levels only co-localized with each other at a significantly high frequency in interactions that were shared between the G1 and G2 phases of the cell cycle (Figure 3.13F(i)). There were also a number of gene sets with no detectable co-localization between chromosomes (Figure 3.13F(i)) or significantly low levels of co-localization within chromosomes (Figure 3.13F(ii)). In comparison to genes with high transcript levels, co-localization of genes with low transcript levels was predominantly due to a significantly higher proportion of intra-chromosomal interactions (>60%) relative to the respective high transcript level genes (Figure 3.14). There was also an increase in the proportion of localized intra-chromosomal co-localization (<50 Kb apart) compared to the highly transcribed genes. These data are consistent with the high and low transcript level genes at each phase of the cell cycle existing in distinct spatial interacting environment. The genes with high transcript levels were more highly connected with other genomic loci and co-localized more frequently than genes with low transcript levels.

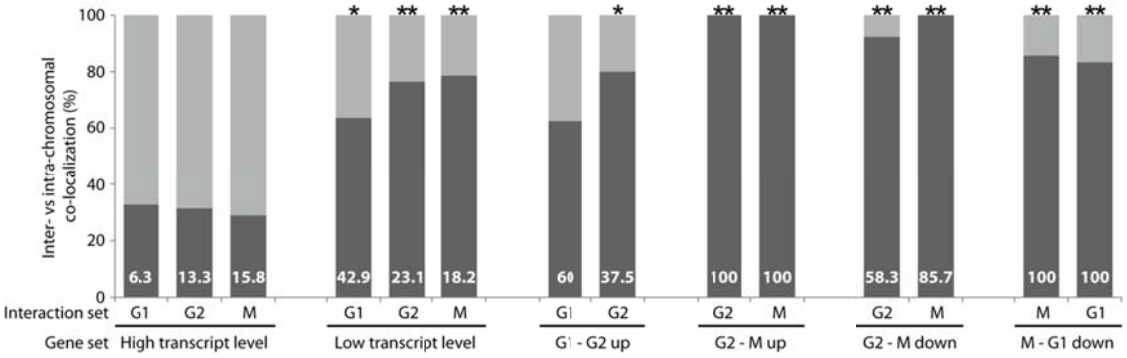


Figure 3.14. The percentage of inter- compared to intra-chromosomal co-localization for genes with high, low and differential transcript levels.

Co-localization between genes with high transcript levels is predominantly due to inter-chromosomal and long distance (>50 Kb) intra-chromosomal gene interactions. Genes with low transcript levels exhibited a significant reduction in co-localization with other lowly transcribed genes on other chromosomes and a corresponding increase in localized (<50 Kb) interactions, when compared to high transcript level gene sets. Co-localization detected between differentially regulated genes (>2-fold change in transcript level) was predominantly intra-chromosomal. Furthermore, changes in intra-chromosomal gene co-localization that correlated with differential gene expression were predominantly due to localized (<50 Kb) interactions. Inter- and intra-chromosomal clustering interactions are depicted for; G1, G2 and M phase high and low transcript level genes, G1 – G2 and G2 – M up regulated genes, and G2 – M and M – G1 down regulated genes. Proportion of inter-chromosomal co-localization; Light grey, Proportion of intra-chromosomal co-localization; Dark grey, The proportion of intra-chromosomal co-localization that was between gene that were <50 Kb from each other in the linear sequence; white numbers.

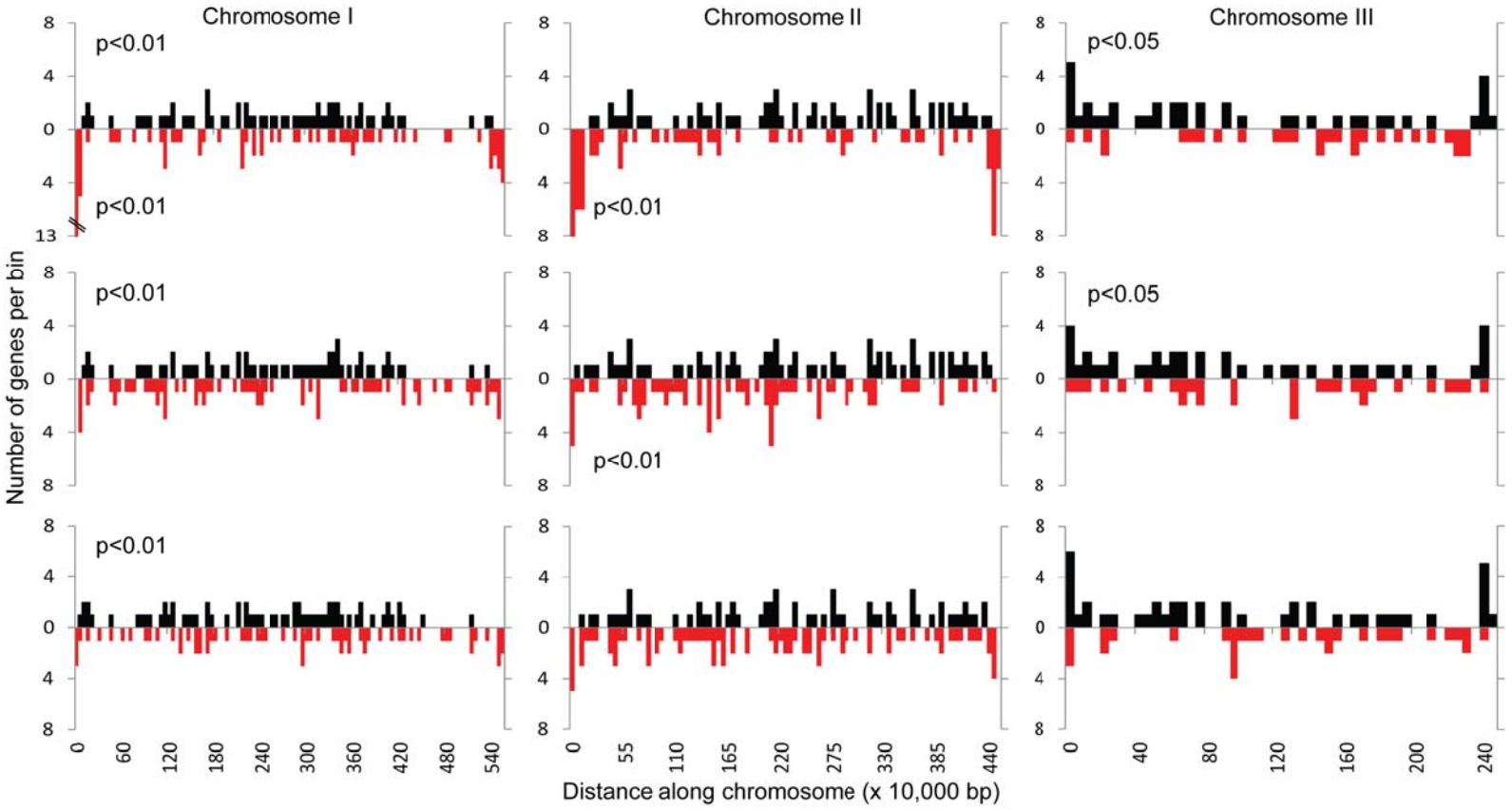


Figure 3.15. Genes with high transcript levels were non-randomly distributed across one or more chromosomes at each stage of the cell cycle.

The chromosomal position of genes with the highest (Black) and lowest (Red) transcript levels within a specified proportion (5%) of the total number of genes tested was plotted as a histogram (50 Kb bins); Top) G1 phase, Middle) G2 phase, and Bottom) M phase. Genes with high transcript levels had a non-random linear chromosomal distribution across one or more chromosomes at each phase of the cell cycle. By contrast, genes with low transcript levels were predominantly non-randomly distributed along chromosomes only in G1 phase. One-sample Kolmogorov-Smirnov tests were performed using the non-binned chromosome distribution data to determine if highly and lowly transcribed genes were non-uniformly distributed across each chromosome.

3.2.5 GENES DIFFERENTIALLY REGULATED DURING THE G1 – G2 – M CELL CYCLE PHASE TRANSITIONS WERE LINEARLY AND SPATIALLY CO-LOCALIZED

Changes in gene transcription have been associated with changes in the three-dimensional position of a gene(s) and the formation or disruption of interactions can contribute to gene regulation. We investigated whether genes that showed a >2-fold change in transcript level during cell cycle transitions (Table 3.4) interacted with themselves and other genomic loci more or less than would be expected at random (Figure 3.16A-F). These gene sets are specific to each cell phase transition and there is no one obvious rule that explains how the genes in these sets connect to each other or to other genomic loci. However, there are numerous interesting examples that demonstrate that the environments these genes are in change during differential expression. For example, genes up regulated during the G2-M phase cell cycle transition are only highly connected with loci on other chromosomes in the G2 phase specific interactions (Figure 3.16B(i)). Moreover, these up regulated genes are found to be lowly connected to other loci within the same chromosome in M phases specific interactions (Figure 3.16B(ii)). Interestingly, a subset of the genes that were down regulated during the G2-M phase cell cycle transition were always highly connected with loci on other chromosomes within the G1-G2-M shared interactions (Figure 3.16C(i)).

To investigate whether the spatial co-localization of genes facilitates their co-regulation, we determined if the up and down regulated gene sets spatially co-localized (Figure 3.16D). For many of the differentially regulated gene sets there was no detectable co-localization, indicating that clustering is not a pre-requisite for co-regulation. However, it was notable that genes that were up regulated during the G1-G2 cell cycle transition had high levels of inter-chromosomal co-localization in interactions specific to the G1 and shared between the G1 and G2 cell cycle phases (Figure 3.16E(i)). These genes were also highly co-localized in the G1 and G2 phase shared and the G2 phase specific intra-chromosomal interactions (Figure 3.16E(ii)). Similarly, clustering of genes that were down regulated during the G2-M phase transition is significantly high in G2 phase specific intra-chromosomal interactions (Figure 3.16F(ii)), but is not detected in inter-chromosomal interactions (Figure 3.16F(i)). Interestingly, as observed for genes with high transcript levels, genes within the G1-G2 up and G2-M down regulated gene sets also had a non-random linear chromosomal distribution on

chromosomes I and II (Figure 3.17). Collectively, these results indicate that co-localization is not required for the co-regulation of genes but may facilitate it at specific cell cycle phases.

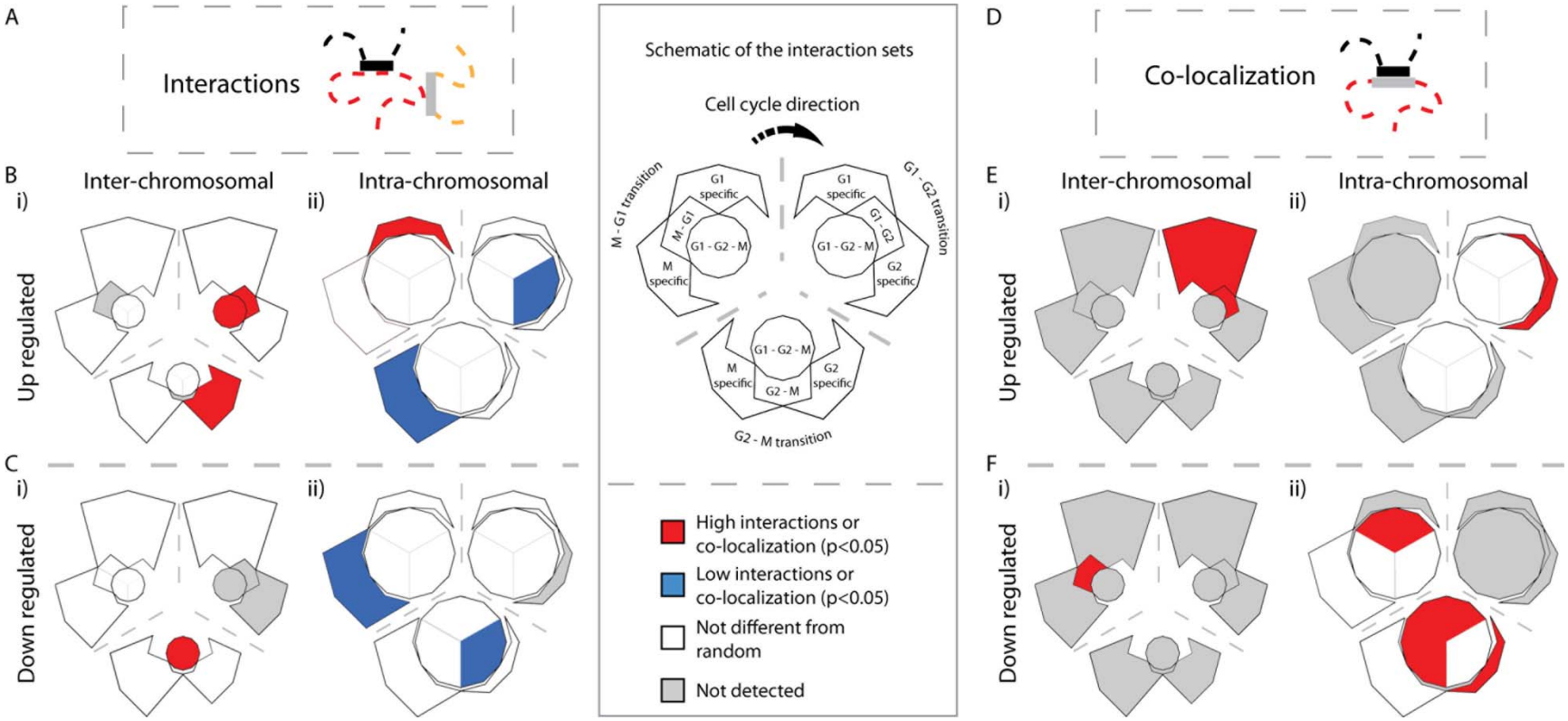


Figure 3.16. Genes that undergo differential expression during cell cycle phase transitions are found in specific interaction environments and co-localize in a cell cycle specific way.

To investigate if differentially regulated genes with a >2-fold change in transcript level occupied specific spatial environments the frequency that these genes interacted with the genome (A) or co-localized with each other (D) was compared to random genomic regions in specific interaction data sets (Key, middle). Proportional Venn-diagrams of the inter- and intra-chromosomal interaction data sets are coloured to represent the level at which the genes interact with the genome (B and C) or co-localize with each other (E and F). Both up and down regulated genes changed their connectivity with the genome in a cell cycle specific inter- (B(i) and C(i)) and intra-chromosomal (B(ii) and C(ii)) interaction data sets. Up regulated genes were found to be highly connected with the genome more often than down regulated genes. The up and down regulated genes showed a low propensity to co-localize with each other with the majority of significantly high co-localization detected for genes up regulated during the G1 – G2 transition (E) and then down regulated during the G2 – M transition (F).

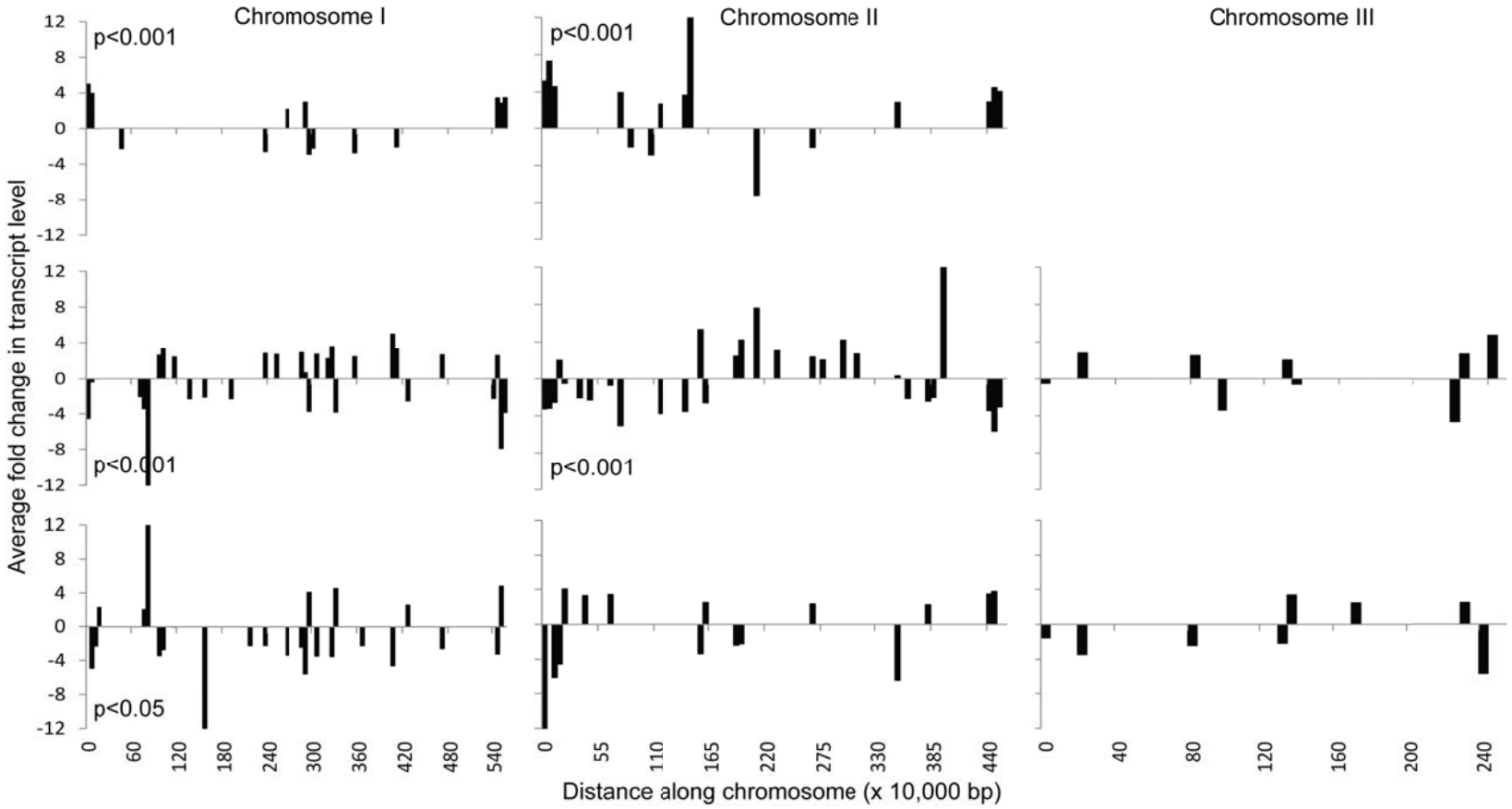


Figure 3.17. Cell cycle specific differentially regulated genes are non-randomly distributed across the linear chromosomal sequences.

The chromosomal position of up and down regulated genes (>2-fold change in transcript level) was used to calculate the average fold change in transcript level per bin (50 Kb), which was plotted as a chromosomal distribution for; G1 – G2 (top), G2 – M (middle), and M – G1 (bottom) cell cycle phase transitions. Genes that were up regulated during the G1 – G2 transition and down regulated during the G2 – M transition had a highly significant non-random chromosomal distribution on chromosomes I and II. One-sample Kolmogorov-Smirnov tests were performed using the non-binned chromosome distribution data to determine if the differentially expressed genes were non-uniformly distributed across each chromosome.

3.3 DISCUSSION

In this study, we identified differences in genome organization between the G1, G2 and M phases of the Fission yeast cell cycle, providing a high-resolution molecular analysis of *in vivo* metaphase chromosome structure. Known hallmarks of Fission yeast genome organization, such as the co-localization of centromeres and telomeres (Alfredsson-Timmins et al., 2009; Olsson & Bjerling, 2011) were present at each phase of the cell cycle. Mixtures of stable and dynamic interactions were detected within and between chromosomes. As is commonly observed, the majority of DNA interactions occurred within chromosomes thus supporting the presence of territories in Fission yeast, consistent with earlier observations (Scherthan et al., 1994; Tanizawa et al., 2010).

Polymer models have been used to investigate how interphase chromosomes decondense during the cell cycle ((Bohn, Heermann, & van Driel, 2007; Cook & Marenduzzo, 2009; Dorier & Stasiak, 2009; Rosa & Everaers, 2008) reviewed in (Dekker, Marti-Renom, & Mirny, 2013)). Intriguingly, we observed that a subset of interactions within and between chromosomes was maintained through the three cell cycle phases. This indicates that the relative spatial positioning of chromosomes and a degree of internal chromosome structure is inherited throughout the cell cycle. The conservation of interactions between the right and left telomeres on chromosomes I and II, and the regions immediately adjacent to the ribosomal repeat regions on chromosome III, means that the chromosomes are effectively circular in all cell cycle phases. This circular organization may contribute to the stabilization of chromosome ends. Moreover, our findings are consistent with polymer models that demonstrate circular chromosome structures are requirements for de-condensation of chromosomes and the formation of chromosome territories, within biologically relevant timescales (Dorier & Stasiak, 2009; Rosa & Everaers, 2008).

The structure of isolated metaphase chromosomes has been extensively investigated (Matsuda et al., 2010; Ohta et al., 2011; Poirier & Marko, 2002). Metaphase chromosomes are thought to contain a continuous chromatin network that is constrained by isolated chromatin-crosslinking elements spaced by ~15 Kb (Poirier & Marko, 2002). The abundance of loops with a length ≤5 Kb in M phase Fission yeast chromosomes is *in vivo* evidence for the formation of a continuous chromatin network. However, the absence of a predominant loop length suggests

that there was no regular coiling of the chromatin fibre. Instead our results are consistent with the chromosomes assuming dynamic 'polymer melt' like structures (Maeshima et al., 2010; Naumova et al., 2013).

Early observations of connections between metaphase chromosomes (Hoskins, 1968; Korf & Diacumakos, 1978) were thought to be an artefact of chromosome isolation (Korf & Diacumakos, 1980). Later work demonstrated that these interactions occurred and were DNA based (Kuznetsova et al., 2007; Maniotis et al., 1997). Our results confirm these findings and provide the first evidence for DNA based connections between metaphase chromosomes within lower eukaryotes. Moreover, our finding that there is a high level of co-localization between LTR elements, from different chromosomes, which are bound by the CENP-B homolog Abp1 (Cam et al., 2008; Lorenz et al., 2012), implicates repeat regions as participating in these M-phase inter-chromosomal linkages. This is consistent with the finding that satellite DNA was involved in thread formation between mitotic chromosome in mouse cell lines and that CENP-B was a component of the thread (Kuznetsova et al., 2007).

Transcriptional silencing of LTR elements and associated genes is achieved by the recruitment of class I and II histone deacetylases (HDACs) to these elements and their association with *Tf* bodies (Cam et al., 2008; Lorenz et al., 2012). We interpret the high level of co-localization between LTR elements on different chromosomes, and at long distances within chromosomes (*i.e.* >50 Kb), as demarcating chromatin domains. Moreover, the correlation between the Ubiquitin-protein ligase E3 - LTR interaction and Ubiquitin-protein ligase E3 transcript levels supports a role for LTRs in the regulation of transcription at a distance. The finding that LTRs exhibit cell cycle phase specific interactions with other loci is consistent with a coordinating role in transcription regulation. This is further supported by the observation of widespread up regulation of transcription in CENP-B mutants (Lorenz et al., 2012).

Highly expressed genes have been shown to preferentially co-localize in Fission yeast (Tanizawa et al., 2010). These findings are often interpreted as indicating that transcription and/or transcription factories are involved in the spatial organization of genomes (Papantonis & Cook, 2013). In our data, the finding of a conserved set of genes that had high transcript levels throughout the cell cycle and their non-random linear chromosomal distribution is consistent with the existence

of regions of increased gene expression (Caron et al., 2001). By contrast, genes with low transcript levels were predominantly cell cycle phase specific and less likely to occur in linear clusters. Moreover, we observed that genes with consistently high transcript levels were highly connected whereas genes with low transcript levels were not. The fact that we did not determine nascent transcription levels prevents us concluding that the act of transcription was responsible for interaction formation. However, the high level of genome connectivity between constitutively highly transcribed genes can be interpreted as reflecting an extended association with transcription factories when compared to individual, cell cycle specific lowly transcribed genes. Moreover, the observation that interactions that formed in the G1 phase were maintained through the S phase into the G2 phase: 1) implicates these interactions in the maintenance of transcript levels and hence transcription; and 2) indicates that interactions are re-established during chromosome replication. Furthermore, the observation that genes with consistently high transcript levels were enriched in interactions that formed in the M phase and were maintained in the G1 phase is consistent with bookmarking facilitating post-mitotic reactivation (Kadauke & Blobel, 2013; Sarge & Park-Sarge, 2005; R. Zhao et al., 2011).

Differentially regulated genes do not exhibit a clear relationship between interactions or connectivity with the genome and the direction of transcriptional change. For example, genes that were up or down regulated during a specific cell cycle transition were highly enriched within specific subsets of interactions. These results indicate that interactions have both positive and negative effects on transcription, consistent with the existence of interactions involving activator and repressor elements or specific chromatin states (Simonis et al., 2006; Smallwood & Ren, 2013).

Co-localization between genes is not a requirement for their co-regulation (Kocanova et al., 2010), but may contribute to it in specific situations (Schoenfelder et al., 2010). Genes that were up regulated during the G1-G2 and down regulated during the G2-M phase cell cycle transitions were co-localized within specific sets of interactions. Moreover, the genes within the G1-G2 up regulated and G2-M down regulated gene sets have a non-random chromosomal distribution, similar to that observed for genes with consistently high transcript levels. This is consistent with the up regulation of specific gene sets being aided by the linear organization of genes in regions of increased gene expression (Caron et al., 2001). It remains

likely that the observed linear clustering of specific gene sets is under evolutionary selection for increased spatial co-localization and thus more efficient co-regulation.

A number of the genes that were differentially regulated during the G1-G2 (transcripts up) and G2-M (transcripts down) cell cycle transitions clustered in sub-telomeric regions (150 Kb) on chromosomes I and II (Including genes identified to form a telomere cluster by Oliva et al. 2005; see Appendix CD: Supplementary Spread sheet S4). The telomere regions of chromosomes I and II also displayed cell cycle specific changes in co-localization. In Fission yeast, the co-localization of telomeric regions is disrupted in RNAi mutants (Hall et al., 2003; Kanoh, Sadaie, Urano, & Ishikawa, 2005). Collectively, these observations suggest an RNAi dependent link between differential telomere gene expression and telomere co-localization. The observation of dynamic changes in telomere clustering throughout the cell cycle in human cells may indicate similar processes occur in these cells (Ramírez & Surrallés, 2009).In conclusion, we show that connections always exist between chromosomes, irrespective of the cell cycle phase, and these interactions are associated with specific elements and nuclear processes (*e.g*. transcription and chromatin de-condensation). This suggests: 1) that the structural connections move in response to, or to allow, cell cycle progression; or 2) that there are separate populations of structural and regulatory connections that participate in the maintenance and/or establishment of cell cycle specific chromatin and gene regulation. Moreover, our results provide support for a commonly emerging theme that high levels of transcription correlate with increased genome connectivity and co-localization. Perhaps most importantly, there are cell cycle specific changes in both the interactions and transcription patterns. Thus, it is clear that the use of asynchronous cultures introduces ambiguity to the understanding of the relationship between the structure and function of the genome. Future work combining specific mutations with the type of analyses we performed will facilitate an increase in our understanding of the spatial and functional organization of the nucleus through time.

3.4 CONCLUSION

Our results demonstrate the existence of cell cycle specific chromosome interactions within the *S. pombe* genome. The dynamic nature of these interactions, and the observed correlation with alterations to transcript levels, indicates that the interactions are regulated and likely regulatory in nature. The observation that subsets of these interactions are maintained even when

chromosomes are fully condensed implicates genome organization in epigenetic inheritance and bookmarking. In addition, our data provides support for *S. pombe* metaphase chromosomes assuming a 'polymer melt' like structure, remaining interconnected and circular even when fully condensed. In conclusion future insights into the role of genome organization in the genotype to phenotype translation will require the integration of temporal and functional data.

3.5 MATERIALS AND METHODS

3.5.1 STRAINS, GROWTH CONDITIONS AND SYNCHRONIZATION

Schizosaccharomyces pombe strains (Table 3.6) were recovered from -80°C on YES (Sabatinosa & Forsburga, 2010) (2% agar) plates (26°C, 4 days). YES medium (12 ml) starter cultures were inoculated and incubated (26°C, 200 rpm) until the OD_{595} measured ~0.8 (~24 h). Synchronization cultures (125 ml EMM2 (Sabatinosa & Forsburga, 2010), in baffled flasks) were inoculated with starter culture to an OD_{595} = ~0.05 and incubated (26°C, 120 rpm). Cultures were grown for four generations (OD_{595} ~0.8) before synchronization was induced by the addition of pre-warmed EMM2 medium (125 ml, 46°C), instantly rising the temperature of the culture to the restrictive temperature (36°C). Cultures were incubated in a hot water bath (36°C, 140 rpm, for 4 h) to complete synchronization. Cells for synchronization efficiency analysis were harvested from cultures before induction and following synchronization (1 ml, 4,000 rpm, 2 min), and snap frozen (dry ice/ethanol (100%) bath).

Table 3.6. *Schizosaccharomyces pombe* strains used in this study.

The strains were obtained from the National BioResource Project – Yeast (http://yeast.lab.nig.ac.jp/nig/index_en.html).

Strain Name	Genotype	Reference
MY291	h- lue1 cdc10-129	(Aves et al., 1985)
MY284	h- lue1 cdc25-22	(Nurse et al., 1976)
MY286	h- lue1 nuc2-663	(Hirano et al., 1988)

3.5.2 SYNCHRONIZATION EFFICIENCY

Cells collected during synchronization were thawed, washed once with ice-cold 1% PBS (500 µl, 4,000 rpm, 2 min) and suspended in PBS (100 µl). Cells were stained with calcofluor white (1g/L with 10% Potassium Hydroxide) and DAPI (25 mg/ml) and photographs were taken of each sample before and following

synchronization using a fluorescence microscope (ZEISS, HBO 100 Axiostart plus). The level of cell cycle phase synchronization was calculated for the G1 and G2 phases by comparing the proportion of cells that had a septum, in >200 cells, in the synchronized cell populations compared to the corresponding pre-synchronized populations (Figure 3.1, Figure 3.18 and Table 3.7). The estimation of >80% synchronization for M phase cells was based on the observation of characteristic traits described for synchronous cultures (Hirano et al., 1988); increased septation index (i.e. increased from ~16% to ~50%), highly condensed chromosomes, and the presence of enucleate cells, following DAPI staining.

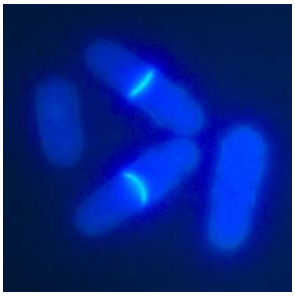


Figure 3.18. The level of cell cycle phase synchronization was calculated using the septation index of *S. pombe* cells before and after synchronization.

Photographs of *S. pombe* cells stained with calcofluor white were taken before and after synchronization using a fluorescence microscope. The proportion of >200 cells that had a visible septum was calculated and used to estimate the synchronization efficiency of each biological replicate.

Table 3.7. An example calculation of the cell culture synchronization efficiency.

The synchronization efficiency for each biological replicate was calculated by comparing the proportion of cells with a septum before and after synchronization.

G2 phase (cdc25-22) biological replicate #1		
	Before synchronization	After synchronization
Total cells counted	225	204
Number with a visible septum	48	2
Percentage	21.33	0.98
Synchronization efficiency	100 – ((0.98 / 21.33) x 100) = **95.41%**	

3.5.3 Chromatin isolation for Genome Conformation Capture (GCC)

Chromatin isolation and GCC were performed as in Rodley et al. 2009 and Cagliero et al. 2013, with modifications; Following synchronization, cultures (200 ml) were cross-linked, washed, and suspended in FA-lysis buffer. Aliquots containing ~$9.5*10^8$ cells were made up to a volume of 330 µl with FA-lysis buffer

and the cell walls were digested with T20 Zymolyase (70 µl at 75mg/ml; 35°C, 40 min with periodic inversion) before heat inactivation (60°C, 5 min). Acid washed glass beads (500 µl) were added to each sample before disruption in a Geno/Grinder (-20°C; 1,750 rpm, 2 x 30 s on 60 s off; SPEX sample prep 2010). Glass beads were removed by the centrifugation of chromatin through a pin hole into a clean tube (2,000 rpm, 1 min). Chromatin was pelleted (13,000 rpm, 15 min, 4°C), washed with FA-lysis buffer, suspended in chromatin digestion buffer and stored (-80°C).

Each chromatin sample was aliquoted into ten sets of $9.5*10^7$ cells. Samples were digested with AseI (100U, New England Biolabs, 37°C, 2 h). A ligation control (sees section 3.5.4 and Table 3.8) was added to the AseI digested chromatin, samples were diluted (~20-fold) and ligated with T4 DNA ligase (20U, Invitrogen). Following ligation, cross-links, protein and RNA were removed. pUC19 plasmid (27.4pg/2ml) was added as a sequence library preparation ligation control before phenol:chloroform (1:1) extraction and column purification. GCC libraries (3 µg of each sample) were sent for paired-end (PE) sequencing (50 bp BGI China).

3.5.4 Production of external ligation controls for GCC library preparation

External ligation controls were produced as in Cagliero et al. 2013 with an AseI restriction enzyme site at one end (Table 3.8) from the *E. coli* genome, Lambda phage genome and pRS426 plasmid (Cagliero et al., 2013). The digested PCR products ($9.5*10^7$ copies) were introduced into the GCC samples (*i.e. E. coli*: G1 phase, pRS426: G2 phase, Lambda: M phase) prior to the ligation step of the GCC protocol. Following sequencing, only one ligation event was detected between the pRS426 ligation control and an AseI fragment in one of the G2 phase biological replicate. A number of ligation events were also detected between the *S. pombe* genome and the pUC19 control (G1 phase: 14, G2 phase: 7, and M phase: 2), indicating that inter-molecular ligation events occurred during preparation for sequencing at the BGI.

Table 3.8. Ligation controls used in this study.

Three short DNA sequences were amplified from the E. coli genome, pRS426 plasmid, and Lambda phage DNA. An AseI site (red) was introduced into each product using the reverse (AseIR) primer. The PCR products were purified digested with AseI and introduced into the GCC samples (at a 1:1 ratio with genome/cell number) before ligation to control for random inter-molecular ligation events.

Primer name	Sequence	Product Length (bp)
E.coli191bp3'AseIF	TAGGCAGGATAAGGCGTTCA	191
E.coli191bp3'AseIR	GTGATTAATGCGGTCTGATGAGTCGTTTC	
pRS426_185bp3'AseIF	TTGGTCTGACAGTTACCAATGC	185
pRS426_185bp3'AseIR	GTGATTAATGATAAATCTGGAGCCGGTGA	
Lambda187bp3'AseIF	TTTACAGCGTGATGGAGCAG	187
Lambda187bp3'AseIR	GTGATTAATACCAATCCAGCCGGTCAG	

3.5.5 NETWORK ASSEMBLY

GCC networks were constructed from 50 bp PE Illumina Genome Analyser sequence reads using the Topography suite v1.19 (Rodley et al., 2009). Topography uses the SOAP algorithm (Ruiqiang Li, Li, Kristiansen, & Wang, 2008) to position PE tags and single ends which contain a AseI restriction enzyme site onto the *S. pombe* (ASM294v2) reference genome, with the inclusion of the pUC19 (SYNPUC19CV) and the E. coli, pRS426 and Lambda phage ligation control sequences. No mismatches or unassigned bases (N) were allowed during positioning.

Except where indicated, all analyses were performed on 'significant' interactions that were detected between sequences that mapped to unique positions on the reference genome. Significant interactions were defined as those that occurred at levels above the significance cut-off value (see section 3.5.6). Unless explicitly stated, all bioinformatics analysis was performed on significant, uniquely mapped, non-adjacent (only interactions between restriction fragments that were not adjacent to each other in the linear sequence) interactions data using in house Perl and Python scripts. Except where indicated, statistical analyses were performed using R and Venn-diagrams were drawn with the Vennerable package (R Development Core Team, 2008).

3.5.6 SIGNIFICANCE CUT-OFF CALCULATIONS

Random ligation events can occur during the two ligation steps in the GCC protocol: 1) the ligation of the cross-linked fragments; and 2) linker addition during

preparation for sequencing. We employed two methods for the identification of significant interactions: 1) a statistical method that calculates a false detection rate (FDR) cut-off as in Rodley et al. 2009; and 2) the external ligation controls during the GCC library preparation allowed us to measure the rates of random inter-molecular ligation events. Only one inter-molecular ligation event was detected and was at a frequency below the calculated significance cut-off value. Therefore, we determined our significance cut-off using the statistical method out-lined below.

3.5.7 COLLECTOR'S CURVE

Collector's curves were generated using the total interaction data sets (including non-significant interactions) for G1, G2 and M phases of the *S. pombe* cell cycle (Cagliero et al., 2013). An interaction between a ligation control (see section 3.5.4) and the *S. pombe* genome was artificially added to each interaction data sets at a frequency of the calculated FDR cut-off. One hundred random subsets of interactions (*e.g.* 10%, 20% etc…) were independently sampled from the total interactions. For each random data set, interactions that occurred at a frequency greater than the number of times the ligation control interaction was detected ((0<=significance cut-off value) +1) were considered significant (random-significant). We then plotted the percentage (average across the 100 replicates) of the significant interaction data sets (>=significance cut-off value, non-adjacent interactions) that were identified in our random-significant populations (>=random-significance cut-off value, non-adjacent interactions) (Figure 3.3). The collector's curves indicated that, despite the high correlation between biological replicates, the interaction network was not sampled to saturation.

3.5.8 HEAT MAPS

GCC interaction networks (Gehlen et al., 2012) were displayed as heat maps using 75 Kb segments. Interaction frequencies were represented as the percentage total interactions for unique (Figure 3.1 and Figure 3.8) and repetitive (Figure 3.7) interactions to allow for cell cycle phase comparisons of interaction frequency. The unique (red) and repeat (green) overly (yellow) heat maps are interaction frequency independent to enable the clear visualisation of overlapping regions (Figure 3.7).

3.5.10 RNA EXTRACTION

For RNA extraction, cells were harvested from 12 ml of each synchronized cell culture prior to cross-linking (4,000 rpm, 2 min, RT), washed (5 ml) and suspended in (400 µl) AE buffer (50 mM Sodium Acetate, 10 mM EDTA, pH 5.3). Cell suspensions were transferred to tubes containing an equal volume of phenol/chloroform/isoamyl alcohol (24/24/1 Ambion) and acid washed glass beads (400 µl; Invitrogen). Cells were mechanically lysed in a Geno/Grinder (SPEX sample prep 2010; block pre-chilled to -20°C; 1,750 rpm, 8 x 30 s on 60 s off). Lysis was completed by freeze (-80oC, ~15 min), thaw before centrifugation (14,500 rpm, 5 min, 4°C). The aqueous phase was extracted three times with an equal volume of phenol/chloroform/isoamyl alcohol. RNA was isolated by precipitation with 1/10th volume ammonium acetate (5 M) and 2 volumes of 100% ethanol at -80°C (>30 min) before centrifugation (14,500 rpm, 10 min, 4°C). RNA was washed with 70% ethanol (350 µl; 14,500 rpm, 5 min, 4°C) and air dried (37°C, ~15 min). RNA pellets were suspended in RNASecure (80 µl; Ambion) and dissolved by heating (60°C, 10 min). RNA concentration was determined by Nano-Drop (ACTGene ASP-3700) and each sample (2 µg) was visually inspected following electrophoresis through a 1% (w/v) agarose gel. RNA was stored at -80°C before RNA sequencing (BGI China, 90 bp PE RNA sequencing analysis).

3.5.11 TRANSCRIPTOME ANALYSIS

RNA sequences (90 bp) were quality assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). To maximize the quality of the sequence reads, 10 bp was trimmed off either end of the sequences using fastx_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/index.html).

Differentially expressed genes were identified using cufflinks (Trapnell et al., 2012) to analyse the trimmed RNA sequences as a time course. This enabled the isolation of effects due to the temperature shift, thus maximizing the chances of identifying genes that are differentially regulated during each cell cycle transition. Briefly, trimmed RNA-seq reads were aligned to the *S. pombe* reference genome (ASM294v2) using Tophat (http://tophat.cbcb.umd.edu/) without providing the *S. pombe* GTF file. This allowed for novel transcript discovery. Aligned reads were assembled for differential expression analysis using cufflinks (http://cufflinks.cbcb.umd.edu/) and merged using cuffmerge (http://cufflinks.cbcb.umd.edu/manual.html#cuffmerge) with an “assemblies” file

containing the transcripts.gtf output files from cufflinks for the two biological replicates of each cell cycle phase in the order G1 – G2 – M – G1. Finally, differential expression analysis was performed using the merged.gtf output file from cuffmerge, the –T option, and the accepted_hits.bam output files from tophat in the time series order G1 – G2 – M – G1.

The raw transcript levels for genes in individual biological replicates were highly correlated (R2 > 0.91) For downstream analyses, transcription data sets were divided into: 1) genes that were in the top and bottom 5% of transcript levels in each cell cycle phase (Table 3.4 and Table 3.5); and 2) genes whose transcript levels were differentially regulated during the three cell cycle transitions (G1 – G2, G2 – M, and M – G1) (Table 3.4). Except where indicated, statistical analyses were performed in R (R Development Core Team, 2008). Venn-diagrams were drawn using the Vennerable package available for R.

3.5.12 The chromosome distribution of genes with high, low, and differential transcript levels

The chromosomal distribution of genes with high, low and differential transcript levels was determined by calculating the centre position for each gene and using this to bin (50,000 bp) the genes along each chromosome (Cagliero et al., 2013). For the genes with high and low transcript levels the number of genes per bin was graphed (Figure 3.15), while for the differentially expressed genes the average fold change in expression per bin was plotted (Figure 3.17). To test whether gene sets had a non-random chromosomal distribution, one-sample Kolmogorov-Smirnov tests were performed using the non-binned chromosome distribution data and significant results are displayed as p-values.

3.5.13 Loop lengths of interactions within chromosomes

The loop length (bp) between interacting fragments was calculations for significant uniquely mapped intra-chromosomal interactions detected at each cell cycle phase. Where two interacting fragments were located on the same chromosome, defined by coordinates Fn_{start} – Fn_{end}. The loop length (L) was defined as:

$$L=|F1_{end}-F2_{start}|$$

Where interacting fragments were ordered so that $F1_{end} < F2_{Start}$.

Loop lengths were plotted as a histogram, with the bp bin size indicated were appropriate.

3.5.14 The determination of genome connectivity and co-localization (clustering) levels

Whether LTR elements and genes with high, low, and differential transcript levels (sets of genomic regions) were connected with the genome and themselves at a level different from random was determined. The total frequency with which each set of genomic regions interacts with the genome and the frequency of interactions that occurred between the genomic regions (*i.e.* region co-localization) were calculated from the GCC interaction networks. We generated random regions sets of the same number and length (bp) as the original region by randomly selecting a start coordinate for each region and then adding the length (bp) of the region to get the end coordinate. Two populations of random sets were generated: 1) the conserved random [CR] sets conserved the number of interacting regions per chromosome. This ensured that significant results were not due to the specific linear or spatial organization of an individual chromosome(s); and 2) the random [R] sets where regions were randomly selected across the entire genome, with chromosome selection determined at a frequency that was relative to the chromosome lengths. One thousand random data sets were generated for the CR and R methods. p-values were estimated as the number of times that the total genome connectivity and co-localization (clustering) frequencies of the original data was higher or lower than the random data sets. The proportions of inter- and long vs. short (<50 Kb) distance intra-chromosomal interactions that occurred for co-localizing regions were determined.

3.5.15 Gene Ontology (GO) analysis

The AmiGO Term Enrichment online resource (http://amigo.geneontology.org/cgi-bin/amigo/term_enrichment) was used to determine if particular gene sets were enriched in GO terms within Pombase. We used a maximum p-value of 0.05 and required a minimum of two gene products for enrichment.

3.5.16 Extracting genes associated with DNA interactions for GO and transcription overlap analyses

Genes that were associated with subsets of interactions were extracted and used for gene ontology (GO) and transcriptome overlap analysis. Genes that were

internal to, overlapped with, and were the nearest up and down stream genes to the interacting restriction fragments were extracted. Gene sets associated with specific interaction sets (*e.g.* G1 or M phase specific) were used for GO enrichment analysis using AmiGO (see Appendix CD: Supplementary Spread sheet S3) and their presence in transcription data sets was determined.

3.5.17 Determining the likelihood of association with specific interactions

Whether genes with high, low, and differential transcript levels were associated with specific interactions at a frequency different from random, was determined. The proportion of genes in each gene set that was found to overlap or be in the vicinity of the interacting fragments in specific interaction sets was calculated (Figure 3.11). One thousand random sets of genes of the same number as the gene set they were being compared to were generated and the p-value was calculated by the number of times the random gene sets had a higher or lower proportion of genes associated with specific interaction sets. If a significantly high proportion of genes were associated with specific interaction sets, then the overlapping genes were used for GO Term Enrichment analysis.

Chapter 4

4 MITOCHONDRIAL-NUCLEAR DNA INTERACTIONS CONTRIBUTE TO THE REGULATION OF NUCLEAR TRANSCRIPT LEVELS AND THE REGULATION OF CELL CYCLE PROGRESSION

Parts of sections 4.1 and 4.4, and sections 4.2.1 and 4.3.1 in this chapter have been published see Appendix III:

Rodley, C.D.M.[a], **Grand, R.S.**[a], Gehlen, L.R.[a], Greyling, G[a], Jones B.M.[b], O'Sullivan, J.M.[a] (2012). Mitochondrial-nuclear DNA interactions contribute to the regulation of nuclear transcript levels as part of the inter-organelle communication system. PLoS ONE, 7(1): e30943.

[a]Institute of Natural Sciences, Massey University, Albany, New Zealand, [b]Centre for Mathematical Biology, Massey University, Albany, New Zealand.

Acknowledgements: The authors wish to thank A. Ganley for critical discussions of this work and manuscript. We also wish to thank, E. Sattlegger, P. Perlman, G. Mohr, and A. Lambowitz for yeast strains. Additionally, the authors would like to thank S. Gasser and M. Tsai (Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland) for help with sequencing.

Author's contributions: CDMR and JMO conceived and directed the project. CDMR performed analysis with LRG and GG contributing to computer analysis. RSG performed RNA expression and 3C experiments. MBJ advised and helped with statistical analysis. CDMR, JMO, LRG, and RSG wrote the manuscript.

Since the submission of this thesis the remaining sections in this chapter have been published in the journal Mitochondrion:

Grand, R.S.[1,2], Martienssen, R.[3], O'Sullivan, J.M.[1] (2014). Functional roles for mitochondrial-nuclear DNA interactions in the *Schizosaccharomyces pombe* cell cycle. Mitochondrion, July; 17:141-149.

[1]Liggins institute, University of Auckland, Grafton Auckland 1032, NZ, [2]Institute of Natural and Mathematical Sciences, Massey University, Albany Auckland 0745, NZ, [3]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

Acknowledgements: The authors would like to thank Lynsey Cree and Julia Horsfield for comments on this manuscript. This work was supported by The Marsden fund (JMOS) and a Massey University Doctoral Scholarship (RSG).

Author's contributions: RSG, RM and JMO conceived and directed the project. RSG performed experiments and analysis. RSG and JMO wrote the manuscript.

4.1 INTRODUCTION

Mitochondria have a central role within the metabolic systems of cells. In yeasts (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe)*, as in other organisms, the mitochondrial organelle contains a genome that encodes an essential subset of the electron transport chain components (Foury et al., 1998; Wood et al., 2002) that are necessary for respiratory growth (Lagunas, 1976; Schäfer, 2003).

Over the course of its evolution into an organelle, most of the ancestral genes present within the mitochondrial genome were transferred and integrated into the host cell genome (Timmis et al., 2004). These transfer events are not all historical; rather, there is evidence for ongoing transfer of mitochondrial DNA to the nucleus in various organisms including the Budding yeast, *Saccharomyces cerevisiae* (Ricchetti et al., 1999, 2004; Rodley et al., 2012, 2009; Thorsness & Fox, 1990, 1993). However, the functional roles of the mitochondrial regions, once in the nuclear compartment, have not been fully elucidated.

Distal regulatory regions (*e.g.* enhancers) are known to loop within chromosomes in order to interact with the promoter region of the genes that they control (Tolhuis et al., 2002). Furthermore, enhancers can also interact with promoters on different chromosomes to control gene expression (Spilianakis & Flavell, 2004; Spilianakis et al., 2005; A. Williams et al., 2010). These types of inter- and intra-chromosomal interactions can be captured using proximity-based ligation methodologies (*e.g.* Chromosome Conformation Capture (3C) (Dekker et al., 2002)) that incorporate high resolution (*i.e.* ~2 Å (Fujita & Wade, 2004)) cross-linking of interacting DNA strands, restriction digestion, dilution, and ligation to identify DNA sequences that interact within a cell.

Using a proximity-based ligation base method developed in the O'Sullivan lab to observe the global set of genome-wide interactions (Genome Conformation Capture (GCC)), nucleic acids of mitochondrial origin had recently been found to interact with nuclear loci (hereinafter referred to as mt-nDNA interactions) in *S. cerevisiae* (Rodley et al., 2009). Surprisingly these inter-organelle, mt-nDNA interactions were frequent and statistically significant suggesting that they perform a hitherto unrecognized role within yeast cells (Rodley et al., 2009). Furthermore, analysis of one of these

interactions demonstrated carbon source dependence (Rodley et al., 2009). Intriguingly, the quality and quantity of mitochondrial DNA has been shown to affect patterns of nuclear transcription (V. Parikh et al., 1987; V. S. Parikh et al., 1989) and replication (Blank et al., 2008) in yeast.

The mt-nDNA interactions involve specific regions of the mitochondrial genome, and together with work presented here (see sections: 4.2.1 and 4.3.1) have been linked to the regulation of transcript levels of nuclear encoded mitochondrial genes (see sections: 4.2.1 and 4.3.1, (Rodley et al., 2012)). Interestingly, just increasing the amount of mitochondrial DNA present in *S. cerevisiae*, by overexpression of the mitochondrial maintenance protein Abf2p, promotes nuclear DNA replication through reducing the Sir2 mediated deacetylation of specific autonomously replicating sequence (ARS) sites (Blank et al., 2008). These observations suggest a function for mtDNA fragments in the regulation of nuclear genes and DNA replication.

Mitochondrial DNA sequences are also often found inserted in eukaryotic nuclear chromosomes (NUclear MiTochondrial sequences (*NUMTs*)) (Lenglez et al., 2010; Ricchetti et al., 2004). NUMTs can form through the use of mitochondrial DNA sequences to repair double-strand DNA breaks in *S. cerevisiae* (Ricchetti et al., 1999) and *S. pombe* (Lenglez et al., 2010), therefore, mitochondrial DNA must be present within the nuclear environment. The mitochondrial sequences that constitute these NUMTs are proposed to have nuclear functions. For example, in the Budding yeast, NUMTs are rich in ARS consensus motifs that promote nuclear DNA replication (Blank et al., 2008; Chatre & Ricchetti, 2011). However, to date there has not been any link between mitochondrial DNA and replication control in *S. pombe*.

It is accepted that successful progression through the cell cycle requires coordination between the nuclear and mitochondrial genomes (Chatre & Ricchetti, 2013; Chu et al., 2007; Crider et al., 2012; Lebedeva & Gerald S. Shadel, 2007; Mandal et al., 2005; Mitra et al., 2009). For instance, in *S. cerevisiae,* the G1 to S phase cell cycle checkpoint is regulated by mitochondrial DNA (Mandal et al., 2005; Mitra et al., 2009). Specifically, in the absence of mtDNA, the Rad53 DNA damage response checkpoint is activated and the G1 to S phase cell cycle transition is

inhibited (Crider et al., 2012). However, the mechanism by which this regulation occurs remains to be determined.

First we explored the hypothesis that inter-organelle interactions respond to the metabolic status of the cell to regulate nuclear transcript levels. It had been demonstrated that Interactions between mitochondrial genes (*i.e. COX1* and *Q0182,* a dubious mitochondrial ORF) and nuclear encoded loci (*i.e. MSY1* and, *RSM7*, respectively), are dependent upon a functional electron transport chain and mitochondrial encoded reverse transcriptase machinery (Rodley et al., 2012). I investigated whether the levels of the nuclear encoded *MSY1* and *RSM7* gene transcripts increase when the interaction frequency of the respective interactions is reduced by the knockout of mitochondrial reverse transcriptase activity. On the basis of these results we propose that reverse-transcription mediated inter-organelle DNA interactions are a novel form of communication between mitochondria and the nucleus.

Second, we set out to characterize the relationship between mt-nDNA interactions and cellular function throughout the *S. pombe* cell cycle. *S. pombe* is a paradigm for cell cycle research sharing many features with higher eukaryotes (Chiron et al., 2007; Coudreuse & Nurse, 2010; Fantes & Nurse, 1978; Nurse et al., 1976), including a dependence upon respiration for survival (Schäfer, 2003; Weir & Yaffe, 2004). *S. pombe* cells have a small nuclear genome and can be synchronised, making it an excellent choice for studying mt-nDNA interactions. Here we characterize the relationship between mt-nDNA interactions and cellular function over the course of the cell cycle in *S. pombe*. We identify specific mt-nDNA interactions with high efficiency, early replicating origins of replication in the G1 phase of the cell cycle. Furthermore, we detect mt-nDNA interactions with genes required for nucleotide synthesis and ribosome biogenesis during metaphase. Our results suggest that the formation of specific mt-nDNA interactions contribute to the regulation of cell cycle progression in *S. pombe* by promoting nuclear DNA replication and protein synthesis following exit from metaphase.

4.2 RESULTS

4.2.1 MT-NDNA INTERACTIONS CONTRIBUTE TO THE REGULATION OF NUCLEAR TRANSCRIPT LEVELS AS PART OF THE INTER-ORGANELLE COMMUNICATION SYSTEM

The number of significant mt-nDNA interactions increased >10-fold in respiring (*i.e.* glycerol lactate grown) cells, relative to glucose or galactose grown cells (Rodley et al., 2012). This increase was not due to a higher number of sequence reads for the respiring sample. Thus, a greater number of unique nuclear loci connect to mtDNA during respiratory growth when the mitochondria are most active. This result, coupled with the need for a functional electron transport chain and reverse transcriptase machinery, led us to hypothesize that the mt-nDNA interactions are functional in nature, and specifically that they are capable of controlling the transcript levels of the nuclear loci with which they interact.

To test this we performed quantitative reverse transcriptase PCR (qRT-PCR) to determine the transcript levels of the nuclear encoded *MSY1* and *RSM7* genes in wild-type (wt) cells, the mitochondrial group-II intron knockout mutant (161-U7 GII-0), and strain 161-U7 GII-0 aI5γ (Figure 4.1A). We found that the population transcript level of the *MSY1* gene is significantly higher (t-test, two-sample unequal variance, one-tail, n=2, p=0.0007) in strain 161-U7 GII-0 (Figure 4.2A), which does not contain the probe site and, therefore, has no detectable *COX1-MSY1* interaction (Figure 4.1A and B), thus identifying the maximum transcript level in the absence of detectable inter-organelle interactions. Critically, we observed a similar population level increase in *MSY1* transcript levels following the removal of the type II introns, except aI5γ (*i.e.* strain 161-U7 GII-0 aI5γ; Figure 4.2A). A similar increase was observed for *RSM7* transcripts in both the 161-U7 GII-0 and 161-U7 GII-0 aI5γ strains relative to the wt (Figure 4.2B), consistent with the effects of intron deletion on the *Q0182-RMS7* interaction level (Figure 4.1C). By contrast deletion of *MRS1*, which is involved in mitochondrial group I intron splicing [36,37], had no effect on either *MSY1* or *RSM7* transcript levels (Figure 4.2C), or the *COX1-MSY1* interaction frequency (Figure 4.3). Thus, strains lacking mitochondrial reverse transcriptase activity have lower frequencies of mt-nDNA interactions and increased levels of nuclear encoded transcripts. These results suggest that cDNA mediated mt-nDNA interactions are

involved in the regulation of the nuclear transcripts, and therefore that the mt-nDNA interactions we observed are biologically relevant.

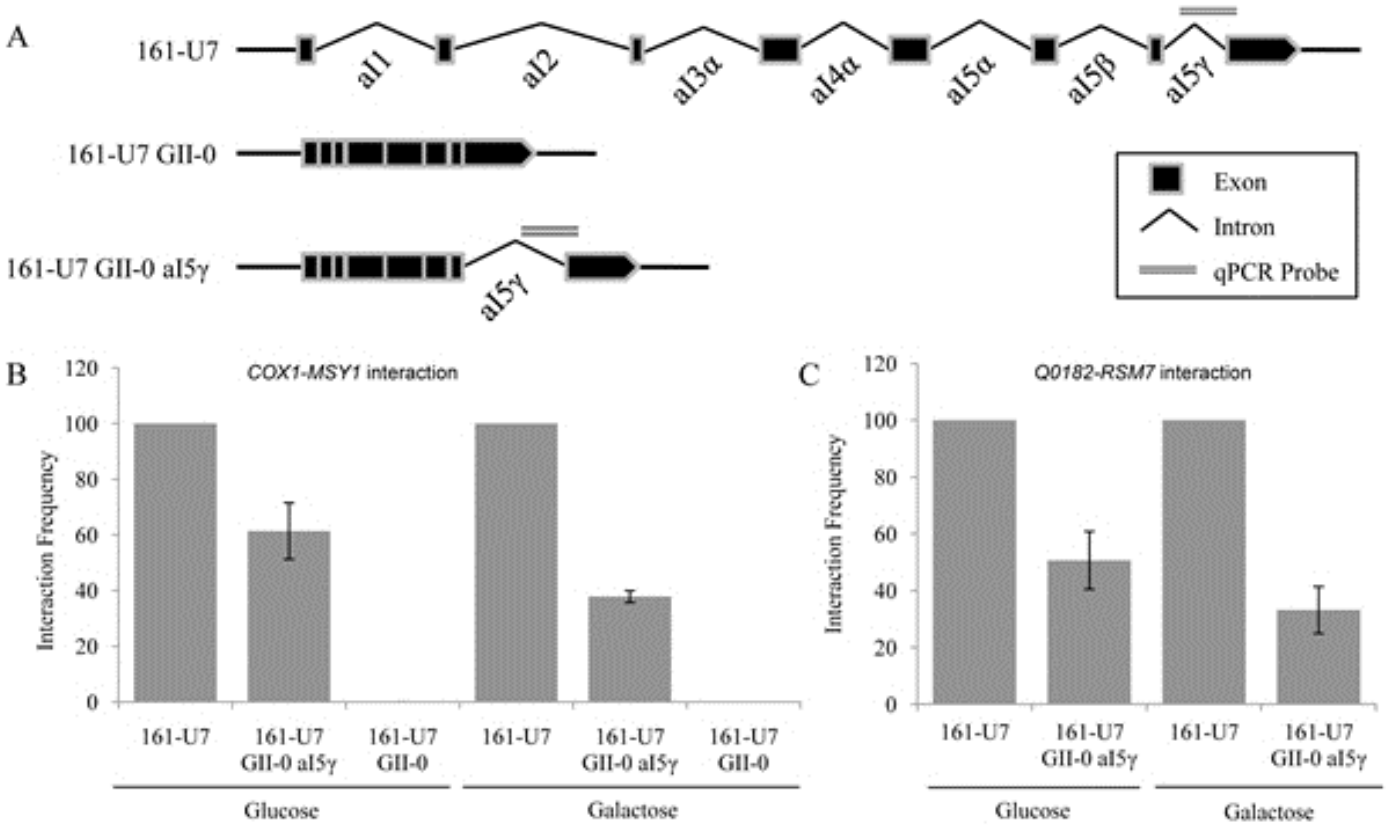


Figure 4.1. Mt-nDNA interactions require active mitochondrial reverse transcriptase machinery.

A) Illustration of *COX1* gene arrangement in the WT (161-U7), intron a15γ (161-U7 GII-0 a15γ), and no mitochondrial group II introns (161-U7 GII-0) strains. Group II introns within the *COX1* gene encode functional reverse transcriptase. The region of *COX1* that participates in the *COX1-MSY1* interaction is indicated (qPCR probe). Strain 161-U7 GII-0 was included as a control to rule out a nuclear sequence, originating from a mitochondrial integration within the nuclear genome (NUMT), being responsible for the observed interaction. B) *COX1-MSY1* interaction frequencies for wt and intron mutants, illustrated in A), grown in glucose or galactose. C) *Q0182-RSM7* interaction frequencies for mitochondrial reverse transcriptase mutant 161-U7 GII-0 a15γ, illustrated in A), grown in glucose or galactose. Interaction frequencies are expressed as percentages of the wild-type *S. cerevisiae* strain 161-U7 for each carbon source (set at 100%) +/- standard error of the mean (n=3). Interaction values in B) and C) were corrected for nuclear genome copy number to facilitate direct comparison.

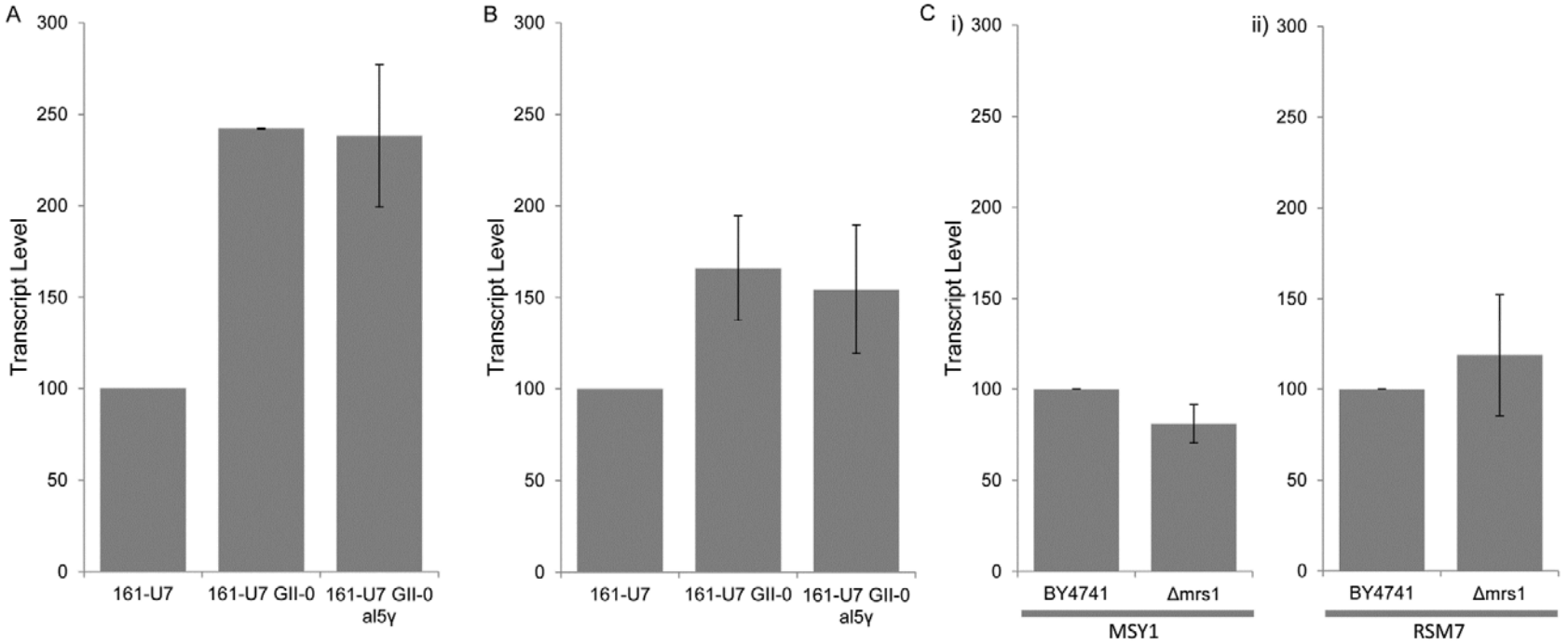


Figure 4.2. Knocking out mitochondrial encoded reverse transcriptase activity results in increased transcript levels of nuclear genes that are involved in mt-nDNA interactions.

A) Nuclear encoded *MSY1* transcript levels were determined by qRT-PCR in WT (strain 161-U7), 161-U7 GII-0 (lacks both the mitochondrial group II introns and the *COX1* interacting region; Figure 4A), and 161-U7 GII-0 a15γ (contains the interacting region and lacks the group II introns; Figure 4A) cells. B) Nuclear encoded *RSM7* transcript levels were determined by qRT-PCR in: WT (strain 161-U7); 161-U7 GII-0; and 161-U7 GII-0 a15γ cells. Neither 161-U7 GII-0 nor 161-U7 GII-0 a15γ has any alteration within the *Q0182* open reading frame. C) Deletion of *MRS1* (BY4741 Δ*mrs1*), a nuclear gene involved in splicing mitochondrial type-I introns, has no effect on i) *MSY1* or ii) *RSM7* transcript levels. All transcript levels were standardized to nuclear *ACT1* and expressed as percentage of wild-type (set at 100%) +/- standard error of the mean (n=2).

The finding that inter-organelle interactions affect nuclear transcript levels necessarily predicts that the deletion of *yme1*, which reduces the frequency of the *COX1-MSY1* interaction (Rodley et al., 2012), should correlate with an increase in nuclear *MSY1* transcript levels. Indeed increases in *MSY1* transcript levels, and other genes involved in mitochondrial gene expression and the biogenesis of the respiratory chain, have been identified within yeast cells containing the *yme1* deletion growing with a mixed respiro-fermentative metabolism (Arnold, Wagner-Ecker, Ansorge, & Langer, 2006).

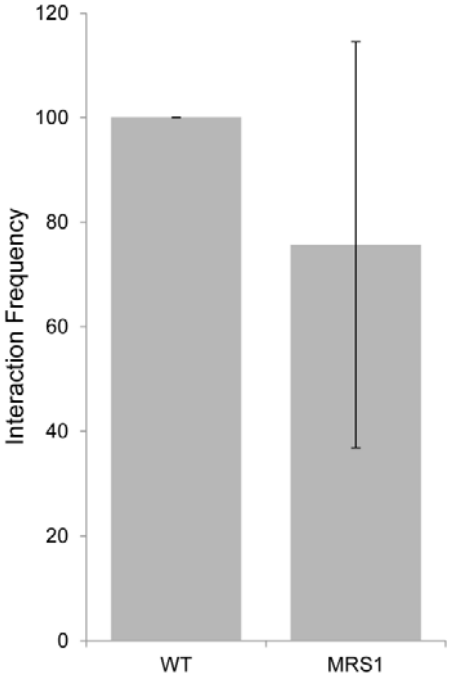


Figure 4.3. Deletion of *MRS1* (BY4741 *Δmrs1*), a nuclear gene involved in splicing mitochondrial type-I introns, has no significant effect on the frequency of the *COX1-MSY1* interaction in glucose grown yeast cells.

Interaction frequency was expressed as percentages of the wild type *S. cerevisiae* strain BY4741 (WT, set at 100%) +/- standard error of the mean (n=3).

4.2.2 MT-NDNA INTERACTIONS VARY THROUGHOUT THE *S. POMBE* CELL CYCLE

Mt-nDNA interactions are dynamic in the Budding yeast (*Saccharomyces cerevisiae*) and change if the yeast is respiring or fermenting (Rodley et al., 2012, 2009). We used GCC to identify connections within the mitochondrial genome(s) and between the mitochondrial genome and the nuclear chromosomes (mt-nDNA interactions) in the G1, G2, and M phases of the Fission yeast cell cycle. Significant

interactions within the individual biological replicates were highly correlated and replicates were combined for further analysis (Figure 4.4A-C).

The number and strength of interactions, both within the mitochondrial genome and between the mitochondrial genome and nuclear chromosomes, varied in the different phases of the cell cycle (Figure 4.4D-F, Figure 4.5, Figure 4.6, and Table 4.1). In fact, the majority of mt-nDNA interactions were specific to each phase of the cell cycle (Figure 4.6). Despite this, the linear distribution of the nuclear interaction partners along the *S. pombe* nuclear chromosomes was not different from random (K-S test), although there were obvious differences in the distributions between cell cycle phases (Figure 4.5A).

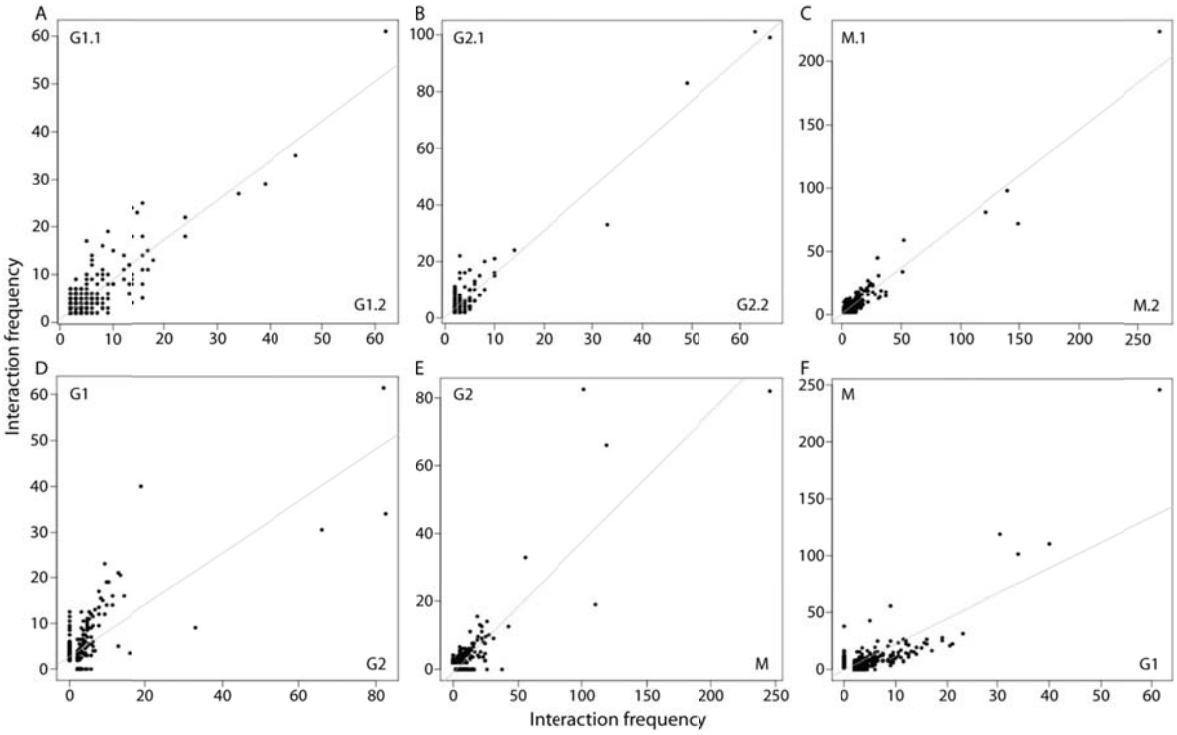


Figure 4.4. Mitochondrial interaction networks for the biological replicates were highly correlated at the AseI restriction fragment level.

Mitochondrial interaction networks for the biological replicates were highly correlated at the AseI restriction fragment level. A) G1 phase biological replicates (G1.1, replicate one and G1.2, replicate two, R2=0.7895), B) G2 phase biological replicates (G2.1, replicate one and G2.2, replicate two, R2=0.9239), B) M phase biological replicates (M.1, replicate one and M.2, replicate two, R2=0.9128). Mt-nDNA interactions that were detected at different phases of the cell cycle were less well correlated than the biological replicates. D) G1 with G2 phase (G1, G1 phase and G2, G2 phase, R^2=0.5853), G2 with M phase (G2, G2 phase and M, M phase, R^2=0.7459), and M with G1 phase (M, M phase and G1, G1 phase, R^2=0.6321). Correlation plots were generated in R using only significant non-adjacent interactions in which the nuclear fragment mapped to a single locus (*i.e.* was uniquely mapping).

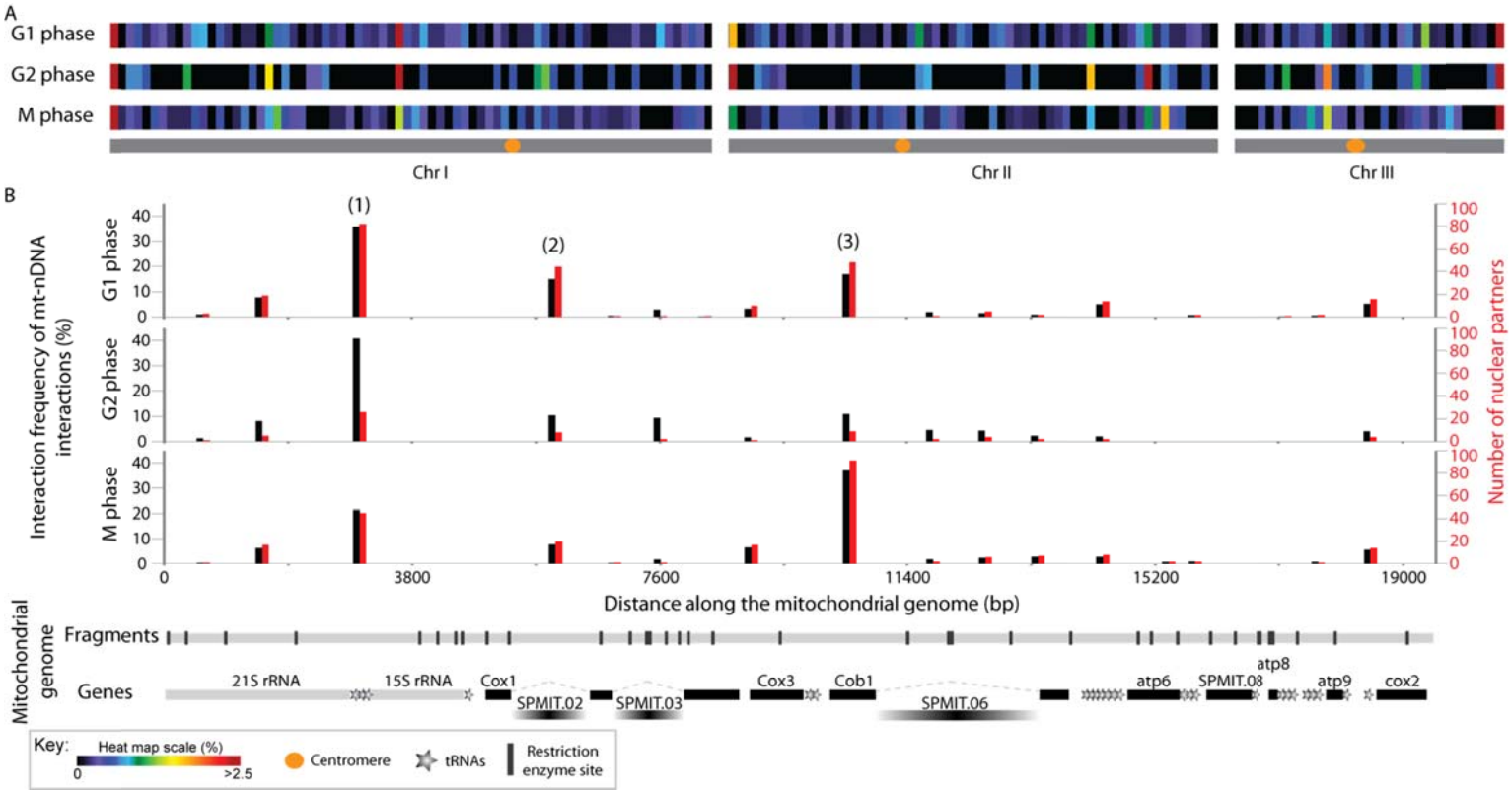


Figure 4.5. Mt-nDNA interactions are cell cycle dependent.

There were differences in the identity, number and contact frequencies of mt-nDNA interactions detected at the G1, G2 and M phases of the Fission yeast cell cycle. Mt-nDNA interactions (% total mt-nDNA interactions) are plotted as heat maps for 75 Kb nuclear fragments (A). The linear chromosomal distribution of the nuclear interacting partners was not different from random (K-S test). (B) Three mitochondrial DNA fragments participate in the majority of the mt-nDNA interactions: 1) 2,021-3,809, 2) 5,243-6,621, and 3) 9,433-11,427; at each phase of the cell cycle (G1 (67%), G2 (62%) and M phase (66%)). However, there is substantial variation in the distribution of the contact frequency and number of nuclear partners among these three fragments. The proportion of the total mt-nDNA contact frequency that each restriction fragment is involved in (black) and number of nuclear partners (red) of the individual mitochondrial fragments was plotted for the three cell cycle phases (G1, G2, and M). The mitochondrial genes and AseI restriction enzyme sites are illustrated.

The mitochondrial DNA fragments that directly participate in interactions with nuclear chromosomes were distributed across the mitochondrial genome (Figure 4.5B). However, three mitochondrial fragments account for the majority of the interactions that occur in the G1, G2 and M phases of the *S. pombe* cell cycle: 1) 21S-15SrRNA (2,021-3,809 bp); 2) Cox1-SPMIT.02 (5,243-6,621 bp); and 3) Cox3-Cob1-SPMIT.06 (9,433-11,427 bp) (Figure 4.5B). Critically, there is cell cycle phase variation in both the frequency with which each of these fragments is involved in mt-nDNA interactions and the number of nuclear loci that they contact. For example, the 21S-15SrRNA fragment accounts for: 36% of the total interaction frequency and contacts 82 nuclear loci in G1 phase; 41% and 26 loci in G2 phase; and 21% and 45 loci in M phase. Thus the frequency of interactions and number of nuclear loci that interact with a particular mtDNA fragment changes throughout the cell cycle.

The number and contact frequency of interactions between all restriction fragments within the mitochondrial genome also varied between the cell cycle phases (Figure 4.6C and Figure 4.7). Of note, there are a large number of within mitochondrial genome interactions that form specifically during metaphase and are shared with the G1 phase of the cell cycle (Figure 4.6C). However, there was no obvious bias in the lengths of the DNA loops that formed between the interacting fragments within the mitochondrial genome (Figure 4.7B-D). Collectively, these results suggest that there is not one particular mitochondrial region that preferentially directs mitochondrial genome structure. Moreover, the connectivity between and within the mitochondrial genome(s) increased during metaphase, in a manner reminiscent of the increased condensation seen for the nuclear chromosomes.

Table 4.1. The number of interactions within the mitochondrial genome and between the mitochondrial and nuclear genomes varies at different phases of the *S. pombe* cell cycle.

The largest numbers of interactions detected within the mitochondrial genome occur in the M phase of the cell cycle. A large number of interactions occur between the mitochondrial and nuclear chromosomes, with the greatest number being detected in the G1 cell cycle phase.

Phase of the cell cycle	Total number of interactions	Interactions within (or between) mitochondria	Interactions between the mitochondria and nuclear genomes
G1 phase	412	167 (40.53%)	245 (59.47%)
G2 phase	153	94 (61.44%)	59 (38.56%)
M phase	559	333 (59.57%)	226 (40.43%)

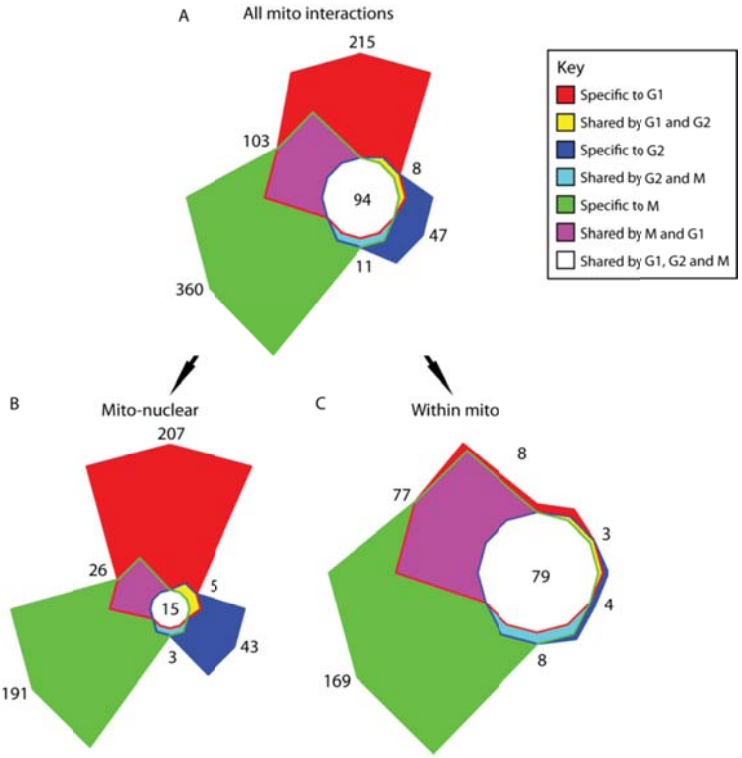


Figure 4.6. Mito-nuclear DNA interactions are largely specific to each cell cycle phase. By contrast, a large number of within mitochondrial interactions were unique to the M phase of the cell cycle.

Three way proportional Venn-diagrams of mt-mtDNA and mt-nDNA interactions highlighting interactions that were specific to or shared by the G1, G2, and M phases of the cell cycle. The numbers of interactions involving uniquely mapping nuclear loci are shown. B) The majority of Mito-nuclear DNA interactions were specific to each cell cycle phase. C) Interactions within the mitochondrial genome (*i.e.* within a single copy or between multiple copies of the mitochondrial chromosome) were enriched in the M phase of the cell cycle or shared between the M and G1 phases. Uniquely mapped data files containing significant non-adjacent interactions were used to generate these plots in R using the Vennerable package.

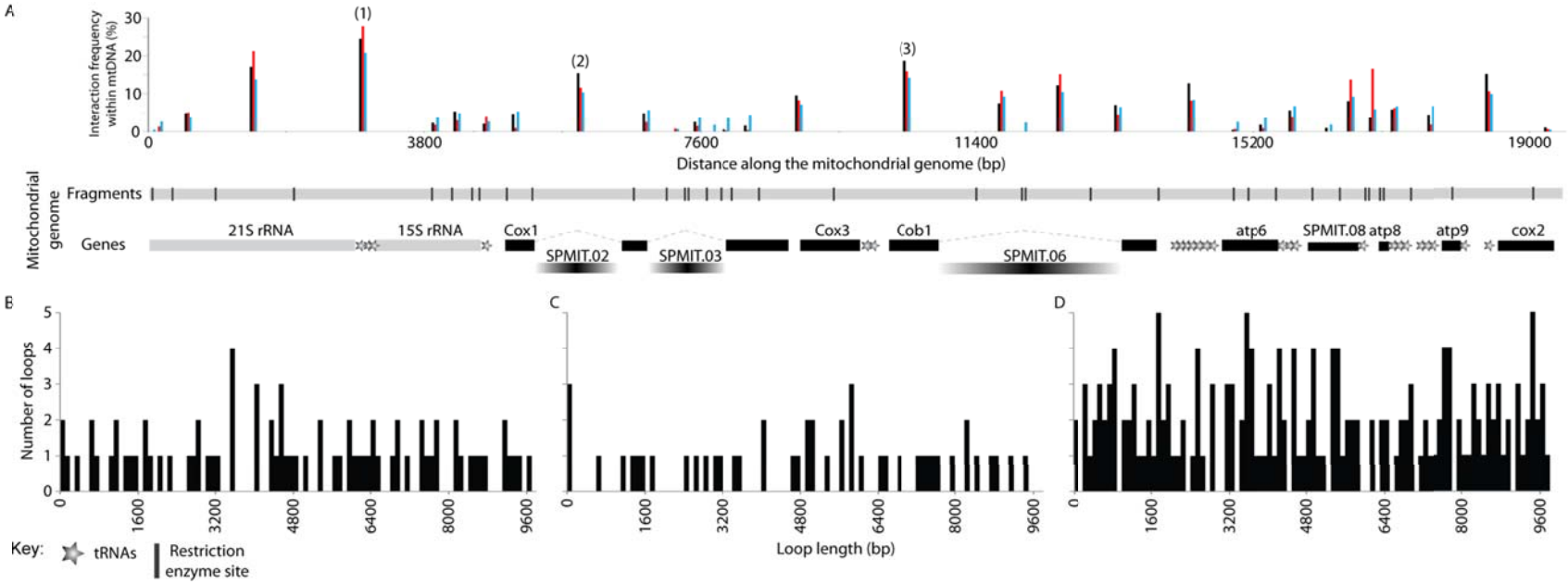


Figure 4.7. Mitochondrial fragments that interact with nuclear chromosomes also interact with other mitochondrial fragments and M phase mitochondrial genomes are the most highly connected.

A) Intra-mitochondrial genome interactions can be detected for nearly all mitochondrial fragments. Fragments 1-3, which are responsible for the majority of mt-nDNA interactions also make a large contribution to intra-mitochondrial interactions. Interactions were plotted as the proportion of the total within mitochondria interactions that each mitochondrial restriction fragment was involved in. Interaction counts are plotted as G1 phase, black; G2 phase, red; and M phase, blue. The lengths of intra-mitochondrial loops between interacting fragments were relatively evenly distributed across the mitochondrial genome: B) G1 phase, C) G2 phase, and D) M phase, suggesting that there is not one particular fragment (or group of fragments) that preferentially interact with each other. The noticeable increase in the number of loops in the M phase is a result of the large number of specific intra-mitochondrial interactions that were detectable during metaphase and may reflect highly condensed mitochondrial nucleoids. The size of the loop between the interacting fragments that defined the intra-mitochondrial interaction was calculated and the number of loops per bin (100bp bins) was plotted for all loop sizes. Significant non-adjacent interactions that were uniquely mapped onto the mitochondrial genome were used in these analyses.

4.2.3 MT-NDNA INTERACTIONS ARE NOT ENRICHED FOR NUCLEAR ENCODED MITOCHONDRIAL GENES.

The majority of genes required for mitochondrial function are encoded in the nuclear chromosomes (Timmis et al., 2004). Thus the transcription of nuclear genes required for mitochondrial function must be precisely coordinated in response to mitochondrial demand (Butow & Avadhani, 2004; Z. Liu & Butow, 2006). We determined whether mt-nDNA interactions preferentially formed with nuclear encoded mitochondrial genes. A subset of the mt-nDNA interactions involved nuclear encoded mitochondrial genes in all data sets (Figure 4.8A). However, in general, the genes involved in mt-nDNA interactions were depleted for nuclear encoded mitochondrial genes when compared to random sets (Figure 4.8A).

Previously, specific mt-nDNA interactions have been shown to influence the transcript levels of nuclear encoded mitochondrial genes (Rodley et al., 2012). We examined whether the formation of mt-nDNA interactions with nuclear encoded mitochondrial genes correlated with their transcript levels (Table 3.4 and Table 3.5). Nuclear encoded mitochondrial genes were not often found in the high or low transcription data sets (Figure 4.8B and C). However of note, mt-nDNA interactions that formed specifically during the G2 phase were highly associated with nuclear encoded mitochondrial genes that were subsequently down regulated during the G2-M phase cell cycle transition (Figure 4.8C(ii)).

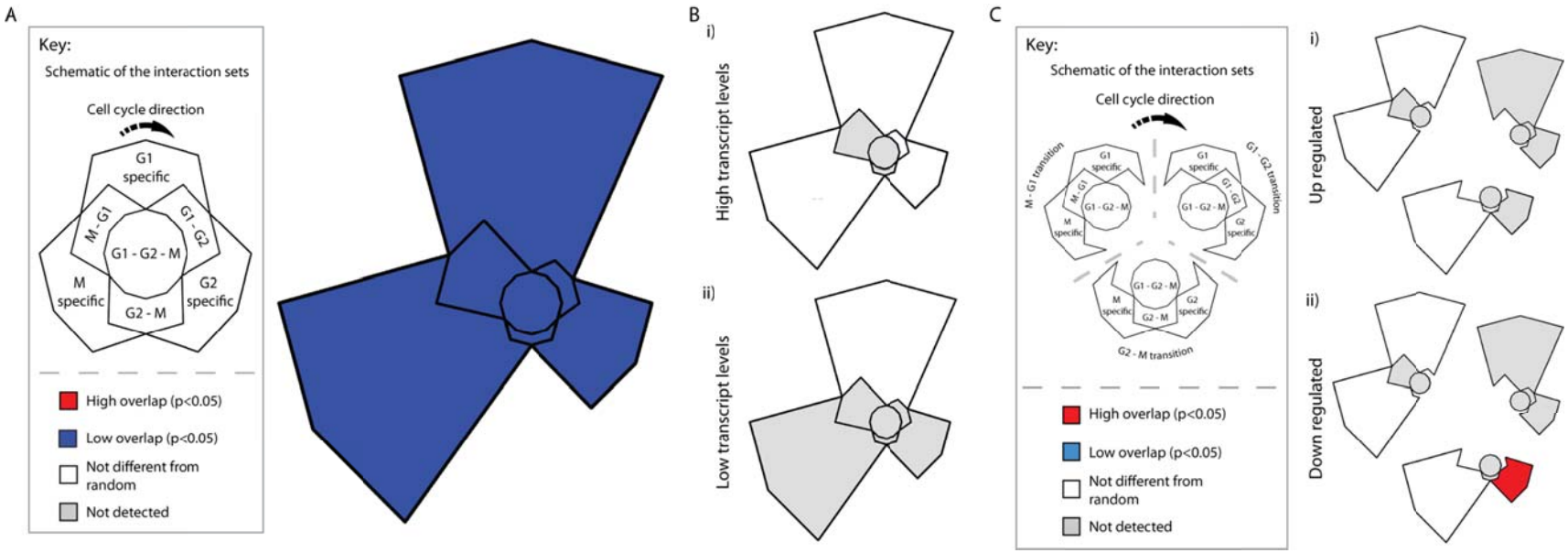


Figure 4.8. Mt-nDNA interactions do not preferentially involve nuclear encoded mitochondrial genes.

Mt-nDNA interactions formed with nuclear encoded mitochondrial genes but at a lower rate than random (A). The proportion of mt-nDNA interactions that overlapped nuclear encoded mitochondrial genes was compared to random chromosomal loci. B) The mitochondrial genes that were associated with mt-nDNA interactions were not more or less likely to have high B(i) or low B(ii) transcript levels compared to random gene sets. C) Mt-nDNA interactions specific to G2 phase were highly associated with nuclear encoded mitochondrial genes that were down regulated during the G2-M phase cell cycle transition.

Table 4.2. The numbers of *S. pombe* genes within the highest, lowest, and differential transcript levels at each stage of the cell cycle.

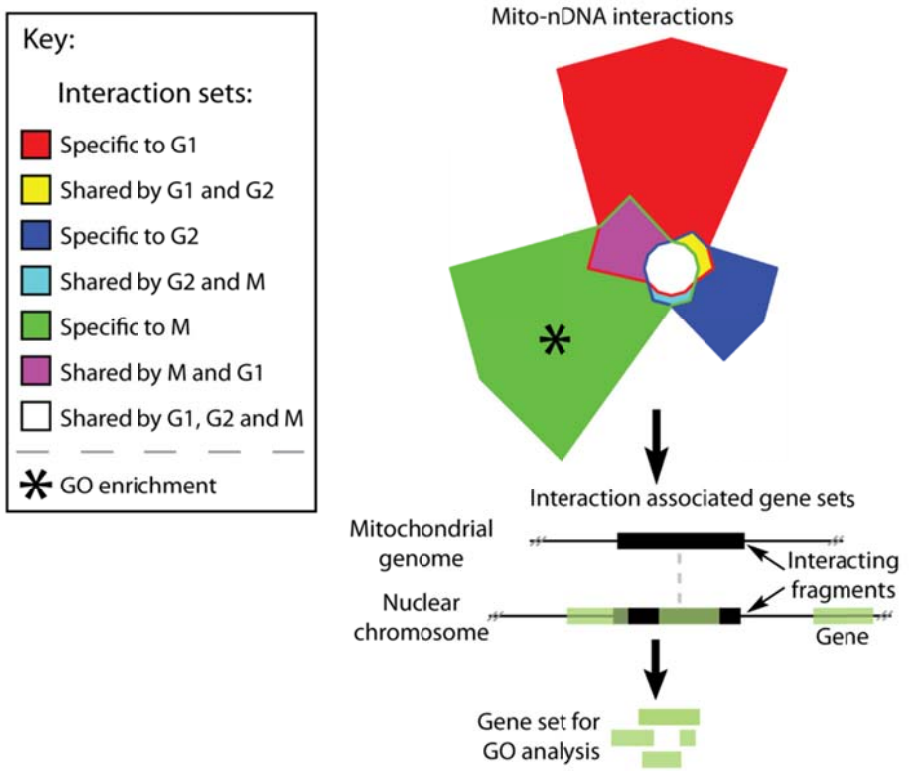
Top) Numbers of genes that had the highest (top 5%) and lowest (bottom 5%; excluding genes that were not expressed) raw transcript levels detected at the G1, G2 and M phases of the cell cycle. Bottom) Numbers of genes whose transcript levels changed significantly (>+/-1.5-fold) during the transition from: G1-G2, G2-M, and M-G1 cell cycle phases.

	G1 phase	G2 phase	M phase
The number of genes with the highest (top 5%) or lowest (bottom 5%) transcript levels	195	197	196
	G1 - G2 phase	G2 - M phase	M - G1 phase
The number of significantly differentially expressed genes	198	346	239
Then number of up regulated genes	102 (51.51%)	138 (39.88%)	150 (62.76%)
The number of down regulated genes	96 (48.49%)	208 (60.12%)	89 (37.24%)

4.2.4 MT-NDNA INTERACTIONS SPECIFICALLY FORMED DURING METAPHASE OCCUR WITH GENES REQUIRED FOR CELL GROWTH AND DNA SYNTHESIS

Mitochondria have a central role in cellular metabolism. Therefore, we hypothesized that this would be reflected in connections to genes that affect metabolic processes central to cell viability. We determined if mt-nDNA interactions that were specific to or shared by the different cell cycle phases involved specific nuclear gene sets. Only mt-nDNA interactions that specifically formed during metaphase were enriched for a gene set (Figure 4.9A). Interestingly, the cellular functions of these genes were associated with molecular binding activity and related to energy availability and growth (Figure 4.9B).

A



B

GO term		Aspect	p-value
GO:0043168	anion binding	F	6.67E-05
GO:0030554	adenyl nucleotide binding	F	1.21E-02
GO:0005524	ATP binding	F	1.53E-02
GO:0032559	adenyl ribonucleotide binding	F	1.63E-02
GO:0017076	purine nucleotide binding	F	1.89E-02
GO:0035639	purine ribonucleoside triphosphate binding	F	2.39E-02
GO:0001883	purine nucleoside binding	F	2.39E-02
GO:0032550	purine ribonucleoside binding	F	2.39E-02
GO:0032555	purine ribonucleotide binding	F	2.52E-02
GO:0032553	ribonucleotide binding	F	2.90E-02
GO:0032549	ribonucleoside binding	F	3.12E-02
GO:0001882	nucleoside binding	F	3.29E-02
GO:0097367	carbohydrate derivative binding	F	4.17E-02

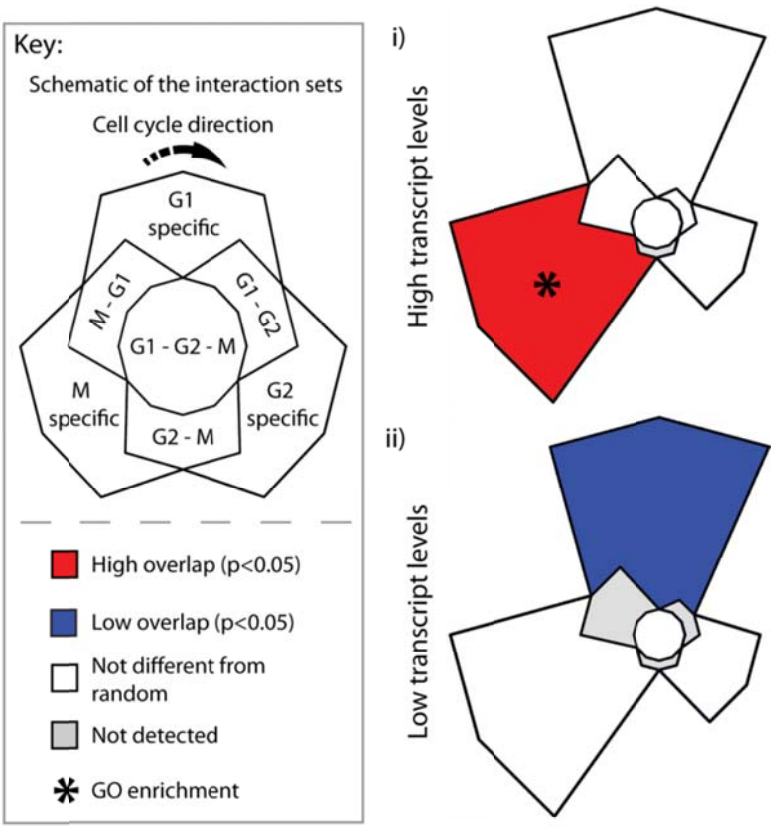
Figure 4.9. Nuclear genes associated with mt-nDNA interactions formed specifically during metaphase are enriched in biological functions related to genomic DNA synthesis and energy regulation.

Genes that were associated with nuclear DNA restriction fragments within specific mt-nDNA interaction sets were extracted and used for gene ontology term enrichment analysis (A). Ontological terms related to genomic DNA synthesis and the regulation of energy availability were enriched within the M phase interactions (B).

Mt-nDNA interactions that specifically formed during metaphase were associated with nuclear genes that had high transcript levels in that phase of the cell cycle (Figure 4.10A(i)). Furthermore, the nuclear gene set involved in the metaphase specific interactions was enriched for genes involved in ribosome structure and function (Figure 4.10B).

By contrast, mt-nDNA interactions were not enriched with nuclear genes that had low transcript levels (Figure 4.10A(ii)) or that were differentially regulated during the G1-G2 or M-G1 phase cell cycle transitions (Figure 4.11). However, a high proportion of mt-nDNA interactions that were present in both the G2 and M phases of the cell cycle, and those formed specifically during M phase, associated with genes whose transcript levels increased significantly (+>1.5-fold) during the G2-M phase cell cycle transition (Figure 4.11A). Despite this, the up regulated genes associated with these mt-nDNA interactions were not enriched for any specific functions.

A



B

GO term		Aspect	p-value
GO:0005840	ribosome	C	1.82E-04
GO:0022626	cytosolic ribosome	C	4.27E-04
GO:0022625	cytosolic large ribosomal subunit	C	8.46E-04
GO:0044391	ribosomal subunit	C	1.18E-03
GO:0015934	large ribosomal subunit	C	1.27E-03
GO:0044445	cytosolic part	C	1.51E-03
GO:0030529	ribonucleoprotein complex	C	2.48E-02
GO:0002181	cytoplasmic translation	P	5.13E-03
GO:0006412	translation	P	6.31E-03
GO:0003735	structural constituent of ribosome	F	4.48E-03
GO:0005198	structural molecule activity	F	9.63E-03

Figure 4.10. Mt-nDNA interactions that formed specifically during metaphase were associated with highly transcribed nuclear genes required for ribosome function.

A high proportion of the nuclear genes that were associated with M phase specific mt-nDNA interactions had high transcript levels compared to random gene sets (A(i)). By contrast, mt-nDNA interactions specific to and shared by the other cell cycle phases were not enriched for genes that had high transcript levels. (A(ii)) Nuclear genes associated with mt-nDNA interactions specifically formed during the G1 phase were less likely than random to involve genes with low transcript levels. Moreover, interactions that were shared by consecutive cell cycle phases did not involve nuclear genes whose transcript levels were within the lowest 5% in each condition. (B) The high transcript level genes that were associated with the M phase specific mt-nDNA interactions were enriched in GO terms related to ribosome structure and function.

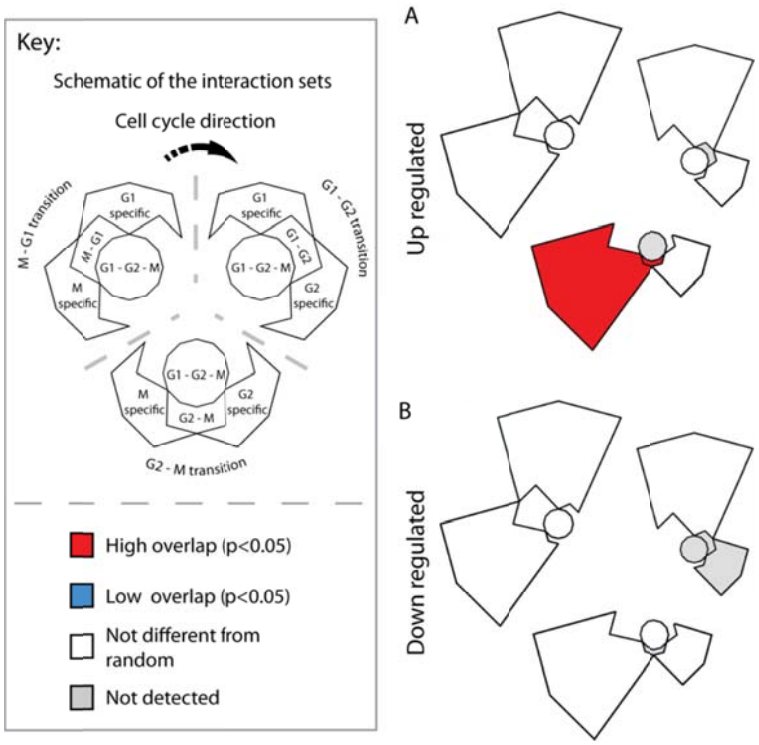


Figure 4.11. Mt-nDNA interactions do not globally correlate with the differential regulation of nuclear genes.

Mt-nDNA interactions that were shared between G2 and M phases, and specific to M phase, significantly overlapped genes whose transcript levels were up regulated during the G2-M phase shift (A). Otherwise, mt-nDNA interactions were not more likely to be associated with nuclear genes whose transcript levels were significantly (A) up or B) down regulated) during the G1-G2, G2-M, or M-G1 cell cycle phase transitions.

4.2.5 G1 PHASE MT-NDNA INTERACTIONS ARE ASSOCIATED WITH HIGH EFFICIENCY, EARLY REPLICATING ORIGINS OF DNA REPLICATION

NUMTs and the cell's mitochondrial DNA concentration have been implicated in promoting nuclear DNA replication in *S. cerevisiae* (Blank et al., 2008; Chatre & Ricchetti, 2011). We investigated if cell cycle specific mt-nDNA interactions preferentially formed with *S. pombe* ARS sites (Heichinger, Penkett, Bahler, & Nurse, 2006). A significantly high proportion of interactions detected during the G1 phase of the cell cycle occurred with ARS sites (Figure 4.12A). By contrast, mt-nDNA interactions that formed in the G2 and M phases of the cell cycle were not enriched for ARS sequences (Figure 4.12A). The nuclear ARS sites that were involved in these interactions were randomly distributed across the linear sequence of all three nuclear chromosomes (Figure 4.12B). However, some 50 Kb regions on chromosomes I and III contain 2-3 ARS sites that were involved in mt-nDNA interactions. Moreover, the nuclear ARS sites involved in the G1 phase mt-nDNA interactions were enriched for high efficiency and early replicating ARS sites (Figure 4.12C). Finally, the 21S-15SrRNA, Cox1-SPMIT.02 and Cox3-Cob1-SPMIT.06 mitochondrial fragments participated in the majority of the ARS site interactions during the G1 phase (Figure 4.13).

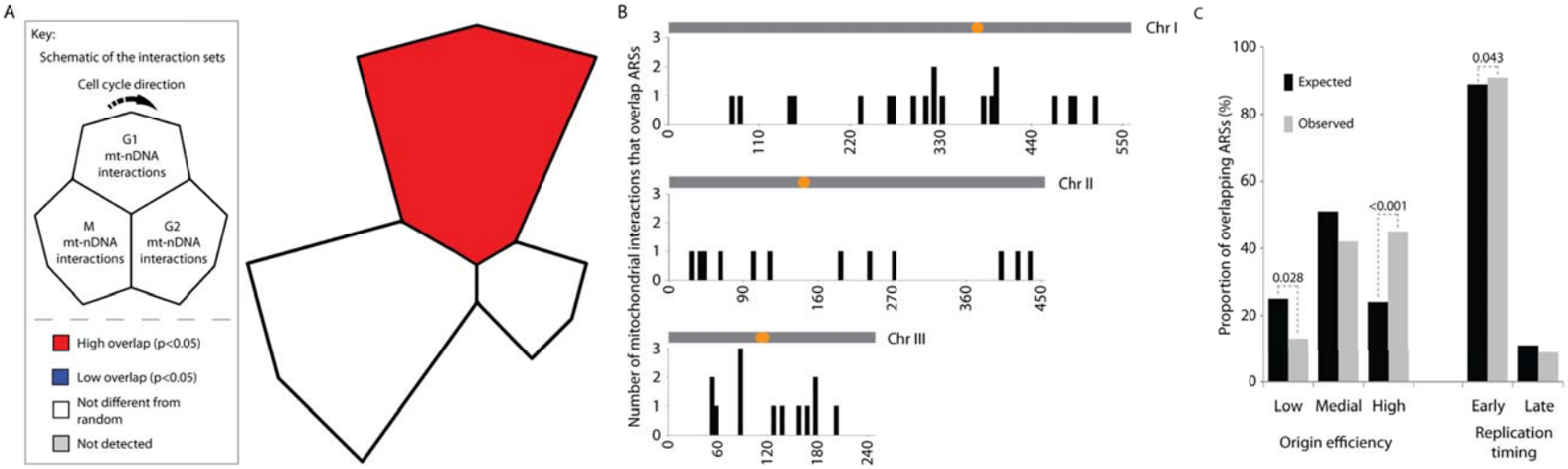


Figure 4.12. G1 phase mt-nDNA interactions are enriched for high efficiency, early replicating autonomously replicating sequences (origins of replication).

(A) mt-nDNA interactions detected during the G1 phase of the cell cycle are enriched for nuclear ARS sites. The proportion mt-nDNA interactions for which the nuclear partner overlapped a nuclear ARS site (+/-500 bp) was compared to an equivalent number of randomly selected nuclear loci. (B) The distribution, within the linear sequences of the nuclear chromosomes, of the ARS sites that were associated with the G1 phase mt-nDNA interactions was not different from random (K-S test). (C(i)) There was a significant enrichment for high efficiency and early replicating ARS sites while interactions with low efficiency ARS sites were depleted within the nuclear loci associated with the G1 phase mt-nDNA interactions.

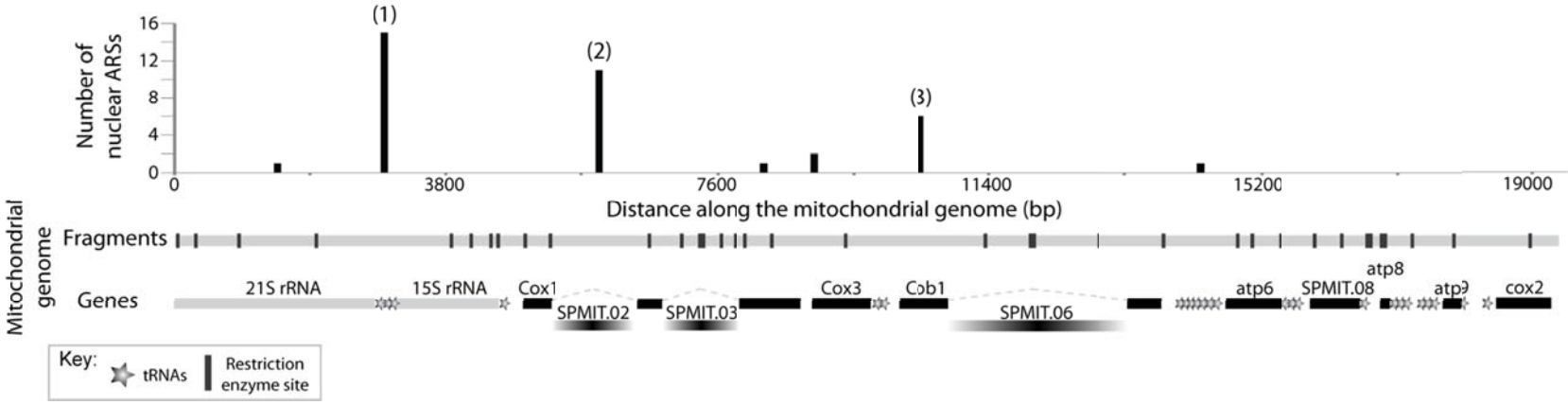


Figure 4.13. Three mitochondrial fragments are involved in the majority of the interactions with ARS sites during the G1 phase of the cell cycle.

The number of ARS sites each mitochondrial fragment interacted with in the G1 phase mt-nDNA interactions was plotted against the position within the mitochondrial genome. The mitochondrial genome AseI fragmentation and genes are also depicted.

4.3 DISCUSSION

4.3.1 MT-NDNA INTERACTIONS REGULATE NUCLEAR ENCODED GENES IN *S. CEREVISIAE*

In this study we have shown that a large number of nuclear loci interact strongly and reproducibly with Mitochondrial DNA in *S. cerevisiae* and that the spectrum of these interactions is dependent upon the carbon source on which the yeast are grown. Interestingly, we find that mt-nDNA interactions are significantly reduced when group II mitochondrial introns that contain reverse transcriptase machinery are deleted. This suggests that the mitochondrial DNA that is involved in these inter-organelle interactions is cDNA that has been reverse transcribed from mitochondrial RNAs. Critically, we demonstrate that suppression of inter-organelle DNA-DNA interactions correlates with elevated transcript levels (Figure 4.2) for the interacting nuclear gene and a reproducible albeit small increase in growth rate (Rodley et al., 2012), suggesting that these interactions are biologically relevant and play a role in regulating nuclear gene expression. This is further supported by previous observations that yeast nuclear transcription responds to the presence or absence of mitochondrial genome sequences (V. Parikh et al., 1987; V. S. Parikh et al., 1989). From these results we propose that the mt-nDNA interactions act as part of an inter-organelle communication system to signal mitochondrial metabolic state and regulate gene expression. While this DNA based inter-organelle communication may seem surprising, there is a large body of evidence demonstrating the presence of mitochondrial DNA in the nucleus and supporting the on-going nature of this transfer (Adams et al., 1999; Adams, Qiu, Stoutemyer, & Palmer, 2002; Brennicke et al., 1993; Campbell & Thorsness, 1998; Farrelly & Butow, 1983; Hazkani-Covo, Zeller, & Martin, 2010; Richly & Leister, 2004; Shirafuji, Takahashi, Matsuda, & Asano, 1997; Thorsness & Fox, 1990, 1993). Thus, it is plausible that the process of mitochondrial DNA transfer has evolved into a functional signaling mechanism. In the case of the glucose and galactose dependent *COX1-MSY1* and *Q0182-RSM7* interactions we have shown a repressive role for mt-nDNA interactions in the control of nuclear transcript levels. However, there is no reason to assume that all interactions are repressive.

4.3.2 MT-NDNA INTERACTIONS DETECTED AT SPECIFIC PHASES OF THE *S. POMBE* CELL CYCLE ARE IMPLICATED IN THE ESTABLISHMENT OF TRANSCRIPTION AND INITIATING DNA REPLICATION

In this study, we detected variations in the number and frequency of interactions between the mitochondrial and nuclear chromosomes during the G1, G2, and M phases of the Fission yeast cell cycle. Variation in connectivity between the mitochondrial and nuclear genomes has also been observed in Budding yeast cells grown under conditions that promote respiration or fermentation (Rodley et al., 2012, 2009). These results support earlier observations that there is continual transfer of specific mitochondrial DNA regions to the nucleus and this transfer varies according to the cell's energy state and cell cycle phase (Ricchetti et al., 1999, 2004; Thorsness & Fox, 1990, 1993). Moreover, the specificity of cell cycle phase mt-nDNA interactions suggests a regulatory role for these connections in cell function and growth.

Mitochondrial genomes exist as a population of nucleoids that are individually comprised of protein complexes and one or more copies of the mitochondrial chromosome (Chen, Butow, & Xin, 2005; Kucej & Butow, 2007). These mitochondrial nucleoids must be segregated during cell division. In the Budding yeast, nucleoids have been shown to be selectively sorted during cell division, however, the molecular mechanism(s) of how this occurs is unknown (Kucej & Butow, 2007; Okamoto, Perlman, & Butow, 1998). The intra-mitochondrial genome interactions that were maintained in all cell cycle phases: G1, G2, and M phase, may reflect the presence of a "stable" nucleoid organization that is maintained throughout the Fission yeast cell cycle. Interestingly, a large number of intra mitochondrial interactions formed specifically during metaphase and were shared with the G1 phase of the cell cycle, consistent with the hypothesis that condensation of a single, or aggregation of multiple copies of the mitochondrial chromosome occurs analogous to, and simultaneously with, nuclear chromosome condensation. Thus, the apparent condensation of mitochondrial genomes during metaphase may reflect mechanisms involved in the faithful segregation of nucleoids during cell division.

The formation of functional mitochondria requires the coordination of gene expression from both mitochondrial and nuclear encoded genes (Butow & Avadhani, 2004; Z. Liu & Butow, 2006). We have previously identified specific contacts between

the mitochondrial genome and nuclear encoded genes in *S. cerevisiae* that modulate the genes' transcript levels (Rodley et al., 2012). The formation of a high proportion of interactions between mtDNA fragments and nuclear encoded mitochondrial genes during the G2-M phase cell cycle transition (Figure 4.8C(ii)) is consistent with our earlier observations of a repressive role for these connections (Rodley et al., 2012). However, the bulk of the mt-nDNA interactions we detected in *S. pombe* did not associate specifically with nuclear encoded mitochondrial genes.

Interactions between mtDNA and other subsets of nuclear genes have positive effects on transcript levels, either by promoting transcription or changing the stability of the transcripts through an as yet undetermined mechanism. This is supported by observations that mt-nDNA interactions that specifically formed during metaphase, contacted genes that were enriched in various molecular binding functions (including anion, ATP, purine and nucleoside binding), and genes that had high transcript levels required for ribosome structure and function. While not specifically annotated as being related to mitochondrial function, these genes are essential for viable cell growth and the synthesis of DNA during S-phase. Considering that the majority of transcription is thought to be supressed during metaphase, it is possible to speculate that these mt-nDNA interactions facilitate/stimulate transcription activation following metaphase exit.

In the Budding yeast, NUMTs are rich in ARS consensus motifs and promote nuclear DNA replication (Blank et al., 2008; Chatre & Ricchetti, 2011). Similarly, NUMTs are also enriched at or near origins of replication in the Fission yeast (Lenglez et al., 2010). Interestingly, merely increasing the amount of mitochondrial DNA has been shown to promote nuclear DNA replication in *S. cerevisiae* by preventing the Sir2 mediated deacetylation of ARS sites (Blank et al., 2008). We determined that a high proportion of mt-nDNA interactions formed during the G1 phase of the cell cycle were associated with strong, high efficiency, early replicating ARS sites in *S. pombe* (Heichinger et al., 2006). Due to the known insertion of NUMTs near ARS sites in the Fission yeast (Lenglez et al., 2010), the analyses was also performed on data that had interactions with NUMTs removed and the same results were obtained. We interpret these results as suggesting a role for mt-nDNA interactions in the regulation of origins of replication. This hypothesis is supported by observations that mitochondria regulate

the G1-S phase cell cycle transition in a number of organisms (Crider et al., 2012; Mandal et al., 2005; Mitra et al., 2009).

Three mitochondrial fragments (*i.e.* 21S-15SrRNA, Cox1-SPMIT.02, Cox3-Cob1-SPMIT.06) were involved in the majority of interactions between the mitochondrial and nuclear chromosomes, including the mt-nDNA interactions formed with ARS sites in the G1 phase. Two of these fragments (*i.e.* Cox1-SPMIT.02, Cox3-Cob1-SPMIT.06) overlapped group II introns (Lang, Ia, Munchen, Davies, & The, 1984). The finding that group II intron containing regions were heavily involved in mt-nDNA interactions in *S. pombe* fits with our earlier observations in *S. cerevisiae* where the group I and II introns within the Cox1 gene were demonstrated to have a role in interaction formation or maintenance (Rodley et al., 2012).

The finding that three mitochondrial regions (*i.e.* 21S-15SrRNA, Cox1-SPMIT.02, Cox3-Cob1-SPMIT.06) were involved in interactions with different nuclear regions in the different stages of the cell cycle indicates that there is a level of regulation of this process which is dependent upon other factors (*i.e.* cell cycle dependent proteins or RNA). This is further supported by the observations that some fragments in the mitochondrial genome predominantly participated in within mitochondrial interactions. For example, the mitochondrial region from SPMIT.08 to atp9, is almost exclusively involved in within mitochondrial genome interactions. Thus there is specific selection for regions of the mitochondrial genome that are transferred to the nucleus. This further emphasizes the existence of a functional role for these fragments once in the nuclear compartment.

In conclusion, we have identified correlations that suggest functional roles for mitochondrial DNA fragments that are undergoing ongoing cell cycle dependent transfer into the *S. pombe* nucleus. Specifically, we identified the formation of mt-nDNA interactions with nuclear genes required for DNA replication, energy availability, and protein synthesis. Moreover, mt-nDNA interactions detected during the G1 phase of the cell cycle are enriched for high efficiency and early replicating origins of DNA replication. Collectively, these results implicate mt-nDNA interactions in the regulation of cell growth and the G1-S phase cell cycle checkpoint. We propose that aberrant mitochondrial to nuclear DNA transfer may contribute to the pleomorphic effects

manifest by mitochondrial dysfunction in complex diseases (Koopman, Distelmaier, Smeitink, & Willems, 2013; Wallace, 2012).

4.4 MATERIALS AND METHODS

4.4.1 STRAINS AND GROWTH CONDITIONS

4.4.1.1 SACCHAROMYCES CEREVISIAE

Saccharomyces cerevisiae strains (Table 4.3) were recovered from -80°C on YES (Sabatinosa & Forsburga, 2010) (2% agar) plates (26°C, 4 days). YES medium (12 ml) starter cultures were inoculated and incubated (26°C, 200 rpm) until the OD_{595} measured ~0.8 (~24 h). Synchronization cultures (125 ml EMM2 (Sabatinosa & Forsburga, 2010), in baffled flasks) were inoculated with starter culture to an OD_{595} = ~0.05 and incubated (26°C, 120 rpm). Cultures were grown for four generations (OD_{595} ~0.8) before synchronization was induced by the addition of pre-warmed EMM2 medium (125 ml, 46°C), instantly raising the culture temperature to a restrictive 36°C. Cultures were incubated (36°C, 140 rpm, for 4 h) to complete synchronization.

Table 4.3. *Saccharomyces cerevisiae* strains used in this study.

Strain	Genotype and Comments	Background
BY4741	*Mata his3**Δ**1 leu2**Δ**0 met15**Δ**0 ura3**Δ**0*	Wild-type
161-U7	$1^{+}2^{+}$ [MATa, ade1, lys1, ura3]. Wild-type (Moran, Zimmerly, Eskes, Kennell, & Lambowitz, 1995)	161-U7
161-U7 GII-0 aI5γ	GII-0 [MATa, ade1, lys1, ura3]. No group II introns aI5γ retained (Boulanger, Belcher, Schmidt, Dib-Hajj, & Schmidt, 1995; Moran et al., 1995; Peebles, Belcher, Zhang, Dietrich, & Perlman, 1993)	161-U7
161-U7 GII-0	GII-0 [MATa, ade1, lys1, ura3]. No group II introns (Moran et al., 1995)	161-U7
mrs1	*Mata his3**Δ**1 leu2**Δ**0 met15**Δ**0 ura3**Δ**0 mrs1Δ::kanMX4*	BY4741

4.4.1.3 SCHIZOSACCHAROMYCES POMBE

For *Schizosaccharomyces pombe* strains and growth conditions see section 3.5.1.

4.4.2 HARVESTING CELLS AND CHROMATIN PREPARATION

S. cerevisiae cells were cultured (30°C, 160rpm) to an OD_{600} of 0.600 (see section: 4.4.1.1). Cultures were cross-linked in formaldehyde (1% final v/v, 10 min, RT). Cells were pelleted (3000 rpm, 3 mins, 4°C) before being washed twice in wash buffer (5 ml, 3000 rpm, 3 mins, 4°C) and suspended in 400 µl FA lysis buffer (0.05 M Hepes KOH [pH 8.0], 0.14 M NaCl, 1 mM EDTA, 1% [v/v] Triton X-100, 0.1% DOC, mini complete EDTS-free protease inhibitors [11836170001 Roche]). Cells were counted and aliquoted into residues of $0.95x10^9$. Acid washed beads (0.4 ml) were added to $0.95x10^9$ cells in FA lysis buffer (400 µl). Cells were vortexed (8 cycles, 30 s, max rpm) and held on ice between cycles (30 s). A hole was punched through the bottom of the tube with a 30 ½ G needle and the sample spun into a new tube to remove the glass beads. The chromatin sample was pelleted (13,000 rpm, 15 mins, 4°C) and washed with FA lysis buffer (400 µl, 13,000 rpm, 15 mins, 4°C) before suspension in 400 µl chromatin digestion buffer (0.01 M Tris-HCl [pH 8.0], 5 mM $MgCl_2$, 0.1% Triton-X100). The sample was then treated with SDS (0.1% final v/v, 37°C, 15 mins) before the addition of 45 µl 11% Triton X-100 (1% final v/v) to remove all unbound SDS.

4.4.3 CHROMOSOME CONFORMATION CAPTURE (3C) SAMPLE PREPARATION

Chromatin (52.6 µl) was digested using the restriction enzyme of choice (100 µl final volume, 100 U, 37°C, 2 hrs) before being inactivated by the addition of SDS (1% final v/v) and heat incubation (65°C, 20 mins). Reactions were diluted in T4 ligation buffer (NEB, 2ml) containing Triton X-100 (1% final v/v). T4 DNA Ligase (20 U, Invitrogen) was added and the reaction incubated (16°C, 2 hrs). Samples were reverse cross-linked (65°C, O/N) in the presence of Proteinase K (3.5-5.5 $ugml^{-1}$, Roche), 20 µl EDTA (0.5 M), 12 µl NaCl (5M), 1.2 µl Tris-HCl (1M). RNase A (20 µg) was added and incubated (37°C, 15 mins) prior to three Phenol:Chloroform (1:1) extractions. DNA was precipitated by addition of absolute ethanol (1 ml) with Na Acetate (40 µl, 3 M) and LPA (0.25% final v/v) and incubation (-20 °C, O/N). DNA was pelleted (13,000 rpm, 25 mins, 4°C) and washed with 70% ethanol (700 µl) before suspension in ddH_2O (40 µl).

4.4.4 QUANTITATIVE 3C ANALYSES

3C samples were prepared as previously described (O'Sullivan et al., 2004). Refer to the supplementary methods for a detailed description. Quantitative 3C analyses (Rodley et al., 2009) were performed using FAM labeled BHQ Probes (BioSearch Technologies; Table 4.4) and Taqman® Gene Expression Master Mix (Applied Biosystems) on an ABI Prism 7000 Sequence Detection System (SDS7000). The mt-nDNA interaction investigated was between the *COX1* gene on the mitochondrial genome (bp 24872 – 26193) and the *MSY1* gene on nuclear Chromosome XVI (bp 365496-365760)). Samples (2 µl in triplicate) were analyzed in a final reaction volume of 20 µl using primers listed in Table 4.4. Assays were performed using a 3-stage program (50°C, 2:00 min; 95°C, 10:00 min; 45x[95°C, 0:15 s; 60°C, 1:00 min]).

Table 4.4. Primers and probes used in this study.

Standard primers were designed using Primer3. Taqman primer and probes were designed using BioSearch Technologies RealTimeDesign online software.

Mt-nDNA 3C TAQMAN	
MitogDNA3CForward	GTGAGCCGTATGCGATGAAAG
MitogDNA3CR13221	GAATCCCTCGCCAACATAGA
MitogDNA3CProbe	FAM-TCGCACGTACGGTTCTTACCGG
RNA qRT-PCR	
RSM7_qRT-PCR_For	TGTCATTCCTGTGCCTCTGA
RSM7_qRT-PCR_Rev	TGGCTGTCTTGTGAATCTGG
MSY1_qRT-PCR_For	CGGCGTATGATGTTTACCAG
MSY1_qRT-PCR_Rev	CCGGAGCCAACTCCATATAA
ACT1_qRT-PCR_cont_F	ACATCGTTATGTCCGGTGGT
ACT1_qRT-PCR_cont_R	AGATGGACCACTTTCGTCGT

4.4.5 RNA EXTRACTION

Total RNA was extracted from *S. cerevisiae* grown in SC (Glucose) to an OD_{600} of 0.600 (see section: 4.4.1.1). Briefly, cells were harvested (4,000 rpm, 4°C, 2 min) and washed with AE buffer (4,000 rpm, 4°C, 2 min; 50mM Sodium Acetate, 10mM EDTA, pH 5.3). The cell pellet was suspended in phenol/chloroform/isoamyl alcohol (400 µl, 24/24/1) and glass beads (400 µl). Cells were lysed in a bead mill (SPEX sample prep 2010, Geno/Grinder; 1,750 rpm, 8 x 30 s cycles with 60 s resting intervals at 4ºC). Lysed cells were frozen (-80°C, 15 min), thawed and pelleted (15,000 rpm, 5 min, 4°C). The aqueous phase was extracted twice with phenol/chloroform/isoamyl alcohol

(400 µl, 24/24/1). Total RNA was pelleted (15,000 rpm, 10 min, 4°C), following addition of 2/3s volume of 8M LiCl and freezing (-20°C, 2h). RNA was washed (70% ethanol), and the pellet air-dried. Total RNA was suspended (60°C, 10 min) in 80µl of DECP treated water (Invitrogen). DNA was removed from the total RNA samples (5µg, 20 µl) by treatment with 1µl of TURBO DNase (TURBO DNA-*free*™ Kit, Ambion) as per manufacturer's instructions. Samples were centrifuged (10,000g, 1.5 min) and the supernatant was retained. Total RNA concentration was measured using a Nano-drop and 50 µl samples (50ng/µl) were stored at -80ºC.

4.4.6 Quantitative Reverse Transcription-PCR

qRT-PCR standards were amplified from *S. cerevisiae* BY4741 genomic DNA (Table 4.4). PCR products were purified (Zymo DNA clean and concentrator^{TM}-5 kit according to manufacturer's instructions). The concentration of each qRT-PCR standard was determined by Nano-drop and used to make dilutions ranging from 4.0 - 4.0X10^{-5} ng/µl. qRT-PCR reactions were performed using One Step SYBR® Ex Taq^{TM} qRT-PCR Kit according to the manufacturer's instructions (TaKaRa). The qRT-PCR was run with the following protocol: 42ºC, 5 min; 95ºC, 10 s; 40x [95ºC, 5 s; 60ºC, 31 s] 95ºC, 15 s; 60ºC, 1min; 95ºC, 15 s. All transcript levels were standardized to nuclear *ACT1* and expressed as percentage of wild-type (set at 100%) +/- standard error of the mean.

4.4.7 The generation of *S. pombe* data

The methods for the generation of the *S. pombe* data pertaining to: Synchronisation efficiency 3.5.2, Chromatin isolation for GCC 3.5.3, Network assembly 3.5.5, Generating heat maps 3.5.8, Loop lengths of interactions within the mitochondrial genome 3.5.13, RNA extraction 3.5.9, Transcriptome analysis 3.5.11, and Gene Ontology analysis 3.5.15, can be seen in the indicated sections.

4.4.8 Significance cut-off calculations

Random ligation events can occur during the two ligation steps in the GCC protocol: 1) the ligation of the cross-linked fragments; and 2) linker addition during preparation for sequencing. External ligation controls (Table 3.8) were produced according to Cagliero et al. 2013 and incorporated into the samples prior to the GCC ligation step to enable the estimation of the frequency of inter-molecular ligation

events. We only detected a single ligation event between the E. coli ligation control, which was included in the G1 phase GCC preparations, and the mitochondrial genome in one of the G1 phase biological replicates. Therefore, interactions that occurred with the mitochondrial genome were deemed significant if they were detected at a frequency >=2 in both biological replicates.

4.4.9 Extracting nuclear genes associated with mt-nDNA interactions for GO and transcription analyses

Nuclear encoded genes associated with mt-nDNA interactions were extracted and used for gene ontology (GO) analysis. Genes that were internal to, overlapped with or were the nearest up and down stream genes to the nuclear interacting fragments of mt-nDNA interactions were extracted and termed mt-nDNA associated gene sets (Figure 4.9). Gene sets associated with specific mt-nDNA interactions were used for GO enrichment analysis using AmiGO (see section 3.5.15) and their presence in transcription data sets was determined (see section 4.4.10).

4.4.10 Determining if nuclear encoded gene sets associated with mt-nDNA interactions were detected in specific transcription data sets

The transcript levels of nuclear encoded genes associated with mt-nDNA interactions and the likelihood that these genes were found in specific transcription data sets (*i.e.* high, low, or differentially regulated during cell cycle transitions) were compared to random gene sets. Briefly, we generated one thousand random sets containing equivalent numbers of genes and used the number of times the random gene sets had a higher or lower proportion of genes associated with the transcription gene sets to estimate significance (Figure 4.9, Table 3.4 and Table 3.5). If there was a significantly high proportion of mt-nDNA associated genes detected in a transcription data set, the shared genes were used for GO Term Enrichment analysis (see section 3.5.15).

4.4.11 Determining whether nuclear restriction fragments involved in mt-nDNA interactions overlap with nuclear encoded mitochondrial genes and ARS sites

We determined if the nuclear restriction fragments involved in the mt-nDNA interactions overlapped nuclear encoded mitochondrial genes more than expected at random. Genomic co-ordinates of nuclear encoded mitochondrial genes were obtained from PomBase using the advance search filter: GO_ID; GO:0005739 - mitochondrion. The proportion of mt-nDNA interactions that overlaped nuclear encoded mitochondrial genes (gene co-ordinates and +/-500 bp) was calculated for specific interaction sets.

The transcript level of nuclear encoded mitochondrial genes, found to be associated with interactions, was compared to random gene sets (see section 4.4.10). We generated sets of random genomic regions containing the same number of identically sized nuclear fragments by randomly selecting a start coordinate for each region and then adding the length (bp) of the restriction fragment to obtain the end coordinate. Two separate populations of random sets were generated: 1) the conserved random [CR] sets. CR sets conserved the number of genomic regions per chromosome and ensured that significant results were not due to the linear distribution of the regions on a particular chromosome(s); and 2) the random [R] sets where regions were randomly selected across the entire genome, with chromosome selection determined at a frequency that was relative to the chromosome lengths. One thousand random data sets were generated for the CR and R methods. The frequency with which the restriction fragments involved in the mt-nDNA interactions overlapped nuclear encoded mitochondrial genes within the actual data was higher or lower than the random data sets was used to estimate significance.

The frequency with which restriction fragments involved in mt-nDNA interactions associated with known DNA replication origins (ARS sites) was compared to sets of random genomic regions to determine if the association was non-random. Chromosomal coordinates for the strongest 401 origins of replication and their efficiency and replication timing were obtained from Heichinger C et al. 2006. ARS regions were defined as the centre coordinate for each ARS (Heichinger et al., 2006) s +/-125 bp or +/-5,000 bp. The frequency with which mt-nDNA interactions overlaped the ARS sites was calculated for specific interaction sets and compared to sets of

random genomic regions as above. When there was a significant association with ARS sites, we performed a Kolmogorov-Smirnov (K-S) test to determine if the contacted sites had a non-random linear chromosomal distribution. ARS sites were further divided into low (<18; 100), medial (18-40; 204), and high (>40; 97) efficiency; and early (<80 min; 360) and late (41) replicating sites. The proportion of contacts with each ARS category in the real and random sets was compared and used to determine significance. Numbers in parentheses represent: the criteria used to divide the origins (*i.e.* Origin efficiency or replication timing) into subgroups, and the number of origins in each subgroup.

Chapter 5

5 DISCUSSION AND FUTURE DIRECTIONS

Over recent years the spatial and temporal organization of chromosomes, in both prokaryotic and eukaryotic cells, has become increasingly recognized as playing an integral role in various cellular processes (Cagliero et al., 2013; Cavalli & Misteli, 2013; de Wit et al., 2013; Denholtz & Plath, 2012; Kuzminov, 2013; Peric-Hupkes et al., 2010; Sherratt, 2003; Umbarger et al., 2011; W. Wang et al., 2011; Woldringh & Nanninga, 2006). Likely one of the clearest demonstrations that specific chromosome conformations participate in cellular processes is the regulated formation of specific DNA-DNA interactions between enhancer elements and genes they control that are distant in the linear sequence (Marsman & Horsfield, 2012; Nolis et al., 2009; Stadhouders et al., 2012). Recently, the O'Sullivan lab developed a method called Genome Conformation Capture (GCC) that enabled the capture of DNA-DNA interactions on a genome-wide scale in an unbiased way (Rodley et al., 2009). I have utilized this method and combined it with transcript level data from a number of organisms (*Escherichia coli*, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae*) to investigate the functional role of changes in genome organization in response to environmental stimuli and during cell cycle progression.

My work has identified that the observable expansion of the *E. coli* nucleoid in response to an induced amino acid starvation, corresponds with a reduction in the number of long distance (>800 bp) chromosomal contacts. Two clear chromosomal domains were observed in wild-type (wt) and starved cells; Ori and Ter, distinguishable based on their level of connectivity with the genome. Furthermore, increasing levels of connectivity towards the origin of replication correlated with an increase in the copy number of origin proximal loci. This Ori dependent interaction frequency and genomic copy number increase was attributed to the continual initiation of DNA replication in wt cells. In addition, the formation of specific interactions, possibly mediated by the hemimethylated DNA binding protein SeqA, was related to

the initiation and progression of chromosomal replication. Combining the 3D maps of the *E. coli* nucleoid in wt and starved cells with gene transcript level data revealed that genes with high transcript levels in wt cells co-localize (cluster) with each other at a high level and remain highly clustered following down regulation. By contrast, up regulated genes were not highly connected to the genome or themselves (clustered) in wt or starved cells. Collectively, these results reveal that the *E. coli* nucleoid is organized to facilitate and optimize replication and transcription in wt cells.

In future experiments it would be interesting to attempt to synchronize *E. coli* cells and ideally reduce the number of genome copies to one. A couple of methods that could be used to obtain synchronous *E. coli* cells are the treatment with Serine Hydroxamate (SHX), which arrests the initiation of chromosome replication, or by growing cultures through one or two rounds of stationary phase (Cashel et al., 1996; Cutler & Evans, 1966). Cells would be subsequently released from the cell cycle arrest (SHX treatment of stationary phase) and grown through a single round of DNA replication. Monitoring synchronized *E. coli* cultures as they progress through a single round of DNA replication and segregation would give further insight into the contribution that NAPs have in nucleoid organization at different growth phases. In particular, it would enable the further investigation into the matS5 and 10 Ter-domain loop and the role that SeqA has in the organization of replicated nucleoids. These two results could be interrogated by the use of quantitative-3C (q3C). For example, the presence of only one copy of the *E. coli* chromosome and a single round of DNA replication would enable the identification of whether the matS5-10 loop occurred within a single chromosome or between replicores, as suggested in chapter 2.

I then investigated whether changes in genome organization in naturally cycling cells participated in the establishment of cell cycle phase specific transcription patterns. Using *S. pombe* as a cell cycle model, clear differences in genome organization were detected between the G1, G2 and M phases of the cell cycle. The characteristic Rabl conformation of yeast chromosomes was detected at all cell cycle phases but the identity of the telomeres that co-localized into clusters varied in a way that appeared to influence the expression of sub-telomeric genes. The characterization of the *in vivo* structure of highly condensed metaphase chromosomes provided evidence that they form a ‘polymer melt’ like structure (Naumova et al.,

2013). Furthermore, my analyses determined that metaphase chromosomes remain connected to other chromosomes, potentially to facilitate the post-mitotic reactivation of essential highly expressed genes. Combining the 3D analysis of the genome with paired gene transcript level data revealed that genes with high transcript levels were highly connected with other regions of the genome and other highly expressed genes in a cell cycle dependent manner. By contrast, genes with low transcript levels were not highly connected. Interestingly, the formation of specific interactions during the G2-M-G1 cell cycle phase transitions associated with groups of genes enriched in gene ontology groups, potentially implicates interactions in gene bookmarking.

The detection of interactions between the mitochondrial and nuclear chromosomes (mt-nDNA interactions) in both *S. cerevisiae* and *S. pombe* identified a potentially unique signalling mechanism that has arisen during the evolution of endosymbiosis to coordinate the two cellular compartments. First I determined that reducing the interaction frequency between regions on the *S. cerevisiae* mitochondrial genome and specific nuclear encoded genes correlated with an increase in gene transcript levels. I then mapped mt-nDNA interactions on a genome-wide scale throughout the *S. pombe* cell cycle. It was found that in *S. pombe* mt-nDNA interactions formed during metaphase occurred with genes required for DNA synthesis and cell growth, and mt-nDNA interactions detected in the G1 phase were enriched for high efficiency, early replicating origins of DNA replication. Collectively, these results suggest a role for mt-nDNA interactions, together with nuclear chromosomal interactions, in facilitating the regulation of gene expression and cell cycle progression, specifically during exit from metaphase and the G1-S phase cell cycle checkpoint.

5.1 Conserved and structural vs dynamic and functional

Many aspects of the results I have presented agree with previous studies that have investigated the three-dimensional organization of genomes. At a broad level interactions can be ascribed to two categories; non-specific – a consequence of stochastic and random processes, and functional – actively and passively facilitating cellular processes. *In vivo* the chromatin fibre is a dynamic molecule that is in continual motion. This motion is governed by, but not limited to, thermodynamics, molecular crowding, physical constraints, and active processes that influence the

flexibility of the chromatin fibre (Cook & Marenduzzo, 2009; de Nooijer, Wellink, Mulder, & Bisseling, 2009; Fritsche & Heermann, 2011; Gehlen et al., 2012; Marenduzzo, Micheletti, & Cook, 2006; Marenduzzo, Micheletti, & Orlandini, 2006; Nicodemi & Prisco, 2009). Stochastic motion inevitably leads to random collisions between distal genomic regions. Alternatively, two regions of a genome may contact randomly as a result of an active process that modifies the topography of the genome. For example, in *E. coli* topoisomerase driven supercoiling of the chromosome produces a local chromatin structure of isolated ~10 Kb topological domains (Postow et al., 2004). These are assigned as non-specific interactions as functional roles for them cannot currently be distinguished from their stochastic nature.

The large number of interactions that I observed between regions of the *E. coli* genome separated by <800 bp may be partly the result of the detection of random contacts formed due to supercoiling. However, the majority of these interactions are conserved between the wt and SHX treated cells. This may suggest that these interactions reflect the formation of a “stable” local domain or fibre structure in the *E. coli* chromosome. Perhaps due to the non-random formation of supercoiled domains that would result in the establishment of similar structures and lead to the reproducible detection of interactions. Similarly, in *S. pombe* the majority of intra-chromosomal interactions were conserved between all three cell cycle phases and were not significantly involved in the establishment of cell cycle specific transcription patterns. This may also reflect the detection of non-specific ‘structural’ interactions formed due to the “stable” local folding of the chromatin fibre. Interestingly, computer modelling of the 3D organization of chromosomes in a number of organisms has revealed that gross level genome organization can be largely reproduced by the formation of non-specific interactions as a result of imposing a number of known spatial constraints on the chromosomes (*i.e.* the nuclear envelope, SPB, nucleolus, and Rabl conformation of the chromosomes) (de Nooijer et al., 2009; Gehlen et al., 2012; Marenduzzo, Micheletti, & Orlandini, 2010; Tanizawa et al., 2010).

In the Budding and Fission yeasts a number of chromosomal constrains have been identified, including the centromeres being bound to the spindle pole body (the microtubule organizing centre) and telomeres are commonly observed clustered at the nuclear periphery (Alfredsson-Timmins et al., 2007, 2009; Funabiki et al., 1993; Heun

et al., 2001; Schober et al., 2008). These may be considered as predominantly structural interactions, shaping the chromosomes into a Rabl conformation in *S. pombe* and contributing to genome function in an indirect – passive – way (Allshire, Nimmo, Ekwall, Javerzat, & Cranston, 1995; Duan et al., 2010; Hall et al., 2003; Schober et al., 2008). In agreement with this I detected a high level of connectivity between the centromeres and telomeres in *S. pombe* supporting earlier observations that the chromosomes assume a Rabl conformation (Tanizawa et al., 2010). However, it is likely that the formation of interactions between centromeres and telomeres is in part the result of passive processes, for example, the entropically favourable state of clustered heterochromatic regions (de Nooijer et al., 2009; Marenduzzo, Micheletti, & Cook, 2006).

In crowded environments such as the nucleus, entropy plays an important role in subdividing large and small particles through the depletion attraction force (Bohrmann, Haider, & Kellenberger, 1993; Marenduzzo, Micheletti, & Cook, 2006; Marenduzzo, Micheletti, & Orlandini, 2006). In a nuclear context, depletion attraction has been implicated in being responsible for the formation of a number of observed features of spatial nuclear organization. Such as, the aggregation of DNA and RNA polymerases into replication and transcription factories, respectively, the co-localization of ribosomal DNA repeats into the nucleolus, and the clustering of telomeres and centromeres into distinct foci (Cook & Marenduzzo, 2009; de Nooijer et al., 2009; Marenduzzo, Micheletti, & Cook, 2006; Marenduzzo, Micheletti, & Orlandini, 2006). However, even if the formation of centromere and telomere clusters may be largely a passive process, the maintenance of heterochromatin at these regions and their clustering appears to play an essential role in chromosome stability and faithful segregation during cell division (Allshire et al., 1995; Kellum & Alberts, 1995; Martienssen, Zaratiegui, & Goto, 2005; Misteli & Soutoglou, 2009).

Evidence also exists that telomere clustering plays a functional role in the regulation of gene expression and DNA repair. For example, in the Budding yeast telomere clustering has been proposed to influence the rate of recombinational repair and to coordinate transcriptional programs that ensure evolutionary advantage (Fabre et al., 2005; Halme, Bumgarner, Styles, & Fink, 2004; Louis, Naumova, Lee, Naumov, & Haber, 1994; Turakainen, Naumov, Naumova, & Korhola, 1993). Furthermore, the

sequestration of SIRs into telomeric foci both favours subtelomeric repression and prevents promiscuous effects on a distinct subset of promoters (Taddei et al., 2009). It has also been found that in *Plasmodium falciparum*, the spatial juxtapositioning of telomeres appears to favour the monoallelic expression of subtelomeric virulence factors that are essential for the parasite to escape the immune system response (Scherf, Lopez-Rubio, & Riviere, 2008). Herein I present evidence that telomeres cluster throughout the cell cycle, but that the identity of the telomeres that cluster together is cell cycle dependent. Due to the functional repercussions of telomere clustering, whether there is preferential juxtaposition of telomeres in *S. cerevisiae* has been a long standing question of interest (Schober et al., 2008). In addition, the formation of specific telomere clusters correlated with cell cycle specific changes in subtelomeric gene expression, providing further support that the specific co-localization of telomeres has a functional role. Therefore, it is likely that, despite a proportion of the interactions detected between heterochromatic regions in *S. pombe* being non-specific as a result of passive entropically favourable clustering, a subset of these interactions are specifically formed and have a functional role.

Entropic effects also appear to have a significant influence on the spatial organization of bacterial genomes (Fritsche & Heermann, 2011; Jun & Mulder, 2006). For example, it has been shown that all topologically distinct domains of a confined polymer complex effectively repel one another to maximize the total conformational entropy. As a result, in a rod-shaped cell, the conformational entropy of a duplicating circular chain alone provides a physical driving force for segregation (Jun & Mulder, 2006). In *E. coli* I identified interactions that were both replication dependent and independent, perhaps implicating a subset of the interactions in the formation of topological conformations that enhance or repress the effect of entropy driven chromosomal segregation. Given that rod-shaped bacterial cells are able to take advantage of conformational entropy to facilitate the segregation of replicating chromosomes; this raises an interesting question as to how chromosome segregation occurs in spherical bacterial cells that are unable to take advantage of the entropic effect. It would be interesting to further investigate the link between the molecular level organization of bacterial nucleoids and the influence of confinement and chromosome segregation on nucleoid structure.

One unique model that may offer the opportunity to study both of these aspects at once: the influence of confinement and chromosome segregation on nucleoid structure, would be *Pseudomonas fluorescens* SBW25 (Leij, Sutton, Whipps, Fenlon, & Lynch, 1995). *P. fluorescens* is able to tolerate the deletion of a generally essential gene in rod shaped bacteria, the rod-shape determining protein: mreB, which is involved in cell shape determination and chromosome segregation (Rainey, PB. Personal communication (Figge, Divakaruni, & Gober, 2004; Gitai, Dye, Reisenauer, Wachi, & Shapiro, 2005; Gitai, Dye, & Shapiro, 2004; L. J. F. Jones, Carballido-López, & Errington, 2001)). The apparent direct result of deleting the mreB gene in *P. fluorescens* is that the cells become spherical, but despite this they are able to continue to grow and divide. Investigating the 3D organization of nucleoids in wt and *Δmre*B *P. fluorescens* cells would shed light on the effect that the rod-shaped confinement has on nucleoid organization. Further, synchronizing the cells and following them through a single round of DNA replication and cell division would potentially enable the distinction between interactions that are due to: 1) the rod-shaped confinement, 2) DNA replication, or are 3) confinement and replication independent interactions.

Even though non-specific interactions and external constrains appear to play a significant role in the spatial organization of genomes they are not sufficient to describe the complexity of 3D genome organization. Indeed the regulated formation and disruption of chromosomal interactions have been shown to make a non-trivial contribution to genome organization and function (Gehlen et al., 2012), with growing evidence for the specific formation of functional interactions: between distal regions within a chromosome, between separate chromosomes, and even between genomes in separate intracellular organelles (*i.e.* the mitochondrial and nuclear chromosomes) (Bartkuhn & Renkawitz, 2008; Cook, 1999; Goetze et al., 2007; Iwasaki et al., 2010; Misteli & Soutoglou, 2009; Rodley et al., 2012; Simonis & de Laat, 2008; G. S. Stein et al., 2003). The formation and disruption of these interactions has been attributed to being involved in the regulation of many cellular processes, with clear interest in their role in gene expression (Bartkuhn & Renkawitz, 2008; Cook, 1999; Goetze et al., 2007; Iwasaki et al., 2010; Misteli & Soutoglou, 2009; Simonis & de Laat, 2008; G. S. Stein et al., 2003). The work presented here provides further support for the specific formation and disruption of interactions affecting a cell's response to environmental

conditions and cell cycle progression, including the regulation of genes required for adaptation. I also provide evidence that interactions between the mitochondrial and nuclear chromosomes in *S. pombe* potentially coordinate the initiation of DNA replication with the energy state of the cell. Specific interactions related to the initiation and progression of the DNA replication fork in *E. coli* were also detected, further implicating the formation of interactions in the regulation of replication. Collectively, this gives rise to the theme that environment and cell cycle phase independent interactions are predominantly "stable" (shared by different conditions) and non-specific or structural in nature. By contrast, environment and cell cycle phase dependent interactions are "dynamic" (specific to different conditions) and functional in nature, facilitating the adaptation to changes through time *e.g.* in response to environmental stimuli or cell cycle progression.

5.2 RESPONSE TO THE ENVIRONMENT AND CELL CYCLE PROGRESSION

The dynamics of chromosome re-organization in response to environmental changes or cell cycle progression can clearly be observed microscopically (Alfredsson-Timmins et al., 2009; Durfee et al., 2008; Funabiki et al., 1993; Heun et al., 2001; Ramírez & Surrallés, 2009; Traxler et al., 2008). However, to what degree the spatial arrangement of chromosomes is re-organized at the molecular level or conserved during these observed changes, remains unresolved. The data presented here supports the idea that the majority of interactions detected are within chromosomes and cover a short linear chromosomal distance (*E. coli*, <800 bp; *S. pombe*, <50 Kb). Furthermore, these short distance interactions are largely conserved between different conditions and are seldom significantly associated with genes required for cell growth and adaption. This suggests that the localized chromatin structure in bacteria and eukaryotes is relatively impervious to changes in spatial chromosome organization in response to the environment and during the cell cycle and do not contribute significantly to cell adaptation. By contrast, there was a large degree of variability in long distance (*E. coli*, >800 bp; *S. pombe*, >50 Kb) intra- and inter-chromosomal interactions. Moreover, genes with high transcript levels, and in specific situations genes that underwent differential regulation in response to environmental change (*E. coli*) or cell cycle progression (*S. pombe*), are frequently highly associated with specific sets of interactions and enriched in GO terms. For

example, the maintenance of high levels of co-localization between essential genes throughout the *S. pombe* cell cycle correlated with stable, high levels of transcription. Therefore, even though only a small proportion of the interactions detected in both bacteria and eukaryotes are environmental or cell cycle progression responsive, they are implicated in a cell's ability to adapt and reproduce.

5.2.1 THE LINEAR AND SPATIAL CLUSTERING OF CO-REGULATED GENES

Over recent years, the increasing interest in the molecular level 3D organization of chromosomes and its role in the regulation of nuclear processes have seen a rapid accumulation in data alongside advances in methodologies (de Wit et al., 2013; Dekker et al., 2002; Dostie et al., 2006; Grand et al., 2011; Lieberman-Aiden et al., 2009; Rodley et al., 2009; Sexton et al., 2012; Tanizawa et al., 2010; Z. Zhao et al., 2006). Of particular interest has been the role in the regulation and coordination of gene transcription. Chromatin looping is now well established as contributing to gene regulation by enabling the co-localization of linearly distant elements, such as enhancers, silencers, imprinting control regions, and locus control regions, with the genes they regulate (Crawford, Davis, et al., 2006; Crawford, Holt, et al., 2006; Follows et al., 2006; Q. Li et al., 2006; Tolhuis et al., 2002). One of the most well characterized examples of specific loop formation regulating gene expression in a developmentally dependent manner is that of the β-globin locus during mammalian development. Looping between upstream Locus Control Regions (LCR), found in DNAseI hypersensitive sites linearly distant from the β-globin gene cluster and gene promoters, in a developmentally dependent manner, regulates β-globin gene expression (Chien et al., 2011; Fromm & Bulger, 2009; Noordermeer & de Laat, 2008; Ragoczy, Telling, Sawado, Groudine, & Kosak, 2003; Stamatoyannopoulos, 2005). At a more gross level, detailed microscopic studies looking at the cellular localization of actively transcribing RNA-polymerase II (pol II), have revealed that it accumulates into highly concentrated foci called transcription factories in both bacterial and mammalian cells (Carter et al., 2008; Papantonis & Cook, 2013; W. Wang et al., 2011). The association of genes with transcription factories is thought to account for the majority of cellular transcription (Bartlett et al., 2006; Papantonis & Cook, 2013). Consistent with this hypothesis, I detected that genes with high transcript levels in both *E. coli* and *S. pombe* contacted other genomic regions and co-localized with other highly

transcribed genes at a high frequency. By contrast, genes with low transcript levels were less frequently highly connected with other genomic regions or other lowly transcribed genes. Therefore, the specific formation and maintenance of interactions between highly transcribed genes correlates with their stable, high level of transcription. The mechanisms for the establishment of specific spatial genome organization to promote transcription remain largely unknown. It has been postulated that depletion attraction may make a contribution with the entropically favourable co-localization of pervasively active euchromatic regions with high concentrations of RNA polymerase (Carter et al., 2008; Marenduzzo, Micheletti, & Orlandini, 2006; Toan, Marenduzzo, Cook, & Micheletti, 2006).

It is interesting to note that genes that were highly transcribed in wt *E. coli* were highly clustered, but also remained highly clustered following the induced amino acid starvation and their coinciding down regulation. Considering that the *E. coli* cells were only treated with SHX for a short period of time (~30 min) this may reflect a form of spatial memory. For example, if the cells are only exposed to a new environment for a short period of time, maintaining a nucleoid organization optimum for the original environment would enable rapid continuation of growth upon return to that environment. Further, it is likely an energetically expensive process to fully re-arrange the spatial organization of the nucleoid and, therefore, should only be done if absolutely necessary. Future work could address this idea by performing pairwise 3D genome organization and transcriptome studies with bacteria that are exposed to a new environment for prolonged periods of time before being returned to the initial environment. The hypothesis would be that once the bacteria had been exposed to the new environment for a certain period of time that the nucleoid would be re-organized to promote optimum transcription for the new condition. Combining such an experiment with the deletion of proteins thought to be involved in the establishment of nucleoid organization, would give insight into the mechanisms of the regulation of nucleoid organization in response to environmental changes.

It has been proposed that not only does transcription occur predominantly at transcription factories but that the level of transcript required may be regulated by the period of time that a gene remains associated with a transcription factory. The level of transcript produced would, however, also depend on the concentration of required

transcription factors found at the transcription factory (Carter et al., 2008; Eskiw et al., 2010; Papantonis & Cook, 2013). Therefore, genes that were expressed at a low level would be infrequently associated with, and spend little time at, transcription factories. Consequently, if they were not in spatial proximity (1-100 nm (Dekker et al., 2002; Grand et al., 2011; van Steensel & Dekker, 2010)) at the time of sample fixation, they would not be detected in our data. Similarly, the detection of differentially regulated genes associating with transcription factories during cell cycle transitions in *S. pombe* would only occur if they were being expressed at the time of fixation. In agreement with such an idea, the relationship between changes in genome organization throughout the *S. pombe* cell cycle and corresponding fluctuations in gene transcript levels is complex. The formation and disruption of interactions correlated with both the up and down regulation of gene sets. Regardless of the perhaps ambiguous nature of these interactions, the enrichment and depletion of differentially regulated genes in specific interaction sets implicates them in the establishment of cell cycle specific transcription patterns.

It is becoming clear that specific region of the genome with different levels of transcriptional activity form contacts at distinct frequencies. In general, genes with high transcript levels contact and co-localize at high frequency and differentially regulated genes show variable contact frequencies depending on for example the phase of the cell cycle. However, these results do not predict the spatial position of the locus in the cell or nucleus. The development of a computer model that includes physical and biological constraints on the folding of a linear polymer(s), which represent chromosome(s), and the incorporation of subsets of experimentally determined chromosomal contacts, would give a depiction of chromosome folding (*e.g.* (Gehlen et al., 2012)). Visualizing the spatial position of specific genomic features (*e.g.* genes with high or differential transcript levels or protein binding sites and epigenetic marks) within an ensemble of genome conformations would enable the investigation of sub-nuclear compartmentalization. Furthermore, the selection of specific loci for experiments using q3C and visualization using high-resolution microscopy would further validate the role of chromosome contacts and reveal the localization of specific loci in single cells (Dekker et al., 2002; B. Huang et al., 2008; Matsuda et al., 2010; Simonis & de Laat, 2008; Volpi & Bridger, 2008; X. Wang et al., 2008).

The spatial co-localization of genes that need to be highly transcribed could be seen to refine the process of transcription by enabling the formation of 'transcriptional hubs'; regions with high concentrations of transcription factors, RNA polymerase, and other regulatory elements that establish an environment to promote efficient transcription (Papantonis & Cook, 2013). If this is the case then an interesting question arises: Is there evolutionary selection pressure imposed upon the linear arrangement of genes by the advantage to have co-regulated genes clustered in the linear sequence and the increased benefit of the additional co-localization in space? The linear clustering of co-regulated genes in the chromosomal sequence is most profoundly illustrated by the operon (polycistronic) arrangement of bacterial genomes (Lawrence, 2002). Despite not being as clearly defined as bacterial operons, gene clusters and regions of increased gene expression (RIDGES) are also observed in mammals (Caron et al., 2001; Lawrence, 2002). I observed that, genes that were highly expressed in wt *E. coli* cells were non-randomly distributed along the chromosome in predictable clusters of up to 32 genes (ACF >0.83). Likewise, the linear chromosomal distribution of genes with high transcript levels in *S. pombe* was non-random across most chromosomes. Interestingly, the detection of a high level of co-localization between differentially regulated genes (specifically: genes up regulated during the G1-G2 phase cell cycle transition, and genes down regulated during the G2-M phase cell cycle transition) appeared to be related to their non-random linear chromosomal distribution. These findings raise two interesting points: 1) even though there is a large difference in the linear chromosomal arrangement of bacterial and eukaryotic genomes, genes that are highly expressed (and differentially regulated at specific times) during natural growth are clustered in the linear sequence, and 2) that the linear clustering appears to be related to the increased likelihood of spatial co-localization. This suggests that there is an advantage, not only in the linear clustering of co-regulated genes, but also likely in the specific positioning of these clusters in a way that promotes spatial co-localization enabling the additional refinement of transcriptional regulation. Therefore, understanding the full significance of the evolutionary selection for particular linear chromosome arrangements will only become evident upon the investigation of their 3D relationship.

5.2.3 SPATIAL BOOKMARKING?

It is generally accepted that during metaphase there is a greatly reduced level of gene transcription coinciding with the observed condensation of chromosomes. Consequently, following metaphase exit and chromosome de-condensation, transcription profiles must be (re)-established for the coming growth and DNA synthesis phases of the cell cycle. Consistent with this idea, I detected the lowest transcript levels of *S. pombe* genes in M phase (data not shown) and the largest number of genes was up regulated during the M-G1 phase cell cycle transition. Over recent years it has become increasingly evident that the transmission of epigenetic marks and the retention of specific transcription factors through metaphase of the cell cycle and even transgenerationally, facilitates the establishment of correct transcription profiles (Blomen & Boonstra, 2011; Kadauke & Blobel, 2013; Sarge & Park-Sarge, 2005; R. Zhao et al., 2011). However, to what degree the 3D organization of genomes is transmitted through metaphase of the cell cycle remains controversial (Blomen & Boonstra, 2011; Essers et al., 2005; Gerlich et al., 2003; Thomson et al., 2004). It has been shown in the Budding yeast that the spatial position of specific loci is heritable and facilitates the rapid initiation of gene expression in response to environmental stimuli (D. G. Brickner et al., 2007; J. H. Brickner, 2009; Laine et al., 2009; Tan-Wong et al., 2009). For example, the Budding yeast *GAL1* locus is repositioned to nuclear pore complexes at the nuclear periphery in response to the presence of galactose. The maintenance of this spatial positioning of the *GAL1* locus at the nuclear pore throughout the cell cycle facilitates the rapid induction of gene expression upon re-exposure to galactose. Data presented herein for *S. pombe* provides support for the idea that there is some degree of conservation of chromosome structure through metaphase of the cell cycle, predominantly consisting of localized chromosome folding (<50 Kb).

The 3D structure of the *S. pombe* metaphase chromosomes are most dissimilar to the other phases investigated with a large increase in within chromosome loops <5 Kb in length and the largest number of phase specific interactions. Very recently, Hi-C analysis and computer simulations of human metaphase chromosomes were performed and in agreement with my prediction for *S. pombe* metaphase chromosomes, they appear to assume a polymer-melt like structure (Naumova et al., 2013). Naumova et.al. (2013) also predict that there is no conservation of

chromosome structure throughout metaphase and that 3D genome organization is established *de novo* in the G1 phase. By contrast, my analysis combining 3D genome organization and paired transcript level data, revealed that despite the grossly different structure of metaphase chromosomes compared to during interphase, the formation and conservation of specific interactions through metaphase correlated with the establishment and maintenance of gene transcription following metaphase exit. In addition, I identified the formation of interactions between the mitochondrial genome and nuclear chromosomes - mt-nDNA interactions - during metaphase. These interactions were associated with genes that had high transcript levels throughout the cell cycle and were enriched in GO terms related to cell growth and DNA synthesis, particularly ribosome structure and function. It is interesting to speculate that the formation of specific interactions during metaphase may assist the establishment of a chromatin environment that promotes the rapid and high level expression of genes following metaphase exit. Effectively, metaphase interactions behave as a form of spatial bookmarking (Blomen & Boonstra, 2011; Kadauke & Blobel, 2013; Sarge & Park-Sarge, 2005).

5.3 REPEAT ELEMENTS MAY BE INVOLVED IN THE ESTABLISHMENT OF SPECIFIC TRANSCRIPTIONAL PROFILES

The genomes of higher eukaryotes are riddled with repeat elements, many of which are remnants of transposable elements (TEs) that are no longer able to transpose (Huda, Bowen, Conley, & Jordan, 2011; Lisch & Bennetzen, 2011). The suppression of TEs has been shown to be facilitated by their targeting for heterochromatin formation and co-localization with other heterochromatic regions (Cam et al., 2008; Huda et al., 2011; Iwasaki & Noma, 2012; Iwasaki et al., 2010; Lorenz et al., 2012). *S. pombe* has a genome structure similar to that of higher eukaryotes and it contains large regions of heterochromatin that encompass repetitive sequences, for example *dg-dh* repeats found at the centromeres (Martienssen et al., 2005; T. A. Volpe et al., 2002). In addition, *S. pombe* has Long Terminal Repeat (LTRs) like elements that are silenced and co-localize together in an *Abp1* (a CENP-B homolog) dependent manner (Cam et al., 2008; Lorenz et al., 2012). Consequently, genes in the vicinity of LTRs are also suppressed and upon *Abp1* deletion there is a global change in transcript levels (Lorenz et al., 2012). Herein I identified the co-localization of a large proportion of known LTRs both within and between nuclear chromosomes. Interestingly, the

identity of the LTRs that participated in the spatial clusters at different stages of the cell cycle varied. Therefore, given that the transcript level of LTR associated genes is repressed by LTR silencing and clustering, perhaps the clustering of LTRs is regulated to facilitate the establishment of cell cycle specific transcription profiles.

If this is true, cells are presented with a trade-off between reduced TE transposition and the requirement to express LTR proximal repressed genes (Hollister & Gaut, 2009). The mobility of TEs can be detrimental and may lead to several pathologies, including cancer (Chénais, 2013). However, genes associated with these same TEs may be required for specific cellular processes. This raises interesting questions about the evolutionary selection pressure on the insertion positions of TEs into genomes in relation to the requirement for nearby gene expression. Indeed TEs are known to be non-randomly inserted into genomes (Behrens, Hayles, & Nurse, 2000; Bownes, 1990; Capel, Montero, Martinez-Zapater, & Salinas, 1993; Jjingo, Huda, Gundapuneni, Mariño-Ramírez, & Jordan, 2011). For example, in *S. pombe Tf1* retrotransposon integration is targeted to the 5' ends of open reading frames (Behrens et al., 2000). Further, it is interesting to speculate that given that silencing of these retrotransposons is facilitated by their co-localization and consequential repression of nearby genes, cells may hijack this as a mechanism for the silencing of a broad range of genes. Utilizing such a process would enable the mechanism used to regulate silencing and co-localization of TEs to also regulate the expression level of a broad range of genes. Alternatively, cells may have adapted to the presence of TE by silencing their transposition through a targeted mechanism and as a consequence nearby genes are silenced. Therefore, the transcriptionally repressive affect that the spatial co-localization of TEs has on nearby genes may contribute to the evolutionary selection pressure on TE insertion positions.

In future experiments it would be interesting to further address the role that clustering of repeat elements have in regulating gene expression. For example, the identification of a specific loop between two regions on *S. pombe* chromosome II, whose disappearance in G2 phase is correlated with the up-regulation of an associated Ubiquitin Ligase gene, should be validated. To confirm the apparent fluctuation in the frequency of loop formation q3C could be used in synchronized *S. pombe* cells. The paired analysis of the Ubiquitin Ligase gene transcript level using

quantitative real-time reverse transcriptase PCR (qRT-RT-PCR) would give an indication of changes in transcription. To investigate whether the clustering of the LTRs into *tf* bodies is what results in the high contact frequencies and gene silencing, these experiments could also be performed in an *Abp1* mutant strain. The deletion of *Abp1* results in the disruption of *tf* bodies and therefore should result in a decreased contact frequency and potentially up-regulation of the Ubiquitin Ligase gene. These results would help validate the importance of LTR clustering in the regulation of associated genes.

5.4 SHARING INFORMATION FOR A MUTUAL BENEFIT

One of the largest explosions in organism diversity on this planet was initiated by an event whereby a primitive eukaryotic cell engulfed a bacterial cell, giving rise to an endosymbiont. The most well-known examples of this endosymbiosis that can be seen today are the mitochondrion and chloroplast. Over the course of these endosymbiosis events the majority of genes that used to exist on the bacteria's genome have been transferred and integrated into the host genome (Timmis et al., 2004). This has resulted in an almost exclusive dependence of the host and endosymbiont on each other for survival, and a signalling system between the two compartments is paramount to ensure their coordination (Butow & Avadhani, 2004; Chae et al., 2013; Z. Liu & Butow, 2006).

5.4.1 MT-NDNA INTERACTIONS PARTICIPATE IN THE COORDINATION OF MITOCHONDRIAL AND NUCLEAR PROCESSES

Given the intimate dependence of the host and endosymbiont on each other for survival, it may not be surprising that evidence presented herein supports a role for the formation of DNA-DNA interactions between the mitochondrial and nuclear chromosomes in coordinating these two cellular compartments. Indeed, retrograde signalling for the coordination of nuclear gene transcription in response to mitochondrial requirement is known (Butow & Avadhani, 2004; Chae et al., 2013; Z. Liu & Butow, 2006).

What is the purpose/ advantage of using fragments of mitochondrial DNA as signalling molecules? Perhaps the levels of mitochondrial fragments that are transferred to the nuclear compartment correspond directly with the stability of the

mitochondrial genome or mitochondrial energetic state and fluctuations therein. The importance of this feedback could be argued to be central to the apparent role that mt-nDNA interactions have in regulating the G1-S phase cell cycle checkpoint. DNA replication is an energetically demanding process and, therefore, can only be initiated under conditions when the cell has sufficient energetic resources, which are provided by the mitochondria. However, my work suggests a more active role for the mtDNA fragments transported to the nucleus than merely acting as a proxy for mitochondrial state. The mt-nDNA interactions apparently were targeted to specific DNA origins of replication, potentially facilitating the formation of a chromatin state that promotes efficient initiation of DNA replication.

The formation of metaphase mt-nDNA interactions with genes required for DNA synthesis and highly transcribed genes involved in ribosome structure and function, implicates these interactions in the establishment of transcription following metaphase exit. It is possible that the formation of mt-nDNA interactions during metaphase may reduce the level of chromatin compaction in specific regions. This could contribute to the rapid re-activation of associated genes by: 1) enabling more rapid access to these regions by transcription factors; and/or 2) marking these regions for rapid de-condensation. Alternatively, it is possible that mt-nDNA interactions promote the maintenance of more 'open' regions of chromatin to facilitate the stable binding of TFs to mitotic chromosomes and, therefore, function in bookmarking (Kadauke & Blobel, 2013). Together, the formation of mt-nDNA interactions that may influence metaphase chromatin accessibility, and the identification of specific interactions with highly transcribed and differentially regulated genes within nuclear chromosomes during the G2-M-G1 phase cell cycle transition (discussed above), support a role for the inheritance of spatial genome organization in epigenetic inheritance and bookmarking in *S. pombe* (Blomen & Boonstra, 2011; Kadauke & Blobel, 2013; Sarge & Park-Sarge, 2005).

5.4.2 Mitochondria in neurons

Neurons have a high energy demand that is provided for by the mitochondria and the mitochondrial organelle is increasingly recognized as a signalling platform involved in fundamental events in the formation and plasticity of neuronal circuits (Ankarcrona, Mangialasche, & Winblad, 2010; Cheng, Hou, & Mattson, 2010; Liesa,

Palacín, & Zorzano, 2009; MacAskill et al., 2009; Stowe & Camara, 2009). Furthermore, increasing evidence points to mitochondria having an important role in neurodegenerative disorders with mitochondrial dysfunction having been shown in Parkinson's disease, Alzheimer's disease, Huntington's disease, and amyotrophic lateral sclerosis (Ankarcrona et al., 2010; Moreira et al., 2010). Given the novel roles I have detected for mt-nDNA interactions in the coordination and regulation of cellular processes in the mitochondrial organelle and nuclear compartment, neurons would provide an attractive model for the investigation of these interactions in mammalian cells. Future work to broaden our understanding of the diversity and functional roles of mt-nDNA interactions could involve the investigation of the presence of these interactions during neuronal differentiation or comparing healthy with diseased neurons. Such analysis will provide further insight into the functional role that mt-nDNA interactions have in the genotype to phenotype relationship. A key feature to the precise understanding of the function that mt-nDNA interactions have will require the identification of the mechanism(s) and regulation of mtDNA transfer to the nucleus.

5.4.3 Implications for the co-evolution of other host-parasite interactions

The establishment of an endysymbiont is an extreme example of the co-evolution of an interaction between organisms that influences the genotype – phenotype relationship. Other examples of close relationships between hosts and parasites or commensalism, shaped by evolution, have also been discovered (Kang et al., 2013; D. M. Matthews & Jenks, 2013; Ridaura et al., 2013). For example, the manipulation of the host by intracellular parasites (*e.g. Plasmodium*, *Leishmania*) (Guerfali et al., 2008; Leirião, Rodrigues, C, Albuquerque, S, & Mota, M, 2004), and the promotion of correct gastrointestinal tract development by commensal bacteria (Clemente, Ursell, Parfrey, & Knight, 2012; R. M. Jones et al., 2013).

The ability of non-host occupants to manipulate cellular processes within the host for selfish or mutual benefits necessitates that there is a form of signalling between the two. Given that my results implicate the formation of interactions between DNA molecules stored in separate cellular compartments as coordinator signals, it remains possible that the transfer of genomic fragments from intracellular parasites or

commensal bacteria into host cell nuclei provide a control/communication system. Indeed gene transfer events between different kingdoms of life have been detected (Fitzpatrick, Logue, & Butler, 2008; Gilbert, Schaack, Ii, & Brindley, 2010; B. F. Sun et al., 2013). Moreover, the formation of biofilms has been shown to increase the transfer rate (Madsen, Burmølle, Hansen, & Sørensen, 2012). In future experiments it would be interesting to investigate if interactions between the genomes of commensal bacteria and mammalian chromosomes can be detected. Gastrointestinal tract cells from mice could be cultured in the absence and presence of bacterial cells identified to influence gastrointestinal tract development. Performing paired 3D genome and transcription analysis on the separate and combined mouse and bacterial cell cultures would enable the elucidation of the effects that the cells had on each other. If as part of this analysis, interactions between the bacterial and mouse cell genomes were detected, this would provide evidence for widespread use of the transfer of DNA fragments as signalling molecules for communication, coordination, and control.

5.5 FINAL REMARKS

Over the past decade it has become increasingly apparent that merely understanding the linear arrangement of genes and regulatory elements on chromosomes will not result in an integrated understanding of genome function and the genotype – phenotype translation. Clearly the epigenetic modification of DNA and histone proteins plays a vital role in the non-heritable and increasingly evident heritable, interpretation of the genetic blueprint, influencing an organism's ability to adapt its phenotype to the environment. The 3D organization of genomes in space and time is emerging as a form of 'epigenetic' variation that plays a vital role in the genotype – phenotype relationship. This is emphasized by the work presented here, which provides further evidence for the functional involvement of intra- and inter-chromosomal interactions in the regulation of gene transcription and DNA replication both in bacteria and eukaryotes. However, non-specific interactions also form as a result of random processes and entropic effects that may also be utilized by cells for the regulation of genome structure and function.

The use of synchronized *S. pombe* cell cultures for the investigation of genome organization throughout the cell cycle revealed substantial variation in interactions at each phase of the cell cycle. Further, these cell cycle dynamic interactions played a

role in the establishment and maintenance of coinciding gene expression profiles. These results highlight the potential importance for future studies to utilize synchronized cell populations for functional 3D genome organization studies as asynchronicity likely clouds the interpretation of the data. Incorporating the pairwise analysis of nascent transcription, DNA and chromatin modification data, and other information (*e.g.* nuclear peripheral association data) in synchronized cell populations, will enable the elucidation of an integrated understanding of genome structure and function. However, performing proximity-based ligation methods on cell populations will only ever provide a probabilistic understanding of the 3D structure-function relationship, necessitating the use of single cell methods like loci tagging (*e.g.* FISH) and high resolution microscopy to elucidate the role of individual interactions in single cells. Furthermore, future studies should look into the cause or consequence question related to 3D genome organization to try and gain an insight into how the 3D organization of genomes is maintained and manipulated. Most definitely, an integrated understanding of the 2D, 3D and 4D organization of genomes is going to be essential if we want to understand the genome in a way that will enable us to combat complex diseases effectively.

Finally, having investigated changes in genome organization and gene transcript levels in response to environmental stimuli and during cell cycle progression in bacterial and eukaryotic cells, I have become aware of a number of common themes: 1) Interactions can be ascribed to two main categories; non-specific and functional; 2) Entropic effects appear to govern a number of aspects of spatial genome organization leading to the formation of non-specific interaction. However, the entropy driven formation of non-specific interactions may be exploited by cells, blurring the lines between non-specific random interactions and non-specific functional interactions; 3) The formation of condition specific 3D genome organization appears to make the most significant contribution to the establishment of transcription profiles required for adaptation to changing environments; and 4) The linear clustering and non-random chromosomal distribution of co-regulated genes appears to facilitate spatial co-localization and, therefore, the 3D organization of genomes may impost evolutionary selection pressure on the linear chromosomal distribution of genes.

6 BIBLIOGRAPHY

Adams, K. L., Qiu, Y.-L., Stoutemyer, M., & Palmer, J. D. (2002). Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci U S A*, *99*, 9905–9912.

Adams, K. L., Song, K., Roessler, P. G., Nugent, J. M., Doyle, J. L., Doyle, J. J., & Palmer, J. D. (1999). Intracellular gene transfer in action: Dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13863–13868. 0033598682&partnerID=40&md5=74961f30ed70de162e5251e44f922c02

Alfredsson-Timmins, J., Henningson, F., & Bjerling, P. (2007). The Clr4 methyltransferase determines the subnuclear localization of the mating-type region in fission yeast. *Journal of Cell Science*, *120*(11), 1935–1943. 34250792759&partnerID=40&md5=1dc70402a63772ec5f71b3f84177cd28

Alfredsson-Timmins, J., Kristell, C., Henningson, F., Lyckman, S., & Bjerling, P. (2009). Reorganization of chromatin is an early response to nitrogen starvation in Schizosaccharomyces pombe. *Chromosoma*, *118*(1), 99–112. 58349120998&partnerID=40&md5=6ea8b73621d018dead3e94654bb2d445

Allan, J., Cowling, G. J., Harborne, N., Cattini, P., & Gould, H. (1981). Regulation of the Higher-order Structure of Chromatin by Histones H1 and H5. *The Journal of Cell Biology*, *90*, 279–288.

Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P., & Cranston, G. (1995). Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes and Development*, *9*(2), 218–233. 0028873187&partnerID=40&md5=aa336f04f7eb4efee352ef38bda22e05

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., … Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457–465. 0019423856&partnerID=40&md5=01ae1917cd8d681941cf99086bd81bcc

Ankarcrona, M., Mangialasche, F., & Winblad, B. (2010). Rethinking Alzheimer's disease therapy: Are mitochondria the key? *Journal of Alzheimer's Disease*, *20*(SUPPL.2), S579–S590.

Ansari, A., & Hampsey, M. (2005). A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*, *19*(24), 2969–2978.

Arnold, I., Wagner-Ecker, M., Ansorge, W., & Langer, T. (2006). Evidence for a novel mitochondria-to-nucleus signalling pathway in respiring cells lacking i-AAA protease and the ABC-transporter Mdl1. *Gene*, *367*, 74–88.

Aula, P., & Saksela, E. (1972). Comparison of areas of quinacrine mustard fluorescence and modified Giemsa staining in human metaphase chromosomes. *Experimental Cell Research*, *71*(1), 161–167.

Aves, S. J., Durkacz, B. W., Carr, A., & Nurse, P. (1985). Cloning, sequencing and transcriptional control of the Schizosaccharomyces pombe cdc10 "start" gene. *EMBO Journal*, *4*(2), 457–463.

Azam, T. A., & Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *Journal of Biological Chemistry*, *274*(46), 33105–33113.

Azam, T. A., Iwata, A., Nishimura, A., Ueda, S., & Ishihama, A. (1999). Growth Phase-Dependent Variation in Protein Composition of the Escherichia coli Nucleoid. *J Bacteriol*, *181*, 6361–6370.

Bach, T., Krekling, M. A., & Skarstad, K. (2003). Excess SeqA prolongs sequestration of oriC and delays nucleoid segregation and cell division. *EMBO Journal*, *22*(2), 315–323.

Baddeley, D., Batram, C., Weiland, Y., Cremer, C., & Birk, U. J. (2007). Nanostructure analysis using spatially modulated illumination microscopy. *Nat Protoc*, *2*(10), 2640–2646. doi:nprot.2007.399 [pii] 10.1038/nprot.2007.399

Bai, L., & Morozov, A. V. (2010). Gene regulation by nucleosome positioning. *Trends in Genetics*, *26*(11), 476–483.

Bartkuhn, M., & Renkawitz, R. (2008). Long range chromatin interactions involved in gene regulation. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1783*(11), 2161–2166.

Bartlett, J., Blagojevic, J., Carter, D., Eskiw, C., Fromaget, M., Job, C., … Cook, P. R. (2006). Specialized transcription factories. *Biochem Soc Symp*, (73), 67–75.

Bates, D., & Kleckner, N. (2005). Chromosome and replisome dynamics in E. coli: Loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, *121*(6), 899–911. 20444390654&partnerID=40&md5=e61affdc9533f57c084fec4765fbac22

Bauer, D. L. V, Marie, R., Rasmussen, K. H., Kristensen, A., & Mir, K. U. (2012). DNA catenation maintains structure of human metaphase chromosomes. *Nucleic Acids Research*, *40*(22), 11428–11434.

Behrens, R., Hayles, J., & Nurse, P. (2000). Fission yeast retrotransposon Tf1 integration is targeted to 5' ends of open reading frames. *Nucleic Acids Research*, *28*(23), 4709–4716.

Bell, S. P. (1995). Eukaryotic replicators and associated protein complexes. *Current Opinion in Genetics and Development*, *5*(2), 162–167. 0028949675&partnerID=40&md5=105e4d1ff4a593658533c496b89e8305

Bell, S. P., & Dutta, A. (2002). DNA replication in eukaryotic cells. *Annual Review of Biochemistry*, *71*, 333–374.

Berezney, R. (2002). Regulating the mammalian genome: the role of nuclear architecture. *Advances in Enzyme Regulation*, *42*, 39–52.

Berger, A. B., Cabal, G. G., Fabre, E., Duong, T., Buc, H., Nehrbass, U., … Zimmer, C. (2008). High-resolution statistical mapping reveals gene territories in live yeast. *Nature Methods*, *5*(12), 1031–1037.

Bernhardt, T. G., & De Boer, P. A. J. (2005). SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over chromosomes in E. coli. *Molecular Cell*, *18*(5), 555–564.

Bielinsky, A. K., & Gerbi, S. A. (2001). Where it all starts: Eukaryotic origins of DNA replication. *Journal of Cell Science*, *114*(4), 643–651.

Binenbaum, Z., Parola, A. H., Zaritsky, A., & Fishov, I. (1999). Transcription- and translation-dependent changes in membrane dynamics in bacteria: Testing the transertion model for domain formation. *Molecular Microbiology*, *32*(6), 1173–1182.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., … De Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816.

Bisht, K. K., Arora, S., Ahmed, S., & Singh, J. (2008). Role of heterochromatin in suppressing subtelomeric recombination in fission yeast. *Yeast*, *25*(8), 537–548.

Blank, H. M., Li, C., Mueller, J. E., Bogomolnaya, L. M., Bryk, M., & Polymenis, M. (2008). An increase in mitochondrial DNA promotes nuclear DNA replication in yeast. *PLoS Genetics*, *4*(4), e1000047. doi:10.1371/journal.pgen.1000047

Blattner, F. R., Plunkett 3rd, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., … Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science*, *277*(5331), 1453–1474.

Blomen, V. A., & Boonstra, J. (2011). Stable transmission of reversible modifications: Maintenance of epigenetic information through the cell cycle. *Cellular and Molecular Life Sciences*, *68*(1), 27–44.

Boccard, F., Esnault, E., & Valens, M. (2005). Spatial arrangement and macrodomain organization of bacterial chromosomes. *Molecular Microbiology*, *57*(1), 9–16.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., … Barillot, E. (2012). Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, *28*(3), 423–425.

Bohn, M., Heermann, D., & van Driel, R. (2007). Random loop model for long polymers. *Physical Review E*, *76*(5), 051805. doi:10.1103/PhysRevE.76.051805

Bohrmann, B., Haider, M., & Kellenberger, E. (1993). Concentration evaluation of chromatin in unstained resin-embedded sections by means of low-dose ratio-contrast imaging in STEM. *Ultramicroscopy*, *49*(1-4), 235–251. 0027542926&partnerID=40&md5=65bdd97cde4b0280ec5366d166941c40

Bolstad, B. M., Irizarry, R. A., Åstrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, *19*(2), 185–193.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., … Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology.*, *3*(5).

Botta, M., Haider, S., Leung, I. X. Y., Lio, P., & Mozziconacci, J. (2010). CTCF-mediated functional chromatin interactome in pluripotent cells. *Molecular Systems Biology*, *6*.

Bouffartigues, E., Buckle, M., Badaut, C., Travers, A., & Rimsky, S. (2007). H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nature Structural and Molecular Biology*, *14*(5), 441–448.

Boulanger, S. C., Belcher, S. M., Schmidt, U., Dib-Hajj, S. D., & Schmidt, T. et al. (1995). Studies of point mutants define three essential paired nucleotides in the domain 5 substructure of a group II intron. *Molecular and Cellular Biology*, *15*, 4479–4488.

Bownes, M. (1990). Preferential insertion of P elements into genes expressed in the germ-line of Drosophila melanogaster. *Molecular and General Genetics*, *222*(2-3), 457–460.

Boyle, S., Gilchrist, S., Bridger, J. M., Mahy, N. L., Ellis, J. A., & Bickmore, W. A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human Molecular Genetics*, *10*(3), 211–219.

Bremer, H., & Dennis, P. (1996). Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli and Salmonella typhimurium. *Neidhardt, F. (ed.). Washington, DC: American Society for Microbiology Press*, 1553.

Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V., & Schuster, W. (1993). The mitochondrial genome on its way to the nucleus: Different stages of gene transfer in higher plants. *FEBS Letters*, *325*(1-2), 140–145.

Brickner, D. G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P. C., Widom, J., & Brickner, J. H. (2007). H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biology*, *5*(4), 704–716.

Brickner, J. H. (2009). Transcriptional memory at the nuclear periphery. *Current Opinion in Cell Biology*, *21*(1), 127–133.

Brickner, J. H. (2010). Transcriptional Memory: Staying in the Loop. *Current Biology*, *20*(1), R20–R21.

Bullerwell, C. E., Leigh, J., Forget, L., & Lang, B. F. (2003). A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Research*, *31*(2), 759–768.

Butow, R. A., & Avadhani, N. G. (2004). Mitochondrial signaling: The retrograde response. *Molecular Cell*, *14*(1), 1–15.

Bystricky, K., Heun, P., Gehlen, L., Langowski, J., & Gasser, S. M. (2004). Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci U S A*, *101*(47), 16495–16500.

Cabal, G. G., Genovesio, A., Rodriguez-Navarro, S., Zimmer, C., Gadal, O., Lesne, A., … Nehrbass, U. (2006). SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, *441*(7094), 770–773.

Cabrera, J. E., Cagliero, C., Quan, S., Squires, C. L., & Ding, J. J. (2009). Active transcription of rRNA operons condenses the nucleoid in Escherichia coli: Examining the effect of transcription on nucleoid structure in the absence of transertion. *Journal of Bacteriology*, *191*(13), 4180–4185.

Cabrera, J. E., & Jin, D. J. (2003). The distribution of RNA polymerase in Escherichia coli is dynamic and sensitive to environmental cues. *Molecular Microbiology*, *50*(5), 1493–1505.

Cabrera, J. E., & Jin, D. J. (2006). Active transcription of rRNA operons is a driving force for the distribution of RNA polymerase in bacteria: effect of extrachromosomal copies of rrnB on the in vivo localization of RNA polymerase. *J Bacteriol*, *188*(11), 4007–4014. doi:188/11/4007 [pii] 10.1128/JB.01893-05

Cagliero, C., Grand, R. S., Jones, M. B., Jin, D. J., & O'Sullivan, J. M. (2013). Genome conformation capture reveals that the Escherichia coli chromosome is organized by replication and transcription. *Nucleic Acids Research*, *41*(12), 6058–6071.

Cagliero, C., & Jin, D. J. (2013). Dissociation and re-association of RNA polymerase with DNA during osmotic stress response in Escherichia coli. *Nucleic Acids Research*, *41*(1), 315–326.

Cai, S., Lee, C. C., & Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*, *38*(11), 1278–1288.

Cam, H. P., Noma, K., Ebina, H., Levin, H. L., & Grewal, S. I. S. (2008). Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature*, *451*(7177), 431–436. doi:10.1038/nature06499

Campbell, C. L., & Thorsness, P. E. (1998). Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. *Journal of Cell Science*, *111*(16), 2455–2464.

Campos, E. I., & Reinberg, D. (2009). Histones: Annotating chromatin. *Annual Review of Genetics*.

Capel, J., Montero, L. M., Martinez-Zapater, J. M., & Salinas, J. (1993). Non-random distribution of transposable elements in the nuclear genome of plants. *Nucleic Acids Research*, *21*(10), 2369–2373.

Caron, H., Van Schaik, B., der Mee, M., Baas, F., Riggins, G., Van Sluis, P., … Versteeg, R. (2001). The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. *Science*, *291*(5507), 1289–1293.

Carroll, J. S., & Brown, M. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, *122*(1), 33–43.

Carter, D. R. F., Eskiw, C., & Cook, P. R. (2008). Transcription factories. *Biochemical Society Transactions*, *36*(4), 585–589.

Cashel, M., Gentry, D., Hernandez, V., & Vinella, D. (1996). The stringent response. *In Neidhardt, F. C. and Curtiss, R. (eds.), Escherichia Coli and Salmonella : Cellular and Molecular Biology. 2nd Ed. ASM Press, Washington, D.C.*, *2nd ed*.

Casjens, S. (1998). The diverse and dynamic structure of bacterial genomes. *Annual Review of Genetics*, *32*, 339–377.

Casjens, S., Palmer, N., Van Vugt, R., Huang, W. M., Stevenson, B., Rosa, P., … Fraser, C. M. (2000). A bacterial genome in flux: The twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi. *Molecular Microbiology*, *35*(3), 490–516.

Cavalli, G., & Misteli, T. (2013). Functional implications of genome topology. *Nature Structural and Molecular Biology*, *20*(3), 290–299. 84875200698&partnerID=40&md5=7f4533f36ec667a8d25614f14c59fb5a

Chae, S., Ahn, B. Y., Byun, K., Cho, Y. M., Yu, M. H., Lee, B., … Park, K. S. (2013). A systems approach for decoding mitochondrial retrograde signaling pathways. *Science Signaling*, *6*(264).

Chatre, L., & Ricchetti, M. (2011). Nuclear mitochondrial DNA activates replication in Saccharomyces cerevisiae. *PloS One*, *6*(3), e17235. doi:10.1371/journal.pone.0017235

Chatre, L., & Ricchetti, M. (2013). Prevalent coordination of mitochondrial DNA transcription and initiation of replication with the cell cycle. *Nucleic Acids Research*, *41*(5), 3068–78. doi:10.1093/nar/gkt015

Chen, X. J., Butow, R. a, & Xin, J. C. (2005). The organization and inheritance of the mitochondrial genome. *Nature Reviews Genetics*, *6*(11), 815–825. doi:10.1038/nrg1708

Chénais, B. (2013). Transposable elements and human cancer: A causal relationship? *Biochimica et Biophysica Acta - Reviews on Cancer*, *1835*(1), 28–35.

Cheng, A., Hou, Y., & Mattson, M. P. (2010). Mitochondria and neuroplasticity. *ASN Neuro*, *2*(5), e00045.

Chien, R., Zeng, W., Kawauchi, S., Bender, M., Santos, R., Gregson, H., … Yokomori, K. (2011). Cohesin mediates chromatin interactions that regulate mammalian β-globin expression. *The Journal of Biological Chemistry*, *286*(20), 17870–8. doi:10.1074/jbc.M110.207365

Chikashige, Y., Ding, D. Q., Imai, Y., Yamamoto, M., Haraguchi, T., & Hiraoka, Y. (1997). Meiotic nuclear reorganization: Switching the position of centromeres and telomeres in the fission yeast Schizosaccharomyces pombe. *EMBO Journal*, *16*(1), 193–202.

Chiron, S., Gaisne, M., Guillou, E., Belenguer, P., Clark-Walker, G. D., & Bonnefoy, N. (2007). Studying mitochondria in an attractive model: Schizosaccharomyces pombe. *Methods in Molecular Biology (Clifton, N.J.)*, *372*(1), 91–105. doi:10.1007/978-1-59745-365-3_7

Cho, H., McManus, H. R., Dove, S. L., & Bernhardt, T. G. (2011). Nucleoid occlusion factor SlmA is a DNA-activated FtsZ polymerization antagonist. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3773–3778.

Chu, Z., Li, J., Eshaghi, M., Karuturi, R. K. M., Lin, K., & Liu, J. (2007). Adaptive expression responses in the Pol-gamma null strain of S. pombe depleted of mitochondrial genome. *BMC Genomics*, *8*(323). doi:10.1186/1471-2164-8-323

Cipak, L., Spirek, M., & Gregan, J. (2008). Sister chromatids caught in the cohesin trap. *Nature Structural and Molecular Biology*, *15*(9), 899–900.

Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell*, *148*(6), 1258–70. doi:10.1016/j.cell.2012.01.035

Clyne, R. K., & Kelly, T. J. (1995). Genetic analysis of an ARS element from the fission yeast Schizosaccharomyces pombe. *EMBO Journal*, *14*(24), 6348–6357. 0029585836&partnerID=40&md5=7be8e1290004218b2bf098c9aeb60aeb

Cook, P. R. (1999). The organization of replication and transcription. *Science*, *284*(5421), 1790–1795.

Cook, P. R., & Marenduzzo, D. (2009). Entropic organization of interphase chromosomes. *The Journal of Cell Biology*, *186*(6), 825–34. doi:10.1083/jcb.200903083

Coudreuse, D., & Nurse, P. (2010). Driving the cell cycle with a minimal CDK control network. *Nature*, *468*(7327), 1074–9. doi:10.1038/nature09543

Crawford, G. E., Davis, S., Scacheri, P. C., Renaud, G., Halawi, M. J., Erdos, M. R., … Collins, F. S. (2006). DNase-chip: A high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature Methods*, *3*(7), 503–509.

Crawford, G. E., Holt, I. E., Whittle, J., Webb, B. D., Tai, D., Davis, S., Collins, F. S. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Research*, *16*(1), 123–131.

Cremer, C., Rappold, G., Gray, J. W., Müller, C. R., & Ropers, H. H. (1984). Preparative dual-beam sorting of the human Y chromosome and in situ hybridization of cloned DNA probes. *Cytometry*, *5*(6), 572–579.

Cremer, T., & Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, *2*(4), 292–301.

Cremer, T., Lichter, P., Borden, J., Ward, D. C., & Manuelidis, L. (1988). Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. *Human Genetics*, *80*(3), 235–246.

Crider, D. G., García-Rodríguez, L. J., Srivastava, P., Peraza-Reyes, L., Upadhyaya, K., Boldogh, I. R., & Pon, L. a. (2012). Rad53 is essential for a mitochondrial DNA inheritance checkpoint regulating G1 to S progression. *The Journal of Cell Biology*, *198*(5), 793–8. doi:10.1083/jcb.201205193

Cullen, K. E., Kladde, M. P., & Seyfred, M. A. (1993). Interaction between Transcription Regulatory Regions of Prolactin Chromatin. *Science*, *261*(5118), 203–206. Retrieved from <Go to ISI>://A1993LL59600038

Cutler, R. G., & Evans, J. E. (1966). Synchronization of Bacteria by a Stationary-Phase Method. *Journal of Bacteriology*, *91*(2), 469–476.

Dai, Z., Dai, X., Xiang, Q., Feng, J., Deng, Y., Wang, J., & He, C. (2009). Transcriptional interaction-assisted identification of dynamic nucleosome positioning. *BMC Bioinformatics*, *10*(SUPPL. 1).

Dame, R. T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology*, *56*(4), 858–870.

Dame, R. T., Kalmykowa, O. J., & Grainger, D. C. (2011). Chromosomal macrodomains and associated proteins: Implications for DNA organization and replication in gram negative bacteria. *PLoS Genetics*, *7*(6).

Dandanell, G., Valentin-Hansen, P., Larsen, J. E., & Hammer, K. (1987). Long-range cooperativity between gene regulatory sequences in a prokaryote. *Nature*, *325*(6107), 823–826. doi:10.1038/325823a0

De Nooijer, S., Wellink, J., Mulder, B., & Bisseling, T. (2009). Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. *Nucleic Acids Research*, *37*(11), 3558–3568.

De Wit, E., Bouwman, B. M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M. M., … de Laat, W. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*, *501*(7466), 227–31. doi:10.1038/nature12420

Dehghani, H., Dellaire, G., & Bazett-Jones, D. P. (2005). Organization of chromatin in the interphase mammalian cell. *Micron*, *36*(2), 95–108.

Dekker, J. (2006). The three “C” s of chromosome conformation capture: Controls, controls, controls. *Nature Methods*, *3*(1), 17–21.

Dekker, J., Marti-Renom, M., & Mirny, L. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics*, *14*(6), 390–403. doi:10.1038/nrg3454

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, *295*(5558), 1306–1311.

Denholtz, M., & Plath, K. (2012). Pluripotency in 3D: Genome organization in pluripotent cells. *Current Opinion in Cell Biology*, *24*(6), 793–801.

Dillon, S. C., & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews. Microbiology*, *8*(3), 185–95. doi:10.1038/nrmicro2261

Djupedal, I., & Ekwall, K. (2009). Epigenetics: Heterochromatin meets RNAi. *Cell Research*, *19*(3), 282–295.

Djupedal, I., Kos-Braun, I. C., Mosher, R. A., Söderholm, N., Simmer, F., Hardcastle, T. J., Ekwall, K. (2009). Analysis of small RNA in fission yeast; Centromeric siRNAs are potentially generated through a structured RNA. *EMBO Journal*, *28*(24), 3832–3844.

Dobi, K. C., & Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. *Mol Cell Biol*, *27*(15), 5575–5586. doi:MCB.00459-07 [pii] 10.1128/MCB.00459-07

Doolittle, W. F. (2013). Is junk DNA bunk? A critique of ENCODE. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(14), 5294–5300.

Dorier, J., & Stasiak, A. (2009). Topological origins of chromosomal territories. *Nucleic Acids Research*, *37*(19), 6316–22. doi:10.1093/nar/gkp702

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., & Richmond, T. J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, *306*(5701), 1571–1573.

Dorman, C. J. (2013). Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nature Reviews Microbiology*, *11*(5l), 349–355. doi:10.1038/nrmicro3007

Dorman, C. J., & Kane, K. A. (2009). DNA bridging and antibridging: A role for bacterial nucleoid-associated proteins in regulating the expression of laterally acquired genes. *FEMS Microbiology Reviews*, *33*(3), 587–592.

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., … Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*, *16*(10), 1299–1309. doi:10.1101/gr.5571506

Dostie, J., Zhan, Y., & Dekker, J. (2007). Chromosome conformation capture carbon copy technology. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, *Chapter 21*.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., … Noble, W. S. (2010). A three-dimensional model of the yeast genome. *Nature*, *465*(7296), 363–367.

Dubey, D. D., Kim, S. M., Todorov, I. T., & Huberman, J. A. (1996). Large, complex modular structure of a fission yeast DNA replication origin. *Current Biology*, *6*(4), 467–473. 6

Dubey, D. D., Zhu, J., Carlson, D. L., Sharma, K., & Huberman, J. A. (1994). Three ARS elements contribute to the ura4 replication origin region in the fission yeast, Schizosaccharomyces pombe. *EMBO Journal*, *13*(15), 3638–3647.

Dunn, T. M., Hahn, S., Ogden, S., & Schleif, R. F. (1984). An operator at -280 base pairs that is required for repression of araBAD operon promoter: Addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(16 I), 5017–5020.

Dupaigne, P., Tonthat, N., Espéli, O., Whitfill, T., Boccard, F., & Schumacher, M. (2012). Molecular Basis for a Protein-Mediated DNA-Bridging Mechanism that Functions in Condensation of the E. coli Chromosome. *Molecular Cell*, *48*(4), 560–571.

Durfee, T., Hansen, A. M., Zhi, H., Blattner, F. R., & Ding, J. J. (2008). Transcription profiling of the stringent response in Escherichia coli. *Journal of Bacteriology*, *190*(3), 1084–1096.

Dworkin, J., & Losick, R. (2002). Does RNA polymerase help drive chromosome segregation in bacteria? *Proceedings of the National Academy of Sciences of the United States of America*, *99*(22), 14089–14094.

Dworsky, P., & Schaechter, M. (1973). Effect of rifampin on the structure and membrane attachment of the nucleoid of Escherichia coli. *Journal of Bacteriology*, *116*(3), 1364–1374.

Edenberg, H. J., & Huberman, J. A. (1975). Eukaryotic chromosome replication. *Annu Rev Genet*, *9*, 245–284.

Elmore, S., Müller, M., Vischer, N., Odijk, T., & Woldringh, C. L. (2005). Single-particle tracking of oriC-GFP fluorescent spots during chromosome segregation in Escherichia coli. *Journal of Structural Biology*, *151*(3), 275–287.

Eskiw, C. H., Cope, N. F., Clay, I., Schoenfelder, S., Nagano, T., & Fraser, P. (2010). Transcription factories and nuclear organization of the genome. *Cold Spring Harbor Symposia on Quantitative Biology*, *75*, 501–506.

Espéli, O., & Boccard, F. (2006). Organization of the Escherichia coli chromosome into macrodomains and its possible functional implications. *Journal of Structural Biology*, *156*(2), 304–310.

Espéli, O., Borne, R., Dupaigne, P., Thiel, A., Gigant, E., Mercier, R., & Boccard, F. (2012). A MatP-divisome interaction coordinates chromosome segregation with cell division in E. coli. *EMBO Journal*, *31*(14), 3198–3211.

Espeli, O., Mercier, R., & Boccard, F. (2008). DNA dynamics vary according to macrodomain topography in the E. coli chromosome. *Molecular Microbiology*, *68*(6), 1418–1427.

Essers, J., Van Cappellen, W. A., Theil, A. F., Van Drunen, E., Jaspers, N. G. J., Hoeijmakers, J. H. J., … Kanaar, R. (2005). Dynamics of relative chromosome position during the cell cycle. *Molecular Biology of the Cell*, *16*(2), 769–775.

Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B., & Fairhead, C. (2005). Comparative genomics in hemiascomycete yeasts: Evolution of sex, silencing, and subtelomeres. *Molecular Biology and Evolution*, *22*(4), 856–873.

Falcon, S., & Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics*, *23*(2), 257–258.

Fan, J. B., Rochet, M., Gaillardin, C., & Smith, C. L. (1992). Triploid meiosis and aneuploidy in Schizosaccharomyces pombe: An unstable aneuploid disomic for chromosome III. *Nucleic Acids Research*, *20*(22), 5943–5945.

Fang, F. C., & Rimsky, S. (2008). New insights into transcriptional regulation by H-NS. *Current Opinion in Microbiology*, *11*(2), 113–120.

Fantes, P. A., & Nurse, P. (1978). Control of the timing of cell division in the fission yeast, *115*, 317–329.

Faraday, D. B. F., Underwood, B. O., Jaumin, F., Sensoy, G., Erichsen, E., & Johnsgaard, A. (1994). Synchronization studies on Schizosaccharomyces pombe. In *Institution of Chemical Engineers Symposium Series* (pp. 154–156).

Farcas, A. M., Uluocak, P., Helmhart, W., & Nasmyth, K. (2011). Cohesin's Concatenation of Sister DNAs Maintains Their Intertwining. *Molecular Cell*, *44*(1), 97–107.

Farrelly, F., & Butow, R. A. (1983). Rearranged mitochondrial genes in the yeast nuclear genome. *Nature*, *301*, 296–301.

Ferullo, D. J., & Lovett, S. T. (2008). The stringent response and cell cycle arrest in Escherichia coli. *PLoS Genetics*, *4*(12), e1000300. doi:10.1371/journal.pgen.1000300

Figge, R. M., Divakaruni, A. V, & Gober, J. W. (2004). MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in Caulobacter crescentus. *Molecular Microbiology*, *51*(5), 1321–1332.

Fitzpatrick, D. A., Logue, M. E., & Butler, G. (2008). Evidence of recent interkingdom horizontal gene transfer between bacteria and Candida parapsilosis. *BMC Evolutionary Biology*, *8*(1).

Follows, G. A., Dhami, P., Göttgens, B., Bruce, A. W., Campbell, P. J., Dillon, S. C., … Green, A. R. (2006). Identifying gene regulatory elements by genomic microarray mapping of DNaseI hypersensitive sites. *Genome Research*, *16*(10), 1310–1319.

Forsburg, S. L., & Nurse, P. (1991). Cell cycle regulation in the yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Annual Review of Cell Biology*, *7*, 227–256.

Foury, F., Roganti, T., Lecrenier, N., & Purnelle, B. (1998). The complete sequence of the mitochondrial genome of Sacchromyces cerevisiae. *FEBS Lett*, *440*, 325–331.

Frenkiel-Krispin, D., Ben-Avraham, I., Englander, J., Shimoni, E., Wolf, S. G., & Minsky, A. (2004). Nucleoid restructuring in stationary-state bacteria. *Mol Microbiol*, *51*(2), 395–405.

Fritsche, M., & Heermann, D. W. (2011). Confinement driven spatial organization of semiflexible ring polymers: Implications for biopolymer packaging. *Soft Matter*, *7*(15), 6906–6913.

Fromm, G., & Bulger, M. (2009). A spectrum of gene regulatory phenomena at mammalian β-globin gene loci. *Biochemistry and Cell Biology*, *87*(5), 781–790.

Fujita, N., & Wade, P. A. (2004). Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. *Methods*, *33*(1), 81–85. doi:10.1016/j.ymeth.2003.10.022 S1046202303002974 [pii]

Fukui, K. (2009). Structural analyses of chromosomes and their constituent proteins. *Cytogenetic and Genome Research*, *124*(3-4), 215–227.

Fullwood, M. J., Han, Y., Wei, C. L., Ruan, X., & Ruan, Y. (2010). Chromatin interaction analysis using paired-end tag sequencing. *Current Protocols in Molecular Biology*, (SUPPL. 89), 21.15.1–21.15.25.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., … Ruan, Y. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, *462*(7269), 58–64.

Fullwood, M. J., & Ruan, Y. (2009). ChIP-based methods for the identification of long-range chromatin interactions. *Journal of Cellular Biochemistry*, *107*(1), 30–39.

Fullwood, M. J., Wei, C. L., Liu, E. T., & Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research*, *19*(4), 521–532.

Funabiki, H., Hagan, I., Uzawa, S., & Yanagida, M. (1993). Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *Journal of Cell Biology*, *121*(5), 961–976.

Gehlen, L. R., Gruenert, G., Jones, M. B., Rodley, C. D., Langowski, J., & O'Sullivan, J. M. (2012). Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus (United States)*, *3*(4), 13.

Gelato, K. A., & Fischle, W. (2008). Role of histone modifications in defining chromatin structure and function. *Biological Chemistry*, *389*(4), 353–363.

Gerlich, D., Beaudouin, J., Kalbfuss, B., Daigle, N., Eils, R., & Ellenberg, J. (2003). Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell*, *112*(6), 751–764.

Gilbert, C., Schaack, S., Ii, J. K. P., & Brindley, P. J. (2010). NIH Public Access. *Nature*, *464*(7293), 1347–1350. doi:10.1038/nature08939.A

Gilli, G., & Gilli, P. (2009). The nature of the Hydrogen bond: Outline of a Comprehensive Hydrogen Bond Theory International Union of Crystallography Monographs on Crystallography. *New York: Oxford University Press. 304.*, *23*.

Gitai, Z., Dye, N. A., Reisenauer, A., Wachi, M., & Shapiro, L. (2005). MreB actin-mediated segregation of a specific region of a bacterial chromosome. *Cell*, *120*(3), 329–341.

Gitai, Z., Dye, N., & Shapiro, L. (2004). An actin-like gene can determine cell polarity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(23), 8643–8648.

Gitai, Z., Thanbichler, M., & Shapiro, L. (2005). The choreographed dynamics of bacterial chromosomes. *Trends in Microbiology*, *13*(5), 221–228.

Goetze, S., Mateos-Langerak, J., Gierman, H. J., De Leeuw, W., Giromus, O., Indemans, M. H. G., … Van Driel, R. (2007). The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Molecular and Cellular Biology*, *27*(12), 4475–4487.

Gómez, E. B., & Forsburg, S. L. (2004). Analysis of the fission yeast Schizosaccharomyces pombe cell cycle. *Methods in Molecular Biology (Clifton, N.J.)*, *241*, 93–111.

Göndör, A., & Ohlsson, R. (2009). Chromosome crosstalk in three dimensions. *Nature*, *461*(7261), 212–217.

Gondor, A., Rougier, C., & Ohlsson, R. (2008). High-resolution circular chromosome conformation capture assay. *Nature Protocols*, *3*(2), 303–313.

Grainger, D. C., Hurd, D., Goldberg, M. D., & Busby, S. J. W. (2006). Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic Acids Research*, *34*(16), 4642–4652.

Grand, R. S., Gehlen, L. R., & O'Sullivan, J. M. (2011). Methods for the investigation of chromosome organization. In K. V Urbano (Ed.), *Advances in Genetics Research* (Vol. 5, pp. 111–129). NOVA.

Greenfeder, S. A., & Newlon, C. S. (1992). A replication map of a 61-kb circular derivative of Saccharomyces cerevisiae chromosome III. *Molecular Biology of the Cell*, *3*(9), 999–1013.

Grond, C. J., Derksen, J., & Brakenhoff, G. J. (1982). The banding pattern of the segment 46A-48C in Drosophila hydei polytene chromosomes as studied by confocal scanning light microscopy (CSLM). *Experimental Cell Research*, *138*(2), 458–466.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., … Van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, *453*(7197), 948–951.

Guerfali, F. Z., Laouini, D., Guizani-Tabbane, L., Ottones, F., Ben-Aissa, K., Benkahla, A., Dellagi, K.. (2008). Simultaneous gene expression profiling in human macrophages infected with Leishmania major parasites using SAGE. *BMC Genomics*, *9*.

Haapala, O., & Nokkala, S. (1982). Structure of human metaphase chromosomes. *Hereditas*, *96*(2), 215–228.

Haber, J. E., Thorburn, P. C., & Rogers, D. (1984). Meiotic and mitotic behavior of dicentric chromosomes in Saccharomyces cerevisiae. *Genetics*, *106*(2), 185–205.

Hahn, S., Dunn, T., & Schleif, R. (1984). Upstream repression and CRP stimulation of the Escherichia coli l-arabinose operon. *Journal of Molecular Biology*, *180*(1), 61–72.

Hahn, S., Hendrickson, W., & Schleif, R. (1986). Transcription of Escherichia coli ara in vitro. The cyclic AMP receptor protein requirement for PBAD induction that depends on the presence and orientation of the araO2 site. *Journal of Molecular Biology*, *188*(3), 355–367.

Hakimi, M. A., Bochar, D. A., Schmiesing, J. A., Dong, Y., Barak, O. G., Speicher, D. W., … Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature*, *418*(6901), 994–998.

Hall, I. M., Noma, K. I., & Grewal, S. I. S. (2003). RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 193–198.

Hall, I. M., Shankaranarayana, G. D., Noma, K. I., Ayoub, N., Cohen, A., & Grewal, S. I. S. (2002). Establishment and maintenance of a heterochromatin domain. *Science*, *297*(5590), 2232–2237.

Halme, A., Bumgarner, S., Styles, C., & Fink, G. R. (2004). Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell*, *116*(3), 405–415.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., … Wei, C. L. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics*, *43*(7), 630–638.

Happel, N., & Doenecke, D. (2009). Histone H1 and its isoforms: Contribution to chromatin structure and function. *Gene*, *431*(1-2), 1–12.

Harrison, C. J., Allen, T. D., Britch, M., & Harris, R. (1982). High-resolution scanning electron microscopy of human metaphase chromosomes. *Journal of Cell Science*, *Vol. 56*, 409–422.

Hazkani-Covo, E., Zeller, R. M., & Martin, W. (2010). Molecular Poltergeists: Mitochondrial DNA Copies in Sequenced Nuclear Genomes. *PLoS Genetics*, *6*, e1000834.

Heard, E., & Bickmore, W. (2007). The ins and outs of gene regulation and chromosome territory organisation. *Current Opinion in Cell Biology*, *19*(3), 311–316.

Heichinger, C., Penkett, C. J., Bahler, J., & Nurse, P. (2006). Genome-wide characterization of fission yeast DNA replication origins. *EMBO Journal*, *25*(21), 5171–5179.

Heitz, E. (1929). Heterochromatin, Chromocentren, Chromomeren. *Ber. Deutsch Bot. Ges*, *47*, 275–284.

Heun, P., Laroche, T., Shimada, K., Furrer, P., & Gasser, S. M. (2001). Chromosome dynamics in the yeast interphase nucleus. *Science*, *294*(5549), 2181–2186.

Hildenbrand, G., Rapp, A., Spori, U., Wagner, C., Cremer, C., & Hausmann, M. (2005). Nano-sizing of specific gene domains in intact human cell nuclei by spatially modulated illumination light microscopy. *Biophys J*, *88*(6), 4312–4318. doi:S0006-3495(05)73480-9 [pii] 10.1529/biophysj.104.056796

Hinnebusch, J., & Tilly, K. (1993). Linear plasmids and chromosomes in bacteria. *Molecular Microbiology*, *10*(5), 917–922. 0027738415&partnerID=40&md5=7024c2e70d73c1d63f4899cf56fcd627

Hirano, T., Hiraoka, Y., & Yanagida, M. (1988). A temperature-sensitive mutation of the Schizosaccharomyces pombe gene nuc2+ that encodes a nuclear scaffold-like protein blocks spindle elongation in mitotic anaphase. *Journal of Cell Biology*, *106*(4), 1171–1183.

Hollister, J. D., & Gaut, B. S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, *19*(8), 1419–1428.

Horike, S. I., Cai, S., Miyano, M., Cheng, J. F., & Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nature Genetics*, *37*(1), 31–40.

Hoskins, G. C. (1968). Sensitivity of micrurgically removed chromosomal spindle fibres to enzyme disruption. *Nature*, *217*(5130), 748–750.

Huang, B., Wang, W., Bates, M., & Zhuang, X. (2008). Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, *319*(5864), 810–813.

Huang, C. Y., Ayliffe, M. A., & Timmis, J. N. (2003). Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature*, *422*(6927), 72–76. doi:10.1038/nature01435 nature01435 [pii]

Huberman, J. A., & Riggs, A. D. (1966). Autoradiography of chromosomal DNA fibers from Chinese hamster cells. *Proceedings of the National Academy of Sciences of the United States of America*, *55*(3), 599–606.

Huda, A., Bowen, N. J., Conley, A. B., & Jordan, I. K. (2011). Epigenetic regulation of transposable element derived human gene promoters. *Gene*, *475*(1), 39–48.

Iborra, F. J., Pombo, A., Jackson, D. A., & Cook, P. R. (1996). Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J Cell Sci*, *109*(6), 1427–1436.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., & Speed, T. P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*, *31*(4), e15.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)*, *4*(2), 249–264.

Ishihama, A. (2009). The nucleoid: an overview. *In Neidhardt, F C et Al. (eds), Escherichia Coli and Salmnella. ASM Press, Washington.*, *2nd versio*.

Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., … Takahashi, K. (2008). Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. *Science*, *321*(5892), 1088–1091.

Ishikawa, F., & Naito, T. (1999). Why do we have linear chromosomes? A matter of Adam and Eve. *Mutation Research - DNA Repair*, *434*(2), 99–107.

Iwasaki, O., & Noma, K.-I. (2012). Global genome organization mediated by rna polymerase iii-transcribed genes in fission yeast. *Gene*, *493*(2), 195–200.

Iwasaki, O., Tanaka, A., Tanizawa, H., Grewal, S. I. S., & Noma, K. I. (2010). Centromeric localization of dispersed Pol III genes in fission yeast. *Molecular Biology of the Cell*, *21*(2), 254–265.

Jackson, D., Wang, X., & Rudner, D. Z. (2012). Spatio-temporal organization of replication in bacteria and eukaryotes (nucleoids and nuclei). *Cold Spring Harbor Perspectives in Biology*, *4*(8).

Jackson, V. (1978). Studies on Histone Organization in Nucleosome Using Formaldehyde as a Reversible Cross-Linking Agent. *Cell*, *15*(3), 945–954.

Jackson, V. (1999). Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods: A Companion to Methods in Enzymology*, *17*(2), 125–139.

Jiang, H., Song, C., Chen, C. C., Xu, R., Raines, K. S., Fahimian, B. P., Miao, J. (2010). Quantitative 3D imaging of whole, unstained cells by using X-ray diffraction microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11234–11239.

Jin, D. J., & Cabrera, J. E. (2006). Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in Escherichia coli. *Journal of Structural Biology*, *156*(2), 284–291.

Jin, D. J., Cagliero, C., & Zhou, Y. N. (2012). Growth rate regulation in Escherichia coli. *FEMS Microbiol Rev*, *36*(2), 269–287.

Jing, H., Vakoc, C. R., Ying, L., Mandat, S., Wang, H., Zheng, X., & Blobel, G. A. (2008). Exchange of GATA Factors Mediates Transitions in Looped Chromatin Organization at a Developmentally Regulated Gene Locus. *Molecular Cell*, *29*(2), 232–242.

Jjingo, D., Huda, A., Gundapuneni, M., Mariño-Ramírez, L., & Jordan, I. K. (2011). Effect of the transposable element environment of human genes on gene length and expression. *Genome Biology and Evolution*, *3*(1), 259–271.

Jones, L. J. F., Carballido-López, R., & Errington, J. (2001). Control of cell shape in bacteria: Helical, actin-like filaments in Bacillus subtilis. *Cell*, *104*(6), 913–922.

Jones, R. M., Luo, L., Ardita, C. S., Richardson, A. N., Kwon, Y. M., Mercante, J. W., … Neish, A. S. (2013). Symbiotic lactobacilli stimulate gut epithelial proliferation via Nox-mediated generation of reactive oxygen species. *The EMBO Journal*, 1–12. doi:10.1038/emboj.2013.224

Jun, S., & Mulder, B. (2006). Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12388–12393.

Jurka, J., Kapitonov, V. V, Kohany, O., & Jurka, M. V. (2007). Repetitive sequences in complex genomes: Structure and evolution. *Annual Review of Genomics and Human Genetics*.

Kadauke, S., & Blobel, G. A. (2013). Mitotic bookmarking by transcription factors. *Epigenetics and Chromatin*, *6*(1)

Kaguni, J. M. (2006). DnaA: controlling the initiation of bacterial DNA replication and more. *Annual Review of Microbiology*, *60*, 351–75. doi:10.1146/annurev.micro.60.080805.142111

Kamakaka, R. T., & Biggins, S. (2005). Histone variants: Deviants? *Genes and Development*, *19*(3), 295–310.

Kang, D. W., Park, J. G., Ilhan, Z. E., Wallstrom, G., LaBaer, J., Adams, J. B., & Krajmalnik-Brown, R. (2013). Reduced Incidence of Prevotella and Other Fermenters in Intestinal Microflora of Autistic Children. *PLoS ONE*, *8*(7).

Kanoh, J., Sadaie, M., Urano, T., & Ishikawa, F. (2005). Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres. *Current Biology*, *15*(20), 1808–1819.

Katayama, T., Ozaki, S., Keyamura, K., & Fujimitsu, K. (2010). Regulation of the replication cycle: Conserved and diverse regulatory systems for DnaA and oriC. *Nature Reviews Microbiology*, *8*(3), 163–170.

Kauffman, S. A. (1993). The origins of order: self organization and selection in evolution, 709.

Kauffman, S. A. (1995). At home in the universe : the search for laws of self-organization and complexity. *Oxford University Press, New York*.

Kellum, R., & Alberts, B. M. (1995). Heterochromatin protein 1 is required for correct chromosome segregation in Drosophila embryos. *Journal of Cell Science*, *108*(4), 1419–1431.

Kim, Y. J., Cecchini, K. R., & Kim, T. H. (2011). Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(18), 7391–7396.

Kite, G. L., & Chambers, J. R. (1912). Vital staining of chromosomes and the function and structure of the nucleus. *Science*, *36*(932), 639–641.

Kloc, A., Zaratiegui, M., Nora, E., & Martienssen, R. (2008). RNA Interference Guides Histone Modification during the S Phase of Chromosomal Replication. *Current Biology*, *18*(7), 490–495.

Kocanova, S., Kerr, E. A., Rafique, S., Boyle, S., Katz, E., Caze-Subra, S., … Bystricky, K. (2010). Activation of estrogen-responsive genes does not require their nuclear co-localization. *PLoS Genetics*, *6*(4).

Koopman, W. J. H., Distelmaier, F., Smeitink, J. A. M., & Willems, P. H. G. M. (2013). OXPHOS mutations and neurodegeneration. *EMBO Journal*, *32*(1), 9–29.

Korf, B. R., & Diacumakos, E. G. (1978). Microsurgically-extracted metaphase chromosomes of the Indian muntjac examined with phase contrast and scanning electron microscopy. *Experimental Cell Research*, *111*(1), 83–93.

Korf, B. R., & Diacumakos, E. G. (1980). Absence of true interchromosomal connectives in microsurgically isolated chromosomes. *Experimental Cell Research*, *130*(2), 377–385.

Kritikou, E. (2005). Exploring chromatin architecture by FISHing in 3D. *Nat Rev Genet*, *6*(6), 429. Retrieved from http://dx.doi.org/10.1038/nrg1641

Kruse, T., Blagoev, B., Løbner-Olesen, A., Wachi, M., Sasaki, K., Iwai, N., … Gerdes, K. (2006). Actin homolog MreB and RNA polymerase interact and are both required for chromosome segregation in Escherichia coli. *Genes and Development*, *20*(1), 113–124.

Kucej, M., & Butow, R. A. (2007). Evolutionary tinkering with mitochondrial nucleoids. *Trends in Cell Biology*, *17*(12), 586–592.

Kuhlman, T. E., & Cox, E. C. (2012). Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Molecular Systems Biology*, *8*.

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., … Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(28), 10684–10689.

Kutschera, U., & Niklas, K. J. (2005). Endosymbiosis, cell evolution, and speciation. *Theory in Biosciences*, *124*(1), 1–24.

Kuzminov, A. (2013). The chromosome cycle of prokaryotes. *Molecular Microbiology*, *90*(2), 214–227.

Kuznetsova, I. S., Enukashvily, N. I., Noniashvili, E. M., Shatrova, A. N., Aksenov, N. D., Zenin, V. V, … Podgornaya, O. I. (2007). Evidence for the existence of satellite DNA-containing connection between metaphase chromosomes. *Journal of Cellular Biochemistry*, *101*(4), 1046–1061.

Lagunas, R. (1976). Energy metabolism of Saccharomyces cerevisiae discrepancy between ATP balance and known metabolic functions. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *440*, 661–674.

Laine, J. P., Singh, B. N., Krishnamurthy, S., & Hampsey, M. (2009). A physiological role for gene loops in yeast. *Genes Dev*, *23*(22), 2604–2609. doi:23/22/2604 [pii] 10.1101/gad.1823609

Laman, A. G., Kurjukov, S. G., Bulgakova, E. V, Anikeeva, N. N., & Brovko, F. A. (2001). Subtractive hybridization of biotinylated DNA in phenol emulsion. *Journal of Biochemical and Biophysical Methods*, *50*(1), 43–52.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lang, B. F., Ia, W., Munchen, D.-, Davies, C. R. W., & The, S. (1984). The mitochondrial genome of the fission yeast Schizosaccharomyces pombe: highly homologous introns are inserted at the same position of the otherwise less conserved cox1 genes in Schizosaccharomyces pombe and Aspergillus nidulans. *The EMBO Journal*, *3*(9), 2129–2136.

Langer, P. R., Waldrop, A. A., & Ward, D. C. (1981). Enzymatic synthesis of biotin-labeled polynucleotides: Novel nucleic acid affinity probes. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(11 II), 6633–6637.

Langowski, J. (2010). Chromosome conformation by crosslinking polymer physics matters. *Nucleus*, *1*(1), 37–39.

Lawrence, J. G. (2002). Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell*, *110*(4), 407–413.

Lebedeva, M. A., & Gerald S. Shadel. (2007). Cell Cycle- and Ribonucleotide Reductase-Driven Changes in mtDNA Copy Number Influence mtDNA Inheritance Without Compromising Mitochondrial Gene Expression. *Cell Cycle*, *6*(16), 2048–2057.

Lee, G. R., Spilianakis, C. G., & Flavell, R. A. (2005). Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. *Nature Immunology*, *6*(1), 42–48.

Leij, F. De, Sutton, E. J., Whipps, J. M., Fenlon, J. S., & Lynch, J. M. (1995). Impact of Field Release of Genetically Modified Pseudomonas fluorescens on Indigenous Microbial Populations of Wheat. *Applied and Environmental Microbiology*, *61*(9), 3443–3453.

Leirião, P., Rodrigues, C, D., Albuquerque, S, S., & Mota, M, M. (2004). Survival of protozoan intracellular parasites in host cells. *EMBO Reports*, *5*(12), 1142–7. doi:10.1038/sj.embor.7400299

Lemmer, P., Gunkel, M., Baddeley, D., Kaufmann, R., Urich, A., Weiland, Y., … Cremer, C. (2008). SPDM: light microscopy with single-molecule resolution at the nanoscale. *Applied Physics B-Lasers and Optics*, *93*(1), 1–12. doi:DOI 10.1007/s00340-008-3152-x

Lenglez, S., Hermand, D., & Decottignies, A. (2010). Genome-wide mapping of nuclear mitochondrial DNA sequences links DNA replication origins to chromosomal double-strand break formation in Schizosaccharomyces pombe. *Genome Research*, *20*(9), 1250–61. doi:10.1101/gr.104513.109

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., … Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature*, *430*(6999), 573–578.

Lennartsson, A., & Ekwall, K. (2009). Histone modification patterns and epigenetic codes. *Biochimica et Biophysica Acta - General Subjects*, *1790*(9), 863–868.

Leonard, A. C., & Grimwade, J. E. (2011). Regulation of DnaA assembly and activity: Taking directions from the genome. *Annual Review of Microbiology*, *65*, 19–35.

Lercher, M. J., Urrutia, A. O., & Hurst, L. D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genetics*, *31*(2), 180–183.

Levi, V., Ruan, Q., Plutz, M., Belmont, A. S., & Gratton, E. (2005). Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope.

Biophysical Journal, *89*(6), 4275–4285. 28444496753&partnerID=40&md5=454cf1ac821418f4654d1ed64730f1a4
Li, G., & Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics and Development*, *21*, 1–12.
Li, Q., Barkess, G., & Qian, H. (2006). Chromatin looping and the probability of transcription. *Trends in Genetics*, *22*(4), 197–202.
Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics Applications Note*, *24*(5), 713–714.
Li, Y., Sergueev, K., & Austin, S. (2002). The segregation of the Escherichia coli origin and terminus of replication. *Molecular Microbiology*, *46*(4), 985–995.
Liang, B. C. (1996). Evidence for association of mitochondrial DNA sequence amplification and nuclear localization in human low-grade gliomas. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *354*(1), 27–33.
Lichter, P., Cremer, T., Tang, C. J. C., Watkins, P. C., Manuelidis, L., & Ward, D. C. (1988). Rapid detection of human chromosome 21 aberrations by in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(24), 9664–9668.
Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293.
Liesa, M., Palacín, M., & Zorzano, A. (2009). Mitochondrial dynamics in mammalian health and disease. *Physiological Reviews*, *89*(3), 799–845.
Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., … Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*, *312*(5771), 269–272.
Lisch, D., & Bennetzen, J. L. (2011). Transposable element origins of epigenetic gene regulation. *Current Opinion in Plant Biology*, *14*(2), 156–161.
Liu, X., Wang, X., Reyes-Lamothe, R., & Sherratt, D. (2010). Replication-directed sister chromosome alignment in Escherichia coli. *Molecular Microbiology*, *75*(5), 1090–1097.
Liu, Z., & Butow, R. A. (2006). Mitochondrial retrograde signaling. *Annual Review of Genetics*, *40*, 159–185.
Liu, Z., & Garrard, W. T. (2005). Long-range interactions between three transcriptional enhancers, active VÎº gene promoters, and a 3â€^{2} boundary sequence spanning 46 kilobases. *Molecular and Cellular Biology*, *25*(8), 3220–3231.
Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal Interactions and Olfactory Receptor Choice. *Cell*, *126*(2), 403–413.
Lorenz, D. R., Mikheyeva, I. V, Johansen, P., Meyer, L., Berg, A., Grewal, S. I. S., & Cam, H. P. (2012). CENP-B cooperates with Set1 in bidirectional transcriptional silencing and genome organization of retrotransposons. *Molecular and Cellular Biology*, *32*(20), 4215–25. doi:10.1128/MCB.00395-12
Louis, E. J., Naumova, E. S., Lee, A., Naumov, G., & Haber, J. E. (1994). The chromosome end in yeast: Its mosaic nature and influence on recombinational dynamics. *Genetics*, *136*(3), 789–802.
Lu, M., Campbell, J. L., Boye, E., & Kleckner, N. (1994). SeqA: A negative modulator of replication initiation in E. coli. *Cell*, *77*(3), 413–426.
Lucas, J. N., Wu, X., Guo, E., Chi, L. E., & Chen, Z. (2006). An efficient chemical method to generate repetitive sequences depleted DNA probes. *American Journal of Medical Genetics, Part A*, *140*(19), 2115–2120.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260.

Luijsterburg, M. S., Noom, M. C., Wuite, G. J. L., & Dame, R. T. (2006). The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *Journal of Structural Biology*, *156*(2), 262–272.

Lukas, J., Lukas, C., & Bartek, J. (2004). Mammalian cell cycle checkpoints: Signalling pathways and their organization in space and time. *DNA Repair*, *3*(8-9), 997–1007.

MacAskill, A. F., Rinholm, J. E., Twelvetrees, A. E., Arancibia-Carcamo, I. L., Muir, J., Fransson, A., … Kittler, J. T. (2009). Miro1 Is a Calcium Sensor for Glutamate Receptor-Dependent Localization of Mitochondria at Synapses. *Neuron*, *61*(4), 541–555.

Madsen, J. S., Burmølle, M., Hansen, L. H., & Sørensen, S. J. (2012). The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunology and Medical Microbiology*, *65*(2), 183–195.

Maeshima, K., Hihara, S., & Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo? *Current Opinion in Cell Biology*, *22*(3), 291–7. doi:10.1016/j.ceb.2010.03.001

Mandal, S., Guptan, P., Owusu-Ansah, E., & Banerjee, U. (2005). Mitochondrial regulation of cell cycle progression during development as revealed by the tenured mutation in Drosophila. *Developmental Cell*, *9*(6), 843–54. doi:10.1016/j.devcel.2005.11.006

Maniotis, A. J., Bojanowski, K., & Ingber, D. E. (1997). Mechanical continuity and reversible chromosome disassembly within intact genomes removed from living cells. *Journal of Cellular Biochemistry*, *65*(1), 114–130.

Marella, N. V, Bhattacharya, S., Mukherjee, L., Xu, J., & Berezney, R. (2009). Cell type specific chromosome territory organization in the interphase nucleus of normal and cancer cells. *Journal of Cellular Physiology*, *221*(1), 130–138.

Marenduzzo, D., Micheletti, C., & Cook, P. R. (2006). Entropy-driven genome organization. *Biophysical Journal*, *90*(10), 3712–3721.

Marenduzzo, D., Micheletti, C., & Orlandini, E. (2006). The depletion attraction: an underappreciated force driving cellular organization. *J. Cell Biol.*, *175*(5), 681–686.

Marenduzzo, D., Micheletti, C., & Orlandini, E. (2010). Biopolymer organization upon confinement. *Journal of Physics Condensed Matter*, *22*(28).

Margolin, W. (2010). Imaging the bacterial nucleoid. *In Dame, R. T. and Dorman, C. J. (eds.), Bacterial Chromatin. Springer, Dordrecht ; New York,*, *x*, 448.

Marshak, A. (1951). Chromosome structure in Escherichia coli. *Experimental Cell Research*, *2*(2), 243–251.

Marsman, J., & Horsfield, J. A. (2012). Long distance relationships: Enhancer-promoter communication and dynamic gene transcription. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1819*(11-12), 1217–1227.

Martienssen, R. A., Zaratiegui, M., & Goto, D. B. (2005). RNA interference and heterochromatin in the fission yeast Schizosaccharomyces pombe. *Trends in Genetics*, *21*(8), 450–456.

Martin, S., & Pombo, A. (2003). Transcription factories: Quantitative studies of nanostructures in the mammalian nucleus. *Chromosome Research*, *11*(5), 461–470.

Martin, W. (2003). Gene transfer from organelles to the nucleus: frequent and in big chunks. *Proc Natl Acad Sci U S A*, *100*(15), 8612–8614. doi:10.1073/pnas.1633606100 1633606100 [pii]

Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P., & Hough, P. V. C. (1991). DNA looping and Sp1 multimer links: A mechanism for transcriptional synergism and

enhancement. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(13), 5670–5674.
Matsuda, A., Shao, L., Boulanger, J., Kervrann, C., Carlton, P. M., Kner, P., … Sedat, J. W. (2010). Condensed mitotic chromosome structure at nanometer resolution using PALM and EGFP- histones. *PLoS ONE*, *5*(9), 1–12.
Matthews, D. M., & Jenks, S. M. (2013). Ingestion of Mycobacterium vaccae decreases anxiety-related behavior and improves learning in mice. *Behavioural Processes*, *96*, 27–35.
Matthews, K. S. (1992). DNA looping. *Microbiological Reviews*, *56*(1), 123–136.
McGovern, V., Higgins, N. P., Chiz, R. S., & Jaworski, A. (1994). H-NS over-expression induces an artificial stationary phase by silencing global transcription. *Biochimie*, *76*(10-11), 1019–1029.
McInerny, C. J. (2004). Cell cycle-regulated transcription in fission yeast. *Biochemical Society Transactions*, *32*(Pt 6), 967–72. doi:10.1042/BST0320972
McInerny, C. J. (2011). Cell Cycle Regulated Gene Expression in Yeasts. 79551634070&partnerID=40&md5=2ed8bc06c905d9471c95b3ff7e4062d9
Meister, P., Gehlen, L. R., Varela, E., Kalck, V., & Gasser, S. M. (2010). Visualizing yeast chromosomes and nuclear architecture. *Methods in Enzymology*, *470*(C), 535–567.
Mercier, R., Petit, M.-A., Schbath, S., Robin, S., El Karoui, M., Boccard, F., & Espéli, O. (2008). The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. *Cell*, *135*(3), 475–85. doi:10.1016/j.cell.2008.08.031
Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiology Reviews*, *26*(4), 355–374.
Mintz, P. J., & Spector, D. L. (2000). Compartmentalization of RNA processing factors within nuclear speckles. *Journal of Structural Biology*, *129*(2-3), 241–251.
Misteli, T. (2007). Beyond the Sequence: Cellular Organization of Genome Function. *Cell*, *128*(4), 787–800.
Misteli, T., & Soutoglou, E. (2009). The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature Reviews Molecular Cell Biology*, *10*(4), 243–254.
Mitchell, J. A., & Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev*, *22*(1), 20–25.
Mitra, K., Wunder, C., Roysam, B., Lin, G., & Lippincott-Schwartz, J. (2009). A hyperfused mitochondrial state achieved at G1-S regulates cyclin E buildup and entry into S phase. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(29), 11960–5. doi:10.1073/pnas.0904875106
Moran, J. V, Zimmerly, S., Eskes, R., Kennell, J. C., & Lambowitz, A. M. et al. (1995). Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol*, *15*, 2828–2838.
Moreira, P. I., Zhu, X., Wang, X., Lee, H. g., Nunomura, A., Petersen, R. B., … Smith, M. A. (2010). Mitochondria: A therapeutic target in neurodegeneration. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, *1802*(1), 212–220.
Morris, G. E. (2008). The Cajal body. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1783*(11), 2108–2115.
Motamedi, M. R., Hong, E. J. E., Li, X., Gerber, S., Denison, C., Gygi, S., & Moazed, D. (2008). HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic Gene Silencing by Nonoverlapping Mechanisms. *Molecular Cell*, *32*(6), 778–790.
Mukherjee, S., Erickson, H., & Bastia, D. (1988). Detection of DNA Looping Due to Simultaneous Interaction of a DNA-Binding Protein with 2 Spatially Separated Binding-Sites on DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(17), 6287–6291.

Müller, R. (1995). Transcriptional regulation during the mammalian cell cycle. *Trends in Genetics : TIG*, *11*(5), 173–8. doi:10.1016/S0168-9525(00)89039-3

Murphy, L. D., & Zimmerman, S. B. (2001). A limited loss of DNA compaction accompanying the release of cytoplasm from cells of Escherichia coli. *Journal of Structural Biology*, *133*(1), 75–86.

Murrell, A., Heeson, S., & Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature Genetics*, *36*(8), 889–893.

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., … Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, *502*(7469), 59–64. doi:10.1038/nature12593

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L., & Dekker, J. (2013). Organization of the Mitotic Chromosome. *Science*, *342*, 948–953. doi:10.1126/science.1236083

Nesic, D., Tanackovic, G., & Krämer, A. (2004). A role for Cajal bodies in the final steps of U2 snRNP biogenesis. *Journal of Cell Science*, *117*(19), 4423–4433.

Nicodemi, M., & Prisco, A. (2009). Thermodynamic pathways to genome spatial organization in the cell nucleus. *Biophysical Journal*, *96*(6), 2168–2177.

Nielsen, H. J., Youngren, B., Hansen, F. G., & Austin, S. (2007). Dynamics of Escherichia coli chromosome segregation during multifork replication. *Journal of Bacteriology*, *189*(23), 8660–8666.

Niki, H., Yamaichi, Y., & Hiraga, S. (2000). Dynamic organization of chromosomal DNA in Escherichia coli. *Genes and Development*, *14*(2), 212–223.

Niwa, O., & Yanagida, M. (1985). Triploid meiosis and aneuploidy in Schizosaccharomyces pombe: An unstable aneuploid disomic for chromosome III. *Current Genetics*, *9*(6), 463–470.

Nolis, I. K., McKay, D. J., Mantouvalou, E., Lomvardas, S., Merika, M., & Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(48), 20222–20227.

Noordermeer, D., & de Laat, W. (2008). Joining the Loops : β-Globin Gene Regulation. *IUBMB Life*, *60*(12), 824–33. doi:10.1002/iub.129

Nosek, J., Kosa, P., & Tomaska, L. (2006). On the origin of telomeres: A glimpse at the pre-telomerase world. *BioEssays*, *28*(2), 182–190.

Nurse, P. (1975). Genetic control of cell size at cell division in yeast. *Nature*, *256*(5518), 547–551.

Nurse, P., Thuriaux, P., & Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. *Molecular and General Genetics*, *146*(2), 167–178.

O'Sullivan, J. M. (2011). Chromosome organizaton in simple and complex unicellular organisms. *Current Issues in Molecular Biology*, *13*(2), 37–42.

O'Sullivan, J. M., Sontam, D. M., Grierson, R., & Jones, B. (2009). Repeated elements coordinate the spatial organization of the yeast genome. *Yeast*, *26*(2), 125–138. doi:10.1002/yea.1657

O'Sullivan, J. M., Tan-Wong, S. M., Morillon, A., Lee, B., Coles, J., Mellor, J., & Proudfoot, N. J. (2004). Gene loops juxtapose promoters and terminators in yeast. *Nat Genet*, *36*(9), 1014–1018.

Odijk, T. (1998). Osmotic compaction of supercoiled DNA into a bacterial nucleoid. *Biophysical Chemistry*, *73*(1-2), 23–29.

Ohniwa, R. L., Ushijima, Y., Saito, S., & Morikawa, K. (2011). Proteomic analyses of nucleoid-associated proteins in escherichia coli, pseudomonas aeruginosa, bacillus subtilis, and staphylococcus aureus. *PLoS ONE*, *6*(4).

Ohta, S., Wood, L., Bukowski-Wills, J. C., Rappsilber, J., & Earnshaw, W. C. (2011). Building mitotic chromosomes. *Current Opinion in Cell Biology*, *23*(1), 114–121.

Okamoto, K., Perlman, P. S., & Butow, R. a. (1998). The sorting of mitochondrial DNA and mitochondrial proteins in zygotes: preferential transmission of mitochondrial DNA to the medial bud. *The Journal of Cell Biology*, *142*(3), 613–23.

Olins, A. L., & Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science*, *183*(4122), 330–332.

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Leatherwood, J. (2005). The Cell Cycle – Regulated Genes of Schizosaccharomyces pombe, *3*(7). doi:10.1371/journal.pbio.0030225

Olsson, I., & Bjerling, P. (2011). Advancing our understanding of functional genome organisation through studies in the fission yeast. *Current Genetics*, *57*(1), 1–12. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-78951488313&partnerID=40&md5=90a115a47a0aa1f956f26b13b809e4cc

Orlando, D. A., Lin, C. Y., Bernard, A., Wang, J. Y., Socolar, J. E., Iversen, E. S., Haase, S. B. (2008). Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature*, *453*(7197), 944–947. doi:nature06955 [pii] 10.1038/nature06955

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genetics*, *36*(10), 1065–1071.

Palstra, R. J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., & de Laat, W. (2008). Maintenance of Long-Range DNA Interactions after Inhibition of Ongoing RNA Polymerase II Transcription. *PLoS ONE*, *3*(2), e1661.

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., & De Laat, W. (2003). The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics*, *35*(2), 190–194.

Papantonis, A., & Cook, P. R. (2013). Transcription Factories: Genome Organization and Gene Regulation. *Chemical Reviews*, *113*(11), 8683–8705. doi:10.1021/cr300513p

Parada, L. A., McQueen, P. G., & Misteli, T. (2004). Tissue-specific spatial organization of genomes. *Genome Biology*, *5*(7).

Parikh, V., Morgan, M., Scott, R., Clements, L., & Butow, R. (1987). The mitochondrial genotype can influence nuclear gene expression in yeast. *Science*, *235*, 576–580.

Parikh, V. S., Conrad-webb, H., Docherty, R., & Butow, R. A. (1989). Interaction between the Yeast Mitochondrial and Nuclear Genomes Influences the Abundance of Novel Transcripts Derived from the Spacer Region of the Nuclear Ribosomal DNA Repeat. *Molecular and Cellular Biology*, *9*(5). doi:10.1128/MCB.9.5.1897.Updated

Peebles, C. L., Belcher, S. M., Zhang, M., Dietrich, R. C., & Perlman, P. S. (1993). Mutation of the conserved first nucleotide of a group II intron from yeast mitochondrial DNA reduces the rate but allows accurate splicing. *Journal of Biological Chemistry*, *268*, 11929–11938.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W. M., Solovei, I., Brugman, W., van Steensel, B. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Molecular Cell*, *38*(4), 603–613.

Pettijohn, D. E., & Hecht, R. (1974). RNA molecules bound to the folded bacterial genome stabilize DNA folds and segregate domains of supercoiling. *Cold Spring Harb Symp Quant Biol*, *38*, 31–41.

Phillips, J. E., & Corces, V. G. (2009). CTCF: Master Weaver of the Genome. *Cell*, *137*(7), 1194–1211.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., & Gray, J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: Detection of trisomy 21 and translocations of chromosome 4. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(23), 9138–9142.

Poirier, M. G., & Marko, J. F. (2002). Mitotic chromosomes are chromatin networks without a mechanically contiguous protein scaffold. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15393–15397.

Postow, L., Hardy, C. D., Arsuaga, J., & Cozzarelli, N. R. (2004). Topological domain structure of the Escherichia coli chromosome. *Genes and Development*, *18*(14), 1766–1779.

Provost, P., Silverstein, R. A., Dishart, D., Walfridsson, J., Djupedal, I., Kniola, B., … Ekwall, K. (2002). Dicer is required for chromosome segregation and gene silencing in fission yeast cells. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(26), 16648–16653.

R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *ISBN 3-900051-07-0, URL http://www.R-project.org.*

Ragoczy, T., Telling, A., Sawado, T., Groudine, M., & Kosak, S. T. (2003). A genetic analysis of chromosome territory looping: Diverse roles for distal regulatory elements. *Chromosome Research*, *11*(5), 513–525.

Rajapakse, I., & Groudine, M. (2011). On emerging nuclear order. *Journal of Cell Biology*, *192*(5), 711–721.

Ramírez, M. J., & Surrallés, J. (2009). Laser confocal microscopy analysis of human interphase nuclei by three-dimensional FISH reveals dynamic perinucleolar clustering of telomeres. *Cytogenetic and Genome Research*, *122*(3-4), 237–242.

Reitzer, L. J., & Magasanik, B. (1986). Transcription of glnA in E. coli is stimulated by activator bound to sites far from the promoter. *Cell*, *45*, 785–792.

Reyes-Lamothe, R., Wang, X., & Sherratt, D. (2008). Escherichia coli and its chromosome. *Trends in Microbiology*, *16*(5), 238–45. doi:10.1016/j.tim.2008.02.003

Ribeiro, S. A., Vagnarelli, P., Dong, Y., Hori, T., McEwen, B. F., Fukagawa, T., Earnshaw, W. C. (2010). A super-resolution map of the vertebrate kinetochore. *Proceedings of the National Academy of Sciences*, *107*(23), 10484–10489. doi:10.1073/pnas.1002325107

Ricchetti, M., Fairhead, C., & Dujon, B. (1999). Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature*, *402*(6757), 96–100. 0033523903&partnerID=40&md5=5f750d7c44149d3941387459f97bfd8c

Ricchetti, M., Tekaia, F., & Dujon, B. (2004). Continued colonization of the human genome by mitochondrial DNA. *PLoS Biology*, *2*(9).

Richly, E., & Leister, D. (2004). NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution*, *21*(6), 1081–1084.

Ridaura, V. K., Faith, J. J., Rey, F. E., Cheng, J., Duncan, A. E., Kau, A. L., Gordon, J. I. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science*, *341*(6150).

Ringrose, L., Chabanis, S., Angrand, P. O., Woodroofe, C., & Stewart, A. F. (1999). Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *Embo J*, *18*(23), 6630–6641.

Rinke, B., Bischoff, A., Meffert, M. C., Scharschmidt, R., Hausmann, M., Stelzer, E. H. K., Cremer, C. (1995). Volume ratios of painted chromosome territories 5, 7 and X in female human cell nuclei studied with confocal laser microscopy and the Cavalieri estimator. *Bioimaging*, *3*(1), 1–11.

Rippe, K. (2001). Making contacts on a nucleic acid polymer. *Trends in Biochemical Sciences*, *26*(12), 733–740.

Rippe, K. (2007). Dynamic organization of the cell nucleus. *Current Opinion in Genetics and Development*, *17*(5), 373–380.

Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., & Belmont, A. S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *Journal of Cell Biology*, *135*(6 II), 1685–1700.

Robinow, C., & Kellenberger, E. (1994). The bacterial nucleoid revisited. *Microbiological Reviews*, *58*(2), 211–232. Retrieved from <Go to ISI>://A1994NP60300003

Robinson, P. J., & Rhodes, D. (2006). Structure of the “30 nm” chromatin fibre: A key role for the linker histone. *Current Opinion in Structural Biology*, *16*(3), 336–343.

Rodley, C. D. M., Bertels, F., Jones, B., & O’Sullivan, J. M. (2009). Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genetics and Biology*, *46*(11), 879–886. doi:doi:10.1016/j.fgb.2009.07.006

Rodley, C. D. M., Grand, R. S., Gehlen, L. R., Greyling, G., Jones, M. B., & O’Sullivan, J. M. (2012). Mitochondrial-nuclear DNA interactions contribute to the regulation of nuclear transcript levels as part of the inter-organelle communication system. *PLoS ONE*, *7*(1), e30943. doi:10.1371/journal.pone.0030943

Rosa, A., & Everaers, R. (2008). Structure and dynamics of interphase chromosomes. *PLoS Computational Biology*, *4*(8).

Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z., & Higgins, N. P. (2012). Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLoS Genetics*, *8*(8), e1002845. doi:10.1371/journal.pgen.1002845

Rustici, G., Mata, J., Kivinen, K., Lió, P., Penkett, C. J., Burns, G., … Bähler, J. (2004). Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics*, *36*(8), 809–17. doi:10.1038/ng1377

Sabatinosa, S. A., & Forsburga, S. L. (2010). Molecular Genetics of Schizosaccharomyces pombe. In *Methods in Enzymology* (pp. 759–795).

Saier Jr., M. H. (2008). The bacterial chromosome. *Crit Rev Biochem Mol Biol*, *43*(2), 89–134. doi:791724494 [pii] 10.1080/10409230801921262

Sambrook, J., & Russell, D. W. (2001). Molecular cloning: A laboratory manual (Vol. Third Edit, p. 157). Cold Spring Harbor Laboratory Press, Clod Spring Harbor, New York.

Sánchez-Romero, M. A., Busby, S. J. W., Dyer, N. P., Ott, S., Millard, A. D., & Grainger, D. C. (2010). Dynamic distribution of SeqA protein across the chromosome of Escherichia coli K-12. *mBio*, *1*(1).

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, *489*(7414), 109–13. doi:10.1038/nature11279

Sarge, K. D., & Park-Sarge, O. (2005). Gene bookmarking: keeping the pages open, *30*(11), 605–610.

Sawitzke, J. A., & Austin, S. (2000). Suppression of chromosome segregation defects of Escherichia coli muk mutants by mutations in topoisomerase I. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(4), 1671–1676.

Schäfer, B. (2003). Genetic conservation versus variability in mitochondria: The architecture of the mitochondrial genome in the petite-negative yeast Schizosaccharomyces pombe. *Current Genetics*, *43*(5), 311–326.

Schalch, T., Duda, S., Sargent, D. F., & Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, *436*(7047), 138–141.

Schardin, M., Cremer, T., Hager, H. D., & Lang, M. (1985). Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Human Genetics*, *71*(4), 281–287.

Scherf, A., Lopez-Rubio, J. J., & Riviere, L. (2008). Antigenic variation in Plasmodium falciparum. *Annual Review of Microbiology*, *62*, 445–470.

Scherthan, H., Bahler, J., & Kohli, J. (1994). Dynamics of chromosome organization and pairing during meiotic prophase in fission yeast. *Journal of Cell Biology*, *127*(2), 273–285.

Schmiedeberg, L., Skene, P., Deaton, A., & Bird, A. (2009). A Temporal Threshold for Formaldehyde Crosslinking and Fixation. *PLoS ONE*, *4*(2), -. doi:Artn E4636 Doi 10.1371/Journal.Pone.0004636

Schober, H., Ferreira, H., Kalck, V., Gehlen, L. R., & Gasser, S. M. (2009). Yeast telomerase and the SUN domain protein Mps3 anchor telomeres and repress subtelomeric recombination. *Genes Dev*, *23*(8), 928–938. doi:23/8/928 [pii] 10.1101/gad.1787509

Schober, H., Kalck, V., Vega-Palas, M. A., Van Houwe, G., Sage, D., Unser, M., Gasser, S. M. (2008). Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res*, *18*(2), 261–271. doi:gr.6687808 [pii] 10.1101/gr.6687808

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., Fraser, P. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics*, *42*(1), 53–61.

Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., … Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, *132*(5), 887–898.

Schueler, M. G., & Sullivan, B. A. (2006). Structural and functional dynamics of human centromeric chromatin. *Annual Review of Genomics and Human Genetics*.

Sekimata, M., Pérez-Melgosa, M., Miller, S. A., Weinmann, A. S., Sabo, P. J., Sandstrom, R., Wilson, C. B. (2009). CCCTC-Binding Factor and the Transcription Factor T-bet Orchestrate T Helper 1 Cell-Specific Structure and Function at the Interferon-γ Locus. *Immunity*, *31*(4), 551–564.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, *148*(3), 458–472.

Sherratt, D. J. (2003). Bacterial chromosome dynamics. *Science*, *301*(5634), 780–785.

Shirafuji, N., Takahashi, S., Matsuda, S., & Asano, S. (1997). Mitochondrial antisense RNA for cytochrome C oxidase (MARCO) can induce morphologic changes and cell death in human hematopoietic cell lines. *Blood*, *90*(11), 4567–4577.

Shoshani, T., Benvenisty, N., Trus, M., & Reshef, L. (1991). Cis-regulatory elements that confer differential expression upon the rat gene encoding phosphoenolpyruvate carboxykinase in kidney and liver. *Gene*, *101*(2), 279–283.

Simonis, M., & de Laat, W. (2008). FISH-eyed and genome-wide views on the spatial organisation of gene expression. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1783*(11), 2052–2060.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., De Wit, E., De Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, *38*(11), 1348–1354.

Singh, B. N., & Hampsey, M. (2007). A transcription-independent role for TFIIB in gene looping. *Mol Cell*, *27*(5), 806–816. doi:S1097-2765(07)00486-8 [pii] 10.1016/j.molcel.2007.07.013

Smallwood, A., & Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. *Current Opinion in Cell Biology*, *25*(3), 387–394.

Smyth, C. P., Lundbäck, T., Renzoni, D., Siligardi, G., Beavil, R., Layton, M., Ladbury, J. E. (2000). Oligomerization of the chromatin-structuring protein H-NS. *Molecular Microbiology*, *36*(4), 962–972.

Snoep, J. L., Van Der Weijden, C. C., Andersen, H. W., Westerhoff, H. V, & Jensen, P. R. (2002). DNA supercoiling in Escherichia coli is under tight and subtle homeostatic control, involving gene-expression and metabolic regulation of both topoisomerase I and DNA gyrase. *European Journal of Biochemistry*, *269*(6), 1662–1669.

Solomon, M. J., & Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci U S A*, *82*(19), 6470–6474.

Spilianakis, C. G., & Flavell, R. A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nature Immunology*, *5*(10), 1017–1027.

Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R., & Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature*, *435*(7042), 637–645.

Splinter, E., Grosveld, F., & de Laat, W. (2004). 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol*, *375*, 493–507.

Splinter, E., Heath, H., Kooren, J., Palstra, R., Klous, P., Grosveld, F., De Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes and Development*, *20*(17), 2349–2354.

Sproul, D., Gilbert, N., & Bickmore, W. A. (2005). The role of chromatin structure in regulating the expression of clustered genes. *Nature Reviews Genetics*, *6*(10), 775–781.

Spurio, R., Durrenberger, M., Falconi, M., La Teana, A., Pon, C. L., & Gualerzi, C. O. (1992). Lethal overproduction of the Escherichia coli nucleoid protein H-NS: Ultramicroscopic and molecular autopsy. *Molecular and General Genetics*, *231*(2), 201–211.

Squarzoni, S., Cinti, C., Santi, S., Valmori, A., & Maraldi, N. M. (1994). Preparation of chromosome spreads for electron (TEM, SEM, STEM), light and confocal microscopy. *Chromosoma*, *103*(6), 381–392.

Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F., & Soler, E. (2012). Transcription regulation by distal enhancers: Who's in the loop? *Transcription*, *3*(4).

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Experimental Hematology*, *33*(3), 259–271.

Stegemann, S., Hartmann, S., Ruf, S., & Bock, R. (2003). High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci U S A*, *100*(15), 8828–8833. doi:10.1073/pnas.1430924100 1430924100 [pii]

Stein, G. S., Zaidi, S. K., Braastad, C. D., Montecino, M., Van Wijnen, A. J., Choi, J. Y., Javed, A. (2003). Functional architecture of the nucleus: Organizing the regulatory machinery for gene expression, replication and repair. *Trends in Cell Biology*, *13*(11), 584–592.

Stein, R. A., Deng, S., & Higgins, N. P. (2005). Measuring chromosome dynamics on different time scales using resolvases with varying half-lives. *Molecular Microbiology*, *56*(4), 1049–1061.

Stowe, D. F., & Camara, A. K. S. (2009). Mitochondrial reactive oxygen species production in excitable cells: Modulators of mitochondrial and cell function. *Antioxidants and Redox Signaling*, *11*(6), 1373–1414.

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45.

Straight, A. F., Belmont, A. S., Robinett, C. C., & Murray, A. W. (1996). GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr Biol*, *6*(12), 1599–1608.

Stuger, R., Woldringh, C. L., der Weijden, C. C., Vischer, N. O. E., Bakker, B. M., Van Spanning, R. J. M., Weterhoff, H. V. (2002). DNA supercoiling by gyrase is linked to nucleoid compaction. *Molecular Biology Reports*, *29*(1-2), 79–82.

Sugino, A., Goodman, H. M., & Heyneker, H. L. (1977). Interaction of bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base paired ends. *Journal of Biological Chemistry*, *252*(11), 3987–3994.

Sun, B. F., Xiao, J. H., He, S., Liu, L., Murphy, R. W., & Huang, D. W. (2013). Multiple Interkingdom Horizontal Gene Transfers in Pyrenophora and Closely Related Species and Their Contributions to Phytopathogenic Lifestyles. *PLoS ONE*, *8*(3).

Sun, Q., & Margolin, W. (2004). Effects of Perturbing Nucleoid Structure on Nucleoid Occlusion-Mediated Toporegulation of FtsZ Ring Assembly Effects of Perturbing Nucleoid Structure on Nucleoid Occlusion-Mediated Toporegulation of FtsZ Ring Assembly, *186*(12). doi:10.1128/JB.186.12.3951

Sunako, Y., Onogi, T., & Hiraga, S. (2001). Sister chromosome cohesion of Escherichia coli. *Molecular Microbiology*, *42*(5), 1233–1241.

Suzuki, K., Iwata, K., & Yoshida, K. (2001). Genome analysis of Agrobacterium tumefaciens: Construction of physical maps for linear and circular chromosomal DNAs, determination of copy number ratio and mapping of chromosomal virulence genes. *DNA Research*, *8*(4), 141–152.

Taddei, A., Van Houwe, G., Nagai, S., Erb, L., Van Nimwegen, E., & Gasser, S. M. (2009). The functional importance of telomere clustering: Global changes in gene expression result from SIR factor dispersion. *Genome Research*, *19*(4), 611–625.

Tanaka, H., Yasuzawa, K., Kohno, K., Goshima, N., Kano, Y., Saiki, T., & Imamoto, F. (1995). Role of HU proteins in forming and constraining supercoils of chromosomal DNA in Escherichia coli. *Molecular and General Genetics*, *248*(5), 518–526.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J. R., Wickramasinghe, P., Lee, M., Noma, K. I. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Research*, *38*(22), 8164–8177.

Tan-Wong, S. M., Wijayatilake, H. D., & Proudfoot, N. J. (2009). Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes and Development*, *23*(22), 2610–2624.

Thanbichler, M., Wang, S. C., & Shapiro, L. (2005). The bacterial nucleoid: A highly organized and dynamic structure. *Journal of Cellular Biochemistry*, *96*(3), 506–521.

Therizols, P., Duong, T., Dujon, B., Zimmer, C., & Fabre, E. (2010). Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc Natl Acad Sci U S A*, *107*(5), 2025–2030. doi:0914187107 [pii] 10.1073/pnas.0914187107

Thiel, A., Valens, M., Vallet-Gely, I., Espéli, O., & Boccard, F. (2012). Long-range chromosome organization in E. coli: a site-specific system isolates the Ter macrodomain. *PLoS Genetics*, *8*(4), e1002672. doi:10.1371/journal.pgen.1002672

Thomson, I., Gilchrist, S., Bickmore, W., & Chubb, J. R. (2004). The Radial Positioning of Chromatin Is Not Inherited through Mitosis but Is Established De Novo in Early G1. *Current Biology*, *14*(2), 166–172. doi:10.1016/j.cub.2003.12.024

Thorsness, P. E., & Fox, T. D. (1990). Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. *Nature*, *346*(6282), 376–379.

Thorsness, P. E., & Fox, T. D. (1993). Nuclear mutations in Saccharomyces cerevisiae that affect the escape of DNA from mitochondria to the nucleus. *Genetics*, *134*(1), 21–28.

Timmis, J. N., Ayliff, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, *5*(2), 123–135.

Tiwari, V. K., Cope, L., McGarvey, K. M., Ohm, J. E., & Baylin, S. B. (2008). A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Research*, *18*(7), 1171–1179.

Toan, N. M., Marenduzzo, D., Cook, P. R., & Micheletti, C. (2006). Depletion effects and loop formation in self-avoiding polymers. *Phys Rev Lett*, *97*(17), 178302.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell*, *10*(6), 1453–1465.

Tonthat, N. K., Arold, S. T., Pickering, B. F., Van Dyke, M. W., Liang, S., Lu, Y., Schumacher, M. A. (2011). Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. *EMBO Journal*, *30*(1), 154–164.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–78. doi:10.1038/nprot.2012.016

Travers, A., & Muskhelishvili, G. (2005). Bacterial chromatin. *Current Opinion in Genetics and Development*, *15*(5 SPEC. ISS.), 507–514.

Traxler, M. F., Summers, S. M., Nguyen, H. T., Zacharia, V. M., Hightower, G. A., Smith, J. T., & Conway, T. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. *Molecular Microbiology*, *68*(5), 1128–1148.

Tremethick, D. J. (2007). Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*, *128*(4), 651–654.

Turakainen, H., Naumov, G., Naumova, E., & Korhola, M. (1993). Physical mapping of the MEL gene family in Saccharomyces cerevisiae. *Current Genetics*, *24*(6), 461–464.

Turner, B. M. (2000). Histone acetylation and an epigenetic code. *BioEssays*, *22*(9), 836–845.

Umbarger, M. A., Toro, E., Wright, M. A., Porreca, G. J., Baù, D., Hong, S.-H., … Church, G. M. (2011). The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation. *Molecular Cell*, *44*(2), 252–264. doi:10.1016/j.molcel.2011.09.010

Valens, M., Penaud, S., Rossignol, M., Cornet, F., & Boccard, F. (2004). Macrodomain organization of the Escherichia coli chromosome. *The EMBO Journal*, *23*(21), 4330–41. doi:10.1038/sj.emboj.7600434

Valenzuela, L., Dhillon, N., & Kamakaka, R. T. (2009). Transcription independent insulation at TFIIIC-dependent insulators. *Genetics*, *183*(1), 131–148.

Van Berkum, N. L., & Dekker, J. (2009). Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods Mol Biol*, *567*, 189–213.

Van Helvoort, J. M. L. M., Huls, P. G., Vischer, N. O. E., & Woldringh, C. L. (1998). Fused nucleoids resegregate faster than cell elongation in Escherichia coli pbpB(Ts) filaments after release from chloramphenicol inhibition. *Microbiology*, *144*(5), 1309–1317.

Van Helvoort, J. M. L. M., Kool, J., & Woldringh, C. L. (1996). Chloramphenicol causes fusion of separated nucleoids in Escherichia coli K-12 cells and filaments. *Journal of Bacteriology*, *178*(14), 4289–4293.

Van Holde, K., & Zlatanova, J. (2007). Chromatin fiber structure: Where is the problem now? *Seminars in Cell and Developmental Biology*, *18*(5), 651–658.

Van Steensel, B., & Dekker, J. (2010). Genomics tools for unraveling chromosome architecture. *Nature Biotechnology*, *28*, 1089–1095.

Van Winkle, Q., Renoll, M. W., Garvey, J. S., Palik, E. S., & Prebus, A. F. (1953). Electron microscopy of isolated chromosomes. *Experimental Cell Research*, *5*(1), 38–47.

Veening, J. W., Smits, W. K., & Kuipers, O. P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. *Annual Review of Microbiology*, *62*, 193–210.

Vellai, T., & Vida, G. (1999). The origin of eukaryotes: The difference between prokaryotic and eukaryotic cells. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1428), 1571–1577.

Versteeg, R., van Schaik, B. D. C., van Batenburg, M. F., Roos, M., Monajemi, R., Caron, H., … van Kampen, A. H. C. (2003). The human transcriptome map reveals extremes in gene dentistry, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Research*, *13*(9), 1998–2004.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S., & Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, *297*(5588), 1833–1837.

Volpe, T., Schramke, V., Hamilton, G. L., White, S. A., Teng, G., Martienssen, R. A., & Allshire, R. C. (2003). RNA interference is required for normal centromere function in fission yeast. *Chromosome Research*, *11*(2), 137–146.

Volpi, E. V, & Bridger, J. M. (2008). FISH glossary: An overview of the fluorescence in situ hybridization technique. *BioTechniques*, *45*(4), 385–409.

Von Freiesleben, U., Rasmussen, K. V, & Schaechter, M. (1994). SeqA limits DnaA activity in replication from oriC in Escherichia coli. *Mol Microbiol*, *14*, 763–772.

Waldminghaus, T., & Skarstad, K. (2009). The Escherichia coli SeqA protein. *Plasmid*, *61*(3), 141–150.

Walker, G. M. (1999). Synchronization of yeast cell populations. *Methods in Cell Science*, *21*(2-3), 87–93.

Wallace, D. C. (2012). Mitochondria and cancer. *Nature Reviews. Cancer*, *12*(10), 685–98. doi:10.1038/nrc3365

Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., & Cremer, T. (2003). Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *Journal of Cell Biology*, *160*(5), 685–697.

Wang, W., Li, G. W., Chen, C., Xie, X. S., & Zhuang, X. (2011). Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, *333*(6048), 1445–1449.

Wang, X., Kam, Z., Carlton, P. M., Xu, L., Sedat, J. W., & Blackburn, E. H. (2008). Rapid telomere motions in live human cells analyzed by highly time-resolved microscopy. *Epigenetics Chromatin*, *1*(1), 4.

Wang, X., Liu, X., Possoz, C., & Sherratt, D. J. (2006). The two Escherichia coli chromosome arms locate to separate cell halves. *Genes & Development*, *20*(13), 1727–31. doi:10.1101/gad.388406

Wang, X., Possoz, C., & Sherratt, D. J. (2005). Dancing around the divisome: Asymmetric chromosome segregation in Escherichia coli. *Genes and Development*, *19*(19), 2367–2377.

Wang, X., & Sherratt, D. J. (2010). Independent segregation of the two arms of the Escherichia coli ori region requires neither RNA synthesis nor MreB dynamics. *Journal of Bacteriology*, *192*(23), 6143–6153.

Weatherall, D. J. (2001). Genotype – Phenotype Relationships. *ENCYCLOPEDIA OF LIFE SCIENCES*.

Weinstock, G. M. (2007). ENCODE: More genomic empowerment. *Genome Research*, *17*(6), 667–668.

Weir, B. A., & Yaffe, M. P. (2004). Mmd1p, a Novel, Conserved Protein Essential for Normal Mitochondrial Morphology and Distribution in the Fission Yeast Schizosaccharomyces pombe. *Molecular Biology of the Cell*, *15*(4), 1656–1665.

Widom, J., & Klug, A. (1985). Structure of the 300Å chromatin filament: X-ray diffraction from oriented samples. *Cell*, *43*(1), 207–213.

Wiedmann, M., Wilson, W. I., Czajka, J., Luo, J., Barany, F., & Batt, C. A. (1994). Ligase chain reaction (LCR) - Overview and applications. *Genome Research*, *3*(4), S51–S64.

Wiggins, P. A., Cheveralls, K. C., Martin, J. S., Lintner, R., & Kondev, J. (2010). Strong intranucleoid interactions organize the Escherichia colichromosome into a nucleoid filament. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 4991–4995.

Wiggins, P. A., Dame, R. T., Noom, M. C., & Wuite, G. J. L. (2009). Protein-mediated molecular bridging: A key mechanism in biopolymer organization. *Biophysical Journal*, *97*(7), 1997–2003.

Williams, A., Spilianakis, C. G., & Flavell, R. A. (2010). Interchromosomal association and gene regulation in trans. *Trends in Genetics*, *26*(4), 188–197.

Williams, S. P., Athey, B. D., Muglia, L. J., Schappe, R. S., Gough, A. H., & Langmore, J. P. (1986). Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophysical Journal*, *49*(1), 233–248.

Woldringh, C. L. (2002). The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Molecular Microbiology*, *45*(1), 17–29.

Woldringh, C. L., Jensen, P. R., & Westerhoff, H. V. (1995). Structure and partitioning of bacterial DNA: Determined by a balance of compaction and expansion forces? *FEMS Microbiology Letters*, *131*(3), 235–242.

Woldringh, C. L., & Nanninga, N. (2006). Structural and physical aspects of bacterial chromosome segregation. *Journal of Structural Biology*, *156*(2), 273–283.

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Gloux, S. (2002). The genome sequence of Schizosaccharomyces pombe. *Nature*, *415*(6874), 871–880.

Woodcock, C. L., & Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev*, *11*(2), 130–135.

Woodcock, C. L. F., Safer, J. P., & Stanchfield, J. E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. *Experimental Cell Research*, *97*(1), 101–110.

Woodcock, C. L. F., Sweetman, H. E., & Frado, L. L. (1976). Structural repeating units in chromatin. II. Their isolation and partial characterization. *Experimental Cell Research*, *97*(1), 111–119. 0017228007&partnerID=40&md5=4472f3ec0ec68d1e9acd0a03440e5946

Woodcock, C. L., & Ghosh, R. P. (2010). Chromatin higher-order structure and dynamics. *Cold Spring Harbor Perspectives in Biology*, *2*(5), a000596. doi:10.1101/cshperspect.a000596

Wu, X., Liu, H., Su, J., Lv, J., Cui, Y., Wang, F., & Zhang, Y. (2013). Z curve theory-based analysis of the dynamic nature of nucleosome positioning in Saccharomyces cerevisiae. *Gene*, *530*(1), 8–18.

Xu, M., & Cook, P. R. (2008). Similar active genes cluster in specialized transcription factories. *Journal of Cell Biology*, *181*(4), 615–623.

Yu, Y., Rambo, T., Currie, J., Saski, C., Kim, H. R., Collura, K., … Dean, R. A. (2003). In-depth view of structure, activity, and evolution of rice chromosome 10. *Science*, *300*(5625), 1566–1569.

Zechiedrich, E. L., Khodursky, A. B., Bachellier, S., Schneider, R., Chen, D., Lilley, D. M. J., & Cozzarelli, N. R. (2000). Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli. *Journal of Biological Chemistry*, *275*(11), 8103–8113.

Zhang, Y., Mccord, R. P., Ho, Y., Lajoie, B. R., Dominic, G., Simon, A. C., Dekker, J. (2012). Chromosomal translocations are guided by the spatial organization of the genome. *Cell*, *148*(5), 908–921. doi:10.1016/j.cell.2012.02.002.Chromosomal

Zhao, R., Nakamura, T., Fu, Y., Lazar, Z., & Spector, D. L. (2011). Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nature Cell Biology*, *13*(11), 1295–1304.

Zhao, W., Wang, L., Zhang, M., Wang, P., Zhang, L., Yuan, C., Gao, C. (2011). NF-κB- and AP-1-mediated DNA looping regulates osteopontin transcription in endotoxin-stimulated murine macrophages. *Journal of Immunology*, *186*(5), 3173–3179.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*, *38*(11), 1341–1347.

Zimmerman, S. B. (2002). Toroidal nucleoids in Escherichia coli exposed to chloramphenicol. *Journal of Structural Biology*, *138*(3), 199–206.

Zimmerman, S. B. (2003). Underlying regularity in the shapes of nucleoids of Escherichia coli: Implications for nucleoid organization and partition. *Journal of Structural Biology*, *142*(2), 256–265.

Zimmerman, S. B. (2006). Shape and compaction of Escherichia coli nucleoids. *Journal of Structural Biology*, *156*(2), 255–261.

Zimmerman, S. B., & Murphy, L. D. (1996). Macromolecular crowding and the mandatory condensation of DNA in bacteria. *FEBS Letters*, *390*(3), 245–248.

Zimmerman, S. B., & Murphy, L. D. (2001). Release of compact nucleoids with characteristic shapes from Escherichia coli. *Journal of Bacteriology*, *183*(17), 5041–5049.

Zusman, D. R., Carbonell, A., & Haga, J. Y. (1973). Nucleoid condensation and cell division in Escherichia coli MX74T2 ts52 after inhibition of protein synthesis. *Journal of Bacteriology*, *115*(3), 1167–1178.

Zwicker, J., & Müller, R. (1995). Cell cycle-regulated transcription in mammalian cells. *Progress in Cell Cycle Research*, *1*, 91–99.

7 APPENDIX I

In: Advances in Genetics Research. Volume 5 ISBN 978-1-61122-072-8
Editor: Kevin V. Urbano © 2010 Nova Science Publishers, Inc.

Chapter 4

METHODS FOR THE INVESTIGATION OF CHROMOSOME ORGANIZATION

Ralph S. Grand, Lutz R. Gehlen and Justin M. O'Sullivan*
Institute of Natural Sciences, Massey University
Auckland, New Zealand

ABSTRACT

Structure is incorporated into the everyday objects, social and genomic networks that surround us providing a delicate balance of flexibility and stability. This chapter will discuss the methods that are currently used to untangle genome structure and the networks of interactions that occur between chromosomes to provide genomic stability. We will concentrate on the recently developed proximity-based ligation approaches which are opening new avenues of investigation into genome architecture. Finally, we will discuss a range of pitfalls that must be avoided when designing and performing these deceptively simple experiments. Although these methods are still in the early stages of development and application, they promise to bring about an explosion in our understanding of genome organization and stability.

* * Corresponding Author
+64 9 4140800 extension 9811
j.m.osullivan@massey.ac.nz

1. INTRODUCTION

Nuclear structure is non random. Rather, eukaryotic chromosomes are organized and folded to occupy zones or territories within the nucleus [1]. These zones are not exclusive and there is a certain amount of chromosome intermingling and interactions which logic dictates must affect genomic stability. Moreover, the relative positions of loci are heritable, move, and rearrange in response to the cell-cycle [2-4], development processes (e.g. [4]), and transcriptional memory [5-8].

Interactions between loci on different chromosomes can be direct, mediated by proteins, or simply due to chance encounters within shared nuclear space. Despite the lack of information on the nature of the chromosomal interactions, it is reasonable to assume that the proximity of loci and the formation of inter- and intra-chromosomal interactions is regulated to prevent inappropriate outcomes resulting from chance encounters [9]. While the form that these regulatory mechanisms take remains to be determined, understanding them is central to models explaining genome stability.

Detailed studies of chromosome organization have expanded alongside light and electron microscopy [10-16] together with advances in staining techniques (*e.g.* Giemsa [17, 18] or silver nitrate stain [12, 19]). The recent development of new proximity-based ligation techniques has provided unprecedented insights into global genome organization by expanding on the local observations made using region specific microscopic methods. These and other studies have provided information on the formation [20, 21], regulation [22] and functional consequences of specific intra-chromosomal interactions in yeast (e.g. [5, 6]), mice (e.g. [23-25]), and humans (e.g. [26, 27]).

In this chapter we will describe some of the methods that are currently used to study genome structure and networks. We will concentrate on the proximity-based ligation methodologies that are currently breaking new boundaries in the study of genome architecture. As part of our description we will highlight the inter-relationships, advantages, and disadvantages that are inherent within these approaches.

2. MICROSCOPY

Staining and microscope methods have revealed many features of chromosome organization (*e.g.* see review [1]). Microscopic measurements can be performed on fixed or living cells. Both approaches have distinct advantages and disadvantages with respect to experimental limitations, image quality, and range of questions which can be addressed. For example, the use of live cells allows the experimenter to follow the actual movement of a locus in real-time [3, 28]. The fluorescent staining of specific genomic loci is also of particular importance in the microscopic study of genome organization as it allows the measurement of a locus' position with respect to other loci or to nuclear structures on a single cell level.

The classical techniques of immunofluorescence (IF) and fluorescence *in situ* hybridization (FISH), which are performed on fixed samples, still play an important role in the investigation of nuclear architecture despite the recent development of *in vivo* live cell fluorescence imaging. IF uses fluorescently labeled antibodies against the target molecule or, more commonly, a primary antibody against the target and a labeled secondary antibody which then recognizes the primary antibody. FISH was pioneered by Cremer [29, 30], Pinkel [31] and Lichter [32] in the 1980's, for the detection of chromosome abnormalities and developmental defects, including trisomies [30, 31]. It relies upon the hybridization and microscopic detection of fluorescently labeled nucleic acid probes to locate specific DNA sequences or even entire chromosomes within cells or nuclei [33]. Since IF and FISH work on fixed samples they generally tolerate higher light intensities and longer exposure times than live cell imaging thus they typically produce higher quality images. The requirement for long exposure times is also shared by the recently developed sub-diffraction techniques, which include variations of spatially modulated illumination (SMI and SPDM) [34-36], photo-activated localization microscopy (PALM) [37], stochastic optical reconstruction microscopy (STORM) [38], and coherent diffraction microscopy [39]. Therefore, these sub-diffraction techniques also only work on fixed samples. While fixed cells enable the use of higher light intensities and exposure times, the preparation of cells for these techniques involves harsh treatments to allow the access of macromolecules that are used to detect the loci [40]. Therefore, great care has to be taken to prevent the appearance of artifacts while preserving as much of the nuclear 3D structure as possible during these treatments.

Live cell imaging techniques using fluorescently tagged repressor/operator combinations were developed to enable *in vivo* observations of loci [41, 42]. These initial studies have been extended to include detailed investigations into, for example, telomere positioning (*e.g.* [3, 43, 44]), artificial chromosomes (*e.g.* [45, 46]), and low and high-throughput studies of locus positioning (*e.g.* [28, 47, 48]). In addition to the obvious advantage that the cells are alive, *in vivo* imaging allows the observation of single cells over time. However, these advantages come at the cost of technical limitations. For example, it is essential to minimize the cellular damage caused by the high intensity illumination. Therefore, light intensity and exposure times are usually kept as low as possible which limits the quality of the images and the number of images that can be taken. For each experiment, the optimum balance between spatial resolution, time resolution, and image quality has to be identified. In addition, image acquisition of live cells is usually more laborious than that for fixed cells which makes it more difficult to image sufficient numbers of cells.

A distinct advantage of fluorescence microscopy techniques compared to most other methods is their ability to monitor cellular and nuclear architecture at the single cell level. This makes it possible to identify positive or negative correlations between internally and externally driven processes (*e.g.* the cell cycle) [3]. These relationships are likely to be averaged out in bulk ensemble measurements. However, the corollary is that even supposed high-throughput microscopic techniques are limited to the observation of a small number of preselected genomic loci in a relatively small number of cells.

3. Molecular Biology Techniques for the Identification of Chromosome Organization

Reductionist approaches to studying genome organization drove the development of molecular techniques that rely upon proximity-based ligation [49, 50] events between spatially adjacent loci. Chromosome Conformation Capture (3C; see section 3.1) was the first widely applied molecular technique developed specifically for this experimental niche [51]. Unfortunately, the *a priori* requirement to know, or suspect, the identity of the spatially proximal loci is a serious limitation to the 3C technique. Circular Chromosome Conformation Capture (4C; see section 3.2) [24, 25] and

Chromosome Conformation Capture Carbon Copy (5C; see section 3.3) [27] were subsequently developed to enable in-depth interrogations of the role of individual regions within the global genome organization (Figure 1). Unlike microscope based technologies, these methods can be applied following standard molecular training and using inexpensive equipment. Therefore, it has not taken long for these proximity-based ligation techniques to gain wide-spread acceptance as rapid and adaptable methodologies for the study of genomic organization at molecular level resolution.

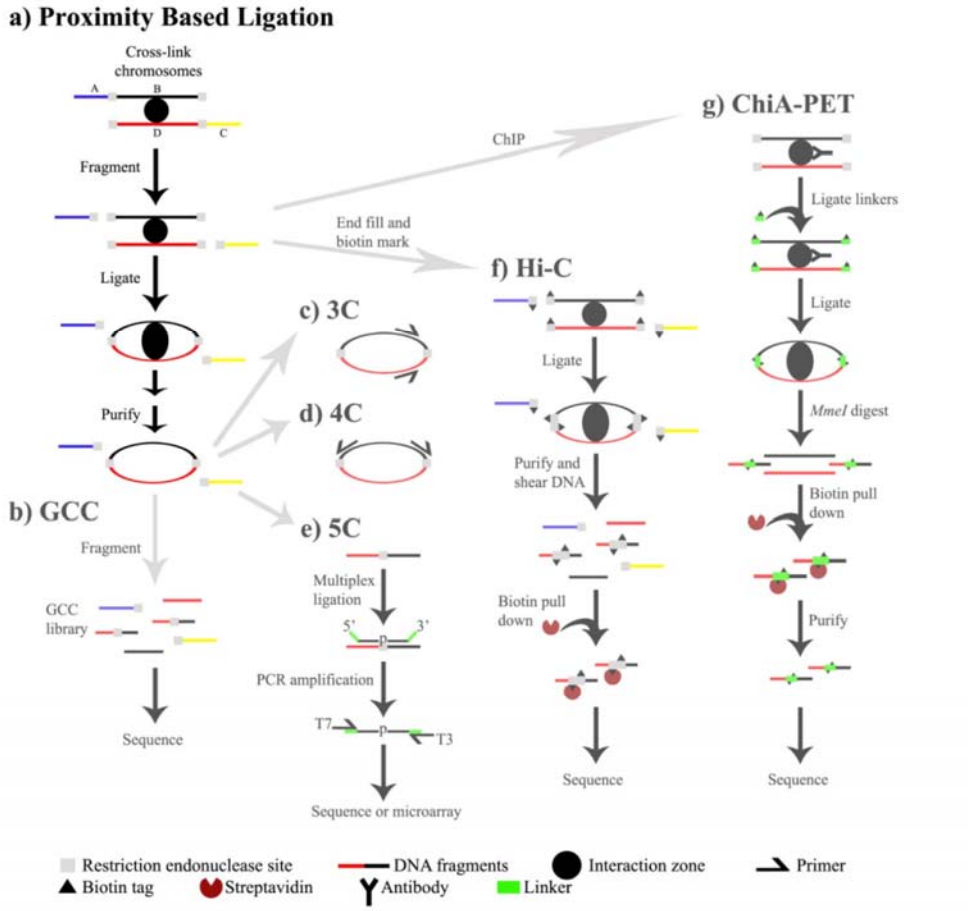


Figure 1. Molecular methods for the detection of chromosomal interactions. a) The chromatin preparation module that forms the core of the proximity-based ligation methodologies. Chromatin is cross-linked, fragmented, and diluted prior to intra-molecular ligation, Cross-links are reversed and the library purified prior to

interaction detection. Fragmentation is achieved by restriction enzyme digestion (*e.g.* [51]) or sonication (*e.g.* [59]). Fragmentation by sonication requires end repair. HiC [26] and ChiA-PET [59] deviate from this general scheme through the filling of the cohesive ends f) and incorporation of linkers and g), respectively. **b)** GCC is the simplest high-throughput methodology for the study of genome organization. GCC combines proximity-based ligation with random fragmentation and high-throughput sequencing in order to determine the genome organization. 3C and 4C incorporate **c)** direct and **d)** inverse PCR amplification to identify specific (*e.g.* [51]) and unknown (*e.g.* [24, 25]) interactions, respectively. Amplified products are sequenced or hybridized to a microarray to confirm their identity. **e)** 5C [27] utilizes primer ligation and amplification with universal T7 forward and reverse primers to amplify interacting segments. Interactions are subsequently identified by micro-array analyses or high-throughput sequencing. **f)** Hi-C [26]: Sticky ends produced by restriction enzyme digestion are end-filled with nucleotides, one of which is biotinylated (black triangle). Intra-molecular ligation is performed prior to DNA purification and shearing. Biotinylated junctions are subsequently isolated and identified by paired-end sequencing. **g)** ChiA-PET[59] deviates from the general chromatin preparation scheme through the integration of chromatin immunoprecipitation to fractionate the restriction enzyme digested chromatin. Subsequent steps include the incorporation of a biotinylated linker, prior to a final purification and high-throughput sequencing. Enhanced 4C [85], 6C, [86] and Associated Chromosome Trap (ACT) [104] are further variations on the proximity ligation method and were not included in this figure for simplicity.

The fundamental basis of all of the following techniques is relatively simple. Genomic organization is captured by a combination of fragmentation and ligation prior to the identification of the interacting loci (Figure 1a).

3.1. Cross-Linking

The internal linkages that hold DNA fragments together may [51] or may not [49] be stabilized by cross-linking. Cross-links are typically short range and reversible. The most commonly used cross-linking agent is formaldehyde as it joins primary amines through a Schiff's intermediate [52-54], resulting in the formation of covalent linkages that are approximately 2 Ångstrom (Å) in length. Alternative cross-linking systems exist (for a brief review and protocols see [54]), but they are not widely employed in the study of genome organization

There are several attributes of cross-linking that must be taken into account when designing experiments and subsequently interpreting the results. Firstly, it is important to recognize that the size of the cross-link (2 Å, the linear distance covered by formaldehyde link) does not mean that the

interaction is that close to begin with, particularly as hydrogen bonds involving amines occur at distances ranging from 1.2 Å to 4 Å [55]. Secondly, cross-linking agents can covalently stabilize large complexes which physically link loci that remain separated by relatively large distances and hence are not clustered within a small volume. Thirdly, it has been demonstrated that interactions that are captured by formaldehyde cross-linking must be stable for periods of approximately 5 seconds or more [53]. This may explain observations that while histones are readily cross-linked to DNA [56], other proteins, including regulatory factors, are not [57]. Even increased times of cross-linking will not necessarily result in a comprehensive and fully representative set of the interactions that were present within the sample.

3.2. Chromosome Conformation Capture (3C)

Cross-linking is the first step in 3C and it is typically performed for 10 minutes [25, 26, 51, 58-62]. Once interactions have been captured by cross-linking, cells or nuclei are lysed and the cross-linked genomes are fragmented, usually by restriction digestion. The choice of restriction enzyme is dependent upon empirical and hypothetical considerations. Firstly, the sequence of the genomic region(s) under investigation automatically limits the choice of restriction enzymes that can be used, as certain restriction sites will or will not be present. Moreover, the desired resolution also affects the choice, as restriction enzymes that have a four nucleotide recognition site typically, but not uniformly, achieve a higher resolution than enzymes that have a six nucleotide recognition site. Secondly, while the use of restriction enzymes that leave overhangs or blunt ends is in many respects a personal preference when performing 3C, cohesive ends result in more efficient ligation [63, 64]. Thirdly, the choice of enzyme is affected by the presence of DNA modification systems within the organism and the chances the modified sites overlap restriction sites. Finally, it is critical that the enzyme is easily inactivated following the digestion as carry-over of active restriction enzyme into the ligation step is counter-productive. Inactivation is generally achieved by the addition of SDS and incubation at moderate temperatures (≤65°C) in order to minimize the temperature dependent reversal of the formaldehyde mediated cross-links [56, 57].

Following digestion, the cross-linked genomic restriction fragments are diluted into a large volume prior to ligation of the free ends. In simplest

terms, ligation requires that free DNA ends and ligase come together in the same place at one time. Diluting DNA samples reduces the chances of the tripartite association, unless the free DNA ends are from the same molecule. However, ligatable free ends can come from two or more linear DNA fragments that are physically held together in one complex by internal linkages. In such a situation, dilution will separate the non linked DNA fragments but not those within a single complex. Importantly, dilution promotes the formation of intra-molecular ligation products by reducing the chances of inter-molecular ligations within the fixed time period. Moreover, the frequency of free DNA end association depends on factors that include the chromatin state and as a consequence the real and apparent length of the DNA fragment(s) [65, 66]. Internal controls are often used to determine the inter-molecular ligation frequency and to standardize between samples. However, the use of internal controls requires evidence that the chosen loci are unlinked, spatially separate [62, 67], and have similar interaction frequencies in the different samples. The use of external controls is more advisable as a means to determine the frequency of inter-fragment ligation events (unpublished results).

Finally, cross-links are reversed and the DNA fragments are purified. Following purification, the mixture of linear and circular DNA molecules constitutes a 3C library. PCR is subsequently used to obtain preliminary confirmation of the presence of an interaction within the 3C library. PCR primers are designed to amplify across the reconstituted restriction site (Figure 1c) formed by the ligation of two interacting restriction fragments. If an interaction occurred a PCR product will be produced. In contrast, an absence of ligation results in the primer annealing sites (Figure 1c) [51] being in a non-convergent orientation that prevents PCR amplification. The identity of the PCR product is subsequently confirmed by restriction digestion and/or sequencing.

3.3. Circular Chromosome Conformation Capture (4C)

4C is a global method suitable for the detection of DNA regions that interact with a predetermined region of interest (the bait). 4C is similar to 3C in its reliance on PCR for the identification of chromosomal interactions, albeit inverse PCR directed out of the bait fragment (Figure 1d). There are two versions of 4C, both of which utilize inverse PCR [24, 25]. As such, 4C enables the identification of 'unknown' interacting DNA fragments and thus

removes the 3C requirement of *a priori* suspecting the identity of both interacting loci.

3.4. Chromosome Conformation Capture Carbon-Copy (5C)

5C ([27], reviewed in [68]) employs a different approach to identify interacting DNA fragments as it is not anchored to a single locus and therefore maps the global interaction networks. As in 3C, 5C requires oligonucleotides that anneal either side of the restriction enzyme site(s) of interest. These oligonucleotides contain universal T7 primer binding sites and are designed to anneal adjacent to each other on the same strand of the ligated template, provided the two DNA fragments interact. This arrangement enables the oligonucleotides to be ligated together and amplified, using the T7 forward and reverse primers which are complementary to the ends of the oligonucleotides. The primer ligation step utilized in 5C is an extension of the ligase chain reaction developed in the early 1990's (reviewed in [69]). Amplification products, and hence chromosomal interactions, are subsequently identified by sequencing or microarray detection (Figure 1e) [27]. 5C provided a step towards the elucidation of global genome organization by enabling multiplexing within a single 3C sample, using forward and reverse oligonucleotides that span a single region or an entire genome.

4. GLOBAL IDENTIFICATION OF DNA INTERACTIONS

The development of 3C enabled genome organization to be readily studied at molecular level resolution. However, 3C remains limited to the analysis of small numbers of chromosomal loci. Moreover, both 3C and 4C require the *a priori* selection of at least one of the two interacting fragments. This bias represents a serious hurdle between these techniques and their application to the study of global genome organization. While the requirement for *a priori* knowledge can be circumvented by the 5C method, the global application of this method requires blanketing the genome with forward and reverse oligonucleotides. Therefore, the genome wide application of these methods is impractical.

Genome Conformation Capture (GCC) [61], Hi-C [26, 58] and Chromatin Interaction Analysis using Paired-End Tag sequencing (ChiA-

PET) [59] overcome the requirements for the *a priori* identification of regions of interest. These techniques rely on cross-linking and intra-molecular ligation [51]. However, rather than designing primers for the detection of the interacting fragments, the purified ligated DNA samples are sequenced using high-throughput genome sequencing technology. The large numbers of sequence tags that are obtained are analyzed to construct genome wide maps of the interactions that were captured at the time of cross-linking. As such, the high-throughput techniques facilitate the unbiased and detailed interrogation of global genome organization. ChiA-PET [59] incorporates chromatin immunoprecipitation (ChIP) and is designed to implicate proteinaceous factors in the maintenance of chromosomal interactions. ChiA-PET will be discussed in more detail in section 5.

ChIA-PET, GCC and Hi-C have been used to produce the first generation of experimentally defined models of global eukaryotic genome organization [26, 58, 59, 61]. As a result, proximity ligation and the proximity-based ligation technologies have increased our ability to perform in-depth and rapid hypothesis driven investigations of genome organization (*e.g.* [70-75]) and its roles in the regulation of gene transcription [76-81] and epigenetic memory [5-8].

4.1. Genome Conformation Capture (GCC)

The first eukaryotic global chromosome interaction map was produced from exponentially growing *Saccharomyces cerevisiae* cells using GCC [61]. GCC is the simplest of the high-throughput techniques. GCC effectively extends the 3C methodology by identifying spatial proximity by direct sequencing (Figure 1b) [61]. Moreover, GCC does not require the specific amplification of the captured interactions, or the incorporation of linkers, or labeled nucleotides (see below). As such, it can be used to produce high resolution, unbiased chromosome interaction maps. However, the lack of enrichment means that large numbers of sequence tags are required, even for relatively small genomes. Enrichment strategies that rely upon immuno-precipitation of the restriction enzyme site were proposed [61] and have subsequently been developed (unpublished). The incorporation of these enrichment strategies will enable the investigation of larger genomes. However, the existing GCC methodology can be easily applied to the study of genomes under 20 Mbp in size.

4.3. Hi-C

Hi-C [26] is similar to GCC [61] except that it incorporates an enrichment step that is specific for the ligated products. Enrichment is achieved by filling the restriction site 5′ overhang with a biotinylated residue prior to blunt-end proximity ligation. The resulting ligation products consist of fragments that were originally in close spatial proximity in the nucleus, marked with biotin at the junction. The ligation product library is subsequently sheared and fragments containing the interaction junction are enriched using streptavidin beads. The library is then analyzed using high-throughput DNA sequencing, producing a catalogue of interacting fragments (Figure 1f).

Hi-C provides a crucial step towards the identification of DNA interactions on a genome wide scale in large genomes albeit with some limitations. Firstly, blunt end ligation is less efficient than sticky ended ligation [63, 64]. Secondly, combining phenol:chloroform extractions with biotinylated DNA is problematic, due to purification biases introduced by the hydrophobic nature of the biotin moiety [82-84]. However, this is likely to only be an issue for smaller DNA fragments, and can be overcome by replacing the phenol:chloroform purification steps with column purifications.

5. METHODS TO LOCALIZE PROTEINS AT CHROMOSOME INTERACTION SITES

3C, 4C and 5C were developed to detect chromosomal interactions and do not address the issue of whether the interactions are direct or mediated by a protein or nucleic acid. Enhanced-ChIP-4C (e4C) [85], 3C-ChIP-cloning (6C) [86], ChiA-PET [59] and variations thereof [87-89] were developed in an attempt to directly address this issue.

5.1. Combined 3C-ChIP-Cloning (6C)

6C combines the cross-linking, digestion, and intra-molecular ligation of chromatin with ChIP and cloning[86]. The incorporation of ChIP requires that *a priori* assumptions are made about the identity of the protein(s) that mediate the chromosomal interactions of interest. The detection of

chromosomal interactions places the antigen targeted by the ChIP within the localized neighborhood of the interaction. The size of the neighborhood is dictated by the fragmentation method (*e.g.* restriction enzyme) and the size of the resulting fragment(s).

A positive result in 6C stops short of proving that the antigen is directly involved in the establishment and/or maintenance of the specified interaction. Proof of involvement in the actual interaction requires the deletion, knock-down or inactivation of the protein of interest and subsequent testing for the presence of the interaction by 3C or another technique.

5.2. Chromatin Interaction Analysis Using Paired-End Tag Sequencing (Chia-PET)

ChiA-PET ([59, 90],[91] reviewed in [92]) was specifically developed to enable the large-scale, *de novo* analysis of higher-order chromatin structure [59]. ChiA-PET[59] uses a similar general strategy to 6C in order to place antigens within the vicinity of chromosomal interactions on a global scale (Figure 1g). As such, it differs from the other global 3C based technologies through the incorporation of a ChIP step. ChIP is introduced following chromatin fragmentation, to fractionate and enrich for cross-linked chromatin DNA segments bound to the protein of interest. The incorporation of this ChIP step reduces the level of complexity of the fragment library prior to the proximity-based ligation step. The proximity-based ligation step was modified to incorporate linkers containing restriction sites that enable the biotin mediated enrichment of the ~27 bp of sequence at the ends of the interacting DNA sequences (Figure 1g) [92]. Enrichment occurs prior to sequencing and hence reduces sample complexity. Thus, ChiA-PET enables the proximity of protein and chromosomal interactions to be determined, within the resolution of the restriction fragment size.

ChIA-PET and 6C are powerful techniques, yet while it is reasonable to assume that mechanisms exist to promote the correct association of interacting loci and prevent accidental interactions [9], there is no *a priori* reason to think that the formation of chromosomal interactions at different sites is mediated by the same signal, nucleic acid, or protein moieties [5, 6, 21, 22, 93]. Moreover, it is possible that 'accidental' interactions have functional roles in certain nuclear processes (*e.g.* DNA repair).

6. Analytical and Experimental Considerations for Global High-throughput Analyses

As with all methods, the proximity-based ligation methods have limitations. Some limitations are specific to particular methods (Table 1) while others are general. For example, all proximity-based ligation protocols rely upon the formation of intra-molecular ligations to capture inter- and intra-chromosomal interactions (Figure 1a). Unfortunately, the incorporation of high-throughput sequencing into the global methods necessitates the inclusion of a second ligation step as part of the sample preparation. Therefore, there are now two [26, 59, 61] or more steps [58] where ligation can potentially join non-cross-linked restriction fragments. Inter-molecular ligations do occur during all ligation steps albeit at low levels(unpublished observations). As such, it is important to include external controls for each of the ligation steps to determine the minimum numbers of sequence tags required to indicate a genuine chromosomal interaction.

Table 1. Advantages and disadvantages of the different molecular methods for the detection of chromosomal interactions

Method	Reference	Advantage(s)	Disadvantage(s)
3C	[51]	1) Detects inter- and intra-chromosomal interactions. 2) Uses both sticky and blunt end ligation.	1) Requires *a priori* predictions of DNA interactions. 2) Restricted to the detection of a single interaction per PCR reaction.
4C	[25]	Detects all DNA fragments interacting with a particular region.	1) Requires *a priori* selection of one interaction partner – the bait. 2) Limited to the identification of interactions with a single locus at a time.
5C	[27]	Allows the global identification of DNA interactions through multiplexing and scaling.	Does not scale well due to the requirement to design oligonucleotides that anneal adjacent to every restriction site in the genome.
6C	[86]	Determines if specific proteins bind in the vicinity of chromosomal interactions.	Requires *a priori* assumptions as to the identity of the proteins in the vicinity of the chromosomal interactions.
ChiA-PET	[59]	Global method to identify chromosomal interactions in the vicinity of specific proteins.	1) Requires *a priori* assumptions as to the identity of the proteins in the vicinity of the chromosomal interactions. 2) Requires the ligations of linker

			sequences before intra-molecular ligation.
GCC	[61]	A global un-biased method for the detection of chromosomal interactions.	Not suitable for the analysis of large genomes as it lacks an enrichment step.
Hi-C	[26]	1) A global un-biased method for the detection of chromosomal interactions. 2) Enriches for interacting DNA fragments.	1) Ligation efficiency is reduced due to the use of blunt end ligation. 2) Potential bias due to use of biotin moiety during DNA purification.

Internal controls[62, 67] are used to demonstrate the specificity of the intra-molecular ligations that occur in 3C and 4C analyses. However, the choice of internal controls during the global screening of a new genome, growth or developmental condition is problematic and relies on an empirical characterization using 3C, 4C or 5C prior to the global analysis. Prior confirmation of an internal control avoids the circularity of choosing a negative internal control from an analysis and using it to justify the results of the same analysis.

In order to determine the statistical significance of the high-throughput techniques two key questions must be addressed. Firstly, for a given number of sequence tags, how often does the interaction between two restriction fragments have to be recorded in order to be considered statistically significant? Based on the high number of fragments in the mixture and the low probability for two specific fragments to be ligated, the single ligation events can be considered as independent of each other. Therefore, the expected number of random ligations of two given fragments is binomially distributed. Based on this observation, an expected false positive rate can be calculated for each number of sequence tags per interaction. The statistical cut-off has to be chosen such that this false positive rate is sufficiently low[58, 61]. This is a conservative cut-off for the noise levels present within these experimental data sets, particularly given the preference for intra-molecular ligation that occurs between the cross-linked restriction fragments [63]. However, subsequent ligation steps are random, occurring in the absence of cross-linking. The inclusion of external controls (see above) demonstrates exact numbers of ligations that can be expected and the frequency with which they occur. It is important to bear in mind that even if the number of occurrences of a specific sequence tag is significant, when compared to random ligation, there is no biological significance ascribed to the observed interaction without further experimentation.

Secondly, how many sequence tags need to be identified in order to be sure that every biologically relevant interaction that can be captured is present within the data set? This coverage can be most simply addressed by demonstrating correlation in comparisons between biological replicates. However, it has been noted that these datasets correlate extremely well when classified into larger zones but not at the level of the individual fragments(unpublished observations and [26]). This limits the applicability of this technique to the estimation of coverage levels.

The choice of fragmentation method affects the resolution that can be achieved in all proximity-based ligation methodologies. If using restriction

enzymes, the choice of enzyme is dictated by genome size (*e.g.* four versus six nucleotide recognition sites), the distribution of sites (and consequently the number and sizes of the fragments), whether the enzyme is affected by overlapping modifications, and the user's preference for cohesive or non-cohesive ends. If fragmenting by sonication or nebulization, the conditions are dictated by the desired resolution and the genome size. Irrespective of how the fragments were generated, it is important to remember that each fragment can participate in a maximum of two ligation reactions at any one time. Therefore, there are ten potential products following a ligation reaction that involves two cross-linked restriction fragments. Six of these ligation products are diagnostic of a chromosomal interaction (Figure 2). However, there are at least 40 potential products resulting from the ligation of three cross-linked restriction fragments. In such a situation, the interactions that are identified will depend on the sizes of the restriction fragments and which products were sequenced. Hence, high coverage is required in order to identify sites of multiple fragment interaction.

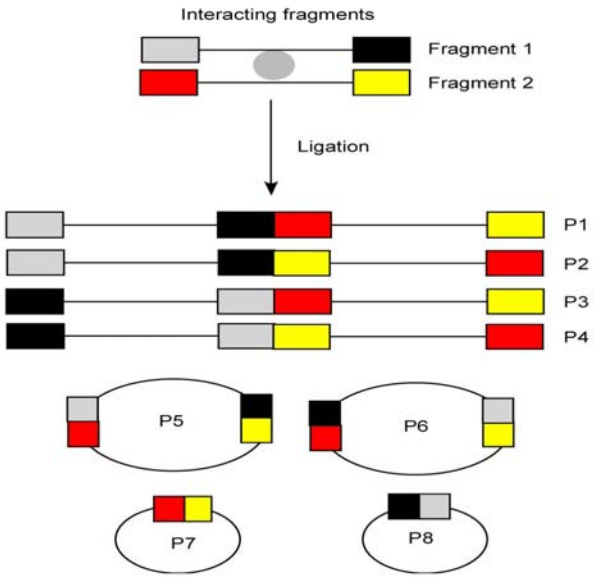


Figure 2. There are eight possible ligation products resulting from the intra-molecular ligation of two interacting restriction sequences. The ends of interacting fragments 1 and 2 are designated different colors to enable identification. Products are designated by P#. P9 and P10 are the linear unligated restriction fragments, as it is always possible that neither fragment is ligated.

By virtue of their design, global analyses interrogate the structure of the complete genetic complement present within an organism. This includes repetitive elements, plasmids, viruses, and the organelle genomes (*e.g.* mitochondrial and plastid genomes) present within the organism or cells under investigation. Repetitive elements pose a serious barrier to mapping the global structure of genome organization. The problems are two-fold: 1) as with all sequencing technologies repetitive sequences cannot be mapped to a unique position within a genome; and 2) the number of each repetitive sequence must be compensated for when determining the level at which the occurrence of specific interactions involving the element are significant or not.

Plasmids, viruses, and organelle genomes must be included within the reference genome against which the empirically derived sequence tags are mapped. The importance of this is illustrated by the finding that mitochondrial DNA interacts with the yeast nuclear genome[61]. At a minimum, failure to include the complete genome could result in the loss of interesting and potentially biologically meaningful interactions. At worst, it results in the identification of high frequency false positive interactions at specific genomic loci that share sequence similarity with the omitted genomes (*e.g.* nuclear mitochondrial insertions) [94]. This problem is particularly acute when combined with an allowance for mismatch errors during the positioning of the sequence tags. Complicated experimental processes have been recently developed to reduce the chances of this occurring by effectively reducing the genome complexity(*e.g.* [58]).

Finally, high-throughput methods for the analysis of global chromosome interaction networks typically use large populations of cells [26, 58, 59, 61]. Therefore, it is important to remember that the interaction networks that are produced are snapshots in time of the global average of synchronous or asynchronous cells. Moreover, the limitations of the current cross-linking techniques (*e.g.* formaldehyde see section 3.1) mean that only stable interactions (>5 seconds) [53] will be captured and represented in the network. As such, it cannot be determined if the identified interactions occur within one cell, or even if they co-exist or are mutually exclusive.

Software packages are available to aid in the planning and analysis of 5C [95], GCC [61], and ChIA-PET [96] experiments. However, with respect to analysis, these packages are designed to deconvolve the sequences that are obtained. This is only the first layer of the onion, considerable targeted bioinformatic analyses are subsequently required to convert the data into information that is able to be tested by hypotheses.

7. Conclusion and Future Prospects

We have presented an overview of the emerging molecular methods that are being used to study chromosome organization. Recent adaptations have been driven by the development of advanced sequencing technologies. At present, these techniques have largely been employed to screen for interactions and have yet to be incorporated into existing studies in a manner that enables identification of the functional aspects of chromosomal interactions. However, doing so is a non-trivial exercise that requires the incorporation of data on genome organization with novel and existing datasets (*e.g.* transcriptome and genome wide ChIP studies) in order to provide testable predictions of the functional roles and biological significance of inter- and intra-chromosomal interactions. It also requires a mind-shift towards the identification of three-dimensional motifs, similar to the arrangement of reactive residues within enzyme active sites, rather than the linear motifs that are typically associated with DNA binding proteins. There is also no reason to assume that all inter- and intra-chromosomal interactions are protein mediated, rather it is possible that some are random, or alternatively are directly mediated by nucleic acids (*e.g.* G-quadruplexes [97-103]).

In conclusion, proximity ligation techniques provide us with some of the tools required to decode the link between genome structure and function. However, there still remain many unanswered questions and we are far from achieving this ambitious goal.

8. Acknowledgments

The authors would like to apologize to colleagues whose work was not cited. Thanks to Dr Sue Mei Tan Wong for reviewing this chapter. The authors would also like to thank A. Ganley, R. McNab, and members of the O'Sullivan laboratory for comments on this chapter. LRG is funded by an SNSF fellowship (PBBSP3-130910). RSG is funded by a Massey University Ph.D. scholarship. Work in JMO's laboratory is funded by the Maurice & Phyllis Paykel trust and the Auckland Medical Research Foundation.

9. REFERENCES

[1] Cremer, T. and C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2001. 2(4): p. 292-301.

[2] Vazquez, J., A.S. Belmont, and J.W. Sedat, Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Curr Biol*, 2001. 11(16): p. 1227-39.

[3] Heun, P., et al., Chromosome dynamics in the yeast interphase nucleus. *Science*, 2001. 294(5549): p. 2181-6.

[4] Costa, S. and P. Shaw, Chromatin organization and cell fate switch respond to positional information in Arabidopsis. *Nature*, 2006. 439(7075): p. 493-6.

[5] Tan-Wong, S.M., H.D. Wijayatilake, and N.J. Proudfoot, Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev*, 2009. 23(22): p. 2610-24.

[6] Laine, J.P., et al., A physiological role for gene loops in yeast. *Genes Dev*, 2009. 23(22): p. 2604-9.

[7] Brickner, D.G., et al., *H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state.* PLoS Biol, 2007. 5(4): p. e81.

[8] Brickner, J.H., Transcriptional memory at the nuclear periphery. *Curr Opin Cell Biol*, 2009. 21(1): p. 127-33.

[9] Dobi, K.C. and F. Winston, Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. *Mol Cell Biol*, 2007. 27(15): p. 5575-86.

[10] Kite, G.L. and R. Chambers Jr, Vital staining of chromosomes and the function and structure of the nucleus. *Science*, 1912. 36(932): p. 639-641.

[11] Marshak, A., Chromosome structure in Escherichia coli. *Experimental Cell Research*, 1951. 2(2): p. 243-251.

[12] Haapala, O. and S. Nokkala, Structure of human metaphase chromosomes. *Hereditas*, 1982. 96(2): p. 215-228.

[13] Van Winkle, Q., et al., Electron microscopy of isolated chromosomes. *Experimental Cell Research*, 1953. 5(1): p. 38-47.

[14] Aula, P. and E. Saksela, Comparison of areas of quinacrine mustard fluorescence and modified Giemsa staining in human metaphase chromosomes. *Experimental Cell Research*, 1972. 71(1): p. 161-167.

[15] Boyle, S., et al., The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human Molecular Genetics*, 2001. 10(3): p. 211-219.
[16] Grond, C.J., J. Derksen, and G.J. Brakenhoff, The banding pattern of the segment 46A-48C in Drosophila hydei polytene chromosomes as studied by confocal scanning light microscopy (CSLM). *Experimental Cell Research*, 1982. 138(2): p. 458-466.
[17] Rodman, T.C., Human chromosome banding by Feulgen stain aids in localizing classes of chromatin. *Science*, 1974. 184(4133): p. 171-173.
[18] Takayama, S., Configurational changes in chromatids from helical to banded structures. *Chromosoma*, 1976. 56(1): p. 47-54.
[19] Pathak, S. and F.F.B. Elder, Silver-stained accessory structures on human sex chromosomes. *Human Genetics*, 1980. 54(2): p. 171-175.
[20] O'Sullivan, J.M., et al., Gene loops juxtapose promoters and terminators in yeast. *Nat Genet*, 2004. 36(9): p. 1014-8.
[21] Ansari, A. and M. Hampsey, A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*, 2005. 19(24): p. 2969-78.
[22] Singh, B.N. and M. Hampsey, A transcription-independent role for TFIIB in gene looping. *Mol Cell*, 2007. 27(5): p. 806-16.
[23] Palstra, R.J., et al., The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics*, 2003. 35(2): p. 190-194.
[24] Simonis, M., et al., Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 2006. 38(11): p. 1348-54.
[25] Zhao, Z., et al., Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 2006. 38(11): p. 1341-7.
[26] Lieberman-Aiden, E., et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009. 326(5950): p. 289-93.
[27] Dostie, J., et al., Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 2006. 16(10): p. 1299-309.
[28] Berger, A.B., et al., High-resolution statistical mapping reveals gene territories in live yeast. *Nat Methods*, 2008. 5(12): p. 1031-7.
[29] Cremer, C., et al., Preparative dual-beam sorting of the human Y chromosome and in situ hybridization of cloned DNA probes. *Cytometry*, 1984. 5(6): p. 572-579.

[30] Cremer, T., et al., Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. *Human Genetics*, 1988. 80(3): p. 235-246.

[31] Pinkel, D., et al., Fluorescence in situ hybridization with human chromosome-specific libraries: Detection of trisomy 21 and translocations of chromosome 4. *Proceedings of the National Academy of Sciences of the United States of America*, 1988. 85(23): p. 9138-9142.

[32] Lichter, P., et al., Rapid detection of human chromosome 21 aberrations by in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 1988. 85(24): p. 9664-9668.

[33] Volpi, E.V. and J.M. Bridger, FISH glossary: An overview of the fluorescence in situ hybridization technique. *BioTechniques*, 2008. 45(4): p. 385-409.

[34] Hildenbrand, G., et al., Nano-sizing of specific gene domains in intact human cell nuclei by spatially modulated illumination light microscopy. *Biophys J*, 2005. 88(6): p. 4312-8.

[35] Baddeley, D., et al., Nanostructure analysis using spatially modulated illumination microscopy. *Nat Protoc*, 2007. 2(10): p. 2640-6.

[36] Lemmer, P., et al., SPDM: light microscopy with single-molecule resolution at the nanoscale. *Applied Physics B-Lasers and Optics*, 2008. 93(1): p. 1-12.

[37] Ribeiro, S.A., et al., A super-resolution map of the vertebrate kinetochore. *Proceedings of the National Academy of Sciences*. 107(23): p. 10484-10489.

[38] Huang, B., et al., Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 2008. 319(5864): p. 810-3.

[39] Jiang, H., et al., Quantitative 3D imaging of whole, unstained cells by using X-ray diffraction microscopy. Proc Natl Acad Sci U S A.

[40] Meister, P., et al., Visualizing Yeast Chromosomes and Nuclear Architecture. Methods in Enzymology, Vol 470: *Guide to Yeast Genetics*:, 2010. 470: p. 535-567.

[41] Straight, A.F., et al., GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr Biol*, 1996. 6(12): p. 1599-608.

[42] Robinett, C.C., et al., In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac

operator/repressor recognition. *Journal of Cell Biology*, 1996. 135(6 II): p. 1685-1700.

[43] Schober, H., et al., Yeast telomerase and the SUN domain protein Mps3 anchor telomeres and repress subtelomeric recombination. *Genes Dev*, 2009. 23(8): p. 928-38.

[44] Wang, X., et al., Rapid telomere motions in live human cells analyzed by highly time-resolved microscopy. *Epigenetics Chromatin*, 2008. 1(1): p. 4.

[45] Levi, V., et al., Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophys J*, 2005. 89(6): p. 4275-85.

[46] Schober, H., et al., Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res*, 2008. 18(2): p. 261-71.

[47] Cabal, G.G., et al., SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, 2006. 441(7094): p. 770-3.

[48] Therizols, P., et al., Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc Natl Acad Sci U S A*, 2010. 107(5): p. 2025-30.

[49] Cullen, K.E., M.P. Kladde, and M.A. Seyfred, Interaction between Transcription Regulatory Regions of Prolactin Chromatin. *Science*, 1993. 261(5118): p. 203-206.

[50] Mukherjee, S., H. Erickson, and D. Bastia, *Detection of DNA Looping Due to Simultaneous Interaction of a DNA-Binding Protein with 2 Spatially Separated Binding-Sites on DNA.* Proceedings of the National Academy of Sciences of the United States of America, 1988. 85(17): p. 6287-6291.

[51] Dekker, J., et al., Capturing chromosome conformation. *Science*, 2002. 295(5558): p. 1306-11.

[52] Jackson, V., Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods: A Companion to Methods in Enzymology*, 1999. 17(2): p. 125-139.

[53] Schmiedeberg, L., et al., A Temporal Threshold for Formaldehyde Crosslinking and Fixation. *PLoS ONE*, 2009. 4(2): p. -.

[54] Fujita, N. and P.A. Wade, Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. *Methods*, 2004. 33(1): p. 81-5.

[55] Gilli, G. and P. Gilli, *The nature of the Hydrogen bond: Outline of a Comprehensive Hydrogen Bond Theory* International Union of Crystallography Monographs on Crystallography, 23. 2009, New York: Oxford University Press. 304.

[56] Jackson, V., *Studies on Histone Organization in Nucleosome Using Formaldehyde as a Reversible Cross-Linking Agent.* Cell, 1978. 15(3): p. 945-954.

[57] Solomon, M.J. and A. Varshavsky, Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci U S A*, 1985. 82(19): p. 6470-4.

[58] Duan, Z., et al., A three-dimensional model of the yeast genome. *Nature*, 2010. 465: p. 363-367.

[59] Fullwood, M.J., et al., An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, 2009. 462(7269): p. 58-64.

[60] O'Sullivan, J.M., et al., Repeated elements coordinate the spatial organization of the yeast genome. Yeast, 2009. 26(2): p. 125-138.

[61] Rodley, C.D., et al., Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genet Biol*, 2009. 46(11): p. 879-86.

[62] Splinter, E., F. Grosveld, and W. de Laat, 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol*, 2004. 375: p. 493-507.

[63] Sambrook, J., and Russell, D. W., Molecular cloning: A laboratory manual. 2001, Cold Spring Harbor Laboratory Press, Clod Spring Harbor, New York. p. 157.

[64] Sugino, A., H.M. Goodman, and H.L. Heyneker, Interaction of bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base paired ends. *Journal of Biological Chemistry*, 1977. 252(11): p. 3987-3994.

[65] Rippe, K., Making contacts on a nucleic acid polymer. *Trends Biochem Sci*, 2001. 26(12): p. 733-40.

[66] Ringrose, L., et al., Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *Embo J*, 1999. 18(23): p. 6630-41.

[67] Dekker, J., The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods*, 2006. 3(1): p. 17-21.

[68] van Berkum, N.L. and J. Dekker, Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods Mol Biol*, 2009. 567: p. 189-213.

[69] Wiedmann, M., et al., Ligase chain reaction (LCR)--overview and applications. *Genome Research*, 1994. 3(4): p. S51-S64.
[70] Shoshani, T., et al., Cis-regulatory elements that confer differential expression upon the rat gene encoding phosphoenolpyruvate carboxykinase in kidney and liver. *Gene*, 1991. 101(2): p. 279-283.
[71] Tolhuis, B., et al., Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, 2002. 10(6): p. 1453-65.
[72] Liu, Z. and W.T. Garrard, Long-range interactions between three transcriptional enhancers, active V{kappa} gene promoters, and a 3' boundary sequence spanning 46 kilobases. *Molecular and Cellular Biology*, 2005. 25(8): p. 3220-3231.
[73] Murrell, A., S. Heeson, and W. Reik, Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature Genetics*, 2004. 36(8): p. 889-893.
[74] Spilianakis, C.G., et al., Interchromosomal associations between alternatively expressed loci. *Nature*, 2005. 435(7042): p. 637-645.
[75] Lomvardas, S., et al., Interchromosomal Interactions and Olfactory Receptor Choice. *Cell*, 2006. 126(2): p. 403-413.
[76] Stein, G.S., et al., Functional architecture of the nucleus: Organizing the regulatory machinery for gene expression, replication and repair. *Trends in Cell Biology*, 2003. 13(11): p. 584-592.
[77] Goetze, S., J. Mateos-Langerak, and R. van Driel, Three-dimensional genome organization in interphase and its relation to genome function. *Seminars in Cell and Developmental Biology*, 2007. 18(5): p. 707-714.
[78] Simonis, M. and W. de Laat, FISH-eyed and genome-wide views on the spatial organisation of gene expression. *Biochimica et Biophysica Acta - Molecular Cell Research*, 2008. 1783(11): p. 2052-2060.
[79] Bartkuhn, M. and R. Renkawitz, Long range chromatin interactions involved in gene regulation. *Biochimica et Biophysica Acta - Molecular Cell Research*, 2008. 1783(11): p. 2161-2166.
[80] Goetze, S., et al., The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Molecular and Cellular Biology*, 2007. 27(12): p. 4475-4487.
[81] Cook, P.R., The organization of replication and transcription. *Science*, 1999. 284(5421): p. 1790-5.
[82] Langer, P.R., A.A. Waldrop, and D.C. Ward, Enzymatic synthesis of biotin-labeled polynucleotides: Novel nucleic acid affinity probes.

Proceedings of the National Academy of Sciences of the United States of America, 1981. 78(11 II): p. 6633-6637.

[83] Laman, A.G., et al., Subtractive hybridization of biotinylated DNA in phenol emulsion. *Journal of Biochemical and Biophysical Methods*, 2001. 50(1): p. 43-52.

[84] Lucas, J.N., et al., An efficient chemical method to generate repetitive sequences depleted DNA probes. *American Journal of Medical Genetics, Part A*, 2006. 140(19): p. 2115-2120.

[85] Schoenfelder, S., et al., Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics*. 42(1): p. 53-61.

[86] Tiwari, V.K., et al., A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Research*, 2008. 18(7): p. 1171-1179.

[87] Cai, S., C.C. Lee, and T. Kohwi-Shigematsu, SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*, 2006. 38(11): p. 1278-88.

[88] Carroll, J.S., et al., Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 2005. 122(1): p. 33-43.

[89] Horike, S., et al., Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, 2005. 37(1): p. 31-40.

[90] Fullwood, M.J. and Y. Ruan, ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem*, 2009. 107(1): p. 30-9.

[91] Fullwood, M.J., et al., Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res*, 2009. 19(4): p. 521-32.

[92] Fullwood, M.J., et al., Chromatin interaction analysis using paired-end tag sequencing. *Curr Protoc Mol Biol*. Chapter 21: p. Unit 21 15 1-25.

[93] Bartlett, J., et al., Specialized transcription factories. *Biochem Soc Symp*, 2006(73): p. 67-75.

[94] Ricchetti, M., F. Tekaia, and B. Dujon, Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol*, 2004. 2(9): p. E273.

[95] Lajoie, B.R., et al., My5C: web tools for chromosome conformation capture studies. Nat Methods, 2009. 6(10): p. 690-1.

[96] Li, G., et al., ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol*. 11(2): p. R22.
[97] Johnson, J.E., et al., In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. *Biochimie*, 2008. 90(8): p. 1250-63.
[98] Huppert, J.L. and S. Balasubramanian, G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research*, 2007. 35(2): p. 406-413.
[99] Waller, Z.A.E., et al., A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. *Journal of the American Chemical Society*, 2009. 131(35): p. 12628-12633.
[100] Rawal, P., et al., Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Research*, 2006. 16(5): p. 644-655.
[101] Hershman, S.G., et al., Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucl. Acids Res*., 2008. 36(1): p. 144-156.
[102] Eddy, J. and N. Maizels, Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Research*, 2006. 34(14): p. 3887-3896.
[103] Du, Z., et al., Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochemical and Biophysical Research Communications*, 2007. 354(4): p. 1067-1070.
[104] Ling, J.Q., et al., CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*, 2006. 312(5771): p. 269-72.

8 APPENDIX II

Nucleic Acids Research Advance Access published April 30, 2013

Nucleic Acids Research, 2013, 1–14
doi:10.1093/nar/gkt325

Genome conformation capture reveals that the *Escherichia coli* chromosome is organized by replication and transcription

Cedric Cagliero[1], Ralph S. Grand[2,3], M. Beatrix Jones[2], Ding J. Jin[1] and Justin M. O'Sullivan[3,*]

[1]Gene Regulation and Chromosome Biology Laboratory, Frederick National Laboratory for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD 21702, USA, [2]Institute of Natural and Mathematical Sciences, Massey University, Auckland 0745, New Zealand and [3]Liggins Institute, University of Auckland, Auckland 1023, New Zealand

Received March 3, 2013; Revised April 7, 2013; Accepted April 8, 2013

ABSTRACT

To fit within the confines of the cell, bacterial chromosomes are highly condensed into a structure called the nucleoid. Despite the high degree of compaction in the nucleoid, the genome remains accessible to essential biological processes, such as replication and transcription. Here, we present the first high-resolution chromosome conformation capture-based molecular analysis of the spatial organization of the *Escherichia coli* nucleoid during rapid growth in rich medium and following an induced amino acid starvation that promotes the stringent response. Our analyses identify the presence of origin and terminus domains in exponentially growing cells. Moreover, we observe an increased number of interactions within the origin domain and significant clustering of SeqA-binding sequences, suggesting a role for SeqA in clustering of newly replicated chromosomes. By contrast, 'histone-like' protein (i.e. Fis, IHF and H-NS) -binding sites did not cluster, and their role in global nucleoid organization does not manifest through the mediation of chromosomal contacts. Finally, genes that were downregulated after induction of the stringent response were spatially clustered, indicating that transcription in *E. coli* occurs at transcription foci.

INTRODUCTION

Our understanding of the spatial organization of bacterial genomes and its relationship to cellular function is limited [for reviews see (1–3)]. Yet it is clear that despite not being enclosed in a nuclear membrane, bacterial nucleoids are spatially organized within a defined sub-fraction of the cell volume (4–11). Various molecular [reviewed in (2)] and recombination-based methodologies have been used to identify the existence of micro- and macrodomains within the *Escherichia coli* nucleoid [e.g. (2,5,8,12,13)]. The four structured macrodomains (~0.5–1 Mb) that have been identified exhibit preferential intra-domain recombination between λ att sites, whereas inter-domain recombination is reduced (5,7,8,12,13). By contrast, microdomains are much smaller (average ~10 kb) and have been linked to the topological isolation of supercoils (2,10). Collectively, micro- and macrodomains are hypothesized to be critical for maintaining global organization while enabling the local levels of compaction required to fit a circular chromosome with an extended diameter of ~490 nm within a cell with a length as small as 1000 nm (2).

Unlike eukaryote chromatin, the bacterial nucleoid does not contain histones. However, nucleoid-associated proteins (NAPs), particularly histone-like proteins, such as histone-like nucleoid structuring (H-NS) protein, heat unstable nucleoid protein (HU), factor for inversion stimulation (Fis) and integration host factor (IHF), are believed to act like histones and play a significant role in the organization of the nucleoid (14–17). These NAPs exhibit DNA bending, looping and bridging properties *in vitro*. However, studies also indicate that *in vivo*, the role of the NAPs could be more regulatory than architectural [e.g. (18,19)]. Non-classical NAPs (i.e. SeqA, SlmA and MatP) have been recently characterized as exhibiting macrodomain-specific DNA-binding properties [reviewed in (16)] and may represent alternative candidates for organizational roles within the nucleoid.

*To whom correspondence should be addressed. Tel: +64 99239868; Fax: +64 9 373 8763; Email: justin.osullivan@auckland.ac.nz

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

The structure of the bacterial nucleoid is dynamic and affected by growth conditions and stress (15,20–23). For example, the relatively compact nucleoid present in fast growing cells is altered by treatment with serine hydroxamate (SHX), which induces the stringent response (24) and inhibits replication initiation through artificial amino acid starvation. In terms of the biology of the *E. coli* nucleoid, the overall effect of the SHX-induced amino acid starvation is an expansion of the nucleoid and a change in transcription patterns (25,26). This suggests a relationship between transcription and the organization of the nucleoid (27). However, the mechanism(s) behind the re-structuring of the nucleoid in response to growth and stress is still largely unknown.

Another long standing question is when and how the nascent nucleoid that arises from DNA replication segregates during bacterial cell growth [reviewed in (1)]. In *E. coli*, the time required for the replication of the nucleoid is fixed at ~40 min (28). To maintain a fast growth rate, cells growing in rich media must initiate multiple rounds of replication before each division. Consequently, a typical cell growing in rich media contains up to 16 origins of replication (29). Whether the nascent nucleoids segregate rapidly (30–32) or remain associated after replication, by a cohesion-dependent mechanism (i.e. the cohesion model) as seen in eukaryotes (33,34), remains unresolved.

Advances in chromosome conformation capture (3C)-related methodologies (35) enable the direct high-resolution detection of chromosome organization [e.g. (36–40)]. Recently, chromosome conformation capture carbon-copy (5C) was used to generate a global DNA:DNA contact map for *Caulobacter crescentus* synchronized swarmer cells (9). Here, we present a high-resolution analysis of the DNA:DNA interactions within *E. coli* nucleoids in rapidly growing and starved cell populations. Using genome conformation capture (GCC), we observe a clear relationship between DNA:DNA interactions, copy number and DNA replication. This suggests that nucleoids remain associated after replication, consistent with the cohesion model. Furthermore, SeqA-binding sites exhibit replication-dependent clustering, whereas binding sites for the major histone-like proteins (Fis, H-NS and IHF) did not. Finally, we observe a correlation between gene regulation and spatial clustering.

MATERIALS AND METHODS

Strains and growth conditions

For GCC analyses (36), *E. coli* strains (Supplementary Table S1) were recovered from −80°C on Luria Bertani (LB) agar (2%) plates (24 h, 37°C). LB medium (3 ml, Gibco) starter cultures were inoculated and grown (37°C, 220 rpm, 16 h). The optical density (OD_{600}) of cultures was measured and used to inoculate LB test cultures to an OD_{600} of ~0.02. The test cultures were grown (37°C, 220 rpm) until the OD_{600} reached ~0.25, and the cells were harvested. For the SHX-treated samples, the cultures were treated with SHX (500 µg/ml, 30 min) before harvesting.

Genome conformation capture

E. coli chromatin was prepared according to Rodley *et al.* (36) with minor modifications. In brief, 5×10^9 formaldehyde cross-linked (1%) cells were lysed (Supplementary Materials and Methods) in the presence of protease inhibitor (Roche), and the chromatin was collected (21 500*g*, 20 min, 4°C). Chromatin was washed and suspended in chromatin digestion buffer (10 mM Tris–HCl, pH 8.0, 5 mM $MgCl_2$ and 0.1% TritonX-100). Chromatin samples were digested with HhaI (100 U, New England Biolabs), diluted (~20-fold) and ligated with T4 DNA ligase (20 U, Invitrogen). A ligation control was added to the digested chromatin (Supplementary Materials and Methods and Supplementary Table S2) before ligation. After ligation, cross-links, protein and RNA were removed. pUC19 plasmid was added as a sequencing control before three extractions with 1:1 phenol:chloroform. DNA was column purified (Zymo, DNA clean and concentratorTM^{-5} kit) according to the manufacturer's instructions and eluted in milliQ H_2O. Three micrograms of purified DNA was sent for paired-end sequencing (100 bp) at the ATC sequencing facility (Rockville, MD, USA) on an Illumina Hi-Seq.

Genome conformation capture network assembly, effects of sample production and processing and bioinformatics analysis

To identify interacting DNA fragments from the paired-end sequence reads, network assembly was performed using the Topography suite v1.19 (41). GCC networks were constructed from 100-bp paired-end Illumina Genome Analyser sequence reads (Supplementary Materials and Methods). Except where indicated, bioinformatics and statistical analyses were performed on interactions identified by sequence reads that were uniquely mapped onto the reference genome and were above the cut-off value derived from the ligation control interactions (Supplementary Materials and Methods). A breakdown of the interactions present in the *E. coli* samples is provided in Supplementary Table S3. The effect of bar-coding, sequencing lane and biological replicates on the correlation between samples was quantified using the Cohen's Kappa statistic, showing that these factors did not strongly affect sample correlations (Supplementary Materials and Methods). All bioinformatics analysis was performed using in house Perl and Python scripts (Supplementary Materials and Methods). Except where indicated, statistical analyses were performed in R (42).

Genome copy number

Copy number was determined across the *E. coli* genome using control-free copy number and genotype caller (Control-FREEC) (43). The *E. coli* input sequences were in the SAM format, genome length was set at 4 639 675 bp, window size = 1000 and telocentromeric = 0. The GC profile was calculated and included.

Transcription microarray

Briefly, similar to GCC, *E. coli* was grown in LB (Gibco, lot 817849) to an OD_{600} ~0.2 and harvested directly, or

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

first treated with SHX before RNA isolation. RNA was isolated using hot phenol and finally suspended in DEPC-treated water (Invitrogen). The cDNA libraries were constructed using a SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) and sent to Roche-Nimblegen for microarray hybridization. Each experiment (exponential or SHX) is a pool of three biological replicates. A total of two technical replicates were performed per condition (exponential and SHX). Genes that were significantly up- or downregulated in SHX-treated compared with exponential samples were identified by calculating the log2 of the SHX/exponential ratio (Supplementary Materials and Methods and Supplementary Tables S4 and S5).

MatS, SeqA, SlmA and NAP clustering analyses

NAP-binding sites were obtained from Grainger *et al.* (18). MatP-binding sites (MatS) were obtained from Mercier *et al.* (5). Regions for analysis were defined by taking a specified number of bases (50, 100 or 250 bp) either side of the peak binding position for NAPs or center of the MatP-binding site for MatS. For SeqA, the strongest 135 confirmed SeqA-binding sites were obtained from Sanchez-Romero *et al.* (44), and the 24 defined SlmA-binding sites were obtained from Cho *et al.* (45). To determine whether these regions could be found in a different interacting environment compared with what would be expected by random chance, the total number of interactions with each of the individual regions and the number of interactions that occurred between the regions of interest (i.e. clustering) was determined from our GCC interaction network. We then generated 1000 random data sets of the same number and length (bp) as the actual region data set using two methods: (i) randomly selecting a start position for each region and then making it the same length as the region for which the random coordinate was being generated [i.e. random spacing (RS)]; or (ii) randomly select the start position for the first region and then sequentially determining the start and end position of all the other regions in the set such that the linear distances between regions were maintained [i.e. conserved linear spacing (CLS)]. This ensured that the particular interaction frequencies we observed were not because of the linear arrangement of the regions around the circular genome. One thousand random data sets were generated for the RS and CLS methods, and the total interaction and clustering frequencies were calculated from our GCC interaction network. The frequency with which the total interaction and clustering frequency of the actual data was higher or lower than the random data sets was used to estimate significance.

Interactions and clustering of genes that significantly change their expression level on SHX treatment

Genomic coordinates of genes that significantly change their expression level on treatment with SHX were obtained from http://regulondb.ccg.unam.mx/data/GeneProductSet.txt. The total number of interactions with each of the individual genes and the number of interactions that occurred between the genes of interest was determined as for MatS, SeqA, SlmA and NAP clustering, as described earlier in the text.

RESULTS

In GCC, the spatial organization of the nucleoid is captured by formaldehyde cross-linking within intact cells before cell lysis and the isolation of the nucleoid (Figure 1A). Once isolated, the nucleoid is digested, diluted and incubated with DNA ligase to enable the capture of spatially proximate but linearly separated loci (Figure 1A) (36). This produces an interaction library that can be sequenced to identify the network of chromosomal interactions occurring at the moment of cross-linking. GCC differs from current competing unbiased 3C technologies in that all DNA material is sequenced without the previous selection of DNA fragments containing ligation products. Therefore, there are no enrichment introduced biases, and DNA copy variation can be determined.

GCC relies on the intra-molecular ligation of cross-linked loci. However, inter-molecular ligation events resulting from random associations during the procedure can also occur, leading to false positives. To reduce the chances of isolating false positives, we (i) induce expansion of the nucleoid by isolation in a high-salt environment [a 'high-salt nucleoid' (2)], following cross-linking of the interacting loci; and (ii) added external ligation controls during GCC library preparations to empirically measure the background level of random inter-molecular ligation events. Thus, we determined a cut-off, for the minimum number of sequences representing any one interaction, above which interactions were deemed significant (Supplementary Materials and Methods). The following analyses were only performed on interactions that were above this significance threshold.

Origin and terminus domains exist within the *E. coli* nucleoid

Chromosome interaction networks were determined for rapidly growing cells in rich medium harvested at early exponential phase and exponential cells treated with SHX (Figure 1B and C). The exponential phase chromosome interaction network (Figure 1B) is dominant in two regions: (i) a high frequency interaction domain surrounding the origin (Ori); and (ii) a low frequency interaction domain surrounding the terminus (Ter). These Ori and Ter domains are also present in the interaction network for the SHX-treated samples, although they are less pronounced (Figure 1C). Higher resolution (i.e. 20 kb) emphasizes that the exponential phase interaction network contains regions that have a demonstrably lower average interaction frequency than the adjacent Ori and Ter domains (Figure 1D). We attribute these reductions to the presence of non-fixed domain boundaries within the population. We predicted that these boundaries would reduce interactions between domains, and that this would be manifested as a reduction in the interactions that cross the boundary regions. However, despite the obvious Ori preference, there is no sharp reduction in the numbers of interactions that cross our apparent domain boundaries (Figure 1E). Despite the diffuse boundaries for the Ori and Ter domains, we observe several noticeable reductions in the interaction frequency at various locations in the chromosome that could represent additional domain boundaries.

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

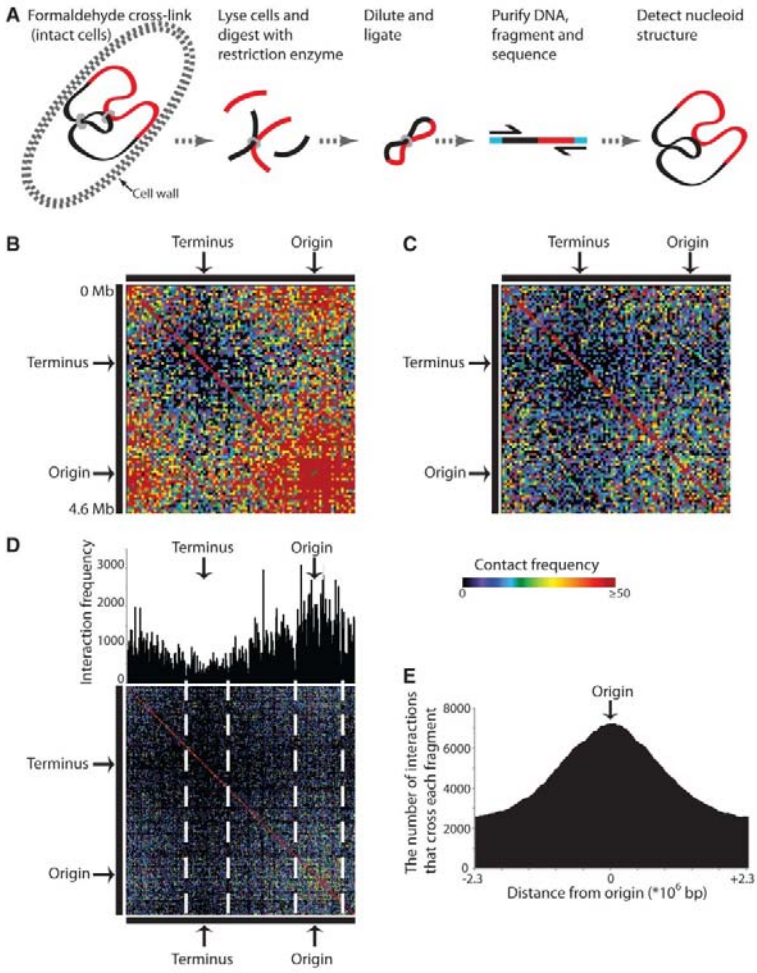


Figure 1. Ori and Ter domains are present within the *E. coli* nucleoid. (**A**) Schematic of the GCC procedure (36). Intact cells are cross-linked with formaldehyde before lysis, and the cross-linked nucleoids are isolated. The nucleoids are restriction digested, diluted and ligated to generate an interaction library. The interaction library is sequenced, after the addition of sequencing adapters (blue bars), and the network of interactions that define the nucleoid organization is determined. (**B**) Genome-wide contact matrix (50-kb bins) for exponentially growing *E. coli* nucleoids. The matrix highlights the Ori (high contact region) and Ter domains (low contact region). (**C**) Genome-wide contact matrix (50-kb bins) for nucleoids isolated from SHX-treated *E. coli*. The Ori and Ter domains remain visible. (**D**) Genome-wide contact matrix (20-kb bins) and bar graph for exponentially growing nucleoids highlighting regions of low interaction frequency ('domain boundaries') surrounding the Ori and Ter regions. (**E**) Frequency of exponential phase interactions that cross each restriction fragment plotted as a function of distance from the Ori (0). Fixed boundaries are not observed. The profile for the SHX-treated cells is not different (data not shown).

Interactions within the Ori and Ter regions are linked to replication

Comparisons of the chromosome networks from the exponential and SHX-treated cells identified similar levels of self and adjacent interactions (Supplementary Table S3). However, SHX treatment results in fewer long distance interactions (between 800 bp and half the length of the genome, respectively; Supplementary Figure S1A), shorter loop lengths (Supplementary Figure S1B) and reduced numbers of partners per fragment (Figure 2A and Supplementary Figure S2) when compared with the exponential network. These observations are consistent with SHX, decreasing the overall compaction of the nucleoid (21–23).

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

The high frequency of replication initiation in rapidly growing cells leads to an enrichment of origin-proximal loci, which could explain the pronounced increase in the number of partners observed in this region in exponentially growing cells (Supplementary Figure S2A). By contrast, treatment with SHX reduces this bias (Supplementary Figure S2B). These results are consistent with the inhibition of replication initiation after SHX treatment leading to a reduction in the Ori:Ter copy number ratio (46) or structural alterations within the origin domain.

To investigate whether interaction frequencies are affected by differences in copy number across the bacterial chromosome because of DNA replication, we compared interaction patterns and copy number before and after SHX treatment. Interactions were grouped according to the linear distance between the interacting loci and occurrence in the different environmental conditions (Figure 2B and Supplementary Table S3). The distribution of interaction strength and copy number relative to the origin was determined (Figure 2C–K). Exponential phase-specific and shared short distance interactions correlate with copy number (Figure 2C, D and F). By contrast, SHX-specific or shared long distance interactions do not correlate with copy number (Figure 2E, G and H). Critically, the ratio of Ori to Ter regions within both the exponential and SHX conditions remains at 3:1 (compare copy number Figure 2C and E). Thus, the observed decrease in the frequency of the interactions within the origin domain (compare Figure 1B and C) is either because of a decrease in the absolute number of origin sequences or because of a structural alteration (e.g. expansion) of the Ori domain.

Correcting the frequency of long distance interactions by copy number, a feature of GCC, indicates that most genomic regions interact with similar frequencies within the exponential-specific and shared interaction sets (i.e. interactions that occur in both the exponential and SHX conditions; Figure 2I and J). However, there are several notable deviations from this trend (labeled peaks within Figure 2I and J). The observed deviations are due to interactions involving multiple fragments within each of the 10 000-bp segments that are plotted (Figure 2I and J). By contrast, copy number correction of the long distance SHX-specific interactions identifies an increase in the interaction frequency within the Ter domain. The remainder of the genome shows relatively even and low interaction frequencies within the SHX-specific interaction set (Figure 2K).

Clustering of MatP- and SeqA-binding sites links nucleoid structure and replication

To further investigate the link between replication and nucleoid organization, we determined the clustering and interaction properties of loci containing characterized protein-binding sites for the MatP, SlmA and SeqA proteins.

MatP is a protein that binds to matS sites and organizes the Ter macrodomain (5). Analyses of matS loci identify significantly ($P < 0.008$) high clustering (i.e. inter-matS loci interactions) within the exponentially growing cells (Supplementary Table S6). In contrast, clustering of matS sites was not detected in the SHX-treated cells. The clustering in the exponentially growing condition was attributed to a single specific interaction between matS10 and matS5 (Figure 3A). This interaction must result from intra- or inter-Ter associations of these matS sites (Figure 3A i–iv).

The finding that SeqA binds as a dimer, which multimerizes to form a left-handed filament [reviewed in (47)], suggests that this protein may link spatially separated binding sites. Clustering of the 135 strongest confirmed SeqA-binding sites present within exponentially growing *E. coli* (44) was significantly higher than the random set ($P < 0.05$) (Supplementary Table S7). Moreover, these sites are significantly more prone to interact with other loci than random sites ($P < 0.05$; Supplementary Table S7). Visualizing the positions of the SeqA–SeqA interactions that formed within the *E. coli* genome showed that they tend to occur toward, and involve, the Ori domain in exponential cells (Figure 3B and C). SeqA interactions that are shared between exponential and SHX-treated nuclei predominantly link the left and right replichores (Figure 3C). By contrast, cells treated with SHX have a reduction in clusters involving SeqA sites surrounding the Ori domain and more inter-replichore interactions toward the terminal domain (Figure 3C and D). This is consistent with the progression of active replication forks that were initiated before SHX treatment.

SlmA binds at 24 defined sites within the genome (45) and acts to prevent FtsZ polymerization and premature cell division before complete chromosome replication. Analyses of the clustering and interaction profiles of *E. coli* SlmA sites demonstrated that clustering of these sites was not different from that observed for randomly selected sites (Supplementary Table S8). However, SlmA sites did exhibit a significantly increased propensity to interact with other genomic loci ($P < 0.05$) compared with randomly spaced elements for both exponential and SHX-treated cells (Supplementary Table S8). The significant increase in interaction frequency was lost when comparisons were made with random sets that have conserved linear spacing (Supplementary Table S8). Note that the differences observed in significance when the test data set was compared with randomly generated data sets (i.e. RS or CLS) confirm that the linear spacing of *E. coli* loci is important. Whether this is an effect or cause of spatial organization remains to be determined.

Intra- or inter-NAP–binding site clustering does not contribute to the global organization of the *E. coli* nucleoid

We investigated the clustering and interaction properties of H-NS-, IHF- and Fis-binding sites, which are not enriched in any particular macrodomain. There is no detectable clustering for the 200-bp regions surrounding the Fis-, H-NS- and IHF-binding sites in either the exponential or SHX-treated nucleoids (Table 1). Moreover, the classical NAP-binding sites have depleted levels of interactions in exponentially growing *E. coli* cells (Table 1).

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

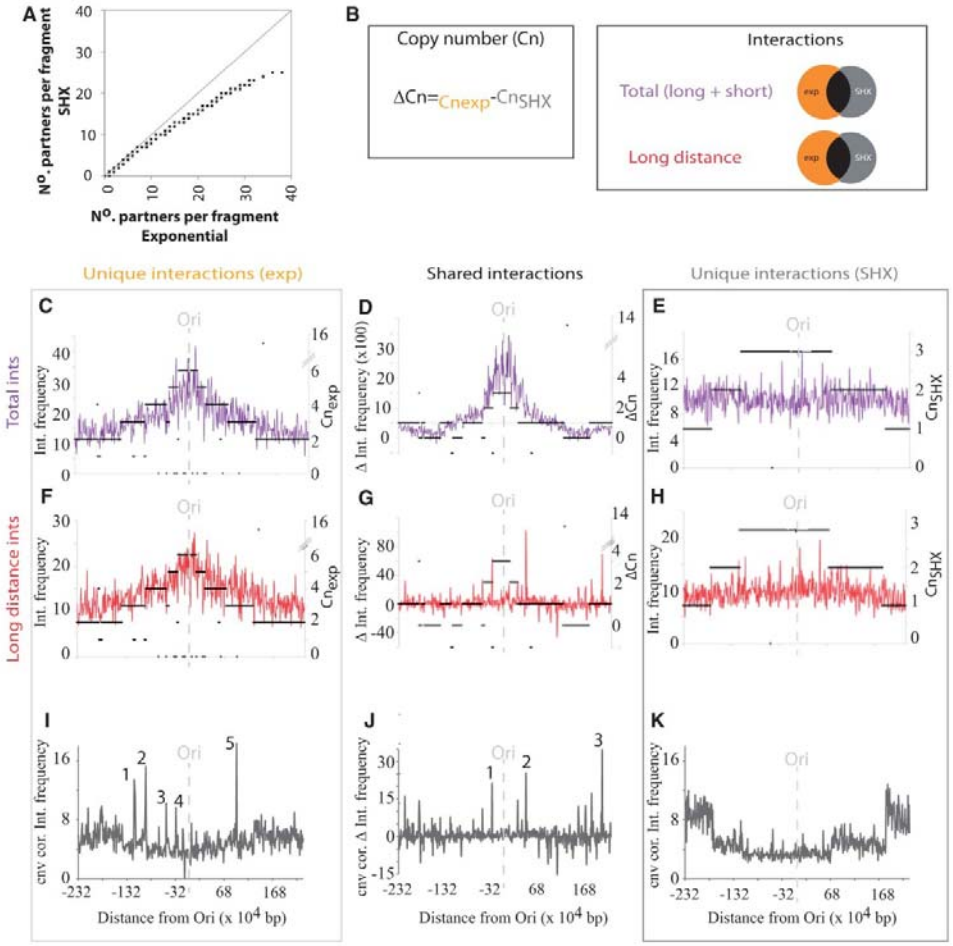


Figure 2. Origin proximal interactions are more frequently detected. **(A)** Fragments that interact have more partners in the exponential nucleoids as opposed to SHX-treated nucleoids. The 45° line shows the expected pattern if the number of partners for each fragment is equal in both conditions. **(B)** Schematic of the copy number and interaction comparisons that were performed. Comparisons between interaction frequency and copy number: **(C–E)**, total observed interactions; **(F–H)**, long distance (>800 bp) interactions. (C) Interactions that are specific to exponential phase growth correlated with copy number. (D) Differences in frequency for shared interactions between exponentially growing and SHX-treated *E. coli* cells indicate a correlation with copy number. (E) Interactions that were specific to SHX-treated cells are copy number independent. (F) Exponential phase-specific long distance interactions correlated with copy number. (G) Removal of short distance interactions (≤800 bp) removed the copy number dependence of the shared interactions. (H) SHX-specific interactions were independent of copy number. **(I)** Correction of exponential-specific long distance interactions identifies five peaks (I1–5) of increased interactions at positions (I1) 2 753 883–2 773 883, (I2) 2 983 883–3 003 883, (I3) 3 413 883–3 423 883, (I4) 3 613 883–3 623 883 and (I5) 224 208–234 208 bp. **(J)** Correction of shared long distance interactions identifies three peaks (J1–3) of increased interactions at positions (J1) 3 643 883–3 653 883, (J2) 4 383 883–4 393 883 and (J3) 1 404 208–1 414 208 bp. **(K)** Correction of SHX-specific long distance interactions for copy number identifies a decrease in the relative frequency of interactions at the origin compared with the terminus. Interactions were tallied for 10 000-bp bins and corrected for the number of fragments per bin. Vertical, gray broken lines denote the position of the origin of replication. Copy number is depicted by black horizontal bars.

These results can be explained by restrictions in the flexibility of the DNA (and, hence, reduced ligation efficiencies) because of the binding of the NAP. However, increasing the length of the region surrounding the binding site has no effect on the clustering (data not shown). Additionally, we do not observe intra-NAP–binding site clustering (Table 1), consistent with the temporal isolation of the expression of these NAPs (48).

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

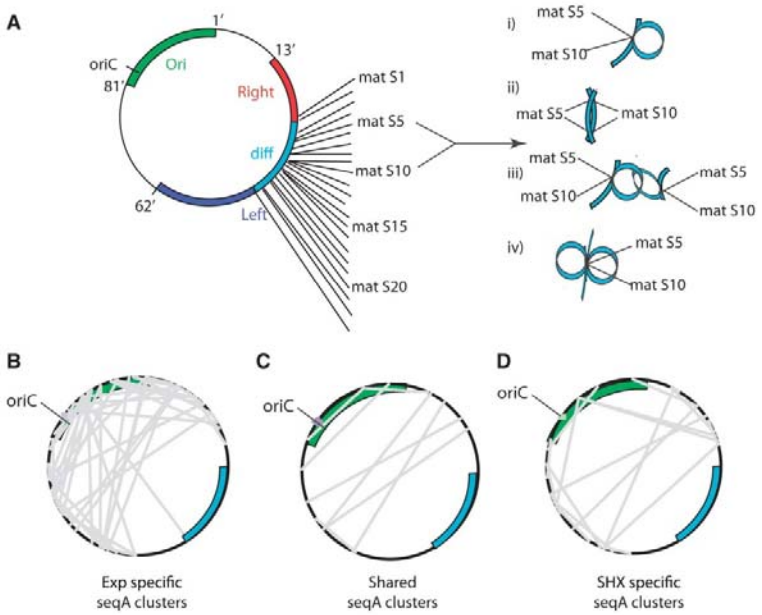


Figure 3. Binding sites for NAPs MatP and SeqA exhibit differing degrees of spatial clustering within the exponential and SHX-treated *E. coli* nucleoids. (**A**) Regions that centered on matS-binding sites [± 50 bp; (5)] show significantly increased clustering in the exponential condition, despite having interaction levels that were no different from random (Supplementary Table S6). MatS site clustering is confined to two matS sites: matS5 and matS10 and may result from (A i) intra-chromosome interactions, or (A ii–iv) inter-chromosomal interactions. Critically, this clustering is not observed in the SHX-treated nucleoid. (**B**) Exponential-specific spatial clustering of SeqA-binding sites was concentrated around the origin. (**C**) Spatial clusters of SeqA-binding sites that were shared between conditions tended to occur between the left and right replichores. (**D**) SHX-specific interactions involved fewer SeqA-binding sites and tended to be toward the terminus (Supplementary Table S7).

Genes up- or downregulated after SHX treatment exist in different spatial environments, confirming functional compartmentalization of the nucleoid

Eukaryotic studies have identified a non-random distribution of gene expression associated with the presence of spatially distinct environments that promote or inhibit nuclear functions [e.g. (49–51)]. Similarly, we observe that *E. coli* genes whose transcript levels increased or decreased in response to SHX treatment are overrepresented in some gene ontology terms (Supplementary Table S5) and are non-randomly distributed across the linear genome (Figure 4A and B) in a manner that does not correlate with GC content (Supplementary Figure S3A). There is no correlation between transcript level and interaction frequency at the level of specific restriction fragments (Supplementary Figure S3B and C). However, the SHX downregulated genes have high average transcript ($P < 0.001$; Supplementary Table S9), clustering and interaction (Figure 3C) levels in exponential phase cells. These results suggest that genes that are highly expressed in exponential phase and downregulated after SHX treatment are not only linearly but also highly spatially clustered. In conjunction with microscopic observations of large RNA polymerase clusters (foci) within exponentially growing *E. coli* cells (21), our results support the hypothesis that the highly expressed exponential phase genes are associated with transcription foci. Despite this, genes downregulated in response to SHX treatment ($P < 0.001$; Supplementary Table S9) remained highly clustered (Figure 4C). Similarly, upregulated genes within lowly clustered regions do not increase their clustering on activation (Figure 4C). As such, the maintenance of the clustering is independent of transcript levels and *ipso facto* transcription.

DISCUSSION

The *E. coli* nucleoid has a complex structure that emerges from the sum of the cellular processes that occur within the bacterial cell. We identified two macrodomains within the *E. coli* chromosome interaction networks corresponding to the Ori and Ter domains that have been previously identified (5,7,8,12,13,52). However, the two remaining macrodomains [Left (L), Right (R)] and the two non-structured domains (NS) are not obvious within our data. Moreover, we did not identify hard boundaries

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

Table 1. H-NS, IHF and Fis sites do not exhibit spatial clustering

	Interaction set	Genic regions		Non-genic regions		Clustering		Interactions	
		Clustering RS	Interactions RS	Clustering RS	Interactions RS	RS	CLS	RS	CLS
NAP[a]									
Fis	Exp	ND	Low**	ND	Low**				
	SHX	ND	NC	ND	Low**				
H-NS	Exp	ND	Low**	ND	Low**				
	SHX	ND	Low*	ND	Low**				
IHF	Exp	ND	Low*	ND	Low**				
	SHX	ND	NC	ND	Low**				
NAP (500 bp)[b]									
Fis, H-NS, IHF coding	Exp					NC	NC	Low**	Low**
	SHX					Low*	NC	Low***	Low***
Fis, H-NS, IHF non-coding	Exp					Low**	Low*	Low**	Low**
	SHX					Low**	NC	Low**	Low**

[a]The summed interaction or clustering strength for each condition was compared with 1000 random data sets with RS or CLS between each element (see 'Materials and Methods' section). Clustering of characterized H-NS-, IHF- and Fis-binding sites [± 100 bp; (18)] located within coding and non-coding sequences was not detected. H-NS-, IHF- and Fis-binding sites had significantly lower interaction frequencies than random. However, treatment with SHX altered the interaction frequencies of genic H-NS and Fis sites such that they were no longer different from random. CLS results were the same as those observed for the RS. Short distance (<800 bp) and self-interactions were excluded from these analyses.
[b]Clustering of loci that contained one or more characterized H-NS, IHF and Fis sites (± 500 bp) was no different to random (coding loci) or lower than random (non-coding loci), confirming that these elements do not cluster either individually or collectively.
*$P < 0.085$; **$P < 0.001$; ***$P < 0.05$.
NC, no change; ND, not detected.

surrounding either the Ori or Ter domain, consistent with earlier predictions (7,12). It remains possible that the L, R and NS domains and the domain boundaries were obscured because of the use of an unsynchronized population of cells. Alternatively, the formation of the macrodomains and the previously observed reductions in inter-domain recombination rates (12) could be achieved by a combination of mechanisms of which physical segregation is only one component. This explanation is supported by the observation that a low level of connectivity remains between the Ter and Ori domains. Critically, this connectivity occurs at levels above those observed for random inter-molecular ligation under our experimental conditions and indicates that although these domains are largely separated, there is some inter-domain mixing during the cell cycle. This is consistent with the observation that recombination rates between λ att sites are reduced but not completely abolished between these domains (12).

The chromosome interaction networks we identified within both exponential and SHX-treated *E. coli* cells contain variable numbers of short and long distance loops. The observation that the number of long distance interactions (long distance loops) reduced after treatment with SHX can be interpreted as indicating that the nucleoid expands under this condition, consistent with microscopic observations (21,22,53). Either the observed expansion is specific and directed as part of the stress response or it is a non-specific consequence of SHX acting on the factors that mediate the interactions (e.g. rapid protein turn over with no replacement). The exact reasons for the loss of interactions remain to be determined. However, the fact that SHX-specific interactions form indicates a directed alteration in nucleoid organization.

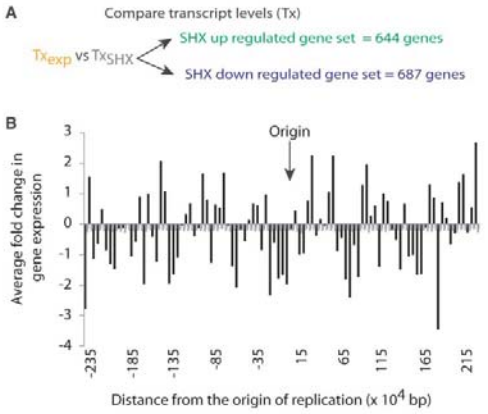
Is the *E. coli* nucleoid shaped as a sausage or rosette?

The presence of short and long distance loops within both networks points to the *E. coli* genome folding into a series of DNA loops connected to a central node (i.e. a rosette). This interpretation agrees with electron microscope observations of isolated nucleoids [reviewed in (2)]. However, our observation that the Ter region has few contacts with itself (i.e. is extended in nature) and is less well connected to the remainder of the genome is consistent with previous observations made by David Sherratt's group (4,54). Therefore, despite differences in growth rate between the studies (4), our data also support the hypothesis that the *E. coli* chromosome is organized as a sausage in which the bulk of the chromosome is organized into a compacted rod that is circularized by the Ter domain [Figure 5A (4,54)]. The apparent dichotomy of these interpretations is reconcilable through the realization that the isolation of a sausage-shaped genome during preparation for electron microscopy would result in the appearance of a rosette. Thus, the sausage model is a variation of the rosette model where the rosette is flattened through confinement or as a result of the biological processes within the live cell.

Replication contributes to nucleoid organization through SeqA

The SeqA and SlmA proteins are implicated in the regulation of replication and chromosome separation

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013



C

Gene set	Interaction set	Clustering RS	Interactions RS
SHX up regulated	exp specific	NC	NC
	SHX specific	NC	Low
	exp shared	Low*	NC
	SHX shared	Low*	NC
SHX down regulated	exp specific	High**	High**
	SHX specific	High**	High**
	exp shared	High*	High**
	SHX shared	High	High*

Key High = increased clustering or interactions ($p<0.1$)
Low = reduced clustering or interactions ($p<0.1$)
** $p<0.001$ * $p<0.05$ NC no change
Locus

Figure 4. Annotated genes with transcripts that were up- (644 genes) or down- (687 genes) regulated after SHX treatment existed in different spatial environments. (**A**) Genes that changed transcript level (Tx) after treatment with SHX were identified. (**B**) Analyses of positions of the up- and downregulated genes across the *E. coli* genome identify non-random clustering within the linear sequence. Average expression levels were calculated for 50-kb bins. Grey bars indicate the average expression across 50-kb bins within a thousand randomized genomes. Autocorrelation analyses on the distribution of gene expression data across the genome demonstrated a strong predictive relationship up to 32 genes away (ACF: >0.83). (**C**) Clustering and interaction patterns for up- or downregulated genes demonstrate that up- and downregulated genes occupy specific spatial environments. The amount of clustering within the up- or downregulated gene sets, and between the up- or downregulated genes and other loci, was compared with 1000 randomly generated sets. One thousand random sets of equivalent size (number and length) to the up- or downregulated sets were generated such that they (i) randomized the spacing between elements (RS) or (ii) conserved the linear spacing between the elements (CLS) involved in the interactions. Clustering and interaction counts were determined individually for the condition specific and shared data sets. Clustering and interaction data are shown for both exponential (exp) and SHX shared interaction sets because despite the interaction being shared, the clustering or interaction frequency was specific for each condition. There were no significant differences for comparisons with either the RS or CLS random sets. These analyses were performed on long distance interactions only.

[reviewed in (16)]. Our results indicate that SlmA-binding sites do not cluster as part of nucleoid occlusion during replication initiation or extension. Therefore, the dimerization necessary to activate SlmA occurs at a single or linearly-adjacent binding site(s) but does not result from spatial associations of distant SlmA sites. Consistent with the supposition by Dame *et al.* (16), the low levels of SlmA clustering observed indicate that any contribution that SlmA-FtsZ makes to nucleoid structure must be facilitated by tethering to an external framework [e.g. shortened preformed FtsZ polymers (45), or non-functional protofilaments (55)] or the cell membrane.

By contrast, the replication-dependent nature and distribution of the exponential phase SeqA-mediated long distance interactions provides support for a role for SeqA clustering in the formation of an intra- and/or inter-chromosomal structure (Figure 5A and B). This is particularly true for SeqA interactions that form over the origin of replication and could function to sequester newly replicated origins and delay chromosome separation [(56–58), reviewed in (16,47)]. As such, the SHX-dependent loss of the long distance interactions is predicted if replication and segregation occur consecutively (29). Thus, the loss of SeqA-mediated interactions within the SHX-treated nucleoid reflects an underlying spatial segregation of the replicated chromosome regions (46). The predominance of SeqA clusters between loci that are approximately equidistant from the Ori within the SHX–specific, and shared interaction data sets represent links between the hemimethylated GATC sites trailing the replisome. We interpret the distinct subset of inter-replichore SeqA clusters as indicating that the DNA polymerases are pausing at specific genomic sites within the cell populations. Finally, there is no correlation between alterations to transcript levels and SeqA clustering (data not shown); therefore, SeqA clustering is independent of transcription. Collectively, these results support a strong linkage between replication and nucleoid organization (4).

For ease of visualization, the chromosomal interactions that we identified are presented as intra-chromosomal connections (Figure 1). This form of presentation is problematic, as the proximity-based ligation data are probabilistic and represent a population average from unsynchronized cells (59). As such, it is impossible to determine which combinations of interactions occur within a single nucleoid. Second, although the sequences we obtain as part of the GCC protocol identify the interacting loci, they do not provide information on whether the interactions occur within or between the chromosome(s). This is an important consideration when investigating nucleoid structure in exponential phase bacterial cells that contain and segregate partially replicated chromosomes (3). Therefore, it is possible that the formation of long distance SeqA-dependent and -independent interactions can be facilitated by overlaps between the replichore arms that result from the chromosome alignment [i.e. inter-chromosomal (Figure 5A, right)]. Interestingly, such a system may contribute to gene dosage control, as well as the control of chromosome segregation. However, it remains possible that interactions also occur within a chromosome [i.e. intra-chromosomal

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

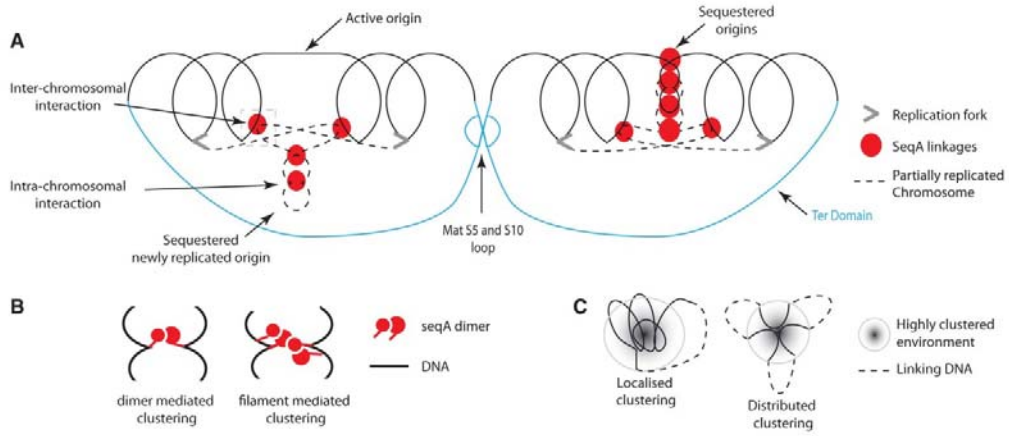


Figure 5. Spatial model of exponential phase nucleoid organization in *E. coli*. (**A**) The exponential phase *E. coli* nucleoid is organized into high interacting domains by nucleoid-associated factors including, but not limited to, SeqA and MatP. SeqA promotes the intra- and inter-chromosomal clustering of hemimethylated GATC sites to sequester recently replicated origins and contribute to chromosome segregation. Newly replicated origins can be sequestered individually (left) or through interactions between the recently replicated origins (right). The matS5–10 loop is hypothesized to form between chromosomes that have almost completed replication. The model illustrates some of the major findings in the study, but for simplicity, overlap between replichores has not been included in this cartoon. Similarly, only one replication process has been illustrated on each chromosome. Moreover, as our data are drawn from an unsynchronized population, and we do not have data on the relative positions of the different elements through the cell cycle, we have not attempted to represent the dynamic nature of the positioning of the different elements throughout the cell cycle. (**B**) SeqA can mediate interactions within or between chromosomes as either a dimer or filament. (**C**) Highly clustered regions form as a result of localized and distributed clustering within and/or between the replicated chromosomes.

(Figure 5A, left)]. Future work should determine the contribution of inter- and intra-chromosomal interactions to the structure of the nucleoid in exponentially growing *E. coli* with a view to understand how structure contributes to gene dosage control in this organism.

What role does the matS5–10 loop play in nucleoid organization?

MatS sites have a role in defining the Ter domain (5,8). *In vivo* experiments indicate that the definition of the Ter domain and condensation of this region are separable events with the condensation dependent on the presence of the MatP C-terminal coiled-coil domain, which is responsible for tetramerization and looping (60). We found that the matS5 and matS10 sites form a specific loop that surrounds the TerA site (1 339 796–1 339 791 bp) and is located away from the dif site (1 589 000 bp) toward the Ori on the right replichore. Note that matS5 is one of two matS sites (the other being matS21) that do not show *in vivo* MatP binding in an *E. coli* K12 derivative of MG1655 (5). The question thus arises as to what contribution the matS5–10 interaction makes to the Ter domain structure and function. It is possible that the matS5–10 loop explains observations of a spatially separable condensed region within the center of the Ter linker domain (4). Furthermore, the absence of detectable matS clustering between the other matS loci raises the possibility of differentiation in the functions of the matS sites. However, further experiments are required to confirm these hypotheses and identify how or if MatP contributes to the formation of the matS5–10 loop.

Do 'histone-like' NAPs play a role in global nucleoid structure?

The spatial clustering of NAP (i.e. H-NS, Fis and IHF) DNA-binding sites is not significant within the gross spatial organization of the *E. coli* nucleoid we identified. Rather our results are consistent with the hypothesis that H-NS, IHF and Fis contribute to compaction through localized structuring [reviewed in (61)], gene regulation or the formation of large protein heterocomplexes [reviewed in (62)]. These results are in contrast to those of Wang *et al.* 2011 (14), who identified H-NS clustering within the *E. coli* nucleoid using microscopic and proximity-ligation–based measurements in slow-growing early log phase cells. This apparent discrepancy may be due to the significant increase in resolution afforded by the use of the HhaI enzyme in our study. This conclusion is supported by our identification of interactions linking HhaI restriction fragments from within the larger EcoRI restriction fragments that were previously characterized as demonstrating an H-NS–dependent association [Supplementary Figure S4 (14)]. Therefore, we propose that the previously recognized relationship between ligation efficiency and the presence/absence of *h-ns* mutants (14) was likely due to a combination of a global reorganization of localized genome structure (63) and epistatic effects resulting from H-NS–dependent transcriptional changes.

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

Do transcription foci have a role in nucleoid organization?

The observed organization of highly transcribed genes into clustered spatial environments is consistent with the hypothesis that some clustering is occurring around transcription foci [e.g. (64)]. Similarly, the copy-number independent long distance interactions may reflect sequence-driven intra-chromosomal nucleoid folding for the coordination of transcription through enhancer-like interactions consistent with previous observations in bacteria (14,65,66) and eukaryotes [e.g. (67–69)]. The existence of these prokaryotic transcription foci is supported by microscopic observations of RNA polymerase foci within *E. coli* cells (20,21). The fact that similar clustering was observed in *Pseudomonas aeruginosa* (data not shown) and among highly transcribed genes in *Schizosaccharomyces pombe* (40) implies that the clustering of highly transcribed genes may be a ubiquitous feature of the control of gene expression.

It is likely that the linear gene clusters (Figure 4A) form into combinations of localized and distributed spatial clusters (Figure 5C). Given that RNA polymerase is redistributed after SHX treatment (21,22), decreases in the number of long distance interactions (i.e. reductions in the extent of distributed clustering), we observed following stress induction could be interpreted as indicating that RNA polymerase mediates some interactions. However, the identification of a core interaction pattern that is conserved within the *E. coli* nucleoid after SHX treatment indicates that at least some of these interactions are stable to a significant redistribution of RNA polymerase. This result agrees with eukaryotic studies that demonstrate long distance interactions are insensitive to inhibition of ongoing RNA polymerase transcription (70). Furthermore, the high levels of clustering and interactions observed at genes that were highly expressed in the exponential phase and subsequently downregulated by SHX treatment indicates that the localized clustering—but not necessarily the identity of the partners—is stable. However, it remains possible that transcription-associated interactions respond slowly to environmental change, allowing for short term fluctuations in environmental conditions without the requirement for major rearrangement of genome organization. This forms an epigenetic memory that is capable of being inherited (71) similar to that observed in yeast (72–76).

Does a nucleolus-like structure form within the *E. coli* nucleoid?

It has been proposed that the formation of transcription factories that include the ribosomal RNA genes and ribosomal protein encoding loci could induce the compaction of the nucleoid through the formation of a nucleolus-like structure (23,77,78). However, we found no evidence that the nucleoid structure promotes the clustering of ribosomal RNA genes and ribosomal protein encoding loci (data not shown). This may be due to technical limitations in the analysis of repetitive loci that cannot be unambiguously positioned onto the reference genome. Alternatively, it may be due to the very high levels of transcriptional activity at these loci interfering with the cross-linking and ligation steps during the preparation of our chromosome interaction libraries. *In silico* modeling of the nucleoid that incorporates biophysical parameters and interaction frequencies [similar to (9,79)] may resolve this issue.

Epistatic interactions and the chromosome interaction network

The bacterial cell is a complex structured entity in which each part exists 'for and by means of the whole' (80). As such nucleoid structure is an integral—inseparable—part of the cells response to environmental challenge. Moreover, the contribution of any one gene to the bacterial phenotype relies on its relationship with other genes on levels that include regulation, transcription, translation, complex formation and function. Therefore, it is likely that the interaction network we have determined contains information on epistatic relationships between multiple genes that occur at the regulatory, transcriptional and translational levels because of the co-dependence of these processes in *E. coli*. Future work should interrogate prokaryotic interaction networks for evidence of epistatic relationships and must address the mechanism(s) governing the organization of global structure.

CONCLUSION

The detection of both long and short distance interactions within the *E. coli* nucleoid is consistent with empirical measures and modeling, which indicated that intra-nucleoid interactions play a dominant role in shaping the *E. coli* nucleoid (11). However, the long distance interactions did not consistently involve loci located equidistant from the Ori on opposite replichores; therefore, it is unlikely that the *E. coli* nucleoid is preferentially structured as ellipsoids as observed in *C. crescentus* (9). Rather our study indicates that the chromosome(s) within exponentially fast-growing *E. coli* cells are structured by interactions that are linked to the ongoing replication and transcription processes within the cell. The specificity of the observed interactions identifies spatial organization as a significant factor in bacterial gene regulation and indicates that the spatial clustering of highly regulated genes is a ubiquitous feature of gene regulation.

ACCESSION NUMBERS

The GCC data has been banked with Gene expression omnibus (GSE40603). Expression data has been deposited GSE40304.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–10, Supplementary Figures 1–6 and Supplementary Materials and Methods.

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

ACKNOWLEDGEMENTS

The authors would like to thank Philippe Collas, Austen Ganley, Gary Greyling, Lutz R. Gehlen, Heather Hendrickson, Julia Horsfield, Rod McNab, Ana Pombo, Tom Schneider and Yan Ning Zhou for helpful discussions.

FUNDING

Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research (to C.C. and J.D.); the Marsden Fund (to J.M.O.S.); Massey University research fund (to J.M.O.S and B.H.A.R.); Massey University scholarship (to R.S.G.). Funding for open access charge: The Liggins Institute Auckland University.

Conflict of interest statement. None declared.

REFERENCES

1. Woldringh,C.L. and Nanninga,N. (2006) Structural and physical aspects of bacterial chromosome segregation. *J. Struct. Biol.*, **156**, 273–283.
2. Ishihama,A. (2009) The nucleoid: an overview. In: Curtiss,R., Kaper,J.B., Squires,C.L., Karp,P.D., Neidhardt,F.C. and Slauch,J.M. (eds), *EcoSal*. ASM Press, Washington, DC.
3. Sherratt,D.J. (2003) Bacterial chromosome dynamics. *Science*, **301**, 780–785.
4. Liu,X., Wang,X., Reyes-Lamothe,R. and Sherratt,D. (2010) Replication-directed sister chromosome alignment in *Escherichia coli*. *Mol. Microbiol.*, **75**, 1090–1097.
5. Mercier,R., Petit,M.A., Schbath,S., Robin,S., El Karoui,M., Boccard,F. and Espeli,O. (2008) The MatP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell*, **135**, 475–485.
6. Wang,X. and Sherratt,D.J. (2010) Independent segregation of the two arms of the *Escherichia coli* ori region requires neither RNA synthesis nor MreB dynamics. *J. Bacteriol.*, **192**, 6143–6153.
7. Boccard,F., Esnault,E. and Valens,M. (2005) Spatial arrangement and macrodomain organization of bacterial chromosomes. *Mol. Microbiol.*, **57**, 9–16.
8. Thiel,A., Valens,M., Vallet-Gely,I., Espeli,O. and Boccard,F. (2012) Long-range chromosome organization in *E. coli*: a site-specific system isolates the Ter macrodomain. *PLoS Genet.*, **8**, e1002672.
9. Umbarger,M.A., Toro,E., Wright,M.A., Porreca,G.J., Baù,D., Hong,S.-H., Fero,M.J., Zhu,L.J., Marti-Renom,M.A., McAdams,H.H. *et al.* (2011) The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol. Cell*, **44**, 252–264.
10. Postow,L., Hardy,C.D., Arsuaga,J. and Cozzarelli,N.R. (2004) Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.*, **18**, 1766–1779.
11. Wiggins,P.A., Cheveralls,K.C., Martin,J.S., Lintner,R. and Kondev,J. (2010) Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament. *Proc. Natl Acad. Sci. USA*, **107**, 4991–4995.
12. Valens,M., Penaud,S., Rossignol,M., Cornet,F. and Boccard,F. (2004) Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J.*, **23**, 4330–4341.
13. Espeli,O., Mercier,R. and Boccard,F. (2008) DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol. Microbiol.*, **68**, 1418–1427.
14. Wang,W., Li,G.W., Chen,C., Xie,X.S. and Zhuang,X. (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, **333**, 1445–1449.
15. Margolin,W. (2010) Imaging the bacterial nucleoid. In: Dame,R.T. and Dorman,C.J. (eds), *Bacterial Chromatin*. Springer, Dordrecht, New York, NY, pp. 13–30.
16. Dame,R.T., Kalmykowa,O.J. and Grainger,D.C. (2011) Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in gram negative bacteria. *PLoS Genet.*, **7**, e1002123.
17. Tanaka,H., Yasuzawa,K., Kohno,K., Goshima,N., Kano,Y., Saiki,T. and Imamoto,F. (1995) Role of HU proteins in forming and constraining supercoils of chromosomal DNA in *Escherichia coli*. *Mol. Gen. Genet.*, **248**, 518–526.
18. Grainger,D.C., Hurd,D., Goldberg,M.D. and Busby,S.J. (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.*, **34**, 4642–4652.
19. Dame,R.T. (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.*, **56**, 858–870.
20. Cagliero,C. and Jin,D.J. (2013) Dissociation and re-association of RNA polymerase with DNA during osmotic stress response in *Escherichia coli*. *Nucleic Acids Res.*, **41**, 315–326.
21. Cabrera,J.E. and Jin,D.J. (2003) The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol. Microbiol.*, **50**, 1493–1505.
22. Cabrera,J.E., Cagliero,C., Quan,S., Squires,C.L. and Jin,D.J. (2009) Active transcription of rRNA operons condenses the nucleoid in *Escherichia coli*: examining the effect of transcription on nucleoid structure in the absence of transertion. *J. Bacteriol.*, **191**, 4180–4185.
23. Jin,D.J., Cagliero,C. and Zhou,Y.N. (2012) Growth rate regulation in *Escherichia coli*. *FEMS Microbiol. Rev.*, **36**, 269–287.
24. Cashel,M., Gentry,D., Hernandez,V. and Vinella,D. (1996) The stringent response. In: Neidhardt,F.C. and Curtiss,R. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn. ASM Press, Washington, DC.
25. Durfee,T., Hansen,A.M., Zhi,H., Blattner,F.R. and Jin,D.J. (2008) Transcription profiling of the stringent response in *Escherichia coli*. *J. Bacteriol.*, **190**, 1084–1096.
26. Traxler,M.F., Summers,S.M., Nguyen,H.T., Zacharia,V.M., Hightower,G.A., Smith,J.T. and Conway,T. (2008) The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol. Microbiol.*, **68**, 1128–1148.
27. Jin,D.J. and Cabrera,J.E. (2006) Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in *Escherichia coli*. *J. Struct. Biol.*, **156**, 284–291.
28. Bremer,H. and Dennis,P.P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt,F.C. and Curtiss,R. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn. ASM Press, Washington, DC, pp. 1553–1569.
29. Nielsen,H.J., Youngren,B., Hansen,F.G. and Austin,S. (2007) Dynamics of *Escherichia coli* chromosome segregation during multifork replication. *J. Bacteriol.*, **189**, 8660–8666.
30. Elmore,S., Muller,M., Vischer,N., Odijk,T. and Woldringh,C.L. (2005) Single-particle tracking of oriC-GFP fluorescent spots during chromosome segregation in *Escherichia coli*. *J. Struct. Biol.*, **151**, 275–287.
31. Li,Y., Sergueev,K. and Austin,S. (2002) The segregation of the *Escherichia coli* origin and terminus of replication. *Mol. Microbiol.*, **46**, 985–996.
32. Wang,X., Possoz,C. and Sherratt,D.J. (2005) Dancing around the divisome: asymmetric chromosome segregation in *Escherichia coli*. *Genes Dev.*, **19**, 2367–2377.
33. Bates,D. and Kleckner,N. (2005) Chromosome and replisome dynamics in *E. coli*: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, **121**, 899–911.
34. Sunako,Y., Onogi,T. and Hiraga,S. (2001) Sister chromosome cohesion of *Escherichia coli*. *Mol. Microbiol.*, **42**, 1233–1241.
35. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

36. Rodley,C.D., Bertels,F., Jones,B. and O'Sullivan,J.M. (2009) Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genet. Biol.*, **46**, 879–886.
37. Dostie,J., Richmond,T.A., Arnaout,R.A., Selzer,R.R., Lee,W.L., Honan,T.A., Rubio,E.D., Krumm,A., Lamb,J., Nusbaum,C. *et al.* (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.*, **16**, 1299–1309.
38. Fullwood,M.J., Liu,M.H., Pan,Y.F., Liu,J., Xu,H., Mohamed,Y.B., Orlov,Y.L., Velkov,S., Ho,A., Mei,P.H. *et al.* (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, **462**, 58–64.
39. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
40. Tanizawa,H., Iwasaki,O., Tanaka,A., Capizzi,J.R., Wickramasinghe,P., Lee,M., Fu,Z. and Noma,K. (2010) Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.*, **38**, 8164–8177.
41. Rodley,C.D., Grand,R.S., Gehlen,L.R., Greyling,G., Jones,M.B. and O'Sullivan,J.M. (2012) Mitochondrial-nuclear DNA interactions contribute to the regulation of nuclear transcript levels as part of the inter-organelle communication system. *PLoS One*, **7**, e30943.
42. Team,R.D.C. (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
43. Boeva,V., Popova,T., Bleakley,K., Chiche,P., Cappo,J., Schleiermacher,G., Janoueix-Lerosey,I., Delattre,O. and Barillot,E. (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, **28**, 423–425.
44. Sanchez-Romero,M.A., Busby,S.J., Dyer,N.P., Ott,S., Millard,A.D. and Grainger,D.C. (2010) Dynamic distribution of seqa protein across the chromosome of *Escherichia coli* K-12. *MBio*, **1**, pii: e00012–10.
45. Cho,H., McManus,H.R., Dove,S.L. and Bernhardt,T.G. (2011) Nucleoid occlusion factor SlmA is a DNA-activated FtsZ polymerization antagonist. *Proc. Natl Acad. Sci. USA*, **108**, 3773–3778.
46. Ferullo,D.J. and Lovett,S.T. (2008) The stringent response and cell cycle arrest in *Escherichia coli*. *PLoS Genet.*, **4**, e1000300.
47. Waldminghaus,T. and Skarstad,K. (2009) The *Escherichia coli* SeqA protein. *Plasmid*, **61**, 141–150.
48. Ali Azam,T., Iwata,A., Nishimura,A., Ueda,S. and Ishihama,A. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, **181**, 6361–6370.
49. Iborra,F., Pombo,A., Jackson,D. and Cook,P. (1996) Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J. Cell Sci.*, **109**, 1427–1436.
50. Cook,P.R. (1999) The organization of replication and transcription. *Science*, **284**, 1790–1795.
51. Versteeg,R., van Schaik,B.D.C., van Batenburg,M.F., Roos,M., Monajemi,R., Caron,H., Bussemaker,H.J. and van Kampen,A.H.C. (2003) The human transcriptome map reveals extremes in gene dentistry, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.*, **13**, 1998–2004.
52. Niki,H., Yamaichi,Y. and Hiraga,S. (2000) Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.*, **14**, 212–223.
53. Kuhlman,T.E. and Cox,E.C. (2012) Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol. Syst. Biol.*, **8**, 610.
54. Wang,X., Liu,X., Possoz,C. and Sherratt,D.J. (2006) The two *Escherichia coli* chromosome arms locate to separate cell halves. *Genes Dev.*, **20**, 1727–1731.
55. Tonthat,N.K., Arold,S.T., Pickering,B.F., Van Dyke,M.W., Liang,S., Lu,Y., Beuria,T.K., Margolin,W. and Schumacher,M.A. (2011) Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. *EMBO J.*, **30**, 154–164.
56. Lu,M., Campbell,J.L., Boye,E. and Kleckner,N. (1994) SeqA: a negative modulator of replication initiation in *E. coli*. *Cell*, **77**, 413–426.
57. Bach,T., Krekling,M.A. and Skarstad,K. (2003) Excess SeqA prolongs sequestration of oriC and delays nucleoid segregation and cell division. *EMBO J.*, **22**, 315–323.
58. von Freiesleben,U., Rasmussen,K.V. and Schaechter,M. (1994) SeqA limits DnaA activity in replication from oriC in *Escherichia coli*. *Mol. Microbiol.*, **14**, 763–772.
59. Grand,R.S., Gehlen,L.R. and O'Sullivan,J.M. (2011) Methods for the investigation of chromosome organization. In: Urbano,K.V. (ed.), *Advances in Genetics Research*, Vol. 5. Nova Publishers, New York, NY.
60. Dupaigne,P., Tonthat,N.K., Espeli,O., Whitfill,T., Boccard,F. and Schumacher,M.A. (2012) Molecular basis for a protein-mediated DNA-bridging mechanism that functions in condensation of the *E. coli* chromosome. *Mol. Cell*, **48**, 560–571.
61. Luijsterburg,M.S., Noom,M.C., Wuite,G.J. and Dame,R.T. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.*, **156**, 262–272.
62. Fang,F.C. and Rimsky,S. (2008) New insights into transcriptional regulation by H-NS. *Curr. Opin. Microbiol.*, **11**, 113–120.
63. Bouffartigues,E., Buckle,M., Badaut,C., Travers,A. and Rimsky,S. (2007) H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.*, **14**, 441–448.
64. Carter,D.R.F., Eskiw,C. and Cook,P.R. (2008) Transcription factories. *Biochem. Soc. Trans.*, **36**, 585–589.
65. Reitzer,L.J. and Magasanik,B. (1986) Transcription of glnA in *E. coli* is stimulated by activator bound to sites far from the promoter. *Cell*, **45**, 785–792.
66. Dandanell,G., Valentin-Hansen,P., Larsen,J.E. and Hammer,K. (1987) Long-range cooperativity between gene regulatory sequences in a prokaryote. *Nature*, **325**, 823–826.
67. Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F. and de Laat,W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, **10**, 1453–1465.
68. Palstra,R.J., Tolhuis,B., Splinter,E., Nijmeijer,R., Grosveld,F. and De Laat,W. (2003) The β-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.*, **35**, 190–194.
69. Sanyal,A., Lajoie,B.R., Jain,G. and Dekker,J. (2012) The long-range interaction landscape of gene promoters. *Nature*, **489**, 109–113.
70. Palstra,R.J., Simonis,M., Klous,P., Brasset,E., Eijkelkamp,B. and de Laat,W. (2008) Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS One*, **3**, e1661.
71. Veening,J.W., Smits,W.K. and Kuipers,O.P. (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.*, **62**, 193–210.
72. Brickner,J.H. (2010) Transcriptional memory: staying in the loop. *Curr. Biol.*, **20**, R20–R21.
73. Brickner,J.H. (2009) Transcriptional memory at the nuclear periphery. *Curr. Opin. Cell Biol.*, **21**, 127–133.
74. Brickner,D.G., Cajigas,I., Fondufe-Mittendorf,Y., Ahmed,S., Lee,P.C., Widom,J. and Brickner,J.H. (2007) H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol.*, **5**, e81.
75. Tan-Wong,S.M., Wijayatilake,H.D. and Proudfoot,N.J. (2009) Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.*, **23**, 2610–2624.
76. Laine,J.P., Singh,B.N., Krishnamurthy,S. and Hampsey,M. (2009) A physiological role for gene loops in yeast. *Genes Dev.*, **23**, 2604–2609.
77. Cabrera,J.E. and Jin,D.J. (2006) Active transcription of rRNA operons is a driving force for the distribution of RNA polymerase in bacteria: effect of extrachromosomal copies of

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

rrnB on the in vivo localization of RNA polymerase. *J. Bacteriol.*, **188**, 4007–4014.
78. O'Sullivan,J.M. (2011) Chromosome organizaton in simple and complex unicellular organisms. *Curr. Issues Mol. Biol.*, **13**, 37–42.
79. Gehlen,L.R., Gruenert,G., Jones,M.B., Rodley,C.D., Langowski,J. and O'Sullivan,J.M. (2012) Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus*, **3**.
80. Kauffman,S.A. (1995) *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*. Oxford University Press, New York, NY.

Downloaded from http://nar.oxfordjournals.org/ by guest on May 1, 2013

9 APPENDIX III

OPEN ACCESS Freely available online

PLoS one

Mitochondrial-Nuclear DNA Interactions Contribute to the Regulation of Nuclear Transcript Levels as Part of the Inter-Organelle Communication System

Chris D. M. Rodley[1], Ralph S. Grand[1], Lutz R. Gehlen[1], Gary Greyling[1], M. Beatrix Jones[2], Justin M. O'Sullivan[1]*

1 Institute of Natural Sciences, Massey University, Albany, New Zealand, **2** Centre for Mathematical Biology, Massey University, Albany, New Zealand

Abstract

Nuclear and mitochondrial organelles must maintain a communication system. Loci on the mitochondrial genome were recently reported to interact with nuclear loci. To determine whether this is part of a DNA based communication system we used genome conformation capture to map the global network of DNA-DNA interactions between the mitochondrial and nuclear genomes (Mito-nDNA) in *Saccharomyces cerevisiae* cells grown under three different metabolic conditions. The interactions that form between mitochondrial and nuclear loci are dependent on the metabolic state of the yeast. Moreover, the frequency of specific mitochondrial - nuclear interactions (*i.e. COX1-MSY1* and *Q0182-RSM7*) showed significant reductions in the absence of mitochondrial encoded reverse transcriptase machinery. Furthermore, these reductions correlated with increases in the transcript levels of the nuclear loci (*MSY1* and *RSM7*). We propose that these interactions represent an inter-organelle DNA mediated communication system and that reverse transcription of mitochondrial RNA plays a role in this process.

Citation: Rodley CDM, Grand RS, Gehlen LR, Greyling G, Jones MB, et al. (2012) Mitochondrial-Nuclear DNA Interactions Contribute to the Regulation of Nuclear Transcript Levels as Part of the Inter-Organelle Communication System. PLoS ONE 7(1): e30943. doi:10.1371/journal.pone.0030943

Editor: Stefan Maas, Lehigh University, United States of America

Received August 25, 2011; **Accepted** December 27, 2011; **Published** January 23, 2012

Copyright: © 2012 Rodley et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Marsden Fund; Auckland Medical Research Foundation; Massey University Research Fund; and Maurice & Phyllis Paykel Trust. HRC Ph.D. scholarship (08/554) to CR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: J.M.OSullivan@massey.ac.nz

Introduction

Mitochondria have a central role within the metabolic systems of cells. In yeast (*Saccharomyces cerevisiae*), as in other organisms, the mitochondrial organelle contains a genome that encodes an essential subset of the electron transport chain components [1] that are necessary for respiratory growth [2].

The mitochondrial genome has drastically reduced in size during the course of evolution to the point that ~98% of the genes required for mitochondrial function are encoded within the nuclear chromosomes [3]. Consequently, mechanisms must exist to co-ordinate and control the expression of the nuclear- and mitochondrial genome- encoded genes in order to maintain and control mitochondrial function [4,5]. Intriguingly, despite the fact that the majority of mitochondrial genes have transferred to the nuclear genome, transfer of mitochondrial DNA (mtDNA) to the yeast nucleus remains an on-going process with mtDNA being used to repair double stranded breaks in yeast nuclear chromosomes under certain conditions [6]. Unstable mitochondrial plasmids have also been observed to transfer into the yeast nucleus [7,8] in a nuclear gene (*e.g. YME1, YME2*) dependent manner [8–10]. The nuclear functions of these transferred mtDNAs are unknown, however elevated mitochondria to nucleus DNA migration rates correlate with accelerated chronological aging in yeast [11].

Distal regulatory regions (*e.g.* enhancers) are known to loop within chromosomes in order to interact with the promoter region of the genes that they control [12]. Furthermore, enhancers can also interact with promoters on different chromosomes to control gene expression [13,14]. These types of inter- and intra-chromosomal interactions can be captured using proximity-based ligation methodologies (*e.g.* Chromosome Conformation Capture (3C) [15]) that incorporate high resolution (*i.e.* ~2 Å [16]) cross-linking of interacting DNA strands, restriction digestion, dilution, and ligation to identify DNA sequences that interact within a cell. Using a proximity-based ligation method that we developed to observe the global set of genome wide interactions (Genome Conformation Capture (GCC)), we previously observed that nucleic acids of mitochondrial origin interact with nuclear loci (hereinafter referred to as Mito-nDNA interactions) in *S. cerevisiae* [17]. Surprisingly these inter-organelle, Mito-nDNA interactions are frequent and statistically significant suggesting that they perform a hitherto unrecognized role within yeast cells [17]. Furthermore, analysis of one of these interactions demonstrated carbon source dependence [17]. Intriguingly, the quality and quantity of mitochondrial DNA has been shown to affect patterns of nuclear transcription [18,19] and replication [20] in yeast.

In this study we explore the hypothesis that inter-organelle interactions respond to the metabolic status of the cell to regulate nuclear transcript levels. Using GCC we identify dramatic differences in both the frequency and identities of inter-organelle interactions occurring in *S. cerevisiae* during growth on glucose, galactose (*i.e.* respiro-fermentation [21,22]), and glycerol lactate

(*i.e.* solely respiration). We also demonstrate that Interactions between mitochondrial genes (*i.e. COX1* and *Q0182*, a dubious mitochondrial ORF) and nuclear encoded loci (*i.e. MSY1* and, *RSM7*, respectively), are dependent upon a functional electron transport chain and mitochondrial encoded reverse transcriptase machinery. Finally, the levels of the nuclear encoded *MSY1* and *RSM7* gene transcripts are increased when the interaction frequency is reduced by the knockout of mitochondrial reverse transcriptase activity. On the basis of these results we propose that reverse-transcription mediated inter-organelle DNA interactions are a novel form of communication between mitochondria and the nucleus.

Results

We previously captured Mito-nDNA interactions in *S. cerevisiae* cells grown in glucose by GCC [17]. A detailed investigation of one of these Mito-nDNA interactions (between the *COX1* gene (Mt: 24872–26193 bp) and the nuclear encoded *MSY1* gene (Chr XVI; 365496–365760 bp), herein after denoted *COX1-MSY*) [17] demonstrated that it was carbon source dependent. Therefore, we hypothesized that Mito-nDNA interactions would alter, on a global scale, according to the cell's metabolic state, and in particular, the carbon source used for growth. Thus, we used GCC to generate comprehensive maps of the Mito-nDNA interactions in *S. cerevisiae* during exponential growth in media containing glucose, galactose, or glycerol lactate. Two biological replicates were prepared and analyzed for each condition. Interaction networks were constructed from 36 bp paired end Illumina Genome Analyzer sequence reads (total reads; glucose 56,167,792, galactose 48,419,385, glycerol lactate 49,134,906) of GCC libraries prepared using *MspI* digested chromatin.

Statistical and experimental methods were used to determine if the Mito-nDNA interaction patterns could have been generated by experimental noise alone, which would be expected to produce random pairings of fragments from the two genomes. *In silico* simulations (100,000) were performed [17] to determine the maximum count of a particular interaction that would be observed under this random noise model, given the same number of sequences, interactions and fragments as in the experimental data. These results showed that the real dataset deviates from a random distribution and, therefore, we conclude that the interaction patterns cannot be attributed to noise alone under any of the conditions, in each case with a p-value less than 10^{-5}. Subsequently, we performed analyses to determine what frequency individual interactions have to achieve before they are deemed to be present at a level above experimental noise (Methods S1). As a result, we identified 8678 statistically significant interactions occurring between the nuclear and mitochondrial genomes during glucose growth, 1780 during galactose growth, and 8153 during growth in glycerol lactate (Table 1). The numbers of interactions in each condition did not correlate with the measured mitochondrial copy number (Table S1). Biological replicates for each condition were highly correlated for statistically significant interactions ($R^2 = 0.78$, 0.93, 0.93, respectively; Figure S1 and Methods S1). Accordingly, sequences from biological replicates were combined and reanalyzed.

To experimentally control for spurious inter-molecular ligation events during the GCC process, samples were spiked with two ligation controls during library preparation. The first ligation control consisted of PCR products amplified from *Escherichia coli* or Lambda bacteriophage (Table S2) that were added (1:1 ratio with the nuclear genome copy number) before the GCC ligation step. These controls were designed to estimate the frequency of random inter-molecular ligation events during GCC library preparation. A maximum of 47 separate ligation events were observed, none of which occurred at levels above the statistically defined experimental noise threshold. The second ligation control consisted of the addition of pUC19 plasmid to the sample following the GCC ligation in order to control for random ligation events during high-throughput sequencing library preparation. We observed a maximum of six interactions between pUC19 and the rest of the genome; again none of these interactions were above the statistically defined experimental noise threshold. These controls show that random inter-molecular ligations occur at very low frequencies that are below our noise threshold for significant interactions. This is true even for the high copy number rDNA and mitochondrial DNA elements and thus provides empirical evidence that random ligation events during sample preparation do not account for the interactions we observe.

Table 1. Inter-organelle interactions are carbon source dependent.

	Glucose	Glycerol Lactate	Galactose
Mito-nDNA Interactions	363	3879	278
Mito-rDNA Interactions	8315	4274	1512
Total	**8678**	**8153**	**1780**

There was a >10 fold increase in the number of Mito-nDNA interactions during growth in glycerol lactate (respiration) as compared to growth in glucose and galactose (respiro-fermentation). Growth on galactose resulted in less Mito-nDNA and Mito-rDNA interactions combined, compared to the other two conditions. Statistically significant DNA-DNA interactions were divided according to whether the mtDNA was interacting with nuclear rDNA, or with unique nuclear loci. Corrections for the copy numbers of the rDNA repeats and the mitochondrial genome were incorporated into the significance calculations (Methods S1).

doi:10.1371/journal.pone.0030943.t001

Significant interactions were separated into two pools, those which occur between the mtDNA and the nuclear ribosomal DNA repeats (Mito-rDNA), and those between mtDNA and unique nuclear loci (Mito-nDNA; Table 1). The rDNA repeats form part of the nucleolus and encode the rRNA component of the cytosolic ribosomes. The rDNA repeats constitute ~9.8% of the yeast genome; yet, the Mito-rDNA interactions constitute 95.8%, 52.4%, and 84.5% of the total interactions between the nuclear and mitochondrial genomes in glucose, glycerol lactate, and galactose, grown cells, respectively. There does not appear to be an interaction "hotspot", with Mito-rDNA interactions evenly spread across the 9.1 kb rDNA repeat (data not shown). Hence, Mito-rDNA interactions are over-represented within the data-set and are carbon source dependent (Table 1). We also observed considerable alterations to the numbers of Mito-nDNA interactions. Moreover, the mitochondrial regions that are involved in interactions with the nDNA are not uniformly distributed across the mitochondrial genome (Figure S2).

In order to determine whether the observed changes in Mito-nDNA interactions are chromosome specific, we asked whether nuclear chromosome length correlates with the number of interactions identified for each individual chromosome. The number of Mito-nDNA interactions per nuclear chromosome is highly correlated with chromosome length in the glycerol lactate condition, but not in glucose or galactose (Figure 1). This discrepancy is mainly due to the deviation of chromosome X from the trend during growth in glucose and galactose. Intriguingly, the increase in mtDNA interactions with chromosome X is accounted for by a single nuclear *MspI* fragment that

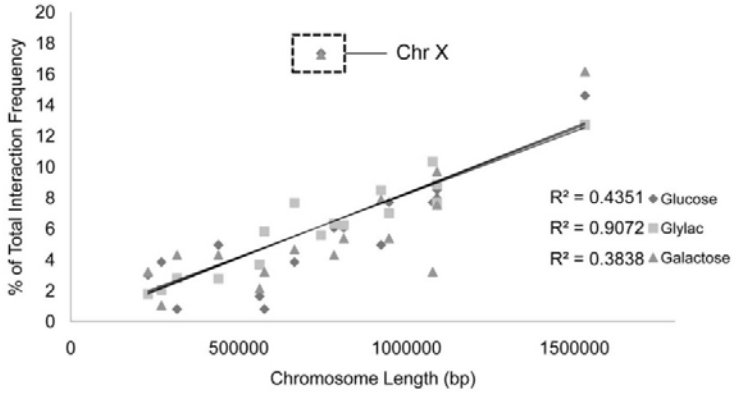


Figure 1. The number of Mito-nDNA interactions correlates with chromosome length, except chromosome X. Statistically significant Mito-nDNA interactions, occurring above the expected noise level (selected to have a false positive rate of between 1 and 3%), have been summed for each nuclear chromosome and expressed as a percentage of the total number of interactions for the particular sample before being plotted according to chromosome length in base pairs. Interactions included in this analysis are between the mitochondrial genome and nuclear chromosomes, with the 2-micron plasmid and rDNA interactions removed. The length of chromosome XII has been reduced to account for the rDNA interactions being removed.
doi:10.1371/journal.pone.0030943.g001

encompasses the promoter region and part of the coding sequences of two divergent ORFs: one uncharacterized ORF (YJR115W), and *RSM7* which encodes a mitochondrial small subunit ribosomal protein. Numerous mtDNA *MspI* fragments, including fragments surrounding or overlapping the *COX1*, *COX3*, *VAR1* and *SCE1* genes, interact with this one region on chromosome X.

Yeast mitochondrial escape mutants (YME) [8] have been previously implicated in an elevated rate of transfer of unstable mitochondrial plasmids to the yeast nucleus [7–10]. Therefore, we predicted that the YME pathway was the source of mtDNA fragments interacting with the nuclear genome, and that mutations within this pathway would result in an increase in the frequency of inter-organelle DNA interactions. To test this prediction, we used quantitative 3 C to compare the frequency of the strong *COX1-MSY1* interaction (identified in [17]) in *S. cerevisiae* YME knockout mutants (*i.e. Δyme1*, *Δyme2*). Contrary to expectations, we observed a significant decrease in the frequency of the *COX1-MSY1* interaction in the Δ*yme1* strain as compared to the wild-type (T-test [Paired P(T<=t) one-tail, n = 4] p = 0.010; Figure 2). Deletion of *YME2* or a functionally unconnected nuclear gene (*ADE2*) did not significantly affect the *COX1-MSY1* interaction frequency (t-test [paired P(T<=t) one-tail, n = 4] p = 0.377 and 0.103 respectively; Figure 2). These results suggest that the source of the mtDNA that participates in the Mito-nDNA interactions is not the unstable mitochondrial plasmids that were previously identified as escaping the mitochondria for the nuclear compartment.

Deletion of *YME1* results in an elevated rate of mitochondrial turn-over as well as an abnormal globular mitochondrial morphology [9,23]. Therefore, it is possible that this fragmented mitochondrial phenotype contributes to the reduction in the *COX1-MSY1* interaction frequency we observed in the *yme1* deletion strain. To test this we arrested yeast cells with α-factor, which results in a fragmented mitochondrial network [24] that is phenotypically similar to the one observed in *yme1* deletion strains [23]. We also deleted the mitochondrial fission gene (*MDV1*) to create strains that are unable to correctly fragment mitochondria [25,26]. Interestingly, we observed a similar reduction in the *COX1-MSY1* interaction frequency upon α-factor induced synchronization (Figure S3). However, the interaction frequency measured in the Δ*mdv1* strain was intermediate between that observed for the wild-type and Δ*yme1* strains, and not significantly different from either (T-test [Paired P(T<=t) one-tail, n = 4] wt-Δ*mdv1* p = 0.143, Δ*mdv1*-Δ*yme1* p = 0.210; Figure 2). Therefore, it is unlikely that mitochondrial fragmentation is directly responsible for the observed changes in *COX1-MSY1* interaction frequency.

We next postulated that an abnormal mitochondrial morphology, coupled with elevated mitochondrial turnover would result in a disturbance of the mitochondrial ATP synthesis pathway, and this may explain the observed reduction in the frequency of the *COX1-MSY1* interaction. Therefore, we tested the inter-organelle *COX1-MSY1* interaction for ATP dependence by treating yeast cells with an electron transport chain uncoupling agent, 2,4-Dinitrophenol (DNP), at a concentration (5 mM) that inhibits respiration but allows fermentation (Figure S4). We observed a significant time-dependent decrease in the frequency of the *COX1-MSY1* interaction in the presence of DNP (t-test $p<0.05$; Figure 3A), as measured by quantitative 3 C. However, an interaction between two nuclear loci (nDNA-nDNA; Chr VII: 868673–873686 bp - Chr IX: 172565–173311 bp) was also shown to be affected by treatment with DNP (Figure 3B). The observed dependence of the nDNA-nDNA interaction on a proton gradient across the mitochondrial membrane, and thus mitochondrial ATP synthesis, suggests that formation of these DNA-DNA interactions is ATP dependent.

The region of the *COX1* gene involved in the *COX1-MSY1* interaction overlaps the non-essential group II mitochondrial aI5γ intron. There are four Group II introns present within yeast mitochondria (aI1, aI2, bI1, and aI5γ) and these introns encode functional splicing, reverse transcriptase, or endonuclease machinery [27–31]. Only the aI1 and aI2 introns encode reverse transcriptase [32] activity while the aI5γ intron encodes endounuclease activity [33,34] but not reverse transcriptase activity [35]. Therefore, we postulated that reverse transcription of the mitochondrial group II introns might be involved in the *COX1-MSY1* interaction. To test this we measured the *COX1-MSY1* interaction using quantitative 3 C on a strain which only

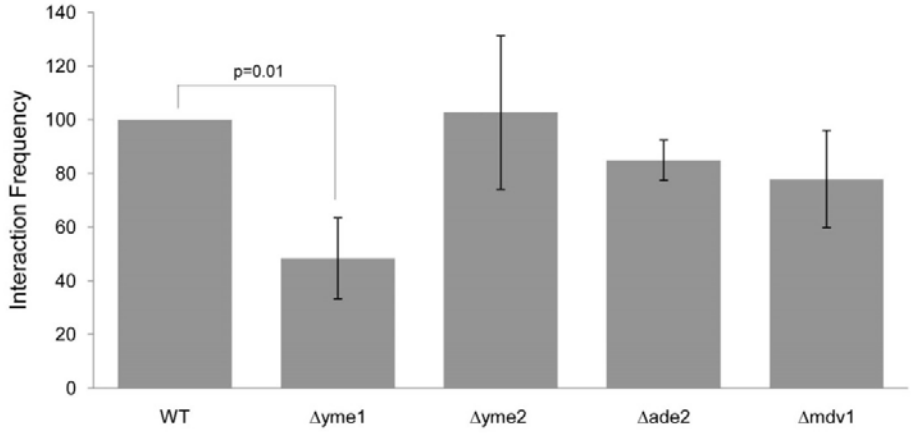


Figure 2. Deletion of *yme1* causes a significant reduction in the frequency of the mitochondrial-nuclear *COX1-MSY1* interaction. Interaction frequency between the mitochondrial *COX1* and nuclear *Msp*I fragments was assayed by quantitative 3 C (Methods S1) in wild-type (*S. cerevisiae* BY4741), Δ*yme1* (BY4741 Δ*yme1*), Δ*yme2* (BY4741 Δ*yme2*), Δ*ade2* (BY4741 Δ*ade2*) and Δ*mdv1* (BY4741, Δ*mdv1*) strains. Interaction values were corrected for mitochondrial genome copy number (see Methods) and are expressed as percentages of wild-type (set at 100%) +/− standard error of the mean (n = 4). Deletion of an unconnected gene (*ade2*) did not significantly affect interaction frequency. T-tests (paired P(T<=t) one-tail, n = 4) were performed to determine the significance of observed variations: wild-type: Δ*yme1* p = 0.01; wild-type: Δ*yme2* p = 0.377; wild-type: Δ*ade2* p = 0.103; wild-type: Δ*mdv1* p = 0.143; Δ*yme1*: Δ*mdv1* p = 0.210. Only Δ*yme1* demonstrated a significant difference.
doi:10.1371/journal.pone.0030943.g002

contained the *COX1* aI5γ intron (*i.e.* 161-U7 GII-0 aI5γ; Figure 4A). We observed a 40–60% decrease in the inter-organelle *COX1-MSY1* interaction in the GII-0 aI5γ strain relative to the wild type (Figure 4B). We concluded that the partial dependence of the *COX1-MSY1* interaction upon the presence of the group II introns reflects a role for reverse transcription in the inter-organelle interactions. However, part of the *COX1-MSY1* interaction remains independent of mitochondrial encoded reverse transcriptase.

It remained to be seen if interactions involving other mitochondrial loci required the presence of the type II introns and hence reverse transcriptase. We examined the interaction frequency between a dubious mitochondrial ORF (*Q0182*; mtDNA [65783–65903 bp]), that does not contain any group II introns and therefore is not altered in the 161-U7 GII-0 strain, and the nuclear encoded *RSM7* (Chr X [638756–640423 bp]) gene. The *Q0182-RMS7* interaction frequency decreased in the absence of the mitochondrial group II introns (Figure 4C). These results

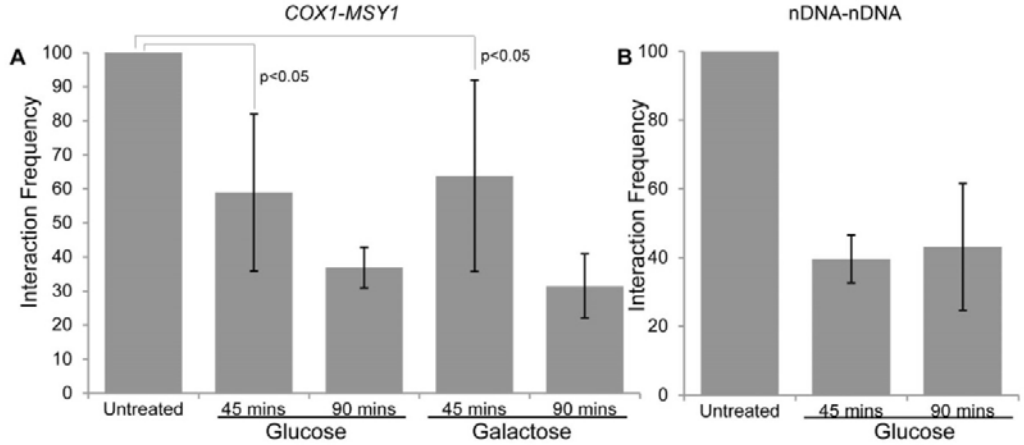


Figure 3. A functional electron transport chain is required to maintain the interaction between the mitochondrial *COX1* and nuclear *MSY1* loci. Uncoupling of the electron transport chain was achieved by 2,4-dinitrophenol (5 mM) treatment of exponentially growing *S. cerevisiae* in synthetic complete media, containing glucose or galactose, for the indicated time (Figure S4). *COX1-MSY1* A) and nDNA-nDNA B) interaction frequencies were determined by quantitative 3 C analyses using fluorescent probes (see Methods S1). Interaction values in A) were corrected for mitochondrial genome copy number while those in B) were corrected for nuclear genome copy number (see Methods). Interaction values were expressed as percentages of the untreated sample (set at 100%) +/− standard error of the mean (n = 3).
doi:10.1371/journal.pone.0030943.g003

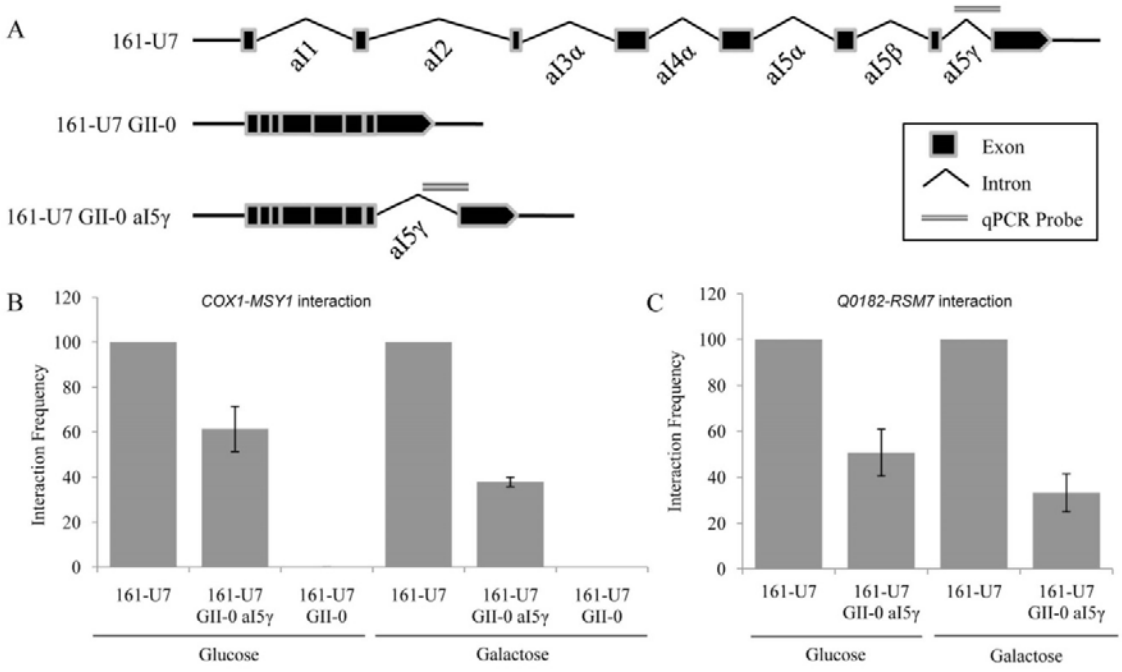


Figure 4. Mito-nDNA interactions require active mitochondrial reverse transcriptase machinery. A) Illustration of *COX1* gene arrangement in the WT (161-U7), intron a15γ (161-U7 GII-0 a15γ), and no mitochondrial group II introns (161-U7 GII-0) strains. Group II introns within the *COX1* gene encode functional reverse transcriptase. The region of *COX1* that participates in the *COX1-MSY1* interaction is indicated (qPCR probe). Strain 161-U7 GII-0 was included as a control to rule out a nuclear sequence, originating from a mitochondrial integration within the nuclear genome (NUMT), being responsible for the observed interaction. B) *COX1-MSY1* interaction frequencies for wt and intron mutants, illustrated in A), grown in glucose or galactose. C) *Q0182-RSM7* interaction frequencies for mitochondrial reverse transcriptase mutant 161-U7 GII-0 a15γ, illustrated in A), grown in glucose or galactose. Interaction frequencies are expressed as percentages of the wild-type *S. cerevisiae* strain 161-U7 for each carbon source (set at 100%) +/− standard error of the mean (n = 3). Interaction values in B) and C) were corrected for nuclear genome copy number to facilitate direct comparison.
doi:10.1371/journal.pone.0030943.g004

confirm that this phenomenon is not restricted to interactions involving *COX1*.

Our results suggested that the nucleic acids of mitochondrial origin which participate in the Mito-nDNA interaction are reverse transcribed from RNA intermediates prior to transfer to the nucleus as cDNAs. However, it remained possible that the Mito-nDNA interactions we observed did not involve inter-organelle transfer. Rather, these interactions might have been completely or partially due to interactions between nuclear loci and mitochondrial sequences that had been integrated into the nuclear genome (*i.e.* nuclear-mitochondrial sequences (NUMTs)). To rule out the possibility that NUMTs were involved, we performed quantitative 3 C analyses in strains without (*i.e.* 161-U7 GII-0) the *COX1* aI5γ intron (Figure 4A and B). Removal of aI5γ, and hence the probe site (Figure 4A), resulted in complete loss of detectable *COX1-MSY1* interactions (Figure 4B). This confirmed that the *COX1-MSY1* interaction involves DNA that is directly derived from the mitochondrial genome and not a NUMT.

The number of significant (Methods S1) Mito-nDNA interactions increased >10-fold in respiring (*i.e.* glycerol lactate grown) cells, relative to glucose or galactose grown cells (Table 1). This increase was not due to a higher number of sequence reads for the respiring sample. Thus, a greater number of unique nuclear loci connect to mtDNA during respiratory growth when the mitochondria are most active. This result, coupled with the need for a functional electron transport chain and reverse transcriptase machinery, led us to hypothesize that the Mito-nDNA interactions are functional in nature, and specifically that they are capable of controlling the transcript levels of the nuclear loci with which they interact. To test this we performed quantitative reverse transcriptase PCR (qRT-PCR) to determine the transcript levels of the nuclear encoded *MSY1* and *RSM7* genes in WT cells, the mitochondrial group-II intron knockout mutant (161-U7 GII-0), and strain 161-U7 GII-0 aI5γ (Figure 4A). We found that the population transcript level of the *MSY1* gene is significantly higher (t-test, two-sample unequal variance, one-tail, n = 2, p = 0.0007) in strain 161-U7 GII-0 (Figure 5A), which does not contain the probe site and, therefore, has no detectable *COX1-MSY1* interaction (Figure 4A and B), thus identifying the maximum transcript level in the absence of detectable inter-organelle interactions. Critically, we observed a similar population level increase in *MSY1* transcript levels following the removal of the type II introns, except aI5γ (*i.e.* strain 161-U7 GII-0 aI5γ; Figure 5A). A similar increase was observed for *RSM7* transcripts in both the 161-U7 GII-0 and 161-U7 GII-0 aI5γ strains relative to the WT (Figure 5B), consistent with the effects of intron deletion on the *Q0182-RMS7* interaction level (Figure 4C). By contrast deletion of *MRS1*, which is involved in mitochondrial group I intron splicing [36,37], had no effect on either *MSY1* or *RSM7* transcript levels (Figure 5C), or the *COX1-MSY1* interaction frequency (Figure S5). Thus, strains lacking

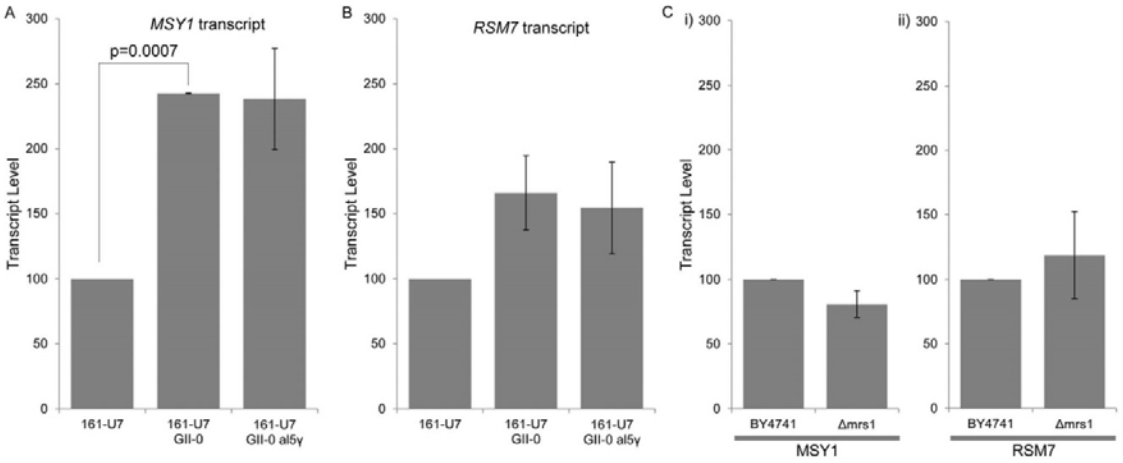


Figure 5. Knocking out mitochondrial encoded reverse transcriptase activity results in increased transcript levels of nuclear genes that are involved in Mito-nDNA interactions. A) Nuclear encoded *MSY1* transcript levels were determined by qRT-PCR in WT (strain 161-U7), 161-U7 GII-0 (lacks both the mitochondrial group II introns and the *COX1* interacting region; Figure 4A), and 161-U7 GII-0 a15γ (contains the interacting region and lacks the group II introns; Figure 4A) cells. B) Nuclear encoded *RSM7* transcript levels were determined by qRT-PCR in: WT (strain 161-U7); 161-U7 GII-0; and 161-U7 GII-0 a15γ cells. Neither 161-U7 GII-0 nor 161-U7 GII-0 a15γ has any alteration within the *Q0182* open reading frame. C) Deletion of *MRS1* (BY4741 Δ*mrs1*), a nuclear gene involved in splicing mitochondrial type-I introns, has no effect on i) *MSY1* or ii) *RSM7* transcript levels. All transcript levels were standardized to nuclear *ACT1* and expressed as percentage of wild-type (set at 100%) +/− standard error of the mean (n = 2).
doi:10.1371/journal.pone.0030943.g005

mitochondrial reverse transcriptase activity have lower frequencies of Mito-nDNA interactions and increased levels of nuclear encoded transcripts. These results suggest that cDNA mediated Mito-nDNA interactions are involved in the regulation of the nuclear transcripts, and therefore that the Mito-nDNA interactions we observed are biologically relevant.

The finding that inter-organelle interactions affect nuclear transcript levels necessarily predicts that the deletion of *yme1*, which reduces the frequency of the *COX1-MSY1* interaction (Figure 2), should correlate with an increase in nuclear *MSY1* transcript levels. Indeed increases in *MSY1* transcript levels, and other genes involved in mitochondrial gene expression and the biogenesis of the respiratory chain, have been identified within yeast cells containing the *yme1* deletion growing with a mixed respiro-fermentative metabolism [38].

Discussion

In this study we have shown that a large number of nuclear loci interact strongly and reproducibly with Mitochondrial DNA in *S. cerevisiae* and that the spectrum of these interactions is dependent upon the carbon source on which the yeast are grown. Interestingly, we find that Mito-nDNA interactions are significantly reduced when group II mitochondrial introns that contain reverse transcriptase machinery are deleted. This suggests that the mitochondrial DNA that is involved in these inter-organelle interactions is cDNA that has been reverse transcribed from mitochondrial RNAs. Critically, we demonstrate that suppression of inter-organelle DNA-DNA interactions correlates with elevated transcript levels for the interacting nuclear gene and a reproducible albeit small increase in growth rate (Figure S6), suggesting that these interactions are biologically relevant and play a role in regulating nuclear gene expression. This is further supported by previous observations that yeast nuclear transcription responds to the presence or absence of mitochondrial genome sequences [18,19]. From these results we propose that the Mito-nDNA interactions act as part of an inter-organelle communication system to signal mitochondrial metabolic state and regulate gene expression. While this DNA based inter-organelle communication may seem surprising, there is a large body of evidence demonstrating the presence of mitochondrial DNA in the nucleus and supporting the on-going nature of this transfer [7–9,39–45]. Thus, it is plausible that the process of mitochondrial DNA transfer has evolved into a functional signaling mechanism. In the case of the glucose and galactose dependent *COX1-MSY1* and *Q0182-RSM7* interactions we have shown a repressive role for Mito-nDNA interactions in the control of nuclear transcript levels. However, there is no reason to assume that all interactions are repressive.

The observation that mito-nDNA interactions correlated with chromosome length would be consistent with non-specific interactions. However, this relationship is also consistent with the hypothesis that the interactions are enriched with elements which are themselves evenly distributed across the genome. Yeast genes, ARS's, and nuclear encoded mitochondrial genes fulfill this criterion (Figure S7). Importantly, we observed that the nuclear fragments involved in the interactions are enriched for regions that overlap genes with mitochondrial functions in glucose ($p = 0.08$) and glycerol-lactate grown cells ($p < 10^{-8}$, Table S3), in agreement with earlier observations [17]. The condition specific significance of these enrichments can be interpreted as reflecting the regulatory roles of these interactions as part of a functional signaling mechanism.

Our global network analyses were obtained using proximity-based ligation methodologies (i.e. GCC and q3C) with a single restriction enzyme (*i.e. Msp*I). Duan *et al.* used different restriction enzymes (*i.e. Eco*RI and *Hind*III) to interrogate yeast genome structure [46]. A comparison of the interaction frequencies

between Duan *et al.*'s *Eco*RI results and ours showed that the data sets share a large portion of total interactions (Table S4) despite the fact that the interaction frequencies do not correlate quantitiatively (Figure S8). Duan *et al.'s* investigation of yeast genome organization did not report any Mito-nDNA interactions [46]. However, the absence of these interactions was due to the author's omission of the mitochondrial genome from their analyses because there is more than one copy of the mitochondrial genome per cell. This form of repetition is surmountable because interactions can be mapped onto a single scaffold, unlike elements which are repeated at different locations within the genome sequence. Re-examination of the datasets published by Duan *et al.* [46] identified a number of conserved Mito-nDNA between our datasets (Table S4). Interactions between *COX1-MSY1* and *Q0182-RSM7* (Data S1 and S2, respectively) were observed in both datasets, at levels above or just below the cut-off. The observation that Mito-nDNA interactions are present in these datasets is important for several reasons: 1) the methodologies while both based on proximity-ligation were considerably different, particularly with respect to enrichment and cell conditions; and 2) *Eco*RI cuts the yeast genome less frequently and at different positions to *Msp*I. While this goes some way towards relieving the likelihood that the interaction network is dependent on the restriction profile it does not completely alleviate this possibility, particularly because only one Mito-nDNA interaction was observed in the *Hind*III datasets (Data S3). The failure to identify more of these interactions in the *Hind*III datasets supports our conclusion that only parts of the mitochondrial genome are transferred into the nucleus and interact with the nuclear DNA. Because *Hind*III only cuts the yeast mitochondrial genome two times it is not unexpected that it does not cut within the transferred parts. Furthermore, the *Hind*III results support the finding that the interactions are not due to random ligation because random interactions would occur irrespective of fragment length.

It has been shown that DNA sequences which share stretches of sequence homology can associate *in vitro* [47,48]. Theoretically it is possible that such a mechanism could allow the sequence dependent association of fragments prior to ligation in our experiments. However, only three pairs of interaction fragments that we detected showed any homology upon blast analysis (Data S6) and these were not the interactions that we investigated. Therefore, this mechanism is unlikely to contribute significantly to our results.

Deletion of the group II introns suggests a role for mitochondrial encoded RNA intermediates in the inter-organelle interactions. This implies that the regions of the mitochondria that are involved in these interactions should predominantly be open reading frames (ORFs). However, there was no bias for mitochondrial ORFs being involved in the Mito-nDNA or Mito-rDNA interactions we observed. Superficially, this argues against the transfer occurring through reverse-transcribed cDNAs. However, yeast mitochondrial genes are transcribed as polycistronic transcripts from 14 ATATAAGTA consensus promoters and possibly another 5 non-consensus promoter sites around the mitochondrial genome [49]. Hence, a large percentage of the mitochondrial genome is physically transcribed [50–52] and therefore available to act as a template for reverse transcription.

The incomplete ablation of the interactions following the deletion of the mitochondrial group II introns raises the possibility that not all Mito-nDNA interactions involve reverse-transcribed mitochondrial sequences. If mitochondrial mini-circles contribute to inter-organelle interactions [8], we should have seen an increase in interaction frequency in the Δ*yme1* mutant. The fact that we did not indicates that this form of mitochondrial DNA transfer has a negligible role in the signaling pathway we are proposing. Ongoing transfer despite the loss of the mitochondrial reverse transcriptase can be explained by: 1) cytoplasmic or nuclear reverse transcription of mRNA released from damaged mitochondria, or 2) the presence of other retrotransposon or retroviral encoded reverse-transcriptase of either mitochondrial or nuclear origin within the mitochondrial matrix. Such a mechanism is supported by the identification of remnants of nuclear derived *copia*-, *gypsy*- and LINE-like retrotransposon elements within *Arabidopsis* mitochondria [53,54].

It is unclear whether the mtDNA that participates in the Mito-nDNA interactions is transferred by a direct connection between the mitochondrial and nuclear organelles or by uptake from the cytoplasm. Uptake from the cytoplasm is feasible given that unstable mitochondrial plasmids are first released into the cytoplasm [7,9], and high success rates are generally attained for yeast transformation which involves passage through the cytoplasm [55,56]. However, Ricchetti *et al.* demonstrated that the mtDNA mediated repair of nuclear double strand breaks is independent of Δ*yme1* mutations [6] and therefore occurs through another, possibly direct, transport mechanism. Direct transport from the mitochondria to nuclear compartments could occur as a result of a tethering/transport complex that physically links mitochondria to the endoplasmic reticulum [57].

In conclusion, several important questions are raised by this work. Firstly, what is the mechanism by which the Mito-nDNA interactions affect changes in transcript levels? Is the effect mediated by physical interaction between the mitochondrial derived cDNA and the nuclear locus or by more indirect means? It is possible that transcription factors may be sequestered or recruited to locations of activity through interactions with the mitochondrial DNA. Secondly, does this inter-organelle signaling pathway operate on a general level or function just to regulate conditionally essential genes (e.g. *RMS7* and *MSY1*)? Thirdly, while we demonstrated that specific interactions are dependent upon reverse transcription, whether this is true for all the Mito-nDNA and Mito-rDNA interactions remains to be determined. Finally, the universal significance of these interactions remains to be determined, particularly given the non-ubiquitous distribution of group II introns within higher eukaryotic mitochondria. The pervasive presence of reverse transcription within higher eukaryotic cells leads us to propose that this phenomenon will be widespread and that it deserves further investigation.

Materials and Methods

Strains and growth conditions

Saccharomyces cerevisiae strains (Table S5) were stored (−80°C) and cultured (30°C, 160 rpm) on synthetic complete (SC) medium containing amino acid supplements and glucose (2% w/v) [58], glycerol lactate (2% glycerol v/v 2% lactic acid v/v with 0.05% glucose w/v), or galactose (2% w/v). For Genome Conformation Capture (GCC) and Chromosome Conformation Capture (3 C) analyses, strains were recovered from −80°C on SC glucose (2% w/v) agar (2%) plates for 48 hours prior to starter culture inoculation. Starter cultures were grown (30°C, 160 rpm, 16 h) in SC glycerol lactate or glucose medium containing amino acid supplements, as indicated. Test cultures were inoculated, from the starter cultures into SC media (containing the indicated carbon source), grown (30°C, 160 rpm) and harvested at an optical density (OD_{600}) of 0.6. Mitochondrial uncoupling was achieved by the addition of 2,4-Dinitrophenol (5 mM final concentration) for 45, 90, or 180 minutes (Figure S4). Cell cycle arrest was achieved by treatment (180 mins, 30°C, 160 rpm) with cell cycle inhibitors

(*i.e.* α-factor (3.4 μm), nocodazole (15 $\mu g ml^{-1}$), or hydroxyurea (100 mM)). Cell cycle arrest was confirmed by microscopy.

Genome Conformation Capture (GCC)

GCC was performed according to [17]. Refer to the supplementary methods for a detailed description. Briefly, chromatin was prepared from 15 sets of 10^8 (*i.e.* a total of 1.36×10^9) cross-linked cells. Chromatin was digested with *MspI* (Fermentas) and ligated (T4 ligase; Invitrogen). Crosslinks were reversed in the presence of proteinase K (final concentration 7–11 μg, Roche). Samples were treated with RNase A (final concentration 10 $\mu g ml^{-1}$) prior to purification by phenol:chloroform (1:1 v/v, three times) and column extraction (Zymo Clean and Concentrator, Zymo Research). Paired-end sequencing (36 bp) was performed on 5 μg DNA using the Illumina Genome Analyzer platform (Allan Wilson Centre, Massey University, New Zealand & Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland).

External controls were added at two steps in the GCC protocol to control for random ligation events. The first ligation control, a linear DNA fragment with a free *MspI* site at one end (Methods S1), was added in a 1:1 ratio with the nuclear genome prior to the addition of ligase. The second ligation control (1×10^6 molecules of pUC19) was included prior to RNase A treatment as a control for the sequencing step ligation.

GCC Network Assembly

Network assembly was performed using Topography v1.19 (available on request [17]). The SOAP [59] algorithm was used to position paired end sequences and single ends, which contain an *MspI* restriction site, onto the *S. cerevisiae* reference genome (Methods S1). No mismatches were allowed.

Bioinformatic analyses

Bioinformatic and statistical analyses (see Methods S1) were performed on chromosomal interactions involving the nuclear and mitochondrial genomes for which the sequences mapped uniquely onto the reference genome. Connections with the ribosomal DNA (rDNA), 2 micron plasmid and mitochondrial genomes were considered as unique because they could be positioned to a ~1 MB region of Chromosome XII, the 6318 bp 2 micron plasmid or the ~85 kbp mitochondrial genome, respectively [17]. All statistical analyses involving 2 micron plasmid, mitochondrial, or rDNA sequences included copy number corrections (Methods S1). Other repetitive elements, such as LTRs and tRNAs, were omitted from the analysis.

Chromosome Conformation Capture (3 C)

3 C samples were prepared as previously described [60]. Refer to Methods S1 for a detailed description. Quantitative 3 C analyses [17] were performed using FAM labeled BHQ Probes (BioSearch Technologies; Table S2) and Taqman® Gene Expression Master Mix (Applied Biosystems) on an ABI Prism 7000 Sequence Detection System (SDS7000). Chromosomal coordinates for the interactions under investigation are listed in Methods S1. Samples (2 μl in triplicate) were analyzed in a final reaction volume of 20 μl using primers listed in Table S2. Assays were performed using a 3-stage program (50°C, 2:00 min; 95°C, 10:00 min; 45×[95°C, 0:15 sec; 60°C, 1:00 min]).

Dedicated interaction standards (concentration from: 2 $ng\ \mu l^{-1}$–2×10^{-15} $g\ \mu l^{-1}$) were prepared by PCR amplification (from *S. cerevisiae* BY4741) of the interacting regions, followed by *MspI* digestion and ligation of the two interacting partner fragments. Mitochondrial and nuclear genome (*i.e. GAL1*) copy number were determined by qPCR (Table S2) using Sybr-green and a five stage program (50°C, 2:00 min; 95°C, 2:00 min; 40× [95°C, 0:15 sec; 59.5°C, 0:30 sec; 72°C, 0:30 sec]; 55°C, 1:00; followed by a dissociation analysis) on an ABI Prism 7000 Sequence Detection System (SDS7000). An *S. cerevisiae* BY4741 genomic DNA sample (concentration from: 2 $ng\ \mu l^{-1}$–7.78125×10^{-4} $ng\ \mu l^{-1}$) was used as a control for all Sybr-green assays.

For comparison, all samples were presented as a percentage of wild-type, following standardization for: 1) the amount of a15γ intron-containing DNA (*i.e.* mitochondrial copy number); or 2) the number of nuclear genomes (determined using the single copy *GAL1* locus; [61]; Primer sequences are listed in Table S2. This standardization was performed to correct for alterations to mitochondrial genome stability and the rates of appearance of rho^- or rho^0 strains. This is critical as inter-organelle interactions are dependent upon the presence of the mitochondrial genome (see 161-U7 GII-0 results). The method of standardization depends upon the interaction being investigated (*i.e. COX1-MSY1* interactions were standardized by mitochondrial genome copy number while nuclear-nuclear locus interactions were standardized by *GAL1* copy number). No significant differences were observed when inter-organelle interactions were standardized by mitochondrial or nuclear copy number (data not shown).

RNA extraction

Total RNA was extracted from *S. cerevisiae* grown in SC (Glucose) to an OD_{600} of 0.600. Briefly, cells were harvested (4,000 rpm, 4°C, 2 min) and washed with AE buffer (4,000 rpm, 4°C, 2 min; 50 mM Sodium Acetate, 10 mM EDTA, pH 5.3). The cell pellet was suspended in phenol/chloroform/isoamyl alcohol (400 μl, 24/24/1) and glass beads (400 μl). Cells were lysed in a bead mill (SPEX sample prep 2010, Geno/Grinder; 1,750 rpm, 8×30 sec cycles with 60 sec resting intervals at 4°C). Lysed cells were frozen (−80°C, 15 min), thawed and pelleted (15,000 rpm, 5 min, 4°C). The aqueous phase was extracted twice with phenol/chloroform/isoamyl alcohol (400 μl, 24/24/1). Total RNA was pelleted (15,000 rpm, 10 min, 4°C), following addition of 2/3 s volume of 8 M LiCl and freezing (−20°C, 2 h). RNA was washed (70% ethanol), and the pellet air-dried. Total RNA was suspended (60°C, 10 min) in 80 μl of DECP treated water (Invitrogen). DNA was removed from the total RNA samples (5 μg, 20 μl) by treatment with 1 μl of TURBO DNase (TURBO DNA-*free*™ Kit, Ambion) as per manufacturer's instructions. Samples were centrifuged (10,000 g, 1.5 min) and the supernatant was retained. Total RNA concentration was measured using a Nano-drop and 50 μl samples (50 ng/μl) were stored at −80°C.

Quantitative Reverse Transcription-PCR

qRT-PCR standards were amplified from *S. cerevisiae* BY4741 genomic DNA (Table S2). PCR products were purified (Zymo DNA clean and concentrator™-5 kit according to manufacturer's instructions). The concentration of each qRT-PCR standard was determined by Nano-drop and used to make dilutions ranging from $4.0–4.0\times10^{-5}$ ng/μl. qRT-PCR reactions were performed using One Step SYBR® Ex Taq™ qRT-PCR Kit according to the manufacturer's instructions (TaKaRa). The qRT-PCR was run with the following protocol: 42°C, 5 min; 95°C, 10 sec; 40× [95°C, 5 sec; 60°C, 31 sec] 95°C, 15 sec; 60°C, 1 min; 95°C, 15 sec. All transcript levels were standardized to nuclear *ACT1* and expressed as percentage of wild-type (set at 100%) +/− standard error of the mean.

Supplementary files

The following additional data are available with the online version of this paper. Supplementary material file contains: Figures S1, S2, S3, S4, S5, S6, S7, S8, Tables S1, S2, S3, S4, S5, Methods S1 and Supplementary references. Additional data files: contain the analysis of the Duan data (Data S1, S2, S3, S4, S5); and the mitochondrial blast analyses (Data S6); a list of all uniquely mapping mito-nDNA and mito-mito interactions and their interaction strength (Data S7); a list of all mapped mito-nDNA and mito-mito interactions and their interaction strength (Data S8); and a list of nuclear genes with a mitochondrion annotation in gene ontology (Data S9). Sequences are available from GEO (accession number GSE34132).

Supporting Information

Figure S1 Biological Repeats correlate well at the *MspI* restriction fragment level. Two biological repeats were performed for each condition; A) glucose, B) glycerol lactate, and C) galactose. R^2 values are as follows; Glucose 0.78, glycerol lactate 0.93, and galactose 0.93. Scatter plots were constructed from statistically significant ($p = <0.0004$) interactions involving only *MspI* fragments which could be uniquely positioned on the reference genome. Adjacent interactions have been omitted as we are unable to distinguish between true adjacent interactions and those which are the result of simply sequencing across an uncut *MspI* site. Circularized fragments (*i.e.* self interactions) have also been omitted.
(DOC)

Figure S2 Inter-organelle interactions vary with metabolic state and do not occur evenly across the mitochondrial genome. Interaction frequency was graphed as a percentage of the total number of interactions in the sample, according to segment length. To test whether Mito-nDNA interactions have a uniform distribution (*i.e*, the total number of interactions in a segment is proportional to its length) we aggregated consecutive restriction fragments to create 58 sections that were expected to have at least 5 interactions under the null hypothesis of uniformity. A Chi-squared goodness of fit test was performed, and the distribution of the interactions was shown to deviate significantly from uniformity ($p<0.0001$, 57 df) for all conditions, Thus, Mito-nDNA interactions are not uniformly distributed across the mitochondrial genome. The linearized mitochondrial genome is shown for comparison of the interaction frequency with mitochondrial ORF and inter-genic sequence positions. Metabolic conditions were as follows: A) respiro-fermentation (glucose), B) respiro-fermentation (galactose), and C) respiration (glycerol lactate). Only statistically significant unique interactions between the mitochondrial genome and nuclear chromosomes were included in this analysis ($p \leq 10^{-5}$; $n = 2$). Interactions with the rDNA and 2-micron plasmid were removed. D) Nuclear genome interactions are not enriched over mitochondrial open reading frames. We compared the numbers of nuclear genome interactions with mitochondrial inter- and intra-genic regions to determine if the interactions across the mitochondrial genome were enriched over the open reading frames (ORFs). Galactose displays a larger number of interactions with mitochondrial ORFs but the difference is not statistically significant. Interactions were assigned proportionally to inter- and intra-genic regions to obtain a ratio of inter-genic to intra-genic interactions and expressed as percentages. tRNAs were not deemed intra-genic. Interestingly, the galactose sample exhibited 7% and 13% more inter-organelle interactions involving the *COX1* ORF than glycerol lactate and glucose, respectively. Thus, while there is no obvious preference for interactions with mitochondrial ORFs, interactions involving *COX1* show differences between the datasets.
(DOC)

Figure S3 The *COX1-MSY1* Mito-nDNA interaction is cell cycle dependent. Cells were synchronized at three different cell cycle phases by treatment with α-factor (3.4 μm), Hydroxyurea (100 mM), or Nocodazole (15 $\mu g ml^{-1}$); G1, S, G2/M, respectively). Mito-nDNA interaction frequency, between the representative mitochondrial and nuclear *MspI* fragments, was assayed by quantitative 3 C (see Methods S1). Interaction values were corrected for mitochondrial genome copy number (see Methods). Interaction values are expressed as percentages of the untreated sample (set at 100%) +/− standard error of the mean ($n = 3$).
(DOC)

Figure S4 5 mM 2,4-Dinitrophenol (DNP) inhibits respiratory growth but does not prevent growth of fermenting *S. cerevisiae* BY4741 cells. *S. cerevisiae* BY4741 cultures were grown (50 ml, 30°C, 160 rpm) on glucose (fermentation) or glycerol/lactate (respiration) to an Optical density (600 nm; OD_{600}) of 0.600. Cultures were diluted to an OD_{600} of 0.150 (50 ml final volume) in their respective media. 5 mM DNP (final concentration) was added to two of the cultures, while two remained untreated. The cell growth was monitored (OD_{600}) for a further 11.5 hours, with the exception of the untreated glucose culture which was only grown for 4 hours.
(DOC)

Figure S5 Deletion of *MRS1* (BY4741 Δ*mrs1*), a nuclear gene involved in splicing mitochondrial type-I introns, has no significant effect on the frequency of the *COX1-MSY1* interaction in glucose grown yeast cells. Interaction frequency was expressed as percentages of the wild type *S. cerevisiae* strain BY4741 (WT, set at 100%) +/− standard error of the mean ($n = 3$).
(DOC)

Figure S6 Deletion of group II introns results in an increase in growth rate. Growth rates were determined for *Saccharomyces cerevisiae* strains (161-U7, 161-U7 GII0, and 161-U7 GII0 +aI5γ; Figure 4a) grown in SC+2% glucose (30°C and 160 rpm). Cultures were inoculated to an initial optical density (OD_{600}) of 0.05 from overnight cultures. The OD_{600} was measured every two hours for 10 hour. Data represent the mean ± SD ($n = 3$).
(DOC)

Figure S7 ARS and ORF numbers correlate with chromosome size. Data on ARS and ORF numbers and chromosome size were taken from the Saccharomyces genome database Genome Inventory (as of Nov 03, 2011). The length of chromosome XII was calculated based on it containing only two copies of the rDNA repeat.
(DOC)

Figure S8 Comparison of the total interaction frequencies for the Glucose derived GCC data (this study) and Duan *et al.* *EcoRI* derived datasets.
(DOC)

Table S1 Mitochondrial copy number calculations.
(DOC)

Table S2 Primers and probes used in this study.
(DOC)

Table S3 Nuclear fragments involved in mito-nDNA interactions are enriched for regions that overlap genes

with mitochondrial functions. The percentage of nuclear fragments that overlap with nuclear encoded mitochondrial genes within the complete genome was calculated and compared to the percentage of nuclear fragments involved in mito-nDNA interactions that overlap with nuclear encoded mitochondrial genes. A test of proportions (prop.test) was performed in R to determine whether the percentage difference is significant, p-values are shown.
(DOC)

Table S4 Comparison between the *Eco*RI interaction set from Duan *et al.* [46] and the glucose set from this study.
(DOC)

Table S5 Strains used in this study.
(DOC)

Methods S1 This file contains supplementary information for methods used in this manuscript.
(DOC)

Data S1 This text file contains the significant interactions that were identified as occurring between the mitochondrial genome and the region surrounding the nuclear *MSY1* locus in the dataset prepared by Duan *et al.* [46]. The number of instances for any interaction that had to be seen was set at >=3 for the *Mse*1 and >=4 for the *Msp*1 datasets.
(TXT)

Data S2 This text file contains the significant interactions that were identified as occurring between the mitochondrial genome and the region surrounding the nuclear *RSM7* locus in the dataset prepared by Duan *et al.* [46]. The number of instances for any interaction that had to be seen was set at >=3 for the *Mse*1 and >=4 for the *Msp*1 datasets.
(TXT)

Data S3 This text file contains the one interaction that was identified as occurring between the mitochondrial and nuclear genomes in chromatin cut with *Hin*dIII and subsequently *Mse*1.
(DAT)

Data S4 This file contains all the interactions that were identified by re-analysis of the Duan *et al.* [46] datasets for chromatin digested with *Eco*RI and subsequently *Mse*1. All listed interactions were all above the cut-off which was set at >=3.
(DAT)

Data S5 This file contains all the interactions that were identified by re-analysis of the Duan *et al.* [46] datasets for chromatin digested with *Eco*RI and subsequently *Msp*1. All listed interactions were all above the cut-off which was set at >=4.
(DAT)

Data S6 This file contains the results of a blastN analysis of mito-nDNA interacting fragments. Only three pairs of interaction fragments that we detected showed any homology upon blast analysis. The smaller restriction fragment from each interaction pair was compared to the longer fragment by blastn using default parameters. The length, score, and evalue for each comparison was recorded. Comparisons that showed no similarity were given an evalue score of 10.
(CSV)

Data S7 This text file contains a list of all uniquely mapped mito-nDNA and mito-mito interactions and their interaction strength.
(XLSX)

Data S8 This text file contains a list of all mapped mito-nDNA and mito-mito interactions and their interaction strength. The interactions that are included in this list include those which mapped uniquely and were repetitive.
(XLSX)

Data S9 This text file contains a list of all genes that have a mitochondrion gene ontology annotation. The gene list was obtained from YeastMine (http://yeastmine.yeastgenome.org/yeastmine/begin.do).
(CSV)

Acknowledgments

The authors wish to thank A. Ganley for critical discussions of this work and manuscript. We also wish to thank, E. Sattlegger, P. Perlman, G. Mohr, and A. Lambowitz for yeast strains. Additionally, the authors would like to thank S. Gasser and M. Tsai (Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland) for help with sequencing.

Author Contributions

Conceived and designed the experiments: CR RG JO. Performed the experiments: CR RG. Analyzed the data: CR RG LG GG MJ JO. Contributed reagents/materials/analysis tools: CR RG LG GG MJ JO. Wrote the paper: CR RG LG JO.

References

1. Foury F, Roganti T, Lecrenier N, Purnelle B (1998) The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae. FEBS Lett 440: 325–331.
2. Lagunas R (1976) Energy metabolism of Saccharomyces cerevisiae discrepancy between ATP balance and known metabolic functions. Biochimica et Biophysica Acta (BBA) - Bioenergetics 440: 661–674.
3. Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5: 123–135.
4. Butow RA, Avadhani NG (2004) Mitochondrial Signaling: The Retrograde Response. Molecular Cell 14: 1–15.
5. Liu Z, Butow RA (1999) A Transcriptional Switch in the Expression of Yeast Tricarboxylic Acid Cycle Genes in Response to a Reduction or Loss of Respiratory Function. Mol Cell Biol 19: 6720–6728.
6. Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 402: 96–100.
7. Thorsness PE, Fox TD (1990) Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. Nature 346: 376–379.
8. Thorsness PE, Fox TD (1993) Nuclear mutations in Saccharomyces cerevisiae that affect the escape of DNA from mitochondria to the nucleus. Genetics 134: 21–28.
9. Campbell CL, Thorsness PE (1998) Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. J Cell Sci 111(Pt 16): 2455–2464.
10. Shafer KS, Hanekamp T, White KH, Thorsness PE (1999) Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae. Curr Genet 36: 183–194.
11. Cheng X, Ivessa AS (2010) The migration of mitochondrial DNA fragments to the nucleus affects the chronological aging process of *Saccharomyces cerevisiae*. Aging Cell 9: 919–923.
12. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10: 1453–1465.
13. Spilianakis CG, Flavell RA (2004) Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nature Immunology 5: 1017–1027.

14. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA (2005) Interchromosomal associations between alternatively expressed loci. Nature 435: 637–645.
15. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295: 1306–1311.
16. Fujita N, Wade PA (2004) Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. Methods 33: 81–85.
17. Rodley CD, Bertels F, Jones B, O'Sullivan JM (2009) Global identification of yeast chromosome interactions using Genome conformation capture. Fungal Genet Biol 46: 879–886.
18. Parikh VS, Conrad-Webb H, Docherty R, Butow RA (1989) Interaction between the yeast mitochondrial and nuclear genomes influences the abundance of novel transcripts derived from the spacer region of the nuclear ribosomal DNA repeat. Mol Cell Biol 9: 1897–1907.
19. Parikh V, Morgan M, Scott R, Clements L, Butow R (1987) The mitochondrial genotype can influence nuclear gene expression in yeast. Science 235: 576–580.
20. Blank HM, Li C, Mueller JE, Bogomolnaya LM, Bryk M, et al. (2008) An increase in mitochondrial DNA promotes nuclear DNA replication in yeast. PLoS Genet 4: e1000047.
21. Lagunas R (1986) Misconceptions about the energy metabolism of Saccharomyces cerevisiae. Yeast 2: 221–228.
22. Otterstedt K, Larsson C, Bill RM, Stahlberg A, Boles E, et al. (2004) Switching the mode of metabolism in the yeast Saccharomyces cerevisiae. Embo Reports 5: 532–537.
23. Campbell C, Tanaka N, White K, Thorsness P (1994) Mitochondrial morphological and functional defects in yeast caused by yme1 are suppressed by mutation of a 26S protease subunit homologue. Mol Biol Cell 5: 899–905.
24. Neutzner A, Youle RJ (2005) Instability of the mitofusin Fzo1 regulates mitochondrial morphology during the mating response of the yeast Saccharomyces cerevisiae. J Biol Chem 280: 18598–18603.
25. Fekkes P, Shepard KA, Yaffe MP (2000) Gag3p, an outer membrane protein required for fission of mitochondrial tubules. J Cell Biol 151: 333–340.
26. Naylor K, Ingerman E, Okreglak V, Marino M, Hinshaw JE, et al. (2006) Mdv1 interacts with assembled dnm1 to promote mitochondrial division. J Biol Chem 281: 2177–2183.
27. Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, et al. (1995) A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell 83: 529–538.
28. Zimmerly S, Guo H, Perlman PS, Lambowitz AM (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. Cell 82: 545–554.
29. Yang J, Zimmerly S, Perlman PS, Lambowitz AM (1996) Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. Nature 381: 332–335.
30. Eskes R, Yang J, Lambowitz AM, Perlman PS (1997) Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. Cell 88: 865–874.
31. Seraphin B, Faye G, Hatat D, Jacq C (1992) The yeast mitochondrial intron aI5 alpha: associated endonuclease activity and in vivo mobility. Gene 113: 1–8.
32. Moran JV, Zimmerly S, Eskes R, Kennell JC, Lambowitz AM, et al. (1995) Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. Mol Cell Biol 15: 2828–2838.
33. Koch JL, Boulanger SC, Dib-Hajj SD, Hebbar SK, Perlman PS (1992) Group II introns deleted for multiple substructures retain self-splicing activity. Molecular and Cellular Biology 12: 1950–1958.
34. Griffin EA, Jr., Qin Z, Michels WJ, Jr., Pyle AM (1995) Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. Chemistry & biology 2: 761–770.
35. Toor N, Hausner G, Zimmerly S (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA 7: 1142–1152.
36. Bassi GS, de Oliveira DM, White MF, Weeks KM (2002) Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. Proc Natl Acad Sci U S A 99: 128–133.
37. Turk EM, Caprara MG (2010) Splicing of yeast aI5beta group I intron requires SUV3 to recycle MRS1 via mitochondrial degradosome-promoted decay of excised intron ribonucleoprotein (RNP). J Biol Chem 285: 8585–8594.
38. Arnold I, Wagner-Ecker M, Ansorge W, Langer T (2006) Evidence for a novel mitochondria-to-nucleus signalling pathway in respiring cells lacking i-AAA protease and the ABC-transporter Mdl1. Gene 367: 74–88.
39. Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular Poltergeists: Mitochondrial DNA Copies in Sequenced Nuclear Genomes. PLoS Genet 6: e1000834.
40. Farrelly F, Butow RA (1983) Rearranged mitochondrial genes in the yeast nuclear genome. Nature 301: 296–301.
41. Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD (2002) Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U S A 99: 9905–9912.
42. Adams KL, Song K, Roessler PG, Nugent JM, Doyle JL, et al. (1999) Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes. Proc Natl Acad Sci U S A 96: 13863–13868.
43. Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. FEBS Letters 325: 140–145.
44. Richly E, Leister D (2004) NUMTs in Sequenced Eukaryotic Genomes. Mol Biol Evol 21: 1081–1084.
45. Shirafuji N, Takahashi S, Matsuda S, Asano S (1997) Mitochondrial Antisense RNA for Cytochrome C Oxidase (MARCO) Can Induce Morphologic Changes and Cell Death in Human Hematopoietic Cell Lines. Blood 90: 4567–4577.
46. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, et al. (2010) A three-dimensional model of the yeast genome. Nature 465: 363–367.
47. Baldwin GS, Brooks NJ, Robson RE, Wynveen A, Goldar A, et al. (2008) DNA double helices recognize mutual sequence homology in a protein free environment. The journal of physical chemistry B 112: 1060–1064.
48. Danilowicz C, Lee CH, Kim K, Hatch K, Coljee VW, et al. (2009) Single molecule detection of direct, homologous, DNA/DNA pairing. Proceedings of the National Academy of Sciences of the United States of America 106: 19824–19829.
49. Christianson T, Rabinowitz M (1983) Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. J Biol Chem 258: 14025–14033.
50. Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G (1982) Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc Natl Acad Sci U S A 79: 7195–7199.
51. Holec S, Lange H, Kuhn K, Alioua M, Borner T, et al. (2006) Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. Mol Cell Biol 26: 2869–2876.
52. Barth C, Greferath U, Kotsifas M, Fisher PR (1999) Polycistronic transcription and editing of the mitochondrial small subunit (SSU) ribosomal RNA in Dictyostelium discoideum. Curr Genet 36: 55–61.
53. Knoop V, Unseld M, Marienfeld J, Brandt P, Sunkel S, et al. (1996) copia-, gypsy- and LINE-like retrotransposon fragments in the mitochondrial genome of Arabidopsis thaliana. Genetics 142: 579–585.
54. Marienfeld J, Unseld M, Brennicke A (1999) The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. Trends in Plant Science 4: 495–502.
55. Costanzo MC, Fox TD (1988) Transformation of yeast by agitation with glass beads. Genetics 120: 667–670.
56. Ito H, Fukuda Y, Murata K, Kimura A (1983) Transformation of intact yeast cells treated with alkali cations. J Bacteriol 153: 163–168.
57. Kornmann B, Currie E, Collins SR, Schuldiner M, Nunnari J, et al. (2009) An ER-mitochondria tethering complex revealed by a synthetic biology screen. Science 325: 477–481.
58. Kaiser CSMMA (1994) Methods in Yeast Genetics, A Cold Spring Harbor Laboratory Course Manual. Cold Spring Harbor, NY.: Cold Spring Harbor laboratory press.
59. Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. Bioinformatics 24: 713–714.
60. O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, et al. (2004) Gene loops juxtapose promoters and terminators in yeast. Nat Genet 36: 1014–1018.
61. O'Sullivan JM, Sontam DM, Grierson R, Jones B (2009) Repeated elements coordinate the spatial organization of the yeast genome. Yeast 26: 125–138.

10 APPENDIX IV

Biosci. Biotechnol. Biochem., **77** (2), 402–404, 2013

Note

Desalting DNA by Drop Dialysis Increases Library Size upon Transformation

Mayank SARASWAT,[1] Ralph. S. GRAND,[1] and Wayne M. PATRICK[1,2,†]

[1]*Institute of Natural Sciences, Massey University, Private Bag 102 904, Auckland 0745, New Zealand*
[2]*Department of Biochemistry, University of Otago, PO Box 56, Dunedin 9054, New Zealand*

Received October 4, 2012; Accepted November 18, 2012; Online Publication, February 7, 2013
[doi:10.1271/bbb.120767]

It is often desirable to obtain gene libraries with the greatest possible number of variants. We tested two different methods for desalting the products of library ligation reactions (silica-based microcolumns and drop dialysis), and examined their effects on final library size. For both intramolecular and intermolecular ligation, desalting by drop dialysis yielded approximately 3–5 times more transformants than microcolumn purification.

Key words: directed evolution; drop dialysis; library construction; ligation; microcolumn

In many molecular biology protocols, it is useful, or even essential, to maximize the number of colonies that result from a cloning experiment (*i.e.*, ligation, desalting, and transformation). This is especially true in the field of directed evolution, in which the Darwinian principles of random mutagenesis and selection are used to identify proteins with new or improved properties. The first step in a directed evolution experiment is to introduce molecular diversity into parental gene sequences. Many random mutagenesis protocols have been developed,[1,2] including methods for random point mutagenesis (*e.g.*, error-prone PCR), random homologous recombination (*e.g.*, DNA shuffling), and random non-homologous recombination (*e.g.*, Incremental Truncation for the Creation of Hybrid enzYmes, ITCHY). The next step is to capture the diversity that has been generated, by cloning the pool of mutagenized DNA molecules into an appropriate expression vector. Transformation of a suitable host, typically *Escherichia coli*, with the cloned DNA yields a library that can be stored for use in downstream screening or selection steps. Often, the library screen or selection is very high-throughput; it is common to design directed evolution experiments with the capacity to interrogate millions, or even billions, of variants. In these high-throughput cases, it is critical to optimize the library cloning and transformation steps, because large, diverse libraries are the most likely to include variants with improvements in the desired property.[3]

Electroporation is the method of choice for transforming *E. coli* with the products of library ligation reactions. Very high transformation efficiencies can be obtained,[4] resulting in large libraries. However, the high electric field strengths that are used (12–18 kV/cm) mean that efficient transformation requires low-conductivity samples, to prevent arcing. Hence, each library ligation reaction must be desalted prior to electroporation. This purification step is commonly carried out with silica-based microcolumns. A previous study that tested electroporation efficiencies with intact plasmid DNA (rather than with the products of ligation reactions) showed the microcolumn desalting method to be highly effective.[5] However, an older study showed that drop dialysis, in which a 5–100 µL drop is placed on a floating membrane filter, can also result in effective desalting, with DNA recovery rates of 98–99%.[6] Here we present a rigorous comparison of the two desalting methods, and show that drop dialysis is preferred in the construction of large libraries.

First, two parental sequences were randomly recombined using ITCHY: (i) the *trpF* portion of the bifunctional *E. coli trpCF* gene (GenBank accession no. NP_415778); and (ii) a cDNA encoding residues 36-367 of the β2 subunit of the *Rattus norvegicus* voltage-gated potassium channel (Kvβ2; GenBank accession no. NM_017304). The TrpF and Kvβ2 proteins share the same $(\beta\alpha)_8$ barrel fold, but their sequences are highly divergent (<10% sequence identity). The genes were each cloned into vector pInSALect[7] and then recombined onto the same linearized vector molecule by PCR, as described previously.[8] Our library was constructed using ITCHY with time-dependent truncation.[9]

At the final stage of library construction, three identical intramolecular ligations (90 µL) were set up, each of which comprised blunt-ended DNA from the ITCHY protocol (180 ng), 1× Fermentas T4 DNA Ligase Buffer, and 30 units of T4 DNA ligase (Fermentas, Vilnius, Lithuania). The ligation reactions were incubated at 16 °C for 16 h, and then heat inactivated (65 °C, 10 min). After heat inactivation, each of the three reactions was split into three 30-µL aliquots. One aliquot was desalted using a microcolumn (EZNA MicroElute Cycle Pure kit; Omega Bio-Tek, Norcross, GA, USA) following the manufacturer's guidelines. The desalted DNA was eluted from the column with 30 µL of elution buffer (10 mM Tris, pH 8.5). The second aliquot was desalted by drop dialysis. A standard Petri dish was half-filled with 30 mL of deionized (Milli-Q) water. A mixed cellulose ester membrane filter (pore size 0.025 µm, diameter 25 mm, MF-Millipore, Billerica, MA, USA) was floated on the water. The 30-µL aliquot of the ligation reaction was pipetted onto the membrane, covered with the lid of the Petri dish, and left to dialyze for 1 h. After dialysis, the desalted sample was recovered

† To whom correspondence should be addressed. Tel: +64-3-4797897; Fax: +64-3-4797866; E-mail: wayne.patrick@otago.ac.nz
Abbreviations: ITCHY, incremental truncation for the creation of hybrid enzymes; PCR, polymerase chain reaction

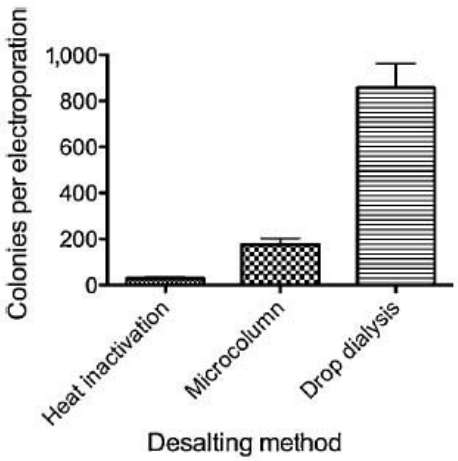


Fig. 1. Comparison of DNA Desalting Methods in the Construction of ITCHY Libraries.

Intramolecular ITCHY ligation reactions were heat inactivated only, purified with a silica-based microcolumn, or purified by drop dialysis. Aliquots of the desalted reaction products were used to transform *E. coli* by electroporation, and the numbers of colonies on dilution plates were used to estimate the total number of transformants resulting from each electroporation. Colony counts plotted are mean ± SEM for three independent ligation reactions.

from the top of the membrane, and the volume of the sample was adjusted to 30 µL with water. The third 30-µL sample from each ligation was not desalted further. Aliquots (2 µL) of each desalted library ligation and the heat inactivated controls were used to transform 50-µL aliquots of *E. coli* DH5α-E (Invitrogen, Carlsbad, CA, USA), by electroporation. SOC medium (500 µL) was added to each aliquot of cells immediately after pulsing. The transformed cells were allowed to recover at 37 °C with shaking for 1 h, and then aliquots were spread on LB-agar plates containing chloramphenicol (34 µg/mL). Colonies were counted after 16 h of incubation at 37 °C.

Figure 1 shows the mean numbers of colonies that resulted when 6.7% of each desalted sample (2 µL of a 30 µL total volume) was used to transform *E. coli* by electroporation. On average, microcolumn purification yielded 6.4-fold more colonies than heat inactivation (without further desalting). However, drop dialysis yielded the greatest number of colonies (4.8-fold more than microcolumn purification).

Had the desalted samples from the triplicate ligations been pooled, there would have been 84 µL of each sample (heat inactivated, microcolumn purified, and dialyzed) remaining. Transforming more aliquots of electrocompetent *E. coli* with all of this material would have yielded libraries with total sizes of approximately 1.3×10^3 variants (heat-inactivated ligation), 8.0×10^3 variants (microcolumn purification), and 3.9×10^4 variants (drop dialysis) respectively. The total number of possible variants in an ITCHY library is given by the product of the lengths of the two parental genes. In this case, the number of possible variants is 597 bp × 996 bp = 594,612. Our library analysis program, GLUE,[10] estimates that 1.8×10^6 clones would be required in our ITCHY library in order to sample 95% of all possible variants. In this example, none of the three desalting methods led to a library of that size. However,

Table 1. Comparison of Methods for Desalting the Products of Intermolecular (Vector + Insert) Ligation Reactions

Desalting method	Colonies per electroporation*
Microcolumn	870 ± 50
Drop dialysis	2,930 ± 540

*Mean ± SEM for three independent ligation reactions.

our analysis with GLUE suggests that the library from drop dialysis will include approximately 6.3% of all possible variants, which is certainly preferable to the other alternatives: 1.3% of all possible variants when microcolumn purification is used; and only 0.2% of all possible variants when the ligation is heat-inactivated but not desalted.

In a second experiment, we tested the effect of varying the desalting method on the outcome of intermolecular (vector + insert) ligation. This mimics the construction of an error-prone PCR library. A 338-bp DNA fragment was ligated with the 4.3-kb expression vector pLAB101[11] after each had been digested with restriction enzymes NdeI and SpeI (both New England Biolabs, Ipswich, MA, USA). Three ligation reactions were performed in a total volume of 60 µL per reaction. Each reaction contained 150 ng of vector DNA, 35 ng of insert DNA (a 3-fold molar excess of insert over vector), 1× T4 DNA Ligase Buffer (Fermentas), and 20 units of T4 DNA ligase (Fermentas). The reactions were incubated at 16 °C for 18 h and then heat inactivated (65 °C, 10 min). Each of the three ligation reactions was split into two 30-µL aliquots. One of the two 30-µL aliquots was desalted with a microcolumn, and the other was desalted by drop dialysis (see above). Aliquots (2 µL) of the desalted reactions were used to transform *E. coli* DH5α-E by electroporation, as described above. Dilutions of the transformed cells were spread on LB-agar plates containing carbenicillin (100 µg/mL) and colonies were counted after 16 h of incubation at 37 °C. Desalting the intermolecular ligation reactions by drop dialysis yielded 3.4-fold more colonies than microcolumn purification (Table 1). This is similar to the 4.8-fold improvement observed for intramolecular ITCHY ligations (Fig. 1).

Our data indicate that drop dialysis is a highly effective method for desalting DNA, confirming earlier results.[6] In both intramolecular and intermolecular ligation tests, we obtained approximately 3–5 times more transformants when we desalted by drop dialysis, as compared with the more commonly used silica-based microcolumn purification. In this study, we tested a single brand of microcolumn (the EZNA MicroElute Cycle Pure kit from Omega Bio-Tek). However, in preliminary desalting tests with intact plasmid DNA (rather than with library ligations), conducted as described previously,[5] we found that this microcolumn and its associated purification protocol yielded identical results to a well-known but more expensive alternative (the QIAquick PCR Purification Kit, Qiagen, Valencia, CA). Hence, drop dialysis remains the superior protocol for desalting ligation reactions, regardless of the microcolumn to which it is compared. Further, a previous study found that varying the membrane filter from one with an average pore diameter of 0.01 µm to one with an average pore diameter of 0.05 µm did not change the

effectiveness of the drop dialysis protocol, although the use of membranes with very small pore diameters can increase the time required for complete removal of buffer salts.[6] The membranes used in our experiments (average pore diameter, 0.025 µm) allow for rapid dialysis while minimizing the likelihood that ligation products are lost.

Finally, drop dialysis requires less hands-on time than microcolumn purification. The membrane filter discs do require careful pipette handling as the samples are loaded onto the membrane. They are also more expensive than microcolumns (NZ$7.48 per membrane *versus* NZ$3.40 per microcolumn at the time of writing), although expert users can reduce the cost by desalting 2–4 reactions on a single membrane simultaneously. Overall, the extra care and costs required (as compared with microcolumn purification) are likely to be warranted for practitioners of directed evolution, for whom the largest possible libraries are often desirable.

Acknowledgments

We gratefully acknowledge financial support for this study from the Marsden Fund. M.S. and R.S.G. were also supported by Massey University Doctoral Scholarships.

References

1) Lutz S and Patrick WM, *Curr. Opin. Biotechnol.*, **15**, 291–297 (2004).
2) Otten LG and Quax WJ, *Biomol. Eng.*, **22**, 1–9 (2005).
3) Patrick WM, Firth AE, and Blackburn JM, *Protein Eng.*, **16**, 451–457 (2003).
4) Dower WJ, Miller JF, and Ragsdale CW, *Nucleic Acids Res.*, **16**, 6127–6145 (1988).
5) Schlaak C, Hoffmann P, May K, and Weimann A, *Biotechnol. Lett.*, **27**, 1003–1005 (2005).
6) Marusyk R and Sergeant A, *Anal. Biochem.*, **105**, 403–404 (1980).
7) Gerth ML, Patrick WM, and Lutz S, *Protein Eng. Des. Sel.*, **17**, 595–602 (2004).
8) Gerth ML and Lutz S, *J. Mol. Biol.*, **370**, 742–751 (2007).
9) Ostermeier M and Lutz S, *Methods Mol. Biol.*, **231**, 129–141 (2003).
10) Firth AE and Patrick WM, *Bioinformatics*, **21**, 3314–3315 (2005).
11) Gerth ML, Nigon LV, and Patrick WM, *Protein J.*, **31**, 359–365 (2012).