Unravelling the genomic structure of *Saccharomyces cerevisiae*

Christopher David Melchoir Rodley

BSc (Hons)

A thesis submitted for the degree of

Doctor of Philosophy

at Massey University, Auckland,

New Zealand

2011

Abstract

Advances in high-throughput sequencing technology have enabled the sequencing of genomes for many organisms. But the ability to describe the linear arrangement of genetic elements on the chromosomes does not tell us much about how all of these elements work in concert to form and maintain a functional cell. To get some way towards a more holistic understanding of how this is achieved requires the elucidation of three-dimensional genome organisation.

The association of chromosomes with each other and other nuclear components plays a critical role in nuclear organisation and genome function. Interactions, which can be structural or functional in nature, form between different parts of the genome. Chromosomal interactions can be broadly divided into two groups, inter- and intra-chromosomal interactions, depending upon whether the interaction forms between different chromosomes or within a single chromosome, respectively.

Here I describe a methodology capable of capturing these interactions on a global scale, Genome Conformation Capture (GCC), and reveal the interaction network for *Saccharomyces cerevisiae.* The inter- and intra- chromosomal interactions detected by GCC are non-random and include contacts between the nuclear chromosomes, 2-micron plasmid, and the mitochondrial genome. These results formed the first global map of chromosomal interactions in a eukaryotic nucleus and demonstrated the highly connected nature of the yeast genome.

I subsequently performed GCC on *S. cerevisiae* cells grown on glucose, galactose, and glycerol lactate to investigate how genome organisation alters depending upon the metabolic regime being employed. I describe the difference in the numbers and types of interactions that form in the three conditions and investigate interactions involving transfer RNAs in detail. Interactions between the mitochondrial and nuclear genomes undergo significant changes depending upon the carbon source on which the yeast is grown.

The nuclear and mitochondrial organelles must maintain a communication system in order to respond effectively to environmental conditions. Previous studies have identified mitochondrial DNA inside the nucleus and interacting with the nuclear chromosomes. How this transfer occurs and what the function of the mitochondrial

DNA is once inside the nucleus remains unclear. Here I isolate interactions between the mitochondrial and nuclear genomes and demonstrate dependence upon mitochondrial encoded reverse transcriptase machinery. Furthermore, the nuclear gene transcript level is altered when the interaction frequency between the mitochondrial and nuclear genome is reduced. I conclude that mitochondrial DNA interactions with the nuclear genome are biologically relevant and that the results argue for a role for reverse transcription in inter-organelle DNA mediated communication.

The results presented in this thesis have significant implications for our understanding of eukaryotic genome organization.

Acknowledgements

I wish to thank my supervisor, Dr. Justin O'Sullivan, for your ideas, continual support, and the passing on of your skills; it goes without saying that this thesis would not have occurred without you. I would also like to thank my secondary supervisor, Prof. Paul Rainey, for critical discussion of the work, your input and support. To my mentor Mr. Donald Ripia, thank you for your support and resources over the years, it was very much appreciated.

I would like to thank the Health Research Council of New Zealand for funding my PhD. And also a big thank you to collaborators Prof. Susan Gasser, and Prof. David Engelke, for DNA sequencing and taking an interest in our research.

A big thank you to Dr. Beatrix Jones for statistical support and always making time for my problems. I'd like thank Dr. Evelyn Sattlegger for the use of the yeast knock out collection and Dr. Andrew Cridge for retrieving strains. I also wish to thank Dr. Austen Ganley and Dr. Wayne Patrick for your support and critical discussions of the research and manuscripts.

I would like to extend a warm thank you to Prof. Alan Lambowitz and Dr. Georg Mohr at the University of Texas, Austin, for providing yeast strains. Additionally, a big thank you must be extended to Prof. Philip S. Perlman at the Howard Hughes Medical Institute in Chevy Chase, for recommending yeast strains.

Soon to be Dr. Ralph Grand, thank you for your friendship, diving, boxing, perl scripts and critical discussion of the research. And thank you to Dr. Lutz Gehlen for your friendship, writing one-off perl scripts for me, critical discussion of research, and help with the statistical analyses. Also I would like to thank Jyothsna Visweswaraiah and Martina Dautel for their help over the years. The remainder of buildings 10, 11, 46 and 48, thank you for your friendship and support over the years.

I would like to give a special thank you to Jarod Young, for your support throughout this process and sharing your worldly views with me. We will have a pre-election test instated at some point.

I must thank the members of Te Huarahi, James Hudson and Veroncia Tawhai, your support was very much needed and appreciated. I hope the Te Huarahi concept continues.

To my parents, Bruce and Sam, without you this PhD would not have been possible; your continuing support of me is much appreciated, and for this I extend the biggest thank you.

I would like to thank Cliff and Amy Dawson for their amazing friendship, warm meals and all the laughs.

And finally, to my fiancée Anne-Sophie, thank you for teaching me how to balance work with life. Thank you for your support, and your love.

Table of Contents

Tables

Figures

Abbreviations

amp	Ampicillin
bp	Base Pairs
Da	Daltons
°C	Degrees Celsius
df	Degrees of freedom
FPR	False positive rate
g	G-Force
hr	Hour
hrs	Hours
kb	Kilobases
µg	Micrgrams
µl	Microlitres
ml	Millilitres
mmol	Millimole
mQ	MilliQ
min	Minute
nm	Nanometres
NTC	No Template Control
O/D	Optical Density
O/N	Overnight
PCR	Polymerase Chain Reaction
rpm	Revolutions per Minute
rm	Room
SDS	Sodium dodecyl sulphate
temp	Temperature
U	Units
V	Volts
v/v	Volume per Volume
H_2O	Water
w/v	Weight per Volume
mtDNA	Mitochondrial DNA
Mito-gDNA	Mitochondrial genome to nuclear genome interaction
Mito-Plas	Mitochondrial genome to 2-micron plasmid interaction
gDNA-gDNA	Interaction between loci within the nuclear genome

Chapter 1 Introduction

In a eukaryotic cell chromosomes exist inside the nucleus, while extra-nuclear DNA is harboured within organelles such as the mitochondria. The collective DNA content of a cell makes up its genome. While the DNA sequence of a genome indicates the linear arrangement of genetic elements (*e.g.* genes, promoters, enhancers *etc.*), it does not allow an understanding of how all of these factors are coordinated to become a functional cell. Recent data has implicated the spatial organisation and three dimensional structure of the genome as important for gene regulation and ultimately realising particular transcriptional profiles [20-27]. In order to gain an understanding of how this nuclear activity, and the regulation of it, is achieved, the spatial organisation of the nucleus and in particular, the organisation of the genome in three-dimensions, needs to be determined.

1.1 Genomes have three-dimensional structure

As early as the 1970s it was evident that genomes had some form of three-dimensional global structure. When bacterial cells were lysed in high salt, non-membrane bound nucleoid structures were visible by microscopy [28]. These nucleoids were shown to consist of polymerases still engaged to their templates, surrounded by a 'halo' of super-helical genomic DNA [28]. Furthermore, in mammalian cells the chromosomes have definite three-dimensional structure during metaphase as they are observed to form a classical X shape. This metaphase organisation results from the heavy compaction of the individual chromosomes as part of the cell division process.

Cells display different transcriptional profiles depending on cell type and environmental conditions [29]. Nuclear processes such as transcription [25, 30-33], replication [4, 26, 34, 35], and DNA repair [36, 37], can theoretically shape genome structure and as a result, genome structure can respond to environmental cues [16, 24, 27, 38-41]. Tom Mistelli (2007) describes the organisation of the genome inside the nucleus as having three hierarchal levels: 1) the spatial organisation of nuclear processes (transcription, replication *ect.*); 2) the organisation of chromatin into domains and; 3) the arrangement of genes and chromosomes in nuclear space [42]. Each of these three levels is dynamic and has the possibility of being regulated. Therefore, one can surmise that the structure of the genome is not fixed, permanent or undeviating [34, 43, 44].

1.1.1 How is DNA packaged within the nucleus?

1.1.1.1 Primary chromatin packaging

In order for all of the factors required for nuclear activity to fit within the nucleus, the nuclear chromosomes must be packaged. Chromosomes are composed of chromatin which is the combination of DNA, proteins, and RNA [45]. The DNA helix is packaged into structural units by being wound around protein complexes called nucleosomes [17, 46, 47]. The nucleosome is the primary DNA packaging unit of chromatin. Each nucleosome monomer organises 147bp of DNA in 1.7 left-handed superhelical turns, while an additional 20 bp acts as a linker between adjacent nucleosomes, although this linker distance is dynamic [17, 48, 49]. Nucleosome occupancy is dynamic in that their positions can be remodelled or they can be completely removed from the DNA strand, according to local nuclear processes. For example, upon the need to activate the transcription of the *GAL7* and *GAL10* loci to subsequently metabolise galactose, the nucleosomes occupying their promoter regions are removed by the Gal4 transcriptional activator, allowing access to the RNA polymerase enzyme [50].

1.1.1.2 Secondary chromatin packaging

Nucleosome bound DNA can be further compacted into secondary structures, such as the much debated 30 nanometre fiber [17, 18, 46, 47]. There are two main classes of the proposed 30 nm fiber, the one-start helix (interdigitated solenoid model) [18], and the two-start helix model ([19]; Figure 1.1). The solenoid model results in ~6 nucleosomes per turn of the helix (Figure 1.1A; [17, 51]). Therefore one complete turn of the solenoid is consistent with between 1000 and 1200 bp of DNA. Crucially, altering the DNA linker length does not adjust the chromatin fiber diameter as the linker DNA is extended into the fiber interior [51].

The two-start helix model consists of zig-zagging nucleosomes which form a helical structure (Figure 1.1B [19, 52]), whose width and length expands and contracts according to linker DNA length [17]. Recent *in vitro* and *in silico* experiments favour the solenoid model as researchers were unable to fit observed electron cryo-microscopy and modelling data to the two-start helical model [17]. However, this contrasts recent *in situ* data which suggest that the 30 nm chromatin fiber does not exist at all, at least for mitotic human chromosomes [53].

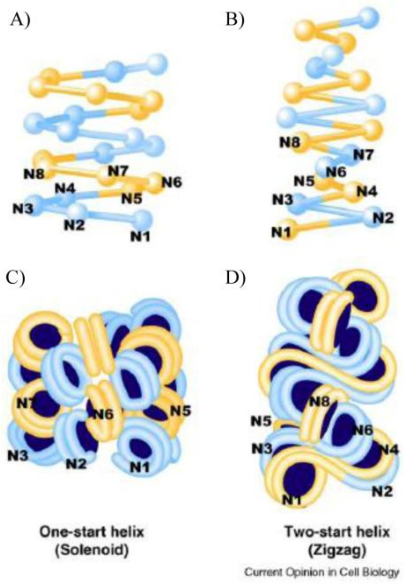

Figure 1.1: 30 nm chromatin fiber model.
There are two well-known structural models for 30-nm chromatin fibres [1]: A) one-start helix (solenoid) [18] and B) two-start helix (zigzag) [19]. Positions from the first (N1) to eighth (N8) nucleosome are labelled. C) In the one-start helix the 30-nm chromatin fibre is an interdigitated solenoid. Alternative helical gyres cores are coloured blue and orange. D) In the two-start model proposed by Woodcock *et al.* (1984), nucleosomes are essentially arranged in a zigzag manner such that alternate nucleosomes form interacting partners. That is, a nucleosome in the fibre binds to the second neighbour nucleosome. Alternate nucleosome pairs are coloured blue and orange. Image reproduced with permission from publisher with license number: 2666800179743.

1.1.1.3 Tertiary chromatin packaging

In mammalian, plant, and insect cells, chromatin fibers have been shown to form large loops [54-56]. These loops have been observed to protrude out from a proteinaceous chromosome scaffold mediated by matrix attachment regions [54-56]. Additionally, the loops vary in size, but can be hundreds of thousands of base pairs in length [56]. The formation of these loops translates to a higher-level of organisation and also an extra packaging dimension. It has been proposed that this chromosome scaffold is folded further into even higher organised structures, the geometry of which is yet to be determined [57].

1.1.1 Active and inactive regions of the genome are spatially separated

Chromatin forms domains within the nucleus according to its activity which is illustrated by electron microscopy studies highlighting electron dense and electron sparse regions [58]. At a basic level, DNA can be designated as hetero-chromatin (transcriptionally inactive) or eu-chromatin (transcriptionally active), depending upon the transcriptional profile of the region of chromatin in question. The distinction between these two forms of chromatin is often blurred [59]. For example, active genes have been shown to be positioned within regions of hetero-chromatin, just as silent

genes have been shown to be located within regions of eu-chromatin [59]. Hetero-chromatin and eu-chromatin tend to differ in their spatial localisation within the nucleus. In mammalian cells heterochromatin localises to the nuclear interior, and euchromatin to the exterior, closer to the nuclear pore complexes [60]. Theoretically, this conformation allows the mRNA transcripts easier access to the cytoplasm for the translation process. For reasons unknown, this conformation is swapped in lower eukaryotes such as *Saccharomyces cerevisiae* [61]

1.1.2 How are the chromosomes structured?

The initial functional unit which sub-divides the mammalian eukaryotic nucleus is the chromosome territory [22, 42, 62-67]. Each individual chromosome folds up upon itself and exists in its own nuclear sub-volume during the non-dividing phases of the cell cycle. Recent experiments, fusing fluorescence in situ hybridisation (FISH) with cryo-sectioning, revealed inter-mingling of the chromatin at the edge of the chromosome territories [21, 68-73]. Interestingly, each chromosome has a preferred position inside the nucleus, either in the interior, surrounded by the other chromosomes, or on the nuclear periphery [62]. These preferred positions change depending upon the cell type [62], indicating that chromosomal positioning is dynamic. Moreover, the organisation of genes within chromosome territories is not random, with gene-poor regions usually located within the territory interior, and gene-rich chromatin located on the periphery [66]. These results suggest that the spatial arrangement of the individual chromosome territories, that is, where they are in relation to one another, other nuclear factors, as well as the internal organisation of the individual genes, affects nuclear processes.

1.1.2.1 The *S. cerevisiae* chromosomes exist in a Rabl-like conformation

S. cerevisiae interphase chromosomes are not thought to form classical chromosome territories, but rather appear to exist in a Rabl-like conformation (Figure 1.2 [74-76]). The Rabl conformation was initially described for anaphase cells (reviewed in [77]), whereby the chromosome folds back upon itself at the centromere, leaving the telomeres juxtaposed at the opposite pole of the cell. While it was not evident at the time of this discovery, we now understand that this conformation arises due to microtubules attaching to the centromeric region and pulling the chromosome into the daughter cell [57, 77]. The Rabl-conformation of the chromosomes persists into interphase in *S. cerevisiae*. This conveys a biologically attractive structure which

theoretically requires little movement from the anaphase organisation of the chromosomes into interphase. This organisation is shared with *Drosophila melanogaster*, wheat, barley, oats, and *S. pombe* [74, 76, 78].

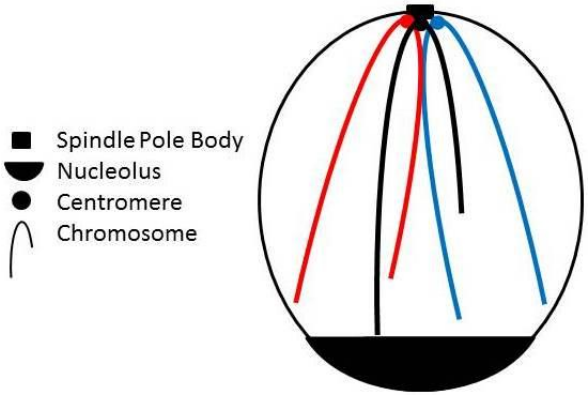

Figure 1.2: The Rabl-conformation
The conformation of chromosomes during cell division was described by Carl Rabl in 1885. The Rabl-conformation describes the arrangement of chromosomes whereby the centromeres are adjacent to the spindle pole body and the chromosome arms protrude out into the nucleus.

1.1.3 How is genome structure visualised?

Early light microscopy enabled researchers to describe large nuclear structures (*e.g.* nucleolus), but no methodology existed which could pinpoint the location of individual loci within the confines of the nucleus. A number of techniques have been developed over the years to elucidate the positions of loci within genome structure.

In situ hybridisation allowed nucleic acid sequences to be detected in cells without altering their morphology, a technique developed independently in 1969 by two competing labs [79, 80]. This technique was largely used for chromosome linkage analysis (*e.g.* banding of metaphase chromosomes [81]). The evolution of this technique, coupled with the introduction of fluorescent labels and higher resolution microscopy, resulted in the development of the fluorescence *in situ* hybridisation (FISH) method. FISH allows researchers to hybridise fluorescently labelled nucleic acid probes to specific regions of the genome in order to ascertain where in the nucleus they reside. This technique, at last, allowed the finer detail of genome structure to be deciphered. As different fluorescent probes became available more than one locus could be visualised at any given time, thus, regions of the genome suspected of co-localising could be confirmed. These days, FISH has evolved into a high-throughput technique, taking advantage of new three-dimensional imaging technology [82-84]. However, FISH and 3D-FISH are cumbersome techniques, requiring expensive equipment and

specialist training. Furthermore, the FISH technique is limited in the number of loci which can be simultaneously investigated.

1.1.3.1 The nucleolus is a prominent genomic structure

The nucleolus, which forms around the ribosomal DNA repeats, is the most prominent structure within the interphase nucleus. In higher eukaryotes the rDNA repeats are normally located on several different chromosomes, resulting in several nucleoli spread throughout the nuclear volume [85]. In *S. cerevisiae* a single nucleolus forms around the rDNA repeats on chromosome XII, at one end of the nucleus opposite the spindle pole body (SPB; [86]). The *S. cerevisiae* rDNA is composed of 150-200 9.1kb tandem repeats. Unusually, not all of the rDNA repeats are transcribed at once, with only a percentage being active at any point in the cell cycle [32]. What dictates which repeats are on or off is yet to be determined. Interestingly, the nucleolus, in cooperation with other repetitive elements has been hypothesised to coordinate the structure of the rest of the genome [87, 88].

1.1.3.2 How do the chromosomes fold?

We can consider a chromosome as a linear polymer and perform folding simulations *in silico*, which may provide insights as to how a chromosome might be structured within the nucleus. Polymers can fold into many different conformations depending on the folding conditions; however, the two which are currently most applicable to chromosome folding are the fractal globule and the equilibrium globule [6, 89-91]. The equilibrium globule is a ‘random walk’ model for polymer folding, which results in a highly tangled structure. This is the conformation a polymer will naturally assume if the polymer is not forced to participate in contacts or interactions with itself. Consequently, comparisons between true chromosome structure and simulated equilibrium globule polymers are a natural extension of chromosomal studies (Figure 1.3A; [6]).

A fractal globule conformation confers a structure which is not tangled like the equilibrium globule making it a more biologically attractive model for chromosome folding and structure. The polymer appears to collapse upon itself, forming a fractal arrangement (Figure 1.3B; [6]). The number of contacts a folded polymer has with itself and the distance between these contacts tends to follow power law dependency. Power law dependence arises when the frequency of an event (individual contacts within the polymer) varies by the power of one of its attributes (linear distance between contact

points). The fractal and equilibrium globule conformations were determined by plotting log-log of polymer folding simulation data and calculating the slope exponent [92]. A fractal globule has a slope exponent of -1, while an equilibrium globule has a slope exponent value of -1.5.

The fractal globule was hypothesised to be a possible model for DNA and chromosome folding as early as 1993 [89]. This hypothesis could only be tested recently when experimental data became available for human chromosomes [90]. Indeed, the human chromosomes follow the fractal globule arrangement well for DNA looping lengths (the distance between contacts within the chromosome) between 500 kb and 7 Mb [90]. This model fits well with other experimental data indicating the presence of chromosome territories in human cells [62, 93-95].

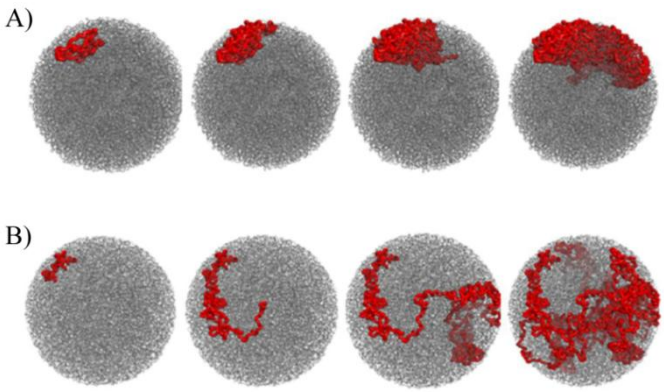

Figure 1.3: Polymer folding into fractal and equilibrium globule models change conformation the longer the polymer.
A) Simulations of the fractal globule with increasing monomer lengths and power law exponent of -1 result in a 'territory' appearance. B) Simulations of the equilibrium globule with increasing monomer lengths and power law exponent of -1.5 result in a perceived lack of territories. This figure has been modified, with permission (license number: 2617790573428), from Mirny, L. 2011 [6].

1.2 Nuclear processes are spatially separated

Nuclear processes occur in distinct, spatially segregated, regions of the nucleus [4]. An example of this is the fact that transcription and replication do not simultaneously occur at the same point in nuclear space [4]. Simultaneous fluorescent labeling of DNA and RNA synthesis sites in mouse 3T3 cells, allows the resultant images to be merged revealing that the processes occur in spatially distinct regions of the nucleus (Figure 1.4). The segregation of transcription and replication is perhaps not a surprise, considering the machinery required for each is cumbersome and therefore cannot co-exist at the exact same point in space.

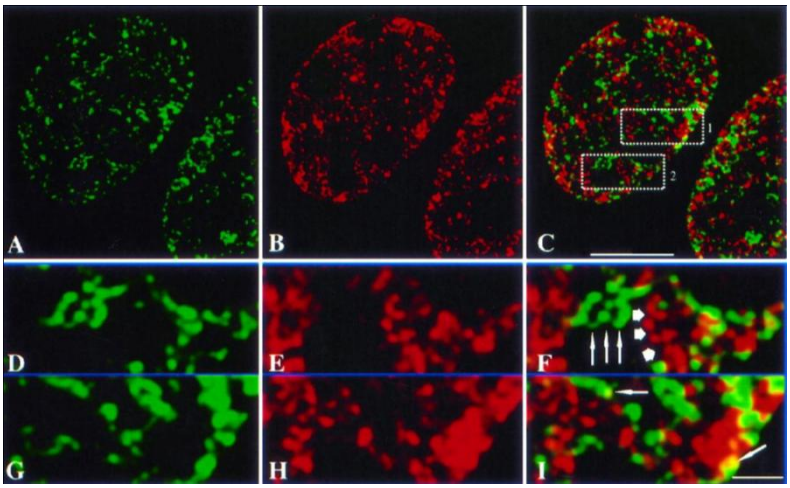

Figure 1.4: Transcription and replication occur in spatially separate regions of the nucleus.
DNA replication (A, D, G) and transcription (B, E, H) sites are visualized in green and red, respectively, in mouse 3T3 cells [4]. C, F, I) Transcription and replication site images merged. F) Thick and thin arrows indicate clusters of transcription and replication sites, respectively. I) Examples of apparent overlap between transcription and replication sites are indicated with arrows. Image reproduced with permission from publisher; license number: 2625111142875.

1.2.1 Transcription occurs in factories

Transcription requires one of three RNA polymerases (RNAPI, II, or III). Transcription of RNAPII genes has been found to occur at RNAPII-rich nuclear foci known as transcription factories [30, 96, 97]. Tagging of newly synthesised mRNA transcripts *de novo* reveals the transcripts originate from these discrete nuclear foci (Figure 1.4B, [4]). Active genes and regulatory elements loop out of their chromosome territories and co-localise at these discrete focal points [30]. Such an arrangement is believed to raise the local concentration of RNA polymerase at these transcription factory sites, and thus enhance RNA production [96]. Interestingly, recent research highlights the importance of transcription factories in mouse erythroid cells [98]. The authors go as far as suggesting that co-regulated genes are localised to specialised transcription factories dedicated to the transcription of genes which are regulated by common transcription factors (*e.g. Klf1*; [98]). It is possible that specific regions are actively 'recruited' to particular transcription factories or that the genome as a whole is arranged or tethered in such a way that co-localising sequences are consistently in close proximity, in that particular cell type.

Osbourne *et al* (2007) demonstrated that, upon gene activation, rapid gene positioning to preassembled transcription factories occurs [97]. Mitchell and Fraser (2008) found that genes repositioned to these transcription factory sites stay associated even when transcription elongation is inhibited, but disassociate if transcription initiation is

inhibited [99]. It is currently unclear what anchors the active DNA at these transcription factory sites, perhaps it is RNA polymerase, or some other structural element. Furthermore, it is also unclear whether the recruitment of genes to these factory sites is an active or passive process.

One model for the localisation of genes and regulatory elements at transcription factory sites is the simple diffusion theory [100]. Cook *et al*. (2002) suggest that aggregation of the large polymerase complexes into transcription factories is a natural phenomenon and results from molecular crowding [63]. Accordingly, if active chromatin has transcription factors attached to it, thus, increasing its size, then by simple Brownian motion it would be natural for these complexes to aggregate with the polymerases.

1.2.2 Gene positioning

Many different factors come into play which affect when and how much a certain gene is able to be transcribed, not least of which is position effect. It has been previously shown that if a gene is transformed into a region of heterochromatin it will not be transcribed nearly as much as if it had been transformed into a euchromatic region [101]. Moreover, the spatial position of a particular DNA locus (whether it be gene, regulatory sequence, gene desert, gene island) within the nucleus appears to correlate with the transcriptional profile for a particular phenotype [102, 103]. The conversion of a gene from inactive to active, in accordance with an environmental cue (*e.g.* yeast GAL locus upon galactose exposure), results in a relocation of the gene within the nuclear volume. Once this initial activation has taken place, reactivation of the locus is quicker, indicating a form of transcriptional memory [40, 43, 103].

Recent experiments have implicated nuclear Sc35 speckle domains in spatially organising active genes [104, 105]. These nuclear speckles are said to mediate long-range physical contacts between different regions of the genome. Conversely, cells which lack these nuclear speckles are still capable of spatially organising active genes, indicating this system is not universally applicable [98].

The spatial architecture of the genome changes upon gene activation. Berger *et al.* (2008) developed a computational imaging technique to screen thousands of *S. cerevisiae* cells for spatial nuclear organisation [64]. *GAL1* and *GAL2* are on different chromosomes, but are co-regulated, and both activated upon the need to metabolise the galactose carbon source. By fluorescently tagging the region surrounding these two loci,

probabilistic maps were generated for their nuclear position by triangulation with the spindle pole body, nucleolus, and the integrated fluorescent tag in thousands of *S. cerevisiae* cells. Upon exposure of the yeast cells to galactose *GAL1* moves from the nuclear interior to the exterior upon activation. The position of *GAL2* does not move as much as *GAL1*, possibly due to it being located on chromosome XII (the chromosome harbouring the rDNA repeats), but it does shift slightly towards the nuclear periphery. These genes appeared to exist in what was described as gene territories, indicating some form of overall genome structure or local chromatin fixture process in *S. cerevisiae*. What was interesting to note was that the nuclear position of the chromosomal region surrounding the *GAL2* locus was not different to wild type when the *GAL2* gene was deleted. These observations suggest that the *GAL2* coding sequence was not necessarily driving its position within the nucleus, leaving open the possibility that its position is governed by the structure of the local chromatin or the overall global genome architecture.

1.3 What are inter- and intra- chromosomal interactions?

Physical contacts between different regions of the genome are called inter- or intra-chromosomal interactions, depending on whether the interaction occurs between or within chromosomes, respectively. These contacts may form as part of general genome compaction (*e.g.* hetero-chromatin and chromosome territories), as genome structure anchor points, or in concert with nuclear processes (*e.g.* transcription or replication).

In mammalian cells, intra-chromosomal interactions might form as a result of the folding of a chromosome into a territory. Alternatively, the establishment of these interactions may drive the chromosome into these territory structures. Inter-chromosomal interactions are physical contacts which form between different chromosomes.

Chromatin looping is well established as being part of the transcriptional regulation of genes [31, 106]. DNA sequences void of nucleosomes are susceptible to DNase I, and are termed DNase I hypersensitive sites (HSs). HSs generally indicate sites which contribute to the regulation of genes [107, 108]. Such sites include, but are not limited to, enhancers, silencers, promoters, imprinting control regions, and locus control regions [109]. Appropriate gene expression is not achieved without these regulatory elements [110]. The mechanism by which these elements regulate their genes is interesting

because large genomic distances can separate the respective factors. DNA can form large loops which can bring regulatory elements in close proximity with the promoter region of the gene which they control, essentially forming an intra-chromosomal interaction. The regulatory elements are not required to be on the same chromosome as the gene which they control [111], with DNA inter-mingling from different chromosomes making it possible for regulation to occur via inter-chromosomal interactions. Thus, inter- and intra- chromosomal interactions form as part of a higher order regulation of genes. It is currently unclear whether regulatory related inter- and intra- chromosomal interactions are set up in an active or passive process or whether they exist as a result of transcription or to promote it. However, inter-chromosomal interactions have been demonstrated to break down in the absence of transcription, their occurrence a function of co-localisation at these transcription factory sites [90, 112].

The lower eukaryote, *S. cerevisiae*, is not thought to harbour classical long-range enhancer sequences. It does, however, have promoter proximal elements called upstream activator sequences (UAS) [113, 114]. Despite the lack of long-range enhancer sequences, inter- and intra- chromosomal interactions still form in *S. cerevisiae* [70, 115]. Inter-chromosomal interactions were detected between homologous chromosomes in pre-mitotic and mitotic cells, just as would be expected due to the sister chromatids being held together during these cell cycle phases [70]. In the same cell cycle phases the centromeres are clustered close to the spindle pole body, and as a result, inter-chromosomal interactions were detected between centromeres on different chromosomes [116]. These same inter-chromosomal interactions, detected between sister chromatids and centromeres, were not detectable during the non-dividing phases of the cell cycle. Inter- and intra- chromosomal interactions were also shown to occur between the *S. cerevisiae* telomeres which are known to cluster in yeast [76].

1.3.1 What are the functions of inter- and intra- chromosomal interactions?

Spilianakis *et al*. (2004) describe developmental inter- and intra- chromosomal interactions in naïve $CD4^+$ T-cells [23]. $CD4^+$ T-cells differentiate into either a T-helper-cell 1 (T_H1) or T-helper-cell 2 (T_H2), and the mutually exclusive expression of two genes dictates the $CD4^+$ T cells fate [23]. T_H1 development requires the expression of the cytokine antigen, interferon-γ (IFN- γ) gene (mouse chromosome 10) while T_H2

development requires expression of cytokines, interleukin-4, 5, and 13 (IL-4, IL-5, IL-13; mouse chromosome 11) [23]. It has been found that the promoter regions of both the IFN-γ and IL-4, 5 and 13 genes participate in inter- and intra- chromosomal interactions, respectively, with the regulatory T_H2 locus control region on chromosome 11 [23]. Crucially, these interactions required for T-helper-cell fates do not occur in non-$CD4^+$ cells [23].

Spilianakis *et al*. propose a model whereby the inter- and intra- chromosomal interactions are formed to coordinate the expression of the linked genes and consequently chose the cell fate [72]. This clustering of developmentally regulated genes with the LCR is suggested to be dynamic, and once a cell fate is chosen, the cluster dissipates.

Genetic imprinting refers to a situation when genes are expressed from only one of the parental alleles (reviewed in [117]). This difference in allelic expression is marked epigenetically by DNA methylation and histone modifications (reviewed in [118]). Zhao *et al*. (2006) searched for interactions with the *H19* imprinting control region (ICR) on mouse chromosome 7 [21]. They subsequently mutated mice at this ICR site such that the CCCTC-binding factor (CTCF) can no longer bind and crossed it with wild type mice. A method design to capture interactions between different regions of the chromosomes called Circular Chromosome Conformation Capture (4C) was performed on the ICR region when the wild type region was maternally or paternally inherited. They found that the maternally inherited wild type ICR region formed the most inter-chromosomal interactions. The authors conclude that the epigenetic status of the *H19* ICR determines patterns of the majority of the inter- and intra- chromosomal interactions. It was also observed that more than two regions interacted with each other at any given time, suggesting a clustering of related or similarly regulated genes [21], in agreement with other observations [119]. Furthermore, CTCF was shown to be involved in the formation or maintenance of inter- and intra- chromosomal interactions [120]. Earlier observations of the *H19* imprinting region found a correlation between the direct interactions of differentially methylated regions and epigenetic regulation of these sites *in trans* [121]. These observations indicate that inter- or intra- chromosomal interactions, which may be structural or regulatory in nature, are functional and important for developmental regulation and general gene transcription in eukaryotic cells.

1.4 What factors shape the eukaryotic genome?

During cell division, when the chromosomes are condensed, proteins such as condensin and cohesin play a major role in forming the characteristic X shape [122]. In fact cohesin has been linked to the formation of large chromosome loops which persist through inter-phase. The function of these loops is hypothesised to separate functional and non-functional chromatin units [123]. In higher eukaryotes, the insulating factor CTCF binds at sites which correlate with cohesin binding and help maintain the boundary between hetero- and eu-chromatin [124]. CTCF and cohesin have been described as the 'master genome weaver' in mammalian cells as it has been shown to mediate cell type-specific inter- and intra- chromosomal interactions including Igf2/H19, β-globin, and IFN-γ [120, 124-128]. CTCF is not present in the lower eukaryotes *S. cerevisiae* and *S. pombe* (or, incidentally in the model plant *Arabidopsis thaliana*), however, factors which bind tRNAs and transcription factor III (TFIIIC) are hypothesised to fulfil these CTCF genome weaving and insulating roles [129].

A number of research groups have uncovered the involvement of transcription factors and thus, transcription, in mediating functional three-dimensional conformations of the genome [130-132]. One example is the mammalian GATA transcription factors shown to assemble DNA loops between enhancer elements and promoter regions of the proto-oncogene *Kit* locus in a, development, transcription, and DNA methylation dependent manner [130, 133]. These types of finds have resulted in the hypothesis that transcription shapes the three-dimensional structure of the genome.

1.5 How are inter- and intra- chromosomal interactions detected?

Inter- and intra- chromosomal interactions can be captured using proximity-based ligation methodologies [70, 134] that incorporate high resolution (*i.e.* ~2 Å) cross-linking of interacting DNA strands [135]. Subsequent restriction digestion, ligation, and detection, allows identification of DNA sequences that interact at a frequency greater than background [5, 134, 136]. Cullen, K.E., *et al* (1993) were the first to develop a method which took advantage of crosslinking and PCR, to detect two regions of the genome in close proximity, which they called the 'nuclear ligation assay'. However, it was not until 2002 before chromosome conformation capture (3C; Figure 1.6A) reintroduced and reinvigorated the proximity based ligation method [70].

Interactions which can be detected can be broken into four distinct groups; adjacent (Figure 1.5A), non-adjacent intra-chromosomal (Figure 1.5B), inter-chromosomal (Figure 1.5C), and self-ligation (Figure 1.5D). Each one of these different types of interactions tells us something about the structure of the genome. For example, an unusually high number of non-adjacent intra-chromosomal interactions might suggest the presence of chromosome territories, as chromosomes fold up upon themselves and thereby interact with themselves at a higher frequency.

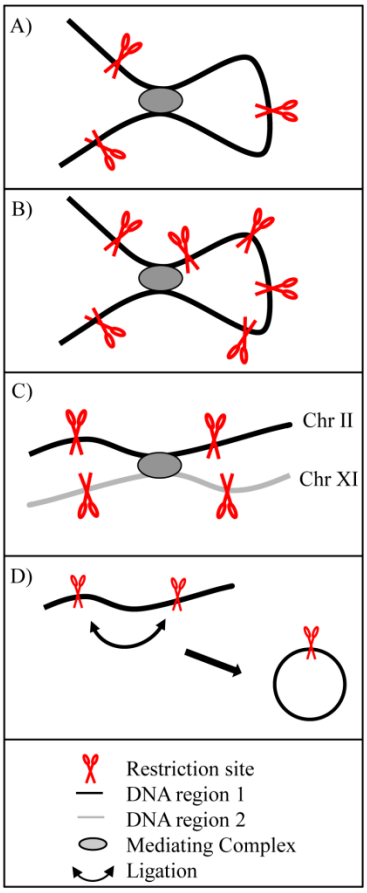

Figure 1.5: Proximity based ligation methods can detect interactions between different regions of the genome.
The chromosomal context of the interactions is depicted. A) Adjacent interactions form when two restriction fragments next to each other in the linear genome are held together. B) Non-adjacent intra-chromosomal interactions are captured between restriction fragments which are not directly next to each other in the linear genome, but between fragments which are looped over to interact. C) Inter-chromosomal interactions form between different chromosomes. D) Self-ligation events are captured either as a result of a restriction fragment circularising and ligating within itself, or between identical repeats. The interaction captured by restriction digestion and ligation, essentially joining the two regions of the genome together which were in close spatial proximity.

1.5.1 Capturing Chromosome Conformation

3C was originally performed on *S. cerevisiae* cells [70] and uses formaldehyde to stabilise nuclear structure by cross-linking primary amines in proteins and nucleic acids.

Formaldehyde is reactive towards primary amine groups, such as are present in the side chains of lysine or arginine. The reaction of formaldehyde with a primary amine results in a Schiffs base intermediate which can subsequently react with other primary amines, thus forming a covalent linkage [135]. Excess formaldehyde is sequestered with glycine, a primary amine, before the chromatin is digested with a restriction enzyme (Figure 1.6A). The sample is then diluted in ligation buffer to minimise the local concentration of digestion fragments, and in turn promote the intra-molecular ligation of cross-linked DNA fragments with T4 DNA ligase. The ligation step covalently joins any interacting DNA fragments before the crosslinks are removed and the protein digested in the presence of high salt, heat and Proteinase K. This allows the now hybrid DNA molecule to be isolated for analysis. 3C primers are designed to amplify across the ligation junction of two fragments of DNA that were originally suspected of interacting with each other (Figure 1.6B; [70]). Because the positions of the restriction enzyme recognition sites are known, the expected PCR product length is also known. Therefore, interactions can be identified by PCR, agarose gel electrophoresis, and subsequently confirmed with DNA sequencing. The presence of the correct band at the correct size indicates that the two sequences were interacting with each other [137].

1.5.2 3C spin-off methodologies

A number of related techniques have subsequently been developed to capture inter- and intra- chromosomal interactions.

1.5.2.1 Circular chromosome conformation capture (4C)

Unlike 3C, Circular Chromosome Conformation Capture (4C) requires no *a priori* knowledge of interacting DNA sequences, apart from the selection of a region of interest to test for interactions; a 'bait' fragment (Figure 1.6C; [138]). Simonis *et al.* (2006) generated a 4C library using mouse primary liver and brain cells [23, 138]. They designed primers for the 4C assay that targeted the β-Globin LCR as the bait region, as previous observations indicated this region could have functional significance for development [139, 140]. Rather than PCR across one ligation junction as in 3C's, primers were designed to perform inverse PCR out of the known restriction fragment around the now circularised DNA (Figure 1.6C). This approach amplifies the unknown fragments of DNA which were interacting with the known fragment [23]. The resulting 4C library was hybridised to a unique microarray which contained probes designed within 100bp of each ligation junction [138]. Using this technique, hundreds of

interactions with the β-Globin LCR were detected, many of which were with inter-chromosomal partners [138]. Simonis *et al's.* results suggest that LCRs may not be restricted to the regulation of just the genes they are linked to, but may regulate genes on other chromosomes. The 4C technique has recently been coupled with ChIP against phosphorylated RNA polymerase II Ser5 residue of the C-terminal domain (CTD), and enrichment of interacting material with biotin labelled 4C primers, thus allowing the isolation of interactions resulting from active transcription and a better signal to noise ratio [98].

1.5.2.2 Chromosome Conformation Capture Carbon Copy (5C)

5C uses a conventional 3C library which is converted to a 5C library by primers annealing either side of the 3C ligation junction, these primers are ligated together, essentially creating a carbon copy of the junction (Figure 1.6D, [20]). These are amplified with universal primers, which were present as linkers on the original primers. The carbon copies can then be analysed by high throughput DNA sequencing or dedicated microarrays.

The frequency with which two restriction fragments interact directly correlates to the frequency of the interaction within the cell population [70]. Dostie *et al*. (2006) suggest this method could be scaled up to form a global map of all interactions in a nucleus. However, the number of primers required to do this effectively restricts this method to relatively small regions of the genome, such as the β-Globin locus, and a conserved gene desert, which were used as proof-of-principle for this high-throughput technique [20].

Other variations of the 3C method have been developed which incorporate chromatin immuno-precipitation to detect interactions mediated with specific proteins, such as CTCF and other transcription factors [111, 141, 142].

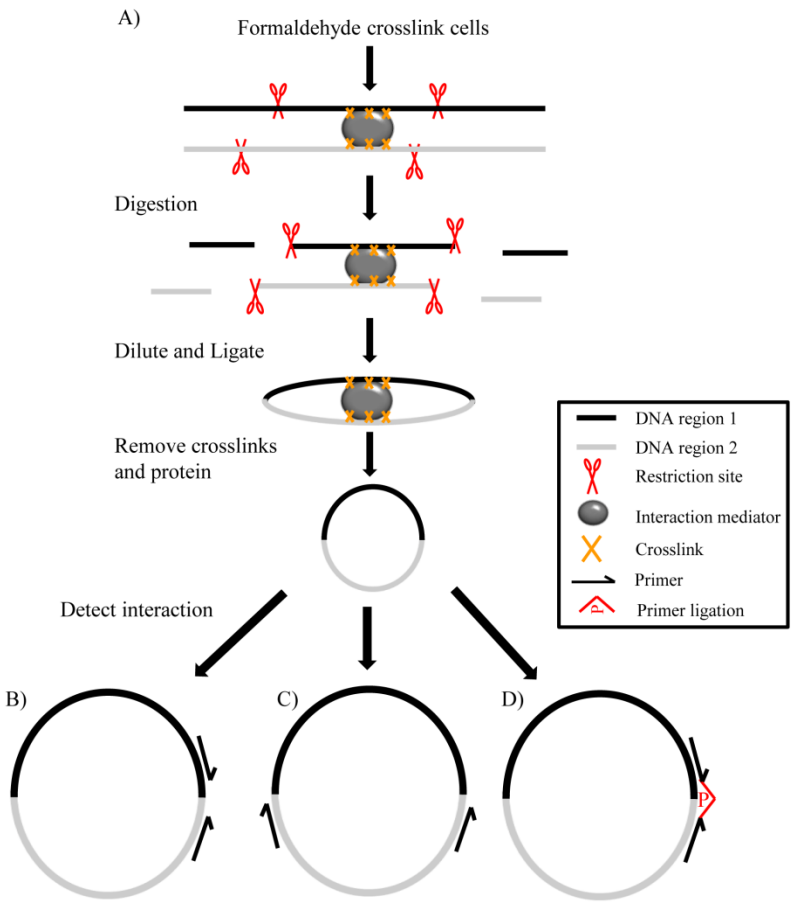

Figure 1.6: Proximity based ligation methods to detect inter- and intra- chromosomal interactions.
A) The chromatin preparation for proximity based ligation methods. The chromatin is crosslinked with formaldehyde before being digested with the restriction enzyme of choice. The digested DNA is diluted prior to ligation in order to promote the intra-molecular ligation of crosslinked DNA molecules. Crosslinks and mediating proteins are removed with high salt and temperature in the presence of Proteinase K, before the ligation library is purified. B) The interaction is detected by conventional 3Cs by amplifying across the ligation junction with PCR. C) 4Cs: amplification of interactions with a chosen bait fragment is achieved by inverse PCR from the bait fragment and detected by hybridising to microarrays or by next generation sequencing. D) 5Cs: primers are designed to bind and ligate across the ligation junction. Ligation products are amplified by universal primers attached to the 5C primers before detection by hybridisation to microarrays or high throughput sequencing.

1.5.3 Problems with proximity based ligation detection methodologies

The 3C technique requires *a priori* knowledge of two regions of the genome that might be coming into contact in order to be able to design primers to amplify across the ligation junction, and thus, detect the interaction. 4Cs requires that a bait sequence is chosen, thus one is restricted to detecting interactions with a certain region of the genome. 4C libraries can be analysed by hybridising to microarrays, or by using more recent technology, high throughput DNA sequencing. 5C libraries can also be analysed by custom microarrays or next generation sequencing. These proximity based ligation techniques incorporate biases due to amplification steps. None of these techniques, with

the exception of 5Cs, are truly high-throughput. At the beginning of this research there was no method to detect inter- and intra- chromosomal interactions on a global scale.

1.5.4 Genome Conformation Capture

In this thesis I will describe the development of Genome Conformation Capture (GCC), the first easily applicable method, which does not rely on *a priori* knowledge, to detect inter- and intra- chromosomal interactions [5]. Following publication of the GCC methodology [5], and throughout the duration of my study, global methods which are applicable to small genomes such as *S. cerevisiae* and *S. pombe* [74, 143], and larger genomes such as Human [90, 141, 144], were independently developed. For example, Hi-C [90], another 3C derivative, used to globally map interactions in human cells, incorporates a restriction enzyme overhang end filling technique using biotin labelled nucleotides before a blunt-end ligation is performed. Thus, Hi-C allows DNA molecules resulting from interactions to be enriched with a simple biotin pull-down assay. Enrichment steps, such as that incorporated into Hi-C, enable inter- and intra-chromosomal interactions to be determined for large genomes with relatively small numbers of high-throughput sequence tags.

1.6 Saccharomyces cerevisiae

S. cerevisiae is a facultative anaerobe and a model eukaryotic organism. Many of the biochemical pathways which occur in *S. cerevisiae* are common to higher eukaryotes, and metazoans (*e.g.* citric acid cycle, oxidative phosphorylation, *etc.* [145]). *S. cerevisiae* can grow successfully in a diploid or haploid form. In this thesis, GCC was performed on a haploid *S. cerevisiae* strain (BY4741). Therefore, interactions were able to be uniquely mapped onto single chromosomes, rather than there being a minimum of two of each locus in the nucleus, as would be the case for a diploid organism.

1.6.1 The *S. cerevisiae* genome

The *S. cerevisiae* nuclear genome consists of 16 chromosomes varying in size from 230,218 bp (chromosome I) to ~2,800,000 bp (chromosome XII, including a full complement of ~200 ribosomal DNA repeats [rDNA]). In 1996 the ~12Mb genome was sequenced, revealing ~6000 open reading frames, 40 genes for snoRNA molecules, 275 transfer RNA genes and 52 complete Ty elements [146].

1.6.1.1 The yeast nucleus is structured

The *S. cerevisiae* nucleus is structured and has non-membrane bound sub-compartments. The *S. cerevisiae* nucleolus forms at one end of the nucleus opposite the spindle pole body [86], is the site of ribosome biogenesis, and the most prominent nuclear structure. Therefore, the region of chromosome XII, which harbours the rDNA repeats, has a fixed position within the nucleus. There is a small amount of evidence which indicates that the remainder of the chromosomes form territories in yeast [64], but as discussed earlier, it is more likely the chromosomes are in a Rabl conformation [76]. A probabalistic structure of chromosome III was mapped by the detection of intra-chromosomal interactions by 3C, which demonstrated a population average conformation for this chromosome consistent with a Rabl-like conformation [70]. This is supported by the observation that the centromeres attach to the spindle pole body [64], and the chromosome arms are free to move throughout the nuclear space, although the telomeres have been shown cluster with each other [147], and be associated with the nuclear envelope. The tRNA genes have also been observed to cluster and co-localise with the nucleolus [88, 148].

All of these results demonstrate that the nucleus and the nuclear genome is structured; however it is not only the 16 nuclear chromosomes which make up the entire *S. cerevisiae* genome complement.

1.6.2 The Mitochondria

The mitochondrial organelle is composed of inner- and outer- membranes, with the inner-mitochondrial membrane being described as a cristae structure [57]. Importantly, the yeast mitochondrion, as in other eukaryotes, is the site of oxidative phosphorylation, and therefore, confers the ability to undergo respiration. It is the inner membrane which harbours the electron transport chain and ATP synthatase machinery [145].

1.6.2.1 Mitochondria have their own genome

Mitochondria have their own DNA genome and in *S. cerevisiae* it is 85,779bp in length and predicted to be of α-proteobacterial origin [149]. Current estimates put the copy number of the yeast mitochondrial genome at ~50 per cell [150], although this number can vary according to the cell's metabolic regime [41, 151].

The *S. cerevisiae* mitochondrial genome harbours 17 verified open reading frames (ORFs) which encode components of the electron transport chain machinery as well as

24 tRNAs. Interestingly, the mitochondrial *COX1* and *COB* genes contain introns, known as group I and group II mobile elements, which encode reverse transcriptase machinery [7]. This machinery provides these mobile elements with a 'homing' ability, that is, they are able to replicate and insert themselves in homologous genes devoid of the introns [7, 152].

1.6.2.2 Nucleic acid is transferred from the mitochondria to the nucleus

The mitochondrial genome is widely recognized as having drastically reduced in size over the course of the evolution of the mitochondrial – host symbiosis, to the point that ~98% of the genes required for mitochondrial function are presently encoded within the nuclear chromosomes [153]. Such transfer is central to endosymbiotic theory, but how this transfer occurs is still an open question. Despite the mechanism of transfer being unknown [154], studies have clearly demonstrated that mitochondrial DNA is transferred to the yeast nucleus at rates of up to 2×10^{-5} per cell per generation [155]. It is currently unclear whether the transfer of mtDNA is an active process or not. Despite this, the mitochondrial organelle has been linked physically with the nucleus through the endoplasmic reticulum, possibly providing a route for this observed DNA transfer [156, 157].

Organelle DNA has also been shown to be transferred to the nucleus in a range of other organisms, including rice [158], tobacco [159], and humans [160, 161]. The transfer of genetic material from the mitochondrial organelle to the nucleus in yeast is hypothesised to be as small unstable plasmids [162]. On the contrary, a complete chloroplast genome has been transferred and integrated into chromosome X of the rice genome indicating that organelle DNA transfer need not be restricted to relatively short lengths of nucleic acid [158].

Mitochondrial DNA escape to the nucleus is partially dependent upon carbon source as a five-fold reduction in the transfer rate was observed in galactose grown, as opposed to, glucose growing *S. cerevisiae* cells [163, 164]. The nucleic acid being transferred is not thought to be integrated into the nuclear genome, possibly indicating another role within the nucleus [162, 165, 166]. Despite this, mtDNA has been shown to be incorporated into the nuclear genome in conjunction with the repair of double stranded breaks [154].

Interestingly, genes involved in the yeast mitochondrial escape (YME) pathway, which increase the rate of transfer, have been characterised in yeast [163, 164].

Nuclear rDNA replication is affected by the concentration of the mtDNA [167]. This result possibly links the transfer of mtDNA to a function which may manifest itself as a form of signalling. Indeed, efficient signalling mechanisms do exist between the mitochondrial and nuclear organelles, in the form of the retrograde and anterograde response [168]. These signalling mechanisms co-ordinate and control the expression of the nuclear and mitochondrial genome encoded genes required to maintain and control mitochondrial function according to cellular metabolism [168, 169].

1.6.2.3 The mitochondrial genome forms nucleoids consistent with its prokaryotic origin

As one might expect, due to their prokaryote origin, the mitochondrial genomes exist as nucleoids [41, 170, 171]. These nucleoids are composed of clusters of between 5 and 7 mitochondrial genomes, along with a number of protein components. The mitochondrial nucleoids are said to undergo rearrangements according to the carbon source being utilised [41]. Interestingly, the mitochondrial organelles undergo regular fission and fusion events which are hypothesised to expose the mitochondrial nucleoids to a greater pool of factors required for their function and maintenance [171].

1.6.3 *S. cerevisiae* metabolism

During growth on glucose (fermentation) the electron transport chain is non-essential as ATP is produced by glycolysis; therefore, the mitochondrial genome is dispensable. Mutations to the nuclear or mitochondrial genomes, which knock out the function of the mitochondria, are possible, as is curing the cells of the mitochondrial genome. Conversely, ATP is produced by oxidative phosphorylation during growth on non-fermentable carbon sources [57]. Consequently, the mitochondria and the mitochondrial genome are essential for growth on non-fermentable carbon sources such as glycerol.

S. cerevisiae is capable of fermenting, respiring, or combining these processes and performing respiro-fermentation in order to generate ATP from a carbon source [172, 173]. Respiro-fermentation is carried out when yeast cells have access to oxygen and the fermentable carbon source is at a concentration exceeding 0.8 mM [172, 173]. Thus, growth on glucose exceeding 0.8 mM results in ~49% of the cellular ATP being produced by oxidative phosphorylation (respiration), while the remaining ~51% is

produced by fermentation [173]. The percentage of ATP produced by oxidative phosphorylation is higher during galactose growth (83%; [173]) despite there being little difference in the downstream metabolism of either of these carbon sources. Growth on galactose requires the expression of only 6 extra nuclear genes in order to pre-process the galactose molecule into glucose-1-phosphate, allowing it to subsequently enter the traditional glucose degradation pathway for fermentation.

1.6.4 The selfish 2-micron plasmid

The *S. cerevisiae* nucleus contains a high copy number (~60 per cell) 6318 bp selfish genetic element of unknown origin; the 2-micron plasmid. Consisting of four open reading frames (ORFs), a stability locus (STB) and two 599 bp inverted repeats (IR), the 2-micron plasmid confers no detectable advantage to the host organism, nor is it a burden on cellular resources [174]. FISH studies reveal the plasmids cluster into several discrete sub-nuclear foci; a result of two plasmid encoded STB locus binding proteins (Rep1 and Rep2) [175]. The STB locus, in conjunction with the Rep DNA binding proteins, are hypothesised to facilitate plasmid segregation during cell division [175, 176].

Yeast 2-micron plasmid clusters have been shown to co-localise with the cohesin complex, indicating an attractive model for segregation during cell division [175]. Cohesin can exist in two forms, either bound to the nuclear chromosomes, or not, and it is unclear which of the two forms the 2-micron plasmid clusters bind [175]. However, plasmid partitioning during cell division is a quandary, with a preference for the bulk of the plasmids to remain in the mother cell [177].

An efficient recombination based amplification system makes certain the ~60 copies are maintained in each cell [174, 178]. Replication normally occurs in theta form from a replication origin (autonomously replicating sequence [ARS]); however, upon the need to increase the 2-micron plasmid copy number, two inverted repeats (IR1 and IR2) undergo FLP recombinase-mediated recombination. This essentially converts theta replication into a ‘chasing polymerases’ conformation commonly known as rolling circle amplification [178]. Therefore, for at least part of the cell cycle, the 2-micron has a definitive structure, as the two IR regions must co-localise in order for this recombination to occur.

1.7 The global determination of genome structure

Our ability to ascertain an organism's linear sequence only goes some way towards understanding how a genome functions. Overlaying onto this linear sequence the positions of ORFs, regulatory elements, transcription factor binding sites, epigenetic marks such as the histone code, DNA methylation, and ultimately genome structure may allow us to determine what factors are important to realise certain phenotypic profiles.

As I have discussed, inter- and intra- chromosomal interactions are important aspects of genome structure, as well as nuclear processes such as transcription, replication and DNA repair. Detecting these interactions on a global scale would allow the identification and classification of another of the multiple layers of regulation of the genome. Individual interactions will provide research leads that elucidate the functional and structural significance of DNA-DNA interactions. Determining how these interactions change under different conditions may provide insights to the growing evidence suggesting three-dimensional genome structure is part of the epigenetic program of the cell.

In this thesis I will present the development of the Genome Conformation Capture (GCC) method which detects and identifies inter- and intra- chromosomal interactions on a global scale. I will subsequently demonstrate the global DNA-DNA interactome network for the yeast, *S. cerevisiae*. As expected, the nuclear genome, including the 2-micron plasmid, is highly connected with inter- and intra- chromosomal interactions. Excitingly, mtDNA was also found to be physically interacting with the nuclear genome.

Following validation of the GCC technique, I performed it on *S. cerevisiae* cells grown on glucose, galactose, and glycerol lactate to ascertain how the interaction network responds to different metabolic conditions. As discussed, the mitochondria are utilised to different extents in the three growth conditions, and this is evident in the numbers and types of interactions that form between the mitochondrial and nuclear genomes. These interactions are investigated further, ultimately demonstrating that mitochondrial cDNAs are transferred to the nucleus as part of a signalling mechanism to regulate genes of mitochondrial origin. Furthermore, inter- and intra- chromosomal interactions within the nuclear genome change depending upon the metabolic condition.

I demonstrate that GCC is a valid method for the detection of inter- and intra-chromosomal interactions, and how these interaction networks respond to different metabolic conditions. Finally, the content of this thesis represents a significant step forward in the field of genome structure.

Chapter 2 Material and Methods

2.1 Materials

2.1.1 Chemicals and Reagents

All chemicals are of analytical grade, unless specified.

Reagents were purchased from the following suppliers:

Invitrogen Corporation
Carlsbad,
USA

Sigma-Aldrich
St. Louis,
USA

Formedium
Norfolk,
United Kingdom

Applichem
Darmstadt,
Germany

Oxoid
Cambridge,
United Kingdom

Univar
Fontenay-sous-Bois,
France

Bacto Laboratories Pty Ltd
NSW,
Australia

J.T. Baker
New Jersey,
USA

BDH AnalaR; VWR International Ltd
Poole,
England

Ajax Finechem Pty Ltd
Auckland,
New Zealand

BD; Becton, Dickinson and Company
Sparks, MD
USA

Labserv
Clayton, VIC,
Australia

2.1.2 Solutions

All solutions were prepared using distilled, deionised Milli-Q™ (Millipore) water (ddH20) and sterilised by autoclaving (15 psi, 121°C for 20 min), or filtering through a 0.2µm filter, prior to use.

Solutions for general molecular biology procedures

0.5 M EDTA	70% (final v/v) Ethanol (EtOH)
10 mg/ml Ethidium Bromide (EtBr)	1 M Magnesium Chloride (MgCl2)
Polyethylene Glycol 8000	Potassium Chloride (KCl)
10 N Sodium Hydroxide (NaOH)	Formaldehyde
5 M Glycine	3 M Sodium Acetate
10% Tween20	5 M Sodium Chloride (NaCl)
10% Sodium Dodecyl Sulfate (SDS)	1.5 M Tris-HCl

5x Ligase Buffer

1M Tris (pH 8.0)	25 ml
1M MgCl2	5 ml
1M Dithiothreitol	0.5 ml
0.5M rATP	1 ml
Polyethylene glycol (8000)	25 g
ddH2O	to 100 ml

Linear Polyacrylamide (LPA)

Acrylamide	5 mg
TE Buffer (pH 8.0)	200 µl
10 % APS	1 µl
Tetramethylethylenediamine (TEMED)	1 µl
Ethanol	505 µl

Acrylamide was dissolved in TE buffer before APS and TEMED were added and left to polymerise. 2.5 volumes of ethanol were added and the polymerised acrylamide pelleted (4000 rpm, 10 mins) and resuspended in mQ H2O (500 µl).

6x Loading Dye (10 ml)

Bromophenol blue (0.25% w/v)	25 mg
Xylene cyanol FF (0.25% w/v)	25 mg
Glycerol (30% v/v)	3 ml

10x NEBuffer 4 (20 ml)

1 M Tris-HCl	10 ml
2 M MgCl2	1 ml
2.5 M NaCl	8 ml
1 M Dithiothreitol	200 µl
ddH2O	800 µl

All solutions were sterilised prior to combining.

Phenol:Chloroform
Buffer saturated phenol (pH 7.5) and chloroform were mixed at a 1:1 ratio.

Solution A (10 ml)

10% Triton X-100 (2% v/v)	2 ml
10% SDS (1% v/v)	1 ml
5M NaCl (0.1 M v/v)	200 µl
1M Tris (pH 8.0, 0.01 M v/v)	100 µl
0.5M EDTA (1 mM v/v)	20 µl
ddH2O	to 10 ml

1x TE Buffer (pH 8.0; 1 l)

1 M Tris-HCl (pH 7.5; 100 mM final v/v)	10 ml
0.5M EDTA (pH 8.0; 10 mM final v/v)	20 ml
ddH2O	to 1 L

5x Tris/Borate/EDTA (TBE) Buffer (1 l)

Tris	54 g
Boric Acid	27.5 g
diNa-EDTA	3.27 g
ddH2O	to 1 L

TBE buffer was used at final concentration of 0.5x [179].

Solutions for chromatin preparation

Chromatin Digestion Buffer

1M Tris-HCl (pH 8.0)	150 µl
1M MgCl2	75 µl
11% Triton X-100	136.5 µl
ddH2O	14.64 ml

FA Lysis Buffer (10 ml)

1M HEPES-KOH	500 µl
5M NaCl	280 µl
0.5M EDTA	20 µl
11% Triton X-100	900 µl
10% DOC	100 µl
ddH2O	8.2 ml

One protease inhibitor cocktail tablet (EDTA-free; Roche) was added per 10 ml.

Wash Buffer (40ml)

11% Triton X-100 (1% final v/v)	3.6 ml
10x PBS	4.0 ml
ddH2O	to 40 ml

2.1.3 Media

Media for bacterial and *S. cerevisiae* cultures was prepared using ddH20 and sterilised by: 1) autoclaving at 15 psi, 121°C for 20 min or 2) pressure filtering using (Millipore 0.2 µm syringe filters), or 3) Millipore 0.2 µm filters for filtering glycerol lactate media after pH adjustment.

Bacterial Media

Antibiotics
When required for selection, Ampicillin (100 $\mu g ml^{-1}$; Roche, Germany) was added to LB media and agar plates.

Agar Plates
When required agar (15 $g L^{-1}$) was added to LB media prior to autoclaving.

Luria-Bertani (LB) Media (1 litre)

Bacto Tryptone	10 g
Bacto Yeast Extract	5 g
NaCl	10 g
ddH2O	to 1 L

The pH was adjusted to 7.0 with NaOH. For LB plates 20 g agar was added. Ampicilin was introduced as a selection agent where appropriate at 100 mg/ml. 20 µl of X-Gal was spread directly onto the plate surface prior to inoculation using a bent stirring rod, where blue:white selection was required.

SOC media (1 litre)

Bacto Tryptone	20 g
Bacto Yeast Extract	5 g
NaCl	0.5 g
ddH2O	to 1 L

10 ml KCl (final concentration 250 mM) was added once solutes had dissolved. The pH was adjusted to 7.0 with 5 N NaOH. After autoclave sterilisation, and just before use, 5ml MgCl2 (2 M) and 20 ml sterile glucose (1 M) were added.

Yeast Media

Yeast media was made in ddH20 and sterilised by autoclaving (15 psi, 121°C for 20 min). Where a single experiment required different media the different batches were made together using the same reagents and autoclaving conditions. For plates, 20 g of agar was added prior to autoclaving.

Synthetic Complete Drop Out Mix

Adenine	0.5 g
Arginine	2.0 g
Aspartic acid	2.0 g
Glycine	2.0 g
Histidine*	2.0 g
Inositol	2.0 g
Isoleucine	2.0 g
Leucine*	10.0 g
Lysine	2.0 g
Methionine	2.0 g
Phenylalanine	2.0 g
Threonine	2.0 g
Tryptophan*	2.0 g
Tyrosine	2.0 g
Uracil*	2.0 g

Amino acids (Formedium, Hunstanton, England) were mixed and stored at 4°C. *Optional amino acids were ‘dropped out’ according to whether there was to be an amino acid selection process. Otherwise they were added separately according to the following concentrations:

Histidine	0.076 g/l
Leucine	0.379 g/l
Tryptophan	0.076 g/l
Uracil	0.076 g/l

Galactose Synthetic Complete (SC) Media

Yeast nitrogen base (w/o amino acids)	6.7 g
SC Drop out mix	2 g
Galactose	20 g
ddH2O	to 950 ml

Galactose (20 g) was dissolved in 50 ml ddH2O, filter sterilised and added after autoclaving.

Glucose Synthetic Complete (SC) Media (1 litre)

Yeast Nitrogen Base (w/o amino acids)	6.7 g
Glucose	20 g
SC Drop Out Mix	2 g
ddH2O	to 1 L

Glycerol Lactate Synthetic Complete (SC) Media

Yeast nitrogen base (w/o amino acids)	6.7 g
Glycerol	10 ml
Lactic acid	23.52 ml
SC Drop out mix	2 g
Glucose	0.5 g
ddH2O	to 1L

The pH was adjusted to 5.4 with NaOH.

YPD Media

Bacto peptone	20 g
Bacto yeast extract	10 g
Glucose	20 g

2.1.4 Enzymes

Restriction enzymes (*i.e. MspI*, *HindIII*) were purchased from Fermentas (Maryland, USA), Invitrogen (Carlsbad, CA USA), and New England Biolabs (Ipswich, MA, USA). Proteinase K (14-22 $mgml^{-1}$) was purchased from Fermentas (Maryland, USA) and Roche (Indianapolis, USA). RNase was purchased from Roche (Mannheim, Germany). Taq Polymerase was purchased from New England Biolabs (Ipswich, MA, USA). Protease inhibitor cocktail tablets (EDTA-free) were purchased from Roche (Mannheim, Germany). T4 DNA Ligase (1 $U\mu l^{-1}$) was purchased from Invitrogen (Carlsbad, CA USA).

2.1.5 Organisms and strains

The organisms used in this study are listed in Table 2.1.

Table 2.1: Organisms

Organism	Genotype
Saccharomyces cerevisiae	
BY4741	*Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*
yme1	*Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yme1Δ::kanMX4*
yme2	*Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yme2Δ::kanMX4*
mdv1	*Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mdv1Δ::kanMX4*
rex2	*Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rex2Δ::kanMX4*
mdm10	*Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mdm10Δ::kanMX4*
ade2	*Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ade2Δ::kanMX4*
161-U7 1+2+	1+2+ [*Mata, ade1, lys1, ura3*]. Wild type [7]
161-U7 GII-0	GII-0 [Mata, ade1, lys1, ura3]. No group II introns [7, 9]
161-U7 GII-0 aI5γ	GII-0 aI5γ [*Mata, ade1, lys1, ura3*]. No group II introns aI5γ retained [7, 9]
yPH499	*MATa ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1*
yDP77	*MATa ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1 tL(UAA)B2:: tY(GUA)J2 (SUP4-1)*
yDP84	*MATa ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1 tL(CAA)G3:: tY(GUA)J2 (SUP4-1)*
Escherichia coli	
DH5-α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

Approvals for the development of genetically modified organisms in containment

GMO Approval was obtained for the cloning of yeast DNA into *E. coli* from the Massey University IBSC and ERMA. Approval number: GMO06/MU009.

2.2 Methods

2.2.1 Cell culture techniques

2.2.1.1 Maintenance of *E. coli* cells

E. coli was maintained in LB media with 10 or 15% glycerol at -80°C.

2.2.1.2 Maintenance of *S. cerevisiae* cells

S. cerevisiae cells were cultured in SC media (10 ml, 37°C, 160 rpm, 16 hrs). Cells were pelleted (4 ml, 4200 rpm, 3 mins, 4°C) and washed with fresh SC media (4200 rpm, 3 mins, 4°C) before being resuspended in 1 ml SC media containing 20% glycerol and transferred to a cryotube. Cryotubes were frozen at -80°C.

2.2.1.3 Seeding agar plates from frozen stocks

Agar plates were inoculated from -80°C freezer stocks by scraping the frozen cells with an inoculation loop and streaking them onto an agar plate.

2.2.1.4 Culturing *S. cerevisiae* cells

A single colony from the agar plates used to rescue the cells from -80°C freezer stocks was used to inoculate the media (10 ml) for culturing (30°C, 16 hrs, 160 rpm). The type of media depended upon the impending experiment. When SC galactose grown cells were ultimately required the starter culture consisted of SC glycerol lactate media and was cultured for 40 hrs (30°C, 160 rpm).

For cultures destined for 3C analysis, 3-80 µl of the 10 ml starter culture was used to inoculate SC media (50 ml, unless otherwise stated) which were subsequently cultured (30°C, 15-20 hrs, 160 rpm) until an OD_{600} of 0.600 was reached.

2.2.2 Molecular biology techniques

2.2.2.1 Polymerase Chain Reaction (PCR)

General PCR was carried out using a typical master mix (Table 2.2) unless otherwise stated. DMSO was added to reduce secondary structure. Tween20, a non-ionic detergent, was added to stabilise the Taq polymerase on the template [180]. Taq Polymerase has an error rate of $1.1x10^{-4}$ errors/bp [181] and has a rate of ~35-150 nts/second. Variations of the master mix are explicitly stated in the text. Variations include changes to primer concentration and DNA template concentration. PCR temperature cycling was tailored for each primer set however; Table 2.3 outlines a typical PCR cycling protocol. Longer extension times were performed for PCR products greater than 500 bp. Temperature gradients (55-65°C) were carried out to determine at what temperature specific primers preferentially annealed to the template when the typical protocol failed.

Table 2.2: General PCR Mastermix Protocol.

10x PCR Buffer	1.50 µl
dNTPs (1.25 nmol each dATP, aGTP, dTTP, dCTP)	2.40 µl
Primer (10 pmol)	0.45 µl
100% DMSO	0.45 µl
10% Tween20	0.15 µl
ddH_2O	7.92 µl
Taq Polymerase	0.13 µl
DNA Template	2.00 µl
Total	15.00 µl

Table 2.3: General PCR temperature cycling protocol.

95°C	2 min	1x
95°C	30 sec	35x
57°C	30 sec	
72°C	30 sec	
72°C	3 min	1x
4°C	∞	

2.2.2.2 Agarose gel electrophoresis

Agarose powder (2% w/v) was dissolved in 1x TBE buffer unless stated otherwise. DNA was prepared by the addition of DNA loading dye (1x final v/v) before electrophoresis at 80 volts until the DNA loading dye was well resolved. The agarose gel was stained in an ethidium bromide bath (15 mins) before visualisation by UV (302 nm) on a BioRad Gel Doc system.

2.2.2.3 Restriction endonuclease digestion

Restriction enzyme digests were typically performed with 10 U in a final volume of 20 µl at 37°C for 2-15 hrs before inactivation at 65°C, unless stated otherwise. Digests were performed in the appropriate buffer according to the specific enzyme being used. The volume of H_2O changed depending on the concentration and volume of DNA added.

2.2.2.4 Ligation

Ligations were typically performed with 5 U in a final volume of 25 µl at 16°C for 2-15 hrs unless otherwise stated. The volume and concentration of DNA changed for each ligation. Where DNA fragments were to be ligated into a plasmid, an insert to vector molar ratio of 3:1 was typically used.

2.2.2.5 Purification of DNA fragments

Purification of PCR products was achieved using either a Roche High Pure PCR Purification Kit (Basel, Switzerland), or a Zymo Research Clean and Concentrator Kit (-5; Irvine, CA, USA), according to manufacturer's instructions. Purification of DNA from agarose gels was accomplished by excising EtBr stained bands under ultra violet light (302 nm), dissolving agarose (50°C) and extracting DNA using the Roche High Pure PCR Purification Kit according to manufacturer's instructions.

2.2.2.6 Preparation of electro-competent DH5-α *Escherichia coli*

E. coli DH5-α (40 µl) cells were removed from storage and inoculated into LB broth (10 ml) and incubated (37°C, 160 rpm, O/N). 250 ml of LB was pre-warmed (30°C), inoculated with 3 ml of the O/N culture, and incubated (37°C, 160 rpm, 2 hrs) until the culture reached an OD_{600} of 0.500. Cells were held on ice for 15 mins before being pelleted (5000 rpm, 15 mins, 4°C), washed three times in reducing volumes of cold 10% glycerol (5000 rpm, 15 mins, 4°C), before resuspension in 0.8 ml cold 10% glycerol. Cells were aliquot (40 µl) and snap frozen in a dry ice-ethanol bath before being and stored (-80°C).

2.2.2.7 Preparation of plasmid for cloning

All plasmid DNA was desalted prior to transformation by purifying through a Zymo Clean and Concentrator column and resuspending in 20 µl H_2O (65°C).

2.2.2.8 Transformation of *E. coli*

Plasmid (25 ng) was mixed with of *E. coli* DH5α electro-competent cells (20 µl) and added to a pre-chilled electroporation curvette. The cells were subjected to one pulse in a Bio-Rad Gene Pulsar II electroporator (2000 volts; 25 µF; 200 Ω) before being resuspended in 1 ml SOC media (4°C) and recovered (37°C, 160 rpm, 1 h). For blue:white selection the transformed cells were spread onto LB Amp X-Gal plates and incubated (37°C, 16 hrs). pUC19, which had not been cut with the restriction enzyme, was used as a positive control, cut pUC19, cut and dephosphorylated pUC19, and ddH_2O were used as negative controls.

2.2.2.9 Harvesting transformed *E. coli*

Transformed *E. coli* (20 µl) was spread on LB Ampicillin X-Gal Plates, which were then incubated (37 °C, O/N). White colonies were picked off the plates using autoclaved toothpicks.

2.2.2.10 Bulk-up and isolation of plasmid DNA

Transformed cells were isolated off blue:white selection plates. 10 ml LB Amp liquid media was inoculated with a chosen colony and incubated (37°C, 16 hrs, 160 rpm). Plasmid DNA was extracted from culture (4 ml) using a Roche High Pure Plasmid Isolation Kit, according to manufacturer's instructions. Plasmid was eluted and resuspended in 100 µl ddH_2O.

2.2.2.11 2,4-Dinitrophenol mediated uncoupling of oxidative phosphorylation

Uncoupling of the electron transport chain was achieved with the proton-ionophore 2,4-dinitrophenol. Initial experiments determined the concentration at which the cells growth under respiratory conditions (glycerol lactate) would cease, but cells which were grown on a fermentable carbon source (glucose) could continue growing (Figure 4.13).

S. cerevisiae cells were cultured (30°C, 160 rpm) to an OD_{600} of 0.600 in SC glucose or SC galactose media before treatment with 0.5 mM 2,4-Dinitrophenol for varying times (45 mins, 90 mins, or 180 mins [galactose only]). Cells were harvested and prepared for quantitative 3C analysis as described in 2.2.3.9.

2.2.2.12 Cell cycle arrest of *S. cerevisiae* cells

S. cerevisiae cells were arrested at different points in the cell cycle using chemical cell cycle inhibitors. Hydroxyurea, nocodazole, and α-factor were used to arrest the cell cycle at S, G2/M, and G1 phase, respectively.

S. cerevisiae (BY4741) cells were cultured (30°C, 160 rpm) in SC glucose media to an OD_{600} of 0.600. Hydroxyurea (0.1 mM final v/v), nocodazole (3.4 µM final v/v), and α-factor (15 $\mu g ml^{-1}$ final v/v) were added to three cultures initially split from one large culture (150 ml) and incubated (30°C, 160 rpm, 3 hrs). Cell cycle arrest was confirmed by light microscopy (Figure 2.1). Cells were harvested and prepared for quantitative 3C analyses as described in 2.2.3.9. All cell cycle samples were compared against an untreated control harvested at the time of cell cycle inhibitor addition.

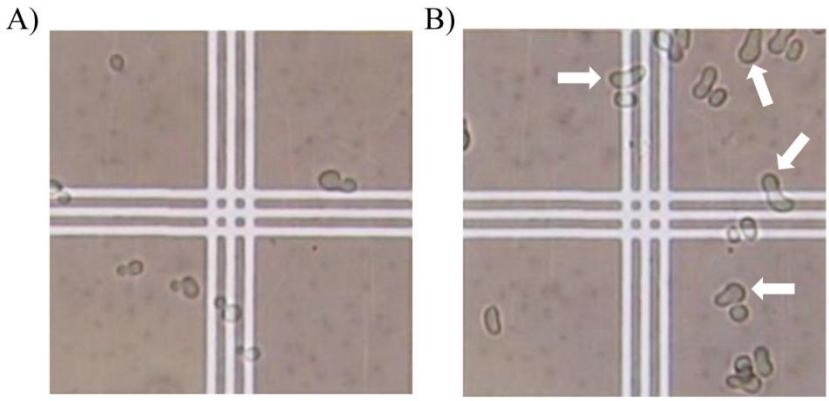

Figure 2.1: *S. cerevisiae* cells 'Schmoo' as a result of alpha-factor induced cell cycle arrest, as observed by light microscopy (400x).
A) Untreated cells display normal budding morphology. B) Cells treated with alpha-factor have a schmoozing morphology. Schmooing cells are highlighted with white arrows.

2.2.2.13 Isolation of *S. cerevisiae* genomic DNA

S. cerevisiae (BY4741) was cultured (100 ml, 30°C, 160 rpm) in SC Media to a final density of $10x10^9$ cells in 60 ml. Cell suspension (60 ml) was pelleted (3000 rpm, 3 mins), washed in ddH_2O (6 ml) and aliquoted into four 1.5 ml microfuge tubes. Cells were pelleted (3000 rpm, 3 mins) before discarding the supernatant. Solution A (200 µl), phenol:chloroform (200 µl) and acid washed glass beads (0.3 g) were added to the cell pellet before vortexing at maximum rpm for 2 mins. TE buffer (200 µl) was added before pelleting the cell debris (13,000 rpm, 5 mins). The aqueous layer was transferred to a new tube and extracted twice with phenol:chloroform (200 µl, 7500 rpm, 7 mins). The aqueous layer was added to 1 ml EtOH (100%) and mixed by inversion. The sample was pelleted (13,000 rpm, 2 mins) and the supernatant removed resuspension in TE buffer (400 µl) and treated with RNaseA (1.2 µl [10 mg/ml], 37°C, 5 mins). NH_4OAc (18 µl, 5M) and EtOH (1 ml) were added prior to precipitation (-20°C, 2 hrs). DNA was pelleted (13,000 rpm, 10 mins) before resuspension in ddH_2O (50 µl).

2.2.3 Capturing Chromosomal Interactions Techniques

2.2.3.1 Harvesting cells and chromatin preparation

S. cerevisiae cells were cultured (30°C, 160rpm) to an OD_{600} of 0.600. Cultures were crosslinked 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 resuspended in FA lysis buffer (400 µl). 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 secs, max rpm) and held on ice between cycles (30 secs). 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 resuspension in chromatin digestion buffer (400 µl). 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.

2.2.3.2 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).

2.2.3.3 Genome Conformation Capture (GCC)

Saccharomyces cerevisiae (BY4741 [Mat**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0]) was cultured (30^oC, 160rpm) to an A_{600} = 0.6 in synthetic complete medium containing amino acid supplements and glucose (2% w/v), glycerol lactate (glycerol 1% v/v, lactate 2% v/v), or galactose (2% w/v). Chromatin was prepared using a modified 3C protocol [70, 88]. Briefly, cells were cross-linked (2% v/v formaldehyde, 10 mins, room temperature, with stirring) and quenched (0.125M glycine). Fifteen samples (residuals of 10^8 cells in 52.6 µl) were digested (37°C, 2 hrs) with *Msp*I (100 U, Fermentas) in a final volume of 100 µl. *Msp*I was inactivated by the addition of 1% SDS (final v/v) and incubation (65°C, 20 mins). Reactions were diluted in 2 ml T4 ligation buffer (NEB) containing 1% Triton

X-100 and external controls added (see section 2.2.3.5) and 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 (7-11 ug, Roche), 20 µl EDTA (0.5 M), 12 µl NaCl (5 M), 1.2 µl Tris-HCl (1 M). The second external control (pUC19; see section 2.2.3.5) and RNase (20 µg) were added and incubated (37°C, 15 mins) prior to three Phenol:Chloroform (1:1) extractions. Samples were purified through 8 columns (Zymo Clean and Concentrator) and each eluted in 15 µl milliQ H_2O. Samples were vacuum concentrated to a volume of 50 µl. DNA (5 µg in 100 µl) was sequenced using the Illumina Solexa platform (Allan Wilson Centre, Massey University and the Friedrich Miescher Institute, Basel, Switzerland).

2.2.3.4 Circular Chromosome Confirmation Capture (4C)

The *S. cerevisiae* strain, yPH499 (wild type), and two mutant isolates, yDP77, and yDP84, were cultured in synthetic complete media containing amino acid supplements and glucose (2% w/v) to an OD_{600} of 0.600. The chromatin was harvested (2.2.3.1) and prepared (2.2.3.2) using the same protocol as a 3C sample. Nested PCR primers (Figure 2.5) were designed to amplify inversely from the 4C 'bait' *MspI* fragments (Figure 2.2). Nested inverse primers were designed for two separate *MspI* fragments, which had the wild type tRNAs intact, in the wild type strain, yPH499 (Chr II: 347,518 – 349,584 bp [Figure 2.2A]; Chr VII: 857,273 – 857,825 bp [Figure 2.2C]). The tRNAs in these regions were replaced with a SUP4-1 suppressor tRNA, creating two mutant *S. cerevisiae* isolates, yDP77 (Figure 2.2B) and yDP84 (Figure 2.2D). The SUP4-1 replacement sequence has an *MspI* site in it, thus the fragment was essentially cut in two. Outside and nested primers were designed to amplify out of fragment 1 and 2 for both of the replacement isolates (yDP77 [Figure 2.2B] and yDP84 [Figure 2.2D]). The nested primers had 6 bp tags attached to them so that the 4C PCR product could be pooled: A) TCTCTG; B) fragment 1 ACAGAG; fragment 2 TAGATC; C) AGAGAC; and D) fragment 1 TGATGC; fragment 2 AGCACG. The 4C PCR product was sent for 100 bp paired end sequencing on an Illumina Genome Analyser (Allan Wilson Centre, Massey University).

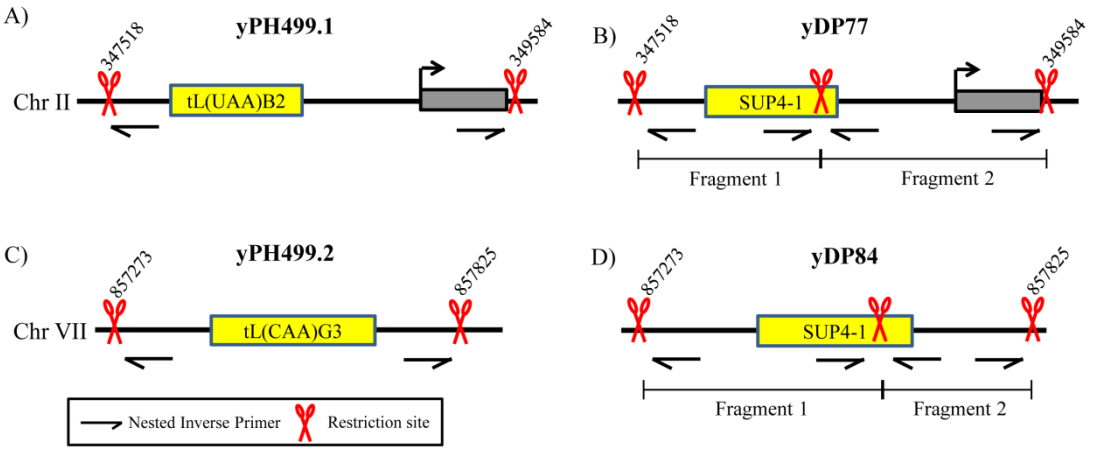

Figure 2.2: Nested inverse 4C primers were designed to amplify from two separate *MspI* fragments in the *S. cerevisiae* genome.
tRNAs which occur in two different regions of the wild type *S. cerevisiae* genome (yPH499; [A) Chr II: 347,518 – 349,584 bp]; [C) Chr VII: 857,273 – 857,825 bp]), have been replaced with the SUP4-1 suppressor tRNA gene; B) yDP77 and D) yDP84, respectively. Outside and nested inverse 4C primers were designed to amplify out of the two *MspI* fragments with the wild type tRNA intact (A) WTyDP84outsideF, WTyDP84outsideR, C) WTyDP84NestedF, WTyDP84NestedR). The SUP4-1 replacement sequence has an *MspI* site within it, essentially cutting the fragment in two, thus, 4C primers were designed to amplify from both fragment 1 and fragment 2 for both replacement regions (B) Fragment 1: F1MUTyDP77outsideF, F1MUTyDP77outsideR, F1MUTyDP77nestedF, F1MUTyDP77nestedR, and Fragment 2: F2MUTyDP77outsideF, F2MUTyDP77outsideR, F2MUTyDP77nestedF, F2MUTyDP77nestedR. D) Fragment 1: F1MUTyDP84outsideF, F1MUTyDP84outsideR, F1MUTyDP84nestedF, F1MUTyDP84nestedR and Fragment 2: F2MUTyDP84outsideF, F2MUTyDP84outsideR, F2MUTyDP84nestedF, F2MUTyDP84nestedR). For primer sequences see Appendix I. The nested primers had 6 bp tags attached to them so that the PCR product could be pooled and sequenced; A) TCTCTG, B) fragment 1 ACAGAG, fragment 2 TAGATC, C) AGAGAC, and D) fragment 1 TGATGC, fragment 2 AGCACG.

2.2.3.5 Generation of controls for GCC

External controls were included during the GCC library preparation to ascertain what frequency of restriction fragment ligation pairing was above experimental noise. The first control consisted of a PCR product with a free *MspI* restriction site at one end, added before the ligation step during GCC. The PCR products have not been cross-linked to any region of the genome, thus if ligations occur we deem them to be experimental noise. External controls were produced by the PCR amplification (for primers see Appendix I) of short regions from the *E. coli* and Lambda phage genomes. PCR products were digested with *MspI* (37°C, 2 hrs) prior to column purification (Zymo Clean and Concentrator). Purified, digested PCR products were introduced into the GCC samples at a 1:1 ratio with the nuclear genome.

During sample preparation for the Illumina Genome Analyser, a ligation step is included. This ligation event is performed to attach linker adapters to the ends of the DNA molecules to be sequenced. This ligation step is performed after a DNA end-repair step, therefore, theoretically, random ligation events could occur during this preparation

for DNA sequencing [182]. I was only interested in ligation events occurring about the chosen restriction enzyme recognition site, nevertheless, the rate at which false positive ligation events occurred during the sequencing preparation of the GCC sample at the Solexa sequencing centre needed to be established. Consequently, GCC required an external control.

pUC19 plasmid was added to the GCC sample prior to phenol:chloroform extraction. As the addition of the pUC19 plasmids was after the GCC ligation step, proteinase k digestion and reverse crosslinking, it was reasoned that no ligation event should be observed between the pUC19 plasmid and the *S. cerevisiae* genome. If pUC19-genome interactions were observed it suggested that: 1) the sequence length attempting to be aligned against the *S. cerevisiae* genome was too short and we were observing mis-alignment of reference genome sequences onto the pUC19 reference sequence or; 2) a ligation had occurred during sequencing preparation. pUC19 plasmid was prepared as indicated (2.2.2.10). A total of $1.5x10^7$ pUC19 plasmids were added per finished GCC library, each of the fifteen residual samples having $1x10^6$ copies added (2 µl of a 13.7 $ngml^{-1}$ solution).

2.2.3.6 Linear vs. circular DNA purification bias

pUC19 was linearised by co-digestion with *Eco*RI and *Hind*III (O/N, 37^{o}C). Linearised and circular pUC19 samples were mixed in an equal ratio (0.1375 $ngml^{-1}$ each). Quantitative real-time PCR to check for a methodological purification bias in the GCC method was performed on an ABI Prism 7000 Sequence Detection System (SDS7000).

Reactions (15 µl) contained: 1x Platinum SYBR Green qPCR Super-Mix UDG (Invitrogen); 1 nmol of each primer (see Appendix I); and 2 µl of GCC library containing 1.375 ng ml^{-1} of cut and uncut pUC19. PCR was performed using a five stage program: 50^{o}C, 2:00 mins; 95^{o}C, 2:00 mins; 40x[95^{o}C, 0:15 secs; 58^{o}C, 0:30 secs; 72^{o}C, 0:30 secs]; 55^{o}C, 1:00 min; followed by a dissociation analysis. Primers 3CpUC19EcoRIfwd and 3CpUC19EcoRIrev crossed the pUC19 multiple cloning site.

2.2.3.7 Sequencing Strategies

The Illumina (Solexa) Genome Analyser platform was used for all sequencing. Single-end and paired-end strategies were employed to sequence the GCC libraries prepared in this study. Briefly, Illumina (Solexa) Genome Analyser sequencing technology is based on reversible dye-terminators ([reviewed in] [183]). DNA molecules are attached to primers

on a slide and amplified to generate clusters. Nucleotides are added to the slide and a camera takes images of the fluorescently labelled nucleotides before the dye and terminal 3' blocker are chemically removed thus allowing the next cycle of synthesis. Initially, the Illumina Genome Analyser was only capable of producing 36 bp reads, but recent advances in high-throughput sequencing technology mean it is now able to produce 150 bp paired-end reads. Paired-end sequencing is achieved by 'flipping' the cluster DNA molecules over and sequencing from the opposite end. The sequencing output dictated what type of analysis was performed on the resulting data (Figure 2.3). Sequencing strategies will be expanded upon in the results chapters and in section 2.2.4.4.

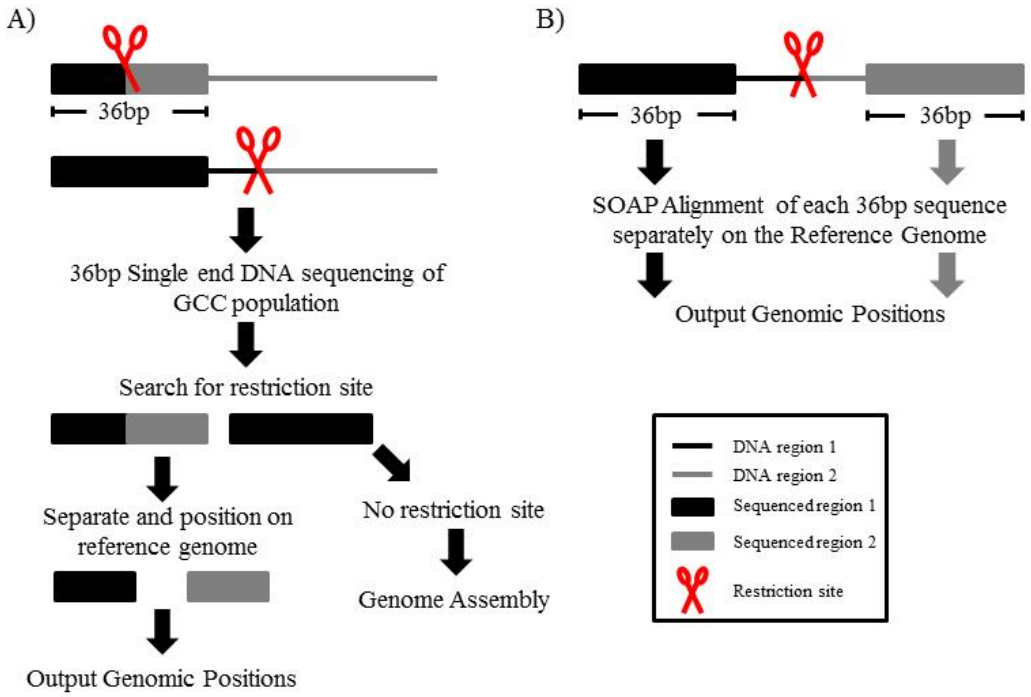

Figure 2.3: GCC Single-end and Paired-end sequencing strategies.
The sequencing strategy employed dictated the necessary analysis required to identify interacting DNA sequences. A) Single-end sequencing requires that the restriction site falls within the sequence tag. The sequence on each side of the restriction site is then aligned against the reference genome. B) Paired-end sequencing results in each sequenced 'end' being aligned against the reference genome. Paired-end sequencing also incorporates the steps in section A) for tags which contain the restriction enzyme site.

2.2.3.8 3C PCR

PCR primers were designed to amplify across suspected ligation junctions. Standard PCR protocols were followed as per section 2.2.2.1. Positive controls were constructed as per section 2.2.3.10.

2.2.3.9 Quantitative 3C

Quantitative 3C PCR analyses were performed using FAM labelled BHQ Probes (BioSearch Technologies; see Appendix I) and Taqman® Gene Expression Master Mix (Applied Biosystems) on an ABI Prism 7000 Sequence Detection System (SDS7000). Primer (900 nM) and probe concentrations (250 nM) were determined by performing concentration gradients (primer gradient 50, 300 or 900 mM; probe gradient 50, 150, 250 nM). Coordinates for interactions under investigation are defined in the results chapters. Samples (2 µl in triplicate) were analysed in a final reaction volume of 20 µl (Table 2.4) using primers listed in Appendix I. Assays were performed using a 3-stage program (50°C, 2:00 mins; 95°C, 10:00 mins; 45x[95°C, 0:15 secs; 60°C, 1:00 min]).

Table 2.4: Quantitative 3C PCR mastermix.

2x Taqman Mastermix	10 µl
Forward Primer	2 µl
Reverse Primer	2 µl
Flurogenic Probe	2 µl
DNA Template	2 µl
Total	20 µl

All results were expressed as a percentage of wild-type, following standardisation for: 1) the number of mitochondrial genomes or 2) the number of genomes (measured using the *GAL1* locus [88]), further details on calculations can be found in Appendix II.

2.2.3.10 Fluorogenic probe real-time PCR standard generation

Dedicated interaction standards were prepared by PCR amplification, from *S. cerevisiae* BY4741 genomic DNA (BY4741), of the *MspI* or *HindIII* (*i.e.* the restriction enzymes most frequently used in this study) fragments involved in interactions under investigation (Figure 2.4). Primers were designed so that the resulting product contained the *MspI* restriction site which the fluorogenic probe and primer set were designed to amplify across (see Appendix I for primers). The PCR product was purified using a Zymo Clean and Concentrator -5 column and eluted in ddH_2O (20 µl, 65°C), digested with *MspI* (37°C, 2 hrs), after which the *MspI* enzyme was inactivated (65°C, 20 mins). The pair of *MspI* fragments which were suspected of interactions were then ligated (10 U, 16°C, 2 hrs) with T4 DNA ligase. Following ligation the sample was purified through a Zymo Clean and Concentrator -5 column and the concentration checked by a nanodrop. The standards were created by an initial dilution of the ligated product to 2 $ngµl^{-1}$ and then subsequent serial dilutions down to $2x10^{-8}$ $ngµl^{-1}$.

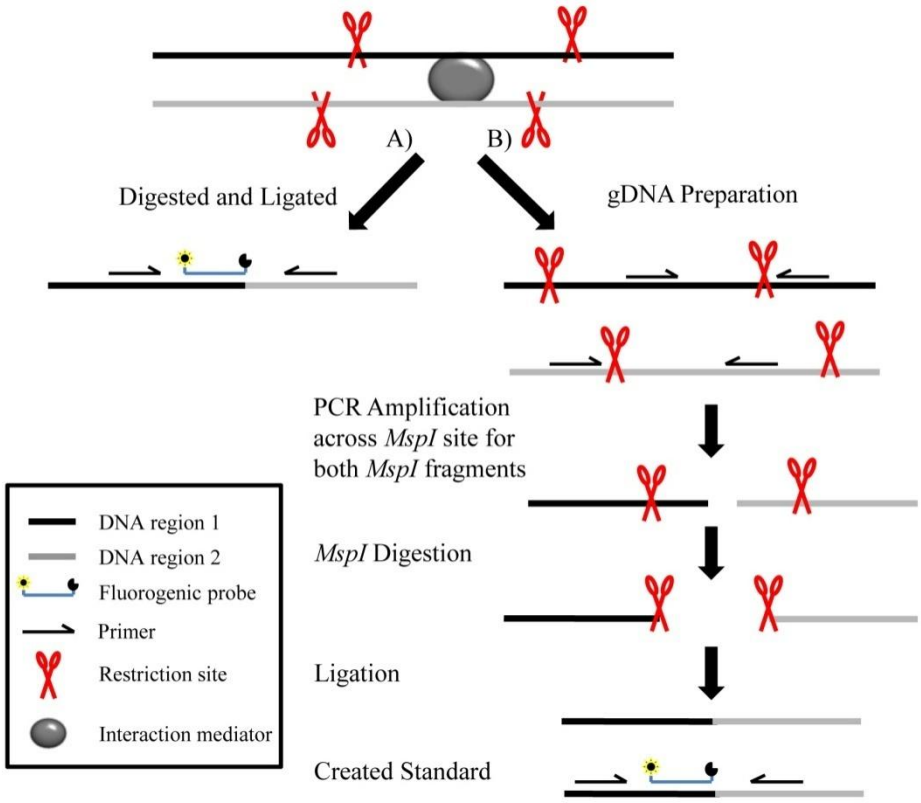

Figure 2.4: Generation of fluorogenic probe PCR standards.
A) An interaction between two regions of the genome can be identified by 3C taqman PCR using a fluorogenic probe. B) PCR standards are created to emulate a real interaction. Briefly, standards are achieved by the amplification of a restriction fragment, digestion, and ligation of the two pieces of DNA. The primers and probe can now bind across the ligation junction and can therefore be used at known concentrations as real-time PCR standards.

2.2.3.11 Copy Number Determination - Sybr green qPCR

Mitochondrial genome and nuclear genome copy number were determined by Sybr-green qPCR in a final volume of 15 µl (Table 2.5). Primers (GAL1F and GAL1R; Appendix I) were designed to amplify a region of the nuclear encoded *GAL1* (Chr 2: 279,790-279,935 bp) gene and also a region of the mitochondrially encoded *COX1* (Mito+ve13909F and Mito(CNC)R13909, Chr Mito: 25,535-25,686 bp). A five stage program was used (50°C, 2:00 mins; 95°C, 2:00 mins; 40x[95°C, 0:15 secs; 59.5°C, 0:30 secs; 72°C, 0:30 secs]; 55°C, 1:00 min; followed by a dissociation analysis). *S. cerevisiae* BY4741 genomic DNA samples (concentration from: 2 ng μl^{-1} – 7.78125x10^{-4} ng μl^{-1}) were used as a standards for all Sybr-green assays unless otherwise stated.

Table 2.5: Sybr green copy number analysis mastermix.

2x Sybr Green Mastermix	10 µl
Primer Pair	2 µl
DNA Template	2 µl
Total	15 µl

2.2.4 Bioinformatic analyses

2.2.4.1 *S. cerevisiae* reference genome

The reference genome used in this study consists of the sixteen *S. cerevisiae* nuclear chromosomes (NC_001133.7, NC_001134.7, NC_001135.4, NC_001136.8, NC_001137.2, NC_001138.4, NC_001139.8, NC_001140.5, NC_001141.1, NC_001142.7, NC_001143.7, NC_001144.4, NC_001145.2, NC_001146.6, NC_001147.5, NC_001148.3) as well as the mitochondrial genome (NC_001224.1) and the 2-micron plasmid (YSCPLASM), pUC19 and the ligation control sequences.

2.2.4.2 Primer Design

DNA sequences of interest were retrieved from the National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/) or the Saccharomyces Genome Database (SGD) (http://www.yeastgenome.org/). PCR primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3/). Primers were designed such that the GC content did not deviate outside 40-60%. The melting temperature (T_m) was normally kept between 52 and 58°C. Primer length ranged from 18 to 25 nt. Secondary structure was kept to a minimum especially at the 3' end. BLAST [184] and PatMatch (http://www.yeastgenome.org/cgi-bin/PATMATCH/nph-patmatch) were used to confirm that each primer did not align back against the reference genome more than once to reduce the occurrence of non-specific PCR products.

Inverse PCR primers were designed by artificially copying one fragment, joining them end-to-end, and designing the primers to PCR across this junction (Figure 2.5). Nested primers for each fragment were designed to enable two rounds of PCR to be performed. Because in many situations fragments of interest ligate back onto themselves (self-ligation), nested primers offer a chance to limit the length of the self-ligated product to enable removal by a purification kit following amplification. Thus the nested PCR product from self-ligated products was designed to be less than 100 nt in length, a small enough molecule to escape purification and be washed through a column.

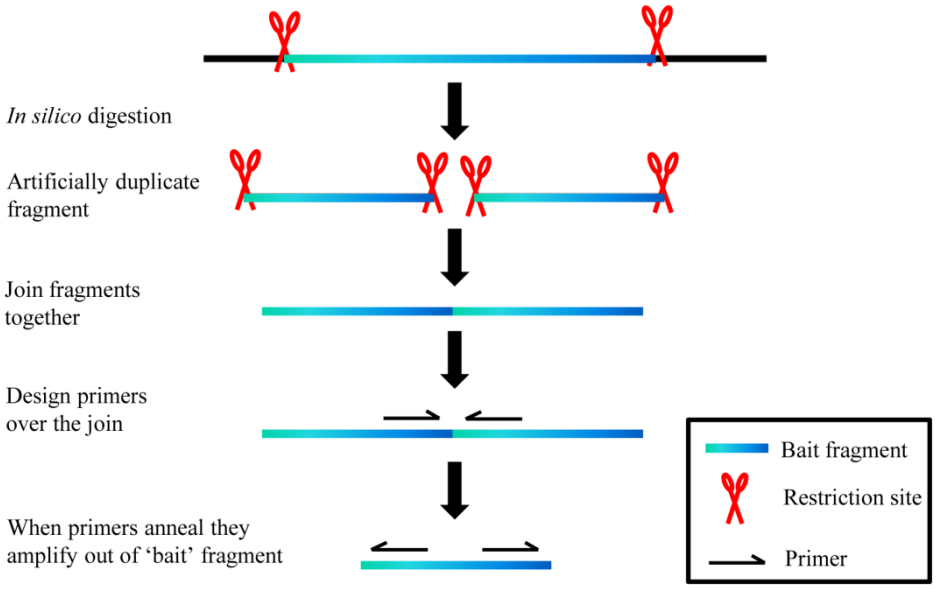

Figure 2.5: Inverse primers are designed to perform 4C analysis from a 'bait' fragment.
A bait fragment of interest is chosen. The goal is to design primers which inversely amplify outwards from the bait fragment. The bait fragment is extracted from the sequence by performing an *in silico* restriction digest before the fragment is artificially duplicated and joined end to end. Primers are designed to amplify across this join junction. When the primers bind during the 4C assay they now bind inversely.

2.2.4.3 *In silico* restriction digest

DNA sequences of interest were retrieved from the National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/) or the Saccharomyces Genome Database (SGD) (http://www.yeastgenome.org/). *In silico* restriction enzyme digests were performed using WebCutter (http://bio.lundberg.gu.se/cutter2/).

2.2.4.4 Topography Program

The topography program was written in Java by computer technician Frederic Bertels, and expanded by Gerd Grunert, to simplify the work flow for analysing GCC sequence data [5]. Topography is able to process the workflow for single and paired-end sequence files from the Illumina Genome Analyser in conjunction with the SOAP algorithm [8].

Single-end sequences

The sequence files, consisting of 36 bp single-end reads, were fed into Topography v1.19, along with the reference genome sequence (saccharomycesGenome_ with_ligation_controls.fasta) and GenBank (soapsgd_with_ligation_controls.gbk) file

(Appendix CD). In effect, Topography searches for sequences that contain the restriction enzyme recognition site, which was used during the GCC library preparation. Those that do not contain the restriction site are discarded; those that do are split into two about this site. The program asks whether each side of the restriction site is greater than 13bp in length (the minimum length that can be uniquely aligned against the *S. cerevisiae* genome), then aligns the two sides against the reference genome using the SOAP algorithm [8]. The SOAP algorithm was incorporated into the topography program as it is ideal for short sequence alignment. Three files (among others), are output which end in the file extensions .dat, .datmult, and .datcomb. The .dat file contains interactions in which both sequences can be uniquely positioned upon the reference genome. The .datmult contains interactions in which at least one of the sequences in an interaction was positioned against what is deemed a repetitive element in the reference genome. The .datcomb file contains all interactions combined from .dat and .datmult files.

Paired-end sequences

Topography follows the same workflow when dealing with paired-end sequences except that it incorporates a further step. Paired end sequencing output files have one 'end' in one file and the other 'end' in another file. These sequence files are interlaced, such that both paired 'ends' are together labelled A and B as well as the pairs unique identifying name. Both 'ends' are aligned against the reference genome and are output as interactions if they originate from more than one restriction fragment.

2.2.4.5 Sequence Population Statistics

Population statistics for the Illumina Genome Analyser sequence data were determined to ensure that we were obtaining a good representation of the *S. cerevisiae* genome and the data was not biased for certain genomic features. This was achieved by first aligning the sequences files against the yeast genome, using the SOAP [8] (huge) algorithm, to determine what percentage of the sequences can be aligned. The sequence files were then individually aligned against different genomic features (centromeric DNA [CEN], telomeric sequences [TEL], an rDNA repeat, the 2-micron plasmid, the mitochondrial genome, and a set of 646 housekeeping genes [see the Appendix CD for list of genes]), to determine whether the percentage of aligned sequences was similar to the percentage these features make up within the *S. cerevisiae* genome (www.yeastgenome.org).

2.2.4.6 GCC Bioinformatic Analyses

Individual bioinformatic and statistical analyses are described within the results chapters.

Bioinformatic and statistical analyses were performed on interactions for which the sequences mapped uniquely onto the reference genome (except where indicated). Connections with the ribosomal DNA, 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 ~85 kbp mitochondrial genome, respectively [5]. All statistical analyses involving 2 micron plasmid, mitochondrial, or rDNA sequences included copy number corrections. Copy number corrections are discussed within the results chapters in context with the data.

Additional repetitive elements (LTRs, tRNAs and repetitive gene families) have been omitted from the analysis unless otherwise stated.

2.2.4.7 GCC Statistical Analyses

Explanations of the statistical analyses of the individual GCC datasets are undertaken in the results chapters.

Briefly, statistical calculations were performed on datasets in which the sequences could be uniquely positioned on the reference genome. Where the analysis was concerned with connections between elements considered repetitive (*i.e*. the rDNA, mtDNA and 2-micron plasmids) the copy numbers of the element in question was corrected for. Connections between unique sequences and repetitive elements were analyzed separately. Values for copy number correction were determined from: 1) published reports of the frequencies of these elements within *S. cerevisiae* [10, 150, 185]; and 2) the frequency of these elements in the GCC sequence files that were returned from the sequencing centres.

During the preparation of the GCC samples random ligation events occur. Thus, our null hypothesis is that all interactions we identified are a result of random ligation events, or that there are only random ligations. These events occur during the ligation following restriction enzyme digestion and during the linker addition step in the sequencing library preparation. In an attempt to control for this, two control steps were included in our analyses: 1) external controls were added during the GCC library

preparation to obtain estimates of the rates of inter-molecular ligation events, and 2) we performed statistical analyses to determine whether our GCC dataset is something other than random (2.2.4.8).

2.2.4.8 GCC simulations for test of randomness

100,000 *in silico* simulations of random pairings were performed to determine whether the GCC datasets as a whole differed from what we would expect at random. Each individual simulation contained the same total number of interactions as the GCC dataset under investigation. This allowed us to observe how many of the same pairings (the same two segments coming together) occur at random. Ultimately, this allowed us to say whether we observed a higher number of pairings in our GCC datasets or not and thus establish whether the GCC datasets consisted of only random ligation events or not.

2.2.4.9 Binomial analysis to establish experimental noise cut-off

We investigated which of the individual interactions were above experimental noise in the GCC dataset by using a binomial distribution. Under the model of experimental noise the number of times one specific pairing occurs is a binomially distributed random variable. Let S1 and S2 be the number of mitochondrial and nuclear segments, respectively, which participate in at least one interaction. We calculate the probability $P(X \geq k)$ where N is the number of observed pairings and p is 1 divided by S1 multiplied by S2, for one specific pairing to occur k or more times. L is then S1 multiplied by S2, being the number of possible pairings and we expect to see $L \cdot P(X \geq k)$ pairings occurring k or more times by chance. This provides us with the expected number of false positives. The value k is chosen to provide an acceptable number of false positives and consequently an acceptable false positive rate typically less than 0.050.

2.2.4.10 Copy number determination

Repetitive elements are those genomic features which occur more than once within the genome (*e.g.*, the mitochondrial genome, rDNA repeats and 2 micron plasmid). We calculated copy numbers for these elements in our samples by aligning the sequence files against sections of these elements (Table 2.6). Three short regions were chosen: 1) a unique nuclear element ([Gal] Chr II: 279790-279909); 2) a short section of rDNA ([rDNA] Chr XII: 460517-460612); and 3) a section of the mitochondrial genome ([Mito] Chr Mito: 25535-25654).

Table 2.6: Calculating Mitochondrial and rDNA copy number using Illumina sequence tags.

Gal (146bp)	**Sequence Reads**	**Aligned Reads**	**Percentage**
Glucose	112335584	733	0.00065251
Glycerol Lactate	98269812	767	0.0007805
Galactose	96838770	621	0.00064127
Mito (152bp)			
Glucose	112335584	8937	0.00795563
Glycerol Lactate	98269812	12156	0.01237002
Galactose	96838770	12808	0.01322611
rDNA (96bp)			
Glucose	112335584	103881	0.09247381
Glycerol Lactate	98269812	95001	0.09667364
Galactose	96838770	85548	0.08834065

Alignments were carried out using the SOAP algorithm [8]. Percentages were calculated by dividing the aligned reads with the sequence reads and multiplying by 100.

I calculated the ratio of rDNA or Mito to the unique nuclear element by dividing the percentage of sequence tags calculated in Table 2.6 for Mito and rDNA against the percentage for Gal. Calculated copy number ratios are outlined in Table 2.7.

Table 2.7: Mitochondrial and rDNA copy numbers calculated from Table 2.6.

	Gal:Mito	**Gal:rDNA**
Glucose	12.19236	141.7203274
Glycerol Lactate	15.848761	123.8604954
Galactose	20.624799	137.7584541

Ratios were calculated by dividing the percentage of unique *GAL1* sequences against the percentage aligned for the rDNA or Mito regions.

Chapter 3 Genome Conformation Capture (GCC)

3.1 Introduction

Interphase chromosomes tend to exist in territories in higher eukaryotes, out of which DNA can loop to inter-mingle and interact with DNA from other chromosomes [66]. *S. cerevisiae* chromosomes are not thought to form classical territories and have been hypothesised to form a Rabl-like conformation [3, 76]. Nevertheless, the *S. cerevisiae* chromosomes have regions of the nucleus they prefer, as reproducible DNA-DNA interactions between different regions of the genome form [70, 74, 88, 115, 186], and gene territories have been shown to be present by live cell imaging [187].

DNA-DNA interactions are thought to form as part of general genome structure [88], or as a result of nuclear processes such as transcription [4, 40, 103], and replication [4]. Individual interactions can be identified with existing methodologies such as FISH [97], 3D-FISH [21], and proximity based ligation techniques (3C [70], 4C [21], 5C [20]) which allow the detection of inter- and intra- chromosomal interactions. FISH techniques are only capable of analysing a limited number of loci simultaneously while 3C and 4C rely upon *a priori* knowledge of interacting sequences, or require the selection of a bait fragment [21, 137]. Furthermore, and the 3C technique is only able to identify one interaction at a time. 5Cs, on the other hand, is approaching a global technique, but would require tens of thousands of primers to be designed and produced, making it unviable as a genome wide interaction detection method. A technique which is able to globally identify inter- and intra- chromosomal interactions without previous knowledge or suspicion about which two regions of the genome might be connecting with each other was lacking in the field of genome and chromatin structure.

Here I describe a method, Genome Conformation Capture (GCC) [5], which does not require *a priori* knowledge of interacting fragments. GCC is a logical extension of previous proximity based ligation methods [21, 70, 90, 138] and utilises high-throughput sequencing of 3C libraries to effectively capture ‘all’ the interactions that

occur in a given cell population. Thus, GCC informs upon the spatial organisation of the genome.

In this chapter GCC is performed on *S. cerevisiae* cells as a proof-of-principle and is established as a valid method for the capture and identification of DNA-DNA interactions. The high connectivity of the yeast genome is demonstrated, which incorporates nuclear episomes and mitochondrial DNA. GCC was performed on an exponential phase population of cells grown on glucose and 1059 novel non-adjacent inter- and intra- chromosomal interactions were identified between the sixteen *S. cerevisiae* nuclear chromosomes. The selfish genetic element, the 2-micron plasmid, was found to interact with the nuclear chromosomes and strikingly, interactions were captured between the mitochondrial genome and the nuclear chromosomes. The GCC method was validated with conventional 3C PCR, quantitative 3C PCR, statistical simulations, and comparisons of the data with previous observations. Essential and non-essential genes are shown to behave differently to each other. Intra- and self- 2 micron plasmid interactions are shown to fold the plasmid into a structure which facilitates its maintenance and copy number. Finally, mitochondrial DNA sequences are shown to have a preference to interact with gene homologues of the mitochondrial ancestor *Paracoccus denitrificans*.

3.2 Results

3.2.1 The Saccharomyces cerevisiae genome

GCC was performed on the haploid yeast *S. cerevisiae* (2.2.4.1), which has a total DNA content of 18,534,409 bp. This consists of 16 nuclear chromosomes, ranging in length from 230,208 bp (Chr I) to 2,879,975 bp (Chr XII), including the 200 rDNA repeats [10]. There is also a nuclear located 2-micron plasmid (6318 bp), at a copy number of 60 [174]. The mitochondrial organelle also has its own genome, which is dynamically maintained at an average copy number of 50 [11], and is 85,779 bp in length [149]. The *S. cerevisiae* genome, including episomes and controls, contains of a total of 14,101 *MspI* fragments (including control sequences) labelled 0 to 14,101. *MspI* restriction fragments are referred to as segments or fragments. Segment 0 is the first *MspI* fragment of chromosome I, the chromosomes are concatenated such that the numbering of the segments is sequential from chromosome to chromosome. However, segments do not cross chromosome boundaries.

3.2.2 Minimum length of DNA accurately positioned upon the genome

During the sample preparation pUC19 was added prior to phenol:chloroform extraction, but crucially, after the ligation step. As a result of this, no ligation events are expected to be identified between the host genome and the pUC19 plasmid. *In silico* simulations were performed to establish the minimum length of a DNA sequence which could be uniquely positioned upon the reference genome (Appendix CD). To establish at what length sequences no longer aligned decreasing lengths (10-20 bp) of *S. cerevisiae* DNA sequence were aligned against the *S. cerevisiae* genome (Figure 3.1); 15 bp was established as the shortest DNA length which the greatest number of sequences could be aligned uniquely. Once the GCC experiment had been carried out it was found that a sequence length of 13 bp resulted in only one interaction between the *S. cerevisiae* genome and the pUC19 plasmid. This interaction involved a pUC19 sequence which could be aligned uniquely against the *S. cerevisiae* genome, thus, it was discarded. Therefore 13 to 23 bp either side of the restriction site of the 36 bp sequence was deemed sufficient for positioning fidelity.

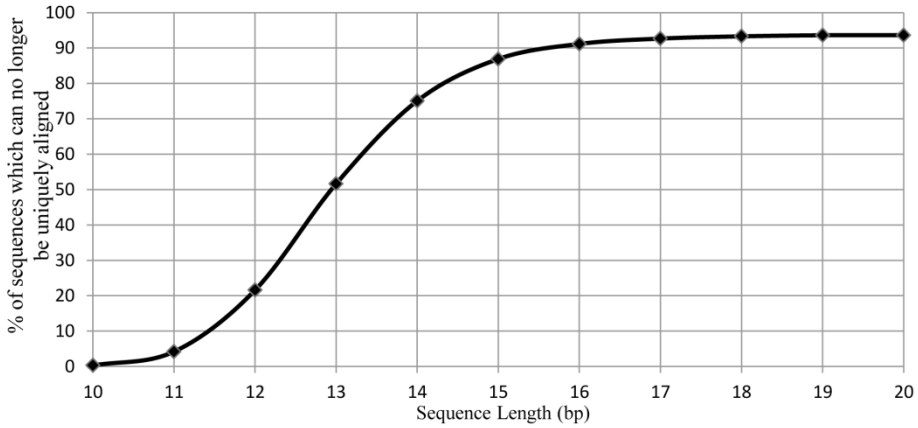

Figure 3.1: Greater than 85% of 15 bp *S. cerevisiae* DNA sequences can no longer be aligned uniquely against the *S. cerevisiae* genome.
In silico simulations were performed whereby decreasing lengths (10-20 bp) of *S. cerevisiae* DNA sequence were aligned against the *S. cerevisiae* genome to establish the length at which the maximum unique alignments take place. The simulated short *S. cerevisiae* sequences were aligned against the *S. cerevisiae* genome and the number of sequences which could be uniquely positioned upon the genome has been plotted as a percentage of the total number of generated sequences.

3.2.3 Choosing an appropriate restriction enzyme

GCC is an extension of the 3C method and therefore requires a restriction enzyme be chosen to cleave the genome into restriction fragments which are subsequently able to be ligated. Table 3.1 outlines the desirable attributes of a restriction enzyme for GCC analyses. The chosen restriction enzyme needed to be active at 37°C and have the ability to be heat inactivated at 65°C. Preferably, the restriction enzyme should not be sensitive to DNA methylation to ensure cleavage across the entire genome, however, this is not an issue in *S. cerevisiae* which lacks DNA methylation [188]. Restriction enzymes which cut the genome into very small fragments (*i.e.* <13bp due to the ability to align uniquely against the *S. cerevisiae* genome) are not suitable as the short sequence length may inhibit ligation around large protein complexes holding two restriction fragments together. On the other hand, while GCC maps can very well be created with infrequent cutting enzymes (resulting in large restriction fragments); the purpose of identifying interactions is to do it on as fine a scale as possible. This is not achieved with infrequent cutting restriction enzymes.

A range of restriction enzyme cleavage maps for the *S. cerevisiae* genome were generated using the topography program (Figure 3.2). The sizes of the digested fragments were plotted as histograms in 500 bp bins. Restriction enzymes that recognise six and eight base pair recognition sites were ruled out as they did not generate a fine-scale interaction map of the *S. cerevisiae* genome. For example, a 6 bp recognition

restriction enzyme only produces between 1448 and 4468 fragments for the *S. cerevisia*e genome with a mean fragment length of 7,232 bp for *BamHI* and just 48 fragments for a restriction enzyme with an 8 bp recognition site with a mean fragment length of 253,459 bp for *AscI* (data not shown). *BfaI*, which has a 4 bp recognition site (Figure 3.2A), was ruled out as it cleaved the genome into 26181 fragments, with 599 fragments below the acceptable 13 bp cut-off determined as the minimum length which can be uniquely aligned against the *S. cerevisiae* genome (3.2.2). Additionally, the inactivation temperature was 80°C, a temperature which promotes reverse crosslinking ultimately leading to a reduced integrity of the complexes holding the two fragments together. *MspI* was chosen as the most suitable restriction enzyme for an even cleavage across the entire *S. cerevisiae* genome. *MspI* resulted in a total of 14,088 fragments, had the least number of fragments below 13bp (382) of all enzymes investigated, and had fragment lengths between 0 and 11,875 bp. Furthermore, *MspI* results in a 5' overhang, and has an activity temperature of 37°C and inactivation temperature of 65°C[1], making it ideal as an enzyme for GCC analysis.

Table 3.1: Ideal restriction enzyme attributes for GCC analysis.

Enzyme attributes	
Activity temperature	37°C
Inactivation temperature	65°C
Overhang	5' or 3'
Minimum fragment length	>13bp
Maximum fragment length	> shortest chromosome

[1] The inactivation temperature (65°C) for the restriction enzyme *MspI* has been revised by the manufacturers twice during the course of this PhD research. The temperature was initially shifted up to 80°C and it was later stated that *MspI* could not be inactivated at all, depending upon the manufacturer in question. Regardless, the restriction enzyme activity is stopped with the addition of SDS. This coupled with the fact that the ligation is performed in a diluted solution, which would probably render the *MspI* enzyme inactive.

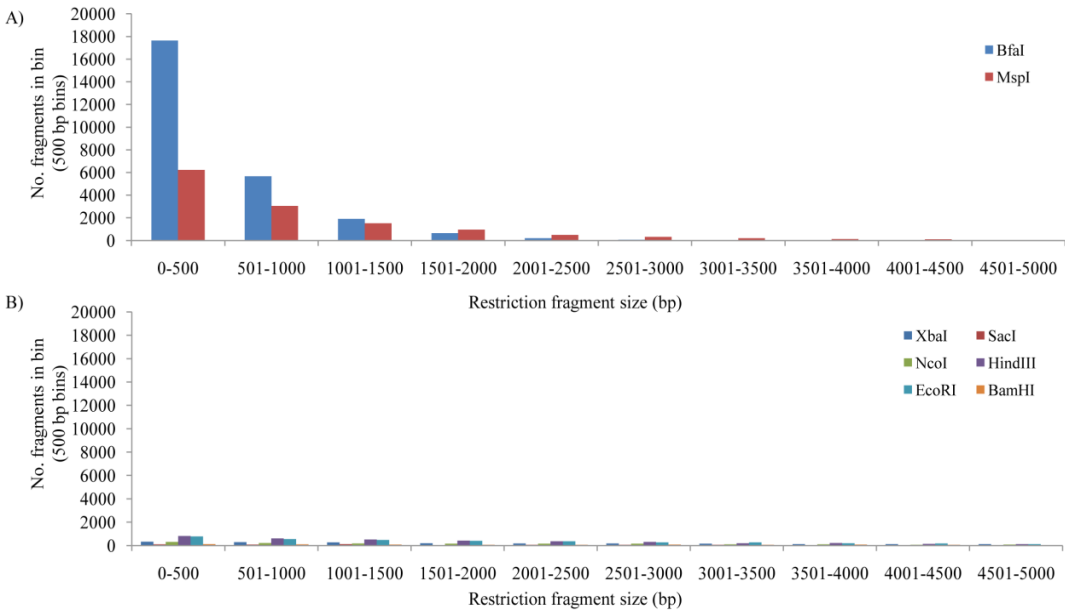

Figure 3.2: The distribution of fragment lengths upon digestion of the *S. cerevisiae* genome with restriction enzymes with 4 or 6 bp recognition sequences
Restriction enzyme cleavage maps of the *S. cerevisiae* genome were generated with the Topography program. The lengths of the restriction fragments have been plotted as histograms. Restriction enzymes with a A) 4 bp, B) 6 bp recognition sequence have been plotted as histograms in 500 bp bins.

3.2.4 GCC library purification

The GCC samples for Solexa sequencing were required to be at a concentration of 100 ngµl^{-1} in a final volume of 50 µl in ddH_2O. To ensure efficient recovery of input DNA a co-precipitating DNA carrier would be need to be used during the DNA purification/precipitation step following phenol:chloroform extraction. I tested linear polyacrylamide (LPA) and tRNAs as carrier molecules, as well as omitting the carrier molecule and running the sample through a column, to establish which was the most effective at purifying the DNA. LPA is an ideal carrier molecule as it does not hamper downstream enzyme activity [189]. However, LPA is a large inert molecule which distorts Nanodrop DNA concentration calculations used for DNA quality control prior to sequencing (Table 3.2). For this reason, it is unsuitable for sample preparation for Solexa sequencing. As a consequence, an alternate DNA carrier molecule was required. LPA (0.25% final v/v) assisted purification was compared against tRNA carrier molecules and a column purification of the DNA that omitted carrier molecules altogether.

tRNA molecules (10 µgµl^{-1}) were incorporated into the preparation of three separate *S. cerevisiae* (BY4741) 3C samples, instead of LPA, and the 3C method carried out as normal. One tRNA containing sample was subjected to RNase treatment following

DNA pellet resuspension. A second was purified through four Zymo Clean and Concentrator columns and eluted in ddH_2O (40 µl total volume) and a third was left unaltered following DNA pellet resuspension. A further 3C sample did not have any carrier molecule added and instead the 2 ml sample was run through four Zymo Clean and Concentrator Columns, each eluted with 10 µl ddH_2O and pooled.

A Nanodrop was used to determine the DNA concentration of the different samples (Table 3.2). As expected, LPA results in a spurious DNA concentration reading from the Nanodrop as well as 260/280 and 260/230 ratios. tRNA molecules result in similar concentrations as LPA assisted purification, even following RNase treatment, and thus are unsuitable as a carrier molecule for GCC sample preparation. Samples which were purified through Zymo Clean and Concentrator columns resulted in 260/280 ratios expected for DNA in a solution with very little protein [179] and more conservative concentration outputs, in line with the expected amount of DNA recovery, albeit slightly higher. Since moving straight to Zymo column purification and omitting the carrier molecule altogether resulted in an output similar to that of the Zymo purified tRNA sample it was decided the GCC samples would be generated using just Zymo column purification.

Table 3.2: Comparison of purification methods for GCC sample preparation.

Sample	Concentration (ngµl^{-1})	260/280 Ratio	260/230 Ratio
LPA	3642.3	5.43	4.31
tRNA	5361.5	1.15	1.14
tRNA + RNase	5360.9	1.12	1.10
tRNA + Zymo	33.3	2.42	2.66
Zymo	55.5	2.03	1.88

The expected concentration for 100% DNA recovery from input cells was ~45.43 ngµl^{-1}. Red indicates a DNA concentration which is too high, whereas green indicates a DNA concentration of an acceptable level.

3.2.5 Linear vs. circular DNA purification bias

Proximity based ligation methods theoretically result in both linear and circular DNA molecules [21, 190]. Therefore, I wanted to determine whether there was any purification bias for linear vs. circular DNA molecules during the preparation of the GCC samples, especially following the incorporation of a column purification step (Figure 3.3). pUC19 plasmid linearised with *EcoRI* and *HindIII* restriction enzymes was mixed with an equal amount of uncut pUC19 plasmid and added to GCC samples prior to phenol:chloroform extraction. Completed GCC libraries were assayed for the presence of cut and uncut plasmids by qPCR using primers that amplified across an

EcoRI site present on the pUC19 plasmid (3cpUC19ecoR1fwd and 3cpUC19ecoR1rev; pUC19 359-559 bp) and primers which amplified within the *EcoRI* fragment (3cpUC19fwd and 3cpUC19rev; pUC19 1487 - 1701 bp), that is they did not cross the *EcoRI* site.

The primers which amplified within the *EcoRI* restriction sites (NonRE) allowed the determination of how much pUC19 was present in the samples. The results obtained from the primers which amplified across the *EcoRI* site (OverRE) allowed the calculation of how much of the total plasmid was cut (linear), by calculating a ratio between OverRE and NonRE. These results were compared against qPCR results of the original input pUC19 cut and uncut sample to determine whether circular or linear DNA molecules had been lost during the purification step. A ratio of the amount of linear vs circular pUC19 plasmids present in the purified samples was calculated by dividing the OverRE by NonRE. A ratio of 1 was expected for the OverRE and NonRE primers when assaying uncut pUC19 because the qPCR should result in equal amounts of product for each sample. We would expect a lower ratio for cut pUC19 if the restriction enzyme had achieved cleavage, while the NonRE primers would continue to amplify from the total pool of pUC19 as it is not hindered by the *EcoRI* cleavage. The input cut and uncut pUC19 resulted in expected OverRE:NonRE ratios of 0.0101 and 0.893, respectively (Table 3.3). The ratios obtained from the input mixed cut and uncut pUC19 and the GCC sample after purification were similar (0.0912 and 0.157, respectively) albeit slightly skewed towards a higher amount of linear DNA meaning that there had been a small loss of circular DNA molecules during the purification steps, this loss was considered negligible.

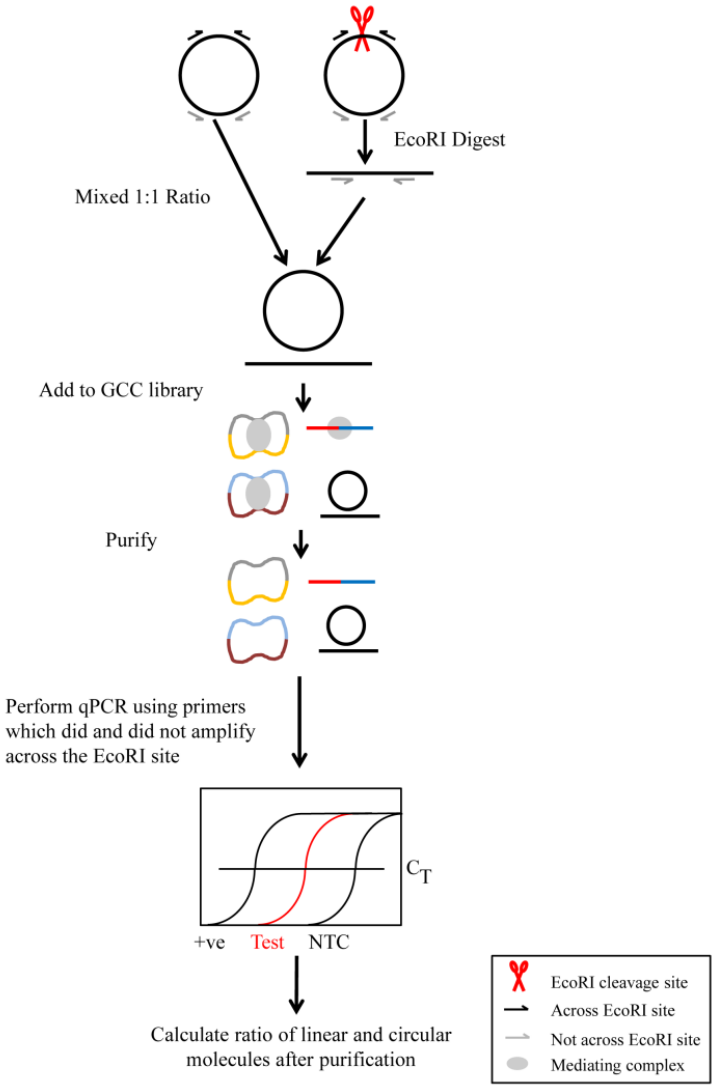

Figure 3.3: Experimental design to determine whether there is a bias for linear of circular DNA molecules.
EcoRI cut and uncut pUC19 plasmid were added to GCC libraries in an equimolar ratio before the sample was purified [5]. Primers (Appendix I) were used to perform sybr green qPCR across the cur *EcoRI* site and also in a region not containing the *EcoRI* site, and the ratios calculated and compared. pUC19 plasmid was isolated as stated previously (2.2.2.10).

Table 3.3: Linear vs. circular DNA purification bias.

Sample	**Ratio (OverRE primers:NonRE primers)**
EcoRI cut pUC19	0.0101
Uncut pUC19	0.893
Input mixed cut and uncut pUC19	0.0912 ± 0.003
GCC sample with cut and uncut pUC19 after purification	0.157 ± 0.03

Ratios are displayed as means ± standard deviation. n=3.

3.2.6 Genome Conformation Capture (GCC)

S. cerevisiae cells were recovered from -80°C glycerol stocks on SC glucose agar plates (30°C, 48 hrs) before a single colony was used to inoculate a starter culture consisting of 10 ml of SC glucose media (30°C, 24 hrs, 160 rpm). 15 µl of the starter culture was used to inoculate 200 ml of SC glucose media before culturing (30°C, 16 hrs, 160 rpm) to an OD_{600} of 0.600.

The GCC protocol was performed as stated (Materials and Methods 2.2.3.7) with the exception that only eight chromatin samples (residuals of 52.6 µl, all from the same replicate chromatin preparation), were prepared and pooled and no external ligation controls were added prior to the ligation step. The GCC library was eluted in 80 µl mQ H_2O before being vacuum concentrated (20 mins, 45°C, Eppendorf) to 60 µl. The final concentration as determined by Nanodrop was 101.1 $ng\mu l^{-1}$ with 260/280 and 260/230 ratios of 2.06 and 2.04, respectively. The GCC library (5 µg) was sent for 36 bp single-end Solexa sequencing at the Allan Wilson Centre Genome Service, Massey University. The sequence was returned with a total of 8.4 million 36 bp sequence tags (or reads).

3.2.1 Sequence Population Statistics

I wanted to test if the sequenced GCC libraries obtained from the Illumina genome analyser were representative of the *S. cerevisiae* genome. Theoretically the Illumina Genome Analyser could introduce sequencing bias for certain *S. cerevisiae* genomic features. Certain genomic elements occur at different copy numbers within the genome (*e.g.* ribosomal DNA repeats, telomeres *etc.*). From published reports of copy number [10, 150, 185] we calculated how many base pairs the whole genome complement is and therefore what percentage specific genomic elements should comprise (Table 3.4). The whole yeast genome (18,563,326 bp) was calculated to contain 16 nuclear chromosomes at a copy number of 1, 200 rDNA repeats (9.1kb each), 50 mitochondrial genomes (85,779 bp each) and 50 2 micron plasmids (6,318 bp each). Sequence files were aligned against the *S. cerevisiae* genome using the SOAP algorithm [8] to determine the total number of sequences which could be aligned with zero mismatches. The raw sequences files were then aligned against different genomic features (ribosomal DNA repeats, reference genes, and telomeres; see Appendix CD). The number of aligned sequences for each category was then compared to the expected number of aligned sequences (calculated by the genomic features base pair composition of the genome) as percentages (Table 3.4). The reference genes set (housekeeping) consisted

of 646 verified ORFs across all chromosomes that were between 400-500 amino acids in size and was obtained by batch download from the SGD website (www.yeastgenome.org). Telomeric (TEL) and rDNA (ribosomal DNA) sequences were also obtained from the SGD website.

Table 3.4: Sequence population statistics

	% of Sequences	(5,737,327)[a]	% of Genome	(13)[b]
Ribosomal DNA	14	(796,683)	14	(1.82)[c]
Telomeric DNA	1.4	(79,824)	1.01	(0.142)[d]
Reference Genes	6.9	(398,463)	6.94	(0.897757)[e]

[a] The subset of sequences which could be positioned onto the reference genome without any mismatches.
[b] The size of the *S. cerevisiae* genome in Mb, including 200 rDNA sequences but excluding the mitochondrial genome and the 2 micron plasmid.
[c] Size of 200 rDNA repeats in Mb, each repeat is 9.1kb in length.
[d] The amount of DNA in Mb the telomeres comprise of the genome.
[e] Reference genes were obtained by batch download from the SGD website. The set consists of 646 verified ORFs, across all chromosomes that were between 400 and 500 amino acids in length.

3.2.2 GCC network construction

The GCC network for exponential glucose grown *S. cerevisiae* cells was assembled from a total of 8,434,102 sequences of 36 bp in length. Sequences were deposited in Gene Expression Omnibus under accession number (GSE13648). Sequences were processed using the topography program (Topography v1.19 [5]), coupled with the SOAP algorithm [8]. The topography program divided the sequences according to whether they did or did not contain an *MspI* site and whether the sequences could be uniquely positioned upon the reference genome or whether they aligned to repetitive elements (Table 3.5). Sequences which did align against the reference genome were assigned to the particular restriction fragment (segment) which they aligned within. 160,169 sequences (1.9%) contained an *MspI* restriction site flanked by restriction fragments of ≥13 bp, the minimum sequence length which resulted in no pUC19-genome interactions (2.2.3.7). No mismatches were allowed during the SOAP algorithm alignment to ensure positioning fidelity.

Table 3.5: Solexa sequence statistics.

	No. sequences	%[a]	%[b]
Solexa output	8,434,102		
With *MspI* site	373,386	4.43	100.00
RE frags ≥13 bp	160,169	1.9	42.90
≥ 13 bp Unique	108,072	1.3	28.94
≥ 13 bp Repeated	52,097	0.6	13.95

[a] Percentage of the original Solexa output.
[b] Percentage of the sequences which contained an *MspI* site flanked by restriction fragments with a length ≥ 13 bp.
With *MspI* site: refers to sequences which contain an *MspI* site (C|CGG) within the 36 bp sequence. RE frags ≥13 bp (Restriction Fragments ≥13 bp) are those 36 bp sequences which contain an *MspI* site and have greater than 13 bp of sequence on each side of the *MspI* site. ≥13 bp Unique refers to sequences from the "RE frags ≥13 bp" category which can be uniquely positioned on the reference genome with no mismatches. ≥13 bp Repeated refers to sequences from the 'RE frags ≥13 bp' category which mapped to repeated genomic elements within the reference genome, with no mismatches.

3.2.3 GCC Statistics

The Topography program output lists all interactions between restriction fragments which were captured by GCC. Before analysis could begin interactions which occur above experimental noise needed to be determined [2]. Initially, random pairing simulations were performed using the statistical software R to establish that the GCC datasets were non-random. Secondly, we used a binomial distribution to calculate at what frequency an individual pairing had to achieve before it was deemed above experimental noise.

3.2.3.1 Statistical analysis of nuclear segments

The null hypothesis is that some segments may not interact, but among those that do the pairings are made at random.

When the segments are identified using 13 bp sequences, there are 12680 distinct interacting segments, occurring in equal amounts. 38,064 pairings were observed in our actual data-set. To test our null hypothesis, we simulated 1000 sets of random pairings of 12,680 segments and took the number of times the most frequent pairing occurs as our test statistic. For all of the replicate simulations, the most frequent pairing occurred twice. In the real data set, the most frequent pairing occurred 199 times and 5376

[2] Statistical analyses were carried out by Dr Beatrix Jones (Centre for Mathematical Biology, Massey University).

pairing occurred 3 or more times. The real data clearly do not follow the null model ($p< 0.001$, based on the most frequent pairing statistic).

It is then of interest to identify which pairs occur more frequently than expected under the null hypothesis; this may indicate that two segments have a particular affinity for each other. Under the random pairing model, the number of times a pairing of two particular distinct segments is observed is a binomial random variable with $p= 2/(12680^2)$ and n=38064. The probability of observing a particular pairing 2 or more times is then $1.1x10^{-7}$. While this number is small, there are 12680*12679/2 pairings under consideration, so we expect (on average) to observe approximately 9.2 pairings that occur two or more times during a particular experiment. In 1000 sets of simulated experiments, the number of pairings that occur 2 or more times ranges from 2 to 18. However, in the actual data we observe 8839 such pairings. Thus declaring all pairings observed 2 or more times as 'non-random' results in an expected false positive rate of 9.2/8839= 0.001; *i.e.* the vast majority of these pairings are due to something other than random pairing of segments.

For segments that self-pair, the number of times a pair is observed is binomial ($1/12680^2$, 38064). The probability a specific segment is observed self-pairing 2 or more times is $2.8 x10^{-8}$. The expected number of false positives using 2 as the threshold for non-random self-pairing is 0.0004. In the 1000 sets of random pairings, no self pairing was observed more than once. In the real data, 74 pairings were observed 2 or more times (maximum was 10 times), so the expected false positive rate is $4.8 x10^{-6}$.

3.2.3.2 Analysis of mitochondrial (repeated) segments

A similar null model for interactions between nuclear and mitochondrial segments can be used. However, the segments now occur with a particular number of copies, so the segments do not have an equal probability of forming pairings; those with more copies are more likely. Again, the test statistic is the maximum count for any pairing.

There are 64 mitochondrial segments and 170 nuclear segments that interact with them; a total of 254 interactions are observed. For these smaller numbers of segments 100,000 simulations can be performed. The maximum count for an observed pairing was 7; only 337 of 100,000 simulations achieved this value. So again, the null hypothesis can be rejected with p = approx. 0.003.

Individual pairings were then considered. The p-value P, for each pairing, was computed (as many or more occurrences of this pairing observed | null hypothesis) and 1/10,710 was used as the significance threshold (that is, the number of mitochondrial segments minus 1, multiplied by the number of nuclear segments). 10,710 tests were computed, so the expected number of false positives is 1. In fact, 39 pairings exceed this threshold; thus a false positive rate of 1/39 = 0.025.

Similar statistics are calculated to determine the probability of the other interactions that included repeated elements.

3.2.4 The GCC network

The *S. cerevisiae* genome is highly connected by statistically significant inter- and intra-chromosomal interactions (Figure 3.4A). 1059 novel inter- and non-adjacent intra-chromosomal interactions occur at levels elevated above random (expected false positive rate = 0.001 in the glucose grown *S. cerevisiae* cells). The most striking observation is that the mitochondrial genome participates in DNA-DNA interactions with the nuclear genome. There are four lipid bi-layer membranes separating the mitochondria and nuclear genomes, therefore, theoretically they should not form physical interactions [57, 191]. The nuclear chromosomes are contained within the nucleus, therefore it is expected that they would form many interactions with each other. Interestingly, the nuclear chromosomes appear to have regions which interact more frequently than others, suggesting that there are interaction faces where two chromosomes ‘fit’ together (Figure 3.4B). While the chromosomes appeared to have chromosome faces, the number of inter-chromosomal interactions it participated in still correlated with its length, in bp (Figure 3.4C).

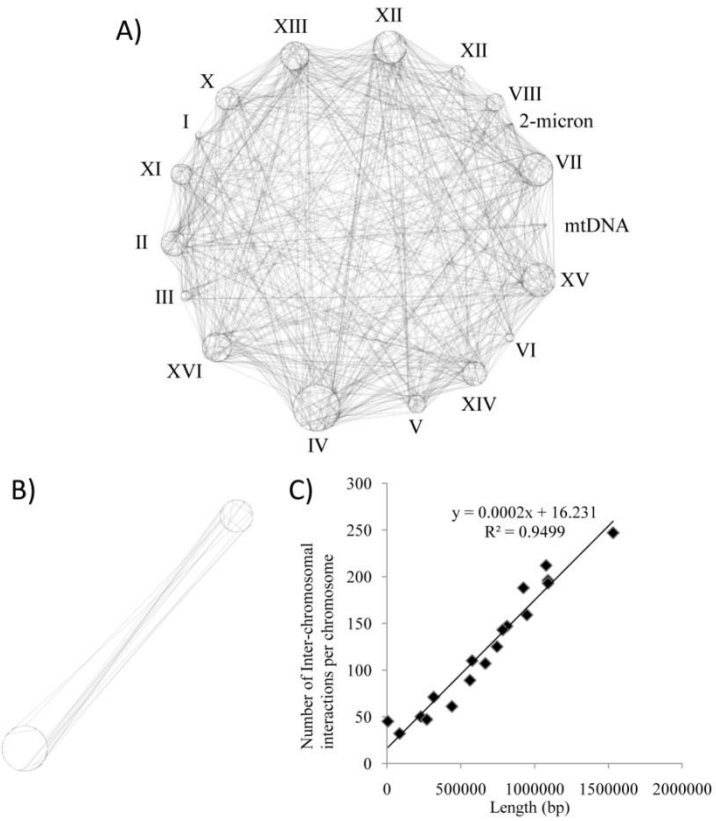

Figure 3.4: The GCC network is constructed from inter- and intra- chromosomal interactions.
A) The GCC network map for exponentially growing yeast cells [5]**.** Chromosomes are depicted as circles around the outside of the network; they have been artificially circularised by positioning the telomeres adjacent to each other. The size of the circle is proportional to the chromosome length. Chromosome XII contains only two rDNA repeats. The black lines indicate interactions between the loci they connect. B) Inter-chromosomal interactions between specific chromosomes occur in clusters, indicating the presence of interaction faces between chromosomes (chromosomes IV and VII are shown to illustrate this). C) The number of inter-chromosomal interactions each chromosome participated in was summed and plotted against the chromosome length, in bp. This illustrated a correlation between the number of inter-chromosomal interactions and chromosome length.

3.2.5 Validation of the GCC method

To confirm the reproducibility of the GCC method and results, six pseudo-replicates each consisting of 10.9 million sequences were randomly sampled from two samples of 23 million sequences. Pseudo-replicate sequences were derived from a subsequent GCC analysis of *S. cerevisiae*, however the growth history of the culture differed slightly (see sections 4.2.1 and 4.2.2). Significant adjacent, non-adjacent intra-chromosomal and inter-chromosomal interactions were highly conserved between six pseudo-replicates (percentage of interactions shared between a pair of replicates ranging from 96.3% to 98.7%, 77.5% to 84.2%, and 67.0% to 77.5% respectively). In contrast, comparisons between the pseudo-replicates and the glucose experimental dataset demonstrated that the adjacent interactions were highly similar (range 81.75–91.04%), but that the non-

adjacent intra-chromosomal (range 9.04–19.2%) and inter-chromosomal interactions (range 0–2.04%) were dependent upon the culture history and growth environment.

To further validate the GCC datasets I compared the interactions with previously observed genomic spatial organisation. Firstly, FISH studies [148, 192, 193] clearly show an association between tRNA genes and the ribosomal DNA (rDNA). By plotting which segments interact with segments containing a tRNA our data confirmed interactions between the tRNA and rDNA repeats (Figure 3.5A). These interactions were not clustered within one rDNA region but rather interacted with different regions of the rDNA repeat (Figure 3.5B).

Secondly, the 32 *S. cerevisiae* telomeres have been demonstrated to cluster in yeast cells [194-196]; unfortunately, telomeric clustering cannot be confirmed in the GCC map, as the reads which mapped to telomeric regions were not long enough to be positioned to specific telomeres. Thus, the telomere group can only be analysed as group and not on an individual basis and therefore cannot be used to confirm telomeric clustering.

Figure 3.5A also demonstrates a high number of interactions between the tRNAs on the mitochondrial genome and the nuclear rDNA. The mitochondrial genome is a high copy number element and has a disproportionate number of tRNAs according to its length which might explain these interaction peaks.

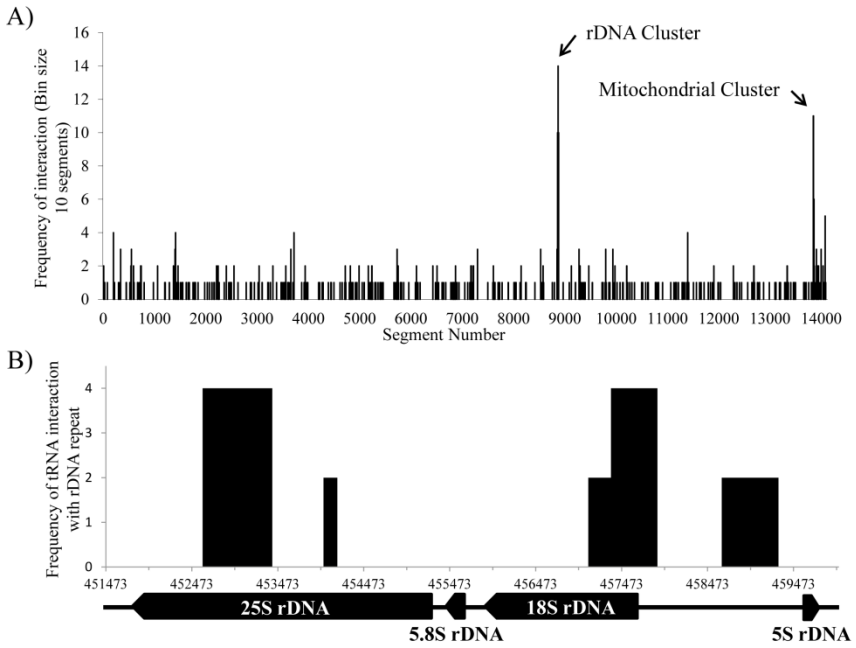

Figure 3.5: Segments containing tRNA genes cluster with the ribosomal DNA.
A) Interactions between tRNA containing segments and their partners were mapped and a histogram plotted according to the segment (bins of 10 segments) with which they interact. That is, the *S. cerevisiae* genome is represented as restriction fragments linearised across the bottom of the graph. Major clusters occur at the rDNA repeats and within the mitochondrial genome. B) The tRNA interaction pattern within the rDNA, limited to statistically significant interactions (*i.e.* allowing for rDNA copy number (statistical methods)). The data plotted in A) and B) include data from repetitive interactions.

3.2.6 Experimental validation of GCC

I established, experimentally, that GCC isolates non-random interactions. One interaction identified by GCC between two genomic loci (gDNA-gDNA; Segment 5313; Chr VII [868673-873666 bp] – Segment 6434; Chr XI [172555-173311 bp]) was chosen to perform classical 3C PCR [70]. *S. cerevisiae* cells were grown in glucose and harvested for 3C sample preparation using the *MspI* restriction enzyme (2.2.3.2). 3C PCR was performed (for primers see Appendix I) to confirm that segments 5313 and 6434 do interact (Figure 3.6). I reasoned that if this same interaction changed depending upon the growth condition then it would add validity to the fact this was a true DNA-DNA interaction. That is, if the interaction was a result of random ligation, then theoretically, no change in interaction frequency would be expected. Quantitative 3Cs was performed between segments 5313 and 6434 in cells grown on glucose, glycerol lactate and galactose containing media (Figure 3.7). The frequency of this interaction reduces during growth on glycerol lactate and galactose when compared to glucose (set at 100%).

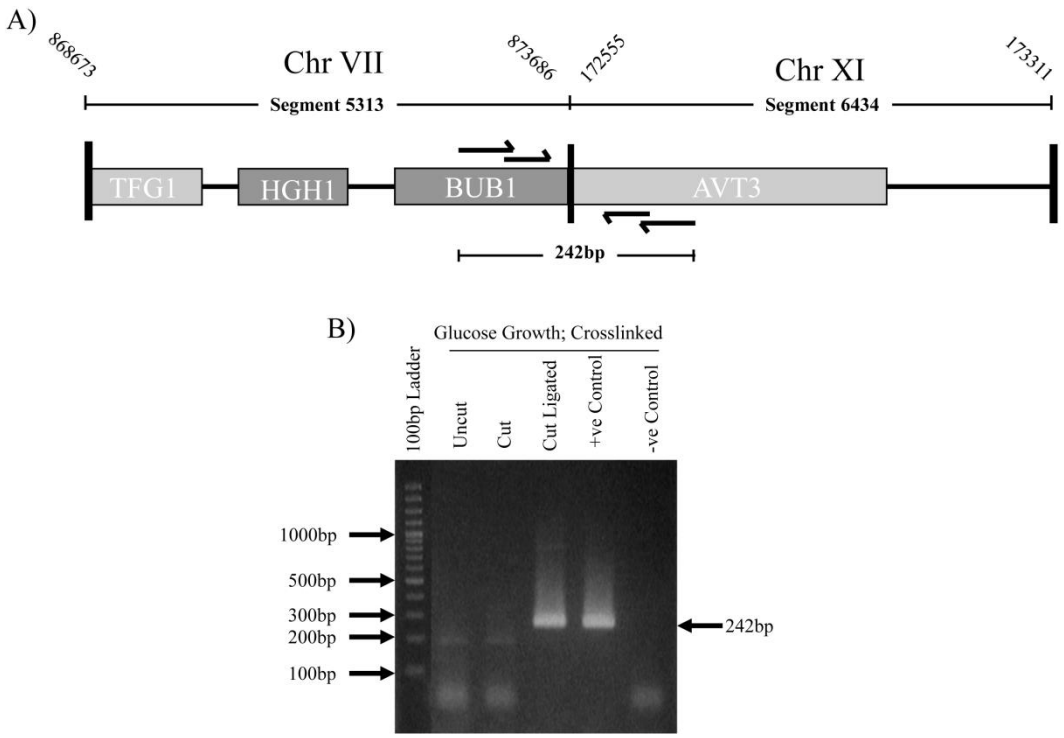

Figure 3.6: Confirmation of an interaction between two nuclear segments originally identified by GCC. A) Chromosomal context of *MspI* segments 5313 and 6434, depicted as the possible interaction ligation product. 3C primers are shown as arrows to amplify across the ligation junction which results in a 242bp product. B) A two round nested 3C PCR results in a product for the 3C cut ligated sample, but no product in the uncut or cut, which are controls. The positive control is *MspI* digested and randomly ligated *S. cerevisiae* BY4741 genomic DNA, encompassing all possible 3C ligation products. -ve control is a no template PCR control.

I went one step further and used quantitative 3C to determine if a subset of interactions occurred at equal frequency in yeast grown in different environmental conditions [197]. Each one of the three interactions was chosen to represent a different class of connection. An interaction between two nuclear chromosomes (gDNA-gDNA; segments 5313-6434) mentioned earlier was chosen as was an interaction between the mitochondrial genome and the nuclear genome (Mito-gDNA) and the 2 micron plasmid with the nuclear genome (Plas-gDNA). These representative interactions identified from the GCC network were confirmed by quantitative 3C (Figure 3.7). These studies demonstrate that the frequencies of the three interactions alter, independently from each other, in the different carbon sources (*i.e.* glucose, glycerol-lactate, or galactose; Figure 3.7).

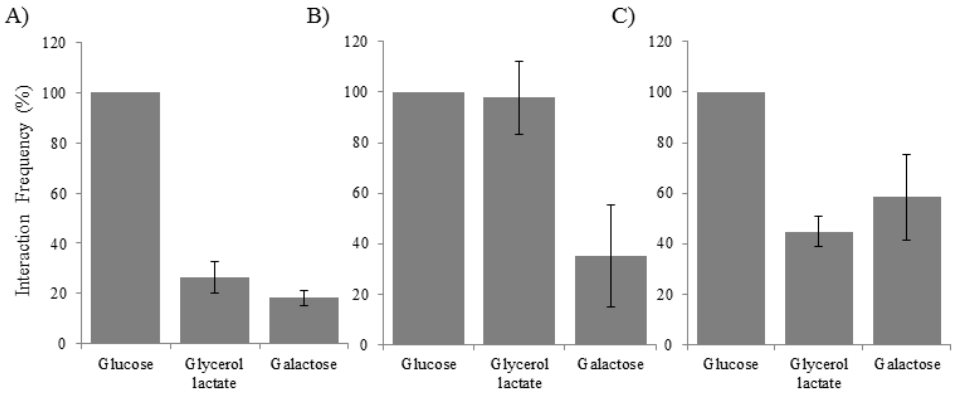

Figure 3.7: Quantitative 3C analyses of three interactions in *S. cerevisiae* BY4741 cells grown on different carbon sources (glucose, glycerol lactate, or galactose) demonstrates that interaction frequencies are independently environmentally dependent.
A) gDNA–gDNA, interaction between loci located on Chromosomes VII (bp 868,673–873,686) and IX (bp 172,565– 173,311). B) Mito-gDNA, interaction between loci located on the mitochondrial genome (bp 24,872–26,193) and Chromosome XVI (bp 365,496–365,760). C) Plas-gDNA, interaction between loci located on the 2 µm plasmid (bp 4586–5008) and Chromosome XI (bp 240,262–240,547) [5]. The 3C interaction was quantified by real-time PCR (methods) using BHQ^{TM} and BHQplus^{TM} probes (Appendix I). Results were standardised according to the number of genomes present in each sample (Appendix II) and are presented as the interaction frequency compared to the glucose sample (100%). Results are the means of three replicates ± StdErr.

Critically, two segments which did not interact in the GCC network also did not interact by conventional 3Cs (Segment 10330; Chr XIII [650,483-654,199 bp]-Segment 6897; Chr X [144,850-146,958 bp]). 3C PCR was performed between segments 10330 and 6897 which did not yield any product (Figure 3.8). Thus, I conclude that GCC identifies true DNA-DNA interactions and is a valid global method for the search of inter- and intra- chromosomal interactions.

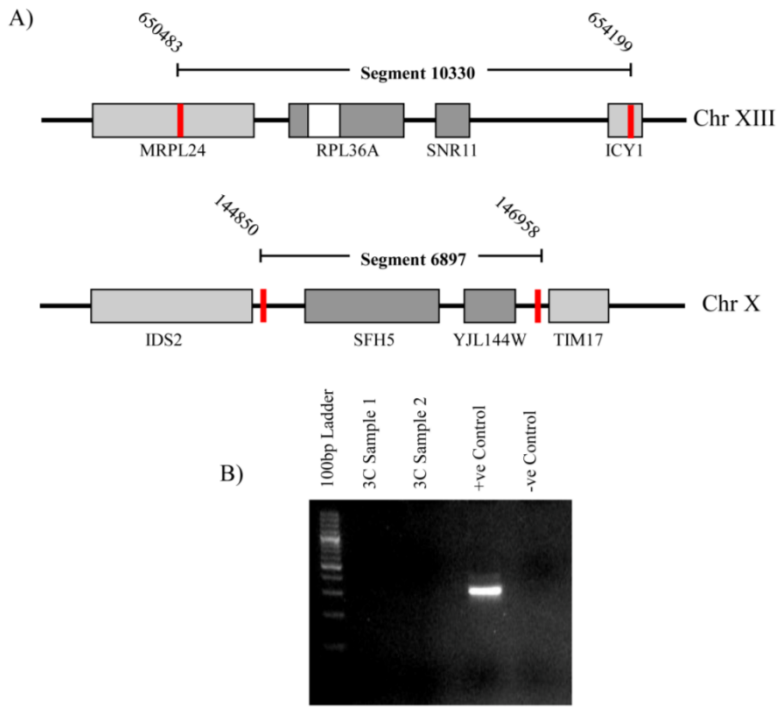

Figure 3.8: Randomly chosen segments, which do not interact in the GCC network, do not interact by 3Cs.
A) Cartoon depicting the chromosomal context of the chosen segments (Segment 10330 Chr XIII [650483-654199] – Segment 6897 Chr X [144850-146958]). Black lines intersecting the DNA strand denote *MspI* sites. B) Positive controls confirmed the efficacy of the PCR reactions however no amplification was observed in any test samples. Four independent 3C samples were screened by PCR. PCR primer concentration and temperature gradients were also performed. Under no conditions did any sample yield a PCR product corresponding to the 10330-6897 interaction; however, other interactions tested positive in the same samples. The positive control was constructed by PCR amplification of each segment from outside of the *MspI* restriction sites; this product was then digested with *MspI* (20 U, 37°C, 2 hrs), the products mixed, and ligated with T4 Ligase (16°C, 16 hrs).

3.2.7 Yeast nuclear chromosomes are structured

The *S. cerevisiae* nucleus has been shown to not contain classical chromosome territories [196, 198], and therefore the GCC interactions between and within chromosomes were searched for individual structure and territorial structure. While the number of inter-chromosomal interactions correlates with chromosome length (Figure 3.4C), those interactions are not evenly spread across the length of the chromosome and thus the GCC network provides evidence for inter-chromosomal interaction faces (Figure 3.4B). Chromosomal pairings were observed which suggests some form of nuclear arrangement for the chromosomes.

A high number of intra-chromosomal loops might be indicative of chromosome territories, or smaller, highly concentrated loops in certain regions could suggest they are forming as part of DNA compaction. Intra-chromosomal loops were investigated as to whether they formed within individual chromosomes. Intra-chromosomal loops form

on average 28 ± 14 (mean ± stdev; range 6 loops on Chr I – 64 loops on Chr IV) times per chromosome. This analysis excludes intra-chromosomal loops within the rDNA locus on chromosome XII as it is not possible to establish whether adjacent segments interact or whether the adjacent interactions were as a result of interactions between different rDNA repeats. The number of intra-chromosomal loops correlates (r^2=0.8594) with chromosome size, in agreement with the observation of a linear relationship between chromosome length and the number of significant unique interactions (Figure 3.4C). These loops are variously arranged as large loops encompassing, and adjacent to, smaller loops and linker DNA sections. I conclude that chromosomal interactions within the nucleus of exponentially growing *S. cerevisiae* cells provide evidence for chromosomal organization and individual structure, but not necessarily chromosome territories.

3.2.8 Adjacent and non-adjacent intra-chromosomal interactions report on chromatin compaction

Non-adjacent intra-chromosomal interactions form as a result of DNA looping within a chromosome. GCC captured and identified 519 statistically significant (FPR = 0.001) non-adjacent intra-chromosomal interactions which ranged in size from ~100 bp to a 1,433,367 bp. The size of the intra-chromosomal loops was plotted as a histogram (500 bp bins; Figure 3.9A) and it was found that the histogram peaked below 5000 bp. I hypothesise that these relatively small loops are forming as part of the normal compaction of DNA into chromatin fibres and thus may report on regions of the genome which are in a heterochromatic state.

Adjacent interactions can also report on local chromatin structure. These interactions form between neighbouring restriction fragments (Figure 1.5A) and theoretically can form in different ways: 1) Sequencing across an uncut *MspI* restriction site will result in an adjacent interaction. Restriction digestion is not 100% efficient across chromatic regions, thus, uncut restriction sites can indicate that restriction sites were shielded, perhaps by protein complexes or by chromatin compaction or nucleosome occupancy of the recognition site [199]. 2) An adjacent interaction can form when two neighbouring restriction fragments are folded into a loop and held together, and while the restriction site is able to be cleaved, the two fragments are re-joined during the ligation step.

I investigated the galactose (*GAL*) locus (Chr II; encompassing *GAL7*, *GAL10*, and *GAL1*), which, during glucose growth, is a repressed region of the genome (reviewed in [200]). Adjacent interactions were mapped across the *GAL* locus to determine whether there was enrichment of adjacent interactions across this locus (Figure 3.9B). The promoter regions of the divergent *GAL1* and *GAL10* genes had high levels of adjacent interactions across them (Figure 3.9B). Interestingly, the *GAL1* and *GAL10* promoter region is the site Gal4p binds to in order to activate transcription of the *GAL1* and *GAL10* genes upon the need to metabolise galactose as a carbon source [201]. Gal4p appears to achieve this activation by remodelling the nucleosomes around this region [50], indicating that this region is compacted during glucose growth. This nucleosome occupancy might explain the high number of adjacent interactions observed across this region. Conversely, Gal4p binds in the promoter region of *GAL7* also, yet no significant adjacent interactions were observed over this region.

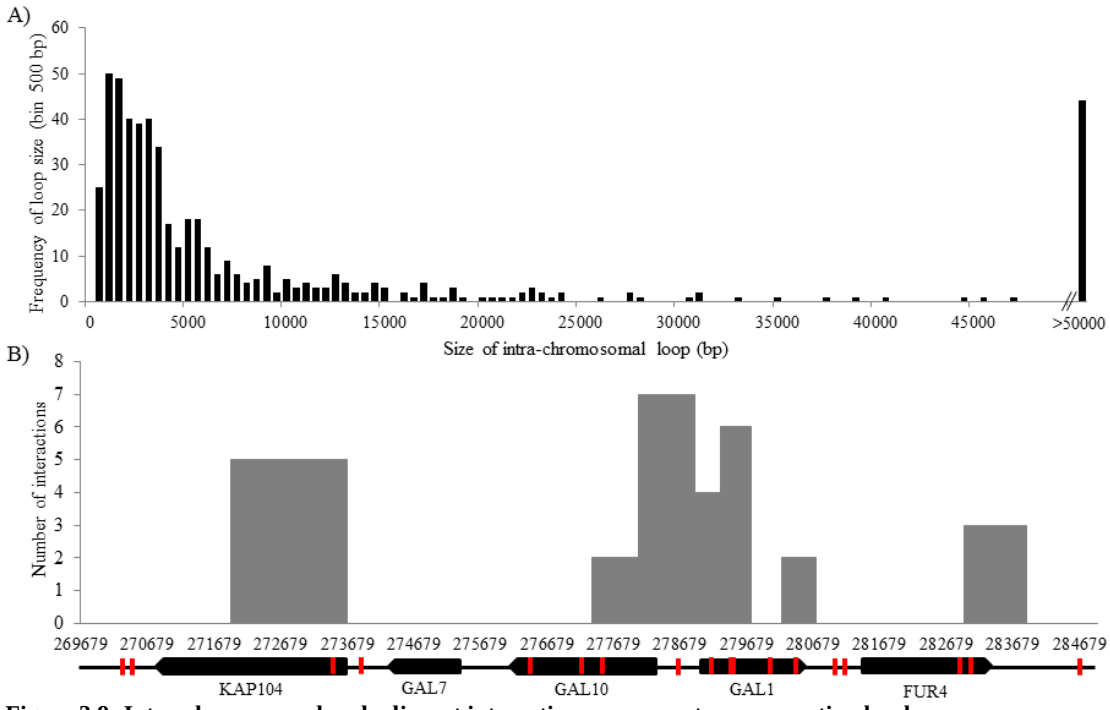

Figure 3.9: Intra-chromosomal and adjacent interactions can report on compaction levels.
A) The majority of the intra-chromosomal loops fall between 1,000 and 5,000 bp in size, sitting well with previous observations [3] for the existence of chromatin fibres [17]. Intra-chromosomal loop size is plotted in bins of 500 bp from significant (FPR = 0.001) intra-chromosomal interactions against the number of loops which fall within each bin. The mean loop size was 32,638 bp with the largest loop being 1,433,367 bp. B) Adjacent interactions can indicate an inability of *MspI* sites to be cut, or suggest the two fragments are being held together. The adjacent interactions are plotted across the GAL locus (Chr 2: 269,679-284,679 bp; genes depicted across the bottom). The red vertical lines indicate *MspI* restriction sites, thus the adjacent interaction frequency is plotted over the *MspI* site.

3.2.9 The 2 micron plasmid is also structured in three dimensions

S. cerevisiae harbours a 2-micron plasmid, which confers no particular advantage or burden on the host [202] and as a result has been described as a selfish DNA element [175]. The 2 micron plasmid is relatively small (6318bp) and encodes only four open reading frames (ORFs). However, it is maintained at 50-100 copies per cell using partitioning and amplification systems [203]. The partitioning mechanism requires that the 2 micron plasmids cluster about a stability locus (STB) [204, 205]. Therefore, I predicted (Figure 3.10A) an uneven distribution of self-ligation products about the STB locus in plasmids isolated as part of the GCC library. The distribution of the self-ligated products does not correlate with the length of the restriction fragment (Figure 3.10B). Rather, the distribution confirms the spatial clustering of identical sequences, which appear as self-ligations within the GCC network due to inter-repeat ligation (Figure 3.10B). Intriguingly, an artist's impression of the plasmid structure reconstructed from interaction data, that was used to fold a scale rubber hose (Figure 3.10C), indicates that the plasmid folds to present the STB locus on a prominent loop, perhaps facilitating plasmid clustering about this locus as previously postulated [175].

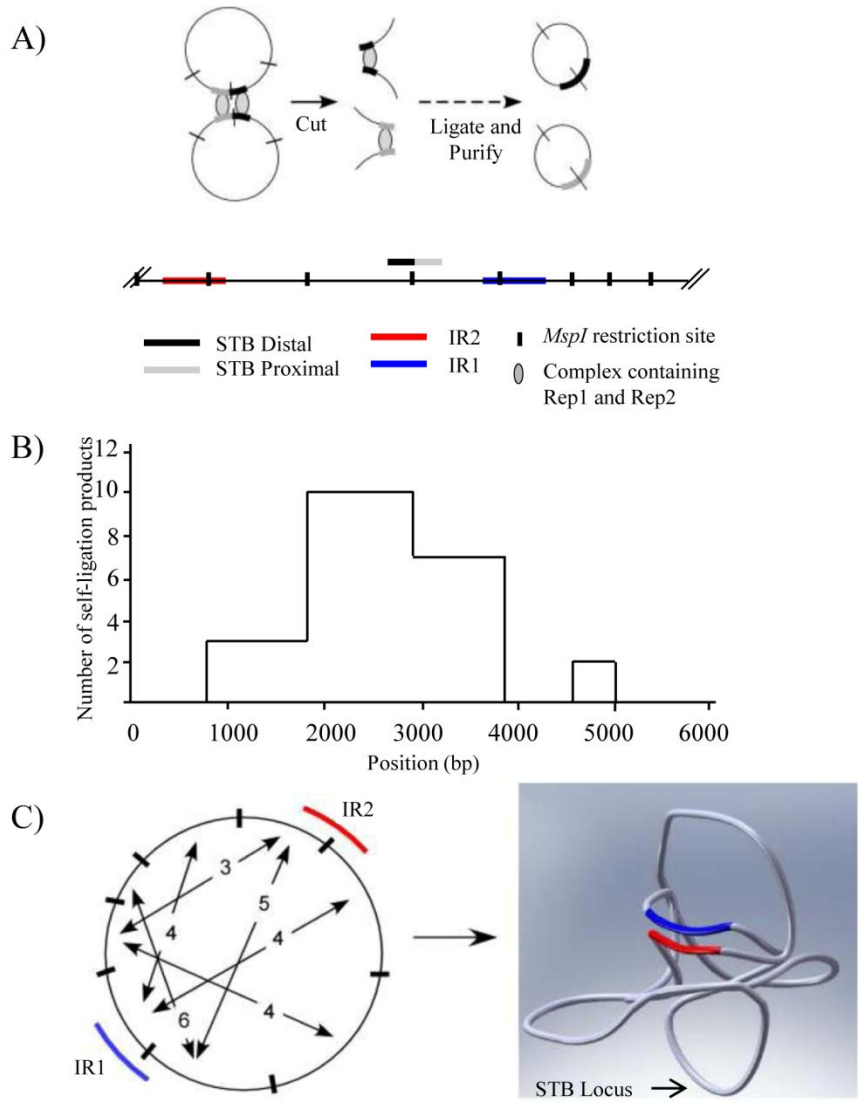

Figure 3.10: The yeast 2 µm plasmid folds to maximize the interactions between the inverted repeats, while the stability locus (STB) locus is on a prominent loop [5].
A) Cartoon illustrating how Rep and STB mediated clustering of the 2 µm plasmid results in self-ligated fragments due to inter-repeat ligation. The GCC method is as shown in Fig. 1. B) The distribution of self-ligated (circularized) restriction fragments indicates functional clustering of the 2 µm plasmid about the STB locus and the inverted repeats. Interaction frequencies were obtained from the combined data file due to the repetitive nature of the inverted repeats. Non-circularizing fragments with similar sizes to the three circularizing fragments were present (1032 bp [3] vs. 1100 bp [10], 950 bp [7] vs. 927 bp [0], and 720 bp [2] vs. 748 bp [0]; number of observed self-ligations in square brackets), within the 2 µm plasmid indicating that size is not the sole criteria for self-ligation. The positions of the circularizing fragments are shown relative to a cartoon of the 2 µm sequence, above the graph. C) Artists rendition of the 2 µm plasmid copied from a physical model which integrated the non-adjacent intra- 2 µm plasmid interactions. Intra-molecular interactions confirm the co-localization of the inverted repeats (IR1 and IR2) and demonstrate organization of the STB locus onto a prominent loop. Restriction fragments are delineated by dividing lines. Interactions are denoted by arrows annotated with the number of isolations of the interaction sequencing product.

3.2.10 Interactions occur between the mitochondrial and nuclear chromosomes

Transfer of mitochondrial genomic DNA to the nuclear genome is frequent and on-going [154, 161, 206]. Therefore, finding elevated levels of interactions between the mitochondrial genome and 39 nuclear chromosome restriction fragments (FPR = 0.025)

containing parts of 51 nuclear ORFs was not unexpected. 34% of these chromosomal ORFs had homologues with the mitochondrial ancestor, *Paracoccus denitrificans*, linking the mitochondrial genome with ORFs that were ancestrally of mitochondrial origin (e.g. *MAK5* and *BNA4*)[3]. The mitochondrial genome does not interact evenly with the nuclear genome (Figure 3.11); there is an obvious preference for particular parts of the mitochondrial genome. Mitochondrial DNA was also identified as interacting with the nuclear rDNA at high frequency.

It could be argued that GCC was capturing nuclear genome inserted mitochondrial sequences. However, this was improbable as firstly, only two interactions mapped to regions that are characterized as having mitochondrial sequences inserted within the genome (NUMTs) [154] but neither NUMT matched the interacting mitochondrial restriction fragment. Secondly, in this analysis an interaction detected by GCC must take place over an *MspI* restriction site, and while mitochondrial insertions could be integrated into the nuclear genome adjacent to *MspI* sites, at the frequency interactions were captured between the nuclear and mitochondrial genomes, by GCC, insertions were unlikely to account them.

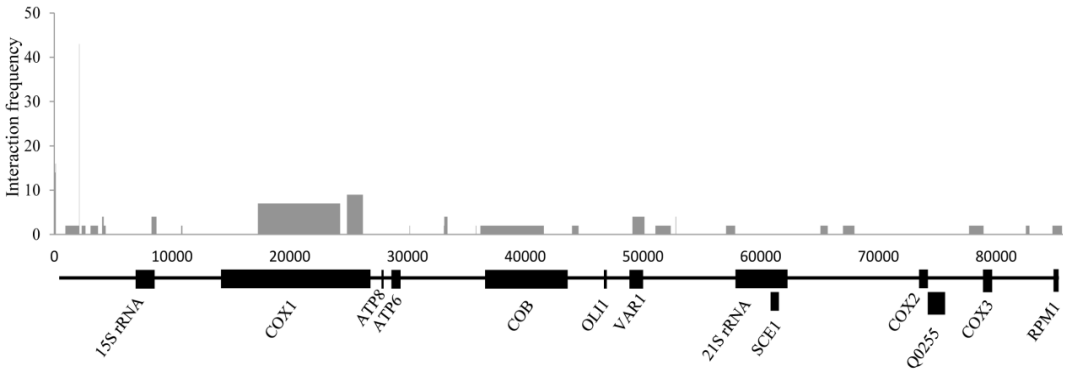

Figure 3.11: Interactions are not evenly spread across the mitochondrial genome [5].
This analysis displays which fragments of the mitochondrial genome interact significantly with the nuclear genome. The mitochondrial genome has been linearised and depicted along the bottom of the graph. Statistically significant (FPR = 0.025) interactions between the mitochondrial and nuclear genomes have been plotted as interaction frequencies for each mitochondrial *MspI* fragment. Interactions are derived from the datcomb file which contains repetitive elements; this analysis is only concerned with the numbers of interactions between the two genomes not their identity, thus their inability to be uniquely aligned upon the nuclear genome is not a concern.

[3] PhD supervisor, Dr Justin O'Sullivan (Institute of Natural Sciences, Massey University) performed analysis of mitochondrial ancestry homologues.

3.2.11 Essential and non-essential genes interact differently

Essential open reading frames (ORFs) are those coding regions which are required for normal laboratory growth. Inter- and intra-chromosomal interactions were assumed to occur with equal frequency at essential and non-essential ORFs. Loops formed within essential and non-essential ORFs at similar frequencies (1% and 1.5%, respectively; Table 3.6). In contrast, essential genes were significantly (t-test $p < 0.01$)[4] less connected to non-adjacent loci than their non-essential counterparts (Table 3.6). A similar difference was observed between intron-containing verses non-interrupted ORFs (t-test $p < 0.01$).

Table 3.6: Essential and non-essential genes are connected to non-adjacent loci at different frequencies.

	ORF Pairing		
% interacting with:	NE-NE	E-E	E-NE
Same ORF	1.50	1.01	N/A
Non-adjacent intra chromosomal locus	3.89	0.27	0.91
Non-adjacent extra chromosomal locus	5.01	0.27	1.46
Inter-genic region	1.66	1.65	N/A

The percentage of non-essential (NE) or essential (E) genes that were interacting: within the same ORF; with non-adjacent ORFs; or with inter-genic regions. Significant interactions within the unique 13-23 data-set were categorized according to whether both partners were essential or non-essential (annotations obtained from the Saccharomyces Genome Database). Percentages were calculated based on the total number of NE and E ORFs, respectively. For the E-NE pairing, data are expressed as percentage of the E genes.

[4] PhD supervisor Dr Justin O'Sullivan performed statistical tests upon the essential and non essential gene analysis.

3.3 Discussion

GCC produces a network map of the genomic interactions occurring within a defined population of cells without *a priori* constraints on the regions of interest. GCC has been used to demonstrate that the yeast genome is a highly integrated interaction network composed of nuclear and mitochondrial chromosomes as well as a nuclear located episome. I have shown, through the use of statistical and experimental methods, that GCC is a valid technique for the global capture and identification of inter- and intra-chromosomal interactions.

The GCC methodology is a 'first step' in the detection and identification of DNA-DNA interactions on a global scale, and as such, future advances in sequencing technology (*i.e.* longer reads and increasing capacities) will allow interaction network maps of larger genomes (e.g. Human) to be created. An enrichment step for sequences containing the chosen restriction site would be necessary to achieve this with current sequencing technology. Following the publication of GCC and the *S. cerevisiae* interaction network [5] two subsequent methodologies, Hi-C and ChIA-PET, were developed that incorporated interaction enrichment steps [90, 144].

The topography software is also geared towards future advances in sequencing technology with the SOAP algorithm [8] capable of positioning longer sequences onto the reference genome. Excitingly, since the publication of GCC the Illumina Genome Analyser has been upgraded to produce paired-end sequencing. Paired-end sequencing increases the useable sequences, enabling the identification of interactions on the basis that the paired sequences are derived from different genomic restriction fragments. In other words, the detection of interactions will no longer require that the restriction site is present within the 36 bp sequence tag, with an appropriate length of sequence either side to align against the reference genome. As the cost of sequencing reduces, GCC interaction networks could be created for the same cell population using different restriction enzymes. This would allow interaction 'points' to be more finely mapped.

Here I used GCC to report on specific features of the genomic interaction network in *S. cerevisiae* grown on glucose. The most striking discovery was that mitochondrial DNA (mtDNA) interacts physically with the nuclear genome. There is no direct evidence to indicate whether the interactions between the mitochondrial and nuclear genomes observed involve complete or fragmented mitochondrial genomes. However,

characterized organelle insertions can be large [159, 207, 208]. It has been proposed that transfer is due to the turn-over or degradation of organelles ([reviewed in] [159]) or regulated by host encoded proteins [166]. Alternatively, nuclear import of organelle genomes may result from a combination of these pathways.

It should be noted that mitochondrial escape has been shown to occur at a higher rate in fermenting yeast cells, where mitochondria only make up 3% of the total volume of the cell [163]. This is in agreement with our observations that the frequency of the mitochondrial-genomic interaction (*i.e. COX1-MSY1:* mitochondrial genome [24872 – 26193 bp] and chromosome XVI [365496-365760 bp]) was highest in fermenting yeast cells (Figure 3.7). Therefore, it is possible that the interactions between the mitochondrial and nuclear chromosomes are regulatory although further work in this area will need to be performed to validate this claim. Such a regulatory role may explain why loss of mitochondrial DNA causes alterations in a diverse range of nuclear genes [209] and similarly why NUMTs are not evenly distributed within the nuclear genomes in which they are found (*e.g.* human [161] and yeast [154]). Further work will be carried out to elucidate the functional role these mitochondrial DNA sequences fulfil once inside the nucleus. Interestingly, the mitochondrial genome copy number affects the rate of replication of the nuclear rDNA repeats [167] and thus, yeast growth rate.

The 2 micron plasmid, a nuclear located selfish genetic element, contains two inverted repeats (IR1 and IR2) which recombine as part of a copy number maintenance mechanism [210]. GCC clearly identified spatial clustering of these repeat sequences (Figure 3.10C), which are predicted to facilitate recombination.

Partitioning of the 2 micron plasmid from mother to daughter cells requires that the clustered plasmids interact with the yeast chromosomes. This process is thought to be analogous to hitchhiking, whereby the 2-micron plasmid ‘hitches’ a ride into the daughter cell upon the chromosomes thus ensuring its segregation during call division [205]. The GCC network clearly identified interactions between the 2 micron plasmid and yeast chromosomes (Figure 3.4A). However, the interacting chromosomal regions did not share any obvious similarities, as determined by gene ontology (data not shown). It remains to be seen whether these interaction sites are dynamic or have fixed positions along the chromosomes. GCC interaction networks created from cells growing

in different conditions may elucidate this question. I propose that these chromosomal loci act as the hitch-hiking sites for 2 micron plasmid segregation.

Previous observations indicate that chromosomal interactions, and hence the structure of the genomic network, are controlled by epigenetic mechanisms [88, 97, 211, 212]. GCC captures and identifies DNA-DNA interactions but does not report upon the nature of the connection, that is, whether it is a regulatory or structural, or some other form of interaction. In order to gain a perspective as to what roles the interactions might play they were analysed based on whether they involved essential or non-essential genes. ORFs are designated as essential based on the function of the encoded product, rather than an inherent property of the DNA sequence. Therefore, the reasons for the observed low connectivity of the essential genes are not immediately obvious. It has been proposed that the essential genes are located in open areas of the nucleus (existing as euchromatin) where they can be easily accessed by DNA regulatory factors and the transcription machinery, and as a result do not connect highly with other regions of the genome, namely their non-essential gene counterparts [5]. This brings us back to non-membrane bound sub-compartmentalisation of the nucleus; however, it is currently unclear by what mechanism the genome achieves these structures. Techniques such as GCC may provide insights into these types of spatial segregation mechanisms.

The function of the intra-chromosomal loops (Figure 3.9A) is largely unknown, although functional roles have been attributed to them [20, 139, 186, 213]. The most well known is that of the β–globin locus control (LCR) region looping over some 20 kb and interacting with the β-globin gene cluster where it regulates and enhances transcription [27]. The β-globin genes are transcribed sequentially according to developmental stage and the LCR interacts with the appropriate gene at the appropriate point in development [106]. I observed many loops in excess of 20 kb within the yeast chromosomes, indicating that similar mechanisms might be at play in *S. cerevisiae*. Further experiments would be needed to elucidate the role, if any, of large loops. By taking into account what loci are involved in the large loops, one can design experimental growth conditions to test hypotheses about the nature of these connections by quantitative 3C analyses. Such experiments are paramount to resolving information inherent in the GCC datasets.

It is interesting to note that the intra-chromosomal loop sizes peaked below 5000 bp (Figure 3.9A). I hypothesise that these loops are different than those which form over large distances in that these interactions form as part of the nucleosome packing of DNA into fibres of varying widths depending upon which model (the one-start helix model [solenoid model] [18] or the two-start helix model [47]) is favoured ([reviewed in [3]). Having said that, our data fits well with both predictions, with interactions occurring over the length of 10-14 nucleosomes per fibre turn, allowing room for variable nucleosome linker regions [3].

The *GAL7*, *GAL1*, and *GAL10* genes are clustered together on chromosome 2, but are transcribed from three separate promoters and regulated directly and indirectly by the transcription factors Gal4p and Gal80p, respectively [200]. Gal4p binds the promoter region of these genes to activate transcription (1,000-fold) after initiation by galactose, but does not bind otherwise [200, 214]. Gal4p removes promoter localised nucleosomes in order to switch on transcription of the *GAL* cluster [50]. Digestion of DNA is inhibited by nucleosomes occupancy, and thus I hypothesised and observed (Figure 3.9B) a high number of adjacent interactions across the promoter regions of *GAL1* and *GAL10*. I predict that the GAL locus will have a lower frequency of adjacent interactions when it is in an actively transcribing state, such as when *S. cerevisiae* are grown in media with galactose as the carbon source. However, further experiments would need to be conducted to test these predictions further, namely a dataset in which the *GAL* locus was in an active state. In conclusion, GCC is not only able to report upon genome structure but also localised chromatin structure.

3.4 Conclusion

GCC was developed, performed and validated on the eukaryote *S. cerevisiae*. The emergence of next-generation sequencing has allowed the GCC method to be carried out, and with further advances in sequencing technology the interaction network of larger genomes could be mapped. If 'all' interactions which occur in an organism could be identified it would allow interaction microarrays to be constructed [88, 138] which would have the advantage that a gain or loss of separate interactions could be observed between environmental conditions.

GCC enables an integrated empirical approach to global investigations of the role that spatial organization has in genome function and evolution. The next challenge will be to determine how the genomic structure is assembled. These advances are crucial to improve our understanding of gene expression, genome maintenance and evolution, given the increasing evidence that the function(s) of the genome are determined by multiple layers of regulatory control processes [5, 42].

Nuclear chromosomes, the 2 micron plasmid and most strikingly the mitochondrial genome were all shown to be involved in DNA-DNA interactions. Structural features of the yeast genome have been reported, which fit well with earlier observations, including the clustering of the tRNAs. Based on the results, the 2 micron plasmid is predicted to fold into a conformation which is hypothesised to facilitate its maintenance. Intra-chromosomal interactions are predicted to form as a result of regulatory processes as well as part DNA compaction. In conclusion, the GCC technique is a step forward in the capture and identification of genomic structure and coupled with existing and future methods, will facilitate our understanding of nuclear processes and how these are linked to genomic structure.

Chapter 4 The Mitochondrial Genome Interacts with the Nuclear Genome

4.1 Introduction

Mitochondria have a central role within the metabolic systems of cells. In yeast, as in other organisms, the mitochondrial organelle is the site of oxidative phosphorylation and the citric acid cycle. As such, the yeast mitochondrial organelle is essential for both the fermentative and respiratory pathways that yeast use to metabolize different carbon sources. Unlike most other organisms, yeast can employ both respiration and fermentation simultaneously (*i.e.* respiro-fermentation) when grown on fermentable carbon sources such as glucose or galactose, at a concentration above 0.8 mM [172, 215].

Saccharomyces cerevisiae mitochondria contain an 85,779 bp genome, which encodes a subset of the electron transport chain components [149]. This genome is essential for respiratory growth on non-fermentable carbon sources like glycerol lactate [173], as the subset of electron transport chain components it encodes are required to generate a proton gradient across the inner mitochondrial membrane, ultimately driving ATP synthesis. In contrast, a functioning electron transport chain, and thus the mitochondrial genome, is non-essential when cells grow solely by fermentation.

Mechanisms are in place to co-ordinate and control the expression of the nuclear and mitochondrial genome-encoded genes required to maintain and control mitochondrial function according to cellular metabolism [168, 169]. As mentioned earlier, unstable mitochondrial plasmids have been observed to transfer into the yeast nucleus [162, 164] in a nuclear gene-dependent (*e.g. YME1, YME2*) manner [162, 165, 166]. The nuclear functions of these transferred mtDNAs are unknown; however, elevated mitochondrial to nuclear migration rates correlate with accelerated chronological aging in yeast [216].

I have previously observed that nucleic acids of mitochondrial origin interact with nuclear loci (hereinafter referred to as Mito-gDNA interactions) in *S. cerevisiae* [5]. The inter-organelle Mito-gDNA interactions formed a subset of the interactions that were identified; however, preliminary data indicated that they may respond to the metabolic status of the yeast cells, reflecting a possible role in mitochondrial-nuclear communication (Figure 3.7; [5]). GCC does not inform on the nature of the contacts or the identity of the complexes that may or may not maintain them [39, 70, 74, 106, 217]. However, inter- and intra- chromosomal interactions have been hypothetically and experimentally linked with transcription and transcriptional memory [40, 97]. Moreover, the quality and quantity of mitochondrial DNA has been shown to affect patterns of nuclear transcription [218, 219] and replication [167] in yeast.

Here I capture Mito-gDNA interactions occurring in *S. cerevisiae* during growth on glucose, galactose (*i.e.* respiro-fermentation) and glycerol lactate (*i.e.* solely respiration) using GCC. In this chapter I compared the Mito-gDNA interactions detected in these three metabolic conditions with each other and find that both the frequency and identities of inter-organelle interactions differ dramatically. Interactions between mitochondrial genes (*i.e. COX1* and *Q0182*, a dubious mitochondrial ORF) and nuclear encoded loci (*i.e. MSY1* and, *RSM7*, respectively), are shown to be dependent upon a functional electron transport chain and mitochondrial encoded reverse transcriptase machinery. Finally, the transcriptional rate of the nuclear encoded *MSY1* and *RSM7* genes is shown to increase when the interaction frequency is reduced by deleting mitochondrial reverse transcriptase activity. My results argue for a role for reverse transcription in inter-organelle DNA-mediated communication.

4.2 Results

4.2.1 Generation of comprehensive Mito-gDNA interaction maps reveals connectedness of the two organelles

Mito-gDNA interactions were previously captured in *S. cerevisiae* cells grown in glucose by GCC (Figure 3.4; [5]). One of these GCC captured Mito-gDNA interactions (*COX1* [Chr Mito: 24,872 – 26,193 bp] - *MSY1* [Chr XVI: 365,496 – 365,760 bp]) was isolated and shown to be carbon source dependent (Figure 3.7; [5]). As a consequence of the difference in the Mito-gDNA interaction frequency that we observed during growth in different carbon sources, I hypothesized that Mito-gDNA interactions would alter, on a global scale, according to the cells metabolic state. Thus, I generated comprehensive maps of the Mito-gDNA interactions in *S. cerevisiae* by GCC (2.2.3.3), during exponential growth in medium with glucose, galactose, or glycerol lactate as carbon source. Importantly, all the starter cultures were grown on glycerol lactate media, which was subsequently used to inoculate the three different conditions; this is in contrast to the earlier GCC chromatin preparation (3.2.6) which was prepared from a glucose starter culture.

The GCC networks were constructed from 36bp, paired end Illumina Genome Analyzer sequence reads from two biological replicates (total pooled reads; glucose 56,167,792, glycerol lactate 49,134,906, and galactose 48,419,385; for sequence file names see Appendix III). The two biological repeats for each condition were highly correlated for statistically significant interactions (R^2=0.78, 0.93, 0.93, respectively; Figure 4.1). Accordingly, sequences from biological repeats were combined and all further analysis was performed with sequence files containing the sequences from both repeats.

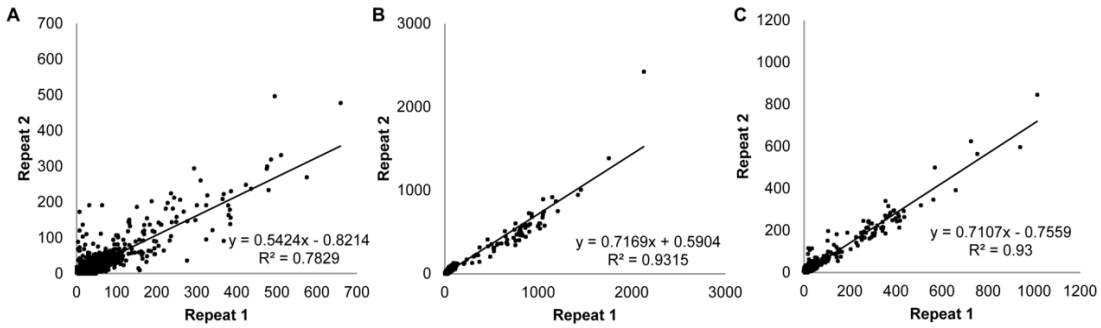

Figure 4.1: Biological repeats correlate well at the *MspI* restriction fragment level.
Two biological repeats were performed for each condition; a) glucose, b) glycerol lactate, and c) galactose. R^2 values are as follows; glucose 0.78, glycerol lactate 0.93, and galactose 0.93. Scatter plots were constructed from statistically significant (FPR = 0.012 – 0.027) interactions involving only *MspI* fragments which could be uniquely positioned on the reference genome. The frequency for identical interactions in different repeats are plotted against each other. Adjacent interactions have been omitted as they could not be distinguished from interactions which are the result of simply sequencing across an uncut *MspI* site. Circularized fragments (*i.e.* self-interactions) have also been omitted.

4.2.2 Sequence Population Statistics

Theoretically the Illumina Genome Analyser could introduce sequencing bias for certain *S. cerevisiae* genomic features. As was performed earlier (3.2.1), I calculated the percentage a number of different genomic elements comprises of the genome and compared it against the percentage each genomic element comprised of the sequence tags (Table 4.1). Sequence files were aligned against the *S. cerevisiae* genome using the SOAP algorithm [8], with zero mismatches. The reference genes set (housekeeping) consisted of 646 verified ORFs across all chromosomes that were between 400-500 amino acids in size and was obtained by batch download from the SGD website (www.yeastgenome.org). Telomeric (TEL), centromeric (CEN), rDNA (ribosomal DNA), and mitochondrial DNA sequences were also obtained from the SGD website.

Table 4.1: The glucose, glycerol lactate and galactose samples sequence files were representative of the genome.

	% of Genome (18,563,326)[a]	% of Glucose Sequences (56,167,792)[‡]	% of Glycerol Lactate Sequences (48,419,385)[‡]	% of Galactose Sequences (49,134,906)[‡]
CEN	0.010	0.004	0.005	0.004
TEL	0.758	1.319	1.379	1.345
rDNA 200	9.804	15.943	15.261	13.535
Housekeeping	4.755	6.702	6.800	7.015
YSCPLASM (50)	1.702	2.428	2.256	2.274
Mitochondria (50)	23.104	3.207	5.791	5.299

Chastity filtered sequence files for each condition (glucose, glycerol lactate and galactose) were aligned against the *S. cerevisiae* genome, using the alignment algorithm SOAP [8] to obtain the total number of sequences which could be aligned with no mismatches. The sequence files were then aligned against genomic features to ascertain whether the sequences were a good representation of the *S. cerevisiae* genome. Using published reports of copy number and the reference genome we calculated the likely percentage these features contribute to the total genome complement (% of genome). We then determined what percentage these features amounted to in the sequence files. Only the centromeric and mitochondrial sequences were lower than expected. [a] The size of the *S. cerevisiae* genome in base pairs including 200 rDNA repeats, 50 mitochondria, and 50 2 micron plasmids. [‡] The subset of sequences which could be positioned on the *S. cerevisiae* genome with no mismatches.

Table 4.1 shows that there is a discrepancy between the percentage of mtDNA that should be present if there was a 100% recovery of the mtDNA during the glucose GCC library preparation and what is actually present in the Illumina genome analyser output sequence files (23.104% compared to between just 3.207-5.791% for the three samples). I hypothesised that during the GCC chromatin preparation the mitochondrial organelle, and thus the mtDNA, were lost. Therefore, I assayed mtDNA content by qPCR at different steps throughout the GCC protocol. A total of seven samples were drawn at different steps during the GCC protocol, with the first being taken immediately after bead lysis, and the last following completion of the protocol. MtDNA content was assayed for each of these samples using primers (Mito+ve13909F and Mito(CNC)R13909; see Appendix I) which amplified within the mitochondria genome. The amount of mtDNA is displayed as a percentage of the first sample (set at 100%; Figure 4.2A), taken directly after cell breaking. A steady loss of mtDNA is observed throughout the GCC preparation such that sample 7 has lost ~50% of the starting mtDNA. Therefore, explaining, partly, the lower amount of mtDNA in the sequence files as compared to the normal genomic complement.

Primers (GAL1F and GAL1R) were also used to amplify within the single-copy nuclear encoded *GAL1* gene (see Appendix I) such that a ratio between nuclear and mitochondrial genomes could be calculated (Figure 4.2B). This was an important step to judge whether the nuclear genomic DNA was also being lost or if enrichment of nuclear to mitochondrial genome ratio was occurring. This analysis highlights the importance of

analysing results in different ways in order to gain a proper perspective; as in this analysis the first sample has only a 3.75:1 mtDNA:nuclear single copy loci ratio instead of the published mitochondrial copy number of ~50 per cell [150]. Therefore, this analysis demonstrates that the bulk of the mtDNA is lost prior to the first sample being taken, but that a further 50% reduction occurs to ultimately result in a 2:1 ratio of mitochondria and nuclear genomes. If the total base pair composition of the *S. cerevisiae* genome is artificially reduced to reflect the mitochondria and nuclear genome ratio of 2:1 the amount it comprises of the genome complement is 1.19%. This brings it more in line with the percentage of sequences we observed in the sequence files for mtDNA.

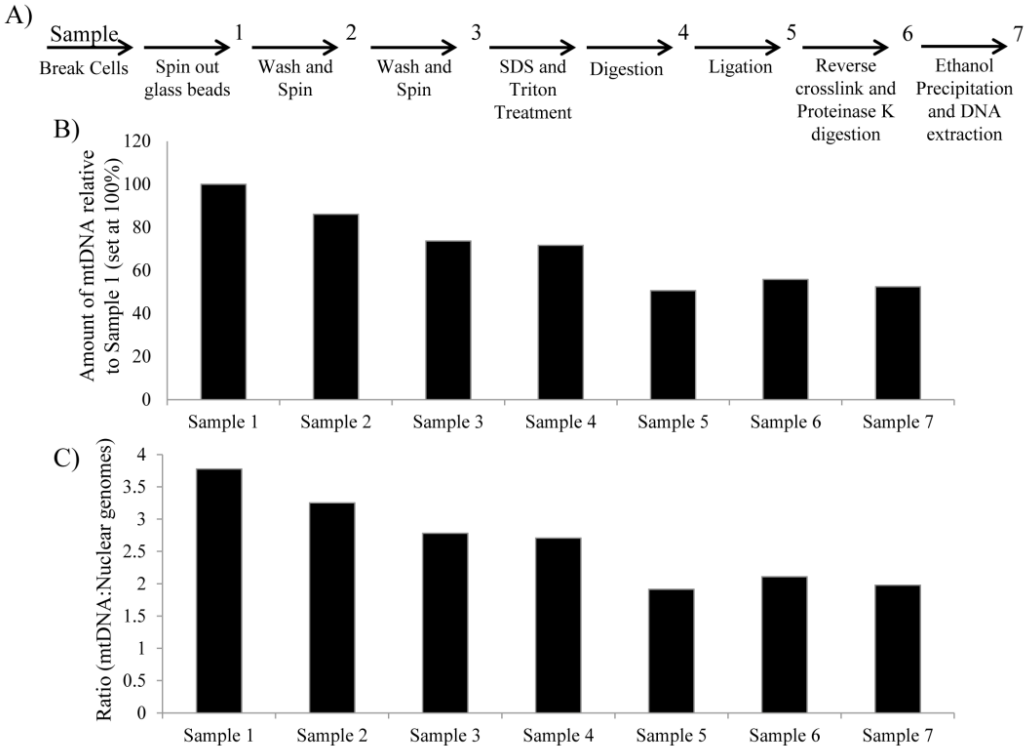

Figure 4.2: Mitochondrial DNA is lost consistently throughout the GCC preparation.
A) Displays the steps in the GCC protocol where a sample was drawn to test for mtDNA content. B) 50% of the mtDNA is lost during the GCC library preparation. Quantitative PCR with primers which amplified within intron aI5γ of the mitochondrial COX1 gene were used to determine mtDNA content at seven steps throughout the GCC library protocol. Values are expressed as a percentage of sample 1 which was taken following cell breakage, and which has been set at 100%. C) The ratio between mitochondrial genomes and nuclear genomes demonstrates that ~96% of mitochondrial genomes are lost during the cell harvesting steps. Quantitative PCR with primers which amplify within intron aI5γ of the mitochondrial *COX1* gene and the nuclear *GAL1* gene were used to determine mitochondrial and nuclear DNA content.

The apparent loss of mitochondrial DNA is occurring during the cell breakage step. A sample was unable to be taken at an earlier stage in the GCC protocol as the cells needed to be separated from the glass beads by centrifugation through a small hole,

before a sample could be taken (Figure 4.3). An alternative lysing method may allow the omission of this separation technique, and may result in a higher yield of mtDNA. Clearly not all mitochondrial organelles are lost. It appears that a portion of the mtDNA in the sequences files will be mtDNA which had been localised to the nucleus and thus was unable to be lost by the apparent sedimentation of the mitochondria from within the samples (Figure 4.3).

Figure 4.3: Cell debris is lost during the GCC chromatin preparation.
The harvested cells are vortexed with glass beads before a hole is punched in the bottom of the 2 ml tube and the sample spun into a new tube (1 min, 2000 rpm). Cell debris are not spun through the hole efficiently, and this is the step where I propose the bulk of the mitochondrial organelles and genomes are being lost.

4.2.3 Validation of Mito-gDNA interactions

Random inter-molecular ligation events can occur during the preparation of the GCC libraries. These inter-molecular events can occur at two stages of the protocol: 1) the ligation immediately following restriction enzyme digestion; and 2) ligation steps during the linker addition stage of the sequencing library preparation.

I wanted to establish that interactions which were detected by GCC were not random inter-molecular ligations occurring at these protocol steps. Statistical and experimental methods were used to determine if the Mito-gDNA interaction patterns could have been generated by experimental noise alone, which would be expected to produce random pairings of fragments from the two genomes. Thus, two steps were included in our analyses: 1) We perform statistical analyses to determine whether our GCC dataset is non-random; and 2) External controls were added during the GCC library preparation to obtain estimates of the rates of inter-molecular ligation events.

In silico simulations (100,000) were performed[5] [5] to determine the maximum count of a particular interaction that would be observed under the random noise model, given the same number of sequences, interactions and fragments as in the experimental data.

For example, an analysis of the glucose grown sample was performed as follows:

There are 133 mitochondrial segments (restriction fragments) and 11059 nuclear segments that participate in a Mito-gDNA interaction; a total of N = 30905 interactions were observed. 100,000 simulations of random pairings, with each segment having equal probability p to pair to another segment, were performed. In an individual simulation the maximum count for an observed pairing was 5 and only 5 of the 100,000 simulations achieved this value. In our real dataset we observe a maximum count of 14 for an observed pairing. We never observed a maximum count of 14 during our simulations and therefore we conclude that our dataset is non-random with a p-value less than 10^{-5} (<1 in 100,000 simulations = $<10^{-5}$; Table 4.2).

Table 4.2: Simulations of random pairings demonstrate the GCC data is non- random.

Sample	**Glucose**	**Glycerol Lactate**	**Galactose**
N^o. Mitochondrial Segments	133	132	131
N^o. Nuclear Segments	11059	12825	10748
N^o. Interactions Observed	30905	88107	27596
N^o. Simulations	100000	100000	100000
Simulated max count	5	6	5
N^o. Simulations Reaching this Count	5	2	2
Observed max count	14	41	32
p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$

p-value was determined as $<10^{-5}$ as none of the 100,000 simulations reached a max count of 14, 41, or 32 which were observed in the glucose, glycerol lactate or galactose samples, respectively.

The maximum count, for any random pairings, over 100,000 simulations was 6 for the glycerol lactate condition, and 5 for the other two conditions. In the real dataset the maximum counts observed for any pairing was 14 for the glucose condition, 32 for galactose and 41 for glycerol lactate. Therefore, I conclude that the interaction patterns we observed cannot be solely attributed to noise alone, under any of the conditions we assayed, in each case with a p-value less than 10^{-5}.

[5] Simulations and statistical analyses were performed myself once the protocols were established in Chapter 3.

4.2.3.1 Which individual interactions are above experimental noise?

Analyses were performed to determine what frequency individual interactions have to achieve before they are deemed to be present at a level above experimental noise. It has to be assumed that the pairings are independent and therefore the number of times one specific pairing occurs is a binomially distributed random variable. Let S1 and S2 be the number of mitochondrial and nuclear segments, respectively, which participate in at least one interaction. We calculate the probability $P(X \geq k)$ where N is number of observed pairings and p is $1/(S1 \cdot S2)$, for one specific pairing to occur k or more times. S1 multiplied by S2 gives L, with L being the number of possible pairings and we expect to see $L \cdot P(X \geq k)$ pairings occurring k or more times by chance. This provides the expected number of false positives. The value k has to be chosen in such a way to provide an acceptable number of false positives and consequently an acceptable false positive rate (Table 4.3). An acceptable noise cut-off value was determined to be a false positive rate of anything less than 0.05, that is, we do not accept a false positive rate to be any higher than 5%. Obviously, the lower the false positive rates the better. However, the interaction frequencies of individual pairings (ligations) can only increase as whole integers, therefore, increasing the noise cut-off to four, for each condition, results in false positive rates which are very low (*e.g.* glucose false positive rate = 0.000176, Table 4.3). Too stringent a noise cut-off would result in meaningful and real interactions being lost in the experimental noise and thus it is a fine balance between including experimental noise and excluding meaningful interactions. k=3 was selected as an acceptable noise cut-off value for each of the glucose, glycerol lactate and galactose mitochondrial samples (in bold text Table 4.3), that is, the interaction has to achieve a frequency of 3 in our output files in order to be deemed significant.

Table 4.3: The experimental noise cut-off can be calculated using a binomial distribution.

Sample	k	N	p	P(X≥k)	L	L·P(X≥k) Expected number of false positives	False Positive Rate
Glucose	2	30905	1/(133*11059)	2.16E-04	1470847	317.9225032	317.92/1590=0.200
Glucose	**3**	**30905**	**1/(133*11059)**	**1.51E-06**	**1470847**	**2.226555129**	**2.23/172=0.012**
Glucose	4	30905	1/(133*11059)	7.95E-09	1470847	0.011694801	0.012/68=1.76E-4
Glycerol Lactate	2	88107	1/(132*12825)	1.29E-03	1692900	2176.467784	2176.47/8884=0.245
Glycerol Lactate	**3**	**88107**	**1/(132*12825)**	**2.23E-05**	**1692900**	**37.75726218**	**37.76/1400 = 0.027**
Glycerol Lactate	4	88107	1/(132*12825)	2.90E-07	1692900	0.491252849	0.49/343=0.001
Galactose	2	27596	1/(131*10748)	1.88E-04	1407988	265.0223343	265.02/1384=0.191
Galactose	**3**	**27596**	**1/(131*10748)**	**1.23E-06**	**1407988**	**1.732331339**	**1.73/138 = 0.013**
Galactose	4	27596	1/(131*10748)	6.03E-09	1407988	0.008487333	0.0085/44=1.93E-4

k = individual interaction frequency. N = number of actual observed interactions in dataset. p = 1/(S1·S2) probability one observed pairing will occur (S1 and S2 are the number of independent segments participating in an interaction). P(X≥k) = result obtained from cumulative binomial function of the binomial distribution given k, N and p (R stats pbinom function). L = maximum possible number of pairings by multiplying S1 with S2. L·P(X≥k) = expected number of false positives. False positive rate includes the numbers of pairings which occur in the dataset for k and above. Acceptable false positive rates are highlighted in bold text.

4.2.3.2 Calculation of false positive rates for repetitive elements

Repetitive elements are those genomic features which occur more than once within the genome (*e.g.* the mitochondrial genome, the rDNA and 2-micron plasmid). I was expressly interested in the interactions between the mitochondrial genome and the nuclear rDNA repeats. Because the mitochondrial genome and the rDNA repeats occur in multiple copies in the genome I needed to ascertain this copy number to calculate acceptable k values for interactions between these elements.

I calculated the copy number of the mitochondrial genomes and rDNA repeats in my samples by aligning the sequence files against sections of these elements using the SOAP [8] algorithm with zero mismatches allowed. Three short regions were chosen, one from a unique nuclear element ([Gal1] Chr II: 279790-279909), a short section of rDNA ([rDNA] Chr XII: 460517-460612) and a section of the mitochondrial genome ([Mito] Chr Mito: 25535-25654). Table 4.4 outlines the number of sequence reads for each condition, and how many of these reads aligned to the specific region (gal, mitochondria, or rDNA), which was expressed as percentage of the total sequence reads. I calculated the ratio of rDNA or Mito to the unique nuclear element for each of the three conditions using the percentages. Calculated copy number ratios are outlined in Table 4.5.

Table 4.4: The sequence files can be used to calculate ratios between different genomic elements.

Gal (146bp)	**Sequence Reads**	**Aligned Reads**	**Percentage of sequence reads**
Glucose	112335584	733	0.00065251
Glycerol Lactate	98269812	767	0.0007805
Galactose	96838770	621	0.00064127
Mitochondria (152bp)			
Glucose	112335584	8937	0.00795563
Glycerol Lactate	98269812	12156	0.01237002
Galactose	96838770	12808	0.01322611
rDNA (96bp)			
Glucose	112335584	103881	0.09247381
Glycerol Lactate	98269812	95001	0.09667364
Galactose	96838770	85548	0.08834065

Sequence reads refers to the number of sequences in the Illumina Genome Analyser sequence output files. The number of sequences which aligned to specific regions (Gal 146bp, Mitochondria 152 bp, or rDNA 96 bp) is displayed in the aligned reads column. The number of sequences which aligned against the specific regions is expressed as a percentage of the total for each of the conditions, glucose, glycerol lactate and galactose.

Table 4.5: The ratio of genomic elements in the three growth conditions.

	Mito:Gal	**rDNA:Gal**
Glucose	12.19236	141.7203274
Glycerol Lactate	15.848761	123.8604954
Galactose	20.624799	137.7584541

Ratios were calculated by dividing the percentages for Mito or Gal against the single copy Gal locus.

Interactions between the mitochondrial genome and the nuclear rDNA repeats were isolated and subjected to binomial analysis with copy number corrections included (Table 4.5). Copy number corrections were integrated into the binomial calculations by altering the number of interacting segments for the probability calculation p (*e.g.*, there were 151 mitochondrial *MspI* fragments involved in an interaction, the copy number was found to be 12.19236, thus 151*12.19236 = 1841 mitochondrial fragments). Thus, the interaction value cut-off for individual pairings above experimental noise, for each condition, has been calculated to be k=3 (Table 4.6).

Table 4.6: Mito-rDNA interactions occur above experimental noise.

Sample	k	N	p	P(X≥k)	L	L·P(X≥k) Expected number of false positives	False Positive Rate
Glucose	3	21358	1/(1841*3330)	$7.028x10^{-9}$	6130530	0.0431	$1.85x10^{-5}$
Glycerol Lactate	3	33576	1/(2425*2911)	$1.787x10^{-8}$	7059175	0.126136304	$2.74x10^{-5}$
Galactose	3	11709	1/(3135*3237)	$2.557x10^{-10}$	10147995	0.00259514	$1.66x10^{-6}$

k = individual interaction frequency. N = number of actual observed interactions in dataset. p = 1/(S1·S2) probability one observed pairing will occur (S1 and S2 are the number of independent segments participating in an interaction). P(X≥k) = result obtained from cumulative binomial function of the binomial distribution given k, N and p (R stats pbinom function). L = maximum possible number of pairings by multiplying S1 with S2. L·P(X≥k) = expected number of false positives. False positive rate includes the numbers of pairings which occur in the dataset for k and above.

4.2.3.3 No interactions with external controls were above experimental noise

To experimentally control for spurious inter-molecular ligation events, during the GCC library preparation protocol, samples were spiked with two ligation controls. The first ligation control consisted of one of two PCR products (either a 211 bp *E. coli* fragment or a 185 bp Lambda fragment; 2.2.3.5) that were added, at a 1:1 ratio with the number of nuclear genomes in the sample, before the GCC ligation step (the glycerol lactate sample did not include this external ligation control). Crucially, the PCR products had a cut *MspI* site at one end, and were therefore free to ligate to any cleaved *S. cerevisiae* DNA fragments. These controls were designed to estimate the frequency of random inter-molecular ligation events during GCC library preparation (2.2.3.5). A maximum of 50 separate ligation events were observed between the ligation control and the *S. cerevisiae* genome in the galactose sample, while the glucose sample only had 10 (data not shown). Crucially, none of the individual ligations occurred at levels above k=3, the statistically defined experimental noise cut-off (Table 4.3).

The second ligation control consisted of the addition of pUC19 plasmid to the sample following the GCC ligation, and prior to purification, in order to control for random ligation events during adaptor ligation, at the sequencing centre. 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 cut off value.

4.2.4 The mitochondrial and nuclear genomes are highly connected

A total of 8678, 8153, and 1780 significant Mito-gDNA interactions were detected by GCC, in exponentially growing (OD_{600} = 0.600) *S. cerevisiae* cells grown on glucose,

glycerol lactate, and galactose media, respectively. Significant Mito-gDNA 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-gDNA; Table 4.7). There were more than 10 times the number of Mito-gDNA interactions characterised in *S. cerevisiae* growing in glycerol lactate as compared to yeast cells growing on glucose or galactose. Similarly, there were more than 10 times the number of nuclear genome interaction positions in yeast cells grown on glycerol lactate. Thus the frequency of the interactions was largely the same in the three conditions, only the number of unique individual interactions increased for glycerol lactate.

Table 4.7: A greater than 10-fold increase in the number of statistically significant ($p=<10^{-5}$) Mito-gDNA interactions is observed for cells undergoing respiration.

	Glucose	**Glycerol Lactate**	**Galactose**
Number of Interactions	363	3879	278
Number of observed *MspI* Fragment Pairings [a]	101	1208	76
Number of Mitochondrial *MspI* fragments	59	126	39
Number of Nuclear *MspI* fragments	94	1136	75

Totals do not include Mitochondrial-Mitochondrial genome interactions, rDNA, or interactions involving the YSCPLASM. [a] There are a total of 135 *MspI* fragments (>=13bp) in the mitochondrial genome at a copy number of 50, and a total of 18265 nuclear *MspI* fragments (including rDNA and YSCPLASM at appropriate copy numbers: Table 4.5). There is then, theoretically, $1.233x10^8$ possible pairings between mitochondrial and nuclear *MspI* fragments. The number of observed pairings between nuclear and mitochondrial *MspI* fragments is listed, as is the number of *MspI* fragments involved in those pairings.

4.2.5 The number of Mito-gDNA interactions correlates with the length of the nuclear chromosome

I wanted to establish whether any chromosomes interacted more or less frequently with the mitochondrial genome than others. One way to achieve this is to plot the length of a nuclear chromosome with the number of Mito-gDNA interactions it participated in, thus allowing me to ascertain whether any chromosomes were different from each other. The number of Mito-gDNA interactions per nuclear chromosome is highly correlated with chromosome length in the glycerol lactate condition, but not in glucose or galactose (Figure 4.4). 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 *MspI* fragment that encompasses the promoter region and partial coding sequence of two divergent ORFs: one uncharacterized ORF (*YJR115W*) and *RSM7* which encodes a mitochondrial small subunit ribosomal protein. Importantly, numerous mtDNA *MspI* fragments, including fragments surrounding or overlapping the *COX1*, *COX3*, *VAR1* and *SCE1* genes,

interact with this one "hotspot" on chromosome 10. Comparing the number of nuclear chromosomal interactions with the mitochondrial genome uncovered nuclear genome interaction hotspots.

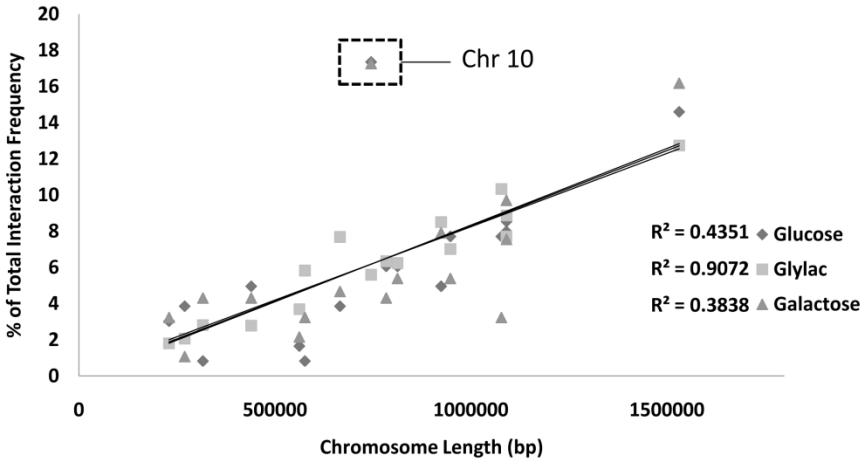

Figure 4.4: The number of Mito-gDNA interactions correlates with chromosome length, with notable exceptions.
Significant Mito-gDNA interactions (Table 4.3) have been summed for each nuclear chromosome and expressed as percentage of the total number of interactions for the particular sample before being plotted according to chromosome length in base pairs. Mito-rDNA interactions have been removed as they result in chromosome 12 as an outlier. Interactions included in this analysis are between the mitochondrial genome and nuclear chromosomes, with the YSCPLASM and rDNA removed. The length of chromosome 12 has been reduced by removing the rDNA repeats to account for the rDNA interactions being removed.

4.2.6 The mitochondrial fragments participating in the Mito-gDNA interactions are not evenly distributed across the mitochondrial genome

I next investigated whether there were any "hotspots" for Mito-gDNA interactions along the length of the mitochondrial genome (Figure 4.5). To test whether Mito-gDNA interactions have a uniform distribution (*i.e.* the total number of interactions in a segment is proportional to its length) consecutive restriction fragments were aggregated to create 58 sections that were expected to have at least five interactions under the null hypothesis of uniformity. A Chi-squared goodness of fit test was performed[6], and the distribution of the interactions was shown to deviate significantly from uniformity ($p < 0.0001$, 57 df) for all conditions, thus, Mito-gDNA interactions are not uniformly distributed across the mitochondrial genome.

[6] This statistical test was performed by Dr. Beatrix Jones at the Centre for Mathematical Biology, Massey University

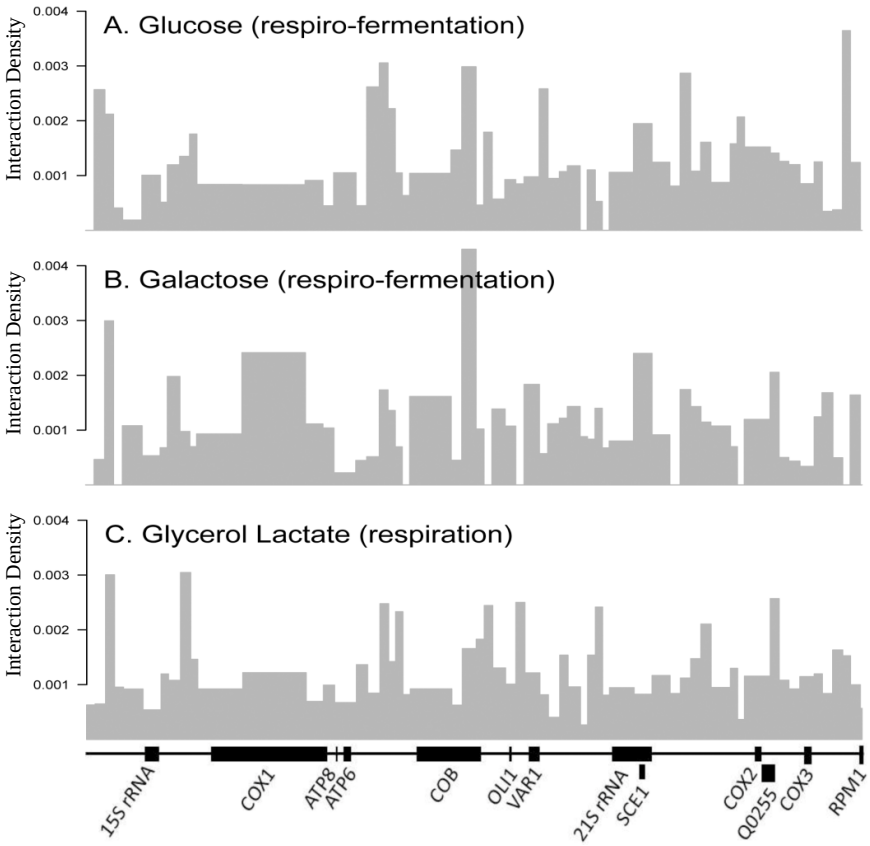

Figure 4.5: Inter-organelle interactions vary with metabolic state and do not occur evenly across the mitochondrial genome.
Interaction frequency is graphed as a percentage of the total number of interactions in the sample, according to segment length (that is, the shaded area is proportional to the interaction frequency and the fragment length). The entire mitochondrial genome is linearised and depicted across the bottom for comparison of the interaction frequency with mitochondrial ORF and inter-genic sequence positions. Metabolic conditions are 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 (FPR = 0.012 – 0.027; n=2). Interactions with the rDNA and 2-micron plasmid were removed.

Statistical smoothing of the data was important in order to remove potential biases evident in the data. Under a random pairing model we expect that the longer the *MspI* fragment the greater the chance it will interact with the nuclear genome, as there are potentially more sites to interact with, which is what tends to be observed in our dataset (Figure 4.6). Similarly, shorter fragments will be expected to interact less frequently (a comparison of Figure 4.5 and Figure 4.8). Figure 4.7 demonstrates this potential bias well, as plotting the raw interaction frequencies for each mitochondrial *MspI* fragment emphasises the greater interaction frequencies associated with longer fragments. The 'long fragment' length bias can be removed from the analysis of the interaction frequencies by dividing the number of base pairs in the *MspI* fragment. This analysis allows the interaction density to be plotted for each base pair (Figure 4.8). Now the opposite problem is evident; if a shorter fragment did indeed participate in an

interaction then because the frequency was only divided by a small number of base pairs, that particular *MspI* fragment appears to interact very highly, in contrast to the results shown in Figure 4.7. This is epitomised by the spikes present in all three conditions. Neither of these analyses is wrong, however the assumptions and biases inherent in the analyses need to be considered. The different plots obtained from the same data illustrate the importance of a full understanding of what is being plotted.

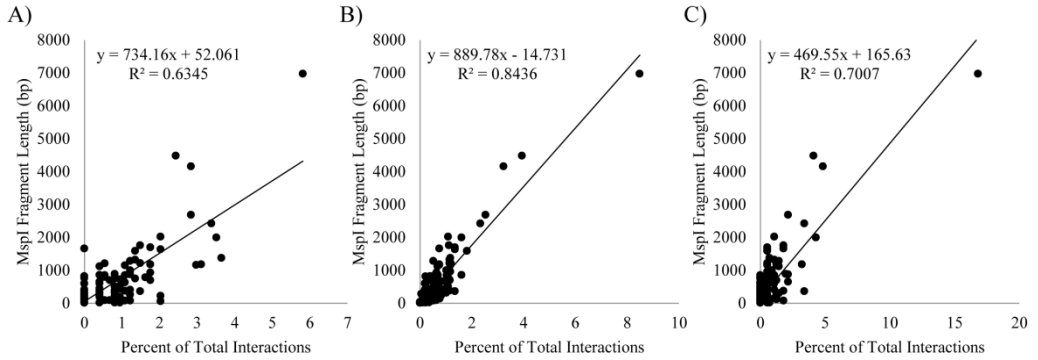

Figure 4.6: The length of the mitochondrial *MspI* fragment correlates with the number of statistically significant interactions it participates in.
The percentage of total statistically significant Mito-gDNA interactions for each condition [A) glucose, B) glycerol lactate, and C) galactose] was plotted as a scatter plot against the length of the *MspI* fragment (bp).

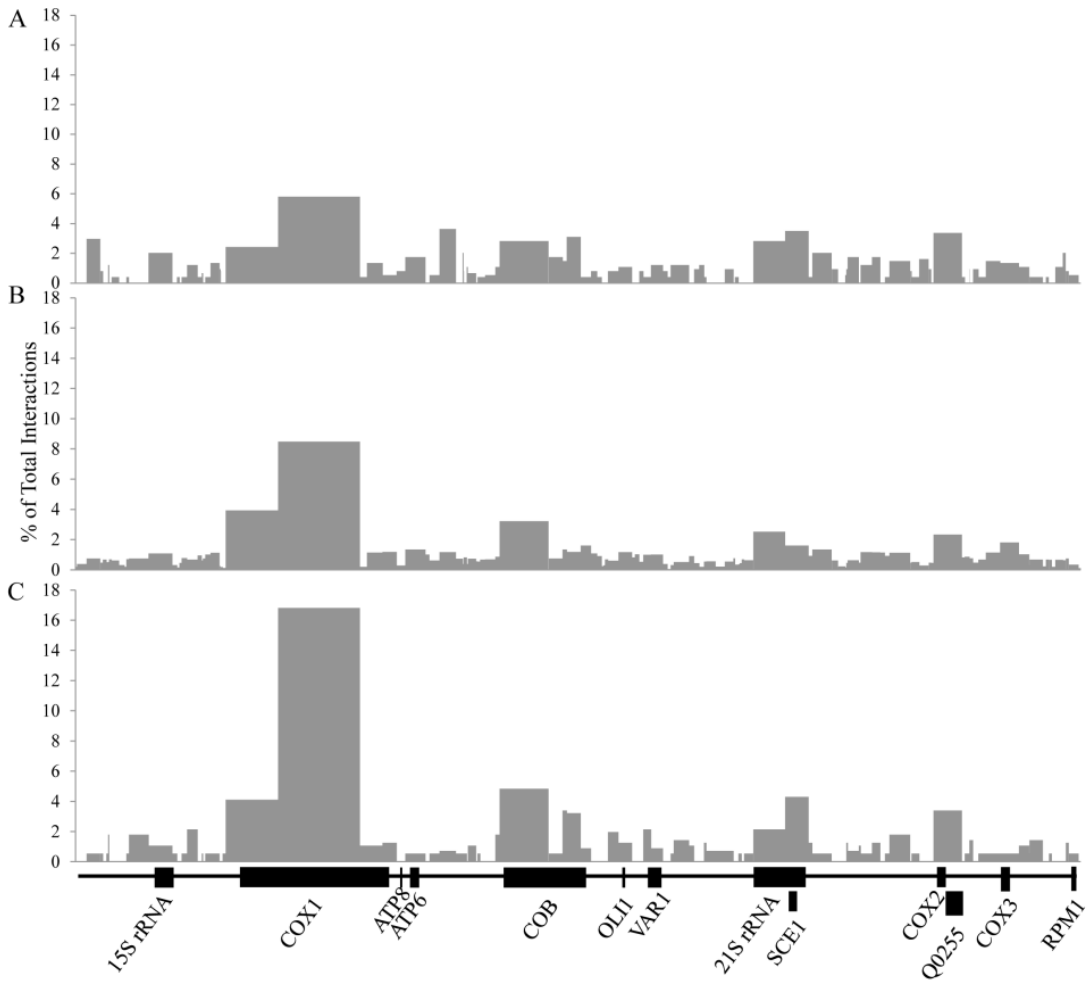

Figure 4.7: Mito-gDNA interactions plotted as percentage of total interactions without statistical smoothing.

The frequencies of statistically significant interactions (same interactions as in Figure 4.5) have been summed for each mitochondrial *MspI* fragment and the totals have been simply plotted as percentages of the total number of Mito-gDNA interactions in a given sample, A) glucose, B) glycerol lactate, and C) galactose. The mitochondrial genome is illustrated in a linearised form below the x axis to illustrate interactions in relation to the positions of the ORFs and inter-genic regions.

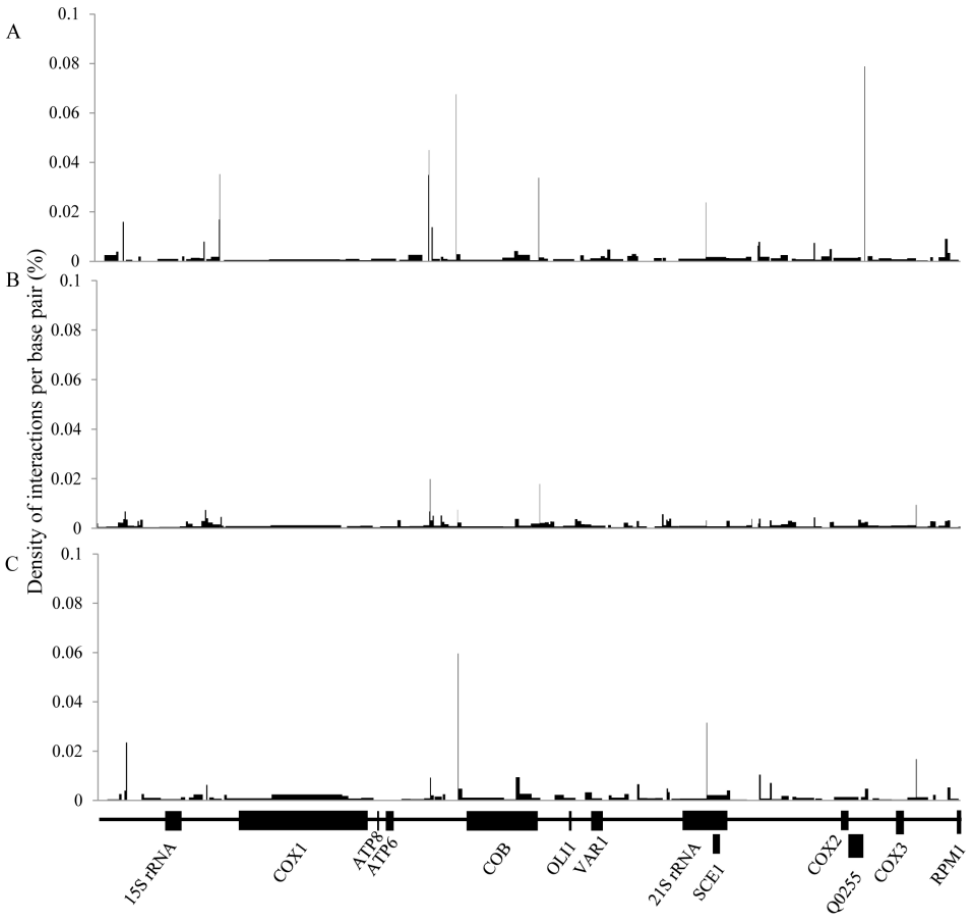

Figure 4.8: Mito-gDNA interactions plotted as interaction density per base pair.
The number of statistically significant Mito-gDNA interactions (the frequencies) have been summed for each mitochondrial *MspI* fragment and divided by the length of the *MspI* fragment in base pairs. Each base pair density is plotted as a percentage of the total density for a given condition; A) glucose, B) glycerol lactate, and C) galactose. The mitochondrial genome is linearised across the bottom of the plots to illustrate interactions in relation to the position of ORFs and inter-genic regions.

4.2.7 Mito-gDNA interactions show no detectable bias for inter- or intra-genic mitochondrial regions

The numbers of inter-organelle interactions occurring with mitochondrial inter- and intra- genic regions were compared to determine if the interactions across the mitochondrial genome were enriched over the open reading frames (ORFs; Figure 4.9). Interestingly, the galactose sample exhibited 7% and 13% more inter-organelle interactions involving the *COX1* ORF than glycerol lactate and glucose, respectively. Despite this, none of the experimental conditions showed significant enrichments over either inter- or intra-genic regions. Thus, while there is no obvious preference for inter- or intra- genic mitochondrial regions in inter-organelle interactions, *COX1* did show differences between the datasets. Therefore, a single interaction, between the mitochondrial *COX1* gene (Chr Mito: 24872–26193 bp) and the nuclear encoded *MSY1*

gene (Chr XVI; 365496–365760 bp), was chosen to be examined in more detail by quantitative 3C.

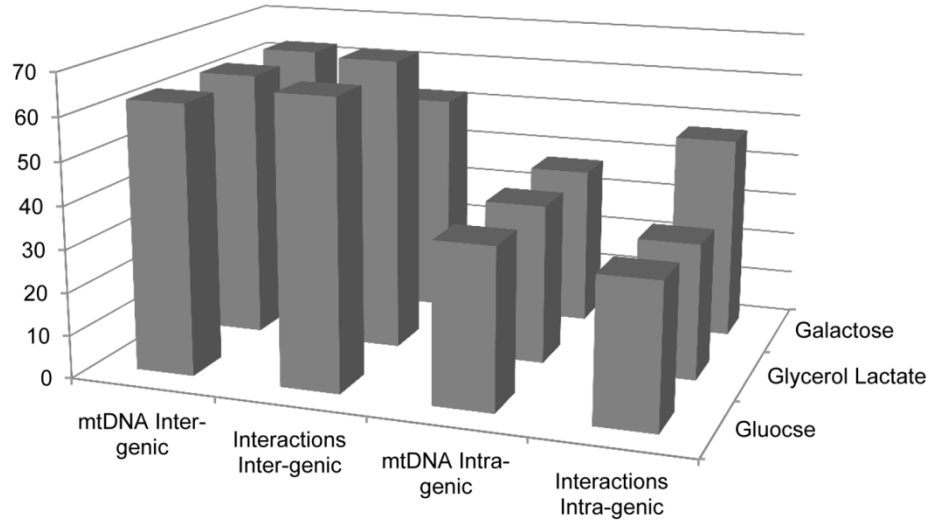

Figure 4.9: Nuclear genome interactions are not enriched over mitochondrial open reading frames.
The base pair percentage of the mitochondrial genome that was deemed inter-genic was calculated (mtDNA Inter-genic). This was compared against the percentage of inter-organelle interactions which interacted with these inter-genic regions (Interactions Inter-genic). The base pair percentage of the mitochondrial genome that was deemed as intra-genic was calculated (mtDNA Intra-genic). This was compared against the percentage of inter-organelle interactions which occurred with the intra-genic regions (Interactions Intra-genic). Galactose displays a larger number of interactions with mitochondrial ORFs (Intra-genic) but the difference is not statistically significant. tRNAs were not deemed intra-genic. 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.

4.2.8 The mitochondrial genome is highly connected to the nuclear rDNA repeats

The ribosomal DNA repeats are the DNA component of the nucleolus, encode the rRNA component of the cytosolic ribosomes and constitute ~9.8% of the yeast genome. Interestingly, the Mito-rDNA interactions constitute 95.8%, 52.4%, 84.5% of the total inter-organelle interactions between the nuclear and mitochondrial genomes in glucose, glycerol lactate, and galactose, grown cells, respectively (calculated as the percentage Mito-rDNA from the total number of interactions summed from Table 4.7 and Table 4.8). Hence, Mito-rDNA interactions are over-represented within the data-set and are carbon source dependent.

Table 4.8: Statistically significant Mito-rDNA interactions are approximately 2-fold higher in yeast cells grown in glucose, as opposed to glycerol lactate, with a further 2.8-fold reduction for cells during growth on galactose.

	Glucose	**Glycerol Lactate**	**Galactose**
Number of Interactions	8315	4274	1512
Number of observed *MspI* Fragment Pairings [b]	283	244	128
Number of Mitochondrial *MspI* fragments	151	153	152
Number of rDNA *MspI* fragments	23	23	23

[b] There are a total of 135 *MspI* fragments (>=13bp) in the mitochondrial genome at a copy number of 50, and a total of 4600 nuclear rDNA *MspI* fragments (at a copy number of 200). Thus there are, theoretically, 3.105×10^7 possible pairings (total mitochondrial fragments [135 multiplied by 50 = 6750] multiplied by total number of rDNA fragments = 4600) between mitochondrial and nuclear rDNA *MspI* fragments. The number of observed pairings between nuclear rDNA and mitochondrial *MspI* fragments is listed, as is the number of *MspI* fragments involved in those pairings.

I wanted to determine whether there were any regions within an rDNA repeat which interacted more prevalently with the mitochondrial genome. I summed the Mito-rDNA interaction frequencies for each rDNA *MspI* fragment and plotted them as percentages of the total number of Mito-rDNA interactions (Figure 4.10). In general, glucose appears to have a greater number of peaks, accounting for a large percentage of the interactions, as opposed to glycerol lactate and galactose whose interactions are more evenly spread across the rDNA repeats. Interaction frequencies at specific points over the rDNA change depending upon the growth condition (highlighted in Figure 4.10). There do not appear to be interaction "hotspots" in the galactose condition, with Mito-rDNA interactions evenly spread across the two 9.1kb rDNA repeats (Figure 4.10C). Short *MspI* fragments (<120 bp) in the glucose and glycerol lactate conditions appear to interact more frequently than other, longer rDNA *MspI* fragments (highlighted in blue and yellow, respectively).

The mitochondrial *TAR1* gene, proposed to be involved in the regulation of respiratory metabolism, is encoded within the 25S rDNA gene [220]. Unfortunately, no specific changes in interaction frequency were observed between this region and the mitochondrial genome, for each of the three conditions.

The interaction between the nuclear rDNA and the mitochondrial genome highlighted in yellow which sits between the 18S and 5S rDNA genes occurs over an rDNA autonomously replicating sequence (ARS1200). This interaction is lowest in glucose and highest in galactose, with an intermediate number of interactions in glycerol lactate. Glycerol lactate and galactose have slower growth rates than glucose, thus the firing of this ARS may not be as frequent, population wide, compared to glucose. Interestingly,

the concentration of mtDNA has been shown promote rDNA replication [167], possibly indicating a functional role for this interaction.

It is not immediately clear from the GCC data why there are specific Mito-rDNA interaction frequency differences between the conditions. Further experiments will be required in order to elucidate what role the rDNA repeats, or the nucleolus, plays in the presence of mtDNA within the nucleus. Data correction was not carried out before graphical display.

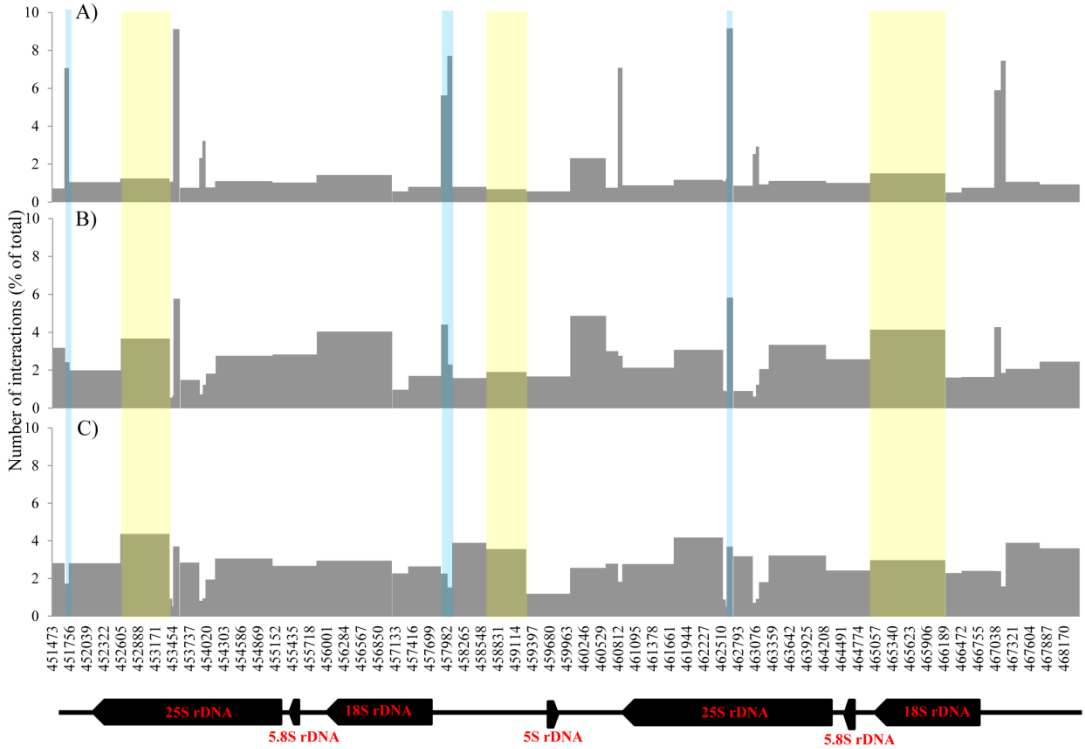

Figure 4.10: The rDNA repeats interact with mtDNA and appear to have interaction hotspots in the glucose and glycerol lactate conditions.
Significant Mito-rDNA interactions have been summed for each rDNA *MspI* fragment. The interactions frequency of each *MspI* fragment has been plotted as percentage of the total number of interactions for the particular condition; A) glucose, B) glycerol lactate, and C) galactose. Only two rDNA repeats are included in the *S. cerevisiae* reference genome file (as opposed to the typically accepted copy number of ~200 [10]) used in the bioinformatic steps of GCC protocol, therefore, two rDNA repeats are linearised and depicted across the bottom of the plot. The inclusion of only two rDNA repeats simplifies the analysis, but ensures interactions are annotated as repetitive for analytical purposes. The base pair position on chromosome 12 is displayed above the rDNA gene depiction. Large *MspI* fragments which change between the three conditions are highlighted in yellow, while shorter fragments are highlight in blue.

4.2.9 How are mitochondrial DNA fragments transferred to the nucleus?

Due to the non-random distributions of the Mito-gDNA interactions within the nuclear genome (Table 4.7, Table 4.8), it was assumed that the Mito-gDNA interactions were occurring within the nucleus and thus I researched the literature for mutants which increased the rate of transfer of mtDNA to the nuclear compartment. Yeast

mitochondrial escape mutants (YME) [162] have been previously implicated in an elevated rate of transfer of unstable mitochondrial plasmids to the yeast nucleus [162-165]. Therefore, I hypothesised that the YME pathway regulated the transfer of the mtDNA fragments that interact with the nuclear genome.

I predicted that mutations within the YME pathway would result in an increase in the frequency of the inter-organelle *COX1-MSY1* DNA interaction. To test this prediction, I used quantitative 3C to compare the *COX1-MSY1* interaction frequency in *S. cerevisiae* YME mutants (*i.e.* Δ*yme1*, Δ*yme2)* and wild-type (BY4741) yeast cells grown on glucose (Figure 4.11). Contrary to expectations, I observed a significant decrease in the frequency of the *COX1-MSY1* interaction in the Δ*yme1* strain as compared to the wild-type (paired P(T<=t) one-tail p = 0.010; Figure 4.11). Deletion of a functionally unconnected nuclear gene (*ade2*) did not significantly affect the *COX1-MSY1* interaction frequency (paired P(T<=t) one-tail p = 0.103; Figure 4.11). These results suggest that the source of the mtDNA that participates in the Mito-gDNA interactions is not the unstable mitochondrial plasmids which have previously been identified as escaping the mitochondria for the nuclear compartment [162].

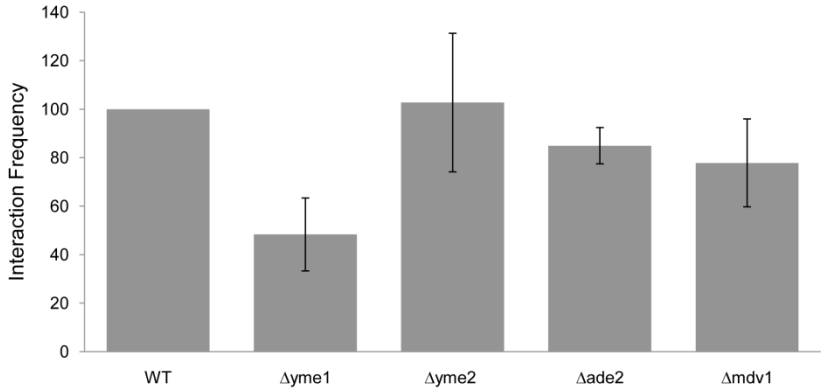

Figure 4.11: Yeast mitochondrial escape mutants do not increase the frequency at which the mitochondrial genome interacts with the nuclear genome.
The COX1-MSY1 interaction frequency was assayed by quantitative 3C PCR (2.2.3.9). Interaction values were corrected for mitochondrial genome copy number and 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.

Deletion of *yme1* results in an elevated rate of mitochondrial turn-over [221] as well as an abnormal globular mitochondrial morphology [165] and reduces the frequency of the inter-organelle *COX1-MSY1* interaction (Figure 4.11). The α-factor mediated arrest of yeast cells in the G1 phase of the cell cycle also causes fragmentation of the lattice-like mitochondrial network [222], a phenotype that is similar to that observed in *yme1*

deletion strains [221]. Therefore, I predicted that treatment of yeast cells with α-factor would cause a reduction in the *COX1-MSY1* interaction frequency (Figure 4.12). It is possible that the fragmented mitochondrial phenotype, which is a common phenotype in both *Δyme1* and α-factor treatment, contributes to the reduction in the *COX1-MSY1* interaction frequency I observed (Figure 4.11 and Figure 4.12). Importantly, strains carrying the *mdv1* deletion, which affects mitochondrial fission and not fusion [223], were also shown to exhibit a small, albeit non-significant (paired P(T<=t) one-tail p = 0.143), decrease in the *COX1-MSY1* interaction frequency. However, deletion of *yme2*, which does not affect mitochondrial morphology, but rather is involved in the maintenance of the mitochondrial nucleoid [224, 225], results in a relatively unchanged *COX1-MSY1* interaction frequency when compared to the wild type.

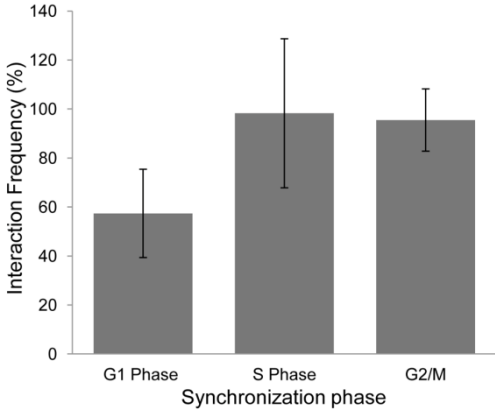

Figure 4.12: The COX1-MSY1 Mito-gDNA interaction is cell cycle dependent.
Cells were synchronized at three different cell cycle phases by treatment with α-factor (3.4 µm), Hydroxyurea (100 mM), or Nocodazole (15 µgml^{-1}); G1, S, G2/M, respectively. *COX1-MSY1* interacting *MspI* fragments, were assayed by quantitative 3C (2.2.3.9). Interaction values are corrected for mitochondrial copy number and expressed as percentages of the untreated sample (set at 100%) +/- standard error of the mean (n=3).

4.2.10 A functional electron transport chain is required to maintain the *COX1-MSY1* Mito-gDNA interaction

I postulated that an abnormal mitochondrial morphology, coupled with elevated mitochondrial turnover would result in a disturbance of the mitochondrial ATP synthesis pathway, that is, oxidative phosphorylation. Therefore, I tested the inter-organelle *COX1-MSY1* interaction for ATP-dependence using the mitochondrial uncoupling agent, 2,4-dinitrophenol (DNP), which does not allow a proton gradient to form across the inner mitochondrial membrane. Fermenting *S. cerevisiae* cells are

capable of producing ATP in the presence of DNP, however respiring cells are not able to. I wanted to compare the interaction frequency of the *COX1-MSY1* in cells producing and not producing ATP. A gradient of DNP (0.1 mM – 1.0 mM final v/v) was added to a series of *S. cerevisiae* cultures grown to an OD_{600} of 0.600 in SC glycerol lactate media to establish the concentration at which the culture ceased to grow. 0.5 mM DNP was sufficient to arrest growth in cells grown in glycerol lactate. I subsequently treated SC glycerol lactate and SC glucose *S. cerevisiae* cultures to ensure 0.5 mM DNP did not arrest growth of *S. cerevisiae* in SC glucose (Figure 4.13). Growth ceased in the DNP treated glycerol lactate containing culture, while the DNP treated glucose culture and untreated cultures continued to grow.

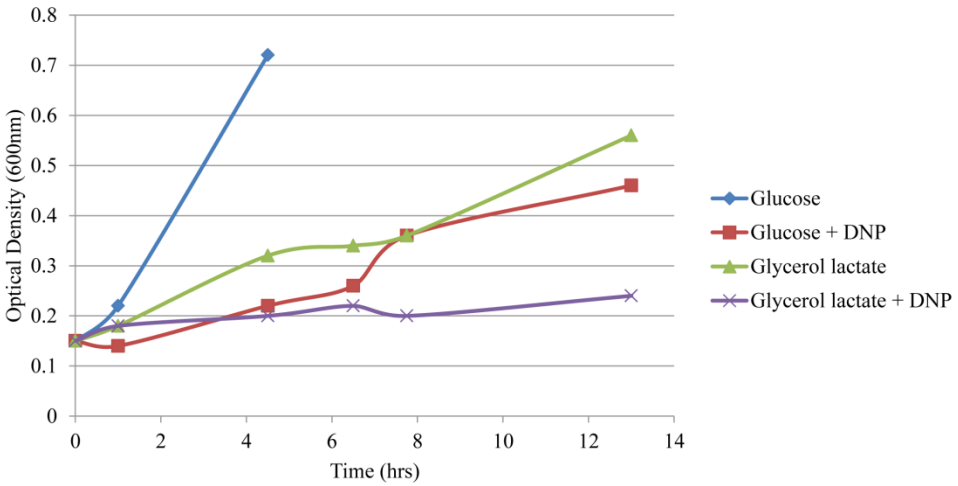

Figure 4.13: 5 mM 2,4-Dinitrophenol (DNP) inhibits respiratory growth but does not prevent growth of fermenting *S. cerevisiae* BY4741 cells.
S. cerevisiae cultures were grown (50 ml, 30°C, 160 rpm) on glucose (fermentation) or glycerol/lactate (respiration) to an OD600 of 0.600. The cultures were diluted to an OD_{600} of 0.150 (50 ml final) in their respective media. 5 mM DNP (final) 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 grown for 4 hours.

Yeast cells were treated with the electron transport chain uncoupling agent, DNP, at a concentration (5 mM) that inhibits respiration but allows fermentation. Quantitative 3C analyses were performed to monitor the effect of increasing DNP treatment times on the *COX1-MSY1* interaction (Figure 4.14). A significant time-dependent decrease was observed in the frequency of the *COX1-MSY1* interaction in the presence of DNP (t-test $p < 0.05$; Figure 4.14A). Additionally, an interaction between two nuclear loci (gDNA-gDNA; Chr VII: 868673- 873686 bp - Chr IX: 172565-173311 bp) was also shown to be affected by treatment with DNP (Figure 4.14B). The observed dependence of these interactions on a proton gradient across the mitochondrial membrane, and thus

mitochondrial ATP synthesis, failed to resolve whether the production and transfer of the mtDNA to the nucleus or the formation of the DNA-DNA interactions was ATP dependent, as indicated for gDNA-gDNA interactions (Figure 4.14B). Both scenarios remained possible. Hence, while it was established that the transfer event was different (Figure 4.11) from the unstable plasmid transfer described previously [162], I was unable to rule out ATP dependence.

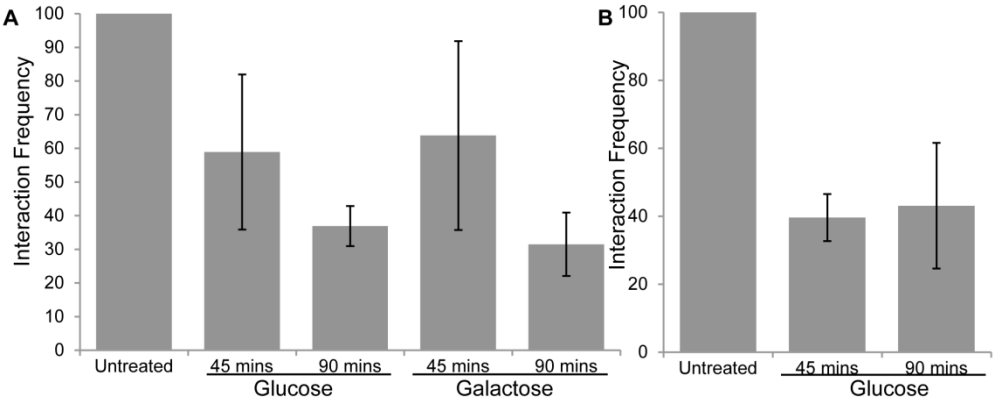

Figure 4.14: A functional electron transport chain is required to maintain the interaction between the mitochondrial *COX1* and nuclear *MSY1* loci.
Uncoupling of the electron transport chain was achieved by 2,4-dinitrophenol (5mM) treatment of exponentially growing *S. cerevisiae* in synthetic complete media, containing glucose or galactose, for the indicated time. Mito-gDNA (a) and gDNA-gDNA (b) interaction frequencies were determined by quantitative 3C analyses (2.2.3.9) and corrected for mitochondrial genome copy number. Interaction values are expressed as percentages of the untreated sample (set at 100%) +/- standard error of the mean (n=3).

4.2.11 Mito-gDNA interactions require active mitochondrial reverse transcriptase machinery

The mitochondrial *COX1* gene contains group II introns which encode functional splicing, reverse transcriptase and endonuclease machinery [226-229]. Therefore, I predicted that the *COX1-MSY1* interaction involved a *COX1* complementary DNA (cDNA) produced by mitochondrial encoded reverse transcriptase. *S. cerevisiae* strains were obtained[7], which had all group II introns removed (strain 161-U7 GII-0), and a strain with all group II introns removed with the exception of aI5γ (strain 161-U7 GII-0 aI5γ; Figure 4.15A; [9]). The aI5γ intron is located within the *COX1* gene and overlaps the region which interacts with *MSY1* (Figure 4.15).

I performed quantitative 3C analyses to establish the frequency of the inter-organelle *COX1-MSY1* interaction in the intron mutant strains growing with a respiro-

[7] Mitochondrial group II intron mutants were obtained from the laboratory of Prof. Alan Lambowitz, University of Texas, Austin. These strains were sourced through Prof. Philip S. Perlman, Howard Hughes Medical Institute.

fermentative metabolism (*i.e.* using glucose or galactose as sole carbon source). The inter-organelle *COX1-MSY1* interaction was shown to be dependent upon the presence of the group II introns and thus the reverse transcriptase machinery (Figure 4.15B). The lack of a measurable *COX1-MSY1* interaction within strain 161-U7 GII-0, which does not contain the fluorogenic probe or reverse primer binding site (Figure 4.15A and 5B), confirmed that the *COX1-MSY1* interaction does not involve a nuclear-mitochondrial sequence (NUMT). An independent interaction involving a dubious mitochondrial ORF (*Q0182;* Mito [65783-65903bp]) and the nuclear encoded *RSM7* locus (Chr X [638756-640423bp]) was also shown to be dependent upon the presence of the group II introns (Figure 4.15C). The dubious mitochondrial ORF involved in this interaction does not contain any group II introns, and thus does not encode any of the components involved in reverse transcriptase activity. These results suggest that the nucleic acids of mitochondrial origin which participate in the Mito-gDNA interaction are reverse transcribed from RNA intermediates prior to transfer to the nucleus as cDNAs. Moreover, since mitochondrial ATP concentration is coupled to mitochondrial transcription in a gene specific manner [230], DNP treatment and Δ*yme1* induced changes to mitochondrial RNA levels may explain the observed differences in Mito-gDNA interaction frequencies.

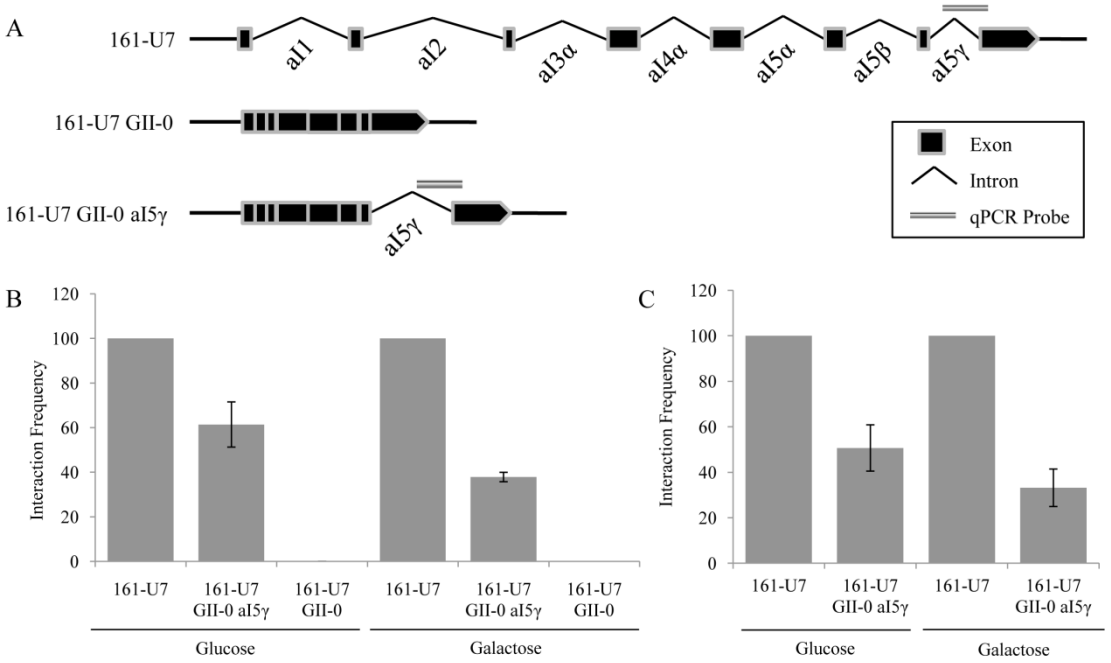

Figure 4.15: Mito-gDNA interactions require active mitochondrial reverse transcriptase machinery.
A) Depiction of the mitochondrial COX1 gene and flanking region in S. cerevisiae strains lacking mitochondrial group II introns. B) Interaction frequencies of the two mitochondrial reverse transcriptase mutants (161-U7 GII-0 aI5γ and 161-U7 GII-0, grown in glucose or galactose SC media) are displayed as a percentage of the wild type S. cerevisiae strain 161-U7 (set at 100%). 161-U7 GII-0 aI5γ is missing all mitochondrial group II introns except intron aI5γ, which overlaps the interaction fragment and taqman probe, while 161-U7 GII-0 contains no mitochondrial group II introns. 161-U7 GII-0 was included as a control for the involvement of a nuclear sequence originating from a mitochondrial integration within the nuclear genome. Interaction values are expressed as percentages of the wild type sample for each carbon source (set at 100%) +/- standard error of the mean (n=3). C) An Interaction between a dubious mitochondrial ORF (Mito [65783-65903bp]) and a nuclear segment incorporating the RSM7 gene (Chr X [638756-640423bp]) requires active mitochondrial reverse transcriptase machinery. Interaction frequencies of the mitochondrial reverse transcriptase mutant (161-U7 GII-0 aI5γ, grown in Glucose or Galactose SC media) are displayed as a percentage of the wild type *S. cerevisiae* strain 161-U7 (set at 100%). 161-U7 GII-0 aI5γ is missing all mitochondrial group II introns except intron aI5γ. Interaction values are expressed as percentages of the wild type sample for each carbon source (set at 100%) +/- standard error of the mean (n=3). Interaction

4.2.12 Nuclear gene transcription is affected by Mito-gDNA interactions

The number of significant (Table 4.7) Mito-gDNA interactions increased by >10-fold in respiring (*i.e.* glycerol lactate grown) cells, as opposed to glucose or galactose grown cells. This increase was not due to a higher number of sequence reads for the respiring glycerol lactate sample (total pooled reads; glucose 56,167,792, glycerol lactate 49,134,906, and galactose 48,419,385). Thus, a greater number of unique nuclear loci connect to mtDNA during respiratory growth when the mitochondria are most active (Table 4.7). This result, coupled with the need for a functional electron transport chain and reverse transcriptase machinery, led to the hypothesis that the Mito-gDNA

interactions are functional in nature, and specifically that they are capable of controlling the transcription of the nuclear loci with which they interact.

Quantitative reverse transcriptase PCR (qRT-PCR) was performed to determine the transcript levels of the nuclear encoded *MSY1* and *RSM7* genes in WT cells[8], the mitochondrial group-II intron knockout mutant (161-U7 GII-0), and strain 161-U7 GII-0 aI5γ (Figure 4.15A). We found that the transcript level of the *MSY1* gene is significantly higher (t-test p=0.0007) in strain GII-0 (Figure 4.16A), which does not contain the probe site or any detectable *COX1-MSY1* interactions (Figure 4.16A and B), thus identifying the maximum level of transcription in the absence of detectable inter-organelle interactions. Critically, removal of only the introns (*i.e.* strain 161-U7 GII-0 aI5γ; Figure 4.16A), caused a similar increase in *MSY1* transcript levels (Figure 4.16A). The mean concentration of *RSM7* transcripts is also higher in both the 161-U7 GII-O and 161- U7 GII-0 aI5γ strains than WT (Figure 4.16B).

Deletion of *MRS1*, which is involved in mitochondrial group I intron splicing [231], had no effect on either *RSM7* or *MSY1* transcript levels (Figure 4.16C), or the *COX1-MSY1* interaction frequency (Figure 4.17). Thus, alterations to mitochondrial reverse transcriptase activity lower the frequency of Mito-gDNA interactions and increase the levels of the nuclear encoded transcripts. Hence, I conclude that, at least in the case of the *COX1-MSY1*, and to a lesser extent *Q0182-RSM7* interactions, cDNA mediated Mito-gDNA interactions are involved in the regulation of the nuclear transcripts.

[8] Mr. Ralph Grand (Institute of Natural Sciences, Massey University) performed qRT-PCR on strains lacking group II introns and determined the *COX-MSY1* interaction frequency by quantitative 3Cs in the Δ*mrs1 S. cerevisiae* strain.

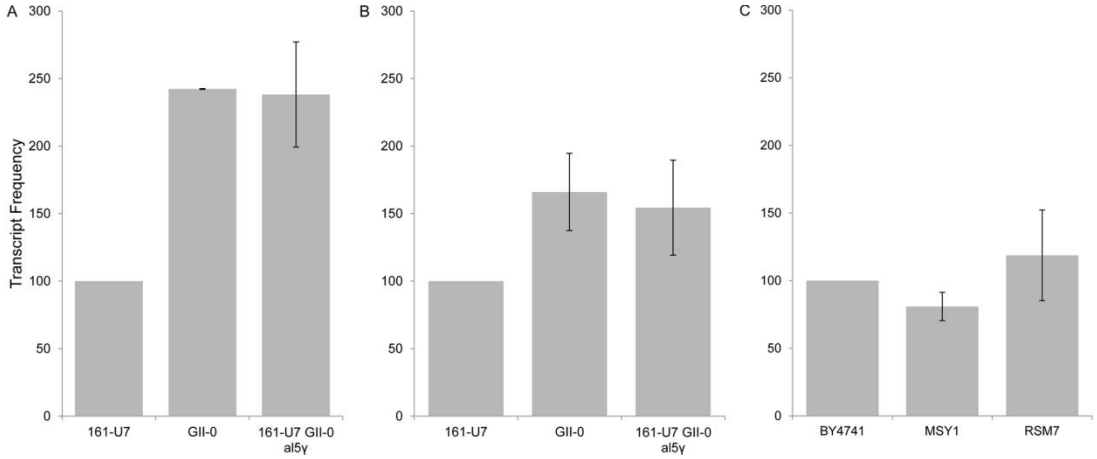

Figure 4.16: Knocking out mitochondrially encoded reverse transcriptase activity results in increased transcription of nuclear genes that are involved in Mito-gDNA interactions.
A) Nuclear encoded *COX1* transcript levels were determined by qRT-PCR in WT (strain 161-U7), GII-0 (lacks both the mitochondrial group II introns and the *COX1* interacting region; Figure 4.15A), and 161-U7 GII-0 a15γ (contains the interacting region and lacks the group II introns; Figure 4.15A) cells. B) Nuclear encoded *RSM7* transcript levels were determined by qRT-PCR in WT (strain 161-U7), strain GII-0 cells and strain 161-U7 GII-0 a15γ. Neither of the GII-0 or 161-U7 GII-0 a15γ strains have any alterations 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 *RSM7* or *MSY1* transcript levels. Similarly, Δ*mrs1* does not affect *COX1-MSY1* interaction frequency (Figure 4.17). 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).

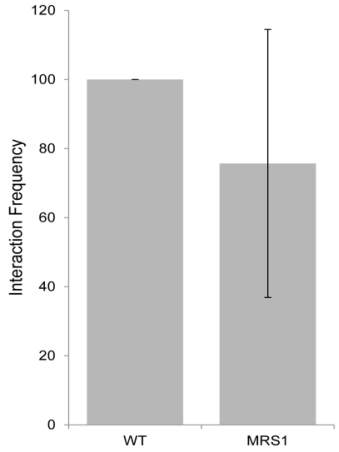

Figure 4.17: Deletion of the *MRS1* gene does not alter COX1-MSY1 interaction frequency.
Quantitative 3C was performed on WT (BY4741) cells and BY4741 cells with the nuclear *MRS1* gene deleted, using primers to assay the *COX1-MSY1* Mito-gDNA interaction. Interaction frequency of the Δ*mrs1* cells is expressed as a percentage of WT (set at 100%). Values are means of three replicates and corrected for nuclear genome copy number using the *ACT1* gene as opposed to *GAL1* previously used.

4.3 Discussion

Despite the fact that inter-organelle nucleic acid transfer events are clearly an on-going [162, 164, 165, 232-238] and predominantly unidirectional process [238, 239], it has been proposed that there is no function for recent transfer events [240]. The observation that the number of mitochondrial interactions per nuclear chromosome is highly correlated with the nuclear chromosome length, with the exception of chromosomes 10 (during growth on glucose and galactose) and 12 (Figure 4.4), appears to support this hypothesis. However, the variation exhibited by chromosome 10 is almost entirely explained by an increase in Mito-gDNA interaction frequency at a single locus: *RSM7*. These data correlate with the fact that *RSM7* encodes a conditionally essential (*i.e.* respiration) small subunit mitochondrial ribosomal protein that is induced during the switch from fermentation to respiration [241, 242], the metabolic condition in which it has the least interactions with mitochondrial DNA. Similarly, the *COX1-MSY1* inter-organelle interaction couples a conditionally essential (respiration) nuclear encoded tyrosyl-tRNA synthetase to mitochondrial activity, and specifically *COX1* transcript processing. Again, this interaction regulates the transcript levels of the nuclear *MSY1* gene. Therefore, in combination with the observation that a reduction in interaction frequency correlates with an increase in the amount of nuclear encoded *RSM7* and *MSY1* transcripts, I propose that the transfer events are functional and form part of the mitochondrial – nuclear communication system.

The yeast mitochondrial genome contains group I and II introns that encode splicing, endonucleases and reverse transcriptase machinery (*i.e.* aI1 and aI2) [243]. Deletion of the aI1 and aI2 introns present within the *COX1* gene, in strain 161-U7 GII-0 aI5γ [7, 9] caused a 40 - 60% reduction in the frequency of the inter-organelle *COX1-MSY1* and *Q0182-RSM7* interactions. These results confirm a role for a mitochondrial encoded RNA intermediate in these interactions. Further support for the role of an RNA intermediate is provided by the finding that disruptions to mitochondrial inner membrane integrity (*i.e.* in Δ*yme1* mutant) and mitochondrial ATP synthesis (*i.e.* DNP experiments) caused reductions in Mito-gDNA interaction frequencies. Both of these experimental conditions reduce the mitochondrial ATP concentration, which has been shown to affect *in vivo* mitochondrial RNA synthesis, in a gene dependent manner [230]. The incomplete ablation of the interactions following the deletion of the *COX1* aI1 and aI2 introns could result from: 1) the presence of other retro-transposon or

retroviral encoded reverse-transcriptase of either mitochondrial or nuclear origin within the mitochondrial matrix, 2) cytoplasmic or nuclear reverse transcription of mRNA released from damaged mitochondria, or 3) the transfer of mitochondrial mini-circles [162]. Mechanisms 1) and 2) are most likely given the identification of remnants of nuclear derived *copia*-, *gypsy*- and LINE-like retro-transposon elements within *Arabidopsis* mitochondria [244, 245], the finding that transcription does not affect mitochondrial mini-circle transfer in yeast [162, 166], and the reductions in interaction frequency observed with the Δ*yme1* mutant which has a disrupted inner mitochondrial membrane.

It is unclear whether the mtDNA, which participates in the Mito-gDNA 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 the high success rates attained for yeast transformation [246, 247], and the finding that the uptake of unstable mitochondrial plasmids occurs by vacuole mediated release into the cytoplasm [164, 165]. However, Ricchetti *et al.* demonstrated that the mtDNA mediated repair of nuclear double strand breaks is independent of Δ*yme1* mutations [154] 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 [157], which is formed from the outer-leaflet of the nuclear membrane [248].

The mtDNA fragments involved in the observed Mito-gDNA interactions were not evenly distributed across the mitochondrial genome. The finding that there was no bias for ORFs appears to argue against the transfer of cDNAs. However, yeast mitochondrial genes are transcribed as polycistronic transcripts (reviewed in [249]), from 14 ATATAAGTA consensus promoters and possibly another 5 non-consensus promoter sites [250]. Hence, as with other genomes, a large percentage of the mitochondrial genome is physically transcribed [251-253] and is not limited to regions immediately surrounding the ORFs.

A large number of metabolism dependent, inter-organelle interactions were observed between mitochondrial genomic loci and the nuclear rDNA repeats. These interactions could be explained in terms of the control of cryptic RNA polymerase II (PolII) transcription from within the rDNA repeats [136, 219, 254]. Previous observations

indicate that the mechanism may involve the formation of *SIR2* dependent inter- and intra- rDNA repeat loops [88] that contact the replication fork block site, present within the repeats, and in so doing regulate cryptic rDNA PolII transcription [255]. Hypothetically, the formation of the Mito-rDNA interactions antagonize the formation of these rDNA loops and thus regulate the rate of cryptic rDNA PolII transcription [255] and replication. Such a mechanism fits with observations that cryptic rDNA transcription [218, 219] and replication [256] are dependent upon the mitochondrial genome. Furthermore, the replication effect was antagonized by *SIR2* [256]. Alternatively, it is possible that the nucleolus sequesters cDNAs derived from the mitochondria in a cell cycle- or environmentally-dependent manner. This situation would be analogous to the role the nucleolus is postulated to play in sequestering nuclear proteins involved in the cell cycle [257] and environmental responses (reviewed in [258]).

Our results establish a role for Mito-gDNA interactions, and hence the mitochondrial cDNA, in inter-organelle signalling and the control of nuclear transcription. It is clear that there is considerable redundancy in the pathways that coordinate the carbon controlled expression of genes, with regulation occurring at all stages of expression both in the nucleus and mitochondria themselves [15, 249, 259]. I propose that a fraction of mitochondrial RNAs are reverse transcribed and these cDNAs are transported to the nucleus where they both negatively and positively regulate nuclear functions, including transcription and replication [256]. While we have shown a repressive role for Mito-gDNA interactions in the control of transcript levels, there is no reason to assume that all interactions are repressive. This proposal is supported by findings that nuclear transcription is positively and negatively altered in response to the presence or absence of mitochondrial genome sequences in yeast [218, 219]. I hypothesise that the targets of these interactions differ depending on metabolism specific nuclear encoded factors (*e.g.* transcription factors [15]), and the steady state levels of the mitochondrial RNAs, which are ATP dependent [230]. In the case of the glucose and galactose dependent *COX1-MSY1* and *Q0182-RSM7* interactions, this communication acts to reinforce the catabolite dependent repression of mitochondrial translation (Figure 4.18).

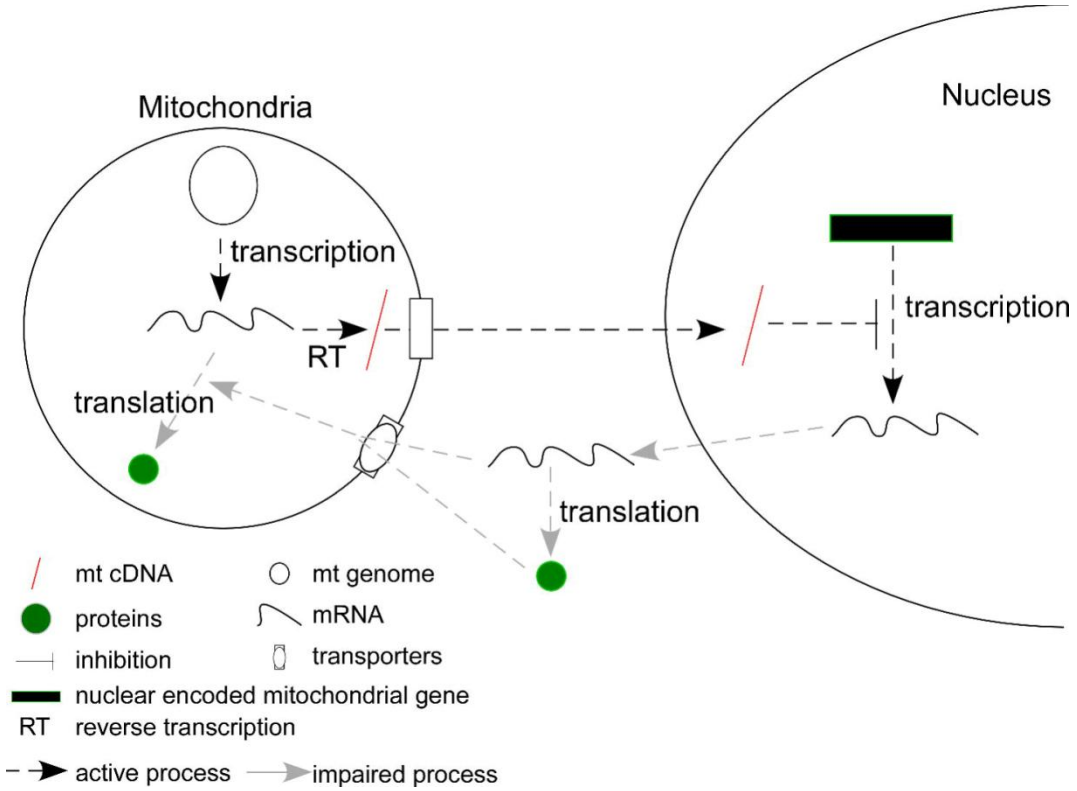

Figure 4.18: Cartoon depicting the role for inter-organelle communication in the refining mitochondrial translation levels.

Transcription levels of the nuclear encoded *MSY1* and *RSM7* genes are modulated by interactions with mitochondrial cDNAs, which are produced from stable non-translated mitochondrial RNA populations. Other Mito-gDNA interactions also contribute to this feed-back system. While negative feedback is emphasized in this model, feed-forward systems are also predicted [2].

4.4 Conclusion

In conclusion, I have demonstrated that mitochondrial cDNA sequences form part of an inter-organelle reverse transcriptase dependent communication system. The inter-organelle DNA-DNA interactions, and hence mitochondrial cDNAs, are capable of altering the transcriptional level of nuclear genes. Therefore, inter-organelle DNA transfer is functional and signals mitochondrial activity according to the transcriptional, translational and bioenergetic state of the mitochondria. Further work is warranted in higher eukaryotes to determine the universal significance of these interactions.

Chapter 5 Metabolic State Affects Genome Structure

5.1 Introduction

Genome conformation capture (GCC) allowed the empirical determination of the first DNA interaction network [5]. The logical next step is to ask how these networks change under different conditions. A lot of work has been carried out over the years in order to understand how *S. cerevisiae* responds to and grows in different carbon sources [13, 14, 260-262]. Therefore, GCC was performed on exponential phase *S. cerevisiae* growing on glucose, glycerol lactate and galactose, to gauge how the interaction network would respond.

S. cerevisiae is capable of fermenting, respiring, or combining these metabolic processes and performing respiro-fermentation [172, 173]. Respiro-fermentation is carried out when cells have access to oxygen and the fermentable carbon source is at a concentration exceeding 0.8mM [172, 173]. Therefore, under normal laboratory growth conditions it is estimated that 49% of the cellular ATP is being produced by oxidative phosphorylation in glucose [173]. A higher percentage of cellular ATP is derived from oxidative phosphorylation during galactose growth (83%; [173]). When all three metabolic conditions are compared (*i.e.* glucose, galactose and glycerol lactate growth), glucose and galactose have the most similar metabolic phenotype as opposed to the solely respiratory metabolism of growth in glycerol lactate.

Respiro-fermentation growth, in galactose, requires the transcription of 6 extra genes (Gene Ontology; SGD website), on top of those already expressed during glucose growth. The *GAL* locus only harbours three of these genes (*GAL1*, *GAL10*, and *GAL7*), and the other three are distributed throughout the rest of the genome. If we assume that transcription, at least in part, is responsible for genomic structure [22, 30, 31, 40, 43, 263-265], then we might expect only a small difference in the numbers and types of inter- and intra- chromosomal interactions occurring between the glucose and galactose interaction network maps. However, it has been shown that spatial rearrangement of the genome is carried out upon the need to metabolise galactose [64].

The mitochondria are much more active during respiratory growth as the mitochondria is the site of oxidative phosphorylation; the sole producer of cellular ATP under this growth condition [266]. The coordinated transcription of some ~1100 nuclear encoded genes is required in order for yeast cells to undergo respiration [15, 29, 215]. If transcription occurs in transcription factories as hypothesised [112, 264, 267], then previously silenced or compacted genes would have to be moved into active or open areas of the nucleus, such as these transcription factory sites [40, 43, 103, 268-270], more than likely leading to structural changes in the genome. Therefore, I hypothesised that the interaction network would be most different during growth in glycerol lactate as opposed to glucose and galactose.

Here I describe the similarities and differences in the numbers and types of significant interactions, detected by GCC, in *S. cerevisiae* during three different metabolic conditions (*i.e.* glucose, glycerol lactate and galactose grown *S. cerevisiae*). Galactose was observed to have the lowest connectivity overall, despite having similar numbers of sequences contributing to the interactions as glucose and glycerol lactate. Adjacent interactions are shown to report on local chromatin structure, and in keeping with this, active genes tend to be more connected by inter- and non-adjacent intra- chromosomal interactions, than when they are inactive. Interestingly, tRNA genes exhibit a preference to interact with the inter-genic spacer (IGS2) region of the ribosomal DNA repeats. A single interaction between a tRNA containing *MspI* fragment and the rDNA is not affected when the tRNA is deleted, suggesting, at least in this instance, the tRNA is not the driving force for the spatial clustering of tRNAs at the nucleolus [148]. Overall, the connectedness of the yeast genome is shown to be dynamic and capable of responding to environmental conditions.

5.2 Results

5.2.1 Construction of the GCC networks

The GCC networks were constructed from 36bp paired end Illumina Genome Analyzer sequence reads (total pooled reads from two biological replicates; glucose 56,167,792, glycerol lactate 49,134,906, and galactose 48,419,385). The two biological replicates for each condition were highly correlated for statistically significant interactions (R^2=0.78, 0.93, 0.93, respectively; Figure 4.1). Accordingly, sequences from biological replicates were combined and all further analyses were performed on these combined sequence files. The Illumina genome analyser sequences were shown to be representative of the genome (Table 4.1).

The topography software (v1.19), coupled to the SOAP algorithm [8], was used to construct GCC networks for the three different conditions. Interactions were once again divided into two files, according to whether; 1) both participating sequences could be uniquely positioned within the *S. cerevisiae* reference genome (unique); or 2) one or both of the participating sequences aligned more than once on the reference genome (repetitive). Statistical analyses were performed on the unique output files to determine which interactions occurred at a frequency above experimental noise.

5.2.2 Validation of GCC results

As mentioned earlier (2.2.4.7), random ligation events can occur during the preparation of the GCC libraries. Simulations of the GCC experiment were performed to demonstrate that sequenced GCC libraries were non-random (2.2.4.8). Additionally, external ligation controls were included into the GCC libraries to empirically measure random inter-molecular interactions (2.2.3.5). Statistical analyses were performed on the uniquely positioned topography output file, to establish that individual interactions with an interaction value k of 3 occur above the experimental noise in glucose, glycerol lactate, and galactose (Table 5.1). All subsequent analyses were performed on the unique topography output file, unless otherwise expressly stated.

Table 5.1: An interaction frequency of ≥ 3 is above experimental noise for unique inter- and non-adjacent intra-chromosomal interactions in glucose, glycerol lactate and galactose.

Sample	k	N	p=1/(S1*S2)	P(X≥k)	L	L·P(X≥k) Expected number of false positives	False Positive Rate
Glucose	3	1957694	2/(20440*20439)	1.36E-07	417773160	56.92	0.000543585
Glycerol Lactate	3	1262970	2/(20387*20386)	3.72E-08	415609382	15.479854	0.0005559
Galactose	3	1916890	2/(20388*20387)	1.30E-07	415650156	53.98464226	0.0008251

The total number of *MspI* fragments involved in interactions (p) was calculated by the addition of unique nuclear genome interacting *MspI* fragments with the number of mitochondrial *MspI* fragments at a genome copy number of 50 [11] and the number of 2-micron plasmid *MspI* fragments at a plasmid copy number of 50 [12-16]. There were 151 mitochondrial *MspI* fragments at a copy number of 50 giving 6750 *MspI* fragments. There were eight 2-micron plasmid *MspI* fragments at a copy number of 50 giving 400 *MspI* fragments. The false positive rate was calculated by dividing the number of false positives by the number of interactions which occur above the k value. The number of different pairings between *MspI* fragments which occur above the interaction frequency (k) of 3 is 104,706, 27,846, or 65,428 for glucose, glycerol lactate, and galactose, respectively.

5.2.3 The numbers and types of interactions change depending upon metabolic state

The total number of significant inter- and non-adjacent intra- chromosomal interactions (hereinafter referred to as intra-chromosomal) is 399,150, 281,867, and 223,654 for glucose, glycerol lactate and galactose, respectively (summed from Table 5.2). The number of paired end sequence reads was 13.8% less in galactose compared to glucose; however the frequency of interactions resulting from these sequences was ~44% less in galactose compared with glucose. Furthermore, glycerol lactate had ~30% less interactions than glucose while coming from a pool of sequences only 12.6% smaller than glucose. The reasons for these interaction discrepancies are not immediately obvious. Although it does imply that the genomes of glycerol lactate and galactose grown yeast are less connected overall than the glucose genome, I investigated these results further by separating the interactions according to whether they occurred within or between chromosomes (intra- or inter- chromosomal interactions, respectively).

Most of the discrepancy between the glucose interactions and the other two conditions were due to reductions in the number of intra-chromosomal interactions present in the glycerol lactate and galactose conditions. What is interesting to note is that while galactose has the fewest intra-chromosomal interactions (215,769), the total intra-chromosomal frequency resulted from a total of 30,862 *MspI* fragment pairings, three times more than in the glycerol lactate condition (9,997).

The ratios of the total intra-chromosomal frequencies to the number of pairings are 7.7:1, 26.5:1, 7.0:1 for glucose, glycerol lactate and galactose, respectively. That is, in

the glucose sample, on average, each intra-chromosomal *MspI* fragment pairing has an interaction frequency of 7.7, while glycerol lactate has a frequency of 26.5 for each intra-chromosomal interaction pairing. There were ~3-5 times fewer intra-chromosomal pairings in the glycerol lactate condition, compared to glucose and galactose, but these fewer pairings resulted in a higher frequency total than galactose. This means that while there are fewer intra-chromosomal loops in the glycerol lactate condition, the ones which do occur, occur at high frequency.

In stark contrast to the reduced number of intra-chromosomal pairings observed in glycerol lactate, the number of inter-chromosomal pairings and the total interaction frequency is more than double that of glucose and galactose. It is important to remember that the inter-chromosomal classification includes mitochondrial to nuclear genome interactions, which are enriched in the glycerol lactate condition (4.2.4).

Table 5.2: The numbers and types of interactions change depending upon the metabolic state of the *S. cerevisiae* cells.

	Glucose	**Glycerol Lactate**	**Galactose**
Total Adjacent Interactions (Total Significant and Non-significant Interaction Frequency)	2,152,251	2,588,702	1,368,342
Number of Pairings	17,226	13,548	16,754
Non-Adjacent Intra-Chromosomal Interaction (Total Significant Interaction Frequency)	390,487	264,903	215,769
Number of Pairings	50,688	9,997	30,862
Inter-Chromosomal Interactions (Total Significant Interaction Frequency)	8,663	16,964	7,885
Number of Pairings	1,665	3,927	1,853

Significant adjacent interactions are not calculated as the adjacent interactions may include interactions which result from protection of the restriction site leading to it being indigestible. Total interaction frequencies are all frequencies for individual *MspI* fragment pairings, summed. Thus the 'Number of Pairings' refers to the number of unique *MspI* pairings.

The interaction networks can be depicted by artificially circularising the chromosomes *in silico*, and drawing a line between interacting *MspI* fragments (Figure 5.1). This graphical view emphasises the reduced number of interactions overall for the galactose condition (Figure 5.1C), and also the high number of interactions and pairings between the mitochondrial genome and the nuclear genome in the glycerol lactate condition (Figure 5.1B). Furthermore, it is evident that specific regions of the chromosomes are more connected than other regions.

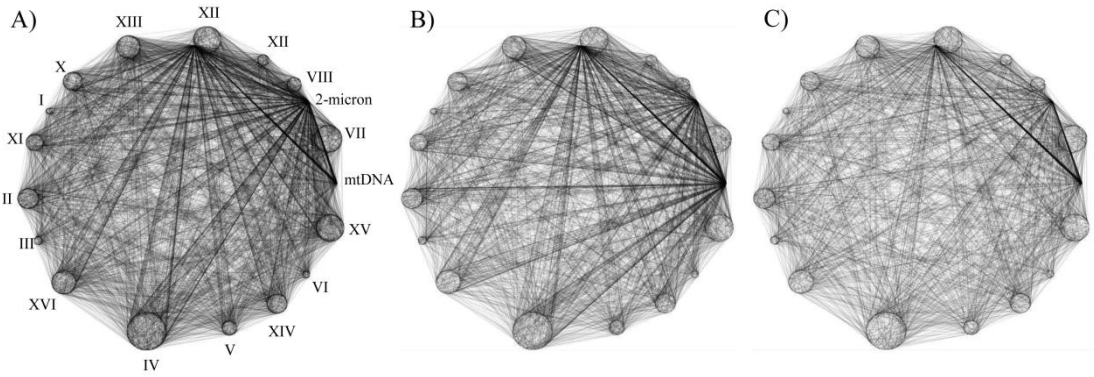

Figure 5.1: The network maps for statistically significant chromosomal interactions change according to metabolic condition.
The interactions networks are displayed for, A) Glucose, B) Glycerol Lactate, and C) Galactose. Chromosomes are depicted as circles around the outside of the network. The size of the circle is proportional to the chromosome length. Chromosome XII appears small as only two rDNA repeats are included in the analysis. Lines indicate interactions between the *MspI* fragments which connect; while the strength of the line depicts the strength of the interaction. Chromosome numbering is the same in all three figures A)-C).

5.2.4 Interaction frequencies change depending upon the growth condition

The total number of interactions in each condition can provide information about the connectedness of the genome as a whole, and the sorts of interactions that are present. I wanted to establish how the interaction frequencies changed in the three different conditions at the individual *MspI* fragment level. Therefore, the interactions were broken into several categories, including, inter-chromosomal (Inter-Chr), non-adjacent intra-chromosomal (Intra-Chr), mitochondrial genome to nuclear genome (Mito-NucGenome), intra-2-micron plasmid (Intra-YSCPLASM), intra-mitochondrial (Intra-Mito), and mitochondrial genome to 2-micron plasmid (Mito-YSCPLASM). The frequencies for individual interactions were plotted as scatter plots to compare the three different conditions; glucose versus galactose, glucose versus glycerol lactate, and galactose versus glycerol lactate (Figure 5.2A, Figure 5.3A, and Figure 5.4A, respectively).

5.2.4.1 Comparing the glucose and galactose interaction networks

Glucose and galactose undergo a mixed respiro-fermentation and thus tend to be the most similar metabolically, as opposed to *S. cerevisiae* growth in glycerol lactate which occurs solely by respiration [172]. Galactose has a number of extra genes activated (*e.g.* the *GAL* locus) in order to pre-process the galactose into glucose-6-phosphate which can be fed into the existing metabolic pathways [271]. Therefore, at least in certain regions of the genome, at or around active loci, we would expect a more open

conformation of the chromatin [50], although recent data indicates that nucleosome occupancy is relatively similar between glucose and galactose [272].

Oxidative phosphorylation activity in the glucose and galactose conditions, is largely the same (albeit slightly more ATP is produced by the mitochondria during galactose growth) [172, 173]. Therefore, I would predict that interactions which form within or between mitochondrial genomes would be similar in these two conditions. Intra-mitochondrial interactions for glucose and galactose appear to loosely follow the x=y line (Figure 5.2), albeit, slightly skewed towards galactose.

The inter-chromosomal interactions which occur in yeast grown on glucose and galactose tend to occur at similar interaction frequencies (Figure 5.2B). This is in stark contrast to intra-chromosomal interactions which occur at a lower frequency during galactose growth. This is evident in Figure 5.2C, where the intra-chromosomal interactions are predominantly below the x=y line.

Many of the same mitochondrial genome to nuclear genome pairings do not occur in both conditions, as seen by the data points sitting on the x and y axes (Figure 5.2D). Those pairings which do occur in the same conditions are inclined to be at low frequency. However, one data point is prominent in Figure 5.2D, with an interaction frequency of over 30 in galactose. This interaction occurs between part of the mitochondrial *COX1* gene and a nuclear *MspI* fragment containing a mitochondrial ribosomal subunit gene, *RSM7* (a pairing described earlier; 4.2.9). This interaction occurs at high levels in galactose even when compared to glycerol lactate (Figure 5.4D). Given the fact that reducing the interaction frequency between these two loci results in an increased transcript level of the *RSM7* gene (Figure 4.16), one might predict that because of the higher interaction frequency in galactose, the *RSM7* gene required more frequent inhibition by the *COX1* cDNA in this growth condition.

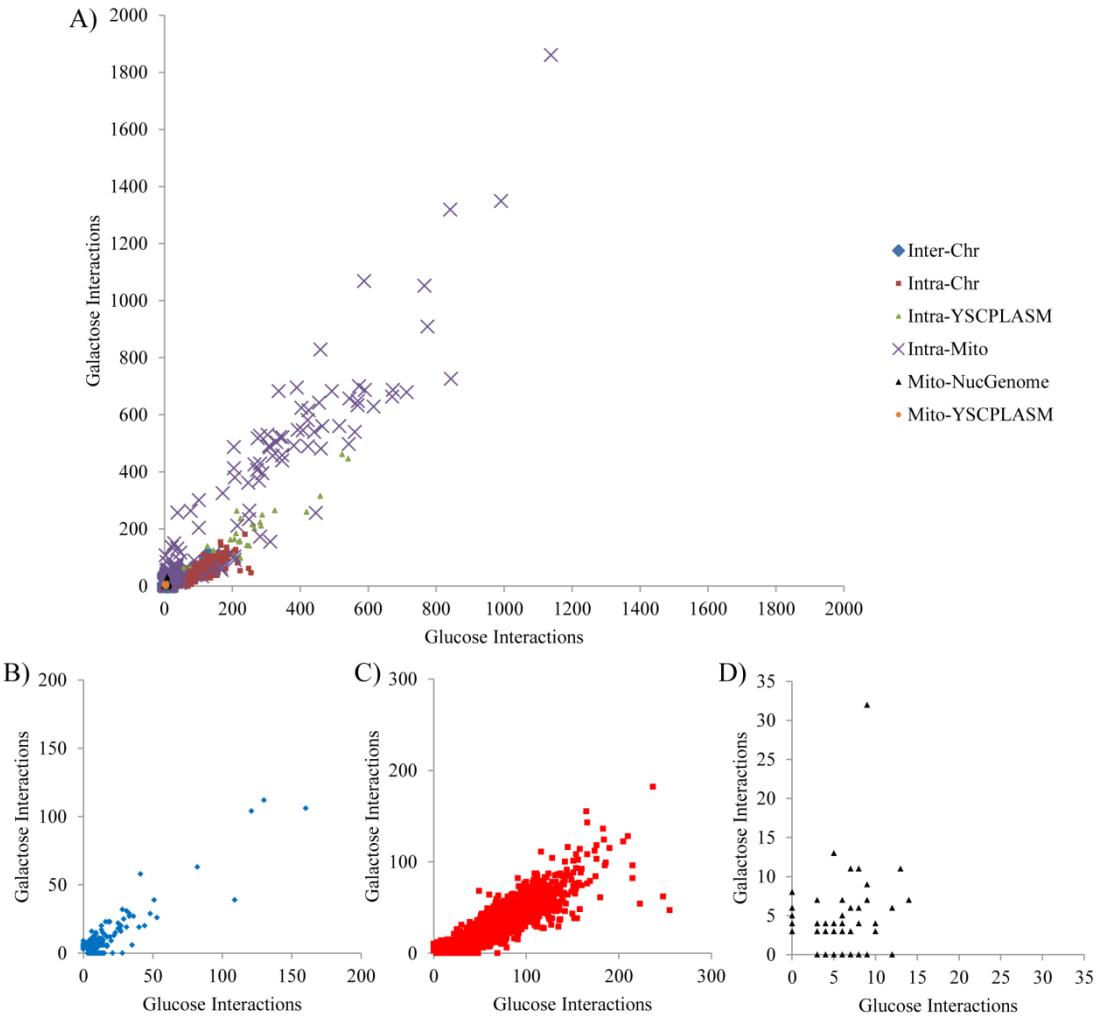

Figure 5.2: Correlations between chromosomal interactions in yeast grown in glucose and galactose conditions.
Significant interactions (FPR = 0.0005 and 0.0008 for glucose and galactose, respectively) have been separated into subsets, and displayed as scatter plots to illustrate similarities and differences between the glucose and galactose conditions. A) All significant interactions in the glucose and galactose conditions were separated into subsets according to whether they were classed as inter-chromosomal (Inter-Chr), non-adjacent intra-chromosomal (Intra-Chr), intra-2 micron plasmid (Intra-YSCPLASM), intra-mitochondrial genome (Intra-Mito), mitochondria to nuclear genome (Mito-NucGenome), and mitochondria to 2-mcron plasmid (Mito-YSCPLASM). Certain subsets are difficult to see in A), therefore inter-chromosomal, intra-chromosomal, and mitochondrial to nuclear genome interactions are plotted in B), C), and D), respectively.

5.2.4.1 Comparing the glucose and glycerol lactate interaction networks

When comparing two of the most divergent metabolic conditions, glucose and glycerol lactate (Figure 5.3) [172, 173], the intra-mitochondrial genome interactions are heavily skewed towards the glycerol lactate condition (Figure 5.3A). These higher values for glycerol lactate are consistent with previous hypotheses concerning mitochondrial nucleoid compaction during respiration, as opposed to fermentation [273]. If the mitochondrial genome is more compact then we would expect more intra-mitochondrial interactions. However, we need to bear in mind that the mitochondrial genome exists

inside the mitochondrial organelle at a copy number of ~50. The mtDNA forms nucleoid structures, consisting of the spatial clustering of ~5 individual mitochondrial genomes [150, 274]. Thus intra-mitochondrial interactions cannot be distinguished from inter-mitochondrial genome interactions. That is, inter-mitochondrial repeat interactions due to clustering of the genomes will result in them being reported as intra-chromosomal interactions (intra-mitochondrial). Therefore, a higher number of intra-mitochondrial interactions in glycerol lactate possibly indicate compactness, resulting from the prevalence of more mitochondrial genomes in each mitochondrial nucleoid.

Intra-YSCPLASM interactions are lower in yeast cells grown on glycerol lactate than those grown on glucose. Once again we have a situation where a high copy number element forms exists in a clustered form, that is, a clustering of 'like' elements [175], therefore, intra-2-micron plasmid interactions can and probably do form as a result of this clustering. There is no reported evidence to suggest that the 2-micron plasmid forms more clusters during glucose growth. One might predict that because of the slower growth rate associated with glycerol lactate, we might observe less clustering of the 2-micron plasmid, purely as a result of there being more time for these clusters to form prior to cell division. Alternatively, the opposite could be the case, a higher proportion of the cells will be in the dividing stages of the cell cycle during a fast growth phenotype, and thus more clustering of the 2-micron plasmid should be evident as the clusters form to promote segregation [177, 275].

Inter-chromosomal and intra-chromosomal interactions which occur in both glucose and glycerol lactate have a tendency to be similar in frequency (Figure 5.3B, Figure 5.3C). However, the mitochondrial genome interactions with the nuclear genome are skewed towards glycerol lactate. The mitochondria are more active during growth on glycerol lactate as they are respiring [173], which may result in a higher mitochondria to nucleus transfer frequency of DNA. As discussed in chapter 4, there was ten times the number of interactions between these two genomes in cells grown on glycerol lactate compared to cells grown on glucose. This explains the numerous data points sitting at zero for glucose on the y-axis (Figure 5.3D).

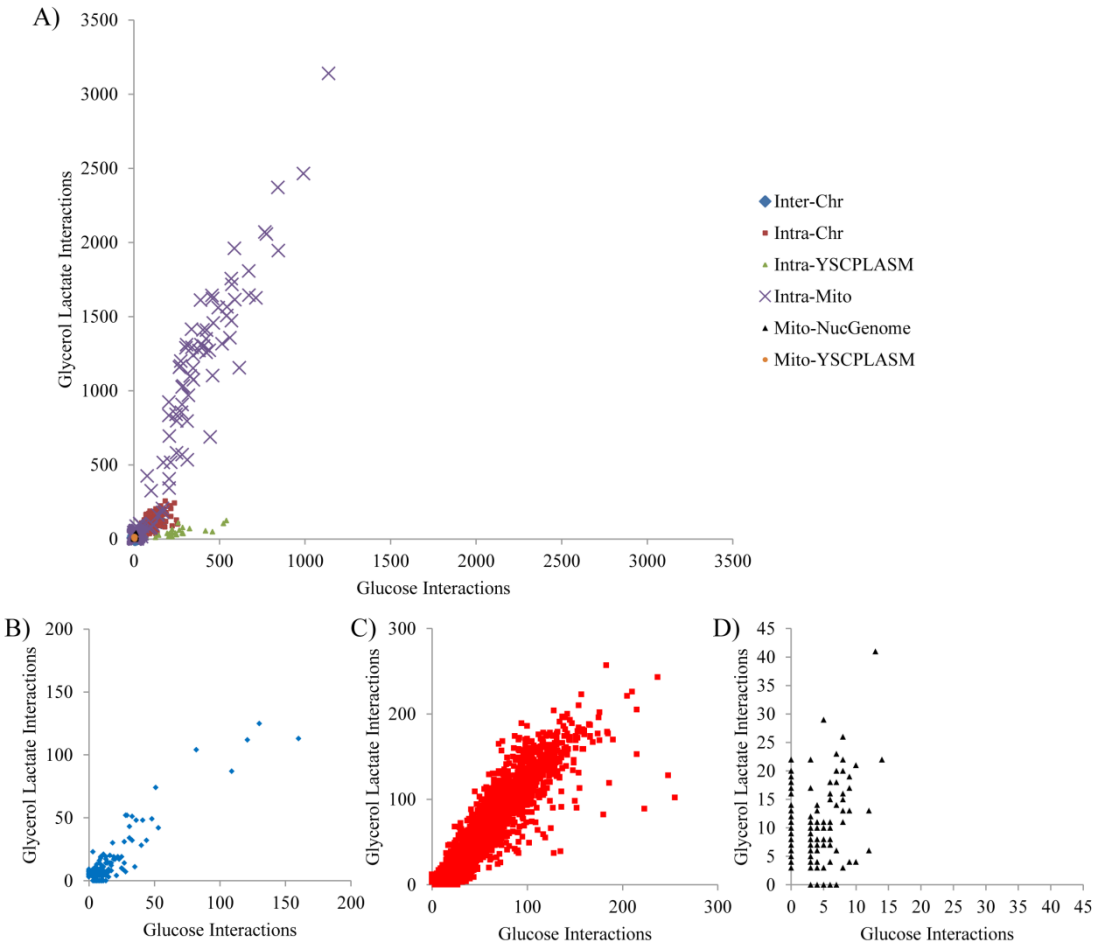

Figure 5.3: Correlations between chromosomal interactions in yeast grown in glucose and glycerol lactate conditions.
Statistically significant interactions (false positive rate 0.0005 and 0.0005 for glucose and glycerol lactate, respectively) have been separated into subsets, and displayed as scatter plots to illustrate similarities and differences between the glucose and glycerol lactate conditions. A) All statistically significant interactions in the glucose and glycerol lactate conditions were separated into subsets according to whether they were classed as inter-chromosomal (Inter-Chr), non-adjacent intra-chromosomal (Intra-Chr), intra-2 micron plasmid (Intra-YSCPLASM), intra-mitochondrial genome (Intra-Mito), mitochondria to nuclear genome (Mito-NucGenome), and mitochondria to 2-mcron plasmid (Mito-YSCPLASM). Certain subsets are difficult to see in A), therefore inter-chromosomal, intra-chromosomal, and mitochondrial to nuclear genome interactions are plotted in B), C), and D), respectively.

5.2.4.1 Comparing the galactose and glycerol lactate interaction networks

Galactose has a similar metabolic state to glucose [172], therefore we would predict that, as glucose and glycerol lactate were different, galactose and glycerol lactate should differ similarly. As expected, the scatter plot comparing mitochondrial genome interactions with the nuclear genome for galactose and glycerol lactate is very similar to that of glucose to glycerol lactate (compare figures Figure 5.3D and Figure 5.4D). When galactose and glycerol lactate interactions are compared; intra-mitochondrial interactions are predominantly skewed towards glycerol lactate, and the intra-2-micron plasmid interactions towards glucose (Figure 5.4).

Overall, comparisons of interactions from the three different conditions behave as expected; glucose and galactose are most different from glycerol lactate. Inter-chromosomal interactions which are conserved in different conditions tend to occur at similar frequencies. However, at least for the comparison of glucose and glycerol lactate, only 6.5% of the total inter-chromosomal pairings occur in both conditions. This leaves 1,759 and 954 pairings for glucose and glycerol lactate, respectively, which are condition dependent. Therefore, when making these comparisons we need to bear in mind that it is just a small fraction of all pairings which are able to be directly compared.

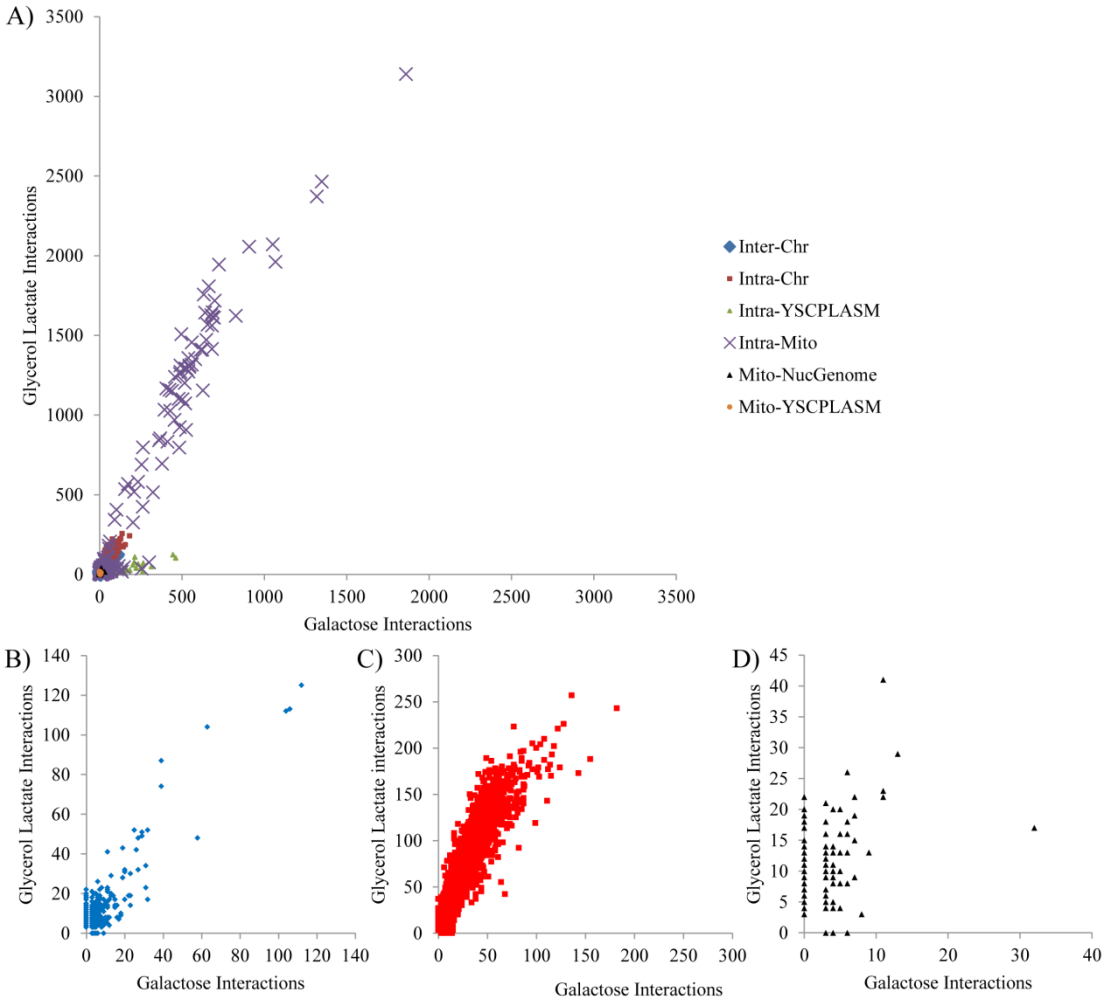

Figure 5.4: Correlations between chromosomal interactions in yeast grown in galactose and glycerol lactate conditions.
Statistically significant interactions (false positive rate 0.0008 and 0.0005 for galactose and glycerol lactate, respectively) have been separated into subsets, and displayed as scatter plots to illustrate similarities and differences between the galactose and glycerol lactate conditions. A) All statistically significant interactions in the galactose and glycerol lactate conditions were separated into subsets according to whether they were classed as inter-chromosomal (Inter-Chr), non-adjacent intra-chromosomal (Intra-Chr), intra-2 micron plasmid (Intra-YSCPLASM), intra-mitochondrial genome (Intra-Mito), mitochondria to nuclear genome (Mito-NucGenome), and mitochondria to 2-mcron plasmid (Mito-YSCPLASM). Certain subsets are difficult to see in A), therefore inter-chromosomal, intra-chromosomal, and mitochondrial to nuclear genome interactions are plotted in B), C), and D), respectively.

5.2.5 The number of inter-chromosomal interactions correlates with chromosome length

If we assume that DNA-DNA interaction positions are non-random then we would not expect the number of inter-chromosomal interactions each chromosome participates in to correlate with the chromosome length. Rather, we would expect that some chromosomes might interact poorly, or conversely, more numerously than others. There is a good correlation between the number of inter-chromosomal interactions a chromosome participates in and its length for all conditions (Figure 5.5). However chromosome 14 deviates from the trend line (highlighted below), although there is no simple feature which explains this deviation.

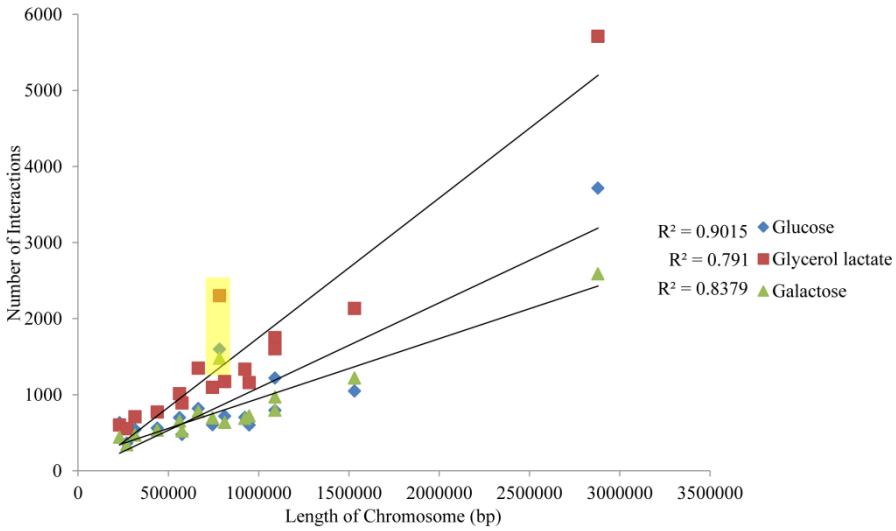

Figure 5.5: The number of inter-chromosomal interactions does not correlate with chromosome length.
The number of statistically significant (FPR = 0.00054 to 0.000082) inter-chromosomal interactions, involving unique loci, for each chromosome, is plotted against chromosome length. Chromosome XII has been altered to its correct length by the addition of the missing 198 rDNA repeats. The mitochondrial genome and the 2-micron plasmid have been excluded from this analysis. Chromosome XIV is highlight in yellow.

5.2.6 The number of intra-chromosomal interactions correlates with chromosome length

Intra-chromosomal interactions form when the DNA within a single chromosome loops and interacts with itself. Alternatively, as GCC was performed on an asynchronous population of cells, there is the possibility that during cell division, when there are two copies of each chromosome, that the intra-chromosomal interactions are being reported due to interactions occurring between sister-chromatids. There is no evidence which would indicate that any one chromosome should have more or less intra-chromosomal

interactions (loops) than any other chromosome. Therefore, I predicted that the number of intra-chromosomal interactions would be proportional to its length. As expected, when the number of intra-chromosomal interactions each chromosome forms is plotted against the length of the chromosome there is a clear correlation (Figure 5.6). It has already been established that there were more intra-chromosomal interactions present during growth in glucose, and this is reflected in Figure 5.6 as the raw number of intra-chromosomal interactions is plotted.

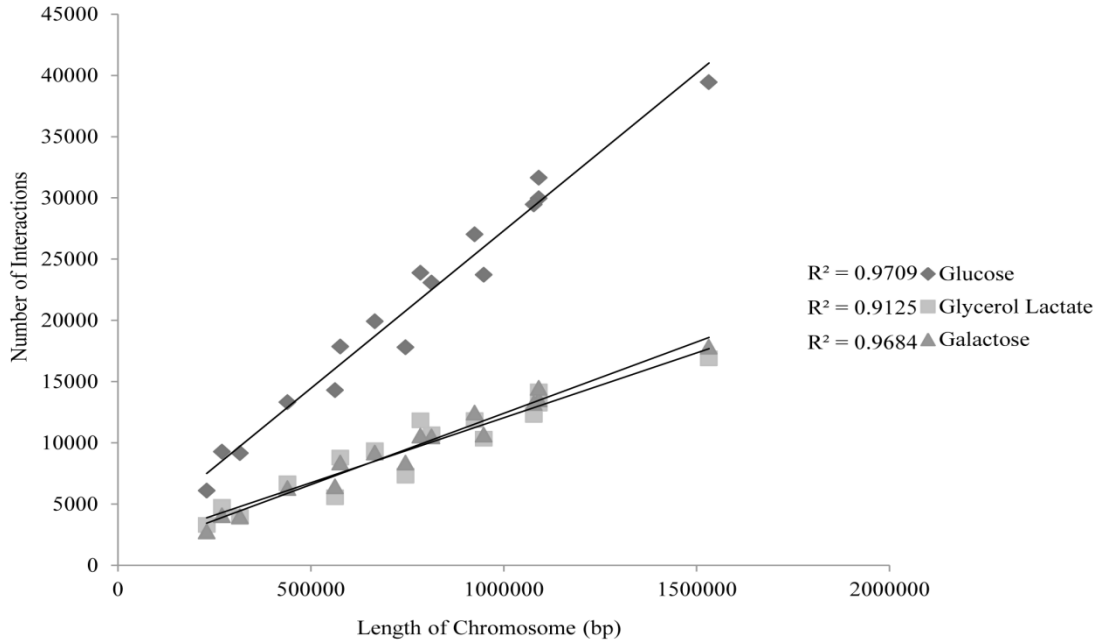

Figure 5.6: Number of intra-chromosomal interactions versus chromosome length.
The number of significant (expected false positive rate ranging from 0.00054 to 0.000082) intra-chromosomal interactions, involving unique loci, for each chromosome, is plotted against chromosome length. Chromosome 12 has not been altered to its correct length by the addition of the missing 198 rDNA repeats as the plotted interactions do not include repetitive elements. The mitochondrial genome and 2-micron plasmid have been excluded from this analysis.

Intra-chromosomal loops can be both large and small; therefore I wanted to establish whether there was a preference for any particular size of loop. The loop size was determined by calculating the distance in base pairs between the two interacting *MspI* fragments. That is, the distance between the end of *MspI* fragment 1 and the start of *MspI* fragment 2. It was calculated this way: 1) to avoid *MspI* fragment length biases for those that were involved in the interaction; and 2) because we are unable to pinpoint exactly where on an *MspI* fragment the interaction is occurring. The loop sizes were placed into bins of 1000bp and plotted as a histogram (Figure 5.7).

Intra-chromosomal interactions predominantly occur between *MspI* fragments less than 10,000bp apart. While GCC does not report upon the nature of the intra-chromosomal

interactions, loops between 1000 and 1200 bp would be consistent with loop sizes predicted for the packaging of DNA into chromatin fibers [17]. That is, one turn of the, much debated, 30nm solenoid fiber is consistent with a length of between 1000 and 1200 bp. Alternatively, gene looping where the promoter and terminator regions of genes interact in order to facilitate RNA polymerase reloading to the promoter during active transcription has been shown to occur in yeast [213, 276]. The average length of all of the verified ORFs in *S. cerevisiae* is 1564 bp. Thus, the intra-chromosomal interactions could result from gene looping. This type of conformation cannot be verified using the GCC data as the exact position, within the *MspI* fragment, where the interaction is occurring is not able to be determined.

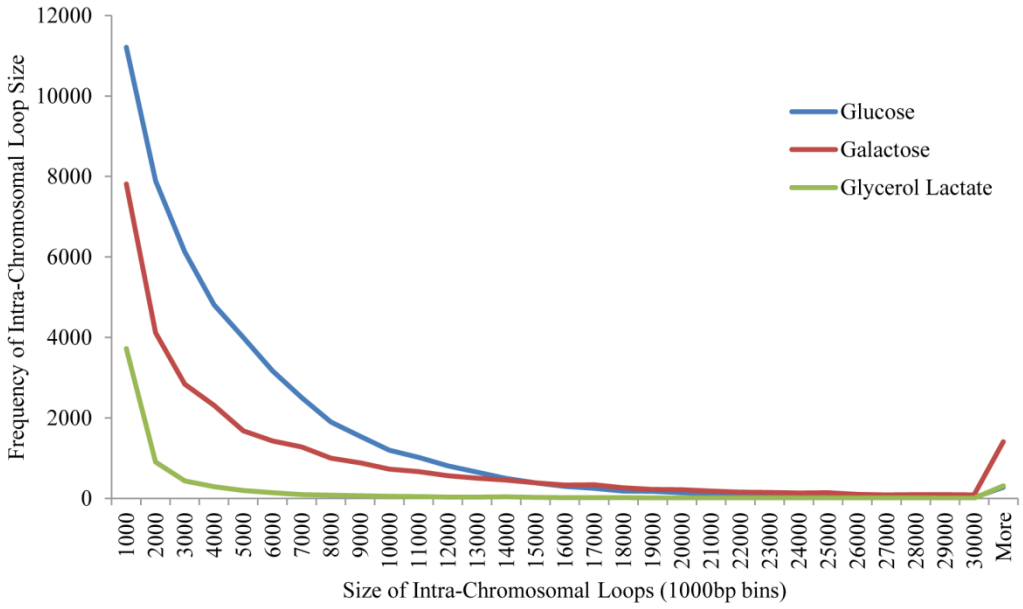

Figure 5.7: Intra-Chromosomal loop size changes depending upon metabolic condition.
The size of intra-chromosomal loops is extracted from statistically significant non-adjacent intra-chromosomal interactions and expressed as the distance between the end of interacting *MspI* fragment 1 and the start of interacting *MspI* fragment 2. Intra-rDNA, intra-mitochondria, and intra-2-micron plasmid interactions have been excluded from this analysis.

Intra-chromosomal loops determined for the human chromosomes, by a GCC related technique (Hi-C [90]), demonstrated that loops between 500 kb and 7 Mb in size followed a prominent power law. The slope exponent of -1.080 for this region was consistent with a fractal globule arrangement for the human chromosomes as opposed to an equilibrium globule (random walk model) simulated to have an exponent of -1.508 [90]. I investigated whether the yeast chromosomes would also follow a power law dependency. The yeast chromosomes are much smaller than human chromosomes, but if a scale-free power law dependency exists, the same law should hold for the *S.*

cerevisiae data. The data from Figure 5.7 was plotted as a log-log plot of loop size versus the number of loops, in bin sizes of 5000bp (Figure 5.8A). Indeed, power-law dependency is evident for loops less than 50,000 bp. The linear portion of the graph, corresponding to loops up to 50,000 bp, was extracted and the slope of the line determined for glucose (Figure 5.8B), glycerol lactate (Figure 5.8C), and galactose (Figure 5.8D). Exponent values of -3.629, -2.617, and -2.159, for glucose, glycerol lactate and galactose, respectively, are not consistent with a fractal globule model for local chromatin structure postulated for the human chromosomes [6, 90].

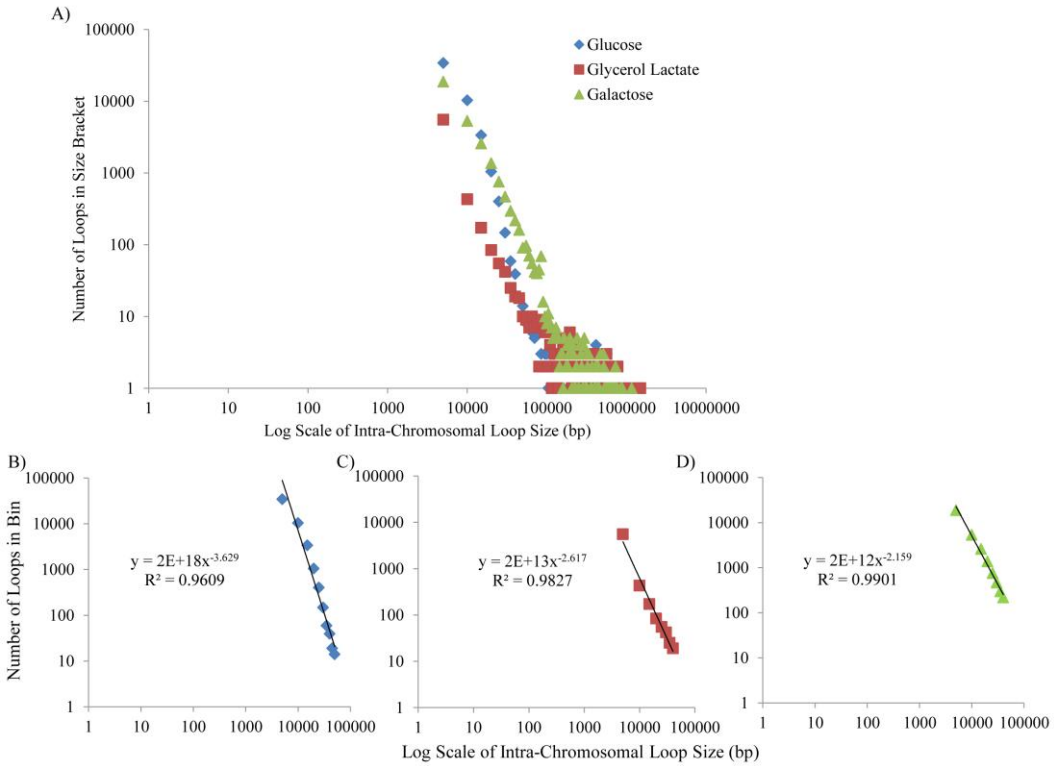

Figure 5.8: Log-log plots of intra-chromosomal loop size versus the number of loops are not consistent with a fractal globule arrangement of the chromosomes.
A) Statistically significant non-adjacent intra-chromosomal loop sizes have been plotted on log-log scatter plots, against the number of loops for each bin size of 5000bp for glucose, glycerol lactate and galactose. Power law scaling is evident up to 50,000bp; therefore, these data points have been plotted separately for B) glucose, C) glycerol lactate, and D) galactose. Power trend lines reveal the slope of the plotted data to be -3.6, -2.6 and -2.1 for glucose, glycerol lactate and galactose, respectively.

5.2.7 Adjacent interactions can report on local genome structure

Adjacent interactions may inform upon local chromatin structure as uncut restriction sites theoretically could result from shielding of the sites by nucleosome occupation (3.2.8; [50]).

The *GAL* operon on chromosome 2, consists of three genes (*GAL1*, *GAL7*, and *GAL10*) which are inactive during growth on glucose, become poised for transcription in

glycerol lactate and are active in galactose [50, 277]. I plotted the adjacent interaction frequencies for *MspI* fragments (Chr 2; 270407-283042bp) occurring across the *GAL* genes (Figure 5.9). The number of adjacent interactions have been corrected at the *MspI* fragment level, for each condition, according to the number of adjacent interactions overall. On the whole, more adjacent interactions occur across this region in glucose compared to galactose (p=0.020), whereas there was no statistically significant difference between glucose and glycerol lactate (p=0.131). Statistical calculations are unable to be carried out to compare the frequencies between the conditions for individual *MspI* fragments as there are only two replicates. Regardless, differences are evident at the individual *MspI* fragment level; the *MspI* fragments, highlighted in yellow, over the *GAL1* promoter region and part of the coding region have different interaction frequencies in the three different conditions.

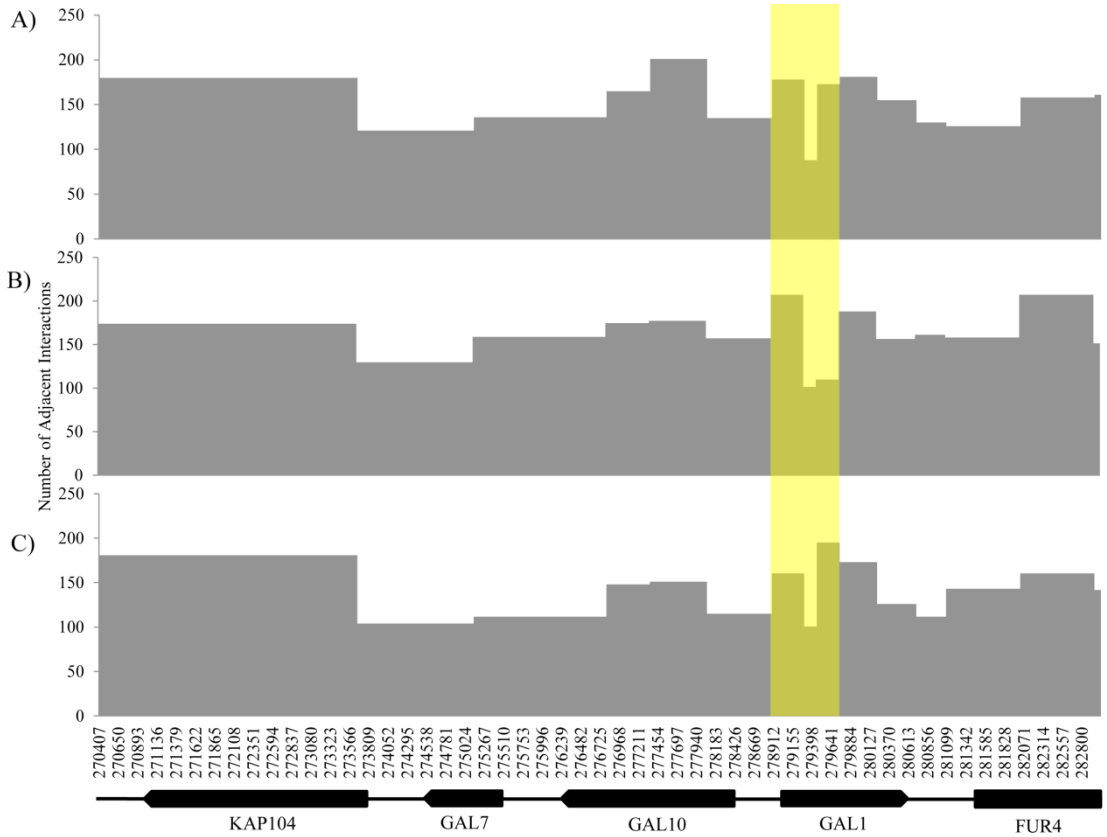

Figure 5.9: The GAL operon has a higher number of adjacent interactions in glucose and glycerol lactate compared to galactose overall.
The number of adjacent interactions across the GAL region is plotted from the middle of one *MspI* fragment across the restriction site junction to the middle of the adjacent downstream *MspI* fragment, for each condition; A) glucose, B) glycerol lactate and C) galactose. The GAL operon region plotted spans across the GAL7, GAL10, and GAL1 genes, as well as two adjacent genes, KAP104 and FUR4 (entire region plotted: Chr 2; 270407-283042bp). The genes are depicted along the bottom. Regions of interest are highlighted in yellow.

In order to gain a better perspective of whether adjacent interactions report upon local chromatin structure, I also plotted the adjacent interactions across the *DAL* locus (Figure 5.10). The *DAL* gene family are involved in allatoin degradation, that is, the transport and catabolism of poor nitrogen sources when good nitrogen sources are depleted [278]. The expression of the *DAL* genes differs depending upon the carbon growth source [29]. However, the genes alter independently of each other in glucose, glycerol lactate or galactose growth [29]. For example, the expression of *DAL1* is slightly enhanced in glucose (0.11 compared to reference pool) and during respiratory growth (ethanol carbon source; 0.68 compared to reference pool), but is repressed in galactose (-1.41 compared to reference pool). At the other extreme, *DAL3* is enhanced during glucose and galactose growth (0.17 and 0.21, respectively) and is slightly repressed during respiration (-0.16) [29]. Overall, the *DAL* genes are more repressed in galactose, as opposed to glucose and respiratory growth [29]. This is a puzzling result given the fact that more nitrogen, in the form of amino acids, is taken up during growth in galactose [279]. The expression changes according to carbon source are negligible when compared to the multiple fold increase in expression observed during nitrogen deprivation [29]. Accordingly, we would not expect any significant changes in the numbers of adjacent interactions across the *DAL* gene region, for the three conditions (paired t-test; glucose and galactose p=0.420; glucose and glycerol lactate p=0.096; glycerol lactate and galactose p=0.218). As expected, no significant difference was observed overall for adjacent interactions across the *DAL* region for the three conditions.

Differences in the interaction frequencies, between the three conditions, were evident (highlighted areas Figure 5.10). Once again, statistical significance cannot be calculated for individual *MspI* fragments; however differences are apparent at the *MspI* level. The *MspI* fragment encompassing the promoter region of *DAL1* has more adjacent interactions than glucose and glycerol lactate, consistent with the fact that *DAL1* is repressed, greater than two-fold, in galactose.

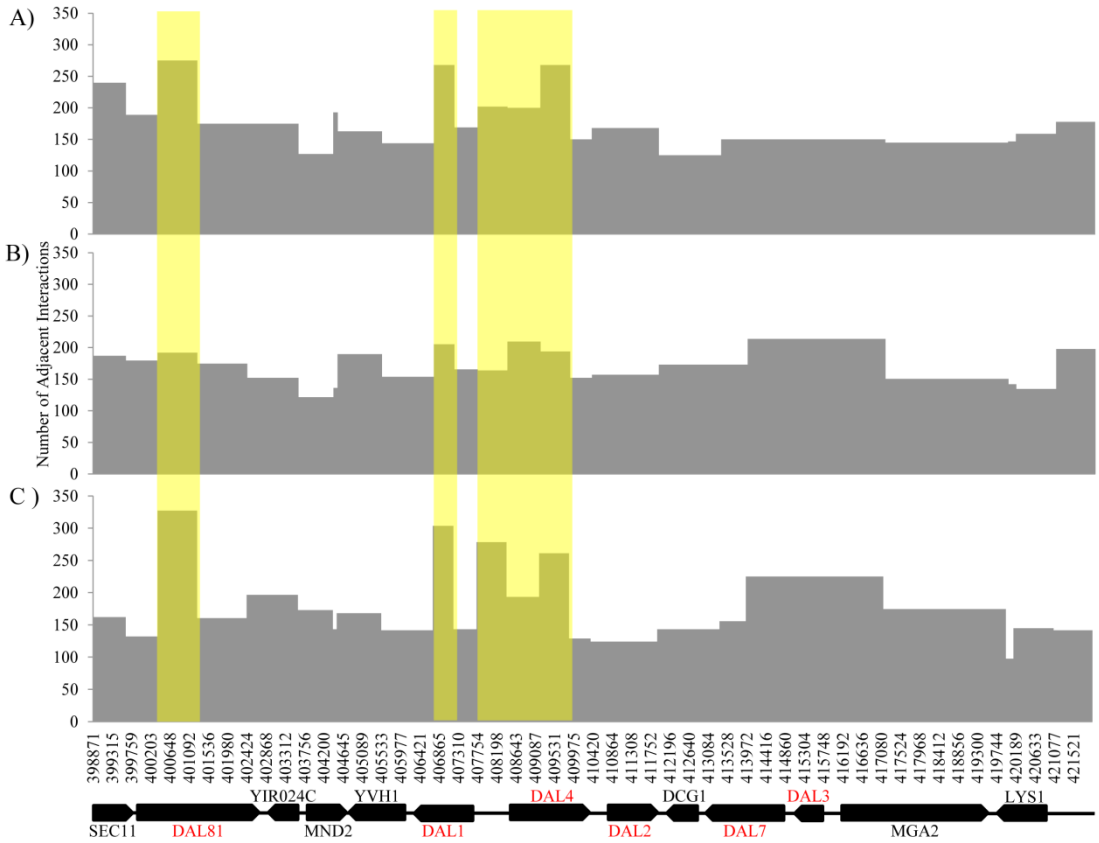

Figure 5.10: Localised changes in adjacent interaction frequency occur across the DAL genes.
The total number of adjacent interactions for each *MspI* fragment has been plotted across a region containing the *DAL* genes (Chr 9; 398697-421959). The number of adjacent interactions is plotted from the middle of one *MspI* fragment across the restriction site junction to the middle of the adjacent downstream *MspI* fragment, for each condition; A) glucose, B) glycerol lactate and C) galactose. The region of chromosome 9 plotted is depicted across the bottom. Areas of interest are high-lighted in yellow.

5.2.8 Does the number and type of interactions with the 2-micron plasmid change depending upon metabolic condition?

The high copy number 2-micron plasmid is a selfish genetic element which forms clusters inside the nucleus [175] and was observed to interact with the nuclear genome in the GCC network of exponentially growing glucose yeast cells (3.2.9; [5]). Interactions between the 2-micron plasmid and the nuclear genome were detected by GCC in all three conditions. Only nine significant pairings (see Table 5.1 for p values) were conserved in all three conditions, out of 848 total pairings. All of the conserved pairings were with the rDNA repeats. In fact there was little similarity when comparing just two conditions. Glucose had 289 pairings in total, and shared just 14 with glycerol lactate which had a total of 208 pairings. Similarly, glucose had 15 pairings in common with galactose, which had a total of 351 pairings, and only 11 in common between glycerol lactate and galactose.

5.2.9 The ribosomal DNA repeats interact highly with the rest of the genome

The ribosomal DNA repeats are located on chromosome 12 and are generally considered to occur at a copy number between 150 and 200 in the wild type *S. cerevisiae* genome [10]. Because of the repetitive nature of the rDNA, interactions involving the rDNA are placed into the repetitive results output file by the topography program. This means, that at the *MspI* fragment level, it is not possible to accurately position the repetitive sequences to one position in the genome. However, because the rDNA repeats only occur in one region on chromosome 12, it is possible to describe interactions with them, despite them being repetitive.

In order to ascertain what interaction frequency the rDNA repeats had to achieve to be statistically significant, the rDNA copy number had to be determined. By doing so, it allowed the total number of rDNA *MspI* fragments which were involved in an interaction to be calculated. The rDNA copy number has been calculated previously by aligning the sequence files against a unique region of the genome (*GAL1*) and against a region of the rDNA. The ratio between these two elements was calculated to be 141.7, 123.9 and 137.8 for glucose, glycerol lactate and galactose, respectively (Table 4.5). There are a total of 47 *MspI* fragments which cover the two rDNA repeats present in the reference genome file used to generate the Topography program output. The number of rDNA associated *MspI* fragments which participated in an interaction with the rest of the genome was 47 of a total of 47. However, because the interactions are assigned across two rDNA repeats, in the GCC dataset, the number of *MspI* fragments needs to be divided by two before the total number of rDNA associated *MspI* fragments can be calculated using the previously calculated rDNA copy numbers. This allowed the total number of rDNA defined *MspI* fragments to be calculated as 3,330, 2,911, and 3,237 (rounded to nearest whole number), for glucose, glycerol lactate and galactose, respectively (Table 5.3). Statistical analyses reveal that interactions between the rDNA and the rest of the nuclear genome need to achieve a frequency of at least 3 to be considered statistically significant (Table 5.4, FPR = 0.00044, 0.00152, and 0.00139 for glucose, glycerol lactate and galactose, respectively).

Table 5.3: The actual number of rDNA segments can be calculated from the rDNA copy number.

	Gal1:rDNA	Interacting rDNA segments divided by 2	rDNA @ Copy Number
Glucose	141.72	23.5	3330
Glycerol Lactate	123.86	23.5	2911
Galactose	137.76	23.5	3237

The rDNA copy number is calculated by a Gal1 to rDNA ratio (Table 4.4). The number of interacting rDNA segments is taken from the file which contains repetitive elements. rDNA at copy number is calculated by multiplying the rDNA copy number with the number of rDNA segments in one repeat involved in interactions.

Table 5.4: Interactions with the rDNA repeats occur above experimental noise.

Sample	k	N	p=1/(S1*S2)	P(X≥k)	L	L·P(X≥k) Expected number of false positives	False Positive Rate
Glucose	3	478978	1/(13622*3330)	$1.95x10^{-7}$	45361260	8.83	0.00044
Glycerol Lactate	3	464729	1/(13407*2911)	$2.79x10^{-7}$	37027777	10.88	0.00152
Galactose	3	408626	1/(13515*3237)	$1.35x10^{-7}$	43748055	5.90	0.00139

Let S1 and S2 be the number of fragments interacting with the rDNA and the number of rDNA segments, respectively, which participate in at least one interaction. We calculate the probability P(X≥k) where N is number of observed pairings and p is 1 divided by S1 multiplied by S2, for one specific pairing to occur k or more times. L is then S1 multiplied by S2, being the number of possible pairings and we expect to see L·P(X≥k) pairings occurring k or more times by chance, using the statistical program 'R' and the 'pbinom' function. An acceptable false positive rate is considered to be less than 0.050 or 5 % of interactions. The false positive rates for tRNA-rDNA interactions ranges from 0.00038 and 0.00041 and for k=3, well within the accepted false positive rate.

Many of the pairings under investigation involve an rDNA defined *MspI* fragment and a second repetitive element (*e.g.* Autonomously Replicating Sequences [ARSs], telomeres, *etc.*). The analysis of non-rDNA repetitive elements is unable to be done on an individual basis as the sequences are unable to be uniquely positioned to one region of the genome. For example, transfer RNAs (tRNAs), of the same type, are multi-copy, have the same sequence, and are distributed throughout the genome. An interaction with a tRNA cannot be positioned uniquely to one region of the genome, rather the tRNAs must be analysed as group.

tRNAs have been shown to co-localise with the nucleolus [88, 148], therefore I hypothesised that tRNA-rDNA interactions would be observed in the GCC dataset. The same rationale was employed, as above, for the incorporation of rDNA copy number into statistical calculations to determine whether tRNA-rDNA interactions occur above experimental noise. The *S. cerevisiae* genome complement has a total of 274 tRNAs in the nuclear genome. The number of *MspI* fragments, containing a tRNA, and participating in an interaction with the rDNA was determined to be 282, 282 and 280, for glucose, glycerol lactate and galactose, respectively. The total number of rDNA-tRNA interactions was summed for each condition, with 6169, 6323 and 5612 for

glucose, glycerol lactate and galactose, respectively. Statistical calculations, as performed in Table 4.3, confirmed that interactions between tRNAs and the nuclear rDNA occur above experimental noise (false positive rate between 0.011 and 0.030; Table 5.5).

Table 5.5: Binomial analysis demonstrates that tRNA-rDNA interactions occur above experimental design.

Sample	**k**	**N**	**p=1/(S1*S2)**	**P(X≥k)**	**L**	**L·P(X≥k) Expected number of false positives**	**False Positive Rate**
Glucose	2	7894	1/(282*3330)	$2.15x10^{-5}$	939060	20.17	0.019
Glycerol Lactate	2	10549	1/(282*2911)	$2.95x10^{-5}$	820902	24.22	0.030
Galactose	2	6871	1/(280*3237)	$1.91x10^{-5}$	906360	17.30	0.011

Let S1 and S2 be the number of tRNA and rDNA segments, respectively, which participate in at least one interaction. We calculate the probability P(X≥k) where N is number of observed pairings and p is 1 divided by S1 multiplied by S2, for one specific pairing to occur k or more times. L is then S1 multiplied by S2, being the number of possible pairings and we expect to see L·P(X≥k) pairings occurring k or more times by chance, using the statistical program 'R' and the 'pbinom' function. An acceptable false positive rate is considered to be < 0.050 or 5 % of interactions. The false positive rates for tRNA-rDNA interactions ranges from 0.0117 and 0.0305 for k=2, well within the accepted false positive rate.

The number of tRNAs per chromosome correlates with its length, with the exception of chromosome 12 (Figure 5.11). In order to ascertain if the tRNA interactions with the rDNA were spread evenly across the chromosomes, as would be expected, the total interaction frequency for the tRNA-rDNA interactions were summed for each chromosome, and plotted as a percentage of the total number of tRNA-rDNA interactions (Figure 5.11). Chromosome 16 deviated most from the general trend (highlighted in Figure 5.11).

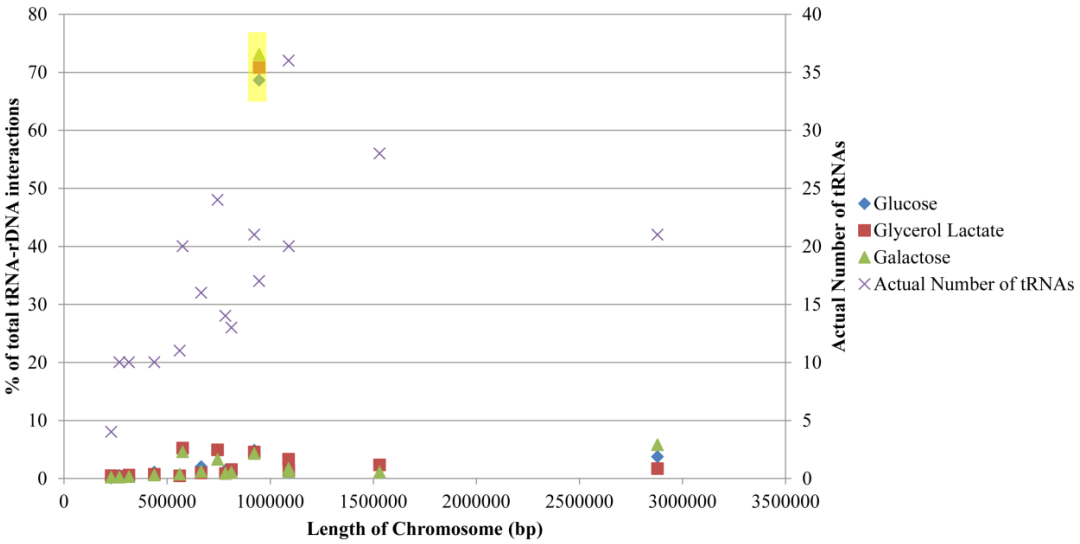

Figure 5.11: The number of tRNA-rDNA interactions correlates with chromosome length with the exception of chromosome XIV.
The number of tRNAs each chromosome has, expressed as a percentage of the total number of tRNAs, is plotted against the chromosome length. The number of nuclear tRNAs (275) was retrieved from the SGD website. Between 280 and 282 different tRNA associated *MspI* fragments participated in interactions with the rDNA (Table 5.5). The number of significant (Table 5.5) interactions the tRNA containing *MspI* fragment (from the tRNA-rDNA interaction subset) was participating in was summed and assigned to its chromosome and plotted against chromosome length as percentage of the total number of interactions for that particular condition. Chromosome 16 (highlighted in yellow) deviated most compared to the other chromosomes, for the three conditions. The 2-micron plasmid and mitochondrial genome were not included in this analysis.

I wanted to establish whether tRNAs interacted evenly across the rDNA repeats, or whether there was a preference for any particular sites. I plotted the number of tRNA-rDNA interactions for *MspI* fragments across two rDNA repeats as percentages of the total number of tRNA-rDNA interactions for glucose (Figure 5.12A), glycerol lactate (Figure 5.12B), and galactose (Figure 5.12C). Strikingly, the tRNAs interact predominantly with an *MspI* fragment situated in what is defined as the inter-genic spacer region (IGS; specifically rDNA non-transcribed spacer 1 [NTS1]) [280].

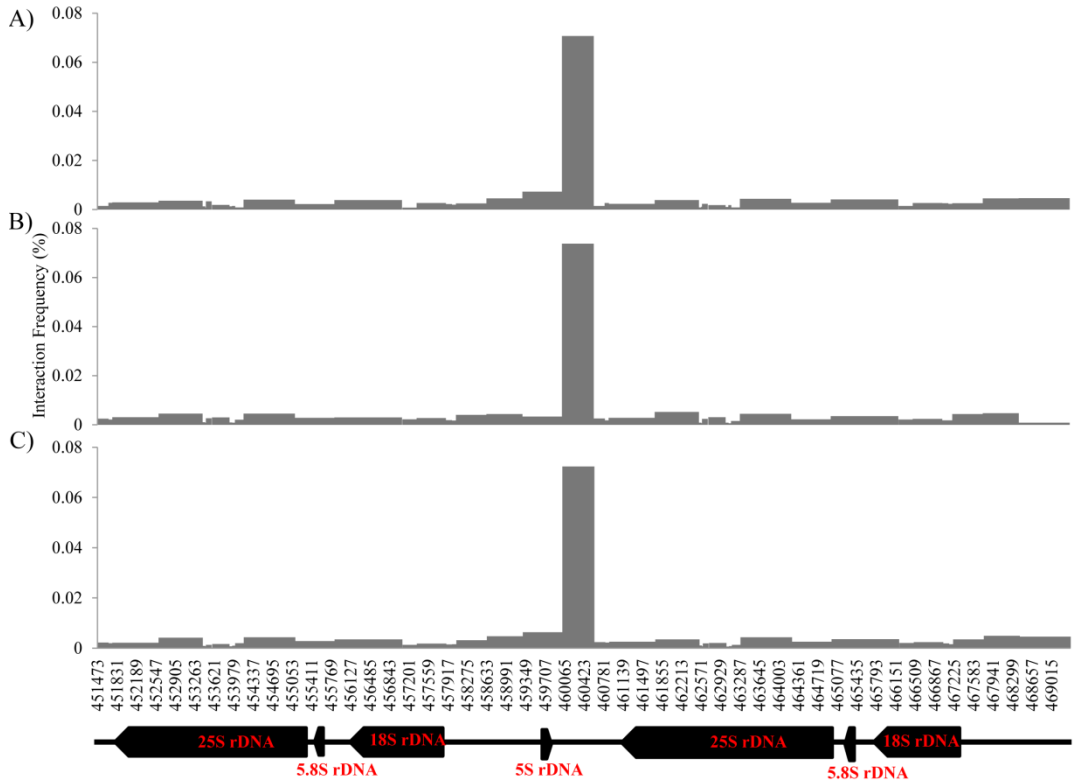

Figure 5.12: tRNA-rDNA interactions do not occur evenly across the rDNA repeats.
Statistically significant (false positive rate ranging from 0.0117 to 0.0306) interactions occurring between *MspI* fragments from the tRNAs and the rDNA repeats are plotted as a percentage of the total number of interactions across two rDNA repeats for each condition; A) glucose, B) glycerol lactate, and C) galactose. The plots have not undergone statistical smoothing to remove length biases. Two rDNA repeats are illustrated across the bottom to allow identification of the rDNA partner fragments.

Because of the observed preference for the tRNAs to interact with the IGS region of the rDNA repeats I isolated all interactions which occur with the IGS region (NTS1 and NTS2; Chr 12; 457,888-460885bp) and broke them into groups according to whether the *MspI* fragment interacting with this defined region contained certain genomic elements (Table 5.6). The IGS region is slightly more connected in glucose and glycerol lactate compared with galactose, as determined by the total number of interactions with this region of 71,052, 75,161 and 52,326, respectively. While not all repetitive elements have been accounted for in Table 5.6 (*e.g.* centromeres, transposable elements *etc.*), the 'Rest of the Genome' category is there to display that it is not only repetitive elements which interact with this region. Overall this analysis demonstrates that the rDNA, and in particular the IGS region, is a highly connected region of the genome, which has been observed previously [88].

Table 5.6: The rDNA IGS region interacts with many different genomic features.

	Glucose		Glycerol Lactate		Galactose	
	Number (total 71052)	**%**	**Number (total 75161)**	**%**	**Number (total 52326)**	**%**
Mitochondria (23.03)	4549	6.40	7710	10.25	1715	3.277
ARS (0.67)	4796	6.74	5631	7.49	4548	8.69
LTR (0.60)	1306	1.84	1340	1.78	1022	1.95
YSCPLASM (2.04)	918	1.29	703	0.93	629	1.20
Telomere (0.76)	137	0.19	223	0.29	60	0.11
Rest of genome	57010	80.23	56477	75.14	42194	80.63

The rDNA inter-genic spacer (IGS) region interacts highly with tRNAs (Figure 5.12). The total number of interactions with the rDNA IGS region (Chr XII 457,888-460885 bp) was calculated from the results files containing interactions involving repetitive elements. The *MspI* fragments interacting with the rDNA IGS *MspI* fragment were separated according to whether they contained, or were derived from, specific genomic features (*e.g.* ARSs, LTRs, 2-micron plasmid [YSCPLASM], telomeres, or the remainder of the genome; percentages these elements comprise of the entire genome complement [in bp] are indicated as percentages next to the names). The repetitive elements could be analysed in this way as there was no concern for their exact position within the genome, only that they were defined as a particular element. The percentage each element consisted of the total has been calculated. The percentages do not add up to 100% as more than one feature can fall within one *MspI* fragment, thus this fragment would be counted twice, once in each feature. The ARS associated with the rDNA repeats is not included in this analysis.

5.2.10 Removal of a tRNA results in a reduced interaction frequency

Upon closer inspection of the tRNA-rDNA interactions it became apparent that one particular interaction was prevalent among all three conditions. This interaction was between an *MspI* fragment on chromosome 16 (581,025-583,522bp) and an *MspI* fragment within the rDNA repeats on chromosome 12 (460,025-460,609bp). In fact, the *MspI* fragment on chromosome 16 interacted highly across the length of the rDNA repeat, not just at this one site. This *MspI* fragment contains the promoter and partial coding region of a ribosomal subunit gene, (*RPA135*), a tRNA (tK[CUU]P), and two uncharacterised ORFs (Figure 5.13). tRNAs have been shown previously to co-localise with the rDNA repeats (Figure 5.11; [148]), therefore this interaction was not a surprise. However, the potential involvement of *RPA135* lent it to further analysis.

3C libraries were created using a *HindIII* restriction enzyme, which essentially cleaves in the middle of the *MspI* fragment, and crucially between the *RPA135* gene and the tRNA to create two fragments (Figure 5.13A and B). 3C primers and a fluorogenic probe (for primer and probes see Appendix I) were designed to assay the interaction between the rDNA *HindIII* fragment (Chr XII 457,910 – 460,634 bp), and each of the two *HindIII* fragments on chromosome XVI (fragment 1, Chr XVI 580,913 – 581,722 bp; fragment 2, Chr XVI 581,723 – 583.733 bp). This assay was done to establish whether the interaction was predominantly forming with the RPA135 gene (fragment 1;

Figure 5.13C) or the tRNA (fragment 2; Figure 5.13D). No significant difference in interaction frequency was observed between fragment 1 and fragment 2 (data not shown).

To determine whether the tRNA sequence was responsible for directing interactions to the rDNA, two yeast strains were assayed by quantitative 3C PCR; a wild type strain (BY4741; Figure 5.13A) and a strain which had the tRNA replaced with a *KanMX* cassette by homologous recombination[9] (yDP97; Figure 5.13B). The frequency of the interaction between the nuclear rDNA and fragment 1 and fragment 2 did not significantly change (Figure 5.13C). This suggests that it is not the tRNA sequence which is driving the interaction with the rDNA. The interaction might be as a result of 'position effect', in that, the region surrounding the tRNA drives its co-localisation with the rDNA repeats.

[9] Gene replacements were carried out by Mr. David Pai at the University of Michigan, Ann Arbor, under the supervision of Prof. David Engelke.

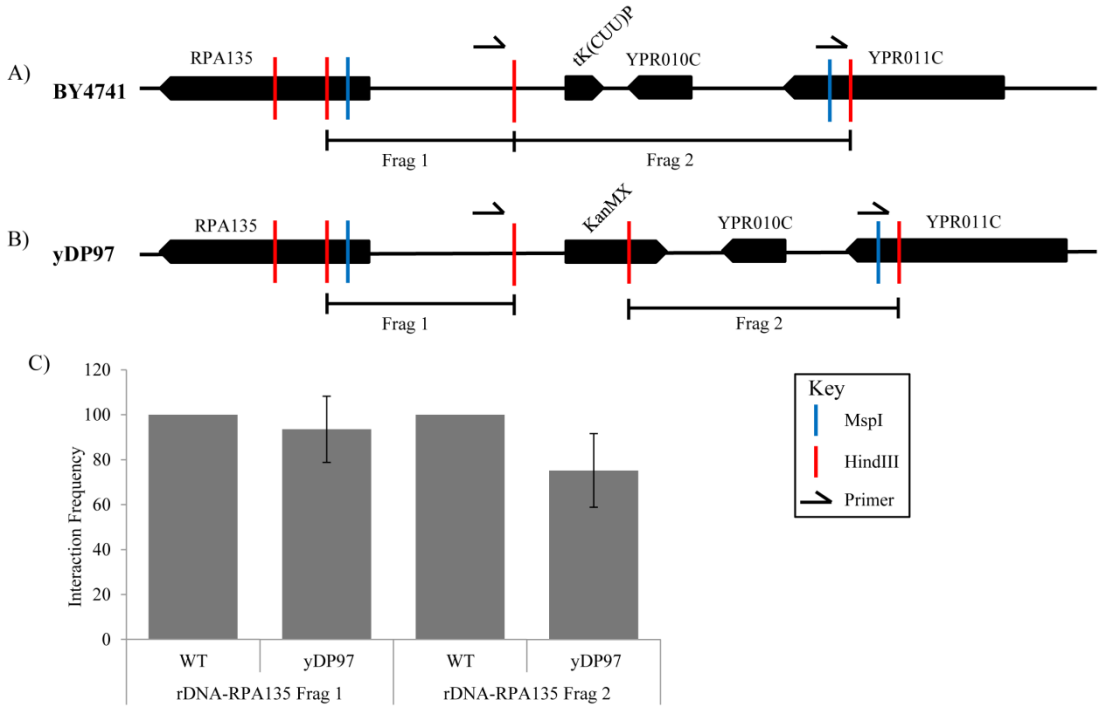

Figure 5.13: Deleting a tRNA off chromosome 16 does not significantly perturb interactions between this region and the ribosomal DNA repeats.
3C primers (rDNAHindIIIF, RT-RPA135Frag1_R, RPA135HindIIIFrag2RorF) and a fluorogenic probe (rDNAHindIIIProbe) were designed to assay interactions between the rDNA (Chr XII 457,910 – 460,634 bp) and chromosome 16 fragments 1 (Chr XVI 580,913 – 581,722 bp) and 2 (fragment 2, Chr XVI 581,723 – 583.733 bp). These primers, coupled with *HindIII* prepared 3C samples and quantitative 3C PCR, were designed to finely map a large *MspI* fragment containing a tK(CUU)P tRNA, a ribosomal subunit gene (RPA135), and two uncharacterised ORFs (YPR010C-A and YPR011C). A) Wild type (WT) *S. cerevisiae* strain (BY4741) with the wild type configuration of the genes. B) The yDP97 strain has had the tK(CUU)P tRNA replaced with the KanMX cassette from the pFA6a plasmid. C) Quantitative 3C PCR demonstrates no significant reduction in interaction frequency for the RPA135 (Frag 1) or tRNA (Frag 2) containing *HindIII* fragments when compared to *S. cerevisiae* (BY4741) wild type (WT; set at 100%). Dedicated standards were produced for the quantitative PCR, as outlined in 2.2.3.10.

5.2.1 Circular Chromosome Conformation Capture (4C) confirms that tRNAs are highly connected to the rest of the genome

Circular Chromosome Conformation Capture (4C) is a powerful technique to detect DNA-DNA interactions between a single restriction fragment and the rest of the genome [21]. It is has been shown that tRNAs cluster [148], therefore, in conjunction with collaborators[10], 4C [21] was used to detect interactions with two separate leucine tRNAs (tL(UAA)B2; Chr II 347,603 – 347,686 bp and tL(CAA)G3; Chr VII 857,491 – 857,378 bp) in a wild type *S. cerevisiae* strain (yPH499) [281]. In addition, 4C was performed on two mutant *S. cerevisiae* strains (*i.e.* yDP77 and yDP84) which had the tL(UAA)B2 or the tL(CAA)G3 replaced with a suppressor tRNA, *SUP4-1*

[10] Collaborators are Prof. David Engelke and Mr David Pai, University of Michigan, Ann Arbor. tRNA replacements were performed by Mr. David Pai.

(tY(GUA)J2), respectively. We predicted that removing the original tRNA would result in the breakdown of tRNA clustering interactions.

I designed nested inverse PCR primers (see Appendix I) to amplify unknown ligation products with each of the two *MspI* fragments (Chr II 347,518 – 349,584 bp; Chr VII 857,273 – 857,825) containing the wild type tRNAs. The replacement SUP4-1 gene contains an *MspI* site, resulting in two *MspI* fragments across each of these regions (Figure 5.14). 4C primers were designed to amplify inversely from both fragments 1 and 2 for the replacement strains (yDP77 and yDP84). The 4C PCRs were performed on the wild type *S. cerevisiae* strain for both of the wild type tRNA positions, as well as the two fragments for each of the mutant strains, resulting in a total of six 4C libraries. The nested PCR primers were designed with 6 bp linker tags, allowing the PCR product from the six separate 4C libraries to be pooled and subjected to 100 bp paired end sequencing on an Illumina Genome Analyser.

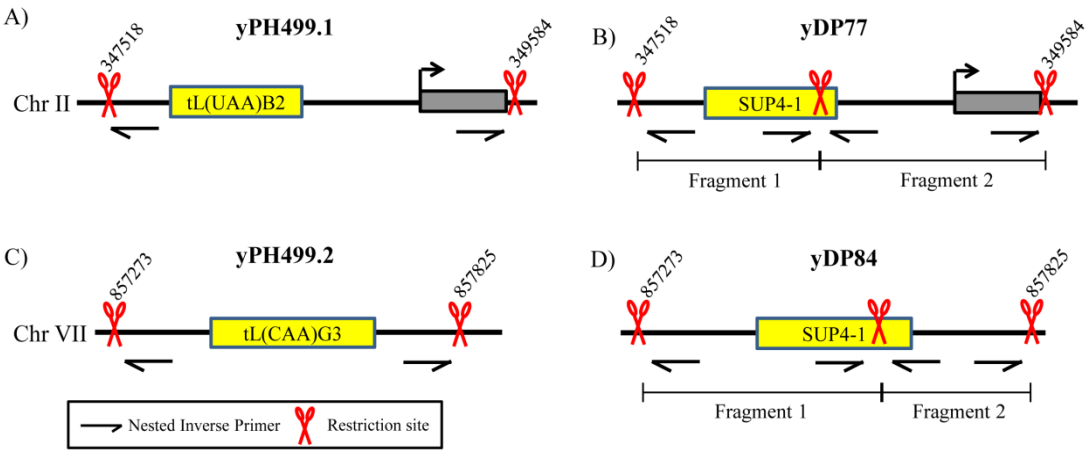

Figure 5.14: Nested inverse 4C primers were designed to amplify from two separate *MspI* fragments in the *S. cerevisiae* genome.
tRNAs, which occur in two different regions of the wild type *S. cerevisiae* genome (yPH499; [A) Chr II: 347,518 – 349,584 bp]; [C) Chr VII: 857,273 – 857,825 bp]), have been replaced with the SUP4-1 suppressor tRNA gene; B) yDP77 and D) yDP84, respectively. Outside and nested inverse 4C primers were designed to amplify out of the two *MspI* fragments with the wild type tRNA intact (A) WTyDP84outsideF, WTyDP84outsideR, C) WTyDP84NestedF, WTyDP84NestedR). The SUP4-1 replacement sequence has an *MspI* site within it, essentially cutting the fragment in two, thus, 4C primers were designed to amplify from both fragment 1 and fragment 2 for both replacement regions (B) Fragment 1: F1MUTyDP77outsideF, F1MUTyDP77outsideR, F1MUTyDP77nestedF, F1MUTyDP77nestedR, and Fragment 2: F2MUTyDP77outsideF, F2MUTyDP77outsideR, F2MUTyDP77nestedF, F2MUTyDP77nestedR. D) Fragment 1: F1MUTyDP84outsideF, F1MUTyDP84outsideR, F1MUTyDP84nestedF, F1MUTyDP84nestedR and Fragment 2: F2MUTyDP84outsideF, F2MUTyDP84outsideR, F2MUTyDP84nestedF, F2MUTyDP84nestedR). For primer sequences see Appendix I. The nested primers had 6 bp tags attached to them so that the PCR product could be pooled and sequenced; A) TCTCTG, B) fragment 1 ACAGAG, fragment 2 TAGATC, C) AGAGAC, and D) fragment 1 TGATGC, fragment 2 AGCACG. ORFs and inter-genic regions are not to scale.

A total of 73,010,074 100 bp sequences were obtained for the pooled 4C libraries, derived from glucose grown *S. cerevisiae*. Each of the individual libraries were isolated

from the sequence files according to its 6 bp tag and primer sequence and trimmed to 34 bp (with the *MspI* recognition sequence in the centre). The sequences for fragment 1 and 2 for each mutant (yDP77 and yDP84) were pooled. The *SUP4-1* replacement gene is from a repetitive gene family, thus, the primer sequences which bound within the *SUP4-1* replacement sequence could not be uniquely positioned upon the reference genome sequence. The reverse primer of fragment 1 and forward primer of fragment 2 were replaced with the forward primer of fragment 1 and reverse primer of fragment 2 so that the bait fragment could be then be uniquely position upon the genome. The pre-processed sequence files were subsequently analysed using Topography v1.19 coupled with the SOAP algorithm [8].

The two wild type tRNA arrangements exhibited similar numbers of interactions with the rest of the genome (Table 5.7). The tRNA replacement on chromosome VII (*i.e.* strain yDP84), had the most interactions associated with it overall (*i.e.* 1,513,787) compared to the wild type arrangement of the tRNA at this locus and the (tL(CAA)G3) tRNA position and its replacement mutant. Determining which interactions occur above experimental noise is difficult with a 4C assay as amplification biases are present, that is, shorter *MspI* fragments will be amplified more than longer ones [282, 283]. Simulations can be performed to gauge how pairings are made at random, and while this was done for the GCC dataset (2.2.4.8), it does not incorporate statistical methods to get around the amplification biases. Therefore, for the purposes of this analysis I will only discuss interactions between the bait *MspI* fragment and the *MspI* fragments in the rest of the genome which occur at a frequency of 10 and above.

Table 5.7: Similar numbers of interactions occur with the wild type tRNAs.

	Total Number of Interactions	**Total Number of Pairings**
yPH499 (tL(UAA)B2)	233,456	683
yDP77 (tY(GUA)J2)	173,546	566
yPH499 (tL(CAA)G3)	250,195	467
yDP84 (tY(GUA)J2)	1,513,787	665

The total interaction frequency was summed for each sample to provide Total Number of Interactions. The total number of pairings refers to the number of connections between the bait *MspI* fragment and other *MspI* fragments in the *S. cerevisiae* genome.

The 4C bait fragment for the yDP84 strain containing the *SUP4-1* replacement sequence interacts more overall than the other strains. Part of this disparity can be explained by a single *MspI* fragment (Chr VIII 219,186- 219,621) which interacted with the yDP84

bait fragment 451,573 times, while the wild type tRNA interacted 565 times. This *MspI* fragment is located within a gene called *MED6* whose protein product is part of the RNA polymerase II mediator complex. The *MspI* fragment adjacent to the *MspI* fragment interacting 451,573 times also interacts 30,471 times in the yDP84 mutant, but not in the wild type at this position. Another RNA polymerase II mediator complex family member, *MED20*, also interacts at high frequency (16,190) with the *SUP4-1* replacement sequence while the wild type tL(CAA)G3 tRNA did not exhibit any interactions with this region.

Given the high number of interactions between these two *MED* gene family members and the *SUP4-1* replacement sequence in the yDP84 strain, I predicted that if the interactions were dependent upon the *SUP4-1* DNA sequence I would observe the yDP77 bait fragment interacting with these *MED* family members also. Indeed, interactions were detected between the *MED6* and *MED20* genes (133 and 386, respectively) and the yDP77 bait fragment containing the *SUP4-1* replacement sequence. The wild type arrangement, containing the tL(UAA)B2 tRNA did not have any interactions with *MED20*, and had only two interactions with *MED6*. This result argues for DNA sequence driven interactions with the *MED* genes, as a result of the *SUP4-1* replacement sequence.

The wild type tRNAs present within both of these bait fragments are leucine tRNAs (tL(UAA)B2 and tL(CAA)G3; Figure 5.14). I predicted that leucine tRNAs would predominantly interact with other leucine tRNAs. Therefore, I isolated interactions with *MspI* fragments that could be uniquely positioned against the genome. No *MspI* fragments interacting with the wild type tL(UAA)B2 tRNA, or the replacement *SUP4-1* (yDP77), contained a leucine tRNA. In actual fact, tL(UAA)B2 paired with two different tRNAs (tQ(UUG)B and tV(CAC)H), and the replacement *SUP4-1* participated in a pairings with the tG(GCC)F2 tRNA. This means that the tRNA interactions were not conserved between the wild type and the mutant. This indicates that while tRNAs cluster, replacing the tRNA sequence results in a change of interaction partners, demonstrating that the tRNA sequence is responsible for the pairings it forms with other specific tRNAs.

On the other hand, the tL(CAA)G3 wild type tRNA interacted with just one other tRNA that could be uniquely positioned; an *MspI* fragment containing the tL(UAA)B2 tRNA

that acts as the bait fragment for the wild type of the yDP77 strain. This interaction was conserved when the *SUP4-1* sequence was present (yDP84). This observation argues against what was presented above, that is, replacing the tRNA sequence does not result in a loss of the interaction with the tL(UAA)B2 tRNA, indicating that for this interaction the tRNA sequence is not responsible for the pairing. The pairing between the two wild type 4C bait fragments was only detected from the tL(CAA)G3 containing *MspI* fragment, and was not captured when the tL(UAA)B2 containing *MspI* fragment was the 4C bait.

Next I investigated whether the *SUP4-1* tyrosine-tRNA replacement sequence interacted with other *SUP* gene family members. Because the primers for the wild type tRNA arrangements can be uniquely aligned against the *S. cerevisiae* genome, interactions with *SUP* family members were deemed to be occurring. The tL(UAA)B2 wild type tRNA had a total number of 504 interactions with *SUP* family members which originated from a total of 8 *MspI* fragments, while the tL(CAA)G3 wild type had a total frequency of 2,092, also originating from a total of 8 *MspI* fragments. It was only possible to uniquely position one of these 8 pairings for the tL(UAA)B2 sample which was with the *SUP6* gene (tY(GUA)F2). No interactions could be uniquely mapped for any of the pairings with *MspI* fragments containing *SUP* genes for the tL(CAA)G3 containing bait fragment. While I was not able to uniquely map the majority of the interactions to one specific point in the *S. cerevisiae* genome, all of the pairings were occurring with suppressor tRNAs of the same family (tRNA-tyrosine). It is not clear from the 4C data whether there was one high frequency pairing with one tRNA-tyrosine gene and due to the inability to uniquely align the frequency was being spread among the other repetitive tRNA-tyrosine genes, or whether the wild type tRNAs were indeed interacting with 8 of the 9 tyrosine tRNAs present in the genome.

Because the *SUP4-1* replacement sequence is repetitive, any interactions which form with other *SUP* family member's may result from the self-ligation of the bait *MspI* fragment or simply sequencing across an uncut 4C bait *MspI* site into the adjacent fragment containing the remainder of the *SUP* sequence. Therefore, I was only able to consider interactions with *MspI* fragments containing *SUP* tRNA genes if the sequence tag was uniquely positioned against the genome, that is, it did not align within the *SUP* tRNA gene, but rather at the opposite end of an *MspI* fragment, in a region which could be uniquely aligned.

Five pairings between the yDP77 strain *MspI* fragment containing the *SUP4-1* replacement and *SUP* family members were able to be uniquely mapped to the genome. These interactions were with SUP2, 4, 5, 6 and 8 (tY(GUA)D, tY(GUA)J2, tY(GUA)M1, tY(GUA)F2, and tY(GUA)M2, respectively), which are all from the tyrosine tRNA family, and more importantly, the same family as the replacement sequence. A further two unique pairings with the yDP84 bait *MspI* fragment containing the *SUP4-1* replacement were detected, which were with *SUP2* and *SUP4*, also of the tyrosine tRNA family. These results are indicative of clustering of the tRNA-tyrosine genes. However, as the 4C libraries were derived from a population of asynchronous cells, it is equally possible that these pairings were occurring independent of each other in different cells, and thus clustering was not occurring. It is not possible to ascertain whether the observed clustering is real or not from the 4C data.

The *SUP2* and *SUP4* pairings were conserved in both mutant strains carrying the *SUP4-1* replacement sequence, that is, the interactions persisted regardless of the *SUP4-1* genomic position. The wild type tRNAs did not participate in unique pairings with the *MspI* fragments containing these two *SUP* tRNA genes, which further indicate the sequence specificity of individual interactions.

I separated the 4C assay results into categories of repetitive elements according to the interacting partner identity (*e.g.* 2-micron plasmid, mitochondria, tRNA, rDNA, ARS, LTR, and telomeres). It is well established that tRNAs cluster at the nucleolus [148], thus, it was not a surprise to detect high frequency interactions between the tRNA containing 4C bait fragments and the rDNA repeats. The pairings were analysed as though there was only one rDNA repeat, as it is not possible to assert exactly which of the possible 200 rDNA repeats the interactions were occurring with. An rDNA repeat consists of 24 *MspI* fragments and 12 of these pair with the wild type tL(UAA)B2 tRNA containing *MspI* fragment and 12 pair with the *SUP4-1* replacement mutant (yDP77), while eight of these pairings are conserved in both the wild type and the mutant. A similar situation plays out for the second wild type 4C bait fragment containing the tL(CAA)G3 tRNA, whereby nine pairings occur with rDNA associated MspI fragments in the mutant, seven occur in the mutant and five of these pairings are conserved between them. This high pairing conservation rate between the wild type and the mutant with the rDNA repeats demonstrates how important it must be for tRNA containing regions to be tethered at the nucleolus. It has been hypothesised that tRNAs

are part of the organising factors that arrange the genome in three dimensions [88, 148, 284], and this high conservation rate supports that hypothesis, as regardless of the tRNA sequence, a high percentage of the pairings with the rDNA prevail.

As mentioned above, I separated the 4C assay results into categories of repetitive elements according to the interacting partner identity. This resulted in a category of interactions which were labelled 'other' and included single copy genes, repetitive gene families, and inter-genic regions. Only 11 of 152 pairings were conserved between the tL(UAA)B2 wild type tRNA, and the mutant yDP77 with the *SUP4-1* replacement, indicating that the original tRNA sequence was important for maintaining these interactions. The 11 interactions which were conserved between the wild type and mutant may form as a result of specific flanking DNA sequence around the tRNA, or as a result of spatial positioning of the partner loci. That is, the partner locus is positioned close to the nucleolus, and is always in crosslinking distance of the 4C bait fragment, regardless of the bait fragments DNA sequence.

The above result was repeated with only 5 of a total of 129 pairings being conserved between tL(CAA)G3 wild type and the *SUP4-1* containing mutant, yDP84. These results suggest again, that only a small percentage of the interactions are occurring due to the flanking DNA sequence. Interestingly, 7 of 138 pairings involving the 'other' category were conserved in both of the mutant strains, indicating that the *SUP4-1* DNA sequence was responsible for directing these interactions, and that the interactions were independent of genomic position.

5.2.2 Interactions which involve inter-genic regions tend to occur further from the ORF start site than expected

The non-coding regions between genes are generally considered to be the sites where transcription factors bind in order to alter the transcription of upstream and/or downstream genes [201]. The formation and maintenance of a number of long range intra-chromosomal interactions in mammalian cells have been shown to be dependent upon two transcription factors, *GATA3* and *STAT6* [72]. These transcription factors mediate interactions between regulatory DNA sequences (*e.g.* enhancers, locus control regions, *etc.*) and the promoter region of genes which they control. With the exception of an enhancer sequence in the rDNA, *S. cerevisiae* is not thought to contain classical long-range enhancer or control sequences [265, 285]. They do, however, have upstream

activation sequences (UAS), which are less flexible in their distance from the transcription start site but can still operate up to ~1200 bp from the promoter [286, 287]. Factors bind within the confines of the inter-genic regions in *S. cerevisiae*, which might mediate inter- and intra- chromosomal interactions. As such, I hypothesised that *MspI* fragments which fell within inter-genic regions and participated in inter- and intra-chromosomal interactions would behave differently than all *MspI* fragments which fall within inter-genic regions but did not necessarily participate in an interaction.

MspI fragments which participated in significant non-adjacent intra-chromosomal or inter-chromosomal interactions and were entirely contained within an inter-genic region of the genome (that is, they did not overlap any defined coding sequence), were isolated from the GCC datasets. The interactions these fragments participated in were broken into three groups, depending on whether the inter-genic regions occurred between divergent genes, tandem genes on the Watson strand, or tandem genes on the Crick strand. A convergent gene conformation is rare in the *S. cerevisiae* genome [288], has an unusually short inter-genic region between the genes [288], and has not been included in the analysis.

The distances from the centres of the interacting *MspI* fragments (fragment centre distance [FCD]) to the ORF start site, in base pairs, was calculated for fragments participating in an interaction (Figure 5.15A, Figure 5.16A, Figure 5.16C). Histograms were plotted (bin size 100 bp) as a percentage of the frequency of all pairings under investigation (Figure 5.15B, Figure 5.16B, and Figure 5.16D). Percentages have been plotted in order to facilitate comparisons of the different conditions. All *MspI* fragments contained within the *S. cerevisiae* genome, which occurred entirely within an inter-genic region, were isolated and separated into the above categories and plotted alongside the data as controls.

Interactions were isolated which occurred between two *MspI* fragment that were both positioned between divergent genes (Figure 5.15A). The distance from the start site of upstream genes (on the crick strand) to the centre of the interacting *MspI* fragment tends to peak at 700-800bp, whereas the distance to the downstream gene start site peaks at ~1000bp (Figure 5.15B). Overall, the *MspI* fragments which are involved in an interaction are significantly (glucose; t-test positive values $p=4x10^{-11}$ and negative values $p=4x10^{-7}$) further from the start site than expected when all *MspI* fragments

which meet the criteria are plotted. No significant difference between the three conditions was observed.

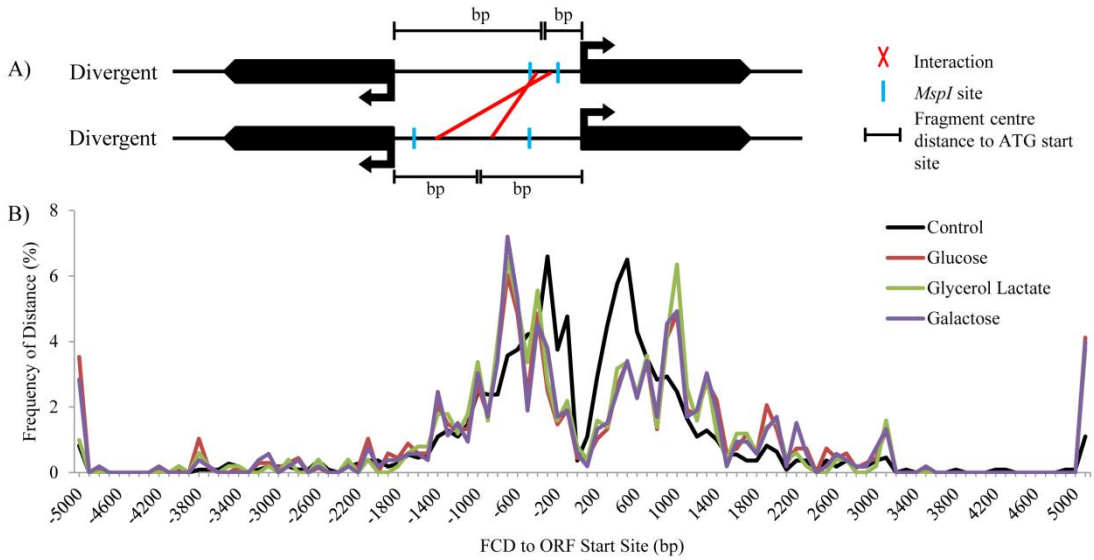

Figure 5.15: Interactions between divergent inter-genic regions are centered 400-1000bp upstream of the ORF start site.
A) Shows the interaction between inter-genic regions falling between divergent genes. Significant interactions which occur between inter-genic regions defined as occurring between divergent genes were isolated. The entire *MspI* fragment had to fall within the inter-genic region, and could not overlap any defined ORF. B) The distance between the centre of the interacting *MspI* fragments (fragment centre distance [FCD]) and the 5' or 3' ORF start site has been plotted in bins of 100 base pairs. The control plot consists of all *MspI* fragments, in the *S. cerevisiae* genome, which fit the criteria of the entire *MspI* fragment falling within an inter-genic region defined as occurring between divergent genes. The histograms are plotted as percentages to facilitate comparisons between the three conditions and also with the control.

Interactions, which occur between inter-genic regions positioned between tandem genes on the Crick strand, were isolated and plotted against a control group of all *MspI* fragments in the *S. cerevisiae* genome which fit these criteria (Figure 5.16B). Only distances to the downstream gene (in the crick orientation) have been included in this analysis to avoid gene length differences if the ORF start site distance was included for the upstream gene. Interactions peaked at ~900bp from the ORF start site in all three conditions. The interactions centred significantly (t-test p=0.0002) further from the ORF start site than the control. Glucose had less interactions occurring closer to the ORF start site than were observed in glycerol lactate and galactose. A similar scenario plays out when plotting interactions between inter-genic regions defined as between tandem genes on the Watson strand (Figure 5.16D). However, the interactions are not significantly (t-test p= 0.351) further from the ORF start site than would be predicted by plotting all *MspI* fragments which fit this criteria (the control plot). There are two interaction peaks at 600bp and 900bp upstream from the ORF start site. It is not clear why the data might

follow this trend, although the control does exhibit a small peak at 600bp upstream from the ORF also.

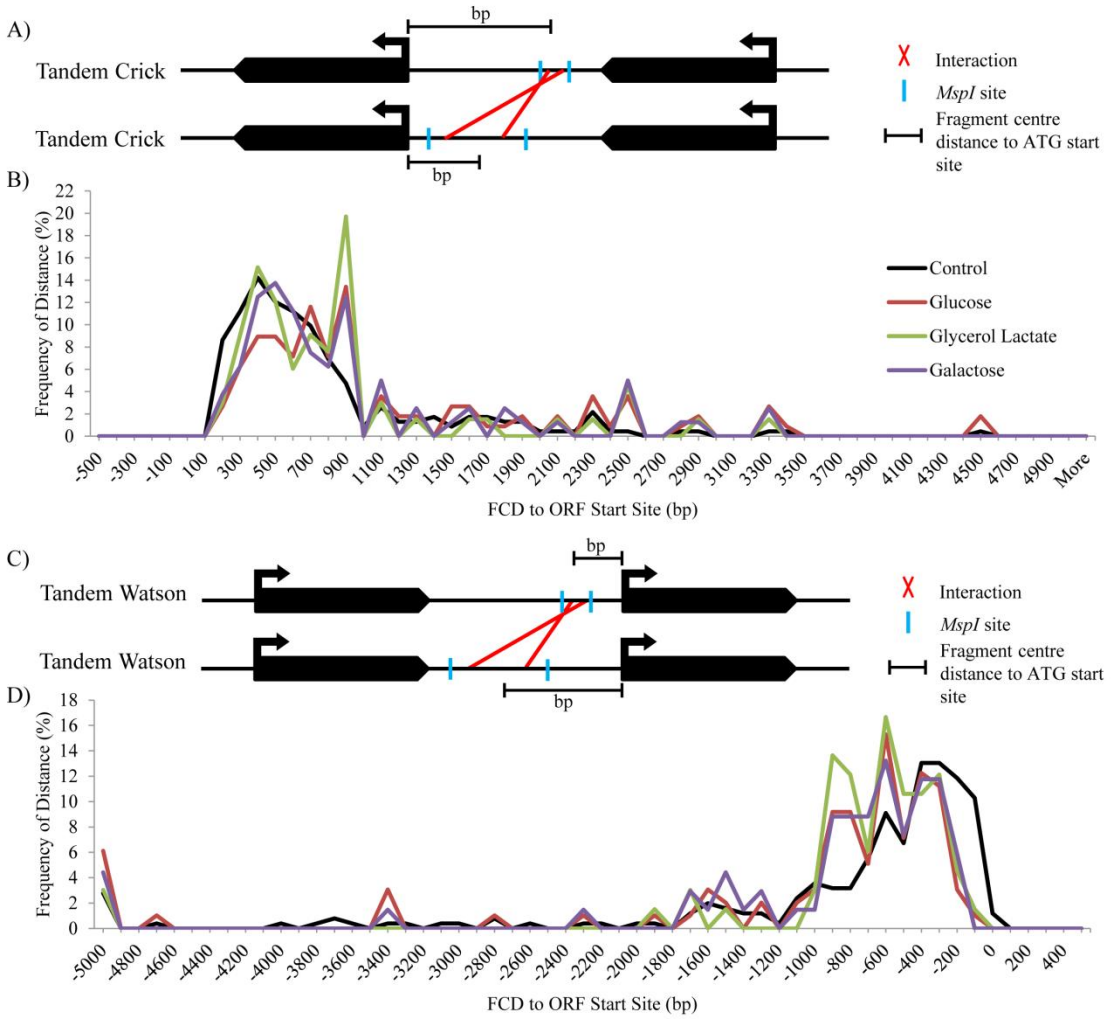

Figure 5.16: Interactions between tandem Watson or tandem Crick inter-genic regions tend to centre 600 and 900 bp upstream or downstream, respectively, of the ORF start site.
A) Depicts an interaction between tandem Crick inter-genic regions. The entire *MspI* fragment had to fall within the inter-genic region, and could not overlap any defined ORF. Significant non-adjacent interactions were isolated which occurred between inter-genic regions between a tandem gene arrangement on the Crick strand. B) The distance between the centre of the interacting *MspI* fragments (fragment centre distance [FCD]) and the 5' or 3' ORF start site has been plotted in bins of 100 base pairs. The control plot consists of all *MspI* fragments, in the *S. cerevisiae* genome, which fit the criteria of the entire *MspI* fragment falling within an inter-genic region defined as occurring between tandem gene arrays on the Crick strand. C) Depicts an interaction between tandem Watson inter-genic regions. Again, the entire *MspI* fragment had to fall within the inter-genic region, and could not overlap any defined ORF. Significant interactions were isolated which occurred between inter-genic regions between a tandem gene arrangements on the Watson strand. D) The distance between the centre of the interacting *MspI* fragments (fragment centre distance [FCD]) and the 5' or 3' ORF start site has been plotted in bins of 100 base pairs. The control plot consists of all *MspI* fragments, in the *S. cerevisiae* genome, which fit the criteria of the entire *MspI* fragment falling within an inter-genic region defined as occurring between tandem gene arrays on the Watson strand. The histograms are plotted as percentages to facilitate comparisons between the three conditions and also with the control.

5.2.3 The connectedness of genes involved in specific pathways changes depending upon the metabolic condition of the cell

Sets of genes are transcribed to different levels under different conditions [29]. There are a multitude of different biochemical pathways in *S. cerevisiae* which respond to environmental stimuli. Accordingly, the transcription levels of the genes in these pathways alter depending upon the environmental stimuli. Different metabolic pathways are utilised during growth in glucose, glycerol lactate and galactose. Therefore, I hypothesised that differences in the number of interactions that genes in a pathway participated in, might tell us something about the nature of the connection.

Three lists of genes, that are involved in specific metabolic and cellular processes (glycolysis, gluconeogenesis, and rDNA coding sequences), were downloaded from Gene Ontology (GO; SGD website; Appendix CD). A Perl program [11] (region_interactions.pl) was used to proportionally assign interaction frequencies to the genes in the GO lists. The Perl program is fed the coordinates of the genes in question. *MspI* fragments can be associated with these coordinates, either entirely, in which the entire frequency is linked to the gene, or partially, at which point the interaction frequency is proportionally assigned to the gene according to the percentage of overlap. The total interaction frequency is calculated for each gene in the list, and comparisons made between the three conditions. Only significant non-adjacent intra-chromosomal and inter-chromosomal interactions were included in this analysis.

There are a number of assumptions which underlie this type of analysis. GCC does not report upon the exact position of the interaction, we are only able to narrow the interaction point down to a single *MspI* fragment. Therefore, we have to assume the interaction frequency is being spread across each base pair in the *MspI* fragment. The actual interaction point might be outside of the gene coordinates we are presenting the program, however, without further experiments this is not possible to determine. The interaction point could be more finely mapped if a second, different, restriction enzyme were used to create an interaction map of the same sample.

The interaction frequencies of genes have been plotted for glycolysis (Figure 5.17A), gluconeogenesis (Figure 5.17B), and nuclear rDNA coding sequences (Figure 5.17C). Many of the GO lists contain too many genes to be plotted as bar graphs.

[11] This Perl script was written by Dr. Lutz Gehlen, Institute of Natural Sciences, Massey University.

Glycolysis is the process by which glucose is converted into pyruvate [289]. Therefore, we would expect this pathway to be most active during *S. cerevisiae* growth on glucose and perhaps slightly reduced in galactose, due to the slower growth rate of *S. cerevisiae* in galactose [145]. Overall, glucose had a significantly higher number of interactions with the glycolysis genes, than glycerol lactate and galactose (Figure 5.17A, t-test p=0.000001 and 0.001, respectively). However, the *GLK1* and *MDH3* genes do not follow this trend, with galactose reduced and glycerol lactate and glucose having similar high interaction frequencies. The *GLK1* gene product is involved in the irreversible first step of glucose metabolism, the phosphorylation of glucose on C6 [289]. I predicted that gene activity would correlate with the number of interactions. However, *GLK1* is repressed in glucose and de-repressed in glycerol lactate [290]; thus, it is not clear why the glucose and glycerol lactate conditions have similar number of interactions. *MDH3* is present in the mitochondria, peroxisomes, and the cytoplasm where it converts malate to oxaloacetate as part of the TCA and glyoxylate cycles. Its expression is slightly enriched during respiratory growth and growth in galactose, but not glucose. Again, no obvious correlation between the interaction frequency of the gene and the expression level can be drawn.

The gluconeogenesis pathway is largely the reverse of glycolysis and is the generation of glucose from non-carbohydrate carbon substrates such as glycerol and lactate [145]. As a consequence of this pathway reversal, gluconeogenesis and glycolysis share many of the same genes [145]. Glycolysis and gluconeogenesis inhibit each other with a reciprocal control system, thus in glucose and galactose growth glycolysis is utilised, while the gluconeogenesis pathway is utilised during glycerol lactate growth. I attempted to find a positive correlation between the interaction frequency of the gluconeogenesis genes and gene activity. Because many of the genes associated with gluconeogenesis are also active during glucose and galactose growth I predicted that the interaction frequencies would be similar between the three conditions. I plotted the interaction frequencies for the genes associated with gluconeogenesis (Figure 5.17B), and observed a significant (t-test p= 0.003 and 0.001 for glucose-galactose and glucose-glycerol lactate comparisons, respectively) enrichment of interactions during glucose growth. The *PYC2* gene did not adhere to this trend, with high interaction frequencies in glucose and glycerol lactate, compared with galactose. The *PYC2* gene is not involved in glycolysis, and I observed a higher interaction frequency in glycerol lactate. *PYC1*

did not exhibit this trend. Again, I did not observe any consistent correlation between interaction frequency and gene activity.

The rDNA repeats are known to be most active when cells are growing at fast rates [291]. The ribosome content is proportional to the cellular growth rate [292], such that the growth in non-fermentable carbon sources (*e.g.* glycerol), is associated with a reduction in ribosome content [293]. Taking into consideration the above trends for glycolysis and gluconeogenesis, one might expect a significantly higher interaction frequency for the rDNA coding genes during glucose growth. On the contrary, there is a significantly (t-test p=0.008) higher number of interactions with rDNA coding genes in glycerol lactate compared with glucose, while there is no statistically significant difference between glucose and galactose (t-test p=0.195). The interactions included in this analysis were from the output file containing repetitive elements.

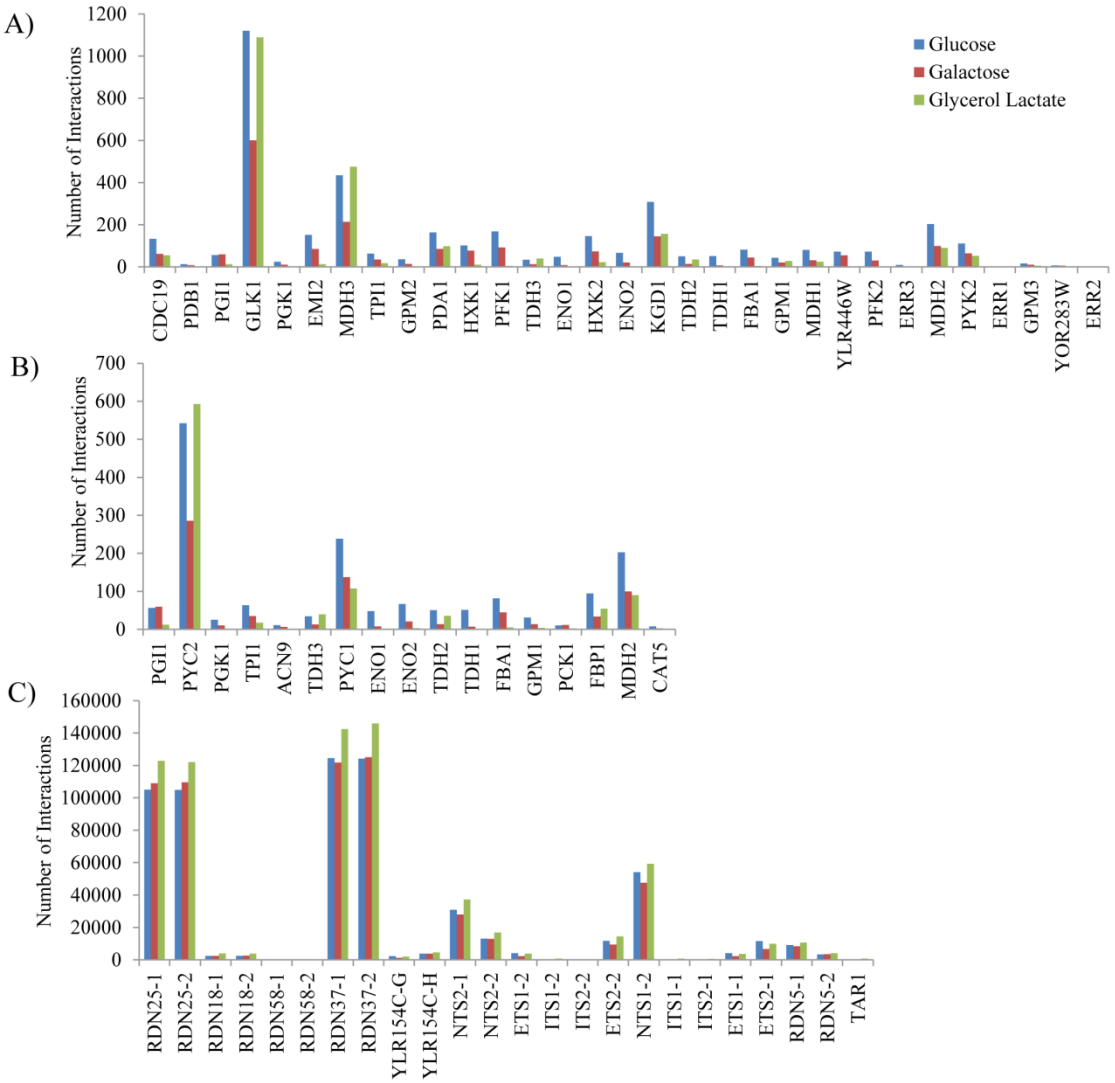

Figure 5.17: The interaction frequency associated with ORFs does not correlate with the ORFs transcriptional activity.
The frequency of statistically significant inter- and intra- chromosomal interactions has been proportionally assigned across the genes for, A) glycolysis related genes, B) gluconeogenesis related genes, and C) nuclear rDNA coding sequences. This proportional assignment of interaction frequency is achieved by calculating the percentage of each *MspI* fragment that overlaps an ORF, and subsequently assigning that percentage of a particular interaction to the ORF. The interaction frequencies have been corrected for the number of sequence tags for each condition to normalise the data.

5.3 Discussion

GCC was performed on non-synchronised exponential phase *S. cerevisiae* cells grown on glucose, glycerol lactate and galactose carbon sources. Advances in sequencing technology allowed the use of high-throughput paired-end sequencing, as opposed to the single-end sequencing originally used in chapter 3.2.6. Paired end sequencing (2.2.3.7) of the GCC libraries increases the likelihood of detecting an interaction. This is because it is no longer a requirement for the restriction enzyme site to fall within the sequence tag. With paired end sequencing, each sequenced 'end' is aligned against the reference genome and if those two 'ends' align to two different *MspI* fragments, then an interaction is recorded. This analysis assumes that the two *MspI* fragments were joined together during the ligation step in the GCC protocol.

5.3.1 The genome of *S. cerevisiae* is the most connected when grown in glucose

Inherent differences between the interaction networks for the three conditions were immediately apparent from the graphical interaction network map output (Figure 5.1). Glucose growth resulted in the most connected genome overall, especially when compared with the galactose network map. It is not immediately obvious why this connection disparity exists between these two conditions, as both conditions employ respiro-fermentative metabolism. Glucose is a repressive phenotype, such that the synthesis of enzymes required for the utilisation of other carbon sources is inhibited [262]. In contrast, galactose is a non-repressive phenotype [13, 50, 262]. Whether catabolite de-repression alone can account for the lower frequency of interactions in galactose, at the inter- and intra- chromosomal interaction level, remains to be determined. The glycerol lactate growth condition is the most different metabolically from the other two conditions, with a solely respiratory metabolism [215], however, its connectedness overall largely matched that of glucose. It was the subtle differences in the types of interactions, which would prove to be the major difference between the three conditions.

The number of inter-chromosomal interactions correlated with chromosome length for the three conditions, with the exception of chromosome XIV. Oddly, chromosome XIV does not contain any extra features, repetitive or otherwise, which might explain this deviation. The chromosomes in higher eukaryotes tend to have preferred positions within the nucleus; either in the interior or to the periphery of the nuclear volume

termed chromosome territories [62]. While yeast chromosomes have not been observed to form territories, it is possible that chromosome XIV exists largely in the interior of the nucleus during the non-dividing phases of the cell cycle, surrounded by the remaining chromosomes. This might lead to more inter-chromosomal interactions, as one 'interaction face' of the chromosome would not be on the periphery of the nucleus unable to interact. Three-dimensional modelling of the *S. cerevisiae* genome from the data generated in this thesis may shed light on chromosome XIV's position within the nucleus and provide clues for the higher number of inter-chromosomal interactions, compared with the other chromosomes.

5.3.2 Intra-chromosomal interactions are hypothesised to form as part of general DNA packaging

Glucose had the highest number of intra-chromosomal interactions overall. Moreover, the number of intra-chromosomal interactions correlates well with chromosome length. As mentioned earlier, the enhanced transcription of ~1100 genes is required during growth on glycerol lactate [168]. Nucleosome depletion has been observed around the sites where transcription factors bind [272], thus the genome in respiring cells might be in a more open conformation than the genome in glucose growing *S. cerevisiae*. The intra-chromosomal loop sizes I observed were generally less than 10,000 bp in length, therefore, in theory, these loops could be forming as a result of general chromatin compaction of inactive regions of the genome [17, 18, 47, 50]. This might explain the higher number of intra-chromosomal interactions observed in the glucose dataset.

Unfortunately, it is not this simple; intra-chromosomal loops do not only form during DNA compaction, but also form around active genes, essentially connecting the promoter and terminator regions [186, 294]. This type of conformation is hypothesised to promote re-initiation of RNA polymerase II back onto the promoter [186, 294]. By performing GCC with different restriction enzymes on the same sample the point of contact for an interaction can be mapped on a fine scale. This would allow gene looping to be investigated by GCC. Coupling GCC with chromatin immuno-precipitation [295, 296] against the phosphorylated active form of RNA-pol II would result in only active genes being isolated. If high numbers of intra-chromosomal interactions were observed in these samples, it might shed light upon the intra-chromosomal interactions observed in the current GCC datasets.

Alternatively, if we look at this issue from a growth rate perspective, we notice that *S. cerevisiae* grows much faster in glucose than in glycerol lactate or galactose. The GCC networks were created from populations of cells, and consequently, differences in the numbers of intra-chromosomal interactions might be born out of cell cycle phase effects. Cells which have a slower growth rate spend a proportionally longer time in the G0/G1 non-dividing phases of the cells cycle [297, 298]. Glucose would, therefore, have a greater number of cells in the dividing phases when the chromosomes are most compacted, as it has a high growth rate [299]. This may account for the greater number of non-adjacent intra-chromosomal interactions seen in glucose. There is currently no data which describes global changes to inter- and intra- chromosomal interactions throughout the cell cycle in *S. cerevisiae*.

5.3.3 Do the *S. cerevisiae* chromosomes form into territories?

Chromosome territories form in higher eukaryotes [62], but have so far eluded detection in *S. cerevisiae* [76, 83, 300]. The organisation of chromosomes into territories has implications for the types of interactions that would form within the nucleus. Theoretically, territories would result in a preference for interactions within a single chromosome as opposed to interactions between chromosomes. This preference was observed, using Hi-C, for human chromosomes [90], which do form territories [62]. Lieberman-Aiden *et al.* concluded that their data supported the existence of chromosome territories. GCC detected a greater number of intra- as opposed to inter-chromosomal interactions in the *S. cerevisiae* datasets. This might suggest some form of yeast chromosome territoriality. However, previous observations, in which double stranded breaks of the yeast chromosomes resulted in subsequent promiscuous end-joining of the DNA breaks conflicts with this hypothesis [37]. Indeed, GCC was performed on an asynchronous population of cells, thus a high number of intra-chromosomal interactions might be explained by the condensation of the chromosomes during cell division. Interestingly, Hi-C was performed on an asynchronous population of human cells as well [90]. GCC performed on *S. cerevisiae* cells arrested in the non-dividing phase of the cell cycle would provide a clearer picture of the types of interactions which predominantly form in the absence of chromosome condensation.

5.3.4 The yeast chromosomes do not adhere to current polymer folding models

The lack of measurable chromosome territories in yeast, lead me to hypothesise that the folding of the *S. cerevisiae* chromosomes would not adhere to a fractal globule arrangement previously shown for Human chromosomes (slope exponent of -1; [6, 90]). Intra-chromosomal loops less than 50,000 bp in size display prominent power law scaling when the frequency of loops is plotted log-log against loop size (in 5000 bp bins) for the three different conditions. Slope exponent values of -3.629, -2.617, and -2.159 for glucose, glycerol lactate and galactose, respectively, support my hypothesis. These values push it well beyond even the equilibrium globule (slope exponent of -1.5 [6]), and therefore I am unable to state exactly what folding characteristics the *S. cerevisiae* chromosomes have.

In many fields the connectivity of large complex networks has been observed to follow (scale-free) power law behaviour [301]. If we think of the intra-chromosomal interactions in the GCC dataset as a network, the slope exponents for large complex networks are similar to those observed for the intra-chromosomal interactions in the glucose, glycerol lactate and galactose GCC datasets [301]. Barabasi *et al*. (1999) explained these large complex networks due to the growth of the network by the addition of vertices over time and the preference for new connections to form with highly connected nodes [301]. Indeed, the GCC data follows this same behaviour, as genetic networks such as the *S. cerevisiae* genome have increased in size over the course of evolutionary time due to the flow of genes from the mitochondria and gene and whole genome duplications [15, 302, 303]. The adherence of the power law slope exponents of the GCC data to the complex network modelling implies that there is a preference for highly connected nodes. This would be consistent with a genome rosette structure with an internal well connected node.

5.3.5 Adjacent interactions can inform upon local chromatin structure

Adjacent interactions are detected by sequencing across the *MspI* site of two neighbouring *MspI* fragments. Since restriction enzyme digestion is not 100% efficient due to site accessibility [199] sequencing across an uncut restriction site results in a reported adjacent interaction in the GCC dataset. However, not all adjacent interactions form as a result of inaccessible cleavage sites, rather they can form as a result of the two

adjacent restriction fragments being held together by some other means, and thus promoting the ligation of the cleaved restriction site.

Active regions of the genome are characterised by open, nuclease-sensitive regions of the genome [304]. Adjacent interactions could then, theoretically, report on local chromatin structure, such as inactive, compacted regions of the genome. Therefore, I plotted the adjacent interaction frequency across two short regions of the genome, the *GAL* and *DAL* loci, to explore whether adjacent interactions provide any information about the local chromatin structure.

The *GAL* genes are required for galactose degradation and utilisation as a carbon source, thus the *GAL* operon is repressed during glucose growth, in a poised non-repressed conformation in glycerol lactate, and active in galactose. Glucose had a significantly higher number of adjacent interactions across this region than galactose, while glycerol lactate and galactose did not differ significantly, as expected. The genes required for allantoin degradation also occur as a cluster called the *DAL* operon. The transcription of the DAL genes is not overly different between the three conditions [29]. Therefore, I predicted and observed no significant difference in the numbers of adjacent interactions across the DAL operon. I conclude that GCC can report upon local chromatin structure, although assumptions (*e.g.* that adjacent interactions are as a result of restriction site inaccessibility) have to be made about the nature of the formation of adjacent interactions in the GCC dataset.

5.3.6 tRNAs cluster with the rDNA

The nucleolus forms around the rDNA repeats at one end of the nucleus [86] reviewed in [305]. Yeast tRNAs have been previously shown to cluster at the nucleolus [88, 147, 148, 306]. In agreement with these previous observations, tRNA-rDNA interactions were detected at high frequency in the GCC dataset. The number of tRNA-rDNA interactions correlated with chromosome length, with the exception of chromosome XVI which deviated from the trend line and interacted much more highly than would be expected for its length (Figure 5.11). A single tRNA containing *MspI* fragment on chromosome XVI (581,025 – 583,522 bp) was responsible for this deviation, with high interaction frequencies with the rDNA repeats in all three conditions.

The tRNA containing *MspI* fragment on chromosome XVI is 2,497 bp in length with the lysine-tRNA essentially in the middle of the fragment. Because of the nature of

GCC, which includes a nebulisation step prior to sequencing (shearing to ~180 bp); this tRNA is not close enough to the *MspI* site to establish this fragment as repetitive. As such, this fragment is considered unique as the sequence tags which map to it, map in the unique regions close to the *MspI* sites. This *MspI* fragment also contains the promoter and partial coding region of the ribosomal subunit gene, *RPA135*. A subsequent 3C library was generated using the *HindIII* restriction enzyme to more finely map this region and the interactions it was involved in. Deleting the tRNA[12] within the *HindIII* fragment did not perturb the interaction with the rDNA repeats to any significant degree. Similarly, the adjacent *HindIII* fragment containing the promoter and partial coding region of *RPA135* was not perturbed by the tRNA deletion. Therefore, I conclude that the interaction of the tRNA containing *MspI* fragment with rDNA is not forming as a result of the tRNA sequence itself. It is possibly forming as part of a regional position effect, or as a result of the *RPA135* gene. It is an exciting prospect for a ribosomal subunit gene (*RPA135*) to be interacting with the rDNA repeats. This interaction may form as part of a co-regulation scheme between the ribosomal subunit gene *RPA135* and the rDNA repeats. This interaction will require further experiments to elucidate its biological context.

Possible experiments could include removing the RPA135 gene from the chromosome and expressing it off a plasmid. The chromosomal region, lacking *RPA135*, could be assayed for the rDNA interaction, as could the plasmid. I predict that the plasmid would interact with the rDNA repeats, while the chromosomal region would interact less frequently.

5.3.7 tRNAs are highly connected to the rest of the genome

It is currently unclear whether DNA-DNA interactions form due to specific DNA sequence or as a result of local or global three dimensional genome structure bringing two regions of the genome in close contact. I performed a 4C analysis of two regions of the genome which contained tRNAs, and in strains where the tRNAs had been replaced with a reporter suppressor tRNA, *SUP4-1*. Overall, little conservation of the interactions existed when the original tRNA was replaced with the *SUP4-1* replacement sequence, indicating that, at least for the regions under investigation, the tRNA sequence itself was responsible for driving the interactions. This implied that the sequence surrounding the

[12] tRNA replacement was performed by Mr. Dave Pai, under the supervision of Prof. David Engelke.

tRNA gene did not contribute extensively to the observed interactions. Obviously not all of the interactions altered when the wild type tRNA sequence was replaced. A small percentage of interactions were conserved between the wild type and mutant strains. These interactions could be as a result of; 1) the DNA sequence flanking the tRNA is the point where the interaction occurs, or 2) the spatial localisation of the particular region. Pairings which occurred with the *SUP4-1* replacement sequence, regardless of its genomic position (*i.e.* yDP77 and yDP84), supported the hypothesis that interactions were sequence dependent.

RNA polymerase III transcription has been shown to occur in discrete foci, thus, Pol III transcribed genes cluster at these discrete foci [307]. These genes include the tRNA genes, previously shown to cluster with one another [148], and the 5S rDNA gene, which exists in the nucleolus where the tRNA genes have been shown to co-localise [148]. The wild type tRNAs and the *SUP4-1* replacement were observed to interact with tyrosine suppressor tRNAs and other tRNAs distributed throughout the genome, possibly suggesting a clustering of these elements in line with previous observations [148]. Furthermore, high frequency interactions between the tRNA containing 4C bait fragments and the rDNA were also observed in agreement with previously published results [88, 148].

tRNA genes along with the promoter *B-box* element, have been shown to function as chromatin barrier elements, that is, they halt the spread of heterochromatin [308]. It is currently unclear exactly how this barrier function is exerted but it is known that the transcriptional competence of the tRNA gene is required. It has been hypothesised that DNA loop formation into higher order chromatin structures, and local chromatin segregation may impart this barrier function, and thus it is not a surprise that tRNAs and *B-box* elements have been implicated in three dimensional genome organisation [309]. tRNAs have a repressive effect on neighbouring Pol II transcribed genes [310] and disruption of tRNA clustering at the nucleolus results in the loss of silencing of these nearby Pol II transcribed genes [311, 312]. Consequently, it is not a surprise to find that interactions between the bait tRNAs and the rDNA repeats are conserved in the wild type and the mutants regardless of the tRNA identity. This suggests that while the replacement *SUP4-1* sequence is also a tRNA, surrounding chromatin ensures co-localisation of these regions to the nucleolus in order to silence neighbouring Pol II transcribed genes and maintain three dimensional genome organisation.

It is not immediately clear why such high frequency interactions occur between the *SUP4-1* replacement sequences and the *MED6* and *MED20* genes. However, as mentioned above, the tRNAs can silence nearby chromatin, thus, physical interactions with tRNAs may result in silencing, or at the very least, regulate the transcription of the *MED6* and *MED20* genes which are required for the activation of Pol II promoters [313]. Alternatively, the MED genes may interact with the tRNAs in order to be co-localised with the nucleolus. Given the importance of Pol I and III transcription on cellular growth rates, it may be a chance for the cell to regulate important Pol II enzyme subunits in accordance with Pol I and III activity.

5.3.8 Interactions between inter-genic regions are centred further from the ORF start site than expected

In higher eukaryotes long range interactions occur between inter-genic regions in order to coordinate the expression of genes [21, 22, 27, 96, 106, 137, 314]. I wanted to determine whether interactions took place between inter-genic regions of the *S. cerevisiae* genome, and whether they differed in location to what would be expected. Because the *MspI* fragments are all different sizes, we are unable to pinpoint exactly which base pairs are involved in the interaction, therefore we calculate the distance to the ORF start site from the centre of the interacting *MspI* fragment. Indeed, I found that interactions between inter-genic regions tend to occur further from the ORF start site than a control, for all conformations of inter-genic regions (defined by surrounding genes; divergent, tandem Watson and tandem Crick).

This analysis, while demonstrating that interactions do occur further upstream, is perhaps too rudimentary to point towards any biological significance. However, the sites which are participating in an interaction may have specific binding proteins which are mediating the interactions. In keeping with this, the binding location of the transcriptional repressor, CTCF, has been shown to correlate with interaction contact points for human chromosomes [5, 72]. Proteins mediating yeast interactions could be general transcription factors, or more pathway specific transcription factors such as Rds2 for gluconeogenesis [315], or Gal4 in galactose metabolism [316]. Further experiments coupling 3C and ChIP [295] to these targets could be performed to elucidate whether they were involved in mediating specific interactions. Alternatively, coupling an immuno-precipitation step, with tagged DNA probes, against the DNA

region of interest, with subsequent mass-spectrometry of bound proteins, would eliminate the requirement for prior knowledge of the mediating factors.

5.3.9 Intra-mitochondrial interactions may be indicative of mitochondrial genome clustering or a novel regulatory system

The mitochondrial organelle is essential during solely respiratory metabolism, as well as fermentative growth [215]. Most of the mitochondrial genes have been transferred to the nucleus over evolutionary time, such that, the mitochondria, in order to function, require the expression of some 1100 nuclear encoded genes ([254, 273, 317], GO SGD). The genes remaining on the mitochondrial genome are mostly required for oxidative phosphorylation; the ATP producing pathway in the mitochondria [149, 266]. As one would expect, during respiration, the mitochondrial genes are very active [230]. If we assume that a high number of intra-chromosomal interactions are indicative of inactive DNA, then we should have observed a lower number of intra-mitochondrial genome interactions during respiratory growth, when compared with glucose. Strangely, this was not the case; glycerol lactate had a greater number of intra-mitochondrial interactions (Figure 5.3).

While I did not observe large differences in the mitochondrial copy number between the three conditions in the GCC datasets (Table 2.7), it is well established that mtDNA content increases during respiratory growth [41]. This might explain the higher number of intra-mitochondrial interactions detected in glycerol lactate. Alternatively, the observed intra-mitochondrial interactions may be occurring within the nucleus between already transferred cDNAs. The intra-mitochondrial interactions may be forming due to clustering of these DNA molecules at the nucleolus. To test if this were the case, nuclei and mitochondrial organelles could be isolated separately by sucrose gradient, the separate fractions could be subjected to a 3C preparation and intra-mitochondrial interactions assayed using 3C primers. Alternatively, performing FISH with probes against the two different regions of the mitochondrial genome [318] might result in florescence from just one of the organelles, indicating in which the intra-mitochondrial interactions are taking place.

It would be an interesting prospect to have transferred mitochondrial cDNAs interacting with each other within the nucleus. As I showed, the mitochondrial cDNAs can alter the transcript levels of genes which they interact with. If clustering of the mitochondrial

cDNAs was observed within the nucleus it might be analogous the RNA granules [319], whereby mRNA is clustered into a granule ready to be released and translated. Often this mRNA has been encoded from stress related genes [320]; releasing the mRNA in response to a stress stimuli, to allow translation, ensures a quick response to the cellular stress. Furthermore, mRNA not required to overcome the cellular stress has been shown to be sequestered to stress granules. The appropriate expression of genes required for mitochondrial function might be controlled by an analogous system. The nucleolus has previously been shown to sequester factors required for other nuclear processes as a form of regulatory management [321].

5.3.10 The *S. cerevisiae* genome is a dynamic structure

The generation of chromosomal interaction networks for *S. cerevisiae*, under three different growth conditions, has shown that genome structure is dynamic. The numbers and types of interactions were different between the three conditions, and these differences may very well support the specific growth phenotypes.

It has been hypothesised previously that repetitive DNA elements, and in particular the rDNA repeats, help to form the structure of the genome [88]. GCC detected extensive interactions between the rDNA repeats, and other repetitive elements spread throughout the genome, in agreement with these earlier observations. Furthermore, tRNAs were also found to interact at high frequency with the rDNA repeats using 4C. GCC and 4C are unable to report on the possible mechanism by which repetitive elements are hypothesised to guide genome structure, although, the numbers of interactions occurring with the rDNA IGS region were similar in all three conditions, suggesting that these interactions were forming regardless of the growth phenotype. It is possible that repetitive DNA elements interact with each other and other specific regions of the genome (*e.g.* the nucleolus) in order to form the primary or base structure of the genome. Further rearrangements might tailor the structure to specific growth phenotypes.

5.4 Conclusion

Interpreting the GCC datasets as a whole is very difficult as the networks are very large, and while we might be able to draw conclusions about certain genomic elements, ultimately the data raises more questions than it answers. GCC can provide clues as to how interaction frequencies change during different conditions, but more in depth experimental work must be undertaken to elucidate any biological role of the individual interactions in question.

Overall, the chromosomal interactions detected in the *S. cerevisiae* genome were shown to change according to growth condition, indicating that the structure is dynamic and can respond to the growth phenotype.

Chapter 6 Final Remarks and Future Directions

Here I presented a novel method, Genome Conformation Capture (GCC) [5], which allowed the global detection and quantification of DNA-DNA interactions. Searching for inter- and intra- chromosomal interactions in the past was laborious and relied upon an *a priori* assumption as to which DNA regions might be interacting (3C [70]) or alternatively, required a bait fragment being chosen and then characterising which regions of genome interacted with it (4C [21] and 5C [20]). By incorporating high throughput sequencing technology, and removing all *a priori* assumptions of interacting fragments, GCC detects, identifies and quantifies DNA-DNA interactions. Where earlier techniques (3C [70] and 4C [21]) could only identify interactions with a single restriction fragment of interest, GCC is able to map the entire network of 'all' interacting DNA fragments in a given cell population.

I used GCC to take a snapshot of the interaction network during *S. cerevisiae* growth on different conditions. These snapshots demonstrated that yeast genome structure is dynamic and changes depending upon the cells metabolic state. The generated GCC interaction networks were most similar for glucose and galactose grown *S. cerevisiae*; the most comparable growth conditions metabolically, that is, they both undergo respiro-fermentation. The GCC interaction network map obtained for *S. cerevisiae* during solely respiratory growth differed the most from the other two growth conditions.

6.1 Extra-nuclear DNA

The mitochondrial genome is essential during respiratory growth and the transcription of the mitochondrial genes as well as ~1100 nuclear encoded genes is enhanced under this condition. Excitingly, GCC detected interactions between the mitochondrial and nuclear genomes, and the numbers and types of these Mito-gDNA interactions changed depending upon the metabolic state of the cells.

Using 3Cs and quantitative PCR, I was able to demonstrate that two Mito-gDNA interactions were dependent upon mitochondrially encoded reverse transcriptase

machinery. These results indicate that the nucleic acid being transferred from the mitochondrial organelle to the nucleus occurs through an RNA intermediate, and thus the interacting mtDNA is a cDNA. Moreover, reducing Mito-gDNA interaction frequency by utilising mitochondrial reverse transcriptase knockout mutants increased the number of detectable mRNA transcripts for the nuclear genes with which the mitochondrial cDNAs were interacting.

S. cerevisiae does not encode RNAi machinery, and thus these observations cannot be attributed to such a mechanism [322]. Data which suggests DNA acts as a signalling molecule has not been presented before. I hypothesise that mitochondrial cDNAs are transported from the mitochondrial organelle to the nucleus where they act as a signalling molecule to control nuclear gene transcript levels. I propose that the transfer and subsequent interactions involving mitochondrial cDNAs form as part of the retrograde response [168]; an important signalling pathway which coordinates the expression of nuclear and mitochondrial genes in concert with the cells current metabolism.

It is not clear how the mitochondrial cDNAs alter the transcript levels of nuclear genes. However, the co-localisation of mitochondrial cDNAs with nuclear coding regions could mediate recruitment and binding of specific transcription factors, thus altering transcription levels of certain genes required for mitochondrial function. Alternatively, the mitochondrial cDNAs might target specific mRNA transcripts for destruction.

6.1.1 The rDNA repeats interact with the mitochondrial cDNAs

The nucleolus forms around the rDNA repeats and is a storehouse of factors required for nuclear function in *S. cerevisiae* and higher eukaryotes [323, 324]. Latent proteins required for cell cycle initiation and regulation are sequestered in the nucleolus which await activation by external stimuli [321]. In the GCC dataset, during respire-fermentative growth the rDNA interacted highly with mtDNA, whereas during respiratory growth the rest of the nuclear genome was much more connected with mtDNA.

A possible explanation for the numbers of interactions forming between mtDNA and the nuclear rDNA repeats might be that the nucleolus is acting as a store for functional mitochondrial cDNAs. The cDNAs might be either sequestered to the nucleolus for storage or released into the rest of the nucleus where they interact with the genes

required for mitochondrial function, according to the metabolic regime being employed by the cell. This type of arrangement is analogous to the p53 protein being sequestered to the nucleolus during specific types of stress [324, 325], or RNA transcripts being sequestered to RNA granules during normal or stressed conditions [319, 320].

Fluorescent labelling of mitochondrial DNA fragments may allow visualisation of the mitochondria cDNAs within the nucleus of *S. cerevisiae* or higher eukaryotes. This has been achieved before, with mitochondrial DNA being observed inside the nucleus of human low-grade gliomas [160]. Furthermore, the nuclear localisation of the mtDNA could be probed under different conditions which may provide clues as to the role the rDNA repeats play in the spatial integration of mtDNA in the nucleus.

6.1.2 Mitochondrial DNA may alter the activity of other nucleolar processes

The amount of mitochondrial DNA within a cell has been shown to influence the rDNA replication rate [167]. The Sir2 protein is involved in silencing at the mating loci, telomeres, and critically, the rDNA repeats [326, 327]. Sir2p, through interactions with other proteins [328], binds to replication initiation sites to negatively regulate the initiation of chromosomal DNA replication [328]. There is an inverse correlation between the amount of mtDNA in the cell and Sir2 bound at the replication initiation sites [167]. These intriguing results could be explained by the mitochondrial cDNA interaction events described above. One could surmise that interactions which form between mitochondrial cDNAs and the rDNA repeats could antagonise Sir2p binding on the rDNA repeats. Thus, a greater amount of cellular mtDNA might lead to a higher transfer rate, a higher interaction frequency with the rDNA repeats, and consequently, a higher replication rate.

NUclear MiTochondrial Sequences (NUMTS) in yeast were found to be rich in ARS consensus sequences [329]. Recent experiments elucidating alternative roles of NUMTs indicated their role in promoting nuclear DNA replication [329]. It is unlikely that GCC is capturing interactions with NUMTS; however, the mitochondrial cDNAs may still promote the recruitment of factors required for replication to the rDNA repeats thus increasing the replication rate.

The measurement of rDNA replication rate in the mitochondrial group I and II intron mutants, whose Mito-gDNA interaction rate is reduced, may demonstrate a correlation between the number of transferred mitochondrial cDNAs and rDNA replication rate.

6.1.3 Mitochondrial DNA is utilised during nuclear chromosome double strand break repair

Mitochondrial DNA is incorporated into the nuclear genome as part of the DNA repair pathway for double stranded breaks [154]. Previous reports [164] of mtDNA transfer to the nucleus as small unstable episomes were assumed be the source of this nucleic acid. The transfer of mitochondrial cDNAs, generated through an RNA intermediate as described in this thesis, could provide an alternative, but not necessarily mutually exclusive explanation for the presence of nuclear located mtDNA fragments that are capable of being incorporated at double strand break sites. Further support for the transfer through an RNA intermediate, comes from the whole *COX2* gene being transferred from the mitochondria to the nucleus during flowering plant evolution [330].

6.1.4 DNA as a signalling molecule

Given the data presented in this thesis, it is not a stretch to imagine how DNA might be used as a signalling molecule, just as RNA has been [331]. DNA and RNA have already been shown to be transported and transferred between cells in multi-cellular organisms [332] as well as between different organisms [333, 334]. I have shown that extra-nuclear DNA, from an intra-cellular symbiont, can alter the transcript levels of genes in *S. cerevisiae*. Therefore, it is not too much of a stretch to imagine that other organisms which invade cells (*e.g.* malaria) might use this machinery in order to alter the cellular environment in its favour. Certainly, proteins are used to this effect by *Agrobacterium tumefaciens* in the infection and transfer of plasmid DNA into plant cells ([reviewed in [335]).

Overall the transfer of mtDNA to the nucleus and subsequent formation of interactions between the transferred mtDNA and the nuclear genome opens up interesting questions for the future of genome organisation and mitochondrial biology. Further experiments are required to gain perspective as to the importance of these interactions. These experiments might include genome-wide transcription analysis in *S. cerevisiae* cells both containing and lacking the mitochondrial genome. Additionally, performing GCC on an *S. cerevisiae* strain which has drastically reduced numbers of rDNA repeats [336]

and comparing it against a strain with a wild type number of rDNA repeats might provide insights as to whether the rDNA repeats are required for sequestration of the mitochondrial cDNAs. These GCC networks would need to be compared against transcriptional profiles for both of these strains. The premise being that there would not be enough rDNA repeats for the transferred mtDNA to interact with, thus, under situations which required sequestration, the mtDNA would be free to interact with the rest of the genome. This experiment assumes that the same number of mtDNA fragments is transferred to the nucleus in both strains. To circumvent this problem one could generate rho^0 cells from the strains with wild type and mutant numbers of rDNA repeats and heat shock the cells with mtDNA fragments. This way the amount of mtDNA could be standardised between the two strains, allowing the measured differences in interaction frequency and transcript levels to be attributed to the rDNA repeat sequestration of the mtDNA.

While *S. cerevisiae* is a great model organism for studying interactions between mtDNA and the nuclear genome due to the mitochondrial genome being non-essential during fermentative growth, it will be interesting to see if these interactions occur in human cells. Early indications suggest this may be the case [160, 337], with mitochondria and the mitochondrial genome becoming more of a focus during the assessment of cancerous tissues [338-342]. My hypothesis, given the data presented in this thesis, is that interactions between the mitochondrial and nuclear genomes play an important role for normal nuclear processes.

6.2 Intra-chromosomal interactions may be indicative of chromosomal structure

Intra-chromosomal interactions detected in the GCC dataset are indicative of chromosome loops which tend to be small in the *S. cerevisiae* genome. I have hypothesised that intra-chromosomal loops were detected in the GCC dataset as a result of general genome compaction and loops forming around active genes [186, 213].

One possibility to determine whether GCC is definitively capable of capturing the structure of inactive regions of the genome would be to couple GCC with chromatin immuno-precipitation (ChIP) against heterochromatic markers (*i.e.* HP1 [269], Sirtuins [115, 147, 327], and heterochromatic histone marks). Given the results presented in section 3.2.11 for the increased connectivity of essential genes, one might predict that

GCC would detect less inter-chromosomal interactions for this dataset, and more short-range intra-chromosomal interactions indicative of chromatin compaction. This result could be correlated with GCC performed on *S. cerevisiae* arrested at metaphase when the chromosomes are condensed for segregation. During metaphase I would expect an enrichment of intra-chromosomal interactions. However, it would be difficult to ascertain whether the interactions resulted from true intra-chromosomal loops or sister chromatid pairing due to cohesion loading [299].

Loops form around active genes which are thought to promote re-initiation of the RNA pol II complex during the process of transcription [186, 213]. Co-transcriptional 3'-end processing requires the CPF complex, components of which (the Pta1 and Ssu72 proteins) have been shown to mediate looping between the promoter and terminator regions of the *BUD3* and *SEN1* genes [213]. ChIP against Pta1 or Ssu72, coupled with GCC, may allow the identification of intra-chromosomal loops which form as part of transcription. The separation of functional as opposed to structural interactions in the GCC dataset would provide researchers with a greater amount of information on the individual interactions. Thus allowing a more informed choice as to the interactions to isolate and investigate further.

6.3 Spatial dynamics of metabolism

The scientific community has amassed a large amount of data and knowledge about how *S. cerevisiae* responds under different conditions. Some of the most well studied conditions are to do with metabolism and the pathways which dictate a carbon sources fate [13, 14, 16, 113, 172, 173, 200, 215, 249, 271, 292, 343, 344]. As previously discussed, the mitochondrial organelle is tightly associated with metabolism, and the interactions which occur between the mitochondrial and nuclear organelles seem to alter in concert with the metabolic regime, challenging our thoughts about the host-symbiont relationship.

The number of interactions with each gene in different metabolic pathways was shown to partially correlate with the genes transcriptional activity. It is not immediately clear why gene activity would result in a higher interaction frequency until you factor in the spatial organisation of nuclear processes [26]. Active genes co-localise at transcription factory sites which have an enriched concentration of the factors required for transcription [112, 345]. Therefore, it is not a surprise to find that active genes connect

more frequently than when the same genes are inactive, especially as it has been shown that in the absence of transcription inter-chromosomal interactions break down [112].

I propose that GCC could also be coupled with a ChIP assay against the active phosphorylated form of RNA pol II to allow the detection of co-regulated interacting genes, or at least interactions resulting from co-localisation at transcription factory sites, on a global scale. This may shed light on the types of genes which co-localise at these transcription factory sites. One might hypothesise that gene families might be spatially clustered in order to synchronise their mRNA output, or be switched on or off in concert with each other. Alternatively, spatially separate transcription factories may have different mRNA output levels, thus genes which require a high transcription rate are recruited to high output factories, and vice versa. We did not observe much evidence for gene family clustering in the GCC dataset, although the two *MED* gene family members were identified to interact with the same tRNA by 4Cs. It is not clear whether the two *MED* gene family members are spatially adjacent to each other in a single cell or whether the interactions are mutually exclusive.

Metabolism is vitally important to a cells development and maintenance and therefore the proper expression of genes involved in metabolic pathways needs to be coordinated with each and every participating element. Coordination of metabolic pathways is typically considered to occur through positive and negative feedback loops. As such, the data presented in this thesis leads me to hypothesise that the three dimensional nature of the nuclear genome in conjunction with extra-nuclear DNA and other signalling molecules may also contribute to the appropriate coordination of condition specific gene expression. Many of the metabolic pathways which occur in *S. cerevisiae* are common to Human. As such, understanding as much as we can about how these pathways are incorporated and regulated is important for a full understanding of metabolic and mitochondrial processes.

6.4 The genome in three dimensions

Two global interaction methods were published shortly after the publication of the GCC methodology [5, 74, 90], however, Duan *et al.* (2010) took the interaction dataset to the next logical step by modelling the nuclear genome in three-dimensions. While Duan *et al.* (2010) focussed their attention on the yeast genome; Liebermann *et al.* (2009) designed an elegant 3C derivative methodology capable of detecting interactions for

large genomes such as Human [74, 90]. Duan *et al.* (2010) used the inter- and intra-chromosomal interactions detected by their methodology to model the *S. cerevisiae* genome in three-dimensions [74]. The *S. cerevisiae* chromosomes were shown to exist in the, not unexpected, Rabl-conformation [74]. As discussed earlier, the interaction data derived from Hi-C is consistent with chromosome territory formation in higher eukaryotes, whereby the chromosomes tend to follow a fractal globule arrangement [90].

Three-dimensional modelling of the *S. cerevisiae* chromosomes, mediated by the GCC datasets outlined in this thesis is currently underway[13]. This modelling will allow comparisons between GCC data and the data derived by Duan *et al.* (2010) and may provide clues as to the existence of chromosome territories, early indications support their existence.

Three-dimensional modelling of genomes is difficult, to say the least. For example, chromosomes are dynamic entities, and static three-dimensional images cannot be accurately produced from interaction data derived from a population of cells. Particularly if the population contains cells at all different stages of the cell cycle, including metaphase where I would predict an overriding number of intra-chromosomal interactions would prevail. By including these interactions one might bias analyses of chromosome folding with the inclusion of these intra-chromosomal interactions. Taking the highest frequency interactions and using those to model, one arrives at a population average structure, which may provide clues as to clustering of elements or the nuclear position of certain genomic components.

A possible alternative to circumvent these population problems is to perform GCC on a single cell. While technically difficult, and possibly incorporating amplification biases, this analysis could arrive at a snap shot structure for the *S. cerevisiae* genome. However, until this analysis had been performed on multiple cells at the same stage of the cell cycle and grown under the same conditions it would not be possible to state that there was one absolute genomic structure which dictates or results from, a particular phenotype [95]. Structures obtained from single cell analysis would be interesting to

[13] Three-dimensional modeling: Mr. Gerd Grunert, Dr. Lutz Gehlen, Myself, Prof. Jorg Langowski, and Dr. Justin O'Sullivan.

compare against the population average structure to determine whether these was any commonality.

No one has yet shown how the structure of a eukaryotic genome responds to different environmental conditions. There are certainly clues in the literature which suggest the genome undergoes rearrangement during the activation of different gene sets [40, 43, 103, 187] and the data I generated in this thesis for the three different growth conditions, glucose, galactose and glycerol lactate could be used to generate the first three-dimensional models of the genome during different conditions. There is existing data which indicates the presence of gene territories and gene movement upon exposure to the carbon source galactose [64], thus correlations with these different nuclear positions could be made in the models.

The ultimate goal of detecting global interaction networks is surely the mapping of interactions in healthy and diseased human cells. Increasing evidence supports aberrant transcriptional regulation as contributing to the development of human cancers [346, 347]. Data suggests that regions of the genome which are in a transcriptionally inactive state in a healthy cell become active in some cancerous tissues [348]. Given the importance of the spatial organisation of the nucleus for the regulation [27, 349, 350] and transcription [351] of genes, it is possible that aberrant genome structure might contribute to the cancer phenotype. In fact, mutations that alter the nuclear matrix, onto which the nuclear genome in human cells is attached, have been verified to occur in cancerous cells [352]. These mutations may prevent regions of the genome from attaching to the nuclear matrix, thus leading to a situation where a ‘base’ genome structure is unable to be achieved. Inactive regions of the genome may enter non-membrane bound sub-compartments set aside for active nuclear processes, leading to the activation of these transcriptionally inactive regions.

Reoccurring chromosomal translocations add weight to the fact that the genome has an ordered structure which can exist across different individuals [353]. For example, the *BRC* region on chromosome XXII and *ABL* gene chromosome IX have been observed to undergo double strand DNA breaks and subsequently join to each other in a well characterised translocation in leukemic cells creating the Philadelphia chromosome [354]. This demonstrates the presence of a reoccurring genomic structure in white blood

cells, as these two regions need to be in close proximity to reproducibly undergo translocation. It also underlines the importance of genome structure and organisation.

6.5 Final remarks

Global methods that detect epigenetic alterations (*e.g.* histone codes, DNA methylation, DNA-DNA interactions *etc.*) allow researchers to gain an overall perspective of how cells respond during different conditions. Merging the results of these global epigenetic experiments with the linear sequence of an organism allows a view of the genome never before seen. A fully integrated picture of the spatial and functional organization of the genome and nucleus is beginning to be put together.

Inter- and intra- chromosomal interactions can tell us something about the structure of the genome as a whole as well as how local chromatin behaves. Capturing these interactions on a global scale has allowed me to ascertain that the numbers and types of these DNA-DNA contacts change depending upon the metabolic state of the cell. At the very least, GCC provides a list of loci participating in inter- and intra- chromosomal interactions that can be interrogated to determine the nature of the individual contacts, be they functional or structural. However, techniques such as quantitative 3C and FISH [355] must be utilised to uncover the possible roles of the individual contacts detected as well as establishing the mechanism by which the contacts form.

The discovery of a novel signalling mechanism between the mitochondria and the nucleus provides exciting leads for the future of mitochondrial biology and understanding the evolution of the symbiosis. By and large this research highlights the importance of examining genomes as fully integrated biological systems into which extra-nuclear DNA must be included. Understanding how the linear DNA sequence traverses multiple layers of organisation to become a functional organism is a biological mystery. The genome structure snapshots GCC provides is a step forward in the quest to unveil how this function manifests itself.

Appendix I. Primer Sequences

Name	Sequence	Tag
MITOCHONDRIA COPY NUMBER		
Mito+ve13909F	TGCTCAACGAAAGTGAATCAA	
Mito(CNC)R13909	GATTTATCGTATGCTCATTTCCAA	
NUCLEAR GENOME COPY NUMBER		
GAL1F	TTGCGAACACCCTTGTTGTA	
GAL1R	CGTGCTCGATCCTTCTTTTC	
MITO-rDNA 3C TAQMAN		
MitogDNA3CForward	GTGAGCCGTATGCGATGAAAG	
Nts1_599R	TTATTCCTTCCCGCTTTCCT	
MitogDNA3CProbe	FAM-TCGCACGTACGGTTCTTACCGG	
MITO-rDNA POSITIVES		
rDNA+ve460025F	CATTATGCTCATTGGGTTGC	
rDNA+ve460025R	AGGAAAGCGGGAAGGAATAA	
Mito+ve13909F	TGCTCAACGAAAGTGAATCAA	
MitoAcross13909R	TCCCGATAGGTAGACCTTTACAA	
MITOCHONDRIAL GROUP I AND II MUTANT GENOTYPE TESTING		
Mito_AI1_E&I_F	AGTGGTATGGCAGGAACAGC	
Mito_AI1_E&I_R	CCCCGTAAAGTTAGCCCCTA	
Mito_AI2_I_F	GGGGATTGTGATTCATGCTT	
Mito_AI2_I_R	CTGTCTTCCTTCCTTGCATTT	
Mito_AI3_I_F	ACTTTCTTCCCCTCCGAATC	
Mito_AI3_I_R	GGCCCTCGTGGGGATAATA	
Mito_AI4_I_F	TGATCAATTTTCATTACAGCGTTC	
Mito_AI4_I_R	TTTTCTTGTAGTCTCTGAGGATCTTTT	
Mito_AI5_I_F	AGGCAAACTCGAGGAAAACC	
Mito_AI5_I_R	AATATCCTCAATTAAGAGGTCGAA	
gDNA-gDNA 3C TAQMAN		
gDNAgDNA3CForward	TGACACCGTCTCTTGTTTAGGA	
gDNAgDNA3CReverse	TTGATCGTATCCTTCTCTAGTGAAC	
gDNAgDNA3CProbe	FAM-TTAACTTCAGTTAAATCTTCAA	
Mito-gDNA 3C TAQMAN		
MitogDNA3CForward	GTGAGCCGTATGCGATGAAAG	
MitogDNA3CR13221	GAATCCCTCGCCAACATAGA	
MitogDNA3CProbe	FAM-TCGCACGTACGGTTCTTACCGG	
LIGATION CONTROLS		
E.coli211bp3'*MspI*F	GCCAGAAATTCGTCGGTAAG	
E.coli211bp3'*MspI*R	AACCGGTCATTGAAGTATTGA	
Lambda185bp3'*MspI*F	TTTACAGCGTGATGGAGCAG	
Lambda185bp3'*MspI*R	ACCAATCCAGCCGGTCAG	
Mito-gDNA POSITIVES		
Mito+ve13909F	TGCTCAACGAAAGTGAATCAA	
MitoAcross13909R	TCCCGATAGGTAGACCTTTACAA	
MitogDNA3CR13221	GAATCCCTCGCCAACATAGA	
gDNA+ve13221F	CATCCATCTCGCAGCAATTA	
gDNA-gDNA POSITIVES		
gDNAAcross5313F	TGATCGTTGGGTAATGCTCTT	
gDNAAcross5313R	AGCCTGTGGTTGATGGAAAC	

gDNAAcross6434F	CCACCACAAATCAAGCCTCT	
gDNAAcross6434R	TATGGTCGCTCCTGCCTATC	
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	
Linear vs Circular Purification Bias		
3cpUC19ecoR1fwd	GTTTTCCCAGTCACGACGTT	
3cpUC19ecoR1fwd	GTTTTCCCAGTCACGACGTT	
3cpUC19ecoR1rev	ATTAGGCACCCCAGGCTTTA	
3cpUC19fwd	ACGCTCAGTGGAACGAAAAC	
3cpUC19rev	GGGAGTCAGGCAACTATGGA	
3C RPA135and tRNA with rDNA – Primers and Probes and Positive primers		
RPA135+veFrag1_F	CTGCAGAAGAAACACCATC	
RPA135+veFrag1_R	CCGGATGGAGGTTTGTTAAA	
RPA135+veFrag2_F	CCACCACACTACAACCACCA	
RPA135+veFrag2_R	GAGGAAATGGTTTGAACTGCAT	
rDNA+veHindIII_F	ACTCATGTTTGCCGCTCTG	
rDNA+veHindIII_R	CGATGAGGATGATAGTGTGTAAGA	
RT-RPA135Frag1_R	CCACTCAATTCTGAAGACGAAA	
rDNAHindIIIF	GCTCCATGAAGCAAACTGTCC	
RPA135HindIIIFrag2RorF	TTCTTGTCCATTATTGCCATTT	
rDNAHindIIIProbe	CAAATCCTTTCACGCTCGGGAAGC	
4C analysis of tRNAs and tRNA replacements		
WTyDP84outsideF	CGTTGTAACCGAGAGGATGG	-
WTyDP84outsideR	GGCGCCTGATTCAAGAAATA	-
WTyDP84NestedF	GTCGATAATGGACTTTGCAC	TCTCTG
WTyDP84NestedR	TCAGGCAAGAAAGGGAACCG	TCTCTG
WTyDP77outsideF	AAGTTTTCAATTATCCAGCTACTGC	-
WTyDP77outsideR	GGCCAACCCTCCGTATAAA	-
WTyDP77NestedF	ACGCAATATCATAAAAGCAC	AGAGAC
WTyDP77NestedR	TTTCTTTAGTAGCTCCTCCG	AGAGAC
F1MUTyDP84outsideF	GTTGGTTTAAGGCGCAAGAC	-
F1MUTyDP84outsideR	GCAGGCCCACCTTTATTTCT	-
F1MUTyDP84nestedF	GATCGGGCGTTCGACTCGC	TGATGC
F1MUTyDP84nestedR	CAGGCAAGAAAGGGAACCG	TGATGC
F2MUTyDP84outsideF	GCCGTTAAATCCAATCATCC	-
F2MUTyDP84outsideR	GGCGCTCCACCTTTATTGTA	-
F2MUTyDP84nestedF	GTCGATAATGGACTTTGCAC	AGCACG
F2MUTyDP84nestedR	CGAACAAAAAAATCTCCCG	AGCACG
F1MUTyDP77outsideF	AGTTGGTTTAAGGCGCAAGA	-
F1MUTyDP77outsideR	TTGCAGCACTCTTTCTTTAGTAGC	-
F1MUTyDP77nestedF	AGATCGGGCGTTCGACTCG	ACAGAG
F1MUTyDP77nestedR	TTTCTTTAGTAGCTCCTCCG	ACAGAG
F2MUTyDP77outsideF	GCGGCACTTCGTAAGTTTTC	-
F2MUTyDP77outsideR	CCCCCTAATTCCCTTCTTTT	-
F2MUTyDP77nestedF	TGATACGCAATATCATAAAAGCAC	TAGATC
F2MUTyDP77nestedR	GTATATTAACAAAAAAATCTCCCG	TAGATC

Negative 3C Confirmation		
10330_3'*MspI*3C	TTGAATGGTCCTTTCGTGGT	
10330_5'*MspI*3C	TTCTTGGTGAGCCGTTTACTA	
6897_3'*MspI*3C	ACATTCATCGGACGCCTTT	
6897_5'*MspI*3C	TTTCATTCCGCCCAATTCTA	
Positive 3C Confirmation (outside and nested primers)		
gDNAgDNA3CFout(868673-172555)	TCGATCGTTGGTTTCATTTG	
gDNAgDNA3CRout(868673-172555)	CAGACCAAAAACTTTACCCTTCC	
(gDNAgDNA3CF(868673-172555)	TTGTTGTGACACCGTCTCTTG	
gDNAgDNA3CR(868673-172555)	GACTTCAATGGCCTCTGC	

Appendix II. Calculations

Standardising the quantitative PCR results

The samples which are subjected to qPCR with primers and fluorescent probes do not have the same amount of DNA in them. Every effort is made to ensure the samples contain the same amount of DNA at the end; however, the concentration of 3C samples cannot be accurately measured. The samples need to be standardised to each other in order to directly compare the fluorescent detection output from the qPCR machine.

This is achieved by performing a standard SYBR green qPCR reaction on the samples using primers designed to amplify within an *MspI* fragment of the *GAL1* gene (primers GAL1F and GAL1R see Appendix I). The *GAL1* gene is a single copy gene on chromosome II. This essentially provides a value for each sample which corresponds to the genome copy number. That is, we assume that *GAL1* occurs at one copy in each genome and therefore it is a measure of the number of genomes.

We are then able to set our wild type sample (or any other sample) to 1 and express our other samples as a ratio of this particular sample set at one. This provides a 'correction value' which the resulting numbers for the qPCR with fluorescent probes can be multiplied by for each sample, thereby standardising our samples for nuclear genome copy number.

This same protocol is used if the data needs to be standardised by mitochondrial or rDNA copy number, obviously using primers designed to these regions.

How much DNA is theoretically in one 3C sample?

Each bp is 650 Daltons. I calculate the base pair composition of the *S. cerevisiae* genome by adding all of the chromosome lengths together, including the rDNA, 2-micron plasmid and mitochondrial genome, at their appropriate copy numbers. This results in 18,534,409 bp of DNA in one *S. cerevisiae* genome. 650 Da multiplied by 18,534,409 bp = 12047365850 Da. A conversion from Daltons to grams (there are $6.02213665168 \times 10^{23}$ Daltons in a gram) gives us a weight for an individual genome complement in grams: $2.000501241 \times 10^{-14}$g. There are 90,854,545 cells in each 3C sample therefore there is $1.8175463002299034 5 \times 10^{-6}$ g of DNA. Converted to

nanograms this is the same as 1817.5463 ng in 40 µl resuspension volume. So theoretically the concentration should be 45.43 ngµl^{-1} of an individual 3C sample.

Appendix III. Sequence files

Chapter 3:

File names for the 36bp single end sequencing output performed at Massey University, Palmerston North of the original GCC library from glucose grown *S. cerevisiae* cells:

Three Illumina Genome Analyser lanes:

Lane 1: s_1_00##
Lane 2: s_2_00##
Lane 3: s_3_00##

Chapter 3 sequence files are deposited with the Gene Expression Omnibus (GSE13648).

Chapters 4 and 5:

File names for the 38bp paired end sequencing output performed at the FMI Institute, Basel, Switzerland of the glucose, glycerol lactate, and galactose GCC libraries:

ETHZ_BSSE_100423_61LNTAAXX_1.fa
ETHZ_BSSE_100423_61LNTAAXX_2.fa
ETHZ_BSSE_100423_61LNTAAXX_3.fa
ETHZ_BSSE_100423_61LNTAAXX_4.fa
ETHZ_BSSE_100423_61LNTAAXX_6.fa
ETHZ_BSSE_100423_61LNTAAXX_7.fa
ETHZ_BSSE_100423_61LNTAAXX_8.fa
A readme file has been deposited on the Appendix CD to deconvolute the file names.

File names for the 100bp paired end sequencing output performed at Massey University, Palmerston North of the 4C libraries generated from the yPH499, yDP77, and yDP84 *S. cerevisiae* strains (same names as above, although stored on different hard drives):

ETHZ_BSSE_100423_61LNTAAXX_1.fa
ETHZ_BSSE_100423_61LNTAAXX_3.fa

All sequence files, Topography output files, and 4C tRNA output and analysis files can be obtained from Dr. Justin O'Sullivan.

Appendix V. Manuscript - In Review

NB: Supplementary Material is located on the Appendix CD.

Mitochondrial-nuclear DNA interactions regulate nuclear transcription.

C.D.M. Rodley[a], R.S. Grand[a], B. Jones[b], J.M O'Sullivan[a†]

[a]Institute of Natural Sciences, Massey University, Albany, New Zealand

[b]Centre for Mathematical Biology, Massey University, Albany, New Zealand

[†] Corresponding Author:

Justin M. O'Sullivan: Email J.M.OSullivan@massey.ac.nz; fax +64 9 441 8142

Address: Building 11
Station Crescent Gate 4, Oteha Rohe
Old Albany Highway
Albany, Auckland
New Zealand

Running head: Mito-nDNA interactions regulate transcription.

Keywords: mitochondria, genome organization, transcription

Abstract

The nuclear and mitochondrial organelles must maintain a communication system in order to respond effectively to environmental conditions. Previous studies have identified mitochondrial DNA inside the nucleus and interacting with the nuclear chromosomes. How this transfer occurs and what the function of the mitochondrial DNA is once inside the nucleus remains unclear. Here, we comprehensively map DNA-DNA interactions between the mitochondrial and nuclear genomes (Mito-nDNA) using Genome Conformation Capture in *Saccharomyces cerevisiae* cells grown under three different metabolic conditions (glucose, galactose and glycerol lactate). Global differences are observed for the numbers and types of Mito-nDNA interactions according to the metabolic state of the cell. Two interactions between the mitochondrial and nuclear genomes (*COX1-MSY1* and *Q0182-RSM7*) showed significant reductions in the absence of mitochondrial encoded reverse transcriptase machinery. Moreover, the *COX1-MSY1* inter-organelle DNA interaction significantly reduced the transcript levels of the nuclear *MSY1* gene while the *Q0182-RSM7* interaction also caused an approximately 30 % reduction in the *RSM7* transcript levels. We conclude that Mito-nDNA interactions are biologically relevant and our results argue for a role for reverse transcription in inter-organelle DNA mediated communication.

Supplementary files

Online supplementary data consists of the following: Supplementary Figures, supplementary methods, supplementary statistics, supplementary interaction tables and the aligned sequence files.

Introduction

Mitochondria have a central role within the metabolic systems of cells. In yeast, as in other organisms, the mitochondrial organelle is the site of oxidative phosphorylation and the citric acid cycle. As such, yeast mitochondria are essential for both the fermentative and respiratory pathways that yeast use to metabolize different carbon sources. Unlike most other organisms, yeast can employ both respiration and fermentation simultaneously (*i.e.* respiro-fermentation) when grown on fermentable carbon sources such as glucose or galactose, at a concentration above 0.8mM (Lagunas 1986; Otterstedt et al. 2004). *Saccharomyces cerevisiae* mitochondria contain an 85,779bp genome, which encodes a subset of the electron transport chain components (Foury et al. 1998). This genome is essential for respiratory growth on non-fermentable carbon sources like glycerol lactate (Lagunas 1976), as the subset of electron transport chain components it encodes are required to generate a proton gradient across the inner mitochondrial membrane, ultimately driving ATP synthesis. In contrast, a functioning electron transport chain, and thus the mitochondrial genome, is non-essential when cells grow solely by fermentation.

The mitochondrial genome is widely recognized as having drastically reduced in size over the course of evolution of the mitochondrial–host symbiosis, to the point that ~98% of the genes required for mitochondrial function are presently encoded within the nuclear chromosomes (Timmis et al. 2004). Consequently, mechanisms must exist to co-ordinate and control the expression of the nuclear- and mitochondrial genome- encoded genes required to maintain and control mitochondrial function (Butow and Avadhani 2004; Liu and Butow 1999). Intriguingly, despite the fact that the majority of mitochondrial genes have transferred to the nuclear genome; the 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 (Ricchetti et al. 1999). Additionally, unstable mitochondrial plasmids have been observed to transfer into the yeast nucleus (Thorsness and Fox 1990; Thorsness and Fox 1993) in a nuclear gene (*e.g. YME1, YME2*) dependent manner (Campbell and Thorsness 1998; Shafer et al. 1999; Thorsness and Fox 1993). The nuclear functions of these transferred mtDNAs are unknown, however elevated mitochondria to nucleus DNA migration rates correlate with accelerated chronological aging in yeast (Cheng and Ivessa 2010).

Distal regulatory regions (*e.g.* enhancers) loop within chromosomes in order to interact with the promoter region of the genes which they control (Tolhuis et al. 2002). In fact, enhancers can also control promoters on different chromosomes (Spilianakis and Flavell 2004; Spilianakis et al. 2005). These types of intra- and inter-chromosomal interactions can be captured using proximity-based ligation methodologies (*e.g.* Genome Conformation Capture (GCC)) that incorporate high resolution (*i.e.* ~2 Å (Fujita and Wade 2004)) cross-linking of interacting DNA strands, restriction digestion (*e.g. MspI*), and ligation to identify DNA sequences that interact at a frequency greater than background (O'Sullivan 2010b; Rodley et al.

2009). We have previously observed that nucleic acids of mitochondrial origin interact with nuclear loci (hereinafter referred to as Mito-nDNA interactions) in *S. cerevisiae* (Rodley et al. 2009). The inter-organelle, Mito-nDNA interactions formed a subset of the interactions that were identified; however, preliminary data indicated that they may respond to the metabolic status of the yeast cells, reflecting a possible role in mitochondrial-nuclear coordination (Rodley et al. 2009). GCC, being a proximity-based ligation methodology, does not inform on the nature of the contacts or the identity of the complexes that may or may not maintain them (Dekker et al. 2002; Duan et al. 2010; Laine et al. 2009; Singh and Hampsey 2007; Tolhuis et al. 2002). However, inter- and intra-chromosomal interactions have been hypothetically and experimentally linked with transcription and transcriptional memory (Brickner 2010; Osborne et al. 2007). Moreover, the quality and quantity of mitochondrial DNA has been shown to affect patterns of nuclear transcription (Parikh et al. 1987; Parikh et al. 1989) and replication (Blank et al. 2008) in yeast.

Here we identify the Mito-nDNA interactions occurring in *S. cerevisiae* during growth on glucose, galactose (*i.e.* respiro-fermentation), and glycerol lactate (*i.e.* solely respiration) using GCC. Inter-condition comparisons indicated that both the frequency and identities of inter-organelle interactions were dramatically different between conditions. Interactions between mitochondrial genes (*i.e. COX1* and *Q0182,* a dubious mitochondrial ORF) and nuclear encoded loci (*i.e. MSY1* and, *RSM7*, respectively), are shown to be 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. Our results argue for a role for reverse transcription in inter-organelle DNA mediated communication.

Results

We previously captured Mito-nDNA interactions in *S. cerevisiae* cells grown in glucose by GCC (Rodley et al. 2009). A detailed investigation of one of these Mito-nDNA interactions demonstrated that it was carbon source dependent. Therefore, we hypothesized that Mito-nDNA interactions would alter, on a global scale, according to the cells metabolic state, and in particular, the carbon source used for growth. Thus, we generated comprehensive maps of the Mito-nDNA interactions in *S. cerevisiae* by GCC, using the *MspI* restriction enzyme, during exponential growth in media containing glucose, galactose, or glycerol lactate. GCC networks were constructed from 36bp paired end Illumina Genome Analyzer sequence reads (total reads; glucose 56,167,792, galactose 48,419,385, glycerol lactate 49,134,906).

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 (Rodley et al. 2009) to determine the maximum count of a particular interaction that would be observed under this noise model, given the same number of sequences, interactions and fragments as in the experimental data. The maximum count over 100,000 simulations was 6 for the glycerol lactate condition, and 5 for the other two conditions. In our real dataset the maximum count we observe for any pairing is 14 for the glucose condition, 32 for galactose and 41 for glycerol lactate. 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}. Secondly, 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 (Supplementary Statistics). 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. Biological repeats for each condition were highly correlated for statistically significant interactions (R^2=0.78, 0.93, 0.93, respectively; Supplementary Figure 1 and Supplementary Statistics). Accordingly, sequences from biological repeats 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 (Supplementary Table 2) 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. 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 preparation of the samples for sequencing, at the

sequencing centre. 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. In conclusion, the fact that the high copy number rDNA and mitochondrial DNA elements do not show significant levels of random inter-molecular interactions with our internal control sequences is empirical evidence that the interactions we observe result from intra-molecular ligation events. Therefore, random ligation events during sample preparation do not account for the interactions we observe.

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 9.1kb rDNA repeat (data not shown). Hence, Mito-rDNA interactions are over-represented within the data-set and are carbon source dependent (Table 1).

We determined whether nuclear chromosome length correlated with the number of Mito-nDNA interactions identified for that 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 *MspI* fragment that encompasses the promoter region and partial coding sequence for two divergent ORFs: one uncharacterized ORF (YJR115W) and *RSM7* which encodes a mitochondrial small subunit ribosomal protein. Interestingly, numerous mtDNA *MspI* fragments, including fragments surrounding or overlapping the *COX1*, *COX3*, *VAR1* and *SCE1* genes, interact with this one "hotspot" on chromosome X. Thus, indicating the presence of hotspots for both interacting partners. The presence of hotspots for interactions within the mitochondrial genome was supported by the finding that the mitochondrial regions that are involved in the interactions are not uniformly distributed across the mitochondrial genome (Supplementary Figure 2).

Yeast mitochondrial escape mutants (YME) (Thorsness and Fox 1993) have been previously implicated in an elevated rate of transfer of unstable mitochondrial plasmids to the yeast nucleus (Campbell and Thorsness 1998; Shafer et al. 1999; Thorsness and Fox 1990; Thorsness and Fox 1993). Therefore, we predicted that the YME pathway was the source of mtDNA fragments interacting with the nuclear genome, and thus mutations within this pathway would result in an increase in the frequency of inter-organelle DNA interactions. To test this prediction, we used quantitative Chromosome Conformation

Capture (3C) to compare the interaction frequency between the *COX1* gene (Mt: 24872–26193bp) and the nuclear encoded *MSY1* gene (Chr XVI; 365496–365760bp), hereinafter denoted *COX1-MSY1,* in *S. cerevisiae* YME knockout mutants (*i.e.* Δ*yme1*, Δ*yme2)*. The *COX1-MSY1* interaction was chosen for further investigation. 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 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=0.103; Figure 2). These results suggest that the source of the mtDNA that participates in the Mito-nDNA interactions is not the unstable mitochondrial plasmids which have previously been 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 (Campbell et al. 1994; Campbell and Thorsness 1998). The α-factor mediated arrest of yeast cells in the G1 phase of the cell cycle also causes fragmentation of the lattice-like mitochondrial network (Neutzner and Youle 2005), a phenotype that is similar to that observed in *yme1* deletion strains (Campbell et al. 1994). Interestingly, we observed a similar reduction in the *COX1-MSY1* interaction frequency upon α-factor induced synchronization (Supplementary Figure 4). Therefore, it is possible that the fragmented mitochondrial phenotype contributes to the reduction in the *COX1-MSY1* interaction frequency we observed in the *yme1* deletion strain. Moreover, deletion of *yme2*, which does not affect mitochondrial morphology but rather is involved in the maintenance of the mitochondrial nucleoid (Hanekamp et al. 2002; Park et al. 2006), results in a relatively unchanged *COX1-MSY1* interaction frequency when compared to the wild-type (T-test [Paired P(T<=t) one-tail, n=4] p=0.377; Figure 2). By contrast, *MDV1* is involved in mitochondrial fission with deletion strains being unable to correctly fragment mitochondria (Fekkes et al. 2000; Naylor et al. 2006). Therefore, if the phenotype were directly due to mitochondrial fragmentation, one would predict that strains carrying the *mdv1* deletion would show the opposite phenotype to the Δ*yme1* strain. However, the *COX1-MSY1* 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-*Δmdv1* p=0.143, *Δmdv1-Δyme1* 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 postulated that an abnormal mitochondrial morphology, coupled with elevated mitochondrial turnover would result in a disturbance of the mitochondrial ATP synthesis pathway and a reduction in the *COX1-MSY1* interactions frequency. We tested the inter-organelle *COX1-MSY1* interaction for ATP dependence. Yeast cells were treated with an electron transport chain uncoupling agent, 2,4-Dinitrophenol (DNP), at a concentration (5mM) that inhibits respiration but allows fermentation

(Supplementary Figure 5). Quantitative 3C analyses were performed to monitor the effect of increasing DNP treatment times on the *COX1-MSY1* interaction (Figure 3). 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). However, an interaction between two nuclear loci (nDNA-nDNA; Chr VII: 868673-873686bp - Chr IX: 172565-173311bp) 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. Hence, while we knew that the transfer event was different (Figure 2) from the unstable plasmid transfer described previously (Thorsness and Fox 1993), we were unable to rule out ATP dependence.

The mitochondrial *COX1* gene contains group II introns that encode functional splicing, reverse transcriptase and endonuclease machinery (Eskes et al. 1997; Yang et al. 1996; Zimmerly et al. 1995a; Zimmerly et al. 1995b). Therefore, we predicted that the *COX1-MSY1* interaction involved a *COX1* complementary DNA (cDNA) produced by mitochondrial encoded reverse transcriptase. We obtained a *S. cerevisiae* strain which had all group II introns removed (strain 161-U7 GII-0), and a strain with all group II introns removed with the exception of aI5γ (strain 161-U7 GII-0 aI5γ; Figure 4A; (Huang et al. 2005)). Crucially, aI5γ does not encode any reverse transcriptase machinery. The aI5γ intron overlaps the *COX1* region which interacts with *MSY1* (Figure 4A). We performed quantitative 3C analyses to establish the frequency of the inter-organelle *COX1-MSY1* interaction in these strains growing with a respiro-fermentative metabolism (*i.e.* using glucose or galactose as sole C-source). The inter-organelle *COX1-MSY1* interaction was shown to be partially dependent upon the presence of the group II introns and thus the reverse transcriptase machinery (Figure 4B). The lack of a measurable *COX1-MSY1* interaction within strain GII-0, which does not contain the fluorogenic probe or reverse primer binding site (Figure 4A and 4B), confirmed that the *COX1-MSY1* interaction does not involve a nuclear-mitochondrial sequence (NUMT). An independent interaction involving a dubious mitochondrial ORF (*Q0182;* mtDNA [65783-65903bp]) and the nuclear encoded *RSM7* (Chr X [638756-640423bp]) was also shown to be partially dependent upon the presence of the group II introns (Figure 4C). The dubious mitochondrial ORF involved in this interaction does not contain any group II introns, and thus does not encode any of the components involved in reverse transcriptase activity. These results suggest 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.

The number of significant (Supplementary Methods) Mito-nDNA interactions increased by >10-fold in respiring (*i.e.* glycerol lactate grown) cells, as opposed 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. 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 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, removal of the type II introns, except aI5γ, (*i.e.* strain 161-U7 GII-0 aI5γ; Figure 4A), caused a similar increase in *MSY1* transcript levels (Figure 5A). The mean level of *RSM7* transcripts is also higher in both the 161-U7 GII-0 and 161-U7 GII-0 aI5γ strains than WT (Figure 5B). By contrast deletion of *MRS1*, which is involved in mitochondrial group I intron splicing (Bassi et al. 2002), had no affect on either *MSY1* or *RSM7* transcript levels (Figure 5C), or the *COX1-MSY1* interaction frequency (Supplementary Figure 6). Thus, alterations to mitochondrial reverse transcriptase activity lower the frequency of Mito-nDNA interactions and increase the levels of the nuclear encoded transcripts. Hence we conclude that, at least in the case of the *COX1-MSY1,* and to a lesser extent *Q0182-RSM7* interactions, cDNA mediated Mito-nDNA interactions are involved in the regulation of the nuclear transcripts.

Discussion

Despite the fact that inter-organelle nucleic acid transfer events are clearly an on-going (Adams et al. 2002; Adams et al. 1999; Brennicke et al. 1993; Campbell and Thorsness 1998; Farrelly and Butow 1983; Hazkani-Covo et al. 2010; Richly and Leister 2004; Shirafuji et al. 1997; Thorsness and Fox 1990; Thorsness and Fox 1993) and predominantly unidirectional process (Shirafuji et al. 1997; Tarassov and Entelis 1992), it has been proposed that there is no function for recent transfer events (Cavalier-Smith 1987). The observation that the number of mitochondrial interactions per nuclear chromosome is highly correlated with the nuclear chromosome length, with the exception of Chromosomes X (during growth on glucose and galactose) and XII, appears to support this hypothesis. However, the variation exhibited by Chromosome X is almost entirely explained by an increase in Mito-nDNA interaction frequency at a single locus: *RSM7*. These data correlate with the fact that *RSM7* encodes a conditionally essential (*i.e.* respiration) small subunit mitochondrial ribosomal protein that is induced during the switch from fermentation to respiration (DeRisi et al. 1997; Roberts and Hudson 2006), the metabolic condition in which it has the least interactions with mitochondrial DNA. Similarly, the *COX1-MSY1* inter-organelle interaction, couples a conditionally essential (respiration) nuclear encode tyrosyl-tRNA synthetase to mitochondrial activity, and specifically *COX1* transcript processing. Again this interaction regulates the

transcript levels of the nuclear *MSY1* gene. Therefore, in combination with the observation that a reduction in interaction frequency correlates with an increase in the amount of nuclear encoded *RSM7* and *MSY1* transcripts, we propose that the transfer events are functional and form part of the mitochondrial – nuclear communication system.

Proximity-based ligation methodologies (*e.g.* GCC) only identify the sequences that are interacting, and do not provide information: 1) on whether or not the entire mitochondrial genome or fragments thereof are transferred to the nucleus; or 2) on the percentage of a population that contain a particular interaction or the absolute level of that interaction within a particular cell. However, our combined 3C and transcript analyses of *S. cerevisiae* 161-U7 GII-0 aI5γ, in which the aI1 and aI2 introns present within the *COX1* gene have been deleted (Huang et al. 2005; Moran et al. 1995), go some way to addressing these issues. Firstly, deletion of the aI1 and aI2 introns caused a 40 - 60% reduction in the frequency of the inter-organelle *COX1-MSY1* and *Q0182-RSM7* interactions. These results confirm a role for a mitochondrial encoded RNA intermediate in these interactions. The incomplete ablation of the interactions following the deletion of the *COX1* aI1 and aI2 introns could result from: 1) the presence of other retrotransposon or retroviral encoded reverse-transcriptase of either mitochondrial or nuclear origin within the mitochondrial matrix, 2) cytoplasmic or nuclear reverse transcription of mRNA released from damaged mitochondria, or 3) the transfer of mitochondrial mini-circles (Thorsness and Fox 1993). Mechanisms 1) and 2) are most likely given the identification of remnants of nuclear derived *copia*-, *gypsy*- and LINE-like retrotransposon elements within *Arabidopsis* mitochondria (Knoop et al. 1996; Marienfeld et al. 1999), the finding that transcription does not affect mitochondrial mini-circle transfer in yeast (Shafer et al. 1999; Thorsness and Fox 1993), and the reductions in interaction frequency we observed with the *Δyme1* mutant. Secondly, the combined analyses of *S. cerevisiae* 161-U7 GII-0 aI5γ identified a significant population effect on nuclear transcript levels correlating with the reduction in the number of mitochondrial to nuclear DNA interactions caused by deletion of the aI1 and aI2 introns.

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 the high success rates attained for yeast transformation (Costanzo and Fox 1988; Ito et al. 1983), and the finding that the uptake of unstable mitochondrial plasmids occurs by vacuole mediated release into the cytoplasm (Campbell and Thorsness 1998; Thorsness and Fox 1990). However, Ricchetti *et al.* demonstrated that the mtDNA mediated repair of nuclear double strand breaks is independent of *Δyme1* mutations (Ricchetti et al. 1999) 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 (Kornmann et al. 2009), which is formed from the outer-leaflet of the nuclear membrane (Shibata et al. 2010).

The mtDNA fragments involved in the Mito-nDNA interactions that we observed were not evenly distributed across the mitochondrial genome. The finding that there is no bias for ORFs over non-coding regions appears to argue against the transfer occurring through reverse-transcribed cDNAs. However, yeast mitochondrial genes are transcribed as polycistronic transcripts (reviewed in (Costanzo and Fox 1990)), from 14 ATATAAGTA consensus promoters and possibly another 5 non-consensus promoter sites (Christianson and Rabinowitz 1983). Hence, as with other genomes, a large percentage of the mitochondrial genome is physically transcribed (Barth et al. 1999; Holec et al. 2006; Montoya et al. 1982).

We observed a large number of metabolism dependent, inter-organelle interactions between mitochondrial genomic loci and the nuclear rDNA repeats. These interactions can be explained in terms of the control of cryptic RNA polymerase II (PolII) transcription from within the rDNA repeats (Butow et al. 1988; Mayan and Aragon 2010; O'Sullivan 2010a; Parikh et al. 1989). Previous observations indicate that the mechanism may involve the formation of *SIR2* dependent inter- and intra-rDNA repeat loops (O'Sullivan et al. 2009) that contact the replication fork block site, present within the repeats, and in so doing regulate cryptic rDNA PolII transcription (Mayan and Aragon 2010; O'Sullivan 2010a). Hypothetically, the formation of the Mito-rDNA interactions antagonize the formation of these rDNA loops and thus regulate the rate of cryptic rDNA PolII transcription (Mayan and Aragon 2010; O'Sullivan 2010a) and replication. Such a mechanism fits with observations that cryptic rDNA transcription (Parikh et al. 1987; Parikh et al. 1989) and replication (Blank et al. 2008) are dependent upon the mitochondrial genome. Furthermore, the replication effect was antagonized by *SIR2* (Blank et al. 2008). Alternatively, it is possible that the nucleolus sequesters cDNAs derived from the mitochondria in a cell cycle- or environmentally-dependent manner. This situation would be analogous to the role the nucleolus is postulated to play in sequestering nuclear proteins involved in the cell cycle (Cerutti and Simanis 2000) and environmental responses (reviewed in (Boisvert et al. 2007)).

Our results establish a role for Mito-nDNA interactions in inter-organelle signaling and the control of nuclear transcript levels. These results are consistent with previous findings that nuclear transcription is positively and negatively altered in response to the presence or absence of mitochondrial genome sequences in yeast (Parikh et al. 1987; Parikh et al. 1989). It is clear that there is considerable redundancy in the pathways that coordinate the carbon controlled expression of genes, with regulation occurring at all stages of expression both in the nucleus and in the mitochondria themselves (Costanzo and Fox 1990; Schuller 2003; Towpik 2005). We propose that a fraction of mitochondrial RNAs are reverse-transcribed and these cDNAs are transported to the nucleus where they regulate nuclear functions, including transcription and replication. While we have shown a repressive role for Mito-nDNA interactions in the control of nuclear transcript levels, there is no reason to assume that all interactions are repressive. We further argue that the targets of these interactions differ depending on metabolism

specific nuclear encoded factors (*e.g.* transcription factors (Schuller 2003)), and the steady state levels of the mitochondrial RNAs, which are ATP dependent (Amiott and Jaehning 2006). In the case of the glucose and galactose dependent *COX1-MSY1* and *Q0182-RSM7* interactions, this communication acts to reinforce the catabolite dependent repression of mitochondrial translation (Figure 6).

The mechanism by which Mito-nDNA interactions control transcript levels is unknown. It is possible that it is mediated by physical interaction between the mitochondrial derived cDNA and the nuclear locus or by more indirect means. Group II introns are not ubiquitous within the mitochondrial genomes of higher eukaryotes. Despite this, it is not essential that the reverse transcriptase be encoded within the mitochondrial genome given that nuclear encoded retrotransposons have been identified within mitochondria (Knoop et al. 1996; Marienfeld et al. 1999). Therefore, further work is warranted in higher eukaryotes to determine the universal significance of these interactions.

In conclusion, we have shown that the inter-organelle DNA-DNA interactions are capable of altering the transcript levels of nuclear genes. Moreover, our results argue for a role for reverse transcription in inter-organelle DNA mediated communication.

Materials and Methods

Strains and growth conditions

Saccharomyces cerevisiae strains (Supplementary Table 1) were stored (-80°C) and cultured (30°C, 160rpm) on synthetic complete (SC) medium containing amino acid supplements and glucose (2% w/v) (Kaiser 1994), 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 (3C) 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, 160rpm, 16h) 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, 160rpm) and harvested at an optical density (OD_{600}) of 0.6. Mitochondrial uncoupling was achieved by the addition of 2,4-Dinitrophenol (5mM final concentration) for 45, 90, or 180 minutes (Supplementary Figure 5). Cell cycle arrest was achieved by treatment (180mins, 30°C, 160rpm) with cell cycle inhibitors (*i.e.* α-factor (3.4µm), nocodazole (15µgml^{-1}), or hydroxyurea (100mM)). Cell cycle arrest was confirmed by microscopy.

Genome Conformation Capture (GCC)

GCC was performed according to (Rodley et al. 2009). Briefly, chromatin was prepared from 15 sets of 10^8 (*i.e.* a total of $1.36x10^9$) 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 7-11ug, Roche). Samples were treated with RNase A (final concentration 10µgml^{-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 5µg 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 (Supplementary Methods), was added in a 1:1 ratio with the nuclear genome prior to the addition of ligase. The second ligation control ($1x10^6$ 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 (Rodley et al. 2009)). The SOAP (Li et al. 2008) algorithm was used to position paired end sequences and single ends, which

contain an *MspI* restriction site, onto the *S. cerevisiae* reference genome (Supplementary Methods). No mismatches were allowed.

Bioinformatic analyses

Bioinformatic and statistical analyses (see Supplementary Methods) 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 (Rodley et al. 2009). All statistical analyses involving 2 micron plasmid, mitochondrial, or rDNA sequences included copy number corrections (Supplementary Methods). Other repetitive elements, such as LTRs and tRNAs, were omitted from the analysis.

Chromosome Conformation Capture (3C)

3C samples were prepared as previously described (O'Sullivan et al. 2004). Quantitative 3C analyses (Rodley et al. 2009) were performed using FAM labeled BHQ Probes (BioSearch Technologies; Supplementary Table 2) 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 Supplementary Methods. Samples (2 µl in triplicate) were analyzed in a final reaction volume of 20 µl using primers listed in Supplementary Table 2. Assays were performed using a 3-stage program (50°C, 2:00 min; 95°C, 10:00 min; 45x[95°C, 0:15 sec; 60°C, 1:00 min]).

Dedicated interaction standards (concentration from: 2 ng μl^{-1} – $2x10^{-15}$ g μl^{-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 (Supplementary Table 2) using Sybr-green and a five stage program (50°C, 2:00 min; 95°C, 2:00 min; 40x [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 μl^{-1} – $7.78125x10^{-4}$ ng μl^{-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; O'Sullivan et al. 2009); Primer sequences are listed in Supplementary Table 2. 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 below). 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; 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 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 2/3s 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, 10min) 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 -80°C.

Quantitative Reverse Transcription-PCR

qRT-PCR standards were amplified from *S. cerevisiae* BY4741 genomic DNA (Supplementary Table 2). 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 - $4.0X10^{-5}$ 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, 5min; 95°C, 10sec; 40x [95°C, 5sec; 60°C, 31sec] 95°C, 15sec; 60°C, 1min; 95°C, 15sec. All transcript levels were standardized to nuclear *ACT1* and expressed as percentage of wild-type (set at 100%) +/- standard error of the mean.

Acknowledgements

The authors wish to thank L. Gehlen, A. Ganley and W. Patrick for critical discussions of this work and manuscript. We wish to kindly 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.

Financial Disclosure

This work was funded by the Auckland Medical Research Foundation, Massey University Research Fund and Maurice & Phyllis Paykel Trust. CR is funded by an HRC Ph.D. scholarship (08/554). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

CR Designed experiments, prepared GCC samples, performed laboratory experiments, bioinformatic analyses, statistical analyses, and wrote the manuscript. RSG prepared and analyzed 3C and RNA samples. BJ performed statistical analyses and advised on statistical analyses. JOS Designed experiments and helped write the manuscript.

Figure Legends

Table 1: Inter-organelle interactions are carbon source dependent. 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 (Supplementary Statistics).

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.

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 3C (Supplementary Methods) 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.

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 (5mM) treatment of exponentially growing *S. cerevisiae* in synthetic complete media, containing glucose or galactose, for the indicated time (Supplementary Figure 5). *COX1-MSY1* A) and nDNA-nDNA B) interaction frequencies were determined by quantitative 3C analyses using fluorescent probes (see Supplementary Methods). 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).

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. 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 mitochondrial reverse transcriptase 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 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 *Δmrs1*), 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).

Figure 6: Cartoon depicting the role for inter-organelle communication in refining mitochondrial translation levels. Transcription levels of the nuclear encoded *MSY1* and *RSM7* genes are modulated by interactions with mitochondrial cDNAs, which are produced from stable non-translated mitochondrial RNA populations. Other Mito-nDNA interactions could also contribute to this feed-back system. While negative feedback is emphasized in this model, feed-forward systems are also predicted.

References

Adams, K.L., Qiu, Y.-L., Stoutemyer, M., and 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. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 9905-9912.

Adams, K.L., Song, K., Roessler, P.G., Nugent, J.M., Doyle, J.L., Doyle, J.J., and Palmer, J.D. 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.

Amiott, E.A. and Jaehning, J.A. 2006. Mitochondrial transcription is regulated via an ATP "sensing" mechanism that couples RNA abundance to respiration. *Mol Cell* **22:** 329-338.

Barth, C., Greferath, U., Kotsifas, M., and Fisher, P.R. 1999. Polycistronic transcription and editing of the mitochondrial small subunit (SSU) ribosomal RNA in Dictyostelium discoideum. *Curr Genet* **36:** 55-61.

Bassi, G.S., de Oliveira, D.M., White, M.F., and Weeks, K.M. 2002. Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 128-133.

Blank, H.M., Li, C., Mueller, J.E., Bogomolnaya, L.M., Bryk, M., and Polymenis, M. 2008. An increase in mitochondrial DNA promotes nuclear DNA replication in yeast. *PLoS Genet* **4:** e1000047.

Boisvert, F.M., van Koningsbruggen, S., Navascues, J., and Lamond, A.I. 2007. The multifunctional nucleolus. *Nat Rev Mol Cell Biol* **8:** 574-585.

Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V., and 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.

Brickner, J.H. 2010. Transcriptional memory: staying in the loop. *Curr Biol* **20:** R20-21.

Butow, R.A. and Avadhani, N.G. 2004. Mitochondrial Signaling: The Retrograde Response. *Molecular Cell* **14:** 1-15.

Butow, R.A., Docherty, R., and Parikh, V.S. 1988. A path from mitochondria to the yeast nucleus. *Philos Trans R Soc Lond B Biol Sci* **319:** 127-133.

Campbell, C., Tanaka, N., White, K., and 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.

Campbell, C.L. and Thorsness, P.E. 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.

Cavalier-Smith, T. 1987. The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Ann N Y Acad Sci* **503:** 55-71.

Cerutti, L. and Simanis, V. 2000. Controlling the end of the cell cycle. *Curr Opin Genet Dev* **10:** 65-69.

Cheng, X. and Ivessa, A.S. 2010. The migration of mitochondrial DNA fragments to the nucleus affects the chronological aging process of *Saccharomyces cerevisiae*. *Aging Cell* **9:** 919-923.

Christianson, T. and 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.

Costanzo, M.C. and Fox, T.D. 1988. Transformation of yeast by agitation with glass beads. *Genetics* **120:** 667-670.

Costanzo, M.C. and Fox, T.D. 1990. Control of mitochondrial gene expression in Saccharomyces cerevisiae. *Annu Rev Genet* **24:** 91-113.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. 2002. Capturing chromosome conformation. *Science* **295:** 1306-1311.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. 1997. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* **278:** 680-686.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. 2010. A three-dimensional model of the yeast genome. *Nature* **465:** 363-367.

Eskes, R., Yang, J., Lambowitz, A.M., and Perlman, P.S. 1997. Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell* **88:** 865-874.

Farrelly, F. and Butow, R.A. 1983. Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* **301:** 296-301.

Fekkes, P., Shepard, K.A., and Yaffe, M.P. 2000. Gag3p, an outer membrane protein required for fission of mitochondrial tubules. *J Cell Biol* **151:** 333-340.

Foury, F., Roganti, T., Lecrenier, N., and Purnelle, B. 1998. The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae. *FEBS Lett* **440:** 325-331.

Fujita, N. and Wade, P.A. 2004. Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. *Methods* **33:** 81-85.

Hanekamp, T., Thorsness, M.K., Rebbapragada, I., Fisher, E.M., Seebart, C., Darland, M.R., Coxbill, J.A., Updike, D.L., and Thorsness, P.E. 2002. Maintenance of mitochondrial morphology is linked to maintenance of the mitochondrial genome in Saccharomyces cerevisiae. *Genetics* **162:** 1147-1156.

Hazkani-Covo, E., Zeller, R.M., and Martin, W. 2010. Molecular Poltergeists: Mitochondrial DNA Copies in Sequenced Nuclear Genomes. *PLoS Genet* **6:** e1000834.

Holec, S., Lange, H., Kuhn, K., Alioua, M., Borner, T., and Gagliardi, D. 2006. Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol Cell Biol* **26:** 2869-2876.

Huang, H.R., Rowe, C.E., Mohr, S., Jiang, Y., Lambowitz, A.M., and Perlman, P.S. 2005. The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc Natl Acad Sci U S A* **102:** 163-168.

Ito, H., Fukuda, Y., Murata, K., and Kimura, A. 1983. Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* **153:** 163-168.

Kaiser, C.S.M.M.A. 1994. *Methods in Yeast Genetics, A Cold Spring Harbor Laboratory Course Manual.* Cold Spring Harbor laboratory press, Cold Spring Harbor, NY.

Kennell, J.C., Moran, J.V., Perlman, P.S., Butow, R.A., and Lambowitz, A.M. 1993. Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell* **73:** 133-146.

Knoop, V., Unseld, M., Marienfeld, J., Brandt, P., Sunkel, S., Ullrich, H., and Brennicke, A. 1996. copia-, gypsy- and LINE-like retrotransposon fragments in the mitochondrial genome of Arabidopsis thaliana. *Genetics* **142:** 579-585.

Kornmann, B., Currie, E., Collins, S.R., Schuldiner, M., Nunnari, J., Weissman, J.S., and Walter, P. 2009. An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science* **325:** 477-481.

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.

Lagunas, R. 1986. Misconceptions about the energy metabolism of Saccharomyces cerevisiae. *Yeast* **2:** 221-228.

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.

Li, R., Li, Y., Kristiansen, K., and Wang, J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24:** 713-714.

Liu, Z. and Butow, R.A. 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.

Marienfeld, J., Unseld, M., and Brennicke, A. 1999. The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. *Trends in Plant Science* **4:** 495-502.

Mayan, M. and Aragon, L. 2010. Cis-interactions between non-coding ribosomal spacers dependent on RNAP-II separate RNAP-I and RNAP-III transcription domains. *Cell Cycle* **9:** 4328-4337.

Montoya, J., Christianson, T., Levens, D., Rabinowitz, M., and 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.

Moran, J.V., Zimmerly, S., Eskes, R., Kennell, J.C., Lambowitz, A.M., Butow, R.A., and Perlman, P.S. 1995. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol* **15:** 2828-2838.

Naylor, K., Ingerman, E., Okreglak, V., Marino, M., Hinshaw, J.E., and Nunnari, J. 2006. Mdv1 interacts with assembled dnm1 to promote mitochondrial division. *J Biol Chem* **281:** 2177-2183.

Neutzner, A. and Youle, R.J. 2005. Instability of the mitofusin Fzo1 regulates mitochondrial morphology during the mating response of the yeast Saccharomyces cerevisiae. *J Biol Chem* **280:** 18598-18603.

O'Sullivan, J.M. 2010a. Nucleolar structure: It's all in a tangle. *Cell Cycle* **9:** 4609-4609.

O'Sullivan, J.M. 2010b. Yeast chromosomal interactions and nuclear architecture. *Current Opinion in Cell Biology* **22:** 298-304.

O'Sullivan, J.M., Sontam, D.M., Grierson, R., and Jones, B. 2009. Repeated elements coordinate the spatial organization of the yeast genome. *Yeast* **26:** 125-138.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. 2004. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36:** 1014-1018.

Osborne, C.S., Chakalova, L., Mitchell, J.A., Horton, A., Wood, A.L., Bolland, D.J., Corcoran, A.E., and Fraser, P. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol* **5:** e192.

Otterstedt, K., Larsson, C., Bill, R.M., Stahlberg, A., Boles, E., Hohmann, S., and Gustafsson, L. 2004. Switching the mode of metabolism in the yeast Saccharomyces cerevisiae. *Embo Reports* **5:** 532-537.

Parikh, V., Morgan, M., Scott, R., Clements, L., and 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., and 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. *Mol Cell Biol* **9:** 1897-1907.

Park, S., Hanekamp, T., Thorsness, M.K., and Thorsness, P.E. 2006. Yme2p is a mediator of nucleoid structure and number in mitochondria of the yeast Saccharomyces cerevisiae. *Curr Genet* **50:** 173-182.

Ricchetti, M., Fairhead, C., and Dujon, B. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402:** 96-100.

Richly, E. and Leister, D. 2004. NUMTs in Sequenced Eukaryotic Genomes. *Mol Biol Evol* **21:** 1081-1084.

Roberts, G.G. and Hudson, A.P. 2006. Transcriptome profiling of Saccharomyces cerevisiae during a transition from fermentative to glycerol-based respiratory growth reveals extensive metabolic and structural remodeling. *Mol Genet Genomics* **276:** 170-186.

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.

Schuller, H.J. 2003. Transcriptional control of nonfermentative metabolism in the yeast Saccharomyces cerevisiae. *Curr Genet* **43:** 139-160.

Shafer, K.S., Hanekamp, T., White, K.H., and Thorsness, P.E. 1999. Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae. *Curr Genet* **36:** 183-194.

Shibata, Y., Shemesh, T., Prinz, W.A., Palazzo, A.F., Kozlov, M.M., and Rapoport, T.A. 2010. Mechanisms determining the morphology of the peripheral ER. *Cell* **143:** 774-788.

Shirafuji, N., Takahashi, S., Matsuda, S., and 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.

Singh, B.N. and Hampsey, M. 2007. A transcription-independent role for TFIIB in gene looping. *Mol Cell* **27:** 806-816.

Spilianakis, C.G. and Flavell, R.A. 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nature Immunology* **5:** 1017-1027.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. 2005. Interchromosomal associations between alternatively expressed loci. *Nature* **435:** 637-645.

Tarassov, I.A. and Entelis, N.S. 1992. Mitochondrially-imported cytoplasmic tRNALys(CUU) of Saccharomyces cerevisiae: in vivo and in vitro targetting systems. *Nucl. Acids Res.* **20:** 1277-1281.

Thorsness, P.E. and Fox, T.D. 1990. Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. *Nature* **346:** 376-379.

Thorsness, P.E. and Fox, T.D. 1993. Nuclear mutations in Saccharomyces cerevisiae that affect the escape of DNA from mitochondria to the nucleus. *Genetics* **134:** 21-28.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* **5:** 123-135.

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.

Towpik, J. 2005. Regulation of mitochondrial translation in yeast. *Cell Mol Biol Lett* **10:** 571-594.

Yang, J., Zimmerly, S., Perlman, P.S., and Lambowitz, A.M. 1996. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381:** 332-335.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S., and Lambowitz, A.M. 1995a. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83:** 529-538.

Zimmerly, S., Guo, H., Perlman, P.S., and Lambowitz, A.M. 1995b. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82:** 545-554.

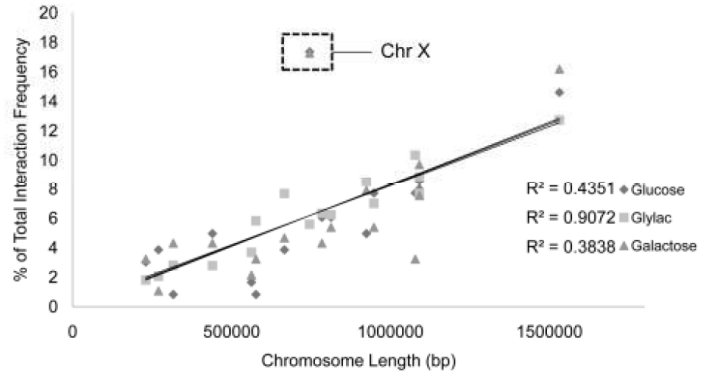

Figure 1

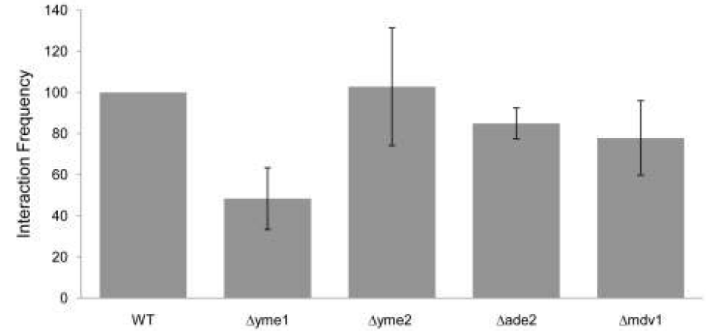

Figure 2

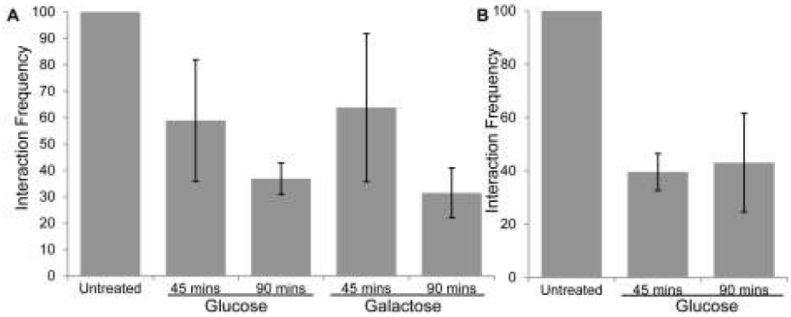

Figure 3

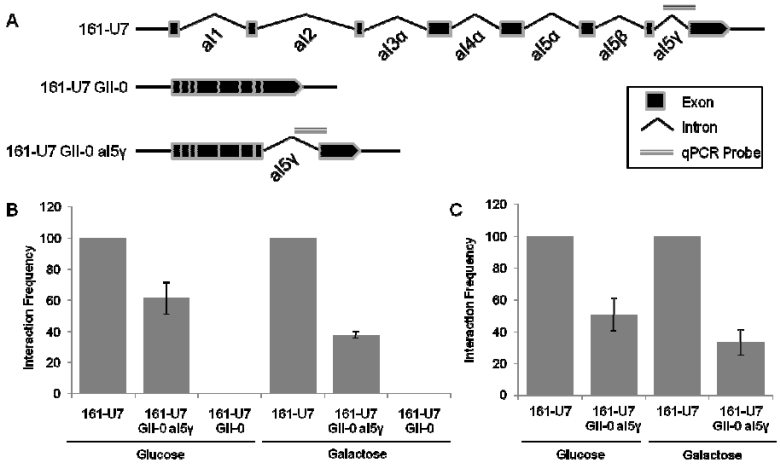

Figure 4

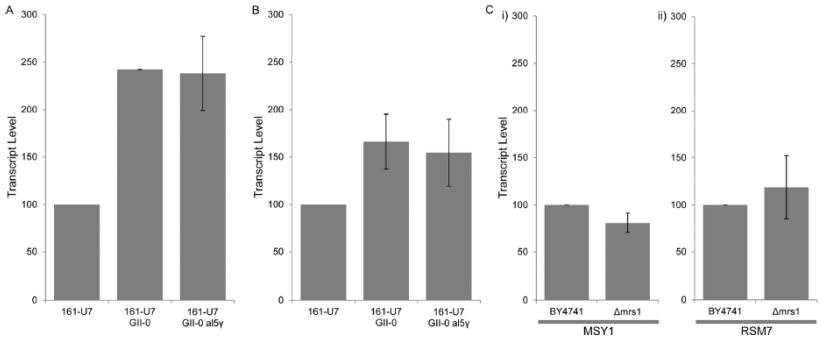

Figure 5

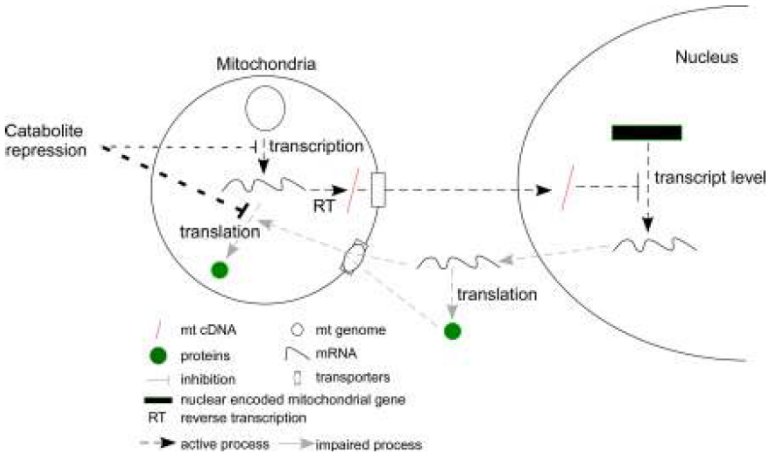

Figure 6

Table 1:

	Glucose	Glycerol Lactate	Galactose
Mito-gDNA Interactions	363	3879	278
Mito-rDNA Interactions	8315	4274	1512
Total	**8678**	**8153**	**1780**

Table 1: Inter-organelle interactions are carbon source dependent. There was a >10 fold increase in the number of Mito-gDNA interactions during growth in glycerol lactate (respiration) as compared to growth in glucose and galactose (respiro-fermentation). Growth on galactose resulted in less Mito-gDNA 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 (Supplementary Statistics).

References

1. Maeshima, K., S. Hihara, and M. Eltsov, *Chromatin structure: does the 30-nm fibre exist in vivo?* Current Opinion in Cell Biology, 2010. **22**(3): p. 291-297.
2. Rodley, C.D.M., et al., *Mitochondrial-nuclear DNA interactions regulate nuclear transcription.* Genome Research, 2011. **In Review**.
3. Wu, C., A. Bassett, and A. Travers, *A variable topology for the 30-nm chromatin fibre.* Embo Reports, 2007. **8**(12): p. 1129-1134.
4. Wei, X., et al., *Segregation of Transcription and Replication Sites Into Higher Order Domains.* Science, 1998. **281**(5382): p. 1502-1505.
5. Rodley, C.D.M., et al., *Global identification of yeast chromosome interactions using Genome conformation capture.* Fungal Genetics and Biology, 2009. **46**(11): p. 879-886.
6. Mirny, L., *The fractal globule as a model of chromatin architecture in the cell.* Chromosome Research, 2011. **19**(1): p. 37-51.
7. Moran, J., et al., *Mobile group II introns of yeast mitochondrial DNA are novel site- specific retroelements.* Mol. Cell. Biol., 1995. **15**(5): p. 2828-2838.
8. Li, R., et al., *SOAP: short oligonucleotide alignment program.* Bioinformatics, 2008. **24**(5): p. 713-714.
9. Huang, H.-R., et al., *The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(1): p. 163-168.
10. Kobayashi, T., et al., *Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I.* Genes & Development, 1998. **12**(24): p. 3821-3830.
11. Grimes, G.W., H.R. Mahler, and P.S. Perlman, *Nuclear gene dosage effects on mitochondrial mass and DNA.* The Journal of Cell Biology, 1974. **61**(3): p. 565-574.
12. Zakian, V.A., B.J. Brewer, and W.L. Fangman, *Replication of each copy of the yeast 2 micron DNA plasmid occurs during the S phase.* Cell, 1979. **17**(4): p. 923-934.
13. Johnston, M., *Feasting, fasting and fermenting: glucose sensing in yeast and other cells.* Trends in Genetics, 1999. **15**(1): p. 29-33.
14. Newcomb, L.L., et al., *Glucose Regulation of Saccharomyces cerevisiae Cell Cycle Genes.* Eukaryotic Cell, 2003. **2**(1): p. 143-149.
15. Schüller, H.-J., *Transcriptional control of nonfermentative metabolism in the yeast Saccharomyces cerevisiae.* Current Genetics, 2003. **43**(3): p. 139-160.
16. Zacharioudakis, I., T. Gligoris, and D. Tzamarias, *A yeast catabolic enzyme controls transcriptional memory.* Curr Biol, 2007. **17**(23): p. 2041-6.
17. Robinson, P.J.J., et al., *EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure.* Proceedings of the National Academy of Sciences, 2006. **103**(17): p. 6506-6511.
18. Finch, J.T. and A. Klug, *Solenoidal model for superstructure in chromatin.* Proc Natl Acad Sci U S A, 1976. **73**(6): p. 1897-901.

19. Woodcock, C.L., L.L. Frado, and J.B. Rattner, *The higher-order structure of chromatin: evidence for a helical ribbon arrangement.* The Journal of Cell Biology, 1984. **99**(1): p. 42-52.
20. 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. 1299-1309.
21. Zhao, Z., et al., *Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions.* 2006. **38**(11): p. 1341-1347.
22. DeLaat, W. and F. Grosveld, *Spatial organization of gene expression: the active chromatin hub.* Chromosome Research, 2003. **11**(5): p. 447-459.
23. Spilianakis, C. and R. Flavell, *Interchromosomal associations between alternatively expressed loci.* Nature Immunology 2004. **5**(10): p. 1017-1027.
24. Chambeyron, S. and W.A. Bickmore, *Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.* Genes Dev. , 2004. **18**(10): p. 1119-1130.
25. Osipov, S., O. Preobrazhenskaya, and V. Karpov, *Chromatin structure and transcription regulation in Saccharomyces cerevisiae.* Molecular Biology, 2010. **44**(6): p. 856-869.
26. Chakalova, L., et al., *Replication and transcription: shaping the landscape of the genome.* Nat Rev Genet, 2005. **6**(9): p. 669-77.
27. Jackson, D.A., J.C. McDowell, and A. Dean, *β-Globin locus control region HS2 and HS3 interact structurally and functionally.* Nucleic Acids Research, 2003. **31**(4): p. 1180-1190.
28. Pettijohn, D.E., *Prokaryotic DNA in nucleoid structure.* CRC Crit Rev Biochem, 1976. **4**(2): p. 175-202.
29. Gasch, A.P., et al., *Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes.* Mol. Biol. Cell, 2000. **11**(12): p. 4241-4257.
30. Osborne, C.S., et al., *Active genes dynamically colocalize to shared sites of ongoing transcription.* Nature Genetics, 2004. **36**(10): p. 1065-1071.
31. Li, Q., G. Barkess, and H. Qian, *Chromatin looping and the probability of transcription.* Trends in Genetics, 2006. **22**(4): p. 197-202.
32. Dammann, R., et al., *Chromatin structures and transcription of rDNA in yeast Saccharomyces cerevisiae.* Nucleic Acids Research, 1993. **21**(10): p. 2331-2338.
33. Cook, P.R., *Predicting three-dimensional genome structure from transcriptional activity.* Nature genetics, 2002. **32**(3): p. 347-352.
34. Cook, P.R., *The Organization of Replication and Transcription.* Science, 1999. **284**(5421): p. 1790-1795.
35. Donaldson, A.D., *Shaping time: chromatin structure and the DNA replication programme.* Trends in Genetics, 2005. **21**(8): p. 444-449.
36. Zhang, Y. and J.D. Rowley, *Chromatin structural elements and chromosomal translocations in leukemia.* DNA Repair 5, 2006.
37. Haber, J.E. and W.-Y. Leung, *Lack of chromosome territoriality in yeast: Promiscuous rejoining of broken chromosome ends.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(24): p. 13949-13954.
38. Qiu, X., et al., *A Complex DNA Looping Configuration Associated with the Silencing of the Maternal Igf2 Allele.* Mol Endocrinol, 2008.
39. Lainé, J.-P., et al., *A physiological role for gene loops in yeast.* Genes & Development, 2009. **23**(22): p. 2604-2609.

40. Brickner, J.H., *Transcriptional Memory: Staying in the Loop.* Current biology : CB, 2010. **20**(1): p. R20-R21.
41. Kucej, M., et al., *Mitochondrial nucleoids undergo remodeling in response to metabolic cues.* J Cell Sci, 2008. **121**(11): p. 1861-1868.
42. Misteli, T., *Beyond the Sequence: Cellular Organization of Genome Function.* Cell, 2007. **128**(4): p. 787-800.
43. Ahmed, S. and J.H. Brickner, *Regulation and epigenetic control of transcription at the nuclear periphery.* Trends in Genetics, 2007. **23**(8): p. 396-402.
44. Schneider, R. and R. Grosschedl, *Dynamics and interplay of nuclear architecture, genome organization, and gene expression.* Genes & Development, 2007. **21**(23): p. 3027-3043.
45. Cammack, R., T.K. Attwood, and A.D. Smith, eds. *Oxford dictionary of biochemistry and molecular biology*. Rev. ed. 2006, Oxford University Press: Oxford New York. xv, 720.
46. Trifonov, E., *The helical model of the nucleosome core.* Nucleic Acids Res, 1978. **5**(4): p. 1371-80.
47. Dorigo, B., et al., *Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber.* Science, 2004. **306**(5701): p. 1571-1573.
48. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8[thinsp]A resolution.* Nature, 1997. **389**(6648): p. 251-260.
49. Luger, K. and J.C. Hansen, *Nucleosome and chromatin fiber dynamics.* Current Opinion in Structural Biology, 2005. **15**(2): p. 188-196.
50. Bryant, G.O., et al., *Activator Control of Nucleosome Occupancy in Activation and Repression of Transcription.* PLoS Biol, 2008. **6**(12): p. e317.
51. Bermudez, A., S. Bartolome, and J. Daban, *Partial denaturation of small chromatin fragments: direct evidence for the radial distribution of nucleosomes in folded chromatin fibers.* J Cell Sci, 1998. **111**(12): p. 1707-1715.
52. Horowitz, R.A., et al., *The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon.* The Journal of Cell Biology, 1994. **125**(1): p. 1-10.
53. Eltsov, M., et al., *Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ.* Proceedings of the National Academy of Sciences, 2008. **105**(50): p. 19732-19737.
54. Tetko, I.V., et al., *Spatiotemporal Expression Control Correlates with Intragenic Scaffold Matrix Attachment Regions (S/MARs) in Arabidopsis thaliana.* PLoS Comput Biol, 2006. **2**(3): p. e21.
55. Gasser, S.M. and U.K. Laemmli, *A glimpse at chromosomal order.* Trends in Genetics, 1987. **3**: p. 16-22.
56. Benham, C., T. Kohwi-Shigematsu, and J. Bode, *Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions.* Journal of Molecular Biology, 1997. **274**(2): p. 181-196.
57. Lodish, H., *Molecular Cell Biology; Fifth Edition*. 2004, New York: Tenney, Sara.
58. Littau, V., et al., *Active and inactive regions of nuclear chromatin as revealed by electron microscope autoradiography.* Proceedings of the National Academy of Sciences of the United States of America, 1964. **52**(1): p. 93.
59. Sudbrak, R., et al., *X chromosome-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications.* Hum Mol Genet, 2001. **10**(1): p. 77-83.

60. Chubb, J.R., *Gene activation at the edge of the nucleus.* EMBO J, 2009. **28**(15): p. 2145-2146.
61. Cockell, M. and S.M. Gasser, *Nuclear compartments and gene regulation.* Current Opinion in Genetics & Development, 1999. **9**(2): p. 199-205.
62. Parada, L., P. McQueen, and T. Misteli, *Tissue-specific spatial organization of genomes.* Genome Biology, 2004. **5**(7): p. R44.
63. Cook, P.R., *Predicting three-dimensional genome structure from transcriptional activity.* Nat Genet, 2002. **32**(3): p. 347-352.
64. Berger, A.B., et al., *High-resolution statistical mapping reveals gene territories in live yeast.* Nat Meth, 2008. **advanced online publication**.
65. DeLaat, W.G., F., *Inter-chromosomal gene regulation in the mammalian cell nucleus.* Differentiation and gene regulation, 2007. **17**(5): p. 456-464.
66. Heard, E.B., W., *The ins and outs of gene regulation and chromosome territory organisation.* Nucleus and Gene Expression, 2007. **19**(3): p. 311-316.
67. Cremer, T. and C. Cremer, *Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells.* Nature Reviews Genetics, 2001. **2**(4): p. 292-301.
68. Branco, M.R. and A. Pombo, *Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations.* PLoS Biology, 2006. **4**(5): p. e138.
69. Carter, D., et al., *Long-range chromatin regulatory interactions in vivo.* 2002. **32**(4): p. 623-626.
70. Dekker, J., et al., *Capturing Chromosome Conformation.* Science, 2002. **295**(5558): p. 1306.
71. Krueger, C. and C.S. Osborne, *Raising the curtains on interchromosomal interactions.* Trends in Genetics, 2006. **22**(12): p. 637-639.
72. Spilianakis, C.G. and R.A. Flavell, *Long-range intrachromosomal interactions in the T helper type 2 cytokine locus.* Nat Immunol, 2004. **5**(10): p. 1017-1027.
73. Branco, M.R. and A. Pombo, *Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations.* PLoS Biol, 2006. **4**(5): p. e138.
74. Duan, Z., et al., *A three-dimensional model of the yeast genome.* Nature, 2010. **465**(7296): p. 363-367.
75. Spector, D.L., *The dynamics of chromosome organization and gene regulation.* Annual Review of Biochemistry, 2003. **72**(1): p. 573-608.
76. Bystricky, K., et al., *Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization.* J. Cell Biol., 2005. **168**(3): p. 375-387.
77. Taddei, A., H. Schober, and S.M. Gasser, *The Budding Yeast Nucleus.* Cold Spring Harbor Perspectives in Biology, 2010. **2**(8).
78. Santos, A.P. and P. Shaw, *Interphase chromosomes and the Rabl configuration: does genome size matter?* Journal of Microscopy, 2004. **214**(2): p. 201-206.
79. John, H.A., M.L. Birnstiel, and K.W. Jones, *RNA-DNA Hybrids at the Cytological Level.* Nature, 1969. **223**(5206): p. 582-587.
80. Gall, J.G. and M.L. Pardue, *Formation and detection of RNA-DNA hybrid molecules in cytological preparations.* Proc Natl Acad Sci U S A, 1969. **63**(2): p. 378-83.
81. Klever, M., et al., *Chromosomal in situ suppression hybridization after Giemsa banding.* Human Genetics, 1991. **86**(5): p. 484-486.

82. Meaburn, K.J., *Fluorescence in situ hybridization on 3D cultures of tumor cells.* Methods Mol Biol, 2010. **659**: p. 323-36.
83. Cremer, M., et al., *Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes.* Methods Mol Biol, 2008. **463**: p. 205-39.
84. Konig, K., et al., *Multiplex FISH and three-dimensional DNA imaging with near infrared femtosecond laser pulses.* Histochem Cell Biol, 2000. **114**(4): p. 337-45.
85. Scheer, U. and R. Benavente, *Functional and dynamic aspects of the mammalian nucleolus.* BioEssays, 1990. **12**(1): p. 14-21.
86. Yang, C.H., et al., *Higher order structure is present in the yeast nucleus: autoantibody probes demonstrate that the nucleolus lies opposite the spindle pole body.* Chromosoma, 1989. **98**(2): p. 123-128.
87. McKeown, P. and P. Shaw, *Chromatin: linking structure and function in the nucleolus.* Chromosoma, 2009. **118**(1): p. 11-23.
88. O'Sullivan, J.M., et al., *Repeated elements coordinate the spatial organization of the yeast genome.* Yeast, 2009. **26**(2): p. 125-138.
89. Grosberg, A. and et al., *Crumpled Globule Model of the Three-Dimensional Structure of DNA.* EPL (Europhysics Letters), 1993. **23**(5): p. 373.
90. Lieberman-Aiden, E., et al., *Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.* Science, 2009. **326**(5950): p. 289-293.
91. Lua, R., A.L. Borovinskiy, and A.Y. Grosberg, *Fractal and statistical properties of large compact polymers: a computational study.* Polymer, 2004. **45**(2): p. 717-731.
92. Grosberg, A., S.K. Nechaev, and E.I. Shakhnovich, *The role of topological limitations in the kinetics of homopolymer collapse and self-assembly of biopolymers.* Biofizika, 1988. **33**(2): p. 247-53.
93. Mateos-Langerak, J., et al., *Spatially confined folding of chromatin in the interphase nucleus.* Proceedings of the National Academy of Sciences, 2009. **106**(10): p. 3812-3817.
94. Meaburn, K.J. and T. Misteli, *Chromosome territories.* Nature 2007. **445(7126):379-381**.
95. Zink, D., et al., *Structure and dynamics of human interphase chromosome territories in vivo.* Human Genetics, 1998. **102**(2): p. 241-251.
96. Ragoczy, T., et al., *The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation.* Genes Dev., 2006. **20**(11): p. 1447-1457.
97. Osborne, C.S., et al., *Myc Dynamically and Preferentially Relocates to a Transcription Factory Occupied by Igh.* PLoS Biology, 2007. **5**(8): p. e192.
98. Schoenfelder, S., et al., *Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells.* Nat Genet, 2010. **42**(1): p. 53-61.
99. Mitchell, J.A. and P. Fraser, *Transcription factories are nuclear subcompartments that remain in the absence of transcription.* Genes Dev. , 2008. **22**(1): p. 20-25.
100. Vazquez, J., A.S. Belmont, and J.W. Sedat, *Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus.* Current Biology, 2001. **11**(16): p. 1227-1239.

101. Pidoux, A.L. and R.C. Allshire, *The role of heterochromatin in centromere function.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2005. **360**(1455): p. 569-579.
102. Morey, C., et al., *Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation.* Development, 2007. **134**(5): p. 909-919.
103. Brickner, J.H. and P. Walter, *Gene Recruitment of the Activated INO1 Locus to the Nuclear Membrane.* PLoS Biol, 2004. **2**(11): p. e342.
104. Brown, J.M., et al., *Association between active genes occurs at nuclear speckles and is modulated by chromatin environment.* The Journal of Cell Biology, 2008. **182**(6): p. 1083-1097.
105. Brown, J.M., et al., *Coregulated human globin genes are frequently in spatial proximity when active.* The Journal of Cell Biology, 2006. **172**(2): p. 177-187.
106. Tolhuis, B., et al., *Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus.* Molecular Cell, 2002. **10**(6): p. 1453-1465.
107. Crawford, G.E., et al., *DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays.* Nature Methods, 2006. **3**(7): p. 503-509.
108. Follows, G.A., et al., *Identifying gene regulatory elements by genomic microarray mapping of DNaseI hypersensitive sites.* Genome Res., 2006. **16**(10): p. 1310-1319.
109. Crawford, G.E., et al., *Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS).* Genome Res., 2006. **16**(1): p. 123-131.
110. Heydemann, A., et al., *A minimal c-fes cassette directs myeloid-specific expression in transgenic mice.* Blood, 2000. **96**(9): p. 3040-3048.
111. Ling, J.Q., et al., *CTCF Mediates Interchromosomal Colocalization Between Igf2/H19 and Wsb1/Nf1.* Science, 2006. **312**(5771): p. 269-272.
112. Dorier, J. and A. Stasiak, *The role of transcription factories-mediated interchromosomal contacts in the organization of nuclear architecture.* Nucleic Acids Research, 2010. **38**(21): p. 7410-7421.
113. West, R.W., Jr, R.R. Yocum, and M. Ptashne, *Saccharomyces cerevisiae GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UASG.* Mol. Cell. Biol., 1984. **4**(11): p. 2467-2478.
114. Dobi, K.C. and F. Winston, *Analysis of Transcriptional Activation at a Distance in Saccharomyces cerevisiae.* Mol. Cell. Biol., 2007. **27**(15): p. 5575-5586.
115. Miele, A., K. Bystricky, and J. Dekker, *Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions.* PLoS Genet, 2009. **5**(5): p. e1000478.
116. Jin, Q., J. Fuchs, and J. Loidl, *Centromere clustering is a major determinant of yeast interphase nuclear organization.* J Cell Sci, 2000. **113**(11): p. 1903-1912.
117. Reik, W. and J. Walter, *Genomic imprinting: parental influence on the genome.* Nat Rev Genet, 2001. **2**(1): p. 21-32.
118. Fuks, F., *DNA methylation and histone modifications: teaming up to silence genes.* Current Opinion in Genetics & Development, 2005. **15**(5): p. 490-495.
119. Dhar, S.S., S. Ongwijitwat, and M.T.T. Wong-Riley, *Chromosome Conformation Capture of All 13 Genomic Loci in the Transcriptional Regulation of the Multisubunit Bigenomic Cytochrome c Oxidase in Neurons.* Journal of Biological Chemistry, 2009. **284**(28): p. 18644-18650.

120. Botta, M., et al., *Intra-and inter-chromosomal interactions correlate with CTCF binding genome wide.* Molecular Systems Biology, 2010. **6**(1).
121. 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.
122. Cipak, L., M. Spirek, and J. Gregan, *Sister chromatids caught in the cohesin trap.* Nat Struct Mol Biol, 2008. **15**(9): p. 899-900.
123. Kim, Y.J., K.R. Cecchini, and T.H. Kim, *Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus.* Proceedings of the National Academy of Sciences, 2011. **108**(18): p. 7391.
124. Handoko, L., et al., *CTCF-mediated functional chromatin interactome in pluripotent cells.* Nat Genet, 2011. **43**(7): p. 630-8.
125. Phillips, J.E. and V.G. Corces, *CTCF: master weaver of the genome.* Cell, 2009. **137**(7): p. 1194-1211.
126. Kurukuti, S., et al., *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, 2006. **103**(28): p. 10684.
127. Splinter, E., et al., *CTCF mediates long-range chromatin looping and local histone modification in the -globin locus.* Genes & Development, 2006. **20**(17): p. 2349.
128. Sekimata, M., et al., *CCCTC-Binding Factor and the Transcription Factor T-bet Orchestrate T Helper 1 Cell-Specific Structure and Function at the Interferon-γ Locus.* Immunity, 2009. **31**(4): p. 551-564.
129. Valenzuela, L., N. Dhillon, and R.T. Kamakaka, *Transcription independent insulation at TFIIIC-dependent insulators.* Genetics, 2009. **183**(1): p. 131.
130. Jing, H., et al., *Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus.* Molecular cell, 2008. **29**(2): p. 232-242.
131. Mastrangelo, I.A., et al., *DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement.* Proceedings of the National Academy of Sciences, 1991. **88**(13): p. 5670.
132. Zhao, W., et al., *NF-κB– and AP-1–Mediated DNA Looping Regulates Osteopontin Transcription in Endotoxin-Stimulated Murine Macrophages.* The Journal of Immunology, 2011. **186**(5): p. 3173-3179.
133. Wall, L., E. DeBoer, and F. Grosveld, *The human beta-globin gene 3'enhancer contains multiple binding sites for an erythroid-specific protein.* Genes & Development, 1988. **2**(9): p. 1089.
134. Cullen, K.E., M.P. Kladde, and M.A. Seyfred, *Interaction between transcription regulatory regions of prolactin chromatin.* Science (New York, N Y), 1993. **261**(5118): p. 203-6.
135. 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. 81-85.
136. O'Sullivan, J.M., *Yeast chromosomal interactions and nuclear architecture.* Current Opinion in Cell Biology, 2010. **22**(3): p. 298-304.
137. Dekker, J., *A closer look at long-range chromosomal interactions.* Trends in Biochemical Sciences, 2003. **28**(6): p. 277-280.

138. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* 2006. **38**(11): p. 1348-1354.
139. Jackson, D.A., J.C. McDowell, and A. Dean, *beta-Globin locus control region HS2 and HS3 interact structurally and functionally.* Nucl. Acids Res., 2003. **31**(4): p. 1180-1190.
140. Pasceri, P., et al., *Full Activity From Human beta -Globin Locus Control Region Transgenes Requires 5'HS1, Distal beta -Globin Promoter, and 3' beta -Globin Sequences.* Blood, 1998. **92**(2): p. 653-663.
141. Fullwood, M.J., et al., *Chromatin Interaction Analysis Using Paired-End Tag Sequencing*. Current Protocols in Molecular Biology. 2010: John Wiley & Sons, Inc.
142. Botta, M., et al., *Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide.* Mol Syst Biol, 2010. **6**.
143. Tanizawa, H., et al., *Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation.* Nucleic Acids Res, 2010. **38**(22): p. 8164-77.
144. Li, G., et al., *ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing.* Genome Biology, 2010. **11**(2): p. R22.
145. Stryer, L., *Biochemistry*. 4th ed. 1995, New York: W.H. Freeman. xxxiv, 1064.
146. Goffeau, A., et al., *Life with 6000 Genes.* Science, 1996. **274**(5287): p. 546-567.
147. Gotta, M., et al., *The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae.* J. Cell Biol., 1996. **134**(6): p. 1349-1363.
148. Thompson, M., et al., *Nucleolar clustering of dispersed tRNA genes.* Science, 2003. **302**(5649): p. 1399-401.
149. Foury, F., et al., *The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae.* FEBS Letters, 1998. **440**(3): p. 325-331.
150. Williamson, D.H. and D.J. Fennell, *Visualization of yeast mitochondrial dna with the fluorescent stain "DAPI"*, in *Methods in Enzymology*, L.P. Sidney Fleischer, Editor. 1979, Academic Press. p. 728-733.
151. Goldthwaite, C.D., D.R. Cryer, and J. Marmur, *Effect of carbon source on the replication and transmission of yeast mitochondrial genomes.* Molecular and General Genetics MGG, 1974. **133**(2): p. 87-104.
152. Belcour, L., A. Sainsard-Chanet, and C.H. Sellem, *Mobile group II introns, DNA circles, reverse transcriptase and senescence.* Genetica, 1994. **93**(1): p. 225-228.
153. Timmis, J.N., et al., *Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes.* Nat Rev Genet, 2004. **5**(2): p. 123-135.
154. Ricchetti, M., C. Fairhead, and B. Dujon, *Mitochondrial DNA repairs double-strand breaks in yeast chromosomes.* Nature, 1999. **402**(6757): p. 96-100.
155. Thorsness, P.E. and T.D. Fox, *Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae.* Nature, 1990. **346**(6282): p. 376-9.
156. Morré, D.J., W.D. Merritt, and C.A. Lembi, *Connections between mitochondria and endoplasmic reticulum in rat liver and onion stem.* Protoplasma, 1971. **73**(1): p. 43-49.
157. Kornmann, B., et al., *An ER-Mitochondria Tethering Complex Revealed by a Synthetic Biology Screen.* Science, 2009. **325**(5939): p. 477-481.
158. Consortium, T.R.C.S., *In-Depth View of Structure, Activity, and Evolution of Rice Chromosome 10.* Science, 2003. **300**(5625): p. 1566-1569.

159. Martin, W., *Gene transfer from organelles to the nucleus: frequent and in big chunks.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8612-4.
160. Liang, B.C., *Evidence for association of mitochondrial DNA sequence amplification and nuclear localization in human low-grade gliomas.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 1996. **354**(1): p. 27-33.
161. Ricchetti, M., F. Tekaia, and B. Dujon, *Continued colonization of the human genome by mitochondrial DNA.* PLoS Biol, 2004. **2**(9): p. E273.
162. Thorsness, P.E. and T.D. Fox, *Nuclear Mutations in Saccharomyces cerevisiae That Affect the Escape of DNA from Mitochondria to the Nucleus.* Genetics, 1993. **134**(1): p. 21-28.
163. Shafer, K.S., et al., *Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae.* Curr Genet, 1999. **36**(4): p. 183-94.
164. Thorsness, P.E. and T.D. Fox, *Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae.* Nature, 1990. **346**(6282): p. 376-379.
165. Campbell, C. and P. Thorsness, *Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments*. 1998. p. 2455-2464.
166. Shafer, K.S., et al., *Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae.* Current Genetics, 1999. **36**(4): p. 183-194.
167. Blank, H.M., et al., *An Increase in Mitochondrial DNA Promotes Nuclear DNA Replication in Yeast.* PLoS Genet, 2008. **4**(4): p. e1000047.
168. Butow, R.A. and N.G. Avadhani, *Mitochondrial Signaling: The Retrograde Response.* Molecular Cell, 2004. **14**(1): p. 1-15.
169. Liu, Z. and R.A. Butow, *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., 1999. **19**(10): p. 6720-6728.
170. Zuo, X.M., G.D. Clark-Walker, and X.J. Chen, *The Mitochondrial Nucleoid Protein, Mgm101p, of Saccharomyces cerevisiae Is Involved in the Maintenance of {rho}+ and ori/rep-Devoid Petite Genomes but Is Not Required for Hypersuppressive {rho}- mtDNA.* Genetics, 2002. **160**(4): p. 1389-1400.
171. Bogenhagen, D.F., D. Rousseau, and S. Burke, *The Layered Structure of Human Mitochondrial DNA Nucleoids.* J. Biol. Chem., 2008. **283**(6): p. 3665-3675.
172. Lagunas, R., *Misconceptions about the energy metabolism of Saccharomyces cerevisiae.* Yeast, 1986. **2**(4): p. 221-228.
173. Lagunas, R., *Energy metabolism of Saccharomyces cerevisiae discrepancy between ATP balance and known metabolic functions.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1976. **440**(3): p. 661-674.
174. Futcher, A.B. and B.S. Cox, *Maintenance of the 2 microns circle plasmid in populations of Saccharomyces cerevisiae.* J. Bacteriol., 1983. **154**(2): p. 612-622.
175. Mehta, S., et al., *The 2 micron plasmid purloins the yeast cohesin complex: a mechanism for coupling plasmid partitioning and chromosome segregation?* J. Cell Biol., 2002. **158**(4): p. 625-637.
176. Yeh, E. and K. Bloom, *Hitching a ride.* EMBO Reports, 2006. **(10): 985–987.**
177. Scott-Drew, S. and J.A. Murray, *Localisation and interaction of the protein components of the yeast 2 mu circle plasmid partitioning system suggest a mechanism for plasmid inheritance.* J Cell Sci, 1998. **111**(13): p. 1779-1789.

178. McLeod, M., S. Craft, and J.R. Broach, *Identification of the crossover site during FLP-mediated recombination in the Saccharomyces cerevisiae plasmid 2 microns circle.* Mol Cell Biol, 1986. **6**(10): p. 3357-67.
179. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular cloning: A laboratory manual, 2nd ed.* . Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., 1989.
180. Varadaraj, K. and D.M. Skinner, *Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases.* Gene, 1994. **140**(1): p. 1-5.
181. Barnes, W.M., *The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion.* Gene, 1992(112): p. 29–35.
182. Quail, M.A., et al., *A large genome center's improvements to the Illumina sequencing system.* Nat Meth, 2008. **5**(12): p. 1005-1010.
183. Mardis, E.R., *The impact of next-generation sequencing technology on genetics.* Trends in Genetics, 2008. **24**(3): p. 133-141.
184. Altschul, S., Gish W, Miller W, Myers EW, Lipman DJ., *Basic local alignment search tool.* J Mol Biol., 1990. **5;215(3):403-10**.
185. Cui, H., S.K. Ghosh, and M. Jayaram, *The selfish yeast plasmid uses the nuclear motor Kip1p but not Cin8p for its localization and equal segregation.* The Journal of Cell Biology, 2009. **185**(2): p. 251-264.
186. O'Sullivan, J.M., et al., *Gene loops juxtapose promoters and terminators in yeast.* Nature genetics, 2004. **36**(9): p. 1014-8.
187. Berger, A.B., et al., *High-resolution statistical mapping reveals gene territories in live yeast.* Nat Meth, 2008. **5**(12): p. 1031-1037.
188. Proffitt, J.H., et al., *5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA.* Mol. Cell. Biol., 1984. **4**(5): p. 985-988.
189. Woda, C.B., et al., *Ontogeny of flow-stimulated potassium secretion in rabbit cortical collecting duct: functional and molecular aspects.* American Journal of Physiology - Renal Physiology, 2003. **285**(4): p. F629-F639.
190. Dugaiczyk, A., H.W. Boyer, and H.M. Goodman, *Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures.* Journal of Molecular Biology, 1975. **96**(1): p. 171-178.
191. Watson, M.L., *The Nuclear Envelope.* The Journal of Biophysical and Biochemical Cytology, 1955. **1**(3): p. 257-270.
192. Thomsen, R., et al., *Localization of nuclear retained mRNAs in Saccharomyces cerevisiae.* Rna, 2003. **9**(9): p. 1049-57.
193. Haeusler, R.A., et al., *Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes.* Genes Dev, 2008. **22**(16): p. 2204-14.
194. Gotta, M., et al., *The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae.* J Cell Biol, 1996. **134**(6): p. 1349-63.
195. Heun, P., et al., *Chromosome Dynamics in the Yeast Interphase Nucleus.* Science 2001. **294**(5549): p. 2181-2186.
196. Bystricky, K., et al., *Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization.* J Cell Biol, 2005. **168**(3): p. 375-87.
197. Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. **295**(5558): p. 1306-11.

198. Haber, J.E. and W.Y. Leung, *Lack of chromosome territoriality in yeast: promiscuous rejoining of broken chromosome ends.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13949-54.
199. Hempel, W.M. and P. Ferrier, *Restriction Endonuclease Accessibility as a Determinant of Altered Chromatin Structure*. 2004. p. 53-63.
200. Johnston, M., *A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae.* Microbiol. Mol. Biol. Rev., 1987. **51**(4): p. 458-476.
201. MacIsaac, K., et al., *An improved map of conserved regulatory sites for Saccharomyces cerevisiae.* BMC Bioinformatics, 2006. **7**(1): p. 113.
202. Yeh, E. and K. Bloom, *Hitching a ride.* EMBO Rep, 2006. **7**(10): p. 985-7.
203. Mehta, S., et al., *A novel role for the mitotic spindle during DNA segregation in yeast: promoting 2 microm plasmid-cohesin association.* Mol Cell Biol, 2005. **25**(10): p. 4283-98.
204. Scott-Drew, S., C.M. Wong, and J.A. Murray, *DNA plasmid transmission in yeast is associated with specific sub-nuclear localisation during cell division.* Cell Biol Int, 2002. **26**(5): p. 393-405.
205. Mehta, S., et al., *The 2 micron plasmid purloins the yeast cohesin complex: a mechanism for coupling plasmid partitioning and chromosome segregation?* J Cell Biol, 2002. **158**(4): p. 625-37.
206. Timmis, J.N., et al., *Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes.* Nat Rev Genet, 2004. **5**(2): p. 123-35.
207. The, R.C.S.C., *In-depth view of structure, activity, and evolution of rice chromosome 10.* Science, 2003. **300**(5625): p. 1566-9.
208. Stegemann, S., et al., *High-frequency gene transfer from the chloroplast genome to the nucleus.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8828-33.
209. Traven, A., et al., *Interorganellar communication. Altered nuclear gene expression profiles in a yeast mitochondrial dna mutant.* J Biol Chem, 2001. **276**(6): p. 4020-7.
210. Futcher, A.B., *Copy number amplification of the 2 micron circle plasmid of Saccharomyces cerevisiae.* J Theor Biol, 1986. **119**(2): p. 197-204.
211. Espada, J. and M. Esteller, *Epigenetic control of nuclear architecture.* Cellular and Molecular Life Sciences (CMLS), 2007. **64**(4): p. 449-457.
212. Kim, S.-I., E.H. Bresnick, and S.J. Bultman, *BRG1 directly regulates nucleosome structure and chromatin looping of the {alpha} globin locus to activate transcription.* Nucl. Acids Res., 2009: p. gkp677.
213. 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. 2969-2978.
214. Harbison, C.T., et al., *Transcriptional regulatory code of a eukaryotic genome.* Nature, 2004. **431**(7004): p. 99-104.
215. Otterstedt, K., et al., *Switching the mode of metabolism in the yeast Saccharomyces cerevisiae.* Embo Reports, 2004. **5**(5): p. 532-537.
216. Cheng, X. and A.S. Ivessa, *The migration of mitochondrial DNA fragments to the nucleus affects the chronological aging process of Saccharomyces cerevisiae.* Aging Cell, 2010. **9999**(999A).
217. Singh, B.N. and M. Hampsey, *A Transcription-Independent Role for TFIIB in Gene Looping.* Molecular Cell, 2007. **27**(5): p. 806-816.
218. Parikh, V., et al., *The mitochondrial genotype can influence nuclear gene expression in yeast.* Science, 1987. **235**(4788): p. 576-580.
219. Parikh, V.S., et al., *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., 1989. **9**(5): p. 1897-1907.
220. Coelho, P.S.R., et al., *A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear 25S rDNA.* Genes & Development, 2002. **16**(21): p. 2755-2760.
221. Campbell, C., et al., *Mitochondrial morphological and functional defects in yeast caused by yme1 are suppressed by mutation of a 26S protease subunit homologue.* Mol. Biol. Cell, 1994. **5**(8): p. 899-905.
222. Neutzner, A. and R.J. Youle, *Instability of the Mitofusin Fzo1 Regulates Mitochondrial Morphology during the Mating Response of the Yeast Saccharomyces cerevisiae.* Journal of Biological Chemistry, 2005. **280**(19): p. 18598-18603.
223. Fekkes, P., K.A. Shepard, and M.P. Yaffe, *Gag3p, an outer membrane protein required for fission of mitochondrial tubules.* J Cell Biol, 2000. **151**(2): p. 333-40.
224. Hanekamp, T., et al., *Maintenance of Mitochondrial Morphology Is Linked to Maintenance of the Mitochondrial Genome in Saccharomyces cerevisiae.* Genetics, 2002. **162**(3): p. 1147-1156.
225. Park, S., et al., *Yme2p is a mediator of nucleoid structure and number in mitochondria of the yeast Saccharomyces cerevisiae.* Current Genetics, 2006. **50**(3): p. 173-182.
226. Eskes, R., et al., *Mobility of Yeast Mitochondrial Group II Introns: Engineering a New Site Specificity and Retrohoming via Full Reverse Splicing.* 1997. **88**(6): p. 865-874.
227. Yang, J., et al., *Efficient integration of an intron RNA into double-stranded DNA by reverse splicing.* Nature, 1996. **381**(6580): p. 332-335.
228. Zimmerly, S., et al., *A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility.* 1995. **83**(4): p. 529-538.
229. Zimmerly, S., et al., *Group II intron mobility occurs by target DNA-primed reverse transcription.* 1995. **82**(4): p. 545-554.
230. Amiott, E.A. and J.A. Jaehning, *Mitochondrial Transcription Is Regulated via an ATP "Sensing" Mechanism that Couples RNA Abundance to Respiration.* Molecular Cell, 2006. **22**(3): p. 329-338.
231. Bassi, G.S., et al., *Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(1): p. 128-133.
232. Adams, K.L., et al., *Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(15): p. 9905-9912.
233. Adams, K.L., et al., *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, 1999. **96**(24): p. 13863-13868.
234. Brennicke, A., et al., *The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants.* FEBS Letters, 1993. **325**(1-2): p. 140-145.
235. Farrelly, F. and R.A. Butow, *Rearranged mitochondrial genes in the yeast nuclear genome.* Nature, 1983. **301**(5898): p. 296-301.

236. Hazkani-Covo, E., R.M. Zeller, and W. Martin, *Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes.* PLoS Genet, 2010. **6**(2): p. e1000834.
237. Richly, E. and D. Leister, *NUMTs in Sequenced Eukaryotic Genomes.* Mol Biol Evol, 2004. **21**(6): p. 1081-1084.
238. Shirafuji, N., et al., *Mitochondrial Antisense RNA for Cytochrome C Oxidase (MARCO) Can Induce Morphologic Changes and Cell Death in Human Hematopoietic Cell Lines.* Blood, 1997. **90**(11): p. 4567-4577.
239. Tarassov, I.A. and N.S. Entelis, *Mitochondrially-imported cytoplasmic tRNALys(CUU) of Saccharomyces cerevisiae: in vivo and in vitro targetting systems.* Nucl. Acids Res., 1992. **20**(6): p. 1277-1281.
240. Cavalier-Smith, T., *The Simultaneous Symbiotic Origin of Mitochondria, Chloroplasts, and Microbodies.* Annals of the New York Academy of Sciences, 1987. **503**(1): p. 55-71.
241. DeRisi, J.L., V.R. Iyer, and P.O. Brown, *Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale.* Science, 1997. **278**(5338): p. 680-686.
242. Roberts, G. and A. Hudson, *Transcriptome profiling of Saccharomyces cerevisiae during a transition from fermentative to glycerol-based respiratory growth reveals extensive metabolic and structural remodeling.* Molecular Genetics and Genomics, 2006. **276**(2): p. 170-186.
243. Kennell, J.C., et al., *Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria.* Cell, 1993. **73**(1): p. 133-146.
244. Knoop, V., et al., *copia-, gypsy- and LINE-Like Retrotransposon Fragments in the Mitochondrial Genome of Arabidopsis thaliana.* Genetics, 1996. **142**(2): p. 579-585.
245. Marienfeld, J., M. Unseld, and A. Brennicke, *The mitochondrial genome of Arabidopsis is composed of both native and immigrant information.* Trends in Plant Science, 1999. **4**(12): p. 495-502.
246. Costanzo, M.C. and T.D. Fox, *Transformation of Yeast by Agitation With Glass Beads.* Genetics, 1988. **120**(3): p. 667-670.
247. Ito, H., et al., *Transformation of intact yeast cells treated with alkali cations.* J. Bacteriol., 1983. **153**(1): p. 163-168.
248. Shibata, Y., et al., *Mechanisms Determining the Morphology of the Peripheral ER.* Cell, 2010. **143**(5): p. 774-788.
249. Costanzo, M.C. and T.D. Fox, *Control of Mitochondrial Gene Expression in Saccharomyces Cerevisiae.* Annual Review of Genetics, 1990. **24**(1): p. 91-108.
250. Christianson, T. and M. Rabinowitz, *Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase.* Journal of Biological Chemistry, 1983. **258**(22): p. 14025-14033.
251. Barth, C., et al., *Polycistronic transcription and editing of the mitochondrial small subunit (SSU) ribosomal RNA in Dictyostelium discoideum.* Current Genetics, 1999. **36**(1): p. 55-61.
252. Holec, S., et al., *Relaxed Transcription in Arabidopsis Mitochondria Is Counterbalanced by RNA Stability Control Mediated by Polyadenylation and Polynucleotide Phosphorylase.* Mol. Cell. Biol., 2006. **26**(7): p. 2869-2876.
253. Montoya, J., et al., *Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA.* Proceedings of the National

Academy of Sciences of the United States of America, 1982. **79**(23): p. 7195-7199.
254. Butow, R.A., R. Docherty, and V.S. Parikh, *A Path from Mitochondria to the Yeast Nucleus.* Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1988. **319**(1193): p. 127-133.
255. Mayan, M. and L. Aragon, *Cis-interactions between non-coding ribosomal spacers dependent on RNAP-II separate RNAP-I and RNAP-III transcription domains.* Cell Cycle, 2010. **9**(21): p. 4328-37.
256. Blank, H.M., et al., *An Increase in Mitochondrial DNA Promotes Nuclear DNA Replication in Yeast.* PLoS Genetics, 2008. **4**(4): p. e1000047.
257. Cerutti, L. and V. Simanis, *Controlling the end of the cell cycle.* Current Opinion in Genetics & Development, 2000. **10**(1): p. 65-69.
258. Boisvert, F.-M., et al., *The multifunctional nucleolus.* Nat Rev Mol Cell Biol, 2007. **8**(7): p. 574-585.
259. Towpik, J., *Regulation of mitochondrial translation in yeast.* Cell Mol Biol Lett, 2005. **10**(4): p. 571-94.
260. Campbell, R.N., et al., *Metabolic control of transcription: paradigms and lessons from Saccharomyces cerevisiae.* Biochem J, 2008. **414**(2): p. 177-187.
261. Sanz, P., *Yeast as a model system to study glucose-mediated signalling and response.* Front. Biosci, 2007. **12**: p. 2358–2371.
262. Gancedo, J.M., *Yeast Carbon Catabolite Repression.* Microbiol. Mol. Biol. Rev., 1998. **62**(2): p. 334-361.
263. Brickner, J.H. and P. Walter, *Gene Recruitment of the Activated <italic>INO1</italic> Locus to the Nuclear Membrane.* PLoS Biol, 2004. **2**(11): p. e342.
264. Bartlett, O., et al., *Specialized transcription factories.* Transcription, 2006: p. 67-75.
265. Petrascheck, M., et al., *DNA looping induced by a transcriptional enhancer in vivo.* Nucleic Acids Research. **33**(12): p. 3743-3750.
266. Bonawitz, N.D., M.S. Rodeheffer, and G.S. Shadel, *Defective Mitochondrial Gene Expression Results in Reactive Oxygen Species-Mediated Inhibition of Respiration and Reduction of Yeast Life Span.* Mol. Cell. Biol., 2006. **26**(13): p. 4818-4829.
267. Ragoczy, T., et al., *The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation*
10.1101/gad.1419506. Genes Dev., 2006. **20**(11): p. 1447-1457.
268. Cheutin, T., et al., *In vivo dynamics of Swi6 in yeast: evidence for a stochastic model of heterochromatin.* Mol Cell Biol, 2004. **24**(8): p. 3157-67.
269. Cheutin, T., et al., *Maintenance of stable heterochromatin domains by dynamic HP1 binding.* Science, 2003. **299**(5607): p. 721-5.
270. Hediger, F. and S.M. Gasser, *Nuclear organization and silencing: putting things in their place.* Nat Cell Biol, 2002. **4**(3): p. E53-E55.
271. Pilauri, V., et al., *Gal80 Dimerization and the Yeast GAL Gene Switch.* Genetics, 2005. **169**(4): p. 1903-1914.
272. Kaplan, N., et al., *The DNA-encoded nucleosome organization of a eukaryotic genome.* Nature, 2009. **458**(7236): p. 362-366.
273. Chen, X.J. and R.A. Butow, *The organization and inheritance of the mitochondrial genome.* Nat Rev Genet, 2005. **6**(11): p. 815-825.

274. MacAlpine, D.M., P.S. Perlman, and R.A. Butow, *The numbers of individual mitochondrial DNA molecules and mitochondrial DNA nucleoids in yeast are co-regulated by the general amino acid control pathway.* EMBO J, 2000. **19**(4): p. 767-775.
275. Scott-Drew, S., C.M.V.L. Wong, and J.A.H. Murray, *DNA plasmid transmission in yeast is associated with specific sub-nuclear localisation during cell division.* Cell Biology International, 2002. **26**(5): p. 393-405.
276. O'Sullivan, J.M., et al., *Gene loops juxtapose promoters and terminators in yeast.* Nature Genetics, 2004. **36**(9): p. 1014-1018.
277. Lohr, D., P. Venkov, and J. Zlatanova, *Transcriptional regulation in the yeast GAL gene family: a complex genetic network.* The FASEB Journal, 1995. **9**(9): p. 777-787.
278. Scott, S., et al., *Functional Domain Mapping and Subcellular Distribution of Dal82p in Saccharomyces cerevisiae.* Journal of Biological Chemistry, 2000. **275**(10): p. 7198-7204.
279. Peter, G.J., L. Düring, and A. Ahmed, *Carbon Catabolite Repression Regulates Amino Acid Permeases in Saccharomyces cerevisiae via the TOR Signaling Pathway.* Journal of Biological Chemistry, 2006. **281**(9): p. 5546-5552.
280. Kobayashi, T., *The Replication Fork Barrier Site Forms a Unique Structure with Fob1p and Inhibits the Replication Fork.* Mol. Cell. Biol., 2003. **23**(24): p. 9178-9188.
281. Rodley, C.D.M., et al., *Working title: tRNA positional effects* In Preparation, 2011.
282. Siebert, P.D., et al., *An improved PCR method for walking in uncloned genomic DNA.* Nucleic Acids Research, 1995. **23**(6): p. 1087-1088.
283. Shagin, D.A., et al., *Regulation of average length of complex PCR product.* Nucleic Acids Research, 1999. **27**(18): p. e23-i-e23-iii.
284. Noma, K., et al., *A role for TFIIIC transcription factor complex in genome organization.* Cell, 2006. **125**(5): p. 859-872.
285. Barberis, A., et al., *Contact with a component of the polymerase II holoenzyme suffices for gene activation.* Cell, 1995. **81**(3): p. 359-368.
286. Tajima, M., Y. Nogi, and T. Fukasawa, *Duplicate upstream activating sequences in the promoter region of the Saccharomyces cerevisiae GAL7 gene.* Mol. Cell. Biol., 1986. **6**(1): p. 246-256.
287. Brand, A.H., G. Micklem, and K. Nasmyth, *A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation.* Cell, 1987. **51**(5): p. 709-719.
288. Prescott, E.M. and N.J. Proudfoot, *Transcriptional collision between convergent genes in budding yeast.* Proceedings of the National Academy of Sciences, 2002. **99**(13): p. 8796-8801.
289. Weaver, R.F., *Molecular biology*. 3rd ed. 2005, Boston: McGraw-Hill. xvii, 894.
290. Rodriguez, A., et al., *The hexokinase 2 protein regulates the expression of the GLK1, HXK1 and HXK2 genes of Saccharomyces cerevisiae.* Biochem J, 2001. **355**(Pt 3): p. 625-31.
291. Morrow, B.E., S.P. Johnson, and J.R. Warner, *The rRNA enhancer regulates rRNA transcription in Saccharomyces cerevisiae.* Mol. Cell. Biol., 1993. **13**(2): p. 1283-1289.
292. Kief, D.R. and J.R. Warner, *Coordinate control of syntheses of ribosomal ribonucleic acid and ribosomal proteins during nutritional shift-up in Saccharomyces cerevisiae.* Mol. Cell. Biol., 1981. **1**(11): p. 1007-1015.

293. Waldron, C. and F. Lacroute, *Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast.* J. Bacteriol., 1975. **122**(3): p. 855-865.
294. Ansari, A. and M. Hampsey, *A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping*
10.1101/gad.1362305. Genes Dev., 2005. **19**(24): p. 2969-2978.
295. 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.
296. Kuo, M.-H. and C.D. Allis, *In Vivo Cross-Linking and Immunoprecipitation for Studying Dynamic Protein:DNA Associations in a Chromatin Environment.* Methods, 1999. **19**(3): p. 425-433.
297. Unger, M. and L. Hartwell, *Control of cell division in Saccharomyces cerevisiae by methionyl-tRNA.* Proceedings of the National Academy of Sciences of the United States of America, 1976. **73**(5): p. 1664.
298. Brauer, M.J., et al., *Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast.* Mol. Biol. Cell, 2008. **19**(1): p. 352-367.
299. Glynn, E.F., et al., *Genome-Wide Mapping of the Cohesin Complex in the Yeast Saccharomyces cerevisiae.* PLoS Biology, 2004. **2**(9): p. e259.
300. Cremer, T. and C. Cremer, *Chromosome territories, nuclear architecture and gene regulation in mammalian cells.* Nat Rev Genet, 2001. **2**(4): p. 292-301.
301. Barabási, A.-L. and R. Albert, *Emergence of Scaling in Random Networks.* Science, 1999. **286**(5439): p. 509-512.
302. Kellis, M., B.W. Birren, and E.S. Lander, *Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae.* Nature, 2004. **428**(6983): p. 617-624.
303. Guan, Y., M.J. Dunham, and O.G. Troyanskaya, *Functional analysis of gene duplications in Saccharomyces cerevisiae.* Genetics, 2006.
304. Felsenfeld, G., *Chromatin Unfolds.* Cell, 1996. **86**(1): p. 13-19.
305. Trumtel, S., et al., *Assembly and Functional Organization of the Nucleolus: Ultrastructural Analysis of Saccharomyces cerevisiae Mutants.* Mol. Biol. Cell, 2000. **11**(6): p. 2175-2189.
306. Taddei, A. and S.M. Gasser, *Multiple pathways for telomere tethering: functional implications of subnuclear position for heterochromatin formation.* Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 2004. **1677**(1-3): p. 120-128.
307. Pombo, A., et al., *Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III.* EMBO J, 1999. **18**(8): p. 2241-2253.
308. Haldar, D. and R.T. Kamakaka, *tRNA genes as chromatin barriers.* 2006.
309. McFarlane, R.J. and S.K. Whitehall, *tRNA genes in eukaryotic genome organization and reorganization.* Cell cycle (Georgetown, Tex.), 2009. **8**(19): p. 3102.
310. Hull, M.W., et al., *tRNA genes as transcriptional repressor elements.* Molecular and cellular biology, 1994. **14**(2): p. 1266.
311. Haeusler, R.A., et al., *Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes.* Genes & Development, 2008. **22**(16): p. 2204-2214.
312. Kendall, A., et al., *A CBF5 mutation that disrupts nucleolar localization of early tRNA biosynthesis in yeast also suppresses tRNA gene-mediated transcriptional silencing.* Proceedings of the National Academy of Sciences, 2000. **97**(24): p. 13108-13113.

313. Lee, Y., et al., *A transcriptional mediator protein that is required for activation of many RNA polymerase II promoters and is conserved from yeast to humans.* Mol. Cell. Biol., 1997. **17**(8): p. 4622-4632.
314. Li, Q., et al., *Locus control regions.* Blood, 2002. **100**(9): p. 3077-3086.
315. Soontorngun, N., et al., *Regulation of Gluconeogenesis in Saccharomyces cerevisiae Is Mediated by Activator and Repressor Functions of Rds2.* Mol. Cell. Biol., 2007. **27**(22): p. 7895-7905.
316. Klar, A.J.S. and H.O. Halvorson, *Effect of GAL4 Gene Dosage on the Level of Galactose Catabolic Enzymes in Saccharomyces cerevisiae.* J. Bacteriol., 1976. **125**(1): p. 379-381.
317. Butow, R.A. and N.G. Avadhani, *Mitochondrial signaling: the retrograde response.* Mol Cell, 2004. **14**(1): p. 1-15.
318. Janes, M.S., et al., *Rapid Analysis of Mitochondrial DNA Depletion by Fluorescence In Situ Hybridization and Immunocytochemistry: Potential Strategies for HIV Therapeutic Monitoring.* Journal of Histochemistry & Cytochemistry, 2004. **52**(8): p. 1011-1018.
319. Anderson, P. and N. Kedersha, *RNA granules.* The Journal of Cell Biology, 2006. **172**(6): p. 803-808.
320. Nilsson, D. and P. Sunnerhagen, *Cellular stress induces cytoplasmic RNA granules in fission yeast.* RNA, 2011. **17**(1): p. 120-133.
321. Bloom, J. and F.R. Cross, *Novel Role for Cdc14 Sequestration: Cdc14 Dephosphorylates Factors That Promote DNA Replication.* Mol. Cell. Biol., 2007. **27**(3): p. 842-853.
322. Shannon, A., et al., *Systemic RNAi mediated gene silencing in the anhydrobiotic nematode Panagrolaimus superbus.* BMC Molecular Biology, 2008. **9**(1): p. 58.
323. Strunnikov, A.V., *A case of selfish nucleolar segregation.* Cell Cycle, 2005. **4**(1): p. 113-7.
324. Bernardi, R., et al., *PML regulates p53 stability by sequestering Mdm2 to the nucleolus.* Nat Cell Biol, 2004. **6**(7): p. 665-672.
325. Klibanov, S., H. O'Hagan, and M. Ljungman, *Accumulation of soluble and nucleolar-associated p53 proteins following cellular stress.* J Cell Sci, 2001. **114**(10): p. 1867-1873.
326. Bell, S.P., R. Kobayashi, and B. Stillman, *Yeast origin recognition complex functions in transcription silencing and DNA replication.* Science, 1993. **262**(5141): p. 1844-9.
327. Gardner, K.A., J. Rine, and C.A. Fox, *A Region of the Sir1 Protein Dedicated to Recognition of a Silencer and Required for Interaction with the Orc1 Protein in Saccharomyces cerevisiae.* Genetics, 1999. **151**(1): p. 31-44.
328. Fritze, C.E., et al., *Direct evidence for SIR2 modulation of chromatin structure in yeast rDNA.* EMBO J, 1997. **16**(21): p. 6495-6509.
329. Chatre, L. and M. Ricchetti, *Nuclear Mitochondrial DNA Activates Replication in Saccharomyces cerevisiae.* PLoS ONE, 2011. **6**(3): p. e17235.
330. Nugent, J.M. and J.D. Palmer, *RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution.* Cell, 1991. **66**(3): p. 473-481.
331. Dinger, M.E., T.R. Mercer, and J.S. Mattick, *RNAs as extracellular signaling molecules.* J Mol Endocrinol, 2008. **40**(4): p. 151-159.
332. Hotopp, J.C.D., et al., *Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes.* Science, 2007. **317**(5845): p. 1753-1756.

333. Watts, J.M., et al., *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-716.
334. Ros, V. and G. Hurst, *Lateral gene transfer between prokaryotes and multicellular eukaryotes: ongoing and significant?* BMC Biology, 2009. **7**(1): p. 20.
335. Zupan, J.R. and P. Zambryski, *Transfer of T-DNA from Agrobacterium to the Plant Cell.* Plant Physiology, 1995. **107**(4): p. 1041-1047.
336. Kobayashi, T., M. Nomura, and T. Horiuchi, *Identification of DNA cis Elements Essential for Expansion of Ribosomal DNA Repeats in Saccharomyces cerevisiae.* Mol. Cell. Biol., 2001. **21**(1): p. 136-147.
337. Liang, B.C. and L. Hays, *Mitochondrial DNA copy number changes in human gliomas.* Cancer Letters, 1996. **105**(2): p. 167-173.
338. Alirol, E. and J.C. Martinou, *Mitochondria and cancer: is there a morphological connection?* Oncogene, 2010. **25**(34): p. 4706-4716.
339. Burzio, V.A., et al., *Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells.* Proceedings of the National Academy of Sciences, 2009. **106**(23): p. 9430-9434.
340. Hadler, H.I., *Comment: mitochondrial genes and cancer.* FEBS Letters, 1989. **256**(1-2): p. 230-231.
341. Shay, J.W. and H. Werbin, *New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging.* Mutation Research/DNAging, 1992. **275**(3-6): p. 227-235.
342. Wallace, D.C., *A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A Dawn for Evolutionary Medicine.* Annual Review of Genetics, 2005. **39**(1): p. 359-407.
343. Ostergaard, S., L. Olsson, and J. Nielsen, *Metabolic Engineering of Saccharomyces cerevisiae.* Microbiol. Mol. Biol. Rev., 2000. **64**(1): p. 34-50.
344. Snoek, I.S.I. and H.Y. Steensma, *Factors involved in anaerobic growth of Saccharomyces cerevisiae.* Yeast, 2007. **24**(1): p. 1-10.
345. Janga, S.C., J. Collado-Vides, and M.M. Babu, *Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes.* Proceedings of the National Academy of Sciences, 2008. **105**(41): p. 15761-15766.
346. Santos-Rosa, H. and C. Caldas, *Chromatin modifier enzymes, the histone code and cancer.* European Journal of Cancer, 2005. **41**(16): p. 2381-2402.
347. Grewal, S.I. and D. Moazed, *Heterochromatin and epigenetic control of gene expression.* Science, 2003. **301**(5634): p. 798-802.
348. Fahrner, J.A., et al., *Dependence of Histone Modifications and Gene Expression on DNA Hypermethylation in Cancer.* Cancer Research, 2002. **62**(24): p. 7213-7218.
349. Bungert, J., et al., *Hypersensitive Site 2 Specifies a Unique Function within the Human beta -Globin Locus Control Region To Stimulate Globin Gene Transcription.* Mol. Cell. Biol., 1999. **19**(4): p. 3062-3072.
350. Cioe, L., et al., *Differential Expression of the Globin Genes in Human Leukemia K562(S) Cells Induced to Differentiate by Hemin or Butyric Acid.* Cancer Res, 1981. **41**(1): p. 237-243.
351. Schoenfelder, S., I. Clay, and P. Fraser, *The transcriptional interactome: gene expression in 3D.* Current Opinion in Genetics & Development, 2010. **20**(2): p. 127 - 133.
352. Davido, T. and R.H. Getzenberg, *Nuclear matrix proteins as cancer markers.* Journal of cellular biochemistry Supplement, 2000. **Suppl 35**: p. 136-41.

353. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation.* Nat Rev Cancer, 2007. **7**(4): p. 233-245.
354. Rieder, H., et al., *High rate of chromosome abnormalities detected by fluorescence in situ hybridization using BCR and ABL probes in adult acute lymphoblastic leukemia.* Leukemia, 1998. **12**(9): p. 1473-1481.
355. Scherthan, H., *FISH Targeting of Chromosomes and Subchromosomal Regions in Yeast*, in *Fluorescence In Situ Hybridization (FISH) — Application Guide*, T. Liehr, Editor. 2009, Springer Berlin Heidelberg. p. 347-359.