

Supplementary Material

Mitochondrial-nuclear DNA interactions regulate nuclear transcription.

C.D.M. Rodley^a, R.S. Grand^a, B. Jones^b, J.M O'Sullivan^{a†}

^aInstitute of Natural Sciences, Massey University, Albany, New Zealand

^bCentre 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

Contents

Supplementary Tables	Page 2
Supplementary Figures	Page 6
Supplementary Methods	Page 12

Supplementary Tables

Supplementary Table 1:

Strain	Genotype and Comments	Background
BY4741	<i>Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	Wildtype
161-U7	1 ⁺ 2 ⁺ [MATa, ade1, lys1, ura3]. Wildtype (Moran et al. 1995)	161-U7
161-U7 GII-0 al5y	GII-0 [MATa, ade1, lys1, ura3]. No group II introns al5y retained (Huang et al. 2005; Moran et al. 1995)	161-U7
161-U7 GII-0	GII-0 [MATa, ade1, lys1, ura3]. No group II introns (Huang et al. 2005; Moran et al. 1995)	161-U7
<i>yme1</i>	<i>Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yme1Δ::kanMX4</i>	BY4741
<i>yme2</i>	<i>Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yme2Δ::kanMX4</i>	BY4741
<i>mdv1</i>	<i>Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mdv1Δ::kanMX4</i>	BY4741
<i>mrs1</i>	<i>Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mrs1Δ::kanMX4</i>	BY4741
<i>ade2</i>	<i>Mata his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ade2Δ::kanMX4</i>	BY4741

Supplementary Table 2:

Name	Sequence	Comments
MITOCHONDRIA COPY NUMBER		
Mito+ve13909F	TGCTCAACGAAAGTGAATCAA	
Mito(CNC)R13909	GATTTATCGTATGCTCATTTCCTAA	
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	CTGTCTTCCTTCCTTGCAATTT	
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	AATATCCTCAATTAAGAGGTCTGAA	
nDNA-nDNA 3C TAQMAN		
gDNAgDNA3CForward	TGACACCGTCTCTTGTTTAGGA	
gDNAgDNA3CReverse	TTGATCGTATCCTTCTCTAGTGAAC	
gDNAgDNA3CProbe	FAM-TTAACTTCAGTTAAATCTTCAA	
Mito-nDNA 3C TAQMAN		
MitogDNA3CForward	GTGAGCCGTATGCGATGAAAG	
MitogDNA3CR13221	GAATCCCTCGCCAACATAGA	
MitogDNA3CProbe	FAM-TCGCACGTACGGTTCTTACCGG	
LIGATION CONTROLS		
E.coli211bp3'MspIF	GCCAGAAATTCGTCCGGTAAG	
E.coli211bp3'MspIR	AACCGGTCATTGAAGTATTGA	
Lambda185bp3'MspIF	TTTACAGCGTGATGGAGCAG	
Lambda185bp3'MspIR	ACCAATCCAGCCGGTCAG	
Mito-nDNA POSITIVES		

Mito+ve13909F	TGCTCAACGAAAGTGAATCAA	
MitoAcross13909R	TCCCGATAGGTAGACCTTTACAA	
MitogDNA3CR13221	GAATCCCTCGCCAACATAGA	
gDNA+ve13221F	CATCCATCTCGCAGCAATTA	
nDNA-nDNA 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	

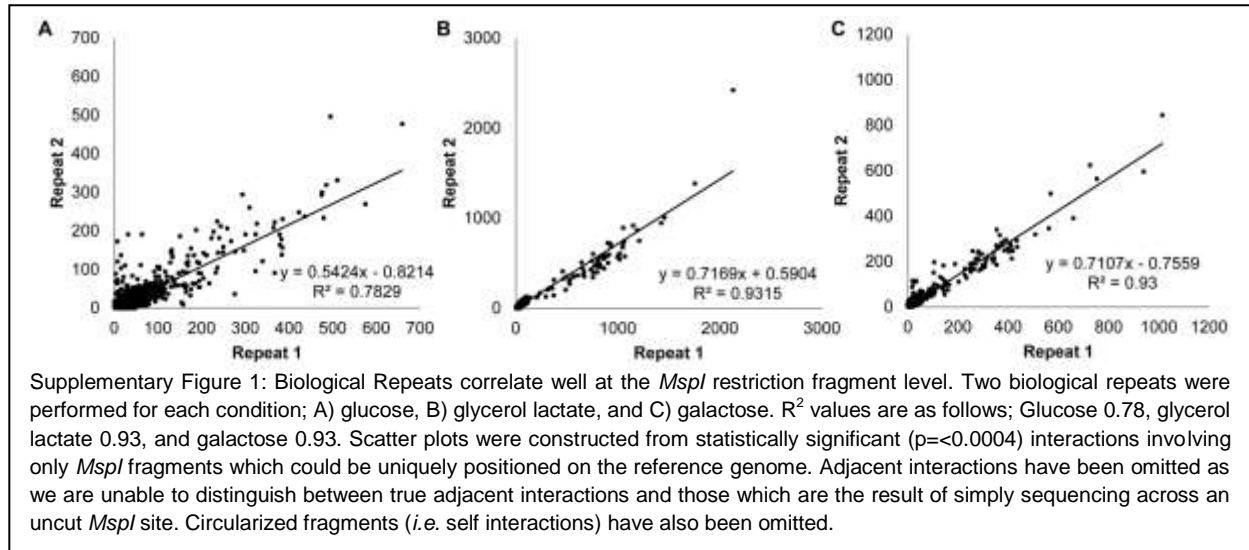
Supplementary Table 2: Standard primers were designed using Primer3. Taqman primer and probes were designed using BioSearch Technologies RealTimeDesign online software.

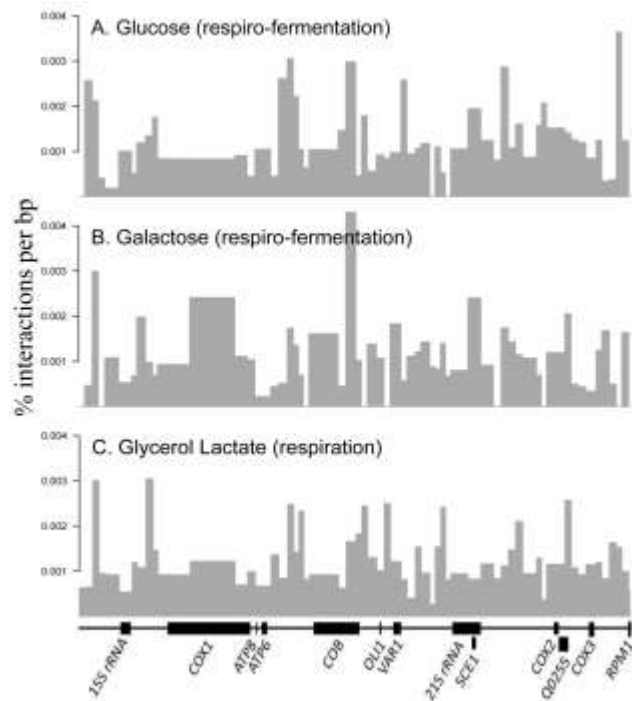
Supplementary Table 3

	% of Genome	Glucose (%)	Glycerol Lactate (%)	Galactose (%)
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 genes	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

Supplementary Table 3: Chastity filtered sequence files for each condition (glucose, glycerol lactate and galactose) were aligned against the *S. cerevisiae* genome, using the alignment algorithm SOAP (Li et al. 2008) to obtain the total number of sequences which could be aligned. 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. We then determined what percentage these features amounted to in the sequence files. Only the centromeric and mitochondrial sequences were lower than expected.

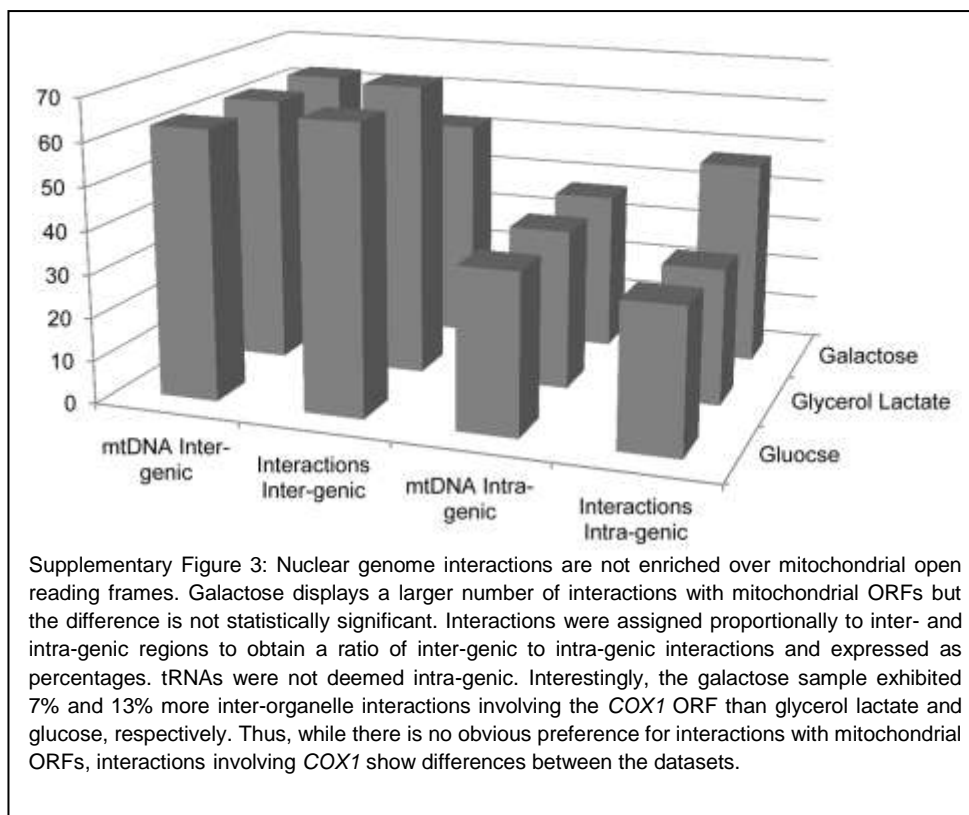
Supplementary Figures

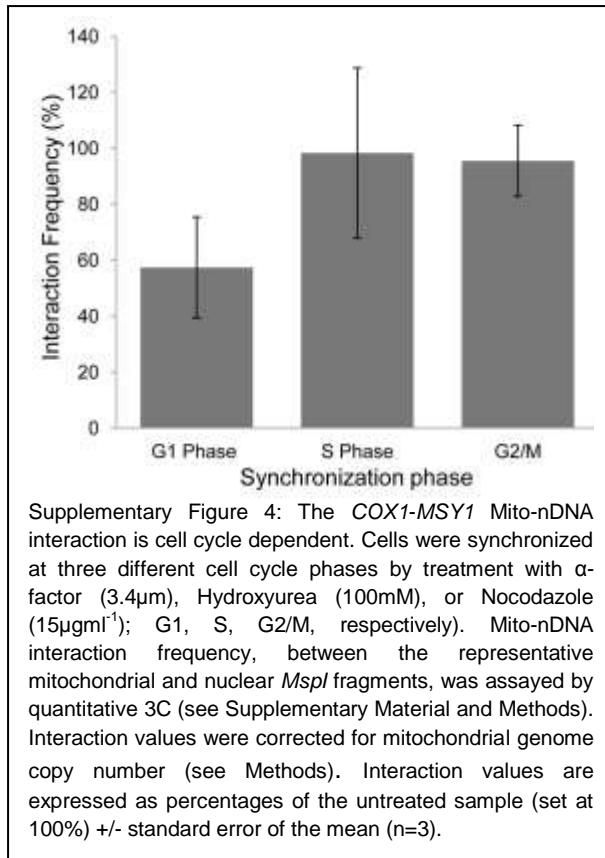


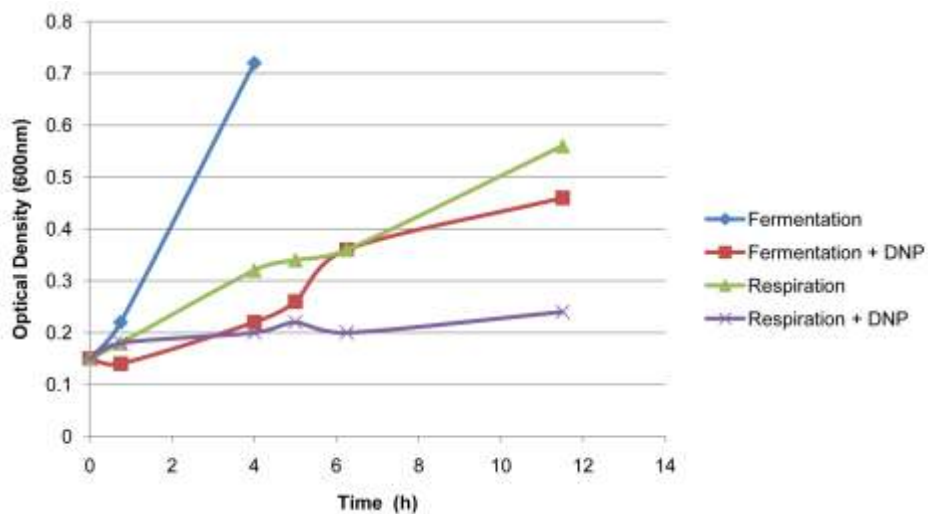


Supplementary Figure 2: Inter-organelle interactions vary with metabolic state and do not occur evenly across the mitochondrial genome. Interaction frequency was graphed as a percentage of the total number of interactions in the sample, according to segment length. To test whether Mito-nDNA interactions have a uniform distribution (*i.e.*, the total number of interactions in a segment is proportional to its length) we aggregated consecutive restriction fragments to create 58 sections that were expected to have at least 5 interactions under the null hypothesis of uniformity. A Chi-squared goodness of fit test was performed, and the distribution of the interactions was shown to deviate significantly from uniformity ($p < 0.0001$, 57 df) for all conditions. Thus, Mito-nDNA interactions are not uniformly distributed across the mitochondrial genome. The linearized mitochondrial genome is shown for comparison of the interaction frequency with mitochondrial ORF and inter-genic sequence positions. Metabolic conditions were as follows: A) respiro-fermentation (glucose), B) respiro-fermentation (galactose), and C) respiration (glycerol lactate). Only statistically significant unique interactions between the mitochondrial genome and nuclear chromosomes were included in this analysis ($p \leq 10^{-5}$; $n=2$). Interactions with the rDNA and 2-micron plasmid were removed.

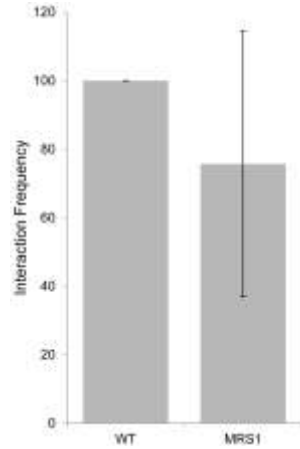
We compared the numbers of nuclear genome interactions with mitochondrial inter- and intra-genic regions to determine if the interactions across the mitochondrial genome were enriched over the open reading frames (ORFs; Supplementary Figure 3).







Supplementary Figure 5: 5mM 2,4-Dinitrophenol (DNP) inhibits respiratory growth but does not prevent growth of fermenting *S. cerevisiae* BY4741 cells. *S. cerevisiae* BY4741 cultures were grown (50 ml, 30°C, 160 rpm) on glucose (fermentation) or glycerol/lactate (respiration) to an Optical density (600nm; OD₆₀₀) of 0.600. Cultures were diluted to an OD₆₀₀ of 0.150 (50 ml final volume) in their respective media. 5mM DNP (final concentration) was added to two of the cultures, while two remained untreated. The cell growth was monitored (OD₆₀₀) for a further 11.5 hours, with the exception of the untreated glucose culture which was only grown for 4 hours.



Supplementary Figure 6: Deletion of *MRS1* (BY4741 $\Delta mrs1$), a nuclear gene involved in splicing mitochondrial type-I introns, has no significant effect on the frequency of the *COX1-MSY1* interaction in glucose grown yeast cells. Interaction frequency was expressed as percentages of the wild type *S. cerevisiae* strain BY4741 (WT, set at 100%) \pm standard error of the mean (n=3).

Supplementary methods

***Saccharomyces cerevisiae* reference genome used in the SOAP Alignment**

The reference genome consists of the 16 *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.

Quantitative 3C Analyses

Quantitative 3C was performed using Taqman PCRs to analyze the interaction frequency between two *MspI* fragments (nDNA-nDNA, interaction between loci located on Chromosomes VII (bp 868673-873686) and IX (bp 172565 – 173311); Mito-nDNA, interaction between loci located on the Mitochondrial genome (bp 24872 – 26193) and Chromosome XVI (bp 365496-365760)) which were identified by GCC analysis (Rodley et al. 2009) (for primer and fluorescent probe sequences see Supplementary Table 2).

Production of external controls for random ligation events

External controls were produced by the PCR amplification of short regions from the *E. coli* and Lambda phage genomes. Primers (Supplementary Table 2) were designed to include an *MspI* site at one end of the final product. PCR products were digested with *MspI* (37°C, 2hrs) 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 (unique region) prior to the ligation step during GCC preparation of the glucose and galactose samples. The *E. coli* fragment was introduced into the glucose samples and resulted in 10 ligation events to *MspI* restriction fragments including eight regions on the nuclear genome, one from the YSCPLASM and one from the mitochondrial genome. The Lambda phage fragment was introduced into the galactose samples and resulted in 50 ligation events with *MspI* fragments including 37 nuclear genome fragments and 13 mitochondrial fragments. No ligation events between external ligation controls and the *S. cerevisiae* genome complement occurred at levels deemed significant.

Statistical Analyses

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 elements was corrected for. Connections between unique sequences and connections involving repetitive elements were analyzed separately. Where we analyzed connections between unique loci on the nuclear genome and the mitochondrial genome we considered the mitochondrial genome as unique, as the nuclear genome is the limiting factor. Values for copy number correction were determined from: 1) published

reports of the frequencies of these elements within *S. cerevisiae* (Cui et al. 2009; Kobayashi et al. 1998; Williamson and Fennell 1979); 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. These events occur during the ligation following restriction enzyme digestion and during the sequencing preparation steps (*i.e.*, linker addition during sequencing library preparation). Thus, our null hypothesis is that all interactions we identified are a result of random ligation events, or that there are only random ligations. In an attempt to control for this, two 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. 2) We perform statistical analyses to determine whether our GCC dataset is something other than random. To achieve this we performed 100,000 simulations of random pairings (we performed these simulations previously (Rodley et al. 2009) to establish the GCC method, but have repeated them here for these new datasets). Each individual simulation contained the same total number of interactions as the GCC dataset. Glucose is used as an example below; for galactose and glycerol lactate see supplementary methods table 1.

Glucose example for establishing whether the GCC dataset differs from random:

There are 133 mitochondrial segments and 11059 nuclear segments that participate in an 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, 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} .

Supplementary Methods Table 1:

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

Next we wanted to know which of our individual interactions are above experimental noise. It is justified to assume the pairings are independent and therefore the number of times one specific pairing occurs is a

binomially distributed random variable. Let S_1 and S_2 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 divided by S_1 multiplied by S_2 , for one specific pairing to occur k or more times. L is then S_1 multiplied by S_2 , 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 has to be chosen in such a way to provide an acceptable number of false positives and consequently an acceptable false positive rate (see Supplementary Methods Table 2). $k=3$ was selected as an acceptable noise cut-off value for each of our samples (Bold).

Supplementary Methods Table 2:

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

Statistical Analysis of 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). We were expressly interested in the interactions between the mitochondrial genome and the nuclear rDNA repeats. We calculated copy numbers for these elements in our samples by aligning the sequence files against sections of these elements (Supplementary Methods Table 3). Three short regions were chosen, one from a unique nuclear element ([Gal]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). We calculated the ratio of rDNA or Mito to the unique nuclear element. Calculated copy number ratios are outlined in supplementary methods table 4.

Supplementary Methods Table 3:

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)	Reads	Aligned Reads	Percentage
Glucose	112335584	8937	0.00795563
Glycerol Lactate	98269812	12156	0.01237002
Galactose	96838770	12808	0.01322611
rDNA (96bp)	Reads	Aligned Reads	Percentage
Glucose	112335584	103881	0.09247381
Glycerol Lactate	98269812	95001	0.09667364
Galactose	96838770	85548	0.08834065

Supplementary Methods Table 4:

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

Analyses involving interactions between the mitochondrial genome and the nuclear rDNA are statistically significant and corrected for copy number. Interactions between the mitochondrial genome and nuclear rDNA repeats were isolated and subjected to binomial analysis with copy number corrections included (Supplementary Methods Table 5). The copy numbers for the mitochondria and rDNA were calculated as above, and 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 $151 \times 12.19236 = 1841$ mitochondrial fragments). The interaction value cut-off for individual pairings above noise, for each condition, is 3 and is indicated in Supplementary Methods Table 5 (bold).

Supplementary Methods Table 5:

Sample	k	N	p	P(X≥k)	L	L·P(X≥k) Expected number of false positives	False Positive Rate
Glucose	3	21358	$1/(1841 \times 6377)$	1.00E-09	11740057	0.011763502	5.06175E-06
Glycerol Lactate	3	33576	$1/(2425 \times 5574)$	2.54948E-09	13516950	0.034461207	7.49157E-06
Galactose 2010	3	11709	$1/(3135 \times 6199)$	3.6427E-11	19433865	0.000707917	4.52632E-07

Sequence Population Statistics

We wanted to know whether our sequences obtained from the Illumina genome analyser were representative of the *S. cerevisiae* genome. From published reports of copy number (Cui et al. 2009; Kobayashi et al. 1998; Williamson and Fennell 1979) we calculated how many base pairs the whole genome complement is and therefore what percentage specific genomic elements should encompass

(Supplementary Methods Table 6). The number of base pairs in the yeast genome complement was calculated by the addition of the 16 yeast chromosomes at a copy number of 1, the rDNA repeats (9.1kb) at a copy number of 200, the mitochondrial genome and 2 micron plasmid at a copy number of 50, for a total of 18,563,326 bp. The sequence files were aligned against the *S. cerevisiae* genome to determine the total number of sequences which could be aligned. We then aligned the sequences against a number of different genomic features to compare the percentage the particular feature should make up against the percentage of sequences which could be aligned against the same feature. The reference genes (housekeeping) were obtained by batch download from the SGD website. The set consisted of 646 verified ORFs across all chromosomes that were between 400-500 amino acids in size. Sequences are available upon request. Supplementary Methods Table 6 summarises the sequence population statistics.

Supplementary Methods Table 6:

	% of Genome (18,563,326) ^a	% of Glucose Sequences (101,767,324) [†]	% of Glycerol Lactate Sequences (89,904,522) [†]	% of Galactose Sequences (88,316,472) [†]
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

^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.

References

- Cui, H., Ghosh, S.K., and Jayaram, M. 2009. The selfish yeast plasmid uses the nuclear motor Kip1p but not Cin8p for its localization and equal segregation. *The Journal of Cell Biology* **185**: 251-264.
- 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. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 163-168.
- Kobayashi, T., Heck, D.J., Nomura, M., and Horiuchi, T. 1998. 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* **12**: 3821-3830.
- Li, R., Li, Y., Kristiansen, K., and Wang, J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**: 713-714.

- Moran, J., Zimmerly, S., Eskes, R., Kennell, J., Lambowitz, A., Butow, R., and Perlman, P. 1995. Mobile group II introns of yeast mitochondrial DNA are novel site- specific retroelements. *Mol. Cell. Biol.* **15**: 2828-2838.
- Rodley, C.D.M., Bertels, F., Jones, B., and O'Sullivan, J.M. 2009. Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genetics and Biology* **46**: 879-886.
- Williamson, D.H. and Fennell, D.J. 1979. Visualization of yeast mitochondrial dna with the fluorescent stain "DAPI". In *Methods in Enzymology* (ed. L.P. Sidney Fleischer), pp. 728-733. Academic Press.