THESIS

A YEAST-BASED ASSAY SYSTEM FOR THE STUDY OF ENVIRONMENTALLY INDUCED COPY NUMBER VARIATION

Submitted by

Jacquelyn Lee Stanton

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2012

Master's Committee:

Advisor: Juan Lucas Argueso

Marie Legare
Laurie Stargell

ABSTRACT

A YEAST-BASED ASSAY SYSTEM FOR THE STUDY OF ENVIRONMENTALLY INDUCED COPY NUMBER VARIATION

Multiple studies have shown that in different individuals, specific genomic segments can occur at a variable copy number relative to the reference human genome. Chromosomal rearrangements resulting in Copy Number Variations (CNVs) have long been recognized as contributing factors in carcinogenesis, and more recently in Autism Spectrum Disorders. The molecular mechanisms underlying the formation of CNVs are not completely understood. The goal of this research project was to complete the development of an assay system to study CNV, and to validate it as a tool to investigate the relationship between environmental exposures and CNV formation. We have optimized a CNV detection assay using the budding yeast *Saccharomyces cerevisiae* as an experimental model system. This CNV reporter contains two yeast genes, *SFA1* and *CUP1* that confer gene dosage-dependent tolerance to formaldehyde and copper, respectively. This system enables the detection of rare clones containing an amplification of the chromosomal segment containing the reporter by selection in media containing high levels of formaldehyde and copper, allowing the estimation of the rate of CNV formation. Results obtained in diploid cells under basal growth conditions (un-induced / un-exposed) showed that most spontaneous CNV events detected in our system were mediated through non-allelic homologous recombination (NAHR) between dispersed repetitive DNA sequences, mainly Ty1 and Ty2 retrotransposable elements and their LTRs. Another set of repeats involved in NAHR included conserved gene family. Single copy sequences and microhomology motifs were detected in our dataset, but were exceedingly rare. The most

abundant classes of CNVs observed involved segmental duplications and non-reciprocal translocations. In order to characterize the effect of environmental factors on CNV, cells were exposed to relatively low doses of three different known mutagens: Hydroxyurea, Methyl Methanesulfonate, and Camptothecin. These exposures resulted in an increase of the CNV rate ranging from 3 to 17 fold over the un-induced cultures. The spectra of chromosomal rearrangements induced by these exposures was analyzed, revealing that not only did exposures result in more chromosomal breaks but often a higher frequency of resulting segmental copies, and allowing further understanding of the CNV mechanisms associated with these exposures.

TABLE OF CONTENTS

ABSTRACT……... ii

TABLE OF CONTENTS……..…. iv

CHAPTER 1: INTRODUCTION……... 1

1.1 Copy Number Variation (CNV) definition and significance…… 1

1.2 Mechanisms of mutations……..... 3

1.3 Yeast as a model organism for genome stability……..... 4

1.4 Inheritance and potential impact of environmental exposure……..... 6

1.5 Objectives and overview……...… 6

CHAPTER 2: MATERIALS AND METHODS…… 7

2.1 Yeast strains used in this study……...…….. 7

2.2 CNV assay and environmental exposures……...….. 7

2.3 Phenotypic re-tests…….. 11

2.4 Pulse field gel electrophoresis (PFGE)……... 11

2.5 Microarray-based comparative genome hybridization (array-CGH)…….. 12

CHAPTER 3: RESULTS…….. 14

3.1 Description of the experimental system……..... 14

3.2 Amplification mutation rate determination and phenotypic retests…… 18

3.3 Spectra of associated chromosomal rearrangements…….. 20

3.4 Specific events associated chromosomal rearrangements…….. 33

3.4.1. Example of a spontaneous Chr5 amplification event……. 35

3.4.2. Example of a spontaneous Chr15 amplification event…….…...… 35

3.4.3. Example of a spontaneous Chr4 amplification event……. 37

3.4.4. Example of a MMS-induced Chr4 amplification event…….. 37

3.4.5. Example of a HU-induced Chr5 amplification event……. 41

3.4.6. Example of a CPT-induced Chr15 amplification event…….. 41

CHAPTER 4: DISCUSSION…… 44

CHAPTER 5: CONCLUSION……..… 51

REFERENCES……. 52

APPENDIX I. Array-CGH analysis…… 57

a. Conversion charts for CGH-array probe levels……. 58

b. CGH-array analysis technique…….. 59

APPENDIX II. Global views of CNVs …… 60

a. Global view of CNV across all tested uninduced Chr5 haploids…….. 61

b. Global view of CNV across all tested uninduced Chr5 diploids……..… 62

c. Global view of CNV across all tested uninduced Chr4 diploids…….…. 63

d. Global view of CNV across all tested uninduced Chr15 diploids……..... 64

e. Global view of CNV across all tested Chr5 diploids exposed to 15 µg/ml of CPT….. 65

f. Global view of CNV across all tested Chr5 diploids exposed to 50 mM of HU……... 66

g. Global view of CNV across all tested Chr5 diploids exposed to 35 µg/ml of MMS... 67

h. Global view of CNV across all tested induced Chr5 diploids…….……. 68

i. Global view of CNV across all tested Chr4 diploids exposed to 15 µg/ml of CPT….. 69

j. Global view of CNV across all tested Chr4 diploids exposed to 50 mM of HU…….. 70

k. Global view of CNV across all tested Chr4 diploids exposed to 35 µg/ml of MMS... 71

l. Global view of CNV across all tested induced Chr4 diploids……... 72

m. Global view of CNV across all tested Chr15 diploids exposed to 15 µg/ml of CPT.. 73

n. Global view of CNV across all tested Chr15 diploids exposed to 50 mM of HU…… 74

o. Global view of CNV across all tested Chr15 diploids exposed to 35 μg/ml of MMS. 75

p. Global view of CNV across all tested induced Chr15 diploids…… 76

APPENDIX III. Initial development of CNV assay……..… 77

a. Assay development……...…… 78

b. Assay development data……... 79

CHAPTER 1

INTRODUCTION

1.1 Copy Number Variation (CNV) definition and significance

Copy Number Variation (CNV) is an alteration of the genome that results in an altered number of copies of one or more sections of DNA. In humans CNVs have been considered to be changes that vary between 50 bp and a few Mb (Girirajan et al., 2011) resulting from segmental deletions or amplifications, insertions or translocations (Feuk et al., 2006).

Beginning in 2004, array-based comparative genomic hybridization (array-CGH)) was used to study variation in the human genome (Iafrate et al., 2004; Sebat et al., 2004). More recently, whole genome and exome sequencing have been extensively used to identify human CNVs at base pair resolution (Meyerson et al., 2010). It was discovered that copy number variants are widespread, and as of April 2012, 66,741 CNVs in the human genome are cataloged in the structural variation database (the Database of Genomic Variants or DGV, http://projects.tcag.ca/variation/), and are the cause of a great deal of genetic diversity, more so than even individual nucleotide changes (single-nucleotide polymorphisms, SNPs) (Sebat et al., 2007). The rate of *de novo* CNV mutations in humans is two to four orders of magnitude higher than the rate of SNP mutations (Lupski, 2007).

A comprehensive study of CNVs involved with autism and asthma found that the mean CNV sizes to be 693Kb and 2Mb, and that on average, 8-25 Kb of *de novo* CNV sequence was observed per genome per transmission (Itsara et al., 2010). While 35-90 substitutions per generation was previously estimated to be the range of *de novo* point mutation rates (Kondrashov, 2003; Nachman and Crowell, 2000; Roach et al., 2010), the autism/asthma study revealed that the overall frequency of *de novo* CNVs is less than point mutations, but that the

number of affected base pairs per generation is about 100-fold higher (Itsara et al., 2010). The rate of somatic CNV is also very high and enough to produce detectable variations even between different tissues in a single healthy individual (Piotrowski et al., 2008) and between identical twins (Bruder et al., 2008). Likewise, somatic CNV mutations in cancer are a major source of intra-tumor genetic mosaicism that is just beginning to come into light (Macosko and McCarroll, 2012).

The link between copy number mutations and Autism Spectrum Disorders (ASDs) is the most notable phenotypic consequence established so far (Abrahams and Geschwind, 2008). This is a category of disorders of which the Center for Disease Control (CDC) estimates 1 in 88 children are identified to possess. This marks a 78% increase of incidence from 2007, and 28% since 2009, leading the CDC to consider it to be "an important public health concern" (http://www.cdc.gov/Features/CountingAutism). While some of the increase in ASDs is due to the increased awareness and improved methods of diagnosis, it is believed that both genetic and non-genetic factors including environmental exposures have also contributed to this increase.

CNVs have also been identified as important contributors in the risk of disease (Sebat et al., 2007). The role of CNVs in carcinogenesis (Abrahams and Geschwind, 2008) and in other diseases, such as Charcot-Marie-Tooth syndrome and Williams syndrome, has long been recognized (Ewart et al., 1993; Ledbetter et al., 1981; Lupski et al., 1991; Sebat et al., 2007).

Though a clear genetic link has been established, the mechanistic connection between CNV and autism is still a wide open question. An active field of investigation is studying which specific genes, when altered in copy number, cause neuro-developmental disorders. A region containing 29 genes on chromosome 16 (16p11.2) has been frequently observed in psychiatric disorders with copy number variants. While the mechanisms and specific genes of the 29

involved are not fully understood, individuals that have loss or gain of 16p11.2 have significant risk of autism and developmental delay (Malhotra and Sebat, 2012b). A study using zebrafish as a brain development model system has provided evidence suggesting that the causal link to autism is the change in the number of copies of three specific genes located within 16p11.2: KCTD13, MAPK3, and MVP (Golzio et al., 2012).

1.2 Mechanisms of mutation

Information from rearrangement breakpoint junctions commonly provides information about the mechanisms underlying the structural variations. Nonallelic homologous recombination (NAHR), nonhomologous end-joining (NHEJ), fork stalling and template switching, and restrotransposition have been identified as the major mechanisms responsible for structural variations in humans (Zhang et al., 2009). The misalignment of homologous (but non-allelic) regions and ensuing crossing over between the sites is what is known as NAHR (Malhotra and Sebat, 2012a). This event occurs most frequently during meiosis (Turner et al., 2008), and to a lesser extent in mitosis (Lam and Jeffreys, 2006, 2007). Regions that possess dispersed highly repetitive DNA sequences such as transposable element insertions, and other low copy repeats (LCRs) are more prone to frequent rearrangements due to NAHR (Malhotra and Sebat, 2012a). In contrast, homology at breakpoints is not essential for NHEJ to occur (Toffolatti et al., 2002).

Fork stalling and template switching, described as microhomology-mediated break-induced replication (FoSTeS/MMBIR), is a rearrangement mechanism that results from errors during the process of DNA replication (Lee et al., 2007; Malhotra and Sebat, 2012a). Rearrangements induced by this mechanism have been observed to vary greatly in size and complexity (Hastings et al., 2009; Zhang et al., 2009).

Results from the 1000 Genomes Project have shown that 70.8% of deletions resulted from NHEJ or FoSTeS/MMBIR, whereas large duplications or deletions were generated by NAHR (Malhotra and Sebat, 2012a; Mills et al., 2011).

1.3 Yeast as a model organism for genome stability

The budding yeast *Saccharomyces cerevisiae* was used in this study of CNVs. S. *cerevisiae* has 16 chromosomes, ranging from 0.2 to 2.5 MB in size. In this model, a CNV event is a deletion or duplication of a segment as small as a single gene (~2Kb) but not greater than one chromosome arm (~500Kb). The yeast genome is 250 fold smaller than the human genome, allowing base-pair resolution genomic analysis at a lower cost. It serves as a powerful model to study fundamental mechanisms of eukaryotic cell and molecular biology, including the formation of CNV events. Though the yeast genome is much simpler than the human genome, it contains many of the same basic properties including: dispersed repetitive DNA sequences, conserved replication and repair pathways, and analogous organization of the chromosomes.

Studies of duplications or deletions before the year 2000 were not genome-wide, but rather focused on single loci and thus the results were gene/strain-specific (Fink and Styles, 1974; Sherman et al., 1974). Haploid strains were used in most of these studies. Hotspots for deletions were identified in two locations in the genome (Liebman et al., 1979). It was determined that these locations were flanked by two 5.9kb Ty retrotransposons in direct orientations, and that non-allelic homologous recombination between the Ty elements was responsible for the deletions that had been previously observed (Liebman et al., 1981). High frequency of deletions was also found to occur due to a similar mechanism of recombination between flanking delta elements (long terminal repeat sequences (LTRs) associated with Ty elements) (Rothstein et al., 1987). Directly oriented retrotransposons have been identified as

being involved in deletion events in several other studies (Argueso et al., 2008; Hoang et al., 2010; McCulley and Petes, 2010).

Studies performed by Chan and Kolodner showed that the presence of a dispersed repeat greatly stimulated the rate of NAHR-mediated deletions, and that in regions without repeats infrequent deletions were often due to NHEJ. This was shown through an assay for gross chromosomal rearrangements (GCR) entailing the simultaneous loss of two genes *CAN1* and *URA3* that were located on the left end of chromosome 5 (Chr5) (Chen and Kolodner, 1999). In more recent versions of the same assay, the insertion of a Ty1 element proximal to these genes showed an increase in the deletion events of about 400-fold (Chan and Kolodner, 2011, 2012).

Duplication events in yeast have not been as frequently studied as deletions, but involve what seem to be more varied mechanisms. For example, duplication events can involve an increase in copy number from a pre-existing duplication or from a duplication of a region originally present in only one copy. Dependent on chromosome context and ploidy, sequences originally present in one copy being duplicated seem to occur at a much lower frequency. A number of reporter systems have been used to detect duplications: medium containing antimycin A can select for duplication of the *ADH2* or *ADH4* (Dorsey et al., 1992), and *RPL20B* duplications in strains with *RPL20A* deleted can be selected as fast-growing descendants of the slow-growing parent strain (Koszul et al., 2004; Payen et al., 2008). Narayanan et al. (2006) selected for amplification of a cassette containing *CUP1* and *SFA1* by use of medium that contained high levels of copper (Cu) and formaldehyde (FA). *SFA1* being formaldehyde dehydrogenase which acts to detoxify formaldehyde, and *CUP1* or metallothionein which has the ability to sequester copper and thus mediate resistance to high concentrations of

copper/cadmium. The Narayanan et al. *CUP1/SFA1* reporter system was the starting point for the development of the assay described in this study.

1.4 Inheritance and potential environmental impact

DNA damage such as double strand breaks are initiators that trigger the mechanisms that can lead to chromosomal rearrangements involved in CNV formation. This damage can be caused by intrinsic cellular factors, or by environmental exposures such as ionizing radiation. The growing rate of sporadic disease in which the genetic mutation responsible for the disease is found in a patient but is absent in the biological parents implies the occurrence of *de novo* CNVs both somatically and in the germline. In addition, the development of somatic cancers are clearly associated to environmental factors (Boffetta, 2006), that stimulation *de novo* CNV. Alarming data from environmental quality studies have shown the presence of concerning concentrations of pesticides, industrial chemicals, pharmaceuticals, and other chemicals in ground water (Richardson, 2008). It is plausible that these environmental pollutants could be partly responsible for an increase in *de* novo and somatic CNV occurrence in the population exposed.

1.5 Objectives and overview

The aim of the research presented in this thesis was to optimize the yeast CNV assay and use it to investigate environmental factors and mechanisms involved in the formation of *de novo* CNV events. *Saccharomyces cerevisiae* cells were exposed to known mutagens not only to study the mechanisms involved, but also confirm the effectiveness of the assay as a tool for screening unknown candidate mutagens. Informative results from these experiments validated this approach of study and will allow for future study of the role of emerging contaminants.

CHAPTER 2

MATERIALS AND METHODS

2.1 Yeast strains used in this study:

All strains used in the CNV assay were isogenic with strain MS71, except for noted locus-specific changes introduced by transformation (Lemoine et al., 2005). MS71 is essentially isogenic to the CG379 strain background (Argueso et al., 2008; Morrison et al., 1991). The specific strains used in this study are listed in Table 1.

The CNV reporter cassette (schematic representation in Figure 1) was integrated in to the yeast genome by PCR-mediated homologous recombination, and transformants were selected in plates containing geneticin. The *SFA1* gene in the reporter cassette contains a dominant mutation that encodes a Valine 208 to Isoleucine substitution (*SFA1-V208I*) predicted to enhance the activity of formaldehyde dehydrogenase enabling higher tolerance to formaldehyde. The reporter cassette was integrated at three different positions (one at a time) in the haploid yeast genome: on Chr5, between the *DDI1* and *UBP5* genes, at SGD coordinate 457,705 (*DDI1::SFA1-V208I-CUP1-Kan*MX4); on Chr15 about 5 Kb distal to the *RPL20B* gene, between the *SFG1* and *COT1* genes, between SGD coordinates 905,937 and 906,020 (*SFG1::SFA1-V208I-CUP1-Kan*MX4); on Chr4 between the *PML2* and *SAM2* genes, at SGD coordinate 1,453,103. Each haploid carrying a specific reporter insertion above was mated to a haploid of the opposite mating type carrying a *TRP1* insertion at the exact same respective positions, therefore forming the diploid strains used in the study.

2.2 CNV assay and environmental exposures

Individual colonies from the CNV reporter parent diploid strains were isolated on YPD plates (rich media; 20 g glucose, 20 g peptone, 10 g yeast extract, and 10 g agar in 1 L of

Table 1. Diploid strains studied.

Strain	Description	Genotype
JAY685	Diploid formed by crossing JAY681 and JAY657. Reporter on Chr15.	
JAY681	MATa	*ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1Δ RSC30 sfa1Δ::hisG RPL20B::SFA1-V208I-CUP1-Kan 3'-5'*
JAY657	MATalpha	*ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1Δ RSC30 sfa1Δ::hisG RPL20B::TRP1*
JAY654	Diploid formed by crossing JAY648 and JAY644. Reporter on Chr4.	
JAY648	MATalpha	*ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1Δ RSC30 sfa1Δ::hisG PLM2::SFA1-V208I-CUP1-Kan*
JAY644	MATa	*ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1Δ RSC30 sfa1Δ::hisG PLM2::TRP1*
JAY538	Diploid formed by crossing JAY484 and JAY520. Reporter on Chr5	
JAY520	MATalpha	*ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1Δ RSC30 sfa1Δ::hisG DDI1::TRP1*
JAY484	MATa	*ade5-1 his7-2 leu2-3,112 LEU+ ura3-52 trp1-289 cup1Δ RSC30 sfa1Δ::hisG DDI1::SFA1-V208I-CUP1-Kan*

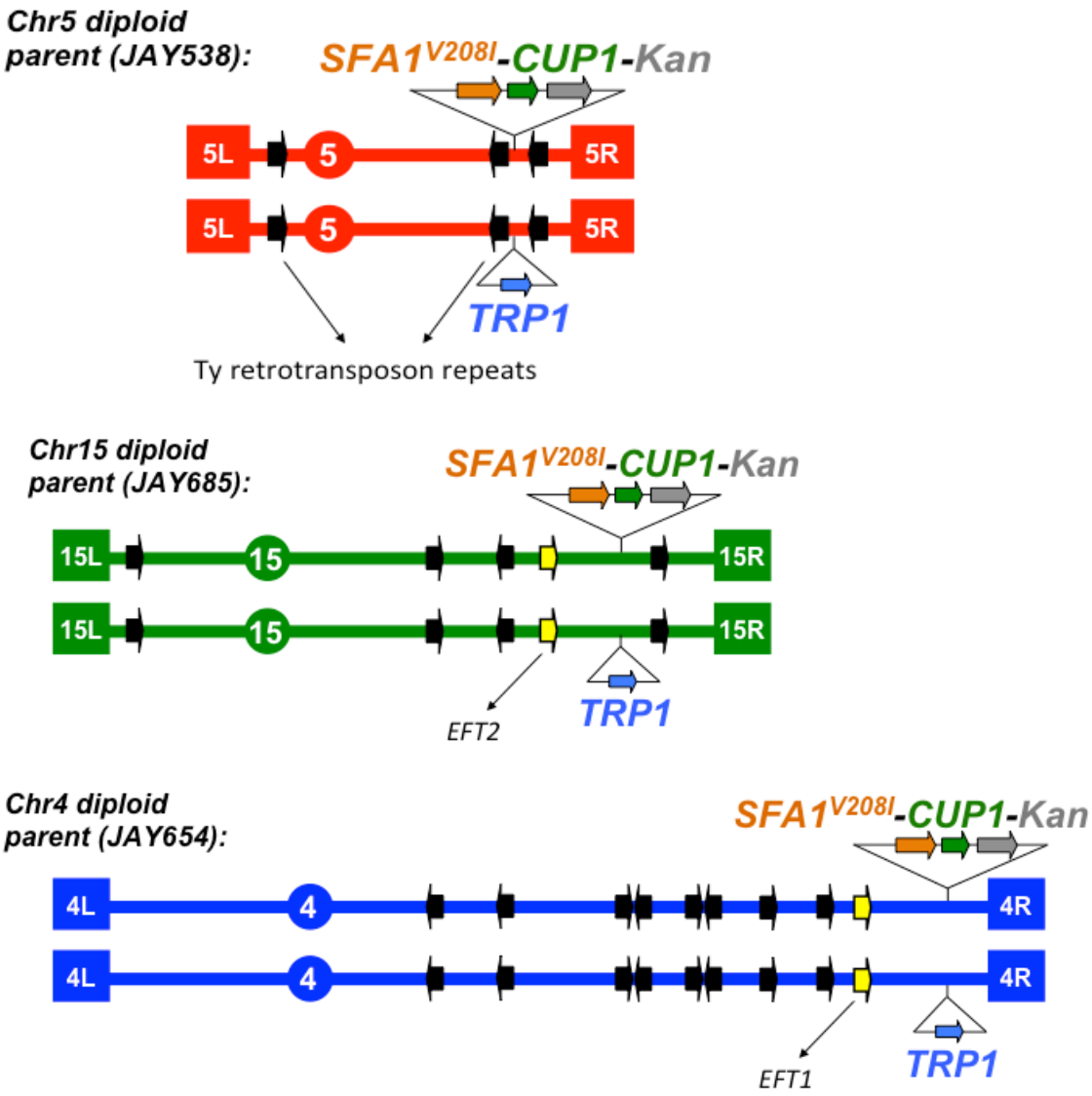


Figure 1. Structure of the CNV reporter and insertion sites on Chr5, Chr15, and Chr4 diploid strains. Circles indicate centromeres and boxes containing "L" and "R" indicate left and right telomeres, respectively. The reporter cassette containing *SFA1, CUP1,* and *KanMX4* genes was inserted on the left arms of each chromosome. The endogenous copies of *CUP1* and *SFA1* were deleted from the strains. Black arrows represent Ty retrotransposons elements. *TRP1* gene was inserted at the allelic position on the homologous chromosome. Yellow arrows represent the *EFT1* and *EFT2* genes that are two copies of the same gene represented twice in the genome.

distilled water), and then incubated in 5 mL liquid YPD for the last 24 hours of growth. In the case of environmental exposures, the specific mutagens were added during the liquid growth phase at the following concentrations: Camptothecin (CPT) 15 μg/ml, Hydroxyurea (HU) 50 mM, and Methyl Methanesulfonate (MMS) 35 μg/ml. 1 mL of each culture was centrifuged down to a pellet and the supernatant was discarded. The pellet was then washed twice using sterile distilled water. The washed pellets were re-suspended in 1mL of water, and the samples were serially diluted. 100 ml of the 10^{-4} dilution was plated on one permissive SC Trp drop-out plate; and four 100 ml aliquots of the 10^{-1} dilution were plated on four different non-permissive SC Trp dropout plates supplemented with 1.8 mM formaldehyde (FA) and 150 μM $CuSO_4$. The SC Trp drop-out plates were prepared with 1.7g YNB w/o a.a., w/o A.S, 1.4g Trp drop-out mix**,** 5 g ammonium sulfate, 20 g glucose, 20 g bacteriological agar, and 1L distilled H_20 were mixed and autoclaved. Cu and FA were added after the media cooled down to 70°C. FA is not stable in water solution, therefore fresh 1 M dilutions from a methanol-stabilized stock were prepared for each new batch of non-permissive media. These plates were poured and used within 48hrs.

Cells were grown for five days and colonies were counted on days three, four and five of growth. The documentation of colonies from each day of growth was compared with the re-tests (section 2.2) to see if the length of time for the original colonies to grow was related to the classes of CNV events detected. The total number of colonies that had formed on the four non-permissive plates after five days were added and compared to the colonies formed on the permissive plate in order to calculate the mutation rate using the Lea and Coulson method of the median (Lea and Coulson, 1949). To ensure the independence of the CNV events analyzed, only one colony from each culture was selected from a non-permissive plate to be re-tested and later used in pulse field gel electrophoresis (PFGE) and array-CGH.

2.3 Phenotypic re-tests

From each culture a colony was selected, streaked on low concentration 0.8 μM CuSO4 and 100 mM FA plates to grow isolated colonies, and was re-tested in order to verify the occurrence of the copy number amplification of the cassette. This was done by plating dilutions of the medium containing the experimental culture, which was inoculated 24-48 hours, on four different plates containing different combinations of copper and formaldehyde. One plate did not contain any copper or formaldehyde, this served as a positive viability reference for the cells plated. The other three plates contained 1.8 mM FA and 150 μM CuSO4 (same as the selective plates used in the CNV assay), 0 mM FA and 300 μM CuSO4, and 2.0 mM FA and 0 μM CuSO4. The plates only containing either Cu or FA provided information about the presence of the two individual genes that were inserted in the reporter cassette. On each of these plates, the growth of the candidate CNV clones was compared to the viability of the parent containing one copy of the cassette.

2.4 Pulse field gel electrophoresis (PFGE)

Agarose-embedded full length chromosomal DNA was prepared for each sample to analyze using PFGE. These samples (plugs) were prepared in sufficient quantity to allow a single source of full length chromosomal DNA to be analyzed repetitively if needed, or in different assays without the need for secondary DNA preparations that might have caused variations in clones carrying unstable chromosomal rearrangements.

To make these plugs, the cultures were grown in 7 mL of YPD for 24-48 hours. The samples were then centrifuged, the supernatant discarded and the pellets were weighed. The appropriate amount of zymolyase and low-melting point agarose was mixed with the cells and transferred to BioRad plug molds. Plugs were incubated in the cold room for about 20-30

minutes, and once they were solidified the plugs were put in a solution made of 500 mM EDTA, 10 mM Tris, pH 7.5, and incubated overnight at 37C. The next day 400 µL of a solution containing 5% sarcosyl, 5 mg/ml proteinase K in 500 mM EDTA pH 7.5, was added and incubated at 50C for five hours. Plugs were then washed 3 to 5 times in 1x TE solution.

Plugs were trimmed and loaded on a 180 ml gel prepared with 1% BioRad Pulse Field Certified agarose in 0.5% TBE buffer. The wide/long BioRad CHEF gel mold was used to cast the gel. The separation program was run at 5 volts, temperature 14C, initial switch time was 47 seconds, final switch time was 170 seconds, and the run time ranged from 50-58 hours. The technique differs from a standard gel electrophoresis in that the voltage switches diagonally resulting in a net forward movement of the DNA. The various sizes of DNA chromosomes react to the change of voltage at different rates, resulting in full-length chromosome separation.

2.5 Microarray-based comparative genome hybridization (array-CGH)

Array-CGH experiments were also used to identify the CNV events. First the DNA was extracted from the agarose-embedded full length chromosomal DNA plugs. For each sample, four plugs were dried on a paper wipe and put in a tube. 250 µL of binding buffer from the GeneJET Extraction Kit #K0702 was added and incubated at 37C until the plugs dissolved. These samples were then sonicated, using the Bioruptor UCD-200 by Diagenode, for 30min. The procedure as stated in the GeneJET Gel extraction kit was followed and 15 µL of elution buffer was used. The dsDNA broad range Quant-iT assay kit and Invitrogen Qubit fluorometer was used to determine the concentration of the DNA that was extracted and a gel was run to confirm the desired size range of the fragmented DNA (0.5 to 2.0 Kb).

An aliquot of 1.5 µg of genomic DNA form each sample was diluted with deionized H_2O to a total volume of 7 µL. 6.7 µL of 2.5x random primers from the Invitrogen BioPrime Array

CGH Genomic Labeling module kit were added to each sample. These samples were then vortex and incubated for five minutes at 95C to denature the genomic DNA. During which time the Cy5 reaction labeling mix and Cy3 labeling mix were made, using dUTP mix and Klenow Pol from the BioPrime Array CGH Genomic Labeling module kit and Cy5-dUTP and Cy3-dUTP. The entire Cy3 mix was then added to the parent strain sample, and the Cy5 mix was evenly distributed among all of the experimental CNV clone samples. The samples were then incubated at 37C for 3 hours. Then the Cy3 mixture was evenly divided into each of the Cy5 reactions and the GeneJET PCR Purification Kit #K0702 was followed to clean up the reaction, using 22µL of elution buffer to recover the purified labeled DNA. A hybridization mixture containing salmon sperm DNA, Agilent 10x blocking agent, and Agilent 2x Hi-RPM buffer was made and 35µL was added to 20 µL of each labeled DNA sample. The samples were then incubated for 3 minutes at 95C and 20 minutes at 37C. The Cy5-labeled experimental DNA and the Cy-3 labeled parental DNA samples were then loaded on Agilent Technologies oligonucleotide microarray slides and co-hybridized at 65C in a rotating oven overnight (at least 14 hours). The arrays contained roughly 15,000 60-mer probes distributed evenly across the entire yeast genome, with a median spacing of about 700 bp.

The next day, the slides were washed for five minutes using Agilent wash buffers 1 and 2 as recommended. The slides were then scanned using the Perkin Elmer Scan Array Express Microarray Scanner. Gene Pix Pro 6.1 was used to grid the arrays and extract the relative Cy5/Cy3 hybridization signals from the images. Nexus Copy Number software was then used to analyze the GenePix data, allowing vitalization of the results in the form of the graphical plots shown in the figures in the Results and Appendices.

CHAPTER 3

RESULTS

3.1 Description of the experimental System

The primary goals of this research project were to complete the development of a yeast assay for the detection of spontaneous chromosomal rearrangements associated with gene amplification and/or deletion (Copy Number Variation; CNV), and to demonstrate its use as a system for the analysis of the role of specific environmental exposures on induction of CNV events. The initial development of this assay was conducted previously and will be presented elsewhere (Zhang et al., 2012a).

The CNV assay system is based on the phenotypic selection of yeast cells containing local genomic increases in gene dosage (Figure 1). Diploid strains were built and studied that had chromosomal integrations of a CNV reporter cassette containing the *SFA1* and *CUP1* genes (Figure 2) that confer gene dosage-dependent tolerance to formaldehyde (FA) and copper (Cu), respectively. These parent strains containing a single copy of the CNV reporter cassette were only able to grow in media that contained low levels of FA and Cu. Selective conditions for the assay were optimized such that at the specific inhibitor concentrations combination of 1.8 mM FA and 150 μM CuSO4 only cells that had undergone genomic amplification events resulting in the presence of two or more copies of the CNV reporter cassette were able to grow.

The cassette also included the *KanMX4* marker (Wach et al., 1994), which confers resistance to geneticin and was used to select for initial integration of the cassette into the genome. Also shown is the auxiliary marker gene *TRP1* inserted at the allelic position on the homologous chromosome. This auxiliary marker further enhanced the sensitivity of the CNV assay because it allowed us to focus our study on the relatively rare non-allelic recombination

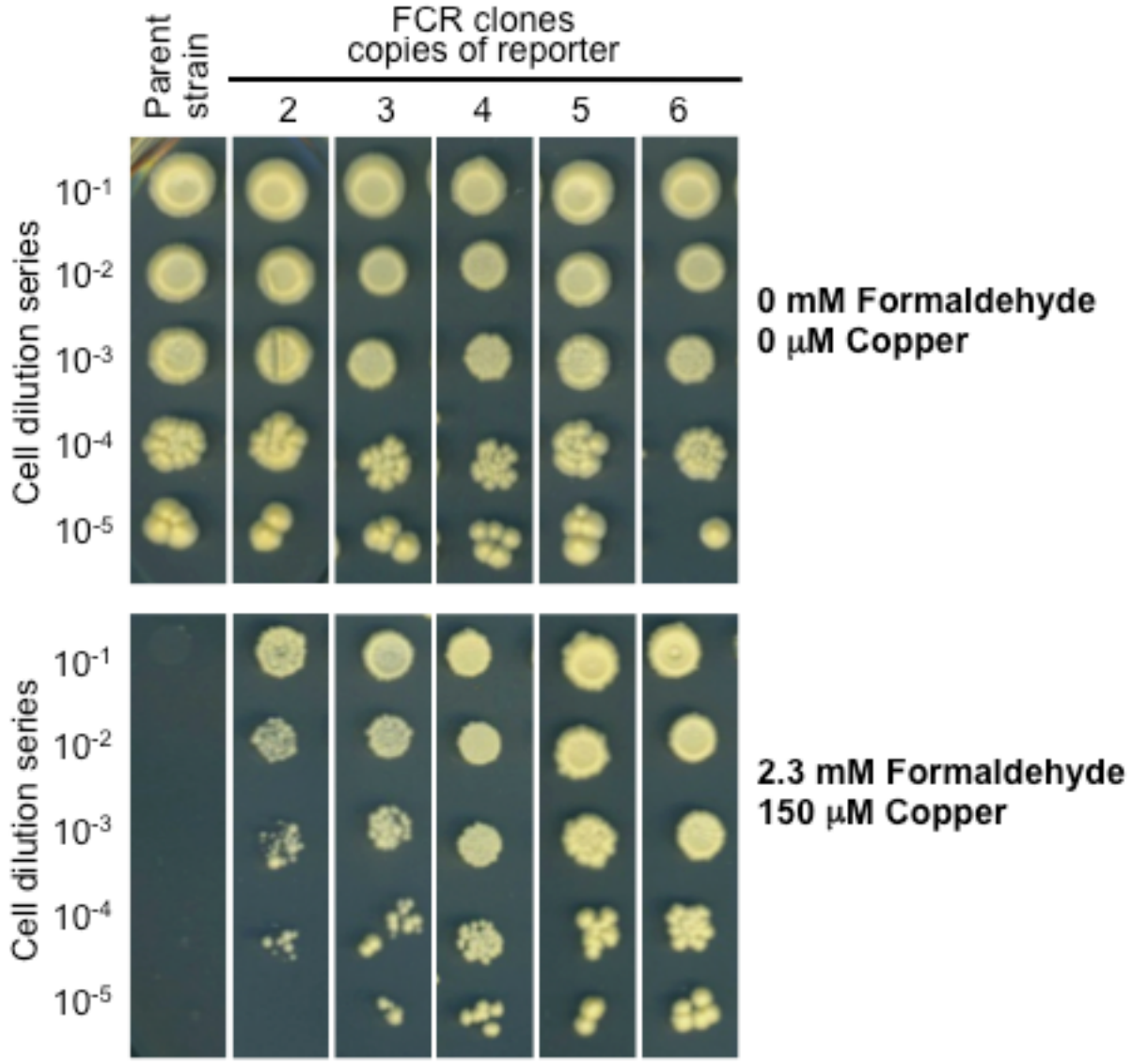


Figure 2. Gene-dosage dependent phenotype. The top image shows the growth of the parent strain and of Formaldehyde and Copper Resistant (FCR) clones containing increasing copies of the CNV reporter cassette on medium containing no copper or formaldehyde. When these same strains are plated on medium containing FA and Cu as shown in the bottom image, the FCRs are able to grow and a higher number of copies of the reporter correlates with better growth. The parent strain, which contains just one copy of the reporter, is unable to grow on the plates containing copper and formaldehyde.

events that are responsible for pathogenic CNV formation in humans. *TRP1* is essential for growth in media lacking tryptophan. Plating cells in Trp drop-out media eliminated the growth of clones that could have acquired a second copy of the CNV reporter by allelic mitotic recombination (Loss-of-Heterozygosity; LOH). This relatively frequent class of allelic double strand break repair mechanism was not the subject of this research project since it does not generate chromosomal rearrangements and is typically benign in phenotypic outcome.

A major advantage of the CNV assay system is that it can be used in diploid cells. Most other assays used previously to model genome stability in yeast were strictly limited to haploid cells. Because our assay uses diploid parent strains, it is not limited in terms of CNV formation mechanisms, meaning that all of the pathways discussed in the Introduction section can be successfully recovered, particularly those events involving large genomic deletions which are lethal in haploids. This allowed for an unbiased sampling of naturally occurring CNV events, and provided a much more relevant model for the pathogenic genomic changes that occur in the human genome.

Three parent diploid strains were built that had the CNV reporter cassette inserted at different genomic sites: Chr15 (JAY685), Chr4 (JAY654), and Chr5 (JAY 538). Schematic representations of the chromosomal insertions in these strains are shown in Figure 1. The genomic context of the reporter insertion differed significantly between the three parent strains. The reporter was inserted between two flanking direct Ty1 repeats on Chr5, which are known to facilitate segmental duplications (Argueso et al., 2008). In addition, Chr5 is a relatively small chromosome in the yeast genome (577 Kb). Chr15 is a larger chromosome (1,091 Kb) and the reporter cassette was inserted farther away from flanking repetitive elements. The specific site of insertion on Chr15 (immediately distal to the *RPL20B* gene) was chosen to allow comparison

of our results to those previously reported using selection for duplication of the *RPL20B* gene (Payen et al., 2008). Chr4 is the second largest chromosome in the genome (1532 Kb) and the cassette was inserted distal to the last annotated dispersed repeat, therefore it was not flanked by any type of homology. The parallel analysis of these different combinations of host chromosomes sizes and reporter cassette locations gave a broad representation of the CNV mechanisms that are possible throughout the yeast genome.

The three diploid parent strains were used to investigate two key parameters of CNV formation: (1) we measured the rate at which gene amplification events occur by using quantitative fluctuation mutation assays; and (2) we characterized the qualitative nature of the genome rearrangements present in the selected CNV clones by molecular karyotyping using a combination of Pulse Field Gel Electrophoresis (PFGE) and microarray based-Comparative Genome Hybridization (array-CGH).

These parameters were first determined in yeast cultures grown under normal (un-induced) conditions. We then set out to validate the assay as a tool for the investigation of the role of environmental exposures on CNV formation by treating the cultures with three known mutagens, all of which had been previously shown to induce the formation of chromosomal rearrangements in other yeast and human cell culture assays (Arlt et al., 2011; Myung and Kolodner, 2002; Payen et al., 2008). These proof-of-concept CNV inducing agents were: Methyl-Methane Sulfonate (MMS), an alkylating agent and known carcinogen; Camptothecin (CPT), a topoisomerase I inhibitor and obsolete chemotherapeutic; and Hydroxyurea (HU), an inhibitor of ribonucleotide reductase (and thus inducer of DNA replication stress), currently used in the treatment of sickle cell disease.

3.2 Amplification mutation rate determination and phenotypic retests

The amplification mutation rates were determined in classic fluctuation assays (Lea and Coulson, 1949) by plating dilutions of several independent yeast cultures on permissive and on non-permissive FA/Cu media, then counting the resulting yeast colonies to estimate the number of mutants per culture. These experiments were carried out for un-induced cultures as well as for cultures exposed to 50 mM of HU, 15 μg/ml of CPT, or 35 μg/ml of MMS during the last 24 hours of their growth. These specific doses were chosen based on preliminary tests that indicated they were sufficient to significantly induce CNV mutations without causing an excessive reduction in cell viability in the cultures. As stated above, the primary intent of this study was to demonstrate the detection of exposure-dependent CNV mutations and to characterize the molecular nature of the resulting genome rearrangements. Since the single dose data points chosen allowed us to successfully reach both goals, the determination of dose response curves over a wide range of doses for each drug was not pursued.

The results of the mutation rate assays, including median rate and 95% confidence intervals, for all conditions in all parent strains are summarized in Table 2. In general, the un-induced CNV mutation rates were quite high (~10^{-6}) when compared to typical base nucleotide mutation rates previously measured in yeast (~10^{-8}). The specific absolute values varied significantly depending on the genomic context of the CNV reporter. The rate measured for Chr15 was quite low when compared to the mutation rates measured with the Chr5 and Chr4 reporters. All cultures exposed to MMS, CPT and HU showed a significant elevation in the CNV mutation rate, ranging from a three to seventeen fold stimulation depending on the drug and the reporter strain combination. The specific CNV inductions provided by each drug were relatively similar for the Chr4 and Chr15 reporter insertions, whereas a lower level of induction was

Table 2. Quantitative measurements of reporter amplification.

Parent Strain	Haploid	Diploids											
	Chr5	Chr5				Chr4				Chr15			
	Unind.	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS
Relative rate	3.1	1.0	2.88	3.39	4.64	1.0	4.98	6.66	12.58	1.0	5.62	5.56	16.85
Absolute rate ($_{x10-6}$)	11	2.5	7.1	8.3	11.4	1.3	6.6	8.8	16.7	0.16	0.91	0.90	2.7
95% confidence ($_{x10-6}$)	.04-.19	1.6-3.0	6.4-11.1	3.9-14.7	7.3-14.4	1.1-1.8	4.5-11.5	5.0-16.4	10.8-27.1	0-0.3	0.6-1.2	0.5-1.4	1.8-3.7
Number of cultures	30	43	15	14	15	42	14	15	15	40	29	29	29

Presented are the median rates of amplification per cell per cell division and the relative rate compared to the respective uninduced cultures, and the 95% confidence intervals for the rates. The total numbers of cultures for each strain containing are also shown.

measured for the Chr5 reporter.

We also monitored the number of viable cells per culture at the time of plating, in un-exposed and mutagen treated cells (Table 3). All exposed cultures had a lower cell titer at the end of the 24 h exposure, ranging from 35% to 75% of the average titer in un-exposed cultures. It is unclear, however, whether the lower titer reflected cell killing, or simply a slower rate of cell division in the presence of the drugs. In either case, the exposures were relatively mild in terms of cell toxicity, but were still capable of inducing a significant increase in CNVs as intended.

Another important parameter that we examined was the frequency of positive phenotypic re-tests for each mutation rate assay. Typical nucleotide mutation assays select for very discrete phenotypes that are straightforward to interpret such as reversion of a mutant allele associated with a nutritional deficiency (*e.g.* Lys^- to Lys^+, growth on Lys drop-out plates), or complete loss of function of a specific gene resulting in drug resistance (*e.g.* Can^S to Can^R, growth on plates containing the drug canavanine). In contrast, the CNV assay is based on scoring the gradual phenotypic enhancement associated with extra copies of the reporter cassette. Previous attempts at amplification detection assays had been plagued by a high percentage of false-positive colonies growing on non-permissive plates due to processes other than true amplification, especially small variations in expression of the reporter genes (Dorsey et al., 1992; Moore et al., 2000). We therefore investigated whether the selection applied in our assay was also susceptible to a similar pitfall. We took at least one individual colony selected on FA/Cu plates from each

Table 3. Final viable cell titer in liquid cultures prior to plating.

	Uninduced	CPT	HU	MMS
Average final titer (x10+7)	5.08	3.80	2.16	1.79
SD final titer (x10+7)	2.07	1.53	1.03	8.98
Relative final titer	1.00	0.75	0.42	0.35
Cultures	125	58	58	59

The final cell titer was calculated based on the number of colonies counted on permissive plate (no FA or Cu). The values for all three parental strains were pooled according to their respective exposures.

culture, re-isolated single colonies non-selectively, and then re-tested their resistance phenotypes to FA individually, Cu individually, and FA and Cu in combination. The majority of the clones re-tested resistant (94.8%), and FA and Cu resistance were nearly always associated, confirming that our double selection regimen is highly specific, and giving us high confidence that the colonies counted on the selective plates indeed correspond to true CNV events.

While the measurement of absolute mutation rates was informative, the interpretation of the reasons responsible for the differences observed between the various parent strains and exposures requires a deeper understanding of the molecular nature of the chromosomal rearrangements. This analysis, presented in the following section, provided useful insight as to why the CNV rates varied depending on where in the genome the reporter was inserted to measure the rates.

3.3 Spectra of associated chromosomal rearrangements

After confirming a positive CNV phenotype on the re-test plates, individual FA/Cu resistant clones (FCRs) were examined using PFGE and array-CGH in order to determine the molecular nature of their chromosomal rearrangements. To ensure independence of origin, only one FCR clone was analyzed per culture. The analysis of both spontaneous and environmentally-induced CNVs in the three parent diploid strains helped us gain a mechanistic understanding of the processes and classes of chromosomal rearrangements associated with each condition. The analysis of the array-CGH patterns was particularly informative in this regard. The summary array-CGH graphical gene dosage plots from all FCR clones analyzed across all conditions in this study are shown in Appendix II a-p.

The first parameter we examined from the array-CGH data was the total number of copies of the reporter cassette present in the FCR clones. Only one additional copy of the

reporter cassette was sufficient for growth on the selective plates, but higher order amplifications may also have been present among the FCR clones. The microarrays used in this study were custom-designed to have a high number of probes within the *SFA1* and *CUP1* genes. We scored the FCR to Parent hybridization ratios for these high-density reporter probes separately from the global copy number changes in the rest of the genome. This allowed for an isolated and reliable method for determining the number of copies of the reporter cassette in the mutated strains. In the array-CGH plots, the signal for the *SFA1* and *CUP1* genes was always displayed at their native positions on Chr4 and Chr8, respectively, regardless of where in the genome the CNV reporter was inserted. Therefore, every FCR clone had a positive spike that appeared at these two locations on the array-CGH analysis. Using the average signal within these spikes, read as the Log_2(FCR-Cy5/Parent-Cy3) value, we were able to estimate the number of copies of the reporter. As seen in Table 3, majority of the un-induced FCR clones contained only one extra copy of the reporter. This confirmed that the FA/Cu resistance selection conditions were optimally adjusted, supporting full viability of clones with low-order amplifications. Another pattern that was observed was that the FCRs derived from environmental exposures contained a higher number of copies of the reporter, particularly in the case of Chr5. This suggested that these exposures not only stimulated CNV formation, but also that those CNVs were more complex in structure.

This analysis also showed that many FCR clones derived from the Chr5 parent strain had higher order amplifications (more than one extra copy), whereas the Chr4 and Chr15 FCRs were mostly limited to low order amplifications. This observation suggested that the Chr5 region is more flexible in terms of tolerating higher amplification levels than the other two sites, which may have reduced viabilities associated with partial aneuploidy.

One important point worth noting is that we cannot rule out the possibility that some of the higher order amplifications may have occurred during the growth of the original FCR colony itself. In order to start growing on the FA/Cu selective plates the cells must have had at least one extra copy of the reporter, but since formaldehyde is a known mutagen (Kumari et al., 2012) it is possible that additional secondary events may have accumulated during the colony growth.

We then examined the array-CGH dataset to identify the patterns of genomic change present in the various samples. Three major mechanisms of reporter amplification were detected: Aneuploidy (full additional copy) of the chromosome, segmental duplication of the insertion region, and non-reciprocal translocation of the insertion chromosome arm. In some cases, FCR clones carrying complex genome rearrangements had more than one of these classes. The summary of CNV mechanisms is shown in Table 4.

The three parental strains analyzed displayed very specific mechanisms of reporter amplification. Under un-induced conditions, aneuploidy was the predominant mechanism for Chr5, and was much less frequent for Chr4, and only observed in one case for Chr15. This difference is likely reflective of the number and the specific genes present in each of these chromosomes, and indicates that gene dosage imbalances on Chr5 may not cause a deleterious effect on cell viability and are more tolerable that in the other larger chromosomes. Interestingly, in the mutagen-exposed FCRs the proportion of Chr5 segmental duplications and translocations increased relative to aneuploidy events, consistently with an increase in DNA damage processes that trigger chromosomal rearrangements.

Several structural properties of the genome rearrangements were noted including the size of the CNV events, and whether single (interstitial) or multiple (translocations) chromosomes

Table 4. CNV reporter amplification levels.

Parent Strain	Haploid	Diploids											
	Chr5	Chr5				Chr4				Chr15			
	Unind.	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS
FCR clones analyzed by array-CGH	29	30	10	12	10	24	12	9	11	24	11	10	8
Reporter copy number:													
2	18	13	1	0	0	22	8	6	6	22	10	5	8
3	11	9	1	2	3	2	4	3	5	1	1	4	0
4	0	4	4	5	3	0	0	0	0	0	0	0	0
5+	0	4	4	5	4	0	0	0	0	1	0	1	0

Array-CGH data was analyzed and the number of copies of the reporter was determined based on the relative hybridization signal between FCR and parent strain for the high density probes at *SFA1* and *CUP1*.

were involved. The breakpoints (defined as the specific positions where gene dosage changes were detected) were also scored in order to determine if recombination occurred at sites containing dispersed repetitive DNA elements or at single copy sequences (micro-homology or non-homology). This data supported the interpretation of the recombination pathways involved with each CNV event, and whether there were specific patterns associated with each exposure or chromosome. As shown in Table 5, the vast majority of breakpoints were found at dispersed Ty or LTR repeats, the most abundant class of yeast repetitive DNA elements, comprising roughly 3% of the yeast genome (Kim et al., 1998). The second most abundant class of breakpoints included those found at sites containing other forms of conserved sequences such as gene family members (*e.g. EFT1* and *EFT2*, discussed in next section). Finally, single copy sequences were also detected, but they rare. These results indicated that the primary mechanism associated with CNV formation in our dataset was the repair of DNA double-strand breaks through non-allelic homologous recombination (NAHR), and that microhomology or non-homologous repair pathways only played a minor (if any) role.

In addition to the chromosomal rearrangements associated with amplification of the CNV reporter, we also detected numerous cases of other gene dosage changes in regions of the genome that were unrelated to FA/Cu resistance. These events (Table 6) involved the same categories of events described above, including aneuploidies, segmental duplications and deletions, and non-reciprocal translocations. Since these events did not directly affect the selected phenotype, it is unclear why they were so abundant in the FCRs. Possible explanations

Table 5. CNV mechanisms and rearrangement breakpoints associated with the reporter.

Parent Strain	Haploid	Diploids											
	Chr5	Chr5				Chr4				Chr15			
	Unind.	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS
CNVs associated with the reporter:													
Aneuploidy:	0	21	5	6	6	5	1	2	2	1	0	0	0
Segmental duplication:	29	10	7	9	6	0	0	0	1	1	1	2	0
Translocation:	1	8	6	7	7	21	11	9	11	23	11	9	7
Rearrangement breakpoints:													
at Ty or LTR repeats:	57	35	27	31	27	42	24	31	27	31	22	27	9
At other repeats:	0	6	0	0	0	0	2	0	0	14	2	2	6
at single copy sequences:	0	0	0	0	0	0	0	0	0	3	0	2	0

CGH-array data of the CNV events associated with the reporter was analyzed; it was determined if the regions of the breakpoints contained dispersed repetitive DNA elements or single copy sequences. The types of events: aneuploidy, segmental duplications, and translocations, were also noted.

Table 6. CNV mechanisms and rearrangement breakpoints not associated with the reporter.

Parent Strain	Haploid	Diploids											
	Chr5	Chr5				Chr4				Chr15			
	Unind.	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS	Unind.	CPT	HU	MMS
CNVs not associated with the reporter:													
Aneuploidy:	0	12	2	1	3	5	2	0	1	1	8	6	4
Segmental duplication:	2	2	0	0	1	1	1	2	2	2	0	2	1
Segmental deletion:	0	2	0	0	0	1	0	0	0	0	0	1	1
Translocation:	0	15	0	1	0	5	1	1	0	2	2	1	2
Rearrangement breakpoints:													
at Ty or LTR repeats:	4	11	0	2	2	10	8	8	2	7	6	8	7
At other repeats:	0	6	0	0	0	0	0	0	0	0	0	0	0
at single copy sequences:	0	0	0	0	0	0	0	0	0	1	0	0	1

CGH-array data of the CNV events not associated with the reporter was analyzed; it was determined if the regions of the breakpoints contained dispersed repetitive DNA elements or single copy sequences. The types of events: aneuploidy, segmental duplications, segmental deletions, or translocations, were also noted.

for this high incidence include: (1) the un-associated rearrangements may have been triggered by the growth in the presence of formaldehyde; (2) these events may have been indirectly selected in the FCR clones as second site suppressors that rescued the negative effects of the gene dosage imbalances caused by the primary (reporter-associated) rearrangements; and (3) recent studies have shown that strains carrying genome rearrangements have higher level of genome instability and progressively accumulate secondary changes (*i.e.* CNV begets more CNV) (Zhu et al., 2012). Regardless of their mechanism of formation, the un-associated rearrangements detected were indistinguishable in structure and breakpoint sequences when compared to the selected CNVs, supporting the conclusion that aneuploidy and NAHR were the primary mechanisms of structural variation in our dataset.

As stated above, two major classes of CNV rearrangements were detected: Segmental duplications and non-reciprocal translocations. Segmental duplications had been characterized in yeast previously as the predominant class of rearrangements detected in haploid cells. This is to be expected since segmental duplications do not involve loss (deletion) of genomic regions, and therefore are more easily recoverable in haploids. In contrast, this class was not highly represented in our diploid strain set, and was almost exclusively found in FCRs derived from the Chr5 reporter, which was flanked by Ty1 repeats. Very few segmental duplications were recovered from the other two parent strains that had either distant flanking repeats (Chr15), or no flanking repeats at all (Chr4).

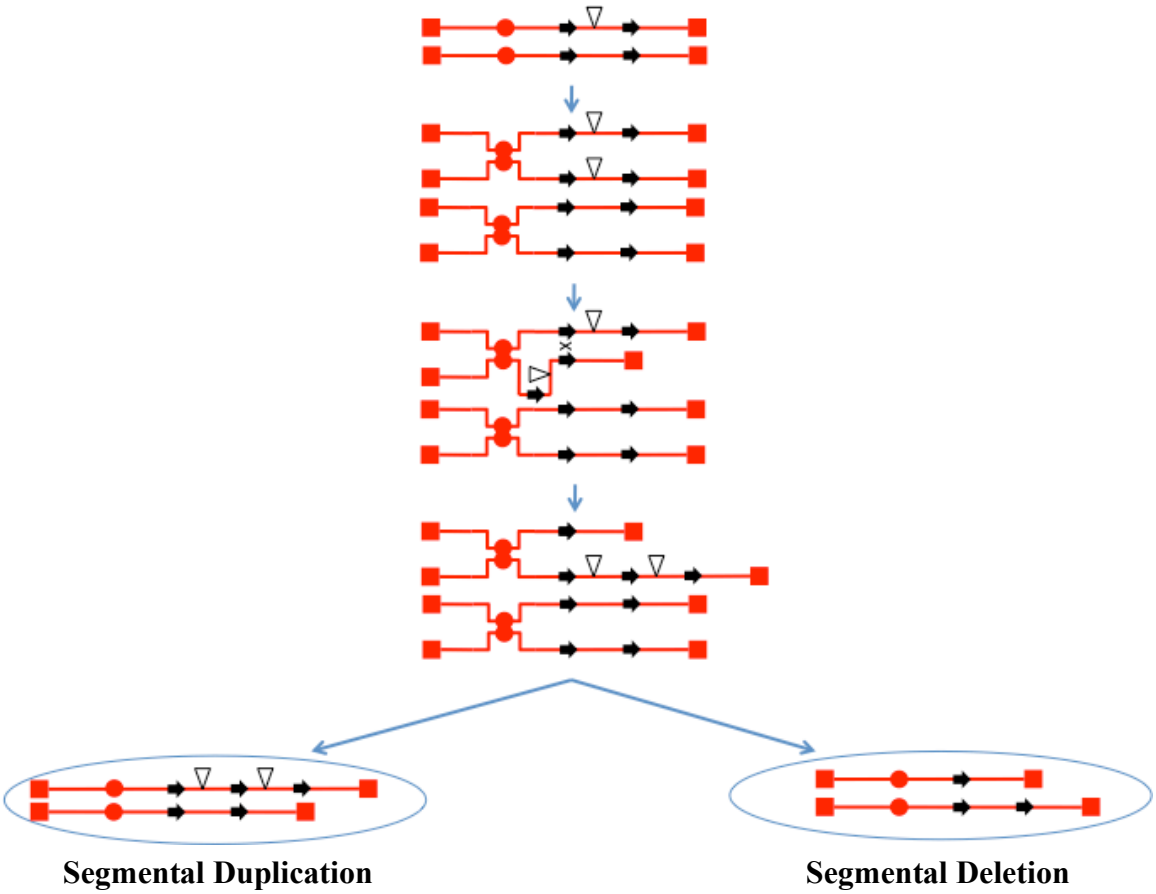


Figure 3. Possible mechanism of segmental duplication for CNV formation. During replication a cross-over type mechanism occurs between non-allelic repeat element (black arrows) on the two sister chromatids resulting in unequal recombination. The end product being two cells: one with a segmental duplication, and the other with a segmental deletion that the assay screens against.

A possible mechanism of formation of CNV by segmental duplication is shown in Figure 3. It is thought that these events occur through an unequal crossover type mechanism between the flanking repetitive elements on sister chromatids or on the homologous chromosome. This mechanism would result in two daughter cells, one with a segmental duplication, which would be selected for with in the CNV assay, and the other with a segmental deletion that would not be detected or inviable.

The majority of the chromosomal rearrangements identified among the FCRs were non-reciprocal translocations. This type of event is characterized by the deletion of a terminal segment of a chromosome, coupled with the amplification of terminal section of the chromosome arm where the CNV reporter was inserted. Figure 4 illustrates a possible mechanism of non-reciprocal translocation formation, in this case by a Break-Induced Replication mechanism, but a conventional mitotic cross over could also produce the same outcome (Paques and Haber, 1999). In diploids, when a DNA double-strand break occurs homologous recombination repair is most often carried out by using the allelic sequence on the sister chromatid or homolog as template. This type of event repairs the break with high fidelity resulting is no change in karyotype. However, if resection of the DSB lesion is long enough to expose a repetitive element, this may result in recombination with a non-allelic copy of the repeat on a different chromosome, in this case the chromosome with the CNV reporter insertion. The outcome is thus non-reciprocal, producing the amplification of one arm of a chromosome and a deletion of the other.

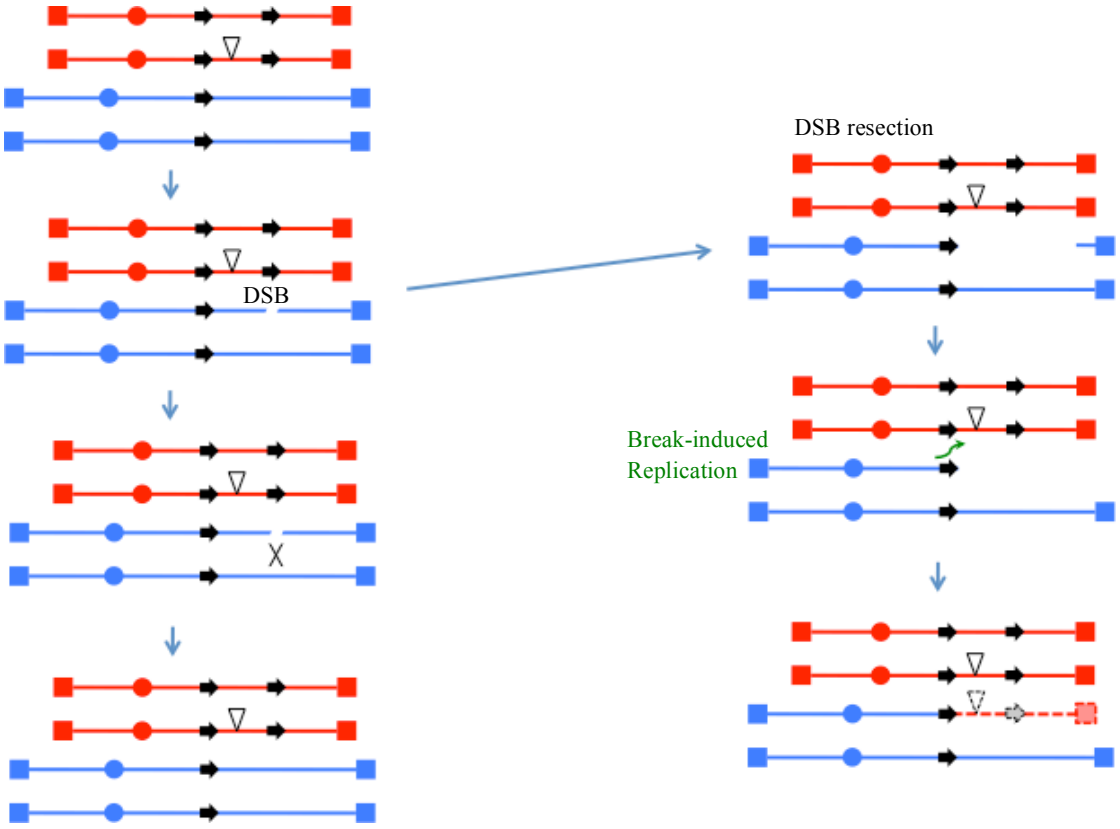


Figure 4. Possible mechanism of non-reciprocal translocations for CNV formation. When a double strand break occurs, allelic repair is the common result and then there is no karyotype change (left side of the figure). While less common, excessive resection of the DSB may occur, thus exposing a repetitive element. This exposed element then anneals to a similar element on a different chromosome, and primes DNA replication (BIR) using the other chromosome as template. This results in an extra copy of the template chromosome (in this case the one containing the reporter) and a deletion of the terminal segment of the broken chromosome.

Since non-reciprocal translocations always involve deletions, they are very difficult to recover in haploids, which have only very limited sections of the genome that are dispensable for viability (Chan and Kolodner, 2011, 2012; Putnam et al., 2009). In contrast, we observed a high incidence of this class of rearrangements in all three of our diploid strains because the cells remained viable even after loosing whole chromosome arms.

We analyzed all the non-reciprocal translocations recovered in our dataset and observed a very striking recurrent pattern: frequent deletion of the terminal section of the right arm of Chr7. This specific deletion accounted for more than half of all terminal deletions, and was observed in similarly high proportion across all parent strains and all exposure conditions (Table 7). An interesting feature of our assay is that the precursor DSB lesions giving rise to the chromosomal rearrangements can occur anywhere in the genome and have a similar opportunity to be recovered as a non-allelic recombination interaction with the chromosome harboring the reporter, regardless of where in the genome the reporter is located. Therefore, the expectation was that if DSB lesions were distributed randomly throughout the genome, then we should have recovered a wide assortment of deletions involving many chromosomes. Instead, the right arm of Chr7 accounted for 62 deletions, while all the 31 remaining chromosome arms combined accounted for only 56 deletions. The strong bias suggested that either breaks are formed on Chr7R much more often than in other regions (possibly due to the presence of a fragile site), or that the breaks that form there are more likely to undergo non-allelic repair. Specific models to account for this bias are presented in the Discussion.
somewhat unusual. These include three spontaneous FCRs (analyzed previously in our

Table 7. Chr7 deletions associated with CNV events amplifying the reporter.

CNV reporter arm amplifications	Associated deletions	
	Chr7 right arm	All other Chr arms
All reporter sites	62	56
Chr5 unexposed	2	3
Chr5 HU	3	3
Chr5 CPT	4	4
Chr5 MMS	2	3
Chr4 unexposed	11	10
Chr4 HU	6	3
Chr4 CPT	9	2
Chr4 MMS	9	3
Chr15 unexposed	8	7
Chr15 HU	3	6
Chr15 CPT	2	5
Chr15 MMS	3	7

3.4 Specific events associated chromosomal rearrangements

The complete set of diploids analyzed by molecular karyotyping in this study added up a total of 170 FCR clones, all of which carried some form of genomic rearrangement. The detailed discussion of each one is beyond the scope of this thesis, however the specific rearrangements found in six representative FCRs is presented below to illustrate some of the more commonly seen events, as well as others that were laboratory by Ane Zeidler), and three drug-induced FCRs that were analyzed as part of this thesis project.

3.4.1. Example of a spontaneous Chr5 amplification event. Figure 5 shows the karyotyping data for clone FCR36. Many of the Chr5 derived strains became resistant, not by chromosomal rearrangement, but by obtaining an extra copy of the entire chromosome (*i.e.* trisomy). In this case a new chromosomal band of about 850 Kb was seen on the PFGE, as well as a more intense band for the parental Chr2 and Chr5 bands. In the array-CGH, there were aneuploidy events that occurred on Chr5 (reporter associated), and on Chr2 (not-associated). In addition, a segment of Chr5 was further amplified from the *YERCTy1-1* repeat element to the right end of the chromosome (*TEL05R*), and segment of Chr13 was deleted from the *YMRCTy1-5* repeat to *TEL13R*. The remaining (left side) of the broken Chr13 was joined to the amplified segment of Chr5 containing the reporter to form a Chr13/Chr5 non-reciprocal translocation predicted to be ~850 Kb in size, matching the new chromosome detected by PFGE. Therefore the FCR36 clones contains two extra copies of the CNV reporter, one associated with the Chr5 trisomy and one associated with the Chr13/Chr5 translocation.

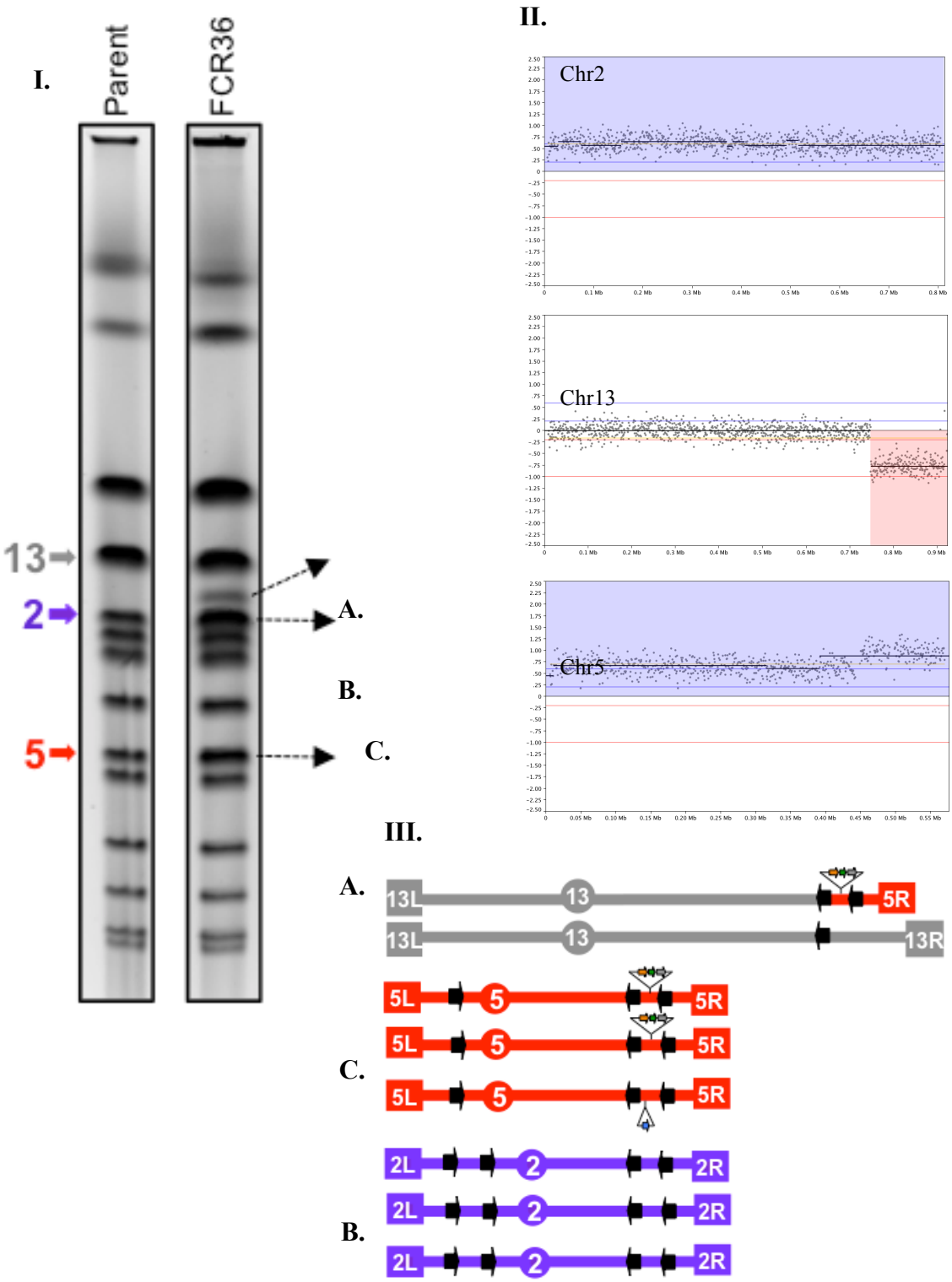


Figure 5. Qualitative analysis of CNV-associated rearrangements in unexposed Chr5 diploid strain. PFGE (**I**) of FCR36 shows three events as marked by A, B, and C. Comparing array-CGH data (**II**) for this sample with the PFGE allows for the interpretation of the structural mutations that occurred. Schematic maps of each corresponding event are shown **III**. The FCR36 karyotype consisted of a Chr13/Chr5 reporter associated translocation event, a Chr5 reporter associated trisomy event, and a Chr2 trisomy event not associated with the reporter.

3.4.2. Example of a spontaneous Chr15 amplification event. While most of the NAHR events were mediated by recombination between Ty or LTR dispersed repeats, we also detected cases of interaction between other classes of repeats. One such case is shown in Figure 6 for the FCR165 clone. The PFGE for FCR165 indicated a new band around 1,800 Kb. The array-CGH detected a deletion on the right arm of Chr4 from the *EFT1* gene to *TEL04R*, and the amplification of the Chr15 containing the reporter from the *EFT2* gene to *TEL15R*. *EFT1* and *EFT2* both encode the translational elongation factor 2 (EF-2) protein, and share a 99.8% nucleotide sequence identity over a 2.5 Kb region. The predicted size for the Chr4/Chr15 translocation at the *EFT* genes matched the size of the new band observed by PFGE. Interestingly, we also detected the exact opposite event in FCRs derived from the strain containing the CNV reporter on Chr4 (data not shown). In those cases the FCR clones had a deletion on Chr15 and an amplification on Chr4 associated with a Chr15/Chr4 translocation with breakpoints at the same *EFT* genes.

3.4.3. Example of a spontaneous Chr4 amplification event. The FCR112 clone, illustrated in Figure 7, displayed a new PFGE band at around 1,300 Kb and the array-CGH indicated amplification on the right arm of Chr4 from *YDRCTy1-3* to *TEL04R*, and a deletion on the right arm of Chr7 from *YGRCTy1-3* to *TEL07R*. This Ty-mediated NAHR Chr7/Chr4 translocation was typical of the events associated with deletions on Chr7 that were over-represented in our dataset.

3.4.4. Example of a MMS-induced Chr4 amplification event. Karyotyping data for FCR271, which was exposed to MMS, is seen in figure 8. A new band around 1,400 Kb was detected in the PFGE, as well as a band indicating the presence of one slightly larger Chr2. The fainter band for the parental size Chr2 was also observed. Array-CGH

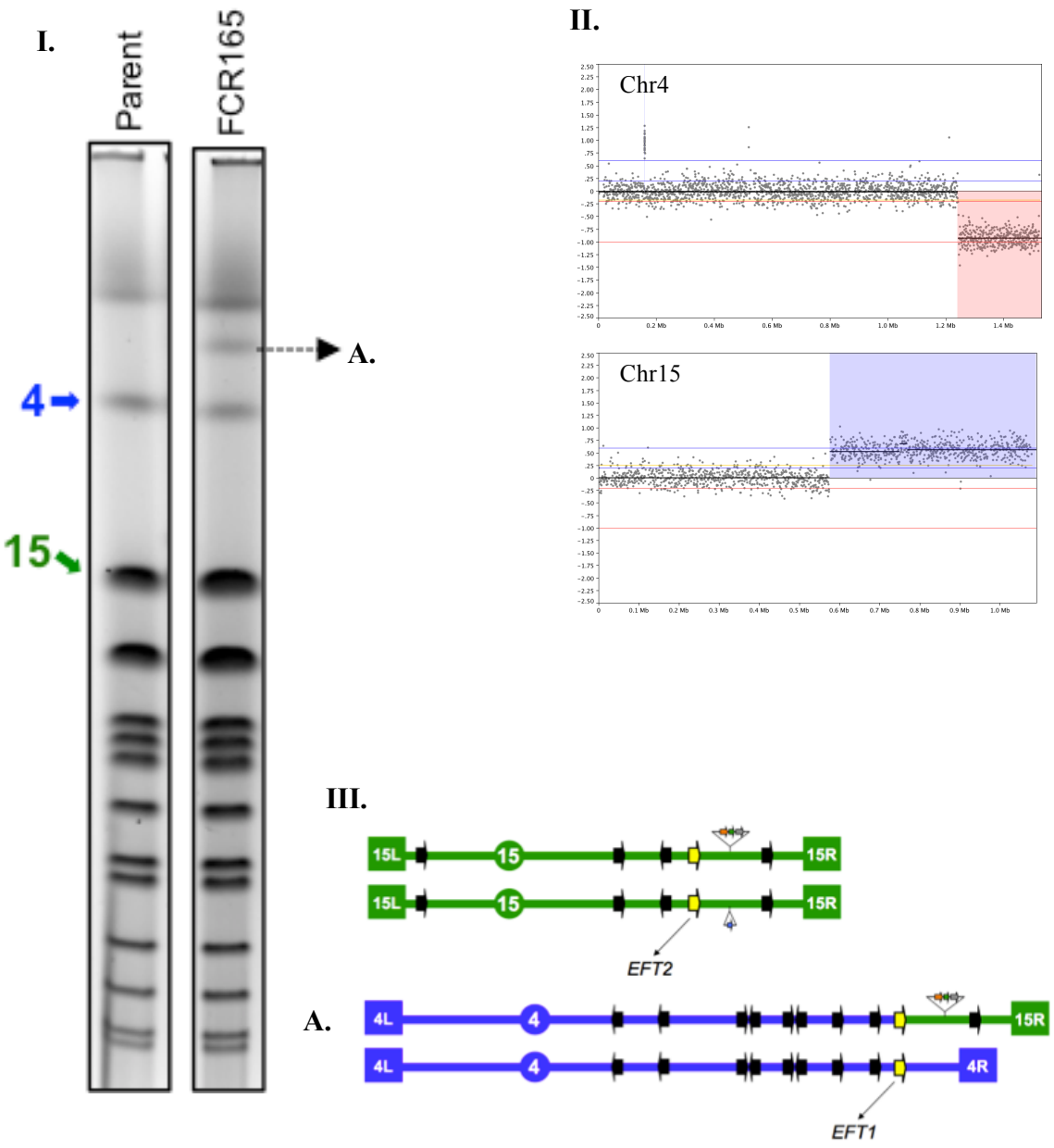


Figure 6. Qualitative analysis of CNV-associated rearrangements in unexposed Chr15 diploid strain. PFGE (**I**) of the sample (right) and the parent (left) show a new band as marked by A. Comparing CGH-array data (**II**) for this sample with the PFGE allows for the interpretation of the mutation that occurred. Schematic map of this event is shown in section **III**. FCR165 karyotype was a Chr15/Chr4 reporter associated translocation event occurring due to recombination at repetitive sites containing *EFT1* and *EFT2* genes.

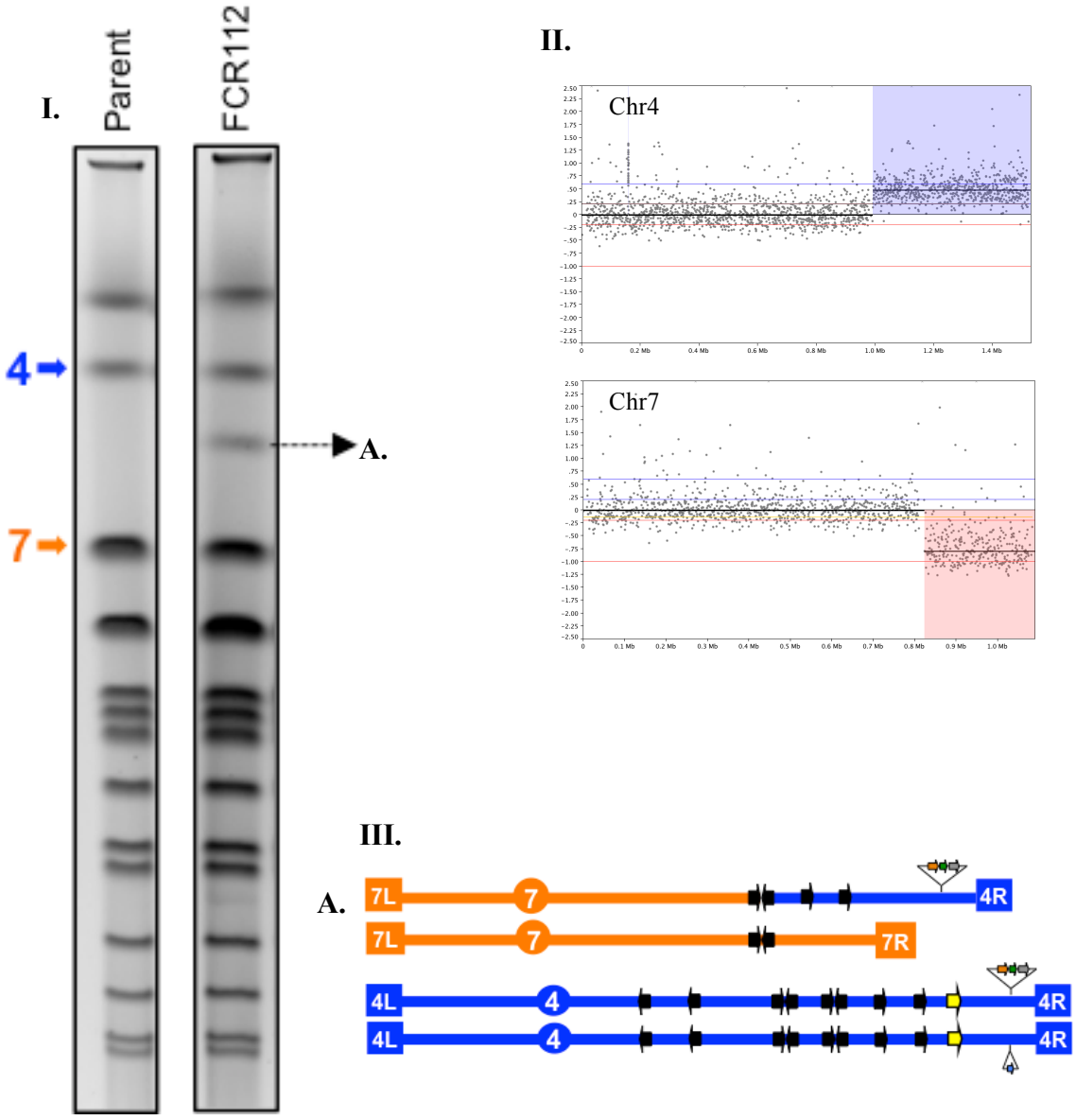


Figure 7. Qualitative analysis of CNV-associated rearrangements in unexposed Chr4 diploid strain. PFGE (**I**) of the sample (right) and the parent (left) show a new band as marked by A. Comparing CGH-array data (**II**) for this sample with the PFGE allows for the interpretation of the mutation that occurred. Schematic map of this event is shown in section **III**. FCR112 karyotype was a Chr7/Chr4 reporter associated translocation event occurring at LTR/TY repetitive sequences.

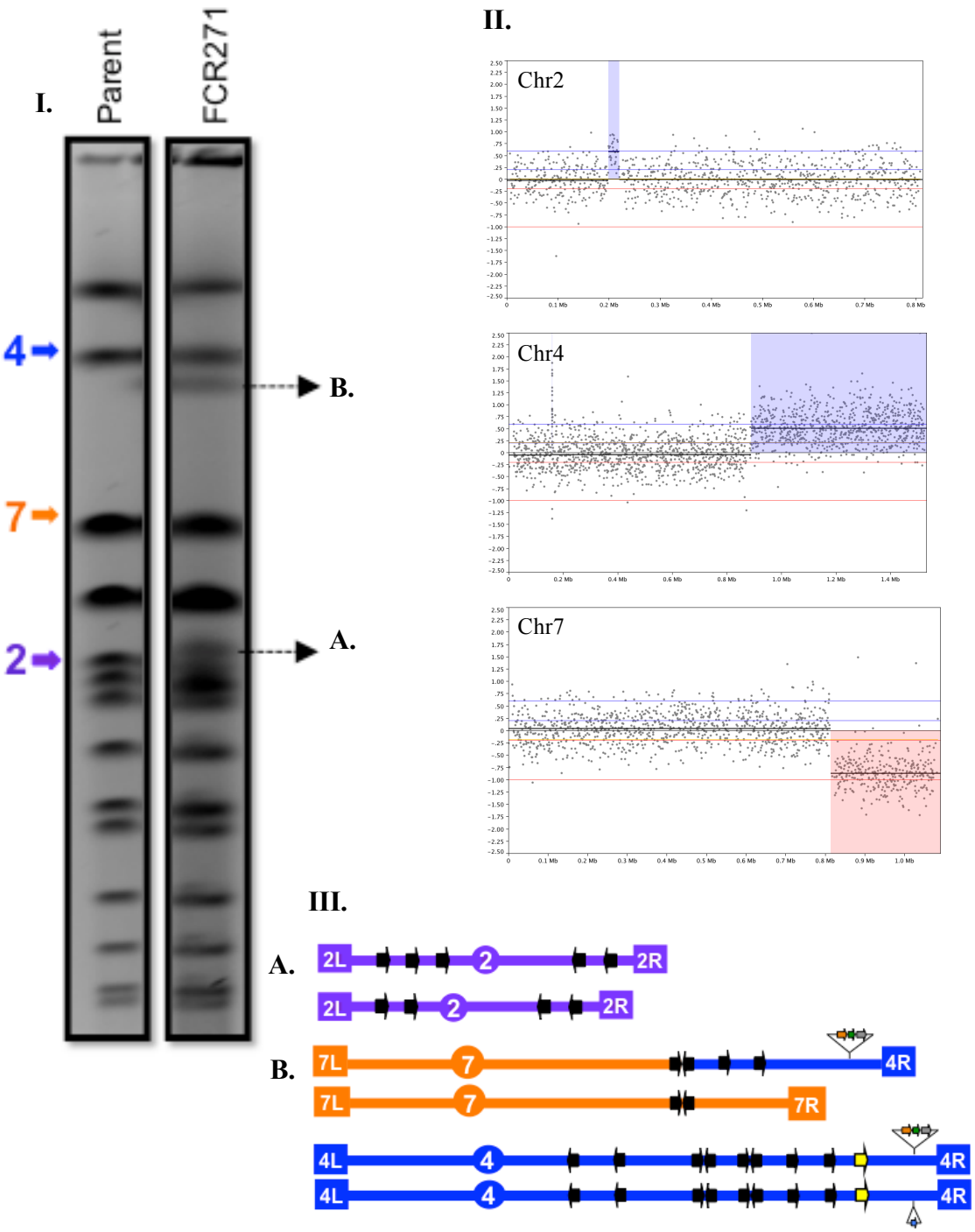


Figure 8. Qualitative analysis of CNV-associated rearrangements in MMS induced Chr4 diploid strain (FCR271). PFGE (**I**) of the sample (right) and the parent (left) show two new events as marked by A and B. Comparing CGH-array data (**II**) for this sample with the PFGE allows for the interpretation of the mutation that occurred. Schematic maps of these events are shown in section **III**. FCR271 karyotype was a Chr7/Chr4 reporter associated translocation event occurring at LTR/TY repetitive sequences, and a Chr2 segmental duplication not associated with the reporter.

data indicated a 600kb amplification on the right arm of Chr4 from *YDRCTy1-2* to *TEL04R*, and deletion on the right arm of Chr7 from *YRGCTy1-3* to *TEL07R*, leaving a 815kb segment of Chr7 to recombine. This Chr7/Chr4 Ty1-mediated NAHR event was another example of the common Chr7 deletion associated events. The array-CGH also indicated a less common ~20 Kb segmental duplication, not associated with the reporter, occurring on Chr2. The amplification on the left arm of Chr2 was from an un-annotated Watson-oriented Ty1 element near the *tG(GCC)B* tRNA gene to the *YBLWTY1-1* repeat element.

3.4.5. Example of a HU-induced Chr5 amplification event. Just as with the un-induced samples, trisomy and terminal translocation events were prevalent throughout the Chr5 exposures. Figure 9 shows the HU-induced Chr5 karyotype of FCR290. The PFGE shows a darker band on Chr5 and a new band around 850Kb. The array-CGH presented a trisomy event on Chr5 as well as a right arm120kb amplification from *YERCTy1-1* to the end of the chromosome (*TEL05R*), and also indicated a deletion on the right arm of Chr13 from *YMRCTy1-5* to *TEL13R*, the remaining (left side) of Chr13 being about 750kb. One extra copy of the reporter was gained through trisomy of Chr5 and another due to a Chr5/Chr13 NAHR terminal translocation. Interestingly, the rearranged karyotype in FCR290 was indistinguishable from that of FCR36 (Figure 5), except for the un-associated Chr2 trisomy.

3.4.6. Example of a CPT-induced Chr15 amplification event. FCR198, as illustrated in figure 10, was a clone induced by CPT and also had a Chr7 deletion associated event. A new band around 1000Kb and a faint band for Chr10 appeared on the PFGE. The array-CGH data had an aneuploidy deletion on Chr10, and a deletion on

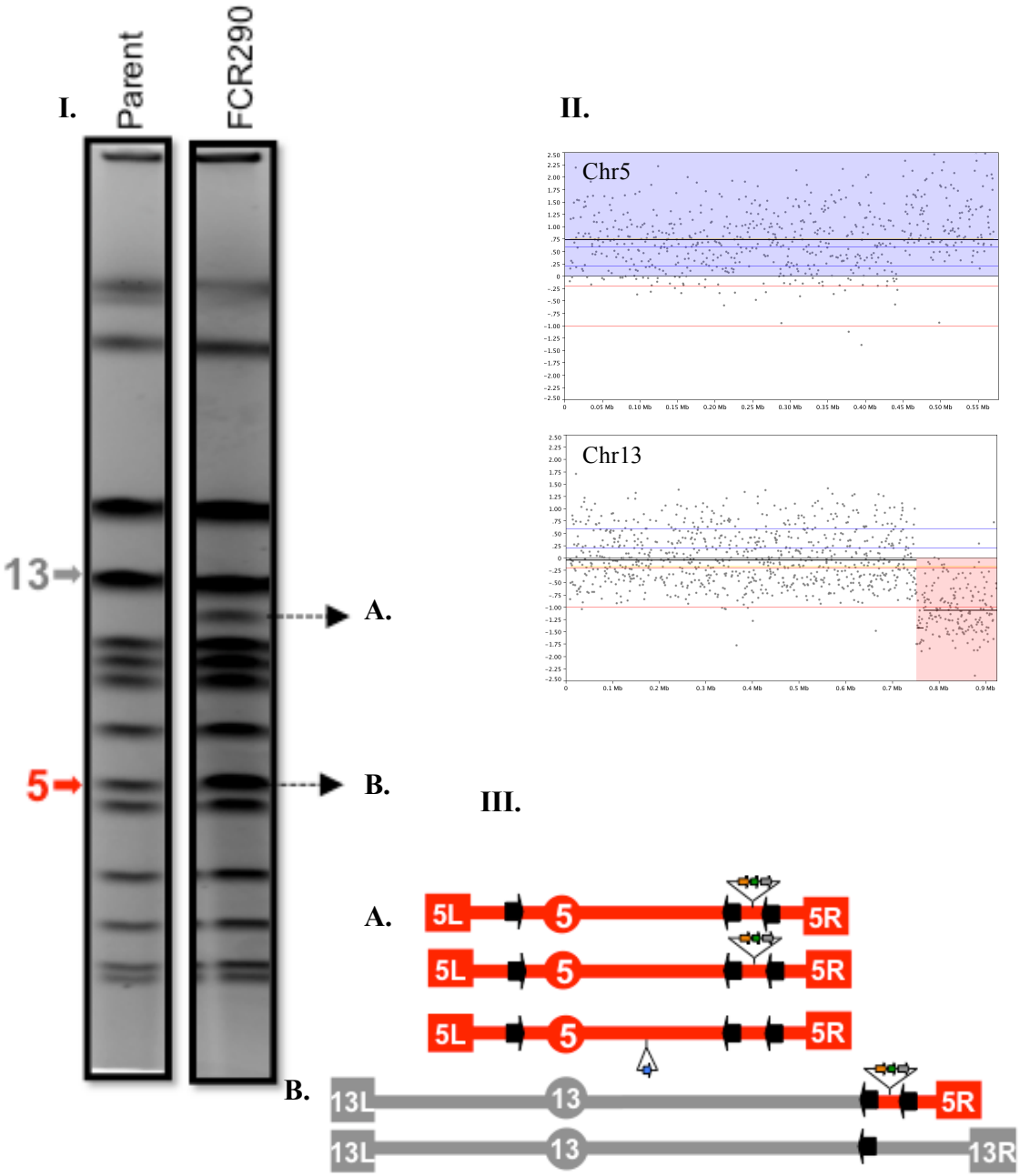


Figure 9. Qualitative analysis of CNV-associated rearrangements in HU induced Chr5 diploid strain (FCR290). PFGE (**I**) of the sample (right) and the parent (left) show two new events as marked by A and B. Comparing CGH-array data (**II**) for this sample with the PFGE allows for the interpretation of the mutation that occurred. Schematic maps of these events are shown in section **III**. FCR290 karyotype was a Chr5/Chr13 reporter associated translocation event occurring at LTR/TY repetitive sequences, and a Chr5 reporter associated trisomy event.

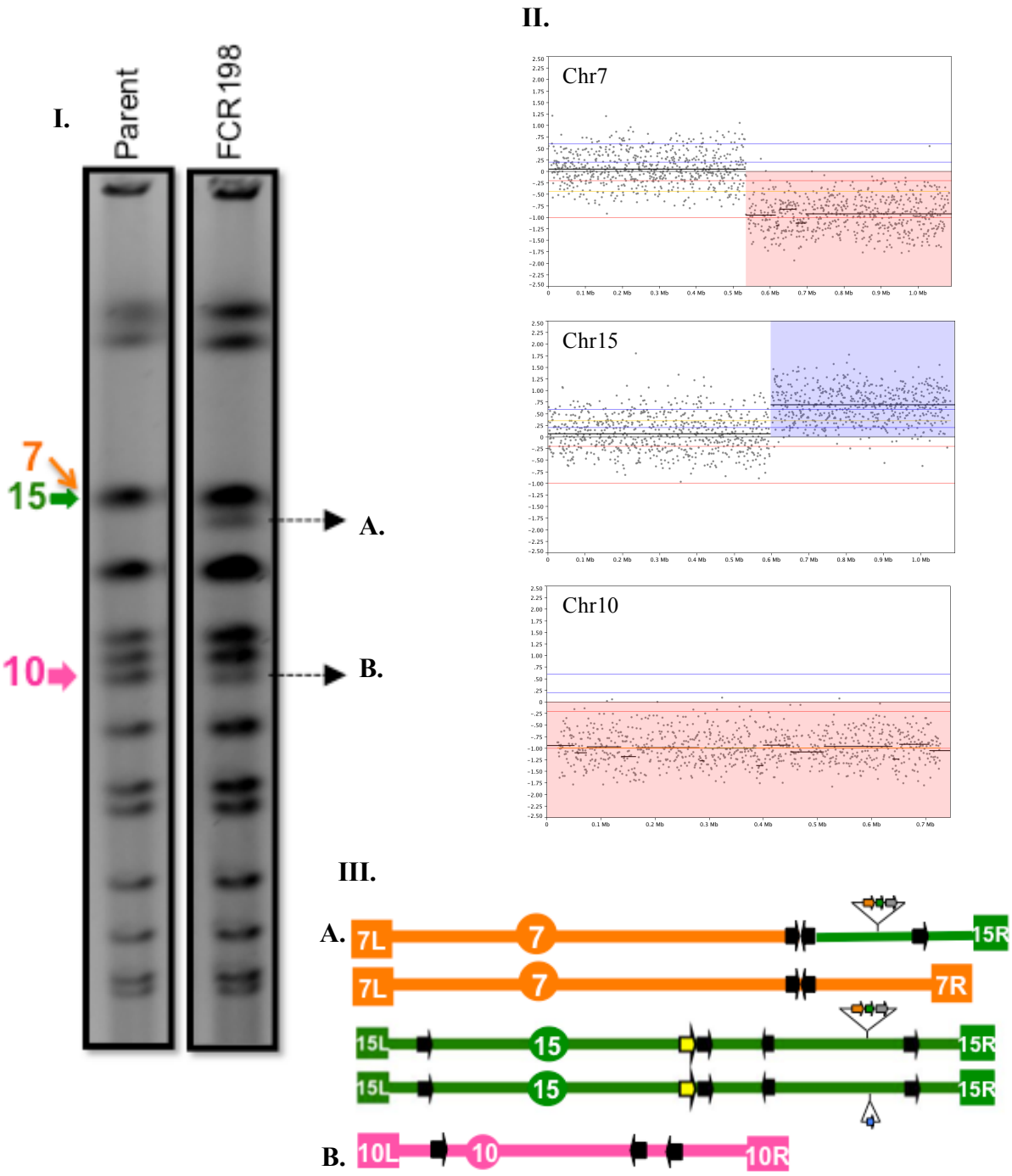


Figure 10. Qualitative analysis of CNV-associated rearrangements in CPT induced Chr15 diploid strain (FCR198). PFGE (**I**) of the sample (right) and the parent (left) show two new events as marked by A and B. Comparing CGH-array data (**II**) for this sample with the PFGE allows for the interpretation of the mutation that occurred. Schematic maps of these events are shown in section **III**. FCR198 karyotype was a Chr15/Chr7 reporter associated translocation event occurring at LTR/TY repetitive sequences, and a Chr10 monosomy event not associated with the reporter.

Chr7's right arm from *YGRWTY1-1* to *TEL07R*, leaving a 535kb region of Chr7. A 490kb amplification on the right arm of Chr15 (including the reporter) from *YORWTy1-2* to *TEL15R* was also detected. Therefore the karyotype changes for this FCR clone included a Chr7/Chr15 NAHR terminal translocation and an aneuploidy event on Chr10.

CHARTER 4

DISCUSSION

Through the early phases of development and validation of the CNV assay we have been able to show that this assay is in many ways significantly more informative than other genome stability assays previously available in the yeast model system. The reporter system described here allowed for an effective method to study chromosomal aberration events in diploids, whereas previous systems were functional only in haploid cells. One such system detected the amplification of *RPL20B,* a gene encoding an essential ribosomal protein (Koszul et al., 2004; Koszul et al., 2006; Payen et al., 2008). Duplication of *RPL20B* rescues the cells from the growth defect caused by a deletion of the homologous *RPL20A* gene. This dose compensation selection system is only effective in haploid cells. Another assay that has been extensively used was developed by the Kolodner laboratory to study gross chromosomal rearrangements (GCR assay) (Chan and Kolodner, 2011, 2012; Chen and Kolodner, 1999; Putnam et al., 2009). The GCR assay, also exclusive to haploids, is based on simultaneous loss of two counter selectable markers that is associated with deletions of a non-essential terminal chromosomal segment. These two assays have been used for several years now, and have contributed enormously to our understanding of the nature and the genetic control of genome stability in yeast, and to model this process in all eukaryotes. Despite their usefulness, both share the limitation of investigating genome rearrangements in haploids that may not fully reflect the mechanisms occurring in the diploid genomes they sought to model. In addition, the conclusions reached from each assay were derived from experiments done at a single specific site in the genome (*RPL20B* at Chr15

right arm, and GCR at Chr5 left arm), therefore they were narrow in scope because they ignored a likely role of genomic context in chromosomal rearrangements.

Our results showed that the three major mechanisms are associated with CNV in diploid cells: Aneuploidy, segmental duplications, and non-reciprocal translocations. In addition, we showed that genomic context plays a key role in determining the balance between these classes. In contrast, each of the haploid assays described above had fairly limited spectra of events: primarily segmental duplications for *RPL20B* and primarily non-reciprocal translocations for GCR. Our own work included the analysis of the *SFA1-CUP1* reporter on Chr5 in a haploid state and also produced a very limited CNV spectrum consisting of only segmental duplications between the flanking Ty1 repeats (the complete haploid CNV dataset will be presented elsewhere).

Another important goal of our study was to identify the recombination breakpoints associated with chromosomal rearrangements that arise spontaneously or after low dose exposure to DNA damaging agents. Previous work in our group had shown that in diploid cells exposed to a high dose of ionizing radiation, nearly all of the resulting chromosomal rearrangements were formed by NAHR at Ty and LTR elements (Argueso et al., 2008). While those results were conclusive, we were unable to definitively demonstrate whether the pronounced abundance of NAHR recombination over NHEJ and MMBIR was observed because this was indeed the predominant mechanism in diploids, or if the result was an artifact of the high radiation doses used to induce the genome rearrangements. This concern was particularly important because previous studies had reported a high incidence of non- or micro-homology events in un-induced haploids (zero dose). The results obtained in the present study, for both un-induced and low exposure to DNA damage, convincingly confirmed that NAHR is in fact the dominant DSB

repair pathway responsible for the formation of chromosomal aberrations in diploid cells, therefore validating our earlier studies.

Although rare, some breakpoints were also detected at other repetitive regions or at single copy sequences indicating that the instability mechanisms of NHEJ and MMBIR also occurred, but our results suggest that these sequences are only used as a last resort when true homology is not available (*i.e.* G1 haploids) for use as template in the repair reaction.

Another interesting result that was unique to our study was the high frequency of aneuploidy events detected on the Chr5 reporter strain. Aneuploidy on Chr5 was substantially higher than on the Chr4 and Chr15. A study by Torres et al. in haploid cells found that aneuploidy of certain chromosomes are associated with loss of cell viability, however, Chr5 appeared to be one of the most tolerable aneuploidies, with only a minor negative effect on growth (Torres et al., 2007). It is reasonable to speculate that Chr5 aneuploidy events in diploids would not have a pronounced decreased viability, especially considering that the effects aneuploidy are thought to be more severe in haploids than in diploid strains (one extra chromosome equals 100% increase in copy number in haploids, but only a 50% increase in diploids), In contrast, Chr4 and Chr15 aneuploidy were the two most taxing to viability in the Torres et al. haploid study, consistently with the low incidence of aneuploidy of these two chromosomes in our diploid CNV clones.

Aside from the expected overall increase in CNV, no major differences in the qualitative classes of un-induced and drug-exposed rearrangements. A good example of this was the comparison between the Chr5-derived clones FCR36 (un-induced; Figure 5) and FCR290 (HU-exposed; Figure 9), which had indistinguishable reporter-associated CNV events. We also did not detect any significant differences in the rearrangements recovered between different DNA

damage exposures. The only noticeable and relevant alteration was in the overall incidence of aneuploidy on Chr5. These aneuploidy events were prevalent in both the spontaneous samples and the three types of exposures. However, the relative contribution of translocations and segmental duplications increased in the FCR clones exposed to HU, CPT, and MMS. Among the un-exposed Chr5 clones we observed a ratio of 21 aneuploidies to 18 translocations plus segmental duplications. This ratio in the combined exposed samples changed to 17:42. This result indicates that while the incidence of Chr5 aneuploidy events remained steady (none of these drugs is involved in the chromosome segregation processes associated with aneuploidy), the rearrangement mechanism that depend on DSBs increased by about three fold when the cells were exposed to the DNA damaging chemicals. This explains why the increase in CNV mutations in exposed culture was not as pronounced for Chr5 when compared to the Chr4 and Chr15 strains, which did not have a significant aneuploidy component (Table 5).

Throughout all three of the chromosomes studied, whether un-induced or exposed to mutagens, a remarkable bias of deletions occurring on the right arm of Chr7 (Table 7). This result was completely unexpected, and understanding it would undoubtedly contribute significant new insight into the processes that affect the formation of chromosomal rearrangements in yeast. Two basic steps are thought to be absolutely necessary for homologous recombination: First, a DSB lesion must form on the DNA to recruit the enzymes that carry out the repair reactions; and second, the broken DNA end (recipient sequence) must be resected by a 5-3' exonucleases and then physically come in contact with a suitable homologous template sequence (donor sequence). In the case of NAHR, the donor should be a homologous sequence from a different (non-allelic) position in the genome. Taking these factors into consideration, we can envision a few possible models to explain why Chr7R was so disproportionately represented in the FCRs.

The first possibility is that a strong chromosomal fragile site exists in the region. These specific genomic regions are prone to the formation of DSBs and have been well characterized in yeast and in humans (Durkin and Glover, 2007). A fragile site generating a frequent DSB on Chr7R would explain the high incidence of deletions that we observed, whereas other (non-fragile) regions of the genome would not appear very frequently. However, if this scenario was true, we would predict that by inducing a large number of additional randomly distributed DSB lesions through exposure to MMS, HU, and CPT might increase the contribution of other regions of the genome relative to Chr7R. This was not observed. The proportion of Chr7R deletions did not change when cells were exposed to DNA damage, thus suggesting that a fragile site may not be the only cause for the recurrent Chr7R deletions.

The next thought we considered is related to the spatial organization of chromosomes in the nucleus. Recent advances in the development of chromosome conformation capture (3C-seq) have shown that the organization of eukaryotic genomes is much more orderly than previously contemplated (Tanizawa and Noma, 2012). The fact that the genome organization in the nuclear space is so well defined leads to the idea that certain genomic regions in close proximity may associate physically more often than others (Hakim and Misteli, 2012). This concept has been recently demonstrated in the mouse genome where DNA spatial co-localization was highly correlated with the formation of recurrent chromosomal rearrangements (Zhang et al., 2012b). Even in lower eukaryotes like *S. cerevisiae*, each chromosome has been shown to occupy preferential territories and certain chromosomes appear to be in closer physical contact with their neighbors, and conversely are more distant to other chromosomes that occupy different territories (Duan et al., 2010). We therefore examined the published 3D structure of the yeast genome and asked whether the recurrent rearrangements observed in our experiments could be

explained though this mechanism. Specifically, we compared the position of Chr7 to that of the three arms where the CNV reporter was inserted. While the Duan et al. 3-D visual models show that Chr5 and Chr7 appear to be located close to each other in the nucleus, Chr7 and Chr15 are located farther apart, and Chr7 and Chr4 are separated even further. Thus the proximity of Chr7 to the CNV reporter, alone, does not provided a robust explanation for the high frequency of its involvement in the observed translocations.

In addition, if the nuclear localization was firmly determined and spatially constrained, then we would expect different and specific sets of recurrent rearrangements for each CNV reporter insertion. For example, the Chr5 insertion would recombine more often with its specific neighbors, while Chr15 and Chr4 should have their own preferential sets of interactions. Instead, we observed that Chr7R was the preferential deletion recovered in all three reporter insertions.

The third possibility is that the spatial position of the normal (intact) chromosomes is not a determinant of inter-chromosomal recombination, but rather the ability of broken chromosomal ends to move around the nuclear space in search of homology is more important. Two recent studies used fluorescence microscopy to follow the motion of yeast chromosomes in live cells before and immediately after the induction of a site-specific DSB (Dion et al., 2012; Mine-Hattab and Rothstein, 2012). These studies found that the broken chromosome ends explored a 10-fold larger nuclear space than the same sequence normally occupied while the chromosome was intact. This intriguing observation, a broken chromosome's increased motility and ability to explore a larger volume of the nucleus (Ira and Hastings, 2012), could possibly explain the high frequency of Chr7 deletions. It could be that this region of the genome is less constrained than other chromosomes, therefore when DSBs occurred on Chr7, the resulting DNA ends explored a larger area of the nucleus, allowing it to come in close proximity to the CNV reporter wherever it

was inserted. This would make a broken Chr7R more able to find a non-allelic repair template, and less likely to engage a nearby sister chromatid for allelic repair.

One final formal possibility is that the bias towards Chr7 deletions is not related to any molecular recombination mechanisms, but instead is due to viability of the deletion itself. In this case Chr7 deletions would be benign in terms of their effect on cell viability (or even somehow beneficial), thus resulting in a higher recovery of clone with Chr7 deletions. In contrast, deletions in other chromosomes would cause the cells carrying them to have comparatively reduced growth rates. While possible, we do not favor this model since it seems unlikely that Chr7's would be the only deletion event benign in phenotype. Other chromosomes such as Chr1, Chr3, Chr8, and Chr10 have been shown to be stably maintained as monosomes (loss of a full copy) in yeast diploids, therefore we would expect to recover a wider assortment of chromosomal deletions than what we actually observed.

While some of the hypotheses above are more attractive than others, we do not have sufficient experimental evidence at this point to directly support or reject any of them. However, all are testable and should provide an interesting avenue of investigation going forward. The search for the reasons behind the high prevalence of Chr7R deletions will be very stimulating and informative, regardless of what the final answer may turn out to be.

CHAPTER 5

CONCLUSION

This study established a standard assay, using the FA and Cu reporter system, to study CNV formation in diploid yeast cells. The assay allowed for the determination of mutation rates and mutation mechanisms associated with CNV formation, and was effective in detecting these events when the reporter cassette was inserted on different contexts in the genome. The exposures of cells to three known mutagens: HU, CPT, and MMS indicated that the assay is able to detect stimulation of CNV events by environmental exposures, even at relatively low toxicity levels.

The verification of the assay's ability to detect and classify CNV events arising from known mutagens, that this study provided, is promising for the prospect of using this assay in future studies of emerging environmental contaminants, and their influence in CNV formation. The relatively recent recognition of CNVs being a source of sporadic diseases in humans, and the continual increase of these sporadic diseases, such as Autism Spectrum Disorders, shows the importance of working to understand the occurrence of CNV events and investigating how environmental pollutants influence the rate and mechanisms involved with these mutations.

REFERENCES

Abrahams, B.S., and Geschwind, D.H. (2008). Advances in autism genetics: on the threshold of a new neurobiology. Nature reviews Genetics *9*, 341-355.

Argueso, J.L., Westmoreland, J., Mieczkowski, P.A., Gawel, M., Petes, T.D., and Resnick, M.A. (2008). Double-strand breaks associated with repetitive DNA can reshape the genome. Proceedings of the National Academy of Sciences of the United States of America *105*, 11845-11850.

Arlt, M.F., Ozdemir, A.C., Birkeland, S.R., Wilson, T.E., and Glover, T.W. (2011). Hydroxyurea induces de novo copy number variants in human cells. Proceedings of the National Academy of Sciences of the United States of America *108*, 17360-17365.

Boffetta, P. (2006). Human cancer from environmental pollutants: the epidemiological evidence. Mutation research *608*, 157-162.

Bruder, C.E., Piotrowski, A., Gijsbers, A.A., Andersson, R., Erickson, S., Diaz de Stahl, T., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A.*, et al.* (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. American journal of human genetics *82*, 763-771.

Chan, J.E., and Kolodner, R.D. (2011). A genetic and structural study of genome rearrangements mediated by high copy repeat Ty1 elements. PLoS genetics *7*, e1002089.

Chan, J.E., and Kolodner, R.D. (2012). Rapid analysis of Saccharomyces cerevisiae genome rearrangements by multiplex ligation-dependent probe amplification. PLoS genetics *8*, e1002539.

Chen, C., and Kolodner, R.D. (1999). Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. Nature genetics *23*, 81-85.

Dion, V., Kalck, V., Horigome, C., Towbin, B.D., and Gasser, S.M. (2012). Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. Nature cell biology *14*, 502-509.

Dorsey, M., Peterson, C., Bray, K., and Paquin, C.E. (1992). Spontaneous amplification of the ADH4 gene in Saccharomyces cerevisiae. Genetics *132*, 943-950.

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.

Durkin, S.G., and Glover, T.W. (2007). Chromosome fragile sites. Annual review of genetics *41*, 169-192.

Ewart, A.K., Morris, C.A., Atkinson, D., Jin, W., Sternes, K., Spallone, P., Stock, A.D., Leppert, M., and Keating, M.T. (1993). Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. Nature genetics *5*, 11-16.

Feuk, L., Marshall, C.R., Wintle, R.F., and Scherer, S.W. (2006). Structural variants: changing the landscape of chromosomes and design of disease studies. Human molecular genetics *15 Spec No 1*, R57-66.

Fink, G.R., and Styles, C.A. (1974). Gene conversion of deletions in the his4 region of yeast. Genetics *77*, 231-244.

Girirajan, S., Campbell, C.D., and Eichler, E.E. (2011). Human copy number variation and complex genetic disease. Annual review of genetics *45*, 203-226.

Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F.*, et al.* (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature *485*, 363-367.
Hakim, O., and Misteli, T. (2012). SnapShot: Chromosome confirmation capture. Cell *148*, 1068 e1061-1062.
Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS genetics *5*, e1000327.
Hoang, M.L., Tan, F.J., Lai, D.C., Celniker, S.E., Hoskins, R.A., Dunham, M.J., Zheng, Y., and Koshland, D. (2010). Competitive repair by naturally dispersed repetitive DNA during non-allelic homologous recombination. PLoS genetics *6*, e1001228.
Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nature genetics *36*, 949-951.
Ira, G., and Hastings, P.J. (2012). DNA breakage drives nuclear search. Nature cell biology *14*, 448-450.
Itsara, A., Wu, H., Smith, J.D., Nickerson, D.A., Romieu, I., London, S.J., and Eichler, E.E. (2010). De novo rates and selection of large copy number variation. Genome research *20*, 1469-1481.
Kondrashov, A.S. (2003). Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Human mutation *21*, 12-27.
Koszul, R., Caburet, S., Dujon, B., and Fischer, G. (2004). Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. The EMBO journal *23*, 234-243.
Koszul, R., Dujon, B., and Fischer, G. (2006). Stability of large segmental duplications in the yeast genome. Genetics *172*, 2211-2222.
Kumari, A., Lim, Y.X., Newell, A.H., Olson, S.B., and McCullough, A.K. (2012). Formaldehyde-induced genome instability is suppressed by an XPF-dependent pathway. DNA repair *11*, 236-246.
Lam, K.W., and Jeffreys, A.J. (2006). Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion. Proceedings of the National Academy of Sciences of the United States of America *103*, 8921-8927.
Lam, K.W., and Jeffreys, A.J. (2007). Processes of de novo duplication of human alpha-globin genes. Proceedings of the National Academy of Sciences of the United States of America *104*, 10950-10955.
Lea, D.E., and Coulson, C.A. (1949). The distribution of the numbers of mutants in bacterial populations. Journal of Genetics *49*, 264-285.
Ledbetter, D.H., Riccardi, V.M., Airhart, S.D., Strobel, R.J., Keenan, B.S., and Crawford, J.D. (1981). Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. The New England journal of medicine *304*, 325-329.
Lee, J.A., Carvalho, C.M., and Lupski, J.R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell *131*, 1235-1247.

Lemoine, F.J., Degtyareva, N.P., Lobachev, K., and Petes, T.D. (2005). Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. Cell *120*, 587-598.
Liebman, S., Shalit, P., and Picologlou, S. (1981). Ty elements are involved in the formation of deletions in DEL1 strains of Saccharomyces cerevisiae. Cell *26*, 401-409.
Liebman, S.W., Singh, A., and Sherman, F. (1979). A mutator affecting the region of the iso-1-cytochrome c gene in yeast. Genetics *92*, 783-802.
Lupski, J.R. (2007). Genomic rearrangements and sporadic disease. Nature genetics *39*, S43-47.
Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A.*, et al.* (1991). DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell *66*, 219-232.
Macosko, E.Z., and McCarroll, S.A. (2012). Exploring the variation within. Nature genetics *44*, 614-616.
Malhotra, D., and Sebat, J. (2012a). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell *148*, 1223-1241.
Malhotra, D., and Sebat, J. (2012b). Genetics: Fish heads and human disease. Nature *485*, 318-319.
McCulley, J.L., and Petes, T.D. (2010). Chromosome rearrangements and aneuploidy in yeast strains lacking both Tel1p and Mec1p reflect deficiencies in two different mechanisms. Proceedings of the National Academy of Sciences of the United States of America *107*, 11465-11470.
Meyerson, M., Gabriel, S., and Getz, G. (2010). Advances in understanding cancer genomes through second-generation sequencing. Nature reviews Genetics *11*, 685-696.
Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K.*, et al.* (2011). Mapping copy number variation by population-scale genome sequencing. Nature *470*, 59-65.
Mine-Hattab, J., and Rothstein, R. (2012). Increased chromosome mobility facilitates homology search during recombination. Nature cell biology *14*, 510-517.
Morrison, A., Bell, J.B., Kunkel, T.A., and Sugino, A. (1991). Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity. Proceedings of the National Academy of Sciences of the United States of America *88*, 9473-9477.
Myung, K., and Kolodner, R.D. (2002). Suppression of genome instability by redundant S-phase checkpoint pathways in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America *99*, 4500-4507.
Nachman, M.W., and Crowell, S.L. (2000). Estimate of the mutation rate per nucleotide in humans. Genetics *156*, 297-304.
Paques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiology and molecular biology reviews : MMBR *63*, 349-404.
Payen, C., Koszul, R., Dujon, B., and Fischer, G. (2008). Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. PLoS genetics *4*, e1000175.
Piotrowski, A., Bruder, C.E., Andersson, R., Diaz de Stahl, T., Menzel, U., Sandgren, J., Poplawski, A., von Tell, D., Crasto, C., Bogdan, A.*, et al.* (2008). Somatic mosaicism for copy number variation in differentiated human tissues. Human mutation *29*, 1118-1124.

Putnam, C.D., Hayes, T.K., and Kolodner, R.D. (2009). Specific pathways prevent duplication-mediated genome rearrangements. Nature *460*, 984-989.
Richardson, S.D. (2008). Environmental mass spectrometry: emerging contaminants and current issues. Analytical chemistry *80*, 4373-4402.
Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M.*, et al.* (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science *328*, 636-639.
Rothstein, R., Helms, C., and Rosenberg, N. (1987). Concerted deletions and inversions are caused by mitotic recombination between delta sequences in Saccharomyces cerevisiae. Molecular and cellular biology *7*, 1198-1207.
Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J.*, et al.* (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445-449.
Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M.*, et al.* (2004). Large-scale copy number polymorphism in the human genome. Science *305*, 525-528.
Sherman, F., Stewart, J.W., Jackson, M., Gilmore, R.A., and Parker, J.H. (1974). Mutants of yeast defective in iso-1-cytochrome c. Genetics *77*, 255-284.
Tanizawa, H., and Noma, K. (2012). Unravelling global genome organization by 3C-seq. Seminars in cell & developmental biology *23*, 213-221.
Toffolatti, L., Cardazzo, B., Nobile, C., Danieli, G.A., Gualandi, F., Muntoni, F., Abbs, S., Zanetti, P., Angelini, C., Ferlini, A.*, et al.* (2002). Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. Genomics *80*, 523-530.
Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916-924.
Turner, D.J., Miretti, M., Rajan, D., Fiegler, H., Carter, N.P., Blayney, M.L., Beck, S., and Hurles, M.E. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. Nature genetics *40*, 90-95.
Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast *10*, 1793-1808.
Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. Annual review of genomics and human genetics *10*, 451-481.
Zhang, H., Zeidler, A.F.B., Song, W., Puccia, C., Malc, E., Greenwell, P.W., Mieczkowski, P.A., Petes, T.D., and Argueso, J.L. (2012a). Gene copy number variation (CNV) in haploid and diploid strains of the yeast Saccharomyces cerevisiae. Genetics, in preparation.
Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012b). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908-921.
Zhu, J., Pavelka, N., Bradford, W.D., Rancati, G., and Li, R. (2012). Karyotypic determinants of chromosome instability in aneuploid budding yeast. PLoS genetics *8*, e1002719.

APPENDIX 1

Array-CGH analysis

APPENDIX Ia. Conversion charts for CGH-array probe levels.

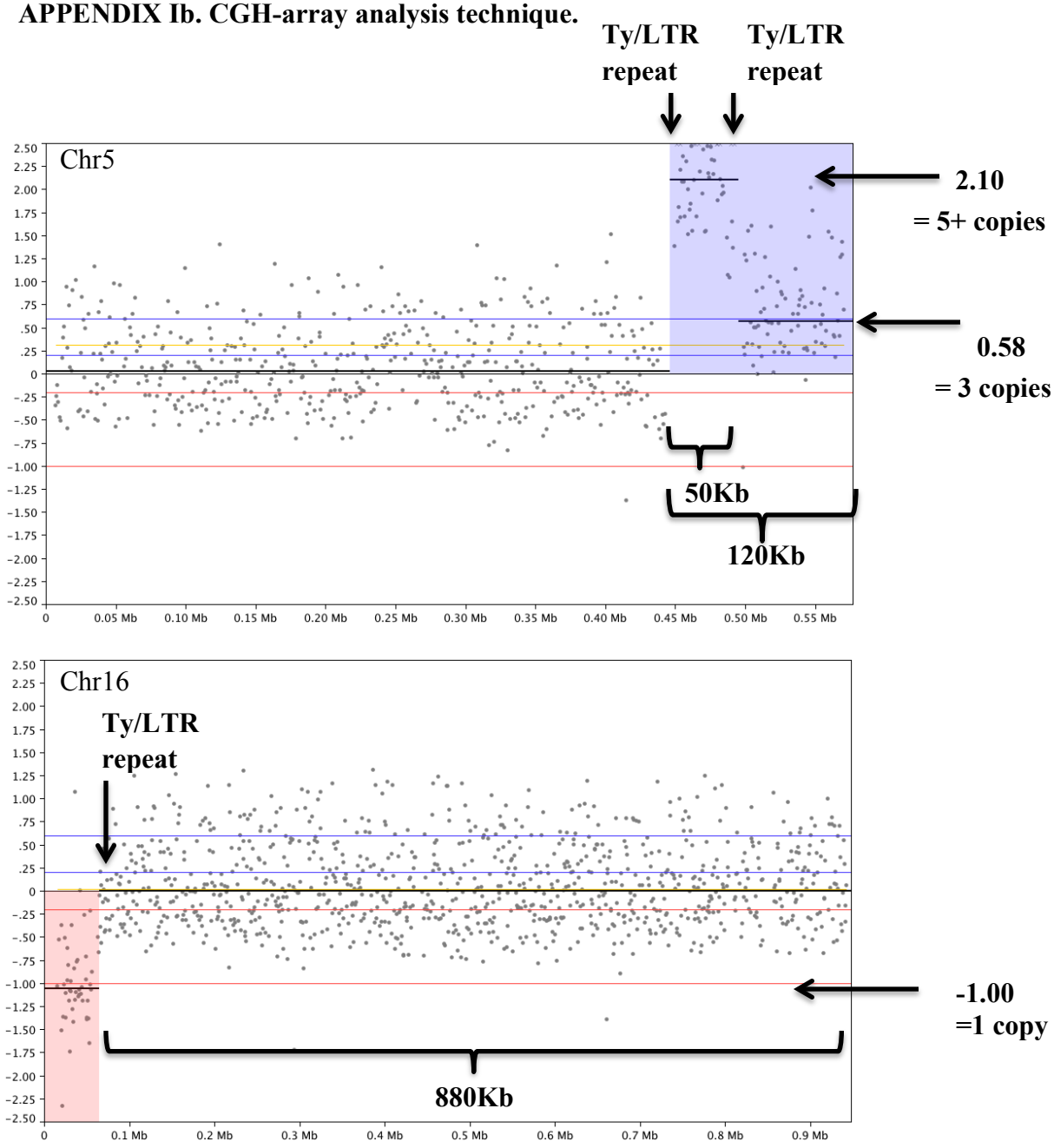
SFA1 and CUP1 copy number in:		R/G ratio	Log2(R/G)
FCR strain	Parent strain		
1	1	1	0.00
2	1	2	1.00
3	1	3	1.58
4	1	4	2.00
5	1	5	2.32
6	1	6	2.58
7	1	7	2.81

Other genomic segment copy number in:		R/G ratio	Log2(R/G)
FCR strain	Diploid parent strain		
1	2	0.5	-1.00
2	2	1.0	0.00
3	2	1.5	0.58
4	2	2.0	1.00
5	2	2.5	1.32
6	2	3.0	1.58
7	2	3.5	1.81

Other genomic segment copy number in:		R/G ratio	Log2(R/G)
FCR strain	Haploid parent strain		
1	1	1	0.00
2	1	2	1.00
3	1	3	1.58
4	1	4	2.00
5	1	5	2.32
6	1	6	2.58
7	1	7	2.81

Log2 tables (FCR/Parent) were used to convert the average signal of genomic segments on the CGH-arrays to determine the number of copies present of the reporter, CNV events in diploid strains, and CNV events in haploid strains, respectively.

APPENDIX Ib. CGH-array analysis technique.

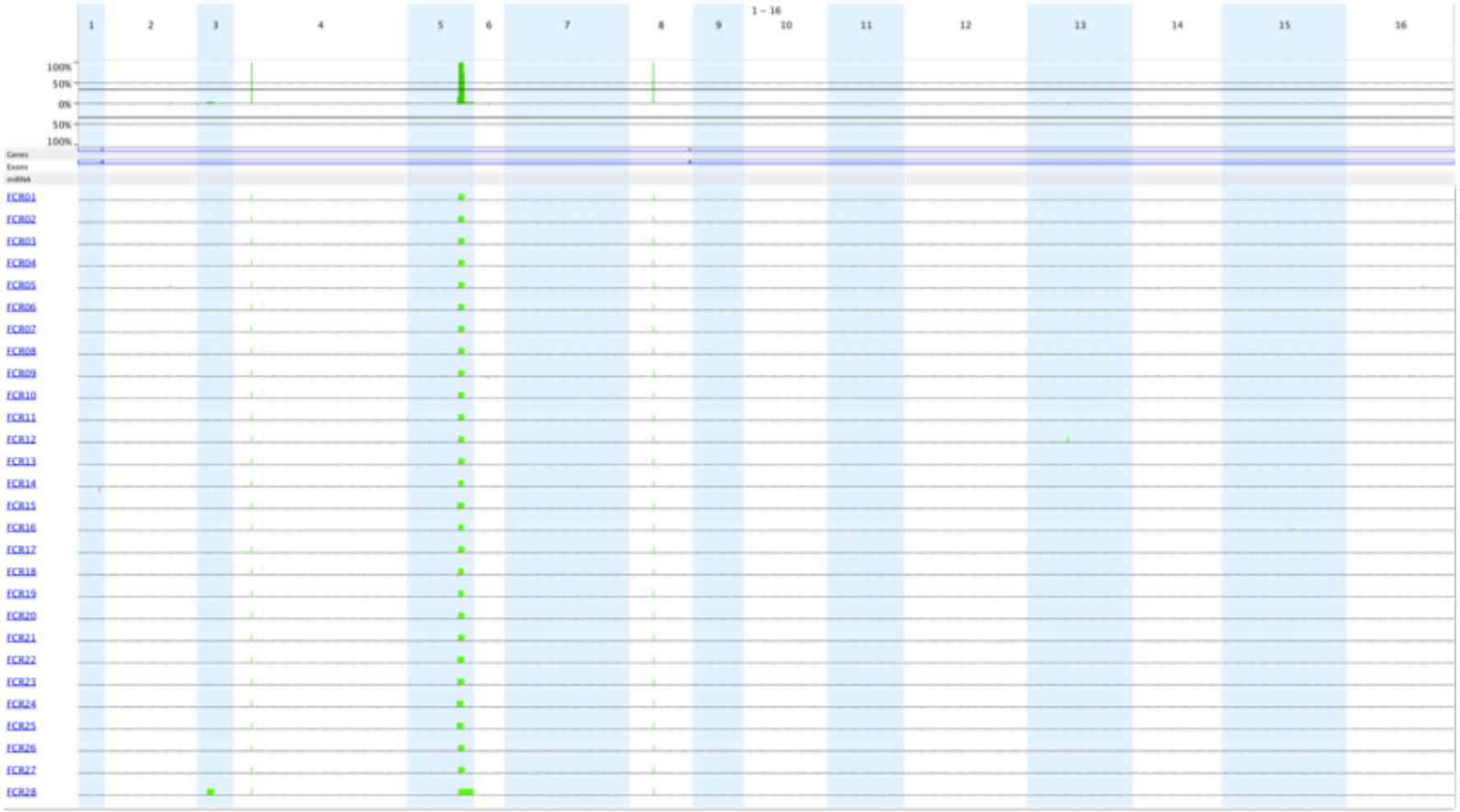


CGH-array data was analyzed by reading the average signal, as seen on Chr5 above as 2.10 and 0.58, and Chr16 as -1.00 and the Log2 tables shown in Appendix I were used to convert the readings to the number of copies present. The size of the CNV events was noted, the segment size for amplification events is shown on Chr5 and for deletions shown on Chr16. It was also determined if the breakpoint regions contain repetitive DNA elements (as were the case in this example), or single copy sequences.

APPENDIX II

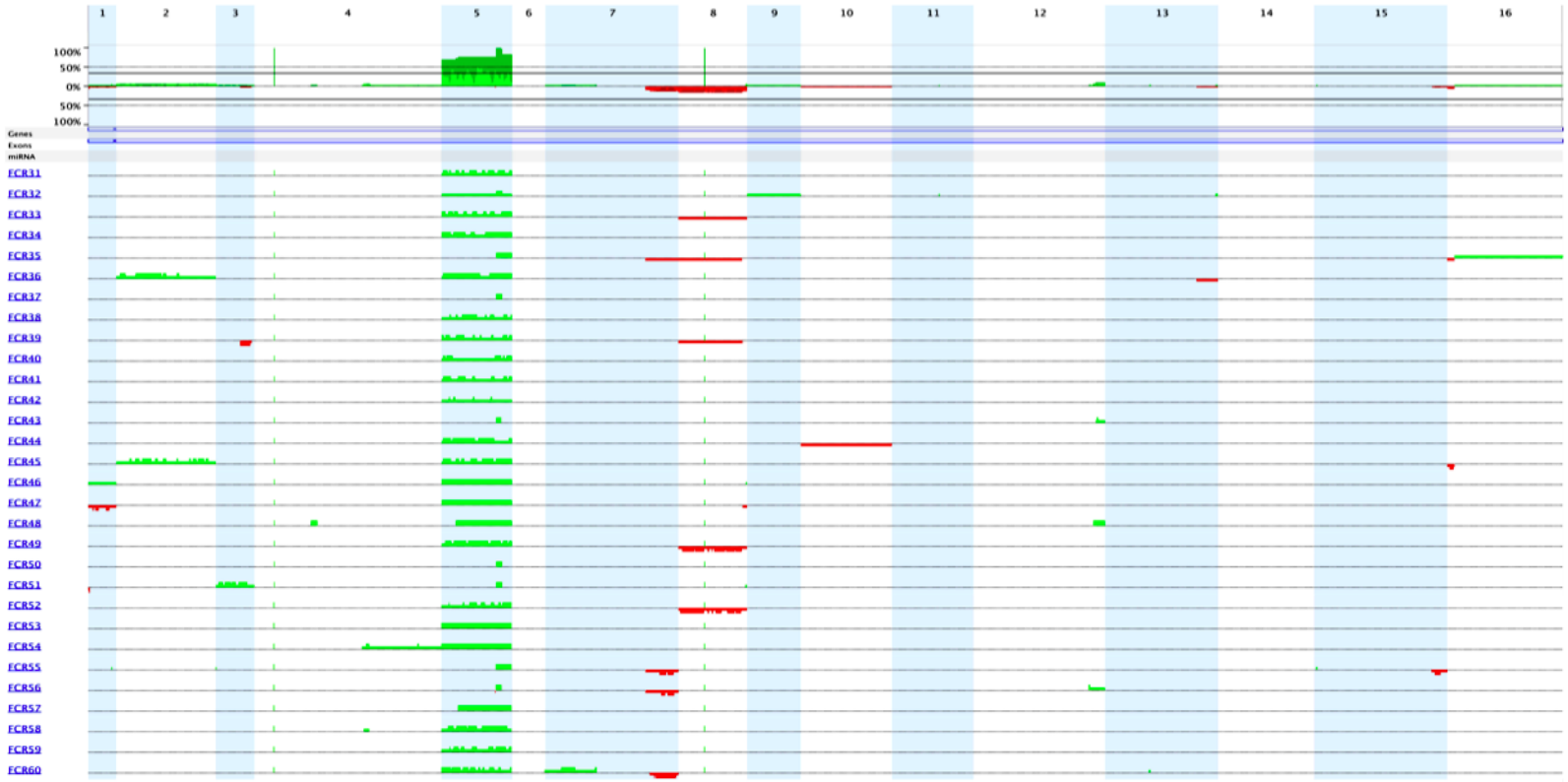
Global views of CNVs

APPENDIX IIa. Global view of CNV across all tested uninduced Chr5 haploids



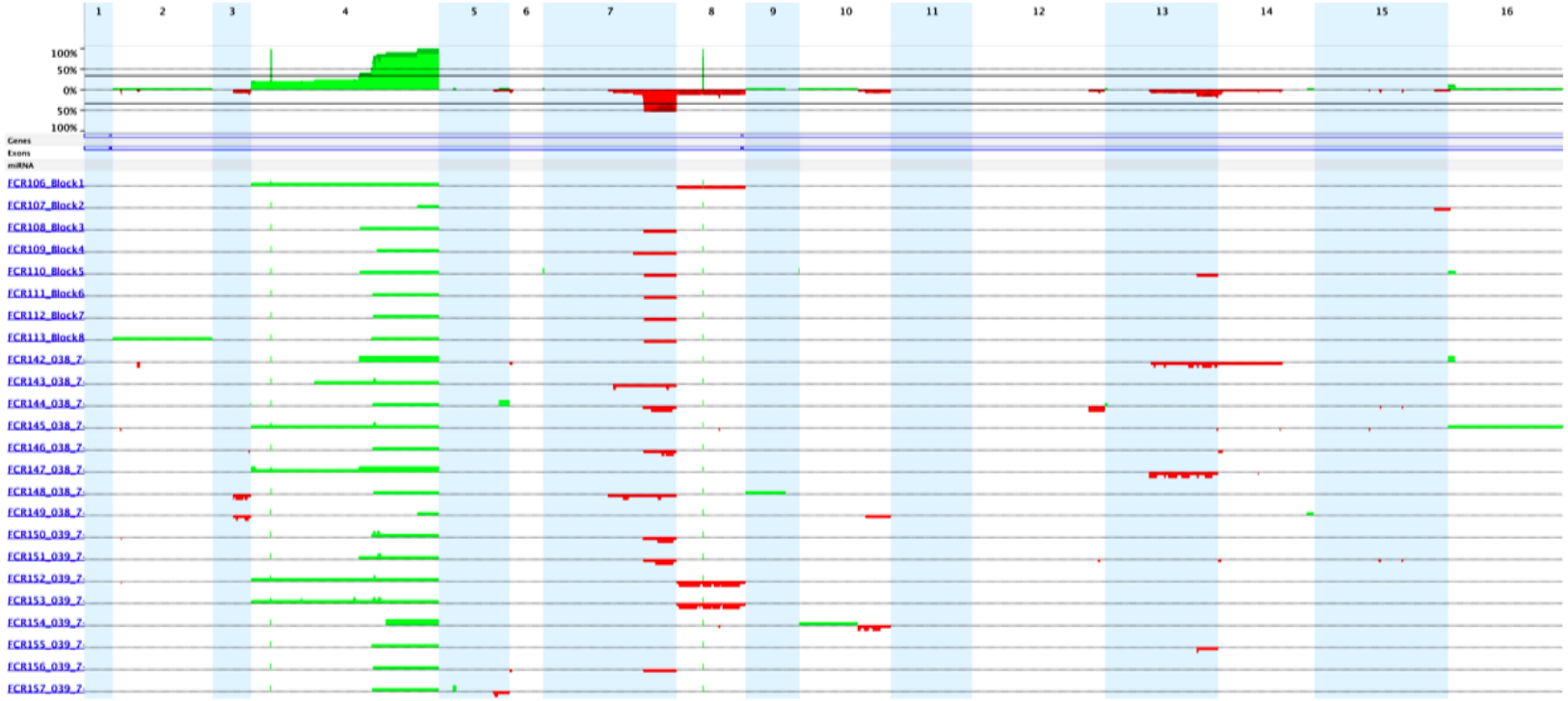
Above is a global view of all of the samples in the studied unexposed Chr5 haploid data set. Amplifications are presented in green and deletions in red. The green spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Chr5 haploids only saw a segmental duplications occurring to amplify the reporter inserted in that region.

Appendix IIb. Global view of CNV across all tested uninduced Chr5 diploids



Above is a global view of all of the samples in the studied uninduced Chr5 diploid data set. Amplifications are presented in green and deletions in red. The green spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Chr5 diploids had deletion events occur and the major method of reporter amplification was aneuploidy.

Appendix IIc. Global view of CNV across all tested uninduced Chr4 diploids

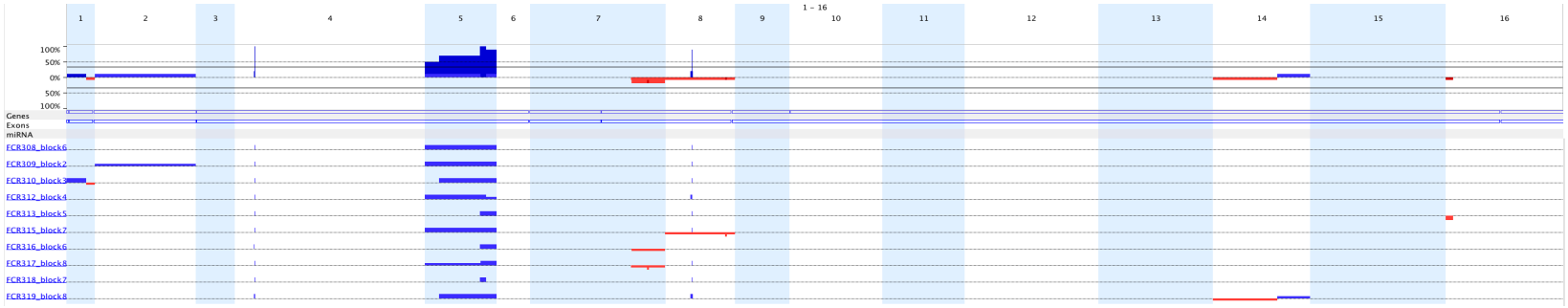


Above is a global view of all of the samples in the studied unexposed Chr4 diploid data set. Amplifications are presented in green and deletions in red. The green spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. A large number of deletions occurred on the right arm of Chr7 and the majority of reporter amplifications occurred through translocations, and a few aneuploidy events.

Appendix IId. Global view of CNV across all tested uninduced Chr15 diploids

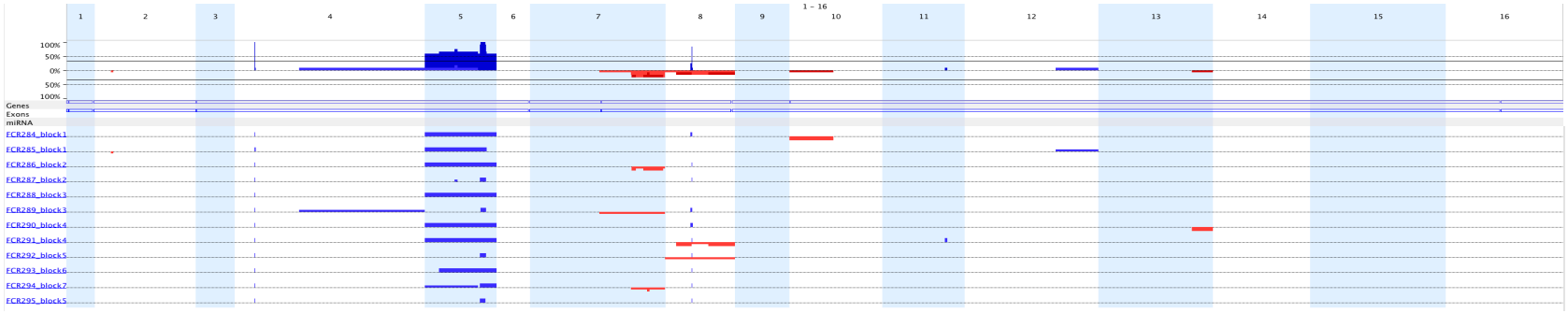
Above is a global view of all of the samples in the studied unexposed Chr15 diploid data set. Amplifications are presented in green and deletions in red. The green spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. The reporter was amplified almost exclusively by translocation events.

Appendix IIe. Global view of CNV across all tested Chr5 diploids exposed to 15 µg/ml of CPT



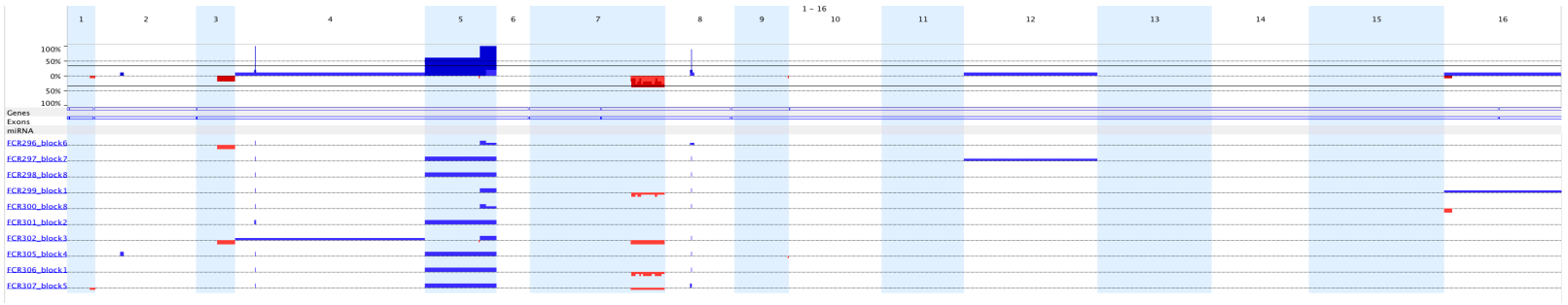
Above is a global view of all of the samples in the studied CPT exposed Chr5 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. A large number of Chr5 aneuploidy events occurred with a few translocations and one segmental duplication.

Appendix IIf. Global view of CNV across all tested Chr5 diploids exposed to 50 mM of HU



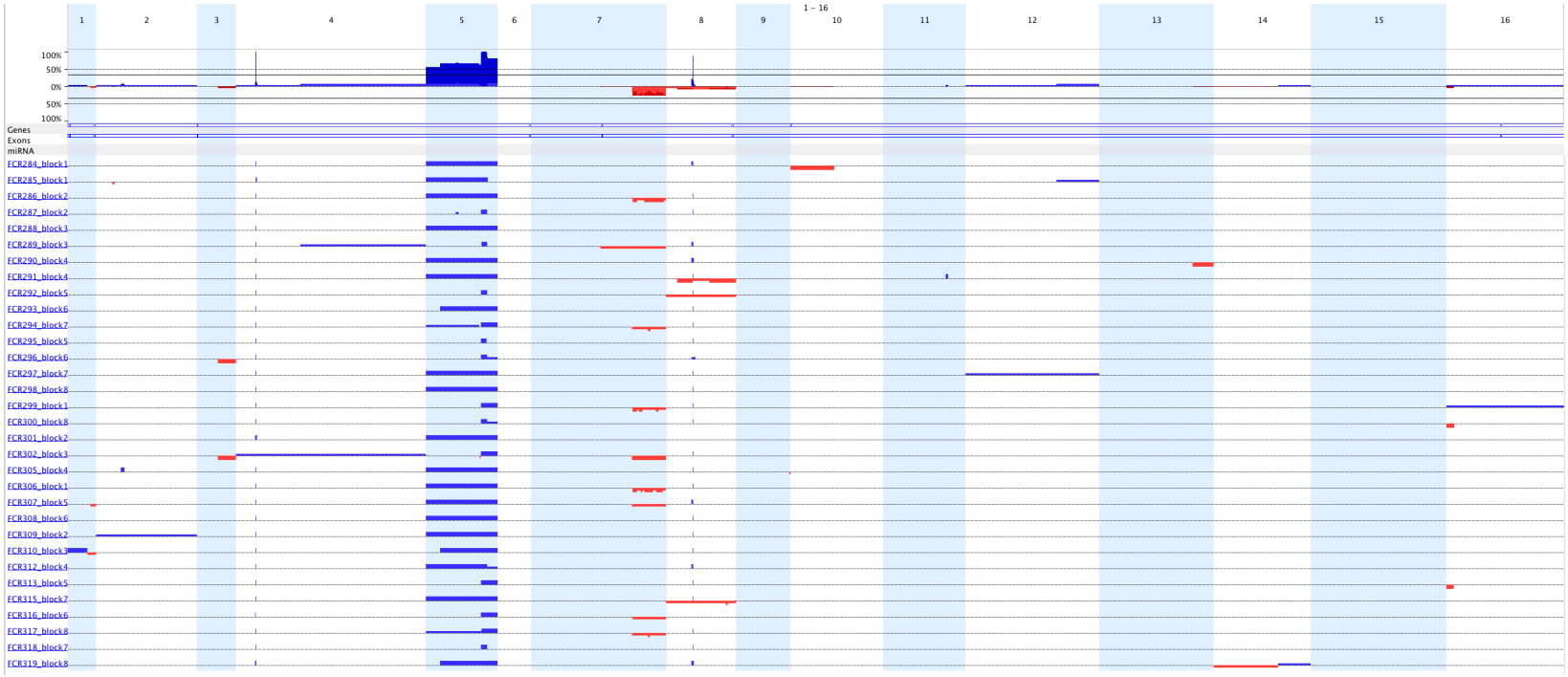
Above is a global view of all of the samples in the studied HU exposed Chr5 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Aneuploidy events are the most common method in amplifying the reporter, with segmental duplications being the method in four of the samples.

Appendix IIg. Global view of CNV across all tested Chr5 diploids exposed to 35 µg/ml of MMS



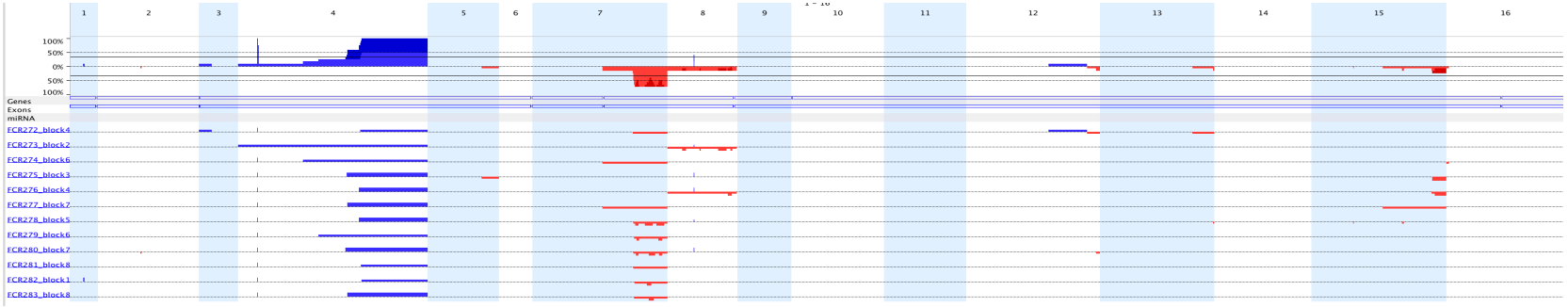
Above is a global view of all of the samples in the studied MMS exposed Chr5 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Aneuploidy events and translocations were the mechanisms amplifying the reporter.

Appendix IIh. Global view of CNV across all tested induced Chr5 diploids



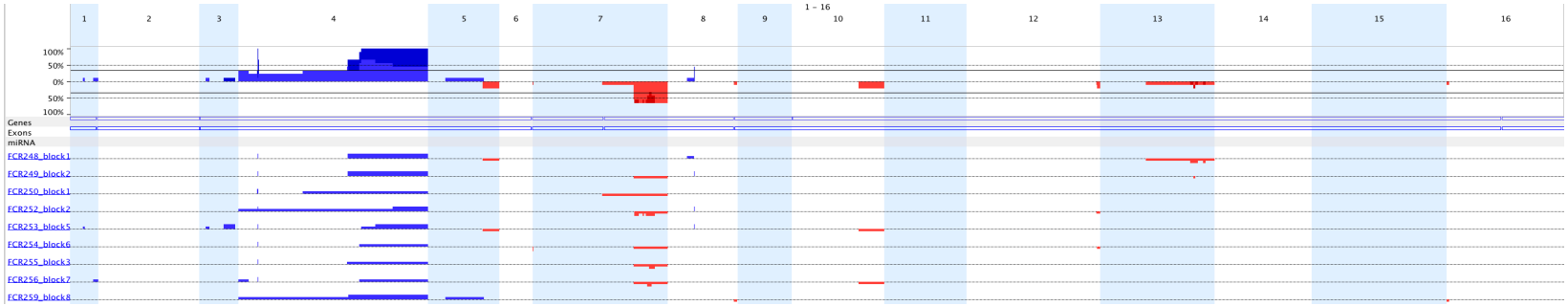
Above is a global view of all of the Chr5 diploid samples exposed to HU, CPT, or MMS. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Most reporter amplifications occur through aneuploidy of Chr5. There are a frequent number of deletions occurring on the right arm of Chr7.

Appendix IIi. Global view of CNV across all tested Chr4 diploids exposed to 15 µg/ml of CPT



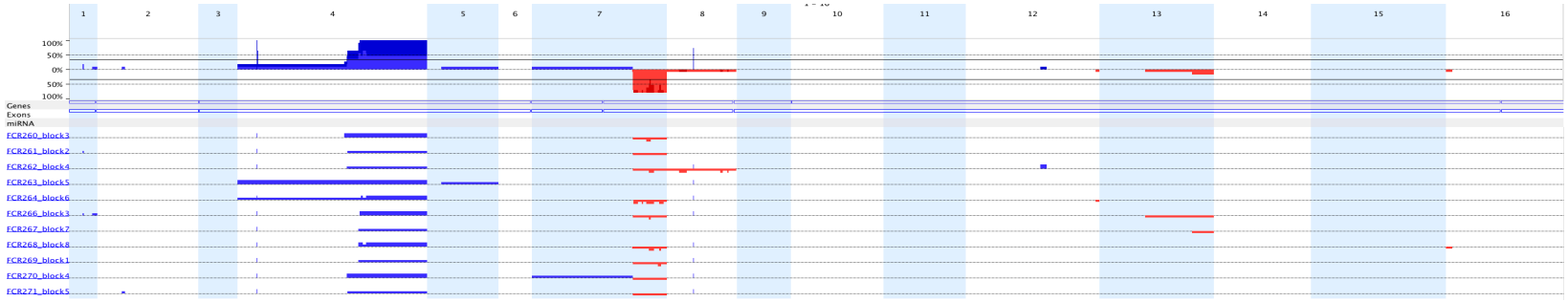
Above is a global view of all of the samples in the studied CPT exposed Chr4 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Majority of reporter amplifications are due to translocations, and a significant number of deletion events occurred on the right arm of Chr7.

Appendix IIj. Global view of CNV across all tested Chr4 diploids exposed to 50 mM of HU



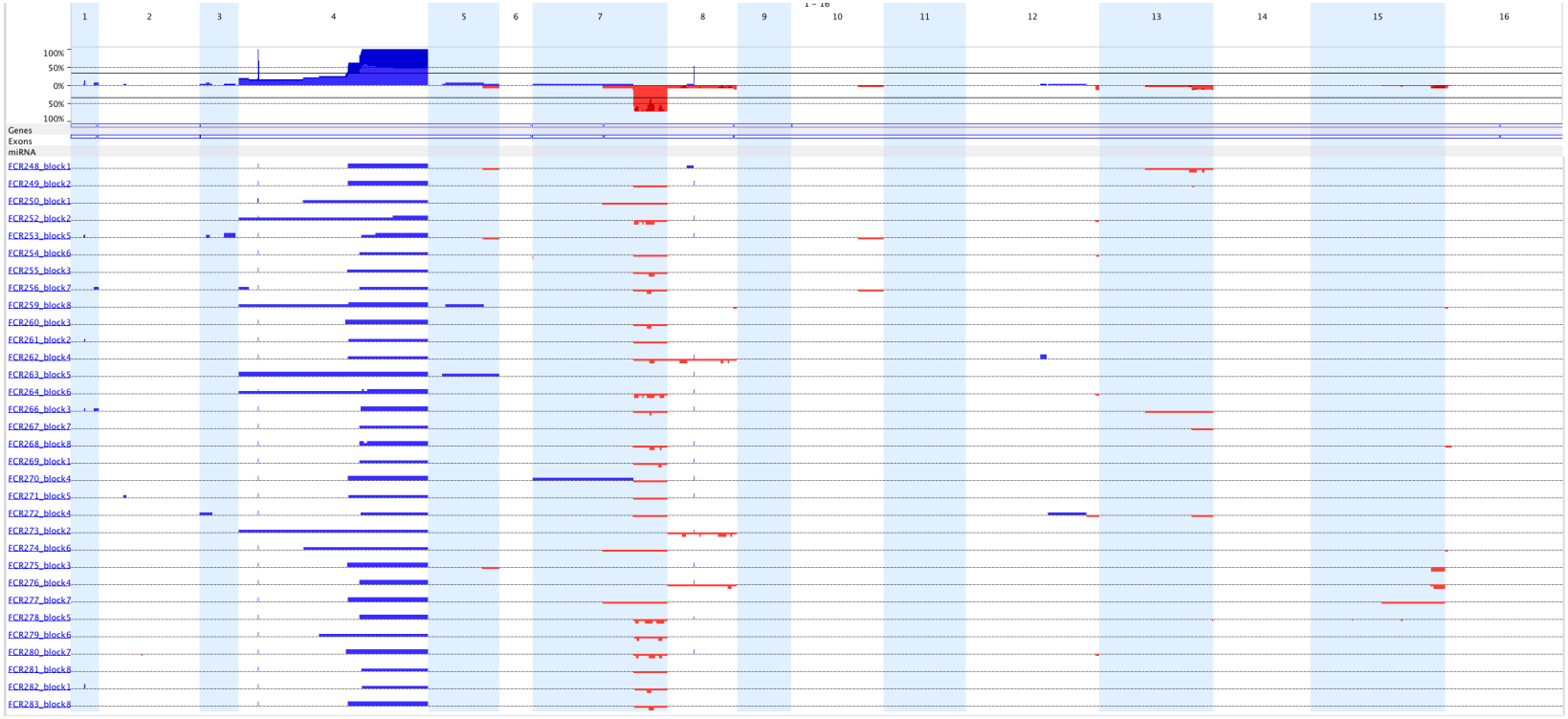
Above is a global view of all of the samples in the studied HU exposed Chr4 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Majority of reporter amplifications are due to translocations, two Chr4 aneuploidy events were responsible as well, and a significant number of deletion events occurred on the right arm of Chr7.

Appendix IIk. Global view of CNV across all tested Chr4 diploids exposed to 35 µg/ml of MMS



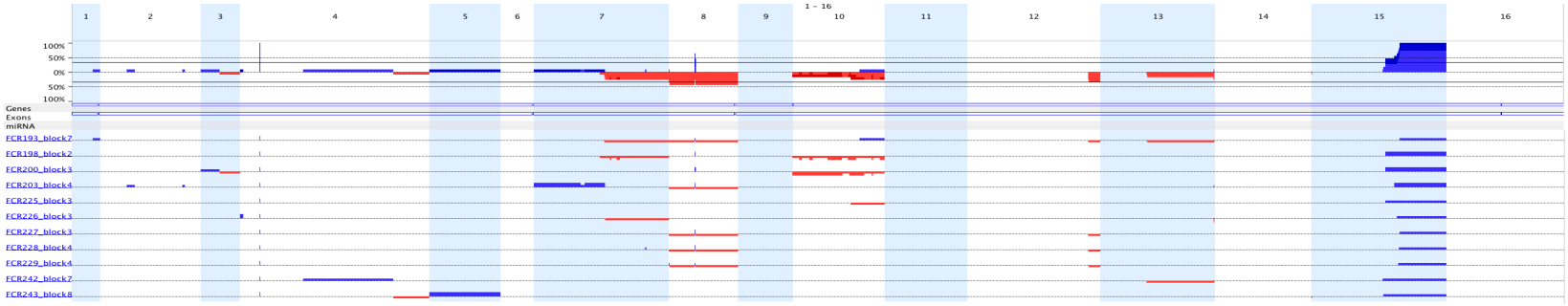
Above is a global view of all of the samples in the studied MMS exposed Chr4 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Majority of reporter amplifications are due to translocations, and a significant number of deletion events occurred on the right arm of Chr7.

Appendix III. Global view of CNV across all tested induced Chr4 diploids



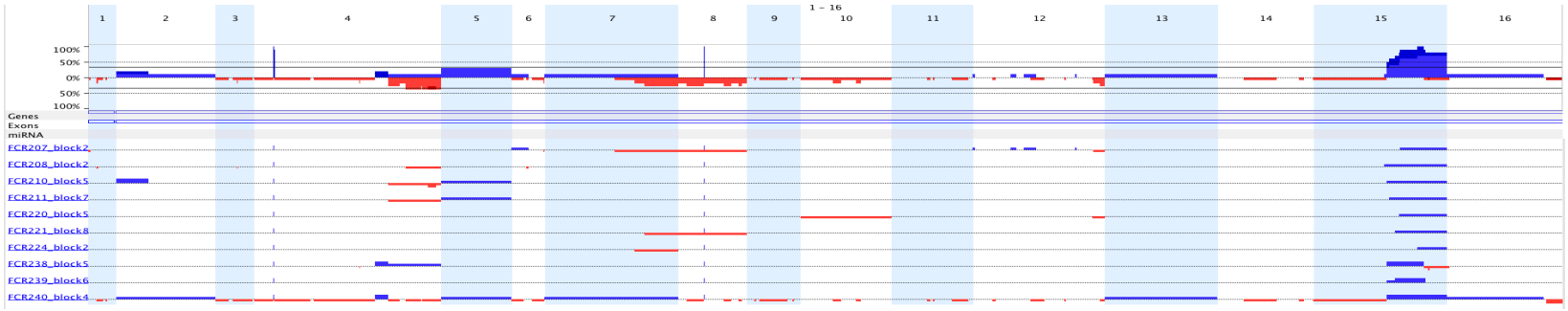
Above is a global view of all of the Chr4 diploid samples exposed to HU, CPT, or MMS. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Most reporter amplifications occur through terminal translocations involving Chr5. There are a frequent number of deletions occurring on the right arm of Chr7.

Appendix IIm. Global view of CNV across all tested Chr15 diploids exposed to 15 µg/ml of CPT



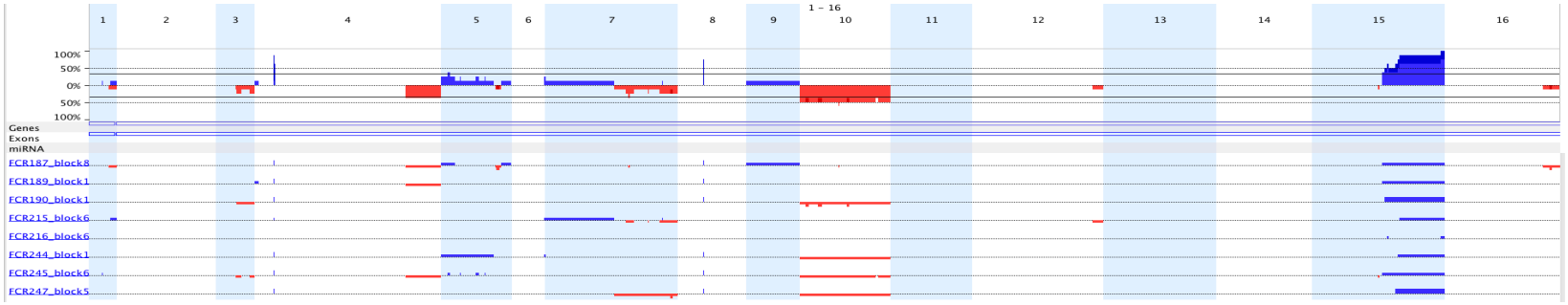
Above is a global view of all of the samples in the studied CPT exposed Chr15 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. All of reporter amplifications are due to translocations.

Appendix IIn. Global view of CNV across all tested Chr15 diploids exposed to 50 mM of HU



Above is a global view of all of the samples in the studied HU exposed Chr15 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Majority of reporter amplifications are due to translocations, amplification events in two samples are due to segmental duplications.

Appendix IIo. Global view of CNV across all tested Chr15 diploids exposed to 35 µg/ml of MMS



Above is a global view of all of the samples in the studied MMS exposed Chr15 diploid data set. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. All of reporter amplifications are due to terminal translocations.

Appendix IIp. Global view of CNV across all tested induced Chr15 diploids

Above is a global view of all of the Chr15 diploid samples exposed to HU, CPT, or MMS. Amplifications are presented in blue and deletions in red. The blue spikes that appear on Chr4 and Chr8 signal for the *SFA1*and *CUP1* genes, respectively. These peaks were used to estimate the number of copies of the reporter were present in each sample. Nearly all of the reporter amplifications occur through terminal translocations. No aneuploidy events occur on Chr15 to cause reporter amplification. There is a lot more events dispersed throughout the genome.

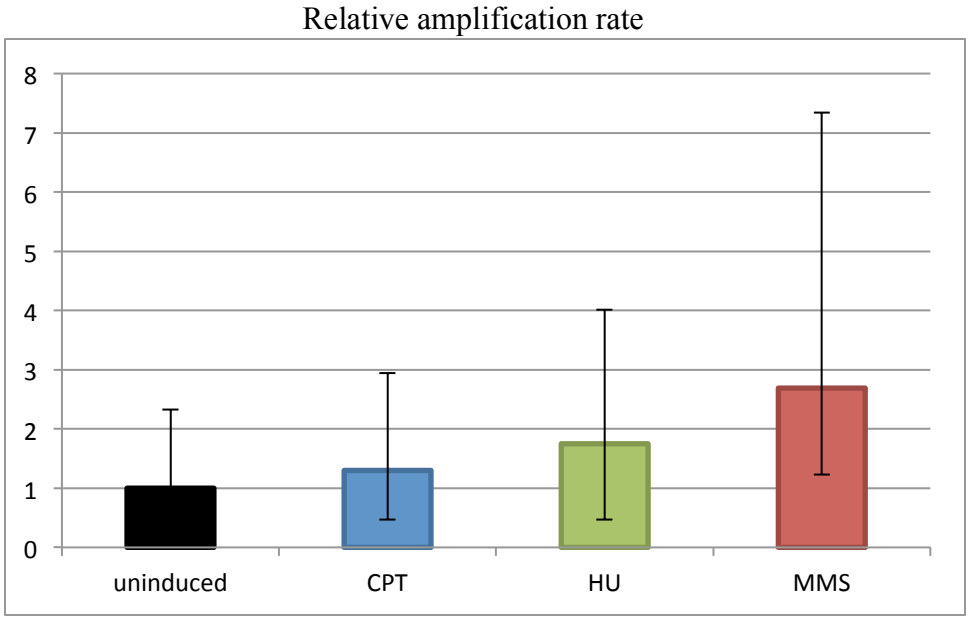
APPENDIX III

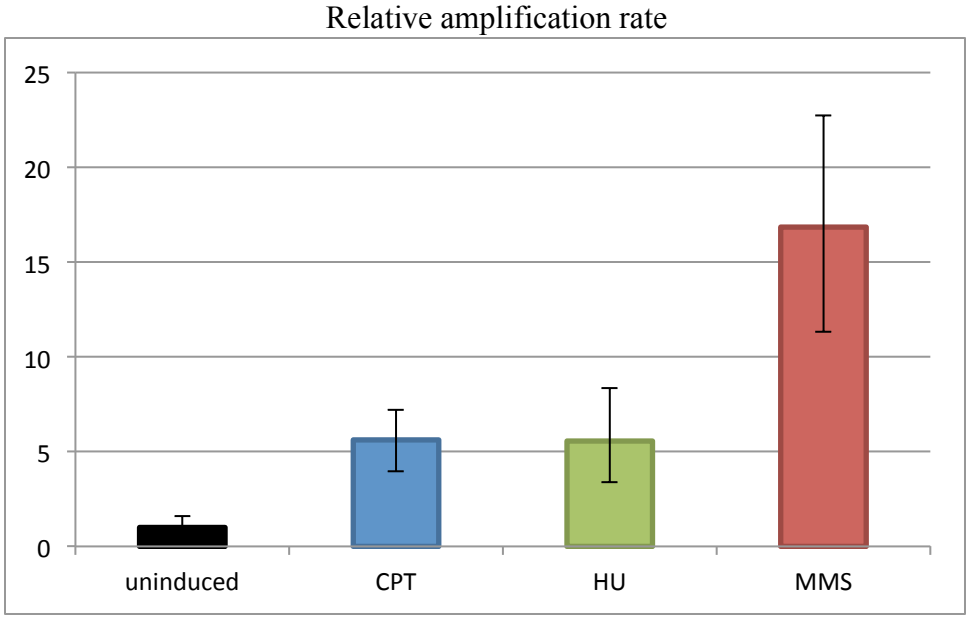
Initial development of CNV assay

APPENDIX IIIa. Assay development

A significant amount of work developing the CNV assay was done previously to my contributions and alterations, and will be explained in a paper entitled *Gene copy number variation (CNV) in haploid and diploid strains of the yeast Saccharomyces cerevisiae* (Zhang et al., 2012a). When researching diploid strain exposures began, the assay consisted of plating 350 µl of 10^-5 dilution of the sample on one plate containing 1.8 mM FA and 150 µM CuSo4, and 300 µl of 10^-1 dilution on a plate without Cu or FA. And then counting the cells, as done in the assay used to obtain the results discussed, on days three, four and five. This method had been successful when studying uninduced CNV events, but when induced by mutagens, resulted in a plague of false-positive colonies to form. The resulting data, part of which is shown in APPENDIX IIIb, contained large error bars, fluctuating results, and overall was unreliable. Thinking that perhaps the FA had been breaking down, new stock solutions of CuSO4 and FA were made or bought, but the results were the same. Other small alterations to the assay were made, such as adding FA and CuSO4 to the autoclaved media at 70°C to provide consistency, and changing the volume of the samples plated on each plate to 150 µl and then finally to 100 µl, as well as plating 10^-4 dilution on the non-permissive plate. It was predicted that there is a maximum capacity regarding cell density that could be plated, therefore the final change to the assay was adding an additional three non-permissive plates. The assay resulting in 100 µl of 10^-1 dilution plated on one permissive plate, and 100 µl, of 10^-4 dilution plated on four non-permissive plates, provided data that was consistently reliable, with a high positive re-test rate, and no observable background growth on the plates.

APPENDIX IIIb. Assay development data





Represented above is the relative amplification rates (uninduced amplification rate set to the value of 1.00) of exposures to strain JAY685, which had the reporter inserted on Chr15. The top chart shows the earliest data set before alterations were made to the assay, and the bottom chart shows the final data set discussed in the results chapter.