File S1

SNP microarrays: specificity of hybridization

To confirm the specificity of the pattern of hybridization to the oligonucleotides, we isolated genomic DNA from the haploid strains PLS2/W303a and PSL5/YJM789. The DNA was labeled using either Cy3-dUTP (YJM789) or Cy5-dUTP (W303a) (LEMOINE *et al.*, 2005). The samples were mixed and competitively hybridized to the microarrays (details below). Our first experiments were done with arrays containing about 60,000 oligonucleotides (Table S3), representing about 15,000 SNP positions. Following scanning of the arrays, we measured the ratio of the signals at wavelengths specific for the Cy3- and Cy5-labeled samples (532 and 635 nm, respectively). The 635 nm/532 nm ratio was analyzed for each oligonucleotide. The average value of the median ratios of the control probes (those that were identical in the two haploid strains listed in Table S4) was calculated and used to normalize the ratios of the experimental probes to a value of 1 by dividing each probe ratio by the average control probe ratio. All W303a probes (designated by an "SF" or "SR" as the last two characters of the ProbeID in Table S3) that had a centered ratio less than 1.5 were discarded. All YJM789 probes (designated by an "YF" or "YR" as the last two characters of the ProbeID in Table S3) that had a centered ratio greater than 0.67 were discarded. These criteria require at least a 50% difference in signal between the two strains for any given probe. We then repeated this experiment, switching the dyes that were used to label the genomic DNA samples. Only probes that satisfied the criteria described above in both experiments were used in our analysis. The final probe set used for analysis is listed in Table S5.

Details of methods used for microarray analysis: sample preparation, hybridization conditions, and data analysis

Five ml YPD yeast cultures were grown at 30° overnight with agitation. Approximately 55 mg of the pelleted culture was resuspended in 500 ml of a melted agarose solution (0.5% low melt agarose, 100 mM EDTA pH 7.5 at approximately 42°), and then 20 μl of a 25 mg/ml Zymolase solution was added. This cell-agarose mixture was distributed into five plug molds and allowed to solidify. The plugs were then incubated in 700 μl of an EDTA/Tris solution (500 mM EDTA, 10 mM Tris, pH 7.5) overnight at 37°. The next day, 400 μl of a sarcosyl/proteinase K solution (5% sarcosyl, 5 mg/ml proteinase K, 500 mM EDTA pH 7.5) was added to the tubes containing the plugs, and they were incubated at 50° for five hours to overnight. DNA was isolated from three plugs per strain analyzed using the Fermentas Life Sciences GeneJet Gel Extraction Kit (#K0692). Following the addition of the "Binding" buffer from the kit, the samples were incubated at room temperature (~25°) until the agarose had melted and were then incubated on ice for 5 minutes. These samples were sonicated using a BioRupter sonicater for two 15-minute sessions of 30-second high pulses, followed by 30 seconds without sonication, resulting in DNA fragments of about 200-300 bp. Following sonication, the samples were treated according to the kit protocol.

We labeled two reactions for each experimental strain and a single reaction for the reference PG311 strain using the Invitrogen Bioprime Array CGH Genome Labeling Module using the kit protocol for all except for the last step (stop buffer). Experimental strains were labeled with Cy5-dUTP and PG311 was labeled with Cy3-dUTP. After incubation at 37°, all of the reactions intended for a single microarray (two reactions of the experimental strain and one reaction of the reference strain) were combined into a single tube and purified using the Fermentas Life Sciences GeneJet PCR Purification kit (#K0702) following the kit protocol but eluting with 79 μl of dH₂O.

The hybridization reactions were prepared using an Agilent Oligo aCGH/ChIP-on-Chip Hybridization kit (5188-5220) following kit instructions. Arrays were incubated for 48 hours at 62°. Following hybridization, the arrays were washed for 5 minutes in Oligo aCGH/ChIP-on-Chip Wash Buffer 1 (Agilent 5188-5221) and 1 minute in Oligo aCGH/ChIP-on-Chip Wash Buffer 2 (Agilent 5188-5222) that was pre-warmed to 37°. The arrays were then scanned at wavelengths of 635 and 532 nm using the GenePix scanner and the GenePix Pro software using settings recommended by the manufacturer.

Microarrays could be re-used approximately 4-6 times by removing the hybridized labeled DNA sequences from the oligonucleotides. Microarrays and gasket slides were stripped separately in 1x stripping buffer (10 mM potassium phosphate, pH6.6). The slides were slowly heated to the boiling point in the stripping buffer for 30-45 minutes. After stripping, they were transferred to deionized water, and then slowly removed and stored in a nitrogen cabinet. The gasket slides were centrifuged at 500 rpm to remove excess liquid. Labels on microarrays were removed prior to stripping.

The data generated by GenePix Pro were exported to text files and analyzed with Microsoft Excel. Probes that were flagged by the software were deleted from the analysis. The ratio of the medians (635 nm/532 nm) for each probe was used for analysis, and replicate probe medians were averaged. The data was centered around one by dividing each probe median by the average of all of the probe medians in order to normalize for differences in the hybridization levels for the reference and experimental strain samples. The data were plotted separately for each haploid parental strain and, in plots showing whole chromosomes, the medians were "smoothed" by averaging over nine consecutive probes.

Generation and analysis of high-throughput sequencing data

Prior to analysis of the experimental strains, a reference genome was compiled from the sequences of the two parental haploid strains of PG311,; this process required several steps. First, from S288c-YJM789 contig alignments (WEI *et al.* 2007), using in-house PERL scripts, we extracted the S288c and YJM789 sequences separately. The PSL2/W303a "reads" were then assembled using these S288c contigs and the PSL5/YJM789 "reads" were assembled using the YJM789 contigs; this analysis was performed using CLC Genomics software with the "random" flag option. From these assemblies, we used CLC Genomic software to generate the first-stage reference genomic sequence for PSL2/W303a and PSL5/YJM789. Second, we re-assembled the PSL2 and PSL5 "reads" using the first-stage reference sequences; CLC Genomics software was used for this process with the "unique" flag option. The resulting assemblies were used to generate a second-stage reference

sequence for PSL2/W303a and PSL5/YJM789. In the construction of the second-stage reference sequences for PSL2/W303a and PSL5/YJM789, we altered the base from the first-stage reference sequence if $\geq$75% of the "reads" in the new alignments had a base that was different from the first-stage reference sequence.

We then aligned the second-stage PSL2/W303a and PSL5/YJM789 contigs with the previously aligned S288c-YJM789 contig using the MAFFT program with the Sequence-to-skeleton alignment with the '-add' option (KATOH *et al.* 2009). The resulting alignments allowed us to translate between positions in the assemblies of the two paternal haploids using in-house PERL scripts. In addition, we used in-house PERL scripts to locate the positions of all of the approximately 55,000 SNPs that distinguished the two haploid strains.

Each of the irradiated diploid strains was independently assembled to the consensus sequences of both of PSL2 and PSL5 using CLC Genomics software and, in addition, the Burrows-Wheeler Alignment Tool (BWA, (LI and DURBIN 2009)). SAMtools was then used to extract "pileup" files out of the assembled files (LI *et al.* 2009), creating files that show the number of bases supported at each position. For each experimental strain, from the "pileup" files, we determined the frequency of each SNP allele at each heterozygous SNP position. Positions where the frequency of one of the two original alleles averaged $\geq$90% of the reads (average of frequencies from PSL2/W303a and PSL5/YJM789 assemblies) were candidates for LOH events. To be regarded as confirmed LOH events, the candidate events had to pass the same criterion for the assemblies obtained with both the CLC Genomics and BWA software. In order to translate the sequence coordinates of each putative LOH event from our analysis of contigs to SGD coordinates, we aligned SGD genomic sequences from the 16 chromosomes with the S288c-YJM789-PSL2-PSL5 alignment by using the MAFFT program. In-house PERL scripts were used to identify the exact SGD coordinates for each putative LOH event and to exclude putative LOH events that were located in repetitive regions of the genome; the coordinates of repetitive regions are given in SGD.

Mutation analysis

In the irradiated experimental diploids, we looked for induced mutations in non-repetitive regions of the genome that were originally identical in the two parental haploids. The mutations were identified using the SNP-calling software of CLC Genomics based on both second-stage parental haploid assemblies. *De* novo mutations were identified if they were supported by $\geq$ 40% of the reads in one experimental strain, and less than 15% of the reads in the other experimental strains. We validated each of the identified mutations by a manual comparison to the BWA assembly using Integrative Genomic Viewer (ROBINSON *et al.* 2011). Finally, we confirmed the existence of the mutations by DNA sequence analysis of PCR fragments containing the relevant region.

Supporting Information: Discussion

Below, we expand our discussion of two topics introduced in the main text: 1) evidence that some of the unselected gene conversion events that are not associated with centromere-distal LOH are associated with crossovers, and 2) the mechanisms involved in forming complex gene conversion events.

Gene conversion events that are unassociated with centromere-distal LOH

In order for a mitotic crossover to produce LOH, the two daughter cells must receive one recombinant chromosome and one non-recombinant chromosome (Fig. 2). If one daughter cell receives both recombinant chromosomes and the other receives both non-recombinant chromosomes, LOH will not be observed (CHUA and JINKS-ROBERTSON 1991). Thus, about half of the gene conversions that <u>are</u> associated with crossovers will not be detectable by our analysis. To look for crossover-associated gene conversion events that did not result in LOH, we looked for changes in linkage of markers flanking the conversion by analyzing the meiotic products that resulted from sporulating diploids derived from sectored colonies.

We examined the meiotic products of eleven sectored colonies (three from γ ray-treated cells and eight from UV-treated cells) that had unselected 4:0 gene conversion events unassociated with LOH. The meiotic segregation patterns of the heterozygous SNPs flanking the 4:0 conversion tract were analyzed using the PCR-based method described previously (Lee *et al.*, 2009). Since the physical distances separating the flanking markers were relatively small (< 20 kb for most intervals), we expected that most of the detected crossovers would reflect mitotic rather than meiotic events in the tested interval.

If no mitotic crossover was associated with the 4:0 tract, we expected that, in the tetrads derived from both sectors, most tetrads would have two spores with the YJM789 form of SNPs flanking the conversion tract and two spores with the W303a form of SNPs flanking the conversion tracts. If a mitotic crossover occurred in G2, we expect that one sector would produce tetrads in which all four spores have the recombinant configuration of the flanking markers, and the other sector would have tetrads in which the spores have flanking markers in the original parental configurations. Finally, if both a gene conversion event <u>and</u> a crossover occur in G1, we would expect to find the recombinant configuration of markers in tetrads derived from both sectors. Of the eleven sectored colonies examined, eight had the pattern expected for gene conversion unassociated with the crossover (Table S12). One had the pattern expected for a crossover in G2 with the segregation of two recombinant chromosomes in one cell and two non-recombinant chromosomes in the other. Interestingly, two sectored colonies had the pattern consistent with a conversion event and a crossover occurring in G1. This pattern, however, could also be a consequence of the repair of two broken chromosomes in G2, both associated with a crossover.

Complex gene conversion tracts

As discussed in the text, many of the conversion tracts associated with crossovers were complex, involving multiple transitions between 3:1, 4:0, and 2:2 regions within the tract or tracts in which 3:1 and 1:3 segments occurred within one

tract. Below, we will first discuss gene conversion events that are not associated with LOH of centromere-distal markers, followed by a discussion of LOH-associated conversions. Diagrams of all recombination events in our study are shown in Tables S1 and S2, and figures showing the patterns of DSB repair required to produce the recombination events are shown in Figs. S1-S40.

Complex gene conversion tracts unassociated with LOH of centromere-distal markers

It is important to note that gene conversion events unassociated with crossovers could occur by three pathways (Fig. 1): 1) SDSA in which there is a single region of heteroduplex, 2) resolution of a double Holliday junction (dHJ) which results in two regions of heteroduplex located in *trans*, and 3) dissolution of the dHJ by topoisomerase leading to two regions of heteroduplex located in *cis* on the chromosome. In an analysis of plasmid-chromosome recombination in yeast, MITCHEL *et al.* (MITCHEL *et al.* 2010) concluded that most mitotic gene conversion events were a consequence of SDSA.

As discussed below, most (about 80%) of the conversion events unassociated with LOH of distal markers are explicable as a consequence of repair of one or two DSBs by the SDSA pathway. These conversion events were divided into Classes A-G. The defining attributes of each class, the mechanism(s) required to produce each class, and the figure number illustrating the class are: 1) Class A (3:1 or 1:3 conversion, SDSA repair of single G2 DSB, Fig. S1), 2) Class B (4:0 or 0:4 conversion, SDSA repair of two DSBs located at the same position on sister chromatids, Fig. S2A), 3) Class C (3:1/4:0 or 1:3/0:4 hybrid conversions, SDSA repair of two DSBs located at the same position on sister chromatids with different length of conversion tracts, Fig. S3), 4) Class D (3:1 tract in which the homozygous region is split between the two sectors, repair of two DSBs by SDSA with the conversion tracts on opposite sides of the DSB, Fig. S4), 5) Class E (3:1/4:0/3:1 or 1:3/0:4/1:3 conversions, repair of two DSBs, one by SDSA and one by the DSB repair pathway as shown in Fig. 1, Fig. S5), 6) Class F (3:1, 1:3 or hybrid conversion tracts are disrupted by segment of 2:2 segregation, repair of DSBs in which there is "patchy" repair (defined below), Fig. S6-S9, and 7) Class G (two closely-linked conversions with different donors (3:1 and 1:3 or 4:0 and 0:4); models to explain Class G (Fig. S10-S12) will be described below.

In summary, Classes A-E, which account for more than 80% of the conversion events, can be simply explained by the repair of one or two DSBs by the mechanisms shown in Fig. 1. Concerning Classes F and G, several additional points need to be discussed. First, although yeast cells can efficiently repair a double-stranded DNA gap (MITCHEL *et al.* 2010; ORR-WEAVER and SZOSTAK 1983), it is likely that most mitotic gene conversion events reflect heteroduplex formation followed by the repair of the resulting mismatches (WENG 1998). Mismatch repair can result in a detectable conversion event or a restoration event. For example, in Fig. 1A, repair of the heteroduplex resulting in a duplex with two "red" strands would represent conversion-type repair, since this pattern produces 3:1 segregation; repair of the mismatch to produce a duplex with two "blue" strands represents restoration-type repair, since this pattern generates two cells that retain heterozygosity at the position of the original heteroduplex. Although multiple mismatches within one heteroduplex are

generally converted in a concerted manner yielding a continuous conversion tract, tracts with mixtures of conversion-type and restoration-type repair have been detected in both meiosis (MANCERA *et al.* 2008; SYMINGTON and PETES 1988) and mitosis (MITCHEL *et al.* 2010; NICKOLOFF *et al.* 1999). Second, we point out that some of the events have more than one plausible interpretation. For example, the F4 recombination event (Table S1) can be explained by patchy repair of two DSBs in G2 (Fig. S8A) or patchy repair of a single DSB in G1 (Fig. S8B).

In the Class G events, two conversions in opposite directions are observed (3:1 and 1:3 or 4:0 and 0:4). Although such events could be explained as a consequence of two independent DSBs, their frequency and the close linkage of the two types of conversion tracts indicate that they likely reflect the repair of a single DNA lesion. There are two different modifications of the models shown in Fig. 1 that can explain the conversion events with two different donors. In one model (Fig. S10A), there are two different rounds of mismatch repair associated with an SDSA event. In the first round, conversion occurs in which information from the invading strand is transferred to the invaded strand. Following the reversal of the strand invasion, a second round of mismatch repair occurs to produce the 3:1/1:3 conversion pattern. In an alternative pathway, a dHJ intermediate is formed, followed by branch migration of one of the junctions (Fig. S10B). During recombination in *E. coli*, a Holliday junction can be translocated by branch migration, resulting in symmetric heteroduplexes (WEST 1997). Although genetic evidence argues against the formation of symmetric heteroduplexes during meiotic recombination in *S. cerevisiae* (PETES *et al.* 1991), symmetric heteroduplexes have been invoked previously to explain certain classes of mitotic gene conversions (ESPOSITO 1978; NICKOLOFF *et al.* 1999; ROITGRUND *et al.* 1993).

As described above, because of the patterns of chromosome segregation following mitotic crossovers, only half of the events lead to LOH of centromere-distal markers. Thus, some of the gene conversion events that are unassociated with LOH for centromere-distal markers could reflect the crossover-associated conversions. Thus, there are likely to be additional pathways of repair other than those shown in the supplementary figures. It is important to stress, however, that about 80% of the conversion events are simply explained as a consequence of the SDSA pathway shown in Fig. 1, and that most of the events (about three-quarters) are most simply explained as a consequence of the repair of two sister chromatids that are broken at approximately the same position.

Complex conversion tracts associated with crossovers

We divided the gene conversion tracts associated with crossovers into the following classes: 1) Class H (no detectable conversion tract or 3:1 conversion associated with repair of single DSB (Fig. S13-S15), 2) Class I (repair of two DSBs by canonical repair pathways (no patch repair or branch migration, Fig. S16-S21), 3) Class J (repair of one or two DSBs involving either patchy repair or branch migration, Fig. S22-S35), and Class K (recombination events that involve more than one independent DSB on one chromosome arm, Fig. S36-S38). Of the crossover-associated conversions, about half were relatively simple (Classes H and I), and half required patchy repair and/or branch migration.

As discussed above, conversion tracts associated with crossovers are significantly more complex than those unassociated with crossovers. MANCERA *et al.* (MANCERA *et al.* 2008) reported that 11% of meiotic crossovers had complex conversion tracts, whereas the frequency of complex tracts among conversions unassociated with crossovers was 3%. One explanation of this difference could be that crossovers that proceed through the pathway shown in Fig. 1 are associated with two regions of heteroduplex, while conversions resulting from SDSA have only a single region of heteroduplex. Second, since gene conversion tracts associated with crossovers are usually longer than those unassociated with crossovers (AGUILERA and KLEIN 1989; MANCERA *et al.* 2008), there may be a greater chance to observe patchy repair in tracts associated with crossovers.

References

AGUILERA, A., and H. L. KLEIN, 1989 Yeast intrachromosomal recombination: long gene conversion tracts are preferentially associated with reciprocal exchange and require the *RAD1* and *RAD3* gene products. Genetics **123:** 683-694.

CHUA, P., and S. JINKS-ROBERTSON, 1991 Segregation of recombinant chromatids following mitotic crossing over in yeast. Genetics **129:** 359-369.

ESPOSITO, M. S., 1978 Evidence that spontaneous mitotic recombination occurs at the two-strand stage. Proc Natl Acad Sci U S A **75:** 4436-4440.

KATOH, K., G. ASIMENOS and H. TOH, 2009 Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol **537:** 39-64.

LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754-1760.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079.

MANCERA, E., R. BOURGON, A. BROZZI, W. HUBER and L. M. STEINMETZ, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature **454:** 479-485.

MITCHEL, K., H. ZHANG, C. WELZ-VOEGELE and S. JINKS-ROBERTSON, 2010 Molecular structures of crossover and noncrossover intermediates during gap repair in yeast: implications for recombination. Mol Cell **38:** 211-222.

NICKOLOFF, J. A., D. B. SWEETSER, J. A. CLIKEMAN, G. J. KHALSA and S. L. WHEELER, 1999 Multiple heterologies increase mitotic double-strand break-induced allelic gene conversion tract lengths in yeast. Genetics **153:** 665-679.

ORR-WEAVER, T. L., and J. W. SZOSTAK, 1983 Yeast recombination: the association between double-strand gap repair and crossing-over. Proc Natl Acad Sci U S A **80:** 4417-4421.

PETES, T. D., R. E. MALONE and L. S. SYMINGTON, 1991 Recombination in yeast, pp. 407-521 in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, edited by J. R. BROACH, JONES, E. W., AND J. R. PRINGLE. Cold Spring Harbor Press, Cold Spring Harbor.

ROBINSON, J. T., H. THORVALDSDOTTIR, W. WINCKLER, M. GUTTMAN, E. S. LANDER *et al.*, 2011 Integrative genomics viewer. Nat Biotechnol **29:** 24-26.

ROITGRUND, C., R. STEINLAUF and M. KUPIEC, 1993 Donation of information to the unbroken chromosome in double-strand break repair. Curr Genet **23:** 414-422.

SYMINGTON, L. S., and T. D. PETES, 1988 Expansions and contractions of the genetic map relative to the physical map of yeast chromosome III. Mol Cell Biol **8:** 595-604.

WEI, W., J. H. MCCUSKER, R. W. HYMAN, T. JONES, Y. NING *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. Proc Natl Acad Sci U S A **104:** 12825-12830.

WENG, Y. S., AND NICKOLOFF, J. A., 1998 Evidence for independent mismatch repair processing on opposite sides of a double-strand break in *Saccharomyces cerevisiae*. Genetics **148:** 59-70.

WEST, S. C., 1997 Processing of recombination intermediates by the Ruv ABC proteins. Annu Rev Genet **31:** 213-244.