Table S1. Mapping Statistics for Spo11 Oligo Sequences

Dataset[a]	No. of reads	No. mapped[b]	No. mapped uniquely to S288C genome [c]	No. mapped uniquely to SK1 genome [c]
1A	425,670	406,328 (95.5%)	398,185 (98.0%)	396,988 (97.7%)
1B	630,115	603,864 (95.8%)	591,686 (98.0%)	589,975 (97.7%)
1C	622,995	583,838 (93.7%)	570,053 (97.6%)	568,169 (97.3%)
2	511,427	490,569 (95.9%)	478,609 (97.6%)	477,106 (97.3%)
Total	2,190,207	2,084,599 (95.2%)	2,038,533 (97.8%)	2,032,238 (97.5%)

a. Datasets 1A, 1B, 1C were from technical replicates obtained from the same starting preparation of Spo11 oligos from a single meiotic culture. Samples for datasets 1A and 1B were generated by independent amplification from the same preparation of tailed and ligated Spo11 oligos. The sample for dataset 1C was generated by repeating the full tailing and ligation procedure on a separate aliquot of the purified Spo11 oligos. The sample for dataset 2 was a biological replicate starting from a separate meiotic culture.

b. Total number of reads that could be mapped to either genome (2,070,408 total were mapped to S288C). The small fraction of sequences that could not be mapped likely reflect sequencing errors, adaptor dimers, PCR primer dimers, and bona fide Spo11 oligos derived from genomic regions that are unique to SK1 (i.e., not found in the S288C reference strain) but that are missing from the current draft assembly of the SK1 genome.

c. Total number of reads that mapped to only one position in either genome

Table S4. *S. cerevisiae* Strains Used in This Study

Strain #	Relevant genotype
NKY611	*MATα/MATa, ho::LYS2/", lys2/", ura3/", leu2::hisG/"*
SKY1510	*MATα/MATa, ho::LYS2/", lys2/", ura3/", leu2::hisG/", his4X::LEU2/", arg4-Nsp/", nuc1Δ::LEU2/", SPO11-HA3-His6::kanMX4/", sae2:KanMX6/"*
SKY1629	*MATα/MATa, ho::LYS2/", lys2/", ura3/", leu2::hisG/", his4X::LEU2/", arg4-Nsp/", nuc1Δ::LEU2/", dmc1Δ::LEU2/", SPO11-HA3-His6::kanMX4/"*
SKY1874	*MATα/MATa, ho::LYS2/", lys2/", ura3/", leu2::hisG/", his4X::LEU2/", arg4-Nsp/", nuc1Δ::LEU2/", SPO11-HA3-His6::kanMX/"*
SKY2222	*MATα/MATa, ho::LYS2/", lys2/", ura3/", leu2::hisG/", arg4-nsp/", his4X::LEU2/", nuc1::LEU2/", spo11Y135F-HA3His6::KanMX/"*

All strains are diploid derivatives of SK1 (Kane and Roth, 1974).

Table S5. Additional Processed Data Files

File Name	Description
S288C_Spo11_reads.full.txt	Map positions, alignment scores, and aligned sequence for all reads that mapped to unique locations in the S288C genome. This file includes reads that mapped to exogenous sequences (*his4LEU2*, etc.), which were not included in the bed-format file available at GEO (GSE26449).
SK1_Spo11_reads.full.txt	Map positions, alignment scores, and aligned sequence for all reads that mapped to unique locations in the SK1 genome, including exogenous sequences (*his4LEU2*, etc.).
SK1_Spo11_Multimap.txt	Map positions, alignment scores, and aligned sequence for all reads that mapped to two or more positions in at least one genome. This file contains the map positions in the SK1 genome.
S288C_Spo11_Multimap.txt	Map positions, alignment scores, and aligned sequence for all reads that mapped to two or more positions in at least one genome. This file contains the map positions in the S288C genome.
S288C_Spo11_Raw_multimap_wiggle.txt	Raw multimap file. Wiggle format file of raw counts for oligos that mapped to multiple positions in one or both genomes. Coordinates are for the S288C genome. This file is formatted for display in the UCSC or other genome browser.
S288C_Spo11_Normalized_multimap_wiggle.txt	Normalized multimap file. As above, but with the count for each oligo normalized for the number of places it mapped.
FullMap_Spo11_SK1_*.txt	Curated SK1 Spo11 oligo map. Raw counts are given of the number of Spo11 oligos whose 5′ ends mapped to the indicated position in the SK1 genome. There are 21 files in this set (* is replaced by the respective chromosome, including exogenous sequences (*his4LEU2*, etc.)), each containing a table with four columns. Column 1: Position, Column 2: Count for the Watson strand; Column 3: Count for the Crick strand; Column 4: Combined Watson and Crick counts.
FullMap_Spo11_S288c_*.txt	Curated S288C map. File structure as above

	for the SK1 map.
his4leu2.fasta	Fasta sequence of the *his4LEU2* hotspot insert. Any oligos that mapped to this sequence were assigned to this region, even if they mapped elsewhere as well.
hisG.fasta	Fasta sequence of the exogenous *hisG* insert present in the strain used as a source of Spo11 oligos.
HA3His6-KanMX4.fasta	Fasta sequence of the *HA3His6-KanMX4* cassette at the 3′ end of the *SPO11* locus.

These .txt files, which augment the data deposited with the GEO repository, can be downloaded from Data S1.

Supplemental Experimental Procedures

Contents:

Detailed protocol for Spo11 oligo purification and amplification

Meiotic cultures. Cells from the SK1 strain background (Kane and Roth, 1974) (strain SKY1874) were streaked from a frozen stock onto a YPD plate and grown at 30°C for two days. A single colony was inoculated into 35 ml liquid YPD medium and grown for ≥24 hrs at 30°C. The saturated YPD culture was used to inoculate 2 l liquid YPA medium (1% yeast extract, 2% Bacto peptone, 2% potassium acetate, 0.001% antifoam 204 (Sigma)) to OD_{600} 0.2 and grown in 2.8 l baffled Fernbach flasks (1 liter per flask) at 225 rpm at 30°C for 14 hrs. Cells were harvested, washed once, resuspended in 2 l sporulation medium (2% potassium acetate and 0.001% antifoam 204) that was pre-warmed at 30°C and incubated in 2.8 l baffled flasks (1 liter per flask) at 225 rpm at 30°C for 4 hrs.

Preparation of nuclear extract. All procedures were performed at 4°C unless otherwise indicated. All buffers were kept at 4°C, and immediately before use, protease inhibitors were added to a final concentration of 1 mM phenylmethylsulphonyl fluoride and 10 μg/ml each of leupeptin, chymostatin and pepstatin A (all from Sigma). Cells from 2 l meiotic cultures were harvested and cell pellets were transferred to a pre-chilled 40 ml glass Dounce homogenizer. Cells were homogenized in 60 ml spheroplasting buffer (1.2 M sorbitol, 0.1 M MES-NaOH, pH

6.4, 10 mM EDTA, plus protease inhibitors). 120 μl of 50 mg/ml zymolyase 100T (US Biological; freshly dissolved in 3 parts of spheroplasting buffer plus 1 part of glycerol) and 400 μl β-mercaptoethanol were added to the cell suspension and incubated at 30°C for 55 min with occasional agitation. The spheroplasts were layered onto three 15 ml sucrose cushions (1 M sucrose, 0.1 M MES-NaOH, pH 6.4, 10 mM EDTA, plus protease inhibitors) in 50 ml tubes and centrifuged for 10 min at 20,000 x *g* at 4°C in a swinging-bucket rotor. The spheroplast pellets were suspended in 180 ml hypotonic lysis buffer (20 mM MES-NaOH, pH 6.4, 10 mM EDTA, 25 mM NaCl, 0.05% Triton X-100, plus protease inhibitors), as follows: spheroplast pellets were transferred to a pre-chilled 100 ml glass Dounce homogenizer and homogenized in 80 ml hypotonic lysis buffer; the spheroplast suspension was then divided into two pre-chilled 100 ml glass Dounce homogenizers; 50 ml hypotonic lysis buffer was added to each and the suspension was further homogenized. The spheroplasts were allowed to lyse for 10 min on ice, followed by further homogenization. Equal volumes of hypotonic lysates were layered onto six 13 ml sucrose cushions (1 M sucrose, 20 mM MES-NaOH, pH 6.4, 10 mM EDTA, 25 mM NaCl, plus protease inhibitors) in 50 ml tubes and centrifuged for 15 min at 20,000 x *g* at 4°C in a swinging-bucket rotor. The pellets containing nuclei were transferred to a pre-chilled 40-ml glass Dounce homogenizer and homogenized in 30 ml 0.5 M Tris-HCl, pH 7.4 plus protease inhibitors. SDS was added to the nuclear suspension to a final concentration of 2%. The suspension was sonicated for 10 s at 100% constant output in a Branson sonicator to reduce viscosity, then heated in a boiling water bath for 6 min, followed by rapid chilling on ice. The nuclear lysate was centrifuged for 15 min at 20,000 x *g* at 4°C in a swinging-bucket rotor. The supernatant was saved as nuclear extracts at -80°C until immunoprecipitation procedures.

Immunoprecipitation (IP) of Spo11-oligo complexes. Nuclear extract was supplemented with 1/10 volume of 10% Triton X-100 and 1/40 volume of 4 M NaCl before IP. 130 μg anti-HA antibody (clone F-7, Santa Cruz) was added to the extract and incubated for 1

hr at 4°C, followed by addition of 1.3 ml 50% Protein G-agarose slurry (Roche) and incubation for 4 hrs at 4°C with rotation. A mock IP control (no antibody addition) was carried out using supernatants of IPs from a prior experiment. Protein G beads were collected by centrifugation and washed four times with wash buffer (WB: 1% Triton X-100, 15 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA). During the washing step, beads were divided into two low-retention tubes (Siliconized G-tube, VWR; low-retention tubes were used for the remaining procedures). Immunocomplexes were eluted from Protein G beads by boiling in an equal volume of 2x Laemmli buffer, followed by rapid chill on ice. Supernatants were saved. Beads were boiled again in an equal volume of 0.5x Laemmli buffer, and supernanants were saved. Beads were then washed with 2 volumes of 2x WB. Supernatants from the wash were combined with the Laemmli buffer eluates, which were subjected to a second round of IP.

To each tube containing the first IP eluates, 15 µg anti-HA antibody was added and incubated for 1 hr at 4°C, followed by incubation with 150 µl Protein G-agarose slurry overnight at 4°C with rotation. To tubes containing mock IP products, only Protain G-agarose slurry was added. Protein G beads were collected by centrifugation and washed three times with 1x WB. During the washing step, Protein G beads were combined into one tube. 1/100 of Protein G beads were reserved in 1x WB for radio-labeling (see next section). The rest of the beads were boiled in an equal volume of 2x Laemmli buffer and the supernatant was saved. The beads were then boiled with an equal volume of 0.5x Laemmli buffer and the supernatant was saved. Elution with 0.5x Laemmli buffer was repeated, and all three eluates were combined.

Labeling of Spo11-oligo complexes. The aliquot of Protein G beads reserved from the second IP was washed twice with 1x labeling buffer (LB, 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 1 mM dithiothreitol at pH 7.9), and then incubated with 10 units terminal deoxynucleotidyl transferase (TdT, Fermentas), 10 µCi [α-^{32}P]-cordycepin triphosphate (5,000 Ci/mmol), and 2.5 µl 10x LB in a total volume of 25 µl for 1 hr at 37°C.

Beads were then washed three times with 1x WB. Spo11-oligo complexes were eluted by boiling for 3 min in 30 µl 2x Laemmli buffer.

Purification of Spo11-associated oligos. Radiolabeled and nonradioactive IP eluates were combined and separated on 7.5% SDS-PAGE. Wet gels were exposed to a phosphor imaging plate for a few hours to detect radiolabeled Spo11-oligo complexes. Gel pieces containing Spo11-oligo complexes were excised, crushed, and incubated in elution buffer (100 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.5% SDS, 1 mM $CaCl_2$) containing 100 µg DNA-free proteinase K overnight at 50°C to elute Spo11-associated oligos (Proteinase K from Fisher Scientific was further purified by HiTrap SP HP column in order to remove any residual contaminating DNA). Gel fragments were removed by filtration through SPIN-X columns (Corning). The cleared eluate was supplemented with 0.3 volume of 9 M ammonium acetate, 10 µg of DNA-free glycogen and 2.5 volumes of 100% ethanol, precipitated for 1 hr on dry ice, and centrifuged for 15 min at 4°C at 16,000 x *g* in a microcentrifuge (Glycogen from Roche was treated with DNase I, RNase A, RecJ and λ exonuclease to remove contaminating DNA or RNA). Tubes were rotated 180 degrees and spun for another 15 min. Tubes were rotated 180 degrees again and spun for a final 15 min. Unless otherwise indicated, ethanol-precipitated DNA was collected by three consecutive centrifugations as described here. Pellets were rinsed with 70% ethanol that was pre-chilled at -20°C, air dried, and then dissolved in 10 mM Tris-HCl, pH 8.0.

Quantification of purified Spo11 oligos. The yield of purified oligos was determined by end-labeling with GTP. A synthetic oligo (1, 2, and 4 fmol) and 1/10 of purified Spo11-oligos were incubated for 3 hrs at 37°C in labeling reactions each containing 1x LB, 10 units TdT, 5 µM cold GTP, and 10 µCi [α-^{32}P]-GTP (3,000 Ci/mmol) in a total volume of 8 µl. An equivalent amount of material from the mock IP was processed in parallel to evaluate specificity. DNA was precipitated with 120 ng tRNA, 0.3 volume of 9 M ammonium acetate, and 2.5 volumes of 100% ethanol on dry ice for 10 min, collected by centrifugation for 20 min at 4°C at 16,000 x *g*, and

dissolved in formamide loading buffer (80% deionized formamide, 10 mM EDTA, pH 8.0, 0.5 mg/ml xylene cyanol FF, 10% saturated bromophenol blue). An aliquot of each reaction was separated on 10% denaturing polyacrylamide gel. The gel was fixed in 10% methanol, 7% acetic acid, vacuum-dried, and exposed to a phosphor imaging screen. The amount of purified Spo11 oligos was determined by comparison to the labeled synthetic oligo.

Adaptor ligation. At least 30 fmol of purified Spo11 oligos were subjected to GTP tailing at their 3' ends. Mock products were processed in parallel, unless otherwise noted. Tailing reaction was carried out in a total volume of 40 µl containing 1x LB, 20 units TdT, and 13.8 µM GTP at 37°C for 5 hrs, followed by heat inactivation of TdT at 75°C for 10 min. TdT adds 3-5 GMP residues under these conditions (data not shown). In experiments with synthetic oligo substrates, the efficiency of TdT to add 3 or more GMP was >75% for oligos ending with dA, dT and dG, while the efficiency toward oligos ending with dC was ~50% (data not shown). The tailed oligos were ligated to double-stranded DNA adaptor I as follows: the tailing reaction was supplemented by addition of 10x T4 RNA ligase 2 buffer (500 mM Tris-HCl, pH 7.6, 50 mM $MgCl_2$, 50 mM β-mercaptoethanol) to 1x, 25 mM ATP to 0.5 mM, 5 pmol double-stranded adaptor I (top strand, 5′-pCCTGCAGGGAATTCCAGCTAGTC-3′-3′-T (i.e., 5′-phosphorylated, with 3′ end blocked with an inverted 3′-3′ linked dT), bottom strand, 5′-AGACTAGCTGGAATTCCCTGCAGGCCC, annealed and purified by non-denaturing polyacrylamide gel electrophoresis) and 300 fmol T4 RNA ligase 2 (gift from Stewart Shuman, MSKCC), and dH_2O to a final volume of 50 µl; the ligation reaction was carried out overnight at room temperature. Complementary strands of Spo11 oligos were synthesized as follows: the ligation reaction was supplemented with 2 mM dNTP to a final concentration of 30 µM and 10 units Klenow polymerase (NEB), and incubated at 25°C for 15 min. For samples in Datasets 1C and 2, Taq DNA polymerase was added directly to the Klenow reaction and incubated at 60°C for 10 min. This additional step was included to try to add an untemplated dA to each 3′ end of

the synthesized strand, to minimize the ambiguity of oligos with 5′-C residues. However, dA addition under these conditions was inefficient (data not shown).

Extension reactions were then supplemented with 0.3 volume of 9 M ammonium acetate, 10 µg of DNA-free glycogen and 2.5 volumes of 100% ethanol. DNA was precipitated for 1 hr on dry ice and centrifuged at 16,000 x *g*. The pellet was rinsed with 70% ethanol, air dried, dissolved in a mixture of 9 µl water and 15 µl formamide loading buffer. Denatured products and 10 bp ladder (radiolabeled with T4 polynucleotide kinase and [γ-^{32}P] ATP) were separated on a 10% denaturing polyacrylamide gel. The region between ~40-80 nt (equivalent to ~10–50 nt Spo11 oligos with $(rG)_{3\sim5}$ tails and ligated adaptor), was excised, crushed, and eluted in 400 µl 10 mM Tris-HCl, pH 8.0 at 37°C overnight with mixing. Elution mixture was spun through a SPIN-X column, then 0.3 volume of 9 M ammonium acetate, 10 µg DNA-free glycogen, and 2.5 volume of 100% ethanol were added. DNA was precipitated on dry ice for 1 hr and centrifuged at 16,000 x *g*. Pellet was rinsed with 70% ethanol and air dried.

The 3′ ends of gel-purified, denatured DNA strands were tailed with GTP by dissolving the dried pellet in 40 µl tailing reaction containing 1x LB, 30 units TdT, and 50 µM GTP, then incubating at 37˚C for 5 hrs. (Note: The original Spo11 oligos ligated to adaptor cannot be further modified because the 3′ end of the first adaptor is blocked by inverted dT. Moreover, Spo11 oligo 5′ ends are blocked by a covalent linkage with residual amino acids of Spo11 protein and thus also cannot be modified by the following procedures.) The tailed oligos were ligated to double-stranded DNA adaptor II as follows: the tailing reaction was directly supplemented by addition of 10x T4 RNA ligase 2 buffer to 1x final, 25 mM ATP to 0.5 mM, 5 pmol double-stranded adaptor II (top strand, 5′-pCCGCTGACGAATTCCGTTGCGTG-3′-3′-T; bottom strand, 5′-ACACGCAACGGAATTCGTCAGCGGCCC, annealed and purified by non-denaturing polyacrylamide gel electrophoresis) and 300 fmol T4 RNA ligase 2, and dH_2O to a final reaction volume of 50 µl; the ligation reaction was carried out at room temperature

overnight. Complementary strands were synthesized as follows: the ligation reaction was supplemented with 2 mM dNTP to a final concentration of 30 µM and 10 units Klenow polymerase (NEB), and incubated at 25˚C for 15 min.

PCR amplification. To estimate the yield, test PCR was carried out in a total of 30 µl containing 1/100 of the final Klenow reaction, 1x PCR buffer (Invitrogen), 2 mM $MgCl_2$, 0.2 mM dNTP, 1.5 units *Taq* polymerase (Invitrogen) and 1 µM JP96 (5′-GCCTCCCTCGCGCCATCAGCGCAACGGAATTCGTCAGCG, comprising 454 sequencing primer and part of adaptor II), 1 µM JP97 (5′-GCCTTGCCAGCCCGCTCAGCTAGCTGGAATTCCCTGCAGGC, comprising 454-specific primer required for emulsion PCR and part of adaptor I). The mixture was divided into three tubes and PCR was initiated by a denaturation step at 94°C for 10 s, followed by 20 cycles of amplification (94°C for 10 s, 60°C for 10 s, and 72°C for 10 s). PCRs were combined and 1/3 of the products were electrophoresed on a 10% non-denaturing polyacrylamide gel with 10 bp DNA ladder and low molecular weight DNA ladder to determine the size and quantity of PCR product. The amount of PCR product was estimated by comparing with the bands in low molecular weight DNA ladder after staining with ethidium bromide.

Template for 454 sequencing was prepared by a large-scale PCR with same conditions as above, scaled up to a total volume of 640 µl containing desired amount of the Klenow-extended products. At the time when we carried out deep sequencing, the 454 platform had a capacity to generate approximately half a million reads of ~100 bp length. In order to maximize the sequencing capacity, we aimed to seed a large-scale PCR with at least an order of magnitude more template molecules (i.e., at least 5 million molecules). The mixture was divided into 10 µl aliquots, denatured at 94°C for 10 s, then amplified for 16 cycles (94°C for 10 s, 60°C for 10 s, and 72°C for 10 s). PCR products were pooled, then 0.3 volume of 9 M ammonium acetate and 2.5 volumes of 100% ethanol were added. DNA was precipitated for 1 hr on dry ice,

and centrifuged at 16,000 x *g*. Pellets were rinsed with 70% ethanol, air dried, dissolved in 20 µl 10 mM Tris, pH 8.0, and separated on a 10% nondenaturing polyacrylamide gel. The gel piece between 100-140 bp was excised, crushed, and eluted in 300 µl 10 mM Tris, pH 8.0 at 37˚C overnight with mixing. The elution mixture was spun through a SPIN-X column, and DNA was precipitated with ammonium acetate and ethanol as above. The DNA pellet was dissolved in 20 µl 10 mM Tris, pH 8.0, and sequenced on the 454 platform (Roche) in the Genomics Core Laboratory at Memorial Sloan-Kettering Cancer Center.

Physical analysis of DSBs

For comparison of Spo11 oligo counts with direct detection of DSBs, genomic DNA was isolated at 6 hr in meiosis from three independent *sae2* cultures (SKY1510), three independent *dmc1* cultures (SKY1629), and one *spo11-Y135F* culture (SKY2222). DNA was prepared in low melting temperature agarose plugs as described (Murakami et al., 2009), except using 1% SDS in place of 1% Sarkosyl. One third of an agarose plug (approximately 600 ng of genomic DNA) was equilibrated twice in 5 ml of TE pH 8.0 and twice in 1 ml of appropriate restriction enzyme buffer (New England Biolabs) at 37°C for 15 min on a rotator. Agarose plugs were melted at 65°C for 10 min, then digested with 10 units of restriction enzyme for 2 hr. Agarose plugs were melted again at 70°C for 10 min and further digested with 10 units of fresh restriction enzyme. After digestion, agarose plugs were melted again at 70°C for 10 min and loaded onto 25 cm-long agarose gels in 1x TBE. Before starting electrophoresis, the agarose gel was equilibrated 15 min with buffer circulation. DNA was separated in 0.6, 0.8 or 1.0 % agarose gels with electrophoresis at 60 V for 16.3 hr, then detected by Southern blot hybridization using a ^{32}P-labeled DNA fragment adjacent to one of the restriction enzyme sites as a probe (Murakami et al., 2009). Hybridization signal was detected by phosphorimager, and DSB frequency was calculated as the percent of radioactivity in DSB fragments relative to total radioactivity in the lane. Data from the *spo11-Y135F* strain were used to subtract background.

Restriction enzymes and primer sequences for amplification of probes for Figure 1E were as follows: blot #1 (*GAT1*), *Pst* I, *GAT1* probe (5′-CGCGCTTCACATAATGCTTCTGG, 5′-TTCAGATTCAACCAATCCAGGCTC); blot #2 (*CCT6* and one additional hotspot), *Pst* I, *SLY1* probe (5′-GCGTCCCGCAAGGACATTAG, 5′-TTGTGGCTAATGGTTTTGCGGTG); blot #3 (*ERG1*), *Nco* I, *ATF2* probe (5′-CTGCCTACTCAAAACAGCAAAG, 5′-GTGAAGGAAGCACGTCAGAAAAAGC); blot #4 (*RSN1*), *Nco* I, *RSN1* probe (5′-GGTATGGCTATCACTCAAGGTGC, 5′-GCTGGATAATCGTAAGCATGAGA); blot #5 (*ERG25*), *Bsi*HKAI, *ADE6* probe (5′-GTTGATTTCTAAGCCGCATCTGG, 5′-GGTAAGCCTCCAAAGATGTCAAG); blot #6 (*YCR048w* and four additional hotspots), *Ase* I, *YCR048w* probe (5-AGTACCTTTCACACCTGAAG, 5-CACCGGGTGCATGAAG); blot #7 (*MET30* and five additional hotspots), *Stu* I, *SYG1* probe (5′-CACTTTCACTATCTGTTTCGGGAG, 5′-TTCATAGTACCTTGTGCGGTCTG); blot #8 (*CYS3* and one additional hotspot), *Ase* I, *CYS3* probe (5′-GAGCAGTTGCCGCTTTAGAG, 5′-TAGTTGGTGGCTTGTTTCAAGG).

Restriction enzymes and primer sequences for probe amplification in Figures 3B, S4C were as follows: *YAT1*, *Xba* I, *SWH1* probe (5′-CCAACTTCGCTCAAGGCTAC, 5′-TTGCAGCAATTCGTTCAAAG); *NAR1*, *Afl* II, *ATG2* probe (5′-GTGCACGAAAATATTGA, 5′-CGCCTTTGAAAGAAGCTTTG); *WHI5*, *Stu* I, *DIA2* probe (5′-TGAATTGGAGATACTGCAGACTTGTCCTCTG, 5′-CACTTATCGATGTCCCCATTAGATCC).

Bioinformatic analysis of Spo11 oligos

Statistical methods. Statistical analyses were performed using R (http://www.r-project.org/) or GraphPad Prism 5.0. Unless otherwise noted, p values were adjusted for

multiple comparisons using the p.adjust function in R (Benjamini and Hochberg, 1995). When thus corrected for false discovery rate, p≤0.05 was considered significant.

Mapping sequenced Spo11 oligos. Mapping of the 454 reads to the target genomes was performed using the following pipeline written specifically for this project. The kernel of the mapping pipeline used *rmapper-ls* (v1.2) from the SHRiMP mapping package (http://compbio.cs.toronto.edu/shrimp) (Rumble et al., 2009), which has a seeded mapping step followed by a local Smith-Waterman alignment. The SHRiMP program allows for gaps in the mapped alignment. The specific mapping parameters we used were:

```
-s 111101111 –n 1 –m -1 –i -1 –g -1 –e -1 –h 10 –o 10000
```

which specifies the seed pattern and Smith-Waterman parameters of 1 for matches and -1 for mismatches, gap openings and gap extensions. The alignment cutoff score was 10 and we set a limit of 10,000 for the number of best matches to store. Mapping was done against two separate genomes: the S288C genome (http://yeastgenome.org/, downloaded April 2009) and a draft assembly of the SK1 genome (Liti et al., 2009). Additional sequence elements present in strain SKY1874 but not in the reference genome assemblies (the *his4LEU2* hotspot construct, *spo11-HA3His6::kanMX4*, and *leu2::hisG*) were also included. These sequences (in fasta format) are available for download (Table S5).

Before mapping, adaptor sequences were removed from both the 5′ and 3′ ends using a custom script that looks for an optimal match to the adaptor sequences at each end of every read. If the adaptor sequences could not be found, the read was marked has having failed adaptor removal but still used in the mapping step. Failure in adaptor removal caused spurious mapping of only a very small number of Spo11 oligos, which were removed in subsequent curation steps (see below). After mapping, the reads were separated into unique and multiple mapping sets. Because of the variable number of rG residues added by terminal transferase to the 3′ end of Spo11 oligos and to the 3′ end of the reverse complementary strands, there is

ambiguity in defining the precise start and end positions for reads that map to positions starting with one or more C residues or ending in one or more G residues (see legend to Figure S1A). In such cases, the 5′ and 3′ ends of each read were defined to provide the longest contiguous sequence match with the genome.

Map curation and evaluation of reproducibility: Spo11 oligo maps were first calculated separately for SK1 and S288C using only reads that mapped to unique positions in both genomes. Maps for the exogenous DNA sequence elements (*his4LEU2*, etc.) were retained separately, rather than being included at their known chromosomal positions. (Their oligo counts were added to those of the chromosome they reside on for whole-chromosome analyses in Figures 2, S3.) The maps were then curated as follows. First, we identified discrepancies between the two maps as reads that mapped to different chromosomes or to different strands on the same chromosome, which accounted for <0.8% of total mappable reads. For each of these reads, we compared the alignment scores against the two genomes and, if the scores differed by ≥20%, the read was deleted from the map with the lower score; otherwise the discrepant reads were retained. Second, the rDNA array is represented in the S288C assembly by only 1.9 copies of the repeat unit, thus oligos that span the boundary between the repeats map to a single position in the genome even though they come from a repetitive sequence. Therefore, reads of this type were moved to the multiple mapping set. Finally, we observed 8,653 reads (<0.42% of total mappable reads) that mapped with similar alignment scores to unique positions in the S288C genome but to multiple positions in SK1. Because the SK1 genome assembly is incomplete and may contain incorrect duplications, we added these reads back to the S288C map.

Reads that mapped to multiple positions in one or both genomes (with exceptions noted above) were handled separately: each such read was assigned to every position to which it mapped, then normalized by the number of mapping sites. Raw and normalized “multi-maps” were assembled separately for each reference genome, and combined with the maps of

uniquely mapped reads where appropriate (e.g., in analysis of oligo counts in subtelomeric regions and analysis of DSBs in repetitive elements).

Raw sequence reads have been deposited in the GEO database (accession number GSE26449). This accession also contains the curated map (unique mapping reads only) against the S288C genome. This map file is in wiggle format to allow direct visualization in appropriate genome browsers, e.g., the UCSC browser (http://genome.ucsc.edu/). Additional processed data files in various formats are available for download (Table S5).

Estimates of total DSB numbers and comparison of DSBs and recombination frequencies. The regression lines in Figure 1E are least-squares fits to $\log_{10}$-transformed data. The regression line for Spo11 oligo counts vs. *dmc1* DSBs is: $\log_{10}(D) = 1.102 \times \log_{10}(S) - 2.992$, and the regression line for *sae2* is: $\log_{10}(D) = 1.162 \times \log_{10}(S) - 3.542$, where D is DSB frequency (percent of DNA cleaved in the population) and S is the Spo11 oligo count (hpM). To estimate DSB numbers from Spo11 oligo counts, we used the *dmc1* regression relationship because *rad50S*-like mutants (of which *sae2* is an example) are reported to have substantially reduced DSBs in many genomic regions (Blitzblau et al., 2007; Buhler et al., 2007). We converted S to D, then estimated the number of DSBs by assuming on average that only one DSB forms at a given site per 4 chromatids, i.e., that a DSB frequency of 1% of DNA corresponds to 0.04 DSB per meiosis. Total DSBs per chromosome were estimated in this manner from the sum of Spo11 oligos for each chromosome; total cellular DSBs were estimated by summing the per-chromosome numbers. The result is an estimate of 126.7 DSBs per meiosis in the *spo11-HA* strain used here. This mutant experiences ~80% of the wild-type number of breaks (Martini et al., 2006), leading to an estimate of 158.3 DSBs per meiosis in wild type, similar to estimates from microarray hybridization of ssDNA from *dmc1* mutants (140–170 DSBs per meiosis) (Buhler et al., 2007).

If we use the *sae2* regression line instead, we estimate 69.6 DSBs per meiosis in *spo11-HA* and 87.0 DSBs per meiosis in wild type. This number is moderately larger than the previous

genome-wide estimate of 44 DSBs per meiosis based on Spo11 microarray ChIP from a *spo11-HA rad50S* mutant (Buhler et al., 2007), but is still not high enough to account for the average of 136.7 total crossovers and noncrossovers detected in wild type meiosis (Mancera et al., 2008). These findings thus support the conclusion (Buhler et al., 2007) that total DSB numbers are suppressed in *rad50S*-like mutants.

Crossover (cM) and noncrossover (rec) frequencies were calculated from previously published studies (Chen et al., 2008; Mancera et al., 2008) as:

cM = number of crossover events ÷ number of tetrads analyzed ÷ 2

rec = number of noncrossover events ÷ number of tetrads analyzed ÷ 4

Densities of crossovers (cM/kb), noncrossovers (rec/kb), and Spo11 oligo counts (hpM/kb) were calculated using the chromosome sizes from the Saccharomyces Genome Database. The full rDNA array was omitted in calculating the size of chromosome XII.

Telomere-proximal and pericentric Spo11 oligos. To evaluate the significance of DSB suppression in the 20 kb telomere-proximal regions, we used a one-sided Mann-Whitney test to compare the oligo counts for the 32 chromosome ends to oligo counts in a sample of 5,000 randomly chosen 20 kb genome segments. To evaluate the significance of the reduced hotspot density and intensity near centromeres (Figure S3G,H), we extracted hotspot densities and intensities in 5 kb segments to the left and right of 1,000 randomly chosen chromosomal positions. (Chromosome ends (defined as first and last 20% of each chromosome) were excluded from being the start position for sampling). The distributions for the random sample were approximately normal (data not shown), so normal distributions with means and standard deviations equal to those of the random sample were used to calculate 99% confidence intervals for the random samples and p values (corrected for multiple comparisons) for the data observed for true pericentric regions. We did not analyze hotspot occurrence near telomeres because hotspots could only be reliably scored within unique sequences, and telomere-proximal zones (unlike pericentric regions) are enriched for repetitive sequences.

To evaluate the width of the zone of pericentric DSB suppression (Figure S3F), Spo11 oligos were summed in 3 kb bins moving outward from a bin centered on the centromeres. Counts were summed in similar bins generated around a collection of 1,000 random positions. The counts within the random bins were not normally distributed (data not shown), so Mann-Whitney tests were performed to evaluate whether the distributions of counts in the bins near the true centromeres were lower than in the random bins. Resulting p values were corrected for multiple comparisons.

Comparison of Spo11 oligo data with higher order chromosome structure features. For correlation analyses in Figure 2F, chromosomes were divided into overlapping bins of the indicated sizes. The GC content or published ChIP occupancy data were averaged in each bin, then tested for correlation with the Spo11 oligo counts. Correlation coefficients (Pearson's *r*) are shown in the figures. The condensin data (D'Ambrosio et al., 2008) and the Scc1 and Smc6 data (Lindroos et al., 2006) were from whole-genome microarrays. Because these proteins show strong enrichment around centromeres, we censored 40 kb pericentric regions from the analysis so we could focus on the patterns in interstitial portions of chromosome arms. Including the pericentric regions resulted in a slight strengthening of the anticorrelation observed with larger window sizes (e.g., with 20 kb windows, Pearson's *r* for Spo11 vs. Scc1 decreased from -0.18 with pericentromeric regions censored, to -0.23 with these regions included). Chr III was also censored from this analysis because the Spo11 oligo data are from a strain carrying the *his4LEU2* hotspot, which substantially alters DSB distributions on this chromosome (Xu and Kleckner, 1995). The Rec8 data (Kugou et al., 2009) were from microarrays covering Chr IV and V only. Because of the more limited coverage, pericentromeric regions were retained.

Hotspot identification. Hotspots were identified as follows. Using uniquely mapped reads pooled from all four datasets, we identified chromosome segments where the smoothed oligo frequency (201 bp Hann window) was >0.4 total hits per bp. We compared results with

various cutoffs, and chose this value because it resulted in hotspot calls that matched well with visual comparison of oligo maps with DSB distributions assayed by Southern blotting of genomic DNA from *sae2* mutants. This cutoff is equivalent to 0.193 hpM/bp, which is 2.3-fold over the genome average oligo density and 175-fold over the mean oligo density within rDNA (i.e., our best estimate for the upper limit of Spo11-independent background). Adjacent segments that were above the cutoff were merged if the distance separating them was ≤200 bp, so that closely spaced oligo clusters were counted as single hotspots. Hotspot boundaries were then defined as the positions within each segment of the first and last base pairs with at least one oligo. The collection of hotspots was subsequently filtered to require a minimum length of 25 bp and a minimum of 25 total mapped oligos. This method yielded 3,604 hotspots: 3,600 in endogenous chromosomal sequences and 4 within heterologous DNA inserts, namely, the *his4LEU2* hotspot, the *Salmonella typhimurium hisG* sequence disrupting *LEU2* (Alani et al., 1987), and the *kanMX4* cassette marking the epitope-tagged *SPO11* construct. No hotspots were identified when we applied these criteria to a randomized map of the same average count density (data not shown).

As part of validating the spatial accuracy of our hotspot calls, we mined the literature for examples of DSB hotspots directly detected by Southern blotting of genomic DNA from meiotic cells of the SK1 background. We found 94 documented hotspots (Sun et al., 1989; Cao et al., 1990; Goldway et al., 1993; Wu and Lichten, 1994; Baudat and Nicolas, 1997; Smith et al., 2001; Yamashita et al., 2004; Buhler et al., 2007; Borde et al., 2009; Kugou et al., 2009). All of these hotspots were successfully identified by our hotspot calling algorithm (data not shown).

Transcription factor analysis and transcription start site compilation. Transcription start sites (TSS) were compiled from previously published studies (Zhang and Dietrich, 2005; Miura et al., 2006; Perocchi et al., 2007; Nagalakshmi et al., 2008). TF binding sites were from (MacIsaac et al., 2006). We limited analysis to the subset of TF sites with strong supporting evidence that these are *bona fide* targets *in vivo*, using the most stringent criteria defined by

MacIsaac et al. (2006): ChIP binding cutoff p value of 0.001 and the stringent cutoff for conservation. We further limited analysis to TFs with ≥10 binding sites annotated according to these criteria (77 TFs and 4233 binding sites in total). To define TF binding sites as hotspot-associated, we required that the TF site overlaps with a Spo11 oligo hotspot. For hotspot ranking plots (e.g., Figure 5E), we counted each hotspot only once even if it had more than one binding site for a given TF.

To examine the spatial distribution of Spo11 oligos around TF binding sites (Figure 5), we calculated a normalized Spo11 oligo count around each binding site by dividing the oligo count at each base pair by the total number of oligos within ±500 bp of the center of the TF binding site. Only binding sites with ≥100 total oligos within this 1 kb window were included, which led to a total number of binding sites of 3,639. These normalized maps were then averaged for each TF to obtain a single Spo11 oligo profile characteristic of that TF. We used normalized rather than raw Spo11 oligo data to prevent these profiles being dominated by the subset of TF sites with the highest numbers of oligos nearby. We performed k-means clustering (Hartigan and Wong, 1979) on the mean normalized Spo11 oligo profiles using the "kmeans" function in R, which is a method of cluster analysis by which a series of observations (in this case, Spo11 oligo profiles for a series of TFs) is partitioned into a defined number (k) of clusters such that each observation belongs to the cluster with the nearest mean. For clustering, the Spo11 profiles across only the central ±200 bp centered on the TF binding site were used because regions further away contained little information to distinguish the TFs from one another (Figure 5A). Clustering was performed with varying k = 2–9. Based on visual inspection, k = 5 provided optimal partitioning of TFs into internally consistent groups that were distinct from one another (data not shown).

Fine-scale sequence analysis. The DNA sequence context ±100 bp around the 5′ end of each Spo11 oligo that mapped to a unique position was extracted, and position-specific base compositions (mono- and dinucleotides) were calculated for the strand from which the Spo11-

oligo was derived (the top strand as diagrammed in Figure 6A). The SK1 genome assembly was used for the analyses presented here, but results were indistinguishable if the S288C genome was used instead (data not shown). To more fully capture the net *in vivo* preferences of Spo11, we included all sequenced Spo11 oligos so that each individual cleavage site in the genome was represented in proportion to the frequency with which it was recovered. However, the conclusions discussed here also were drawn if only unambiguously mapped oligos were used (i.e., those that do not map next to a C residue) or if each cleavage position in the genome was only considered once (data not shown).

The position-specific score for net deviation of dinucleotide frequency from the local average (Figure 6B,E) was calculated as:

$$S_n = \sum_{i=1}^{16} \left| \log_2 \left(\frac{x_{n,i}}{\bar{x}_i} \right) \right|$$

where S_n is the deviation score for dinucleotide position n, i is the dinucleotide (1=AA, 2=AC, …, 16=TT), $x_{n,i}$ is the fraction of dinucleotide i at position n, and $\bar{x}_i$ is the mean fraction of dinucleotide i for the entire region ±100 bp from 5′ end of the Spo11 oligo. The randomized sample of 50-mers (gray line in Figure 6B) was generated by choosing a start position at random from within ±100 bp of the 5′ end of 100,000 randomly selected Spo11 oligos, then extracting the dinucleotide composition from the top (Spo11-oligo containing) strand and calculating the deviation score as above.

DNA structural parameters defined for each of the 16 dinucleotides were obtained from the Dinucleotide Property Database http://diprodb.fli-leibniz.de (Friedel et al., 2009). In order to calculate possible structural properties of the DNA sequences favored for Spo11 cleavage, we took a weighted average of the value for each property (major groove width, etc.) across all 16 possible dinucleotides, weighted according to the observed dinucleotide frequency at each position relative to the dyad axis. Similar overall trends for major and minor groove dimensions

were obtained using all mapped reads, only unambiguous reads (i.e., Spo11 oligos without a C at the 5′ end), or only unique cleavage positions (data not shown).

Structural depictions of *Methanococcus jannaschii* Top6A (PDB ID: 1D3Y) in Figures 6C and S7D were created using the UCSF Chimera package (http://www.cgl.ucsf.edu/chimera (Pettersen et al., 2004)).

Strand asymmetry analysis. Strand-specific oligo counts were tallied for all Spo11 cleavage positions for which the total count (Watson + Crick) was ≥8. Weaker cleavage sites were excluded because they provided insufficient statistical power to evaluate strand asymmetry. Cleavage sites were defined based on a 2-nt 5′ overhang (Liu et al., 1995; Murakami and Nicolas, 2009) and, where necessary because of ambiguity of mapping oligos with 5′-C residues, clusters of potential sites were merged (see example in Figure 7B). For each cleavage site, Watson and Crick counts were compared by Poisson test, which evaluates the probability that differences as large as those observed could have arisen by chance if the true frequencies in the population were equal on the two strands. P values were corrected for false discovery rate (Benjamini and Hochberg, 1995). We cannot exclude the possibility that observed asymmetry arises from sampling bias imposed by the ligation, amplification, or sequencing methods. However, many strongly asymmetric sites (including those highlighted in Figure 7B) are closely flanked by sites that show little to no evidence of asymmetry. Because the oligos that arose from these adjacent sites share most of their DNA sequence with oligos that arose from asymmetric sites, this finding suggests that sequence of the Spo11 oligo itself causes little bias in recovery in our assay. This further suggests that the observed strand-specific asymmetries reflect biases in oligo frequencies in the population rather than systematic sampling bias.

Nucleosome occupancy.

Sequencing mononucleosomal DNA. Synchronized meiotic SK1 cultures (NKY611) were incubated in 0.3% potassium acetate at OD_{600} ~1.9 (Dataset N1) or 2% potassium acetate at OD_{600} ~6.5 (Dataset N2). Samples of 50 ml were harvested at 0, 1, 2, 3 or 4 hr after transfer to sporulation conditions, and immediately crosslinked by addition of formaldehyde to 1% and incubation for 15 min at room temperature. Glycine was then added to a final concentration of 125 mM, cells were pelleted and resuspended in 18.5 ml Buffer Z (0.5 M sorbitol, 50 mM Tris-HCl, pH 7.4, 10 mM β-mercaptoethanol) containing 5 mg zymolyase 100T (US Biological) and incubated at 30°C for 30 min. Spheroplasted cells were pelleted, supernatants removed, and treated with a dilution series of micrococcal nuclease (MNase) from 5 to 80 Units (Worthington Biochemical, Corp.) in NP buffer (1 M sorbitol, 50 mM NaCl, 10 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5 mM spermidine, 1 mM β-mercaptoethanol, 0.075% NP-40) and incubated at 37°C for 20 min. Digestions were stopped by the addition of 5% SDS and 50 mM EDTA followed by the addition of 100 μg proteinase K (Fisher Scientific) and incubated at 65°C for 18 hr. DNA was then extracted with 1.5 M ammonium acetate, pH 5.5 and 500 μl of phenol-chloroform-isoamyl alcohol (25:24:1), washed with 100% isopropanol and then 70% ethanol, and treated with RNase. Mononucleosome-sized DNA generated by treatment with 20 (Dataset N1) or 40 (Dataset N2) units MNase was then purified by size fractionation on agarose gels and prepared for sequencing using the Genomic DNA Sample Prep Kit (Illumina, Inc.) and sequenced using the Illumina Genome Analyzer Systems at the MIT Biopolymers Facility (Dataset N1) or at the Genomics Resource Center at The Rockefeller University (Dataset N2).

Mapping. For mapping of Illumina reads (1–9 million reads per dataset) the rmapper-ls program from SHRiMP was again used, but with more stringent “gapless” parameters:

```
-s 111101111 -o 10001 -n 1 -m 1 -i -6 -g -5000 -e -5000 -h 24
```

The match score was 1, mismatches were -6 and gap openings and extensions were 5,000

which effectively makes them infinite on this scale. The Smith-Waterman cutoff was 24; so a read would need at least 24 matches (with zero mismatches) to pass. We recorded 10,001 optimal hits. This number was chosen to insure that all multiple mappings were enumerated and we checked the mapping results to verify that the most degenerate multiple mapping read was below this threshold. Reads were mapped to both the S288C and SK1 genomes, but only the S288C map was used in this study.

The mapping data were post-processed following previously described methods (Kaplan et al., 2009) (Figure S5A). We first looked across all datasets to identify regions containing reads that mapped to other locations as well. These regions, extended by 150 bp on each side, were excluded from further analysis. We next computed a histogram of distances by taking for each read the distances to the 5′ ends of all other reads that were downstream and on the opposite strand. The first mode of this distribution of distances was defined as the mononucleosome DNA length. Each read was computationally extended to this length, then an occupancy table was computed by summing at each base pair the number of extended reads that overlapped that base pair. The raw values for occupancy at each base pair were normalized by dividing them by the genome average occupancy per bp (mean raw occupancies ranged from 20–59 per bp, depending on the dataset). Raw sequence reads have been deposited in the GEO database (accession number GSE26452). This accession also contains a text file containing the calculated raw occupancy score for each bp (S288C genome coordinates).

For comparison between the seven nucleosome occupancy maps (six meiotic samples plus a vegetative dataset from Kaplan et al., 2009) shown in Figures S5B-D, Pearson's product-moment correlation coefficients were determined from the $\log_2$-transformed normalized occupancy values at each mapped position in the genome. Genomic positions that were not mapped in all datasets were excluded from the analysis. Annotated +1 nucleosomes (Figure 6F,G) were from (Jiang and Pugh, 2009).

Supplemental References

Alani, E., Cao, L., and Kleckner, N. (1987). A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains. Genetics *116*, 541-545.

Baudat, F., and Nicolas, A. (1997). Clustering of meiotic double-strand breaks on yeast Chromosome III. Proc. Natl. Acad. Sci. USA *94*, 5213-5218.

Ben-Aroya, S., Mieczkowski, P.A., Petes, T.D., and Kupiec, M. (2004). The compact chromatin structure of a Ty repeated sequence suppresses recombination hotspot activity in *Saccharomyces cerevisiae*. Mol. Cell *15*, 221-231.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. Series B *57*, 289-300.

Blitzblau, H.G., Bell, G.W., Rodriguez, J., Bell, S.P., and Hochwagen, A. (2007). Mapping of meiotic single-stranded DNA reveals double-strand-break hotspots near centromeres and telomeres. Curr. Biol. *17*, 2003-2012.

Borde, V., Robine, N., Lin, W., Bonfils, S., Geli, V., and Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. EMBO J. *28*, 99-111.

Buhler, C., Borde, V., and Lichten, M. (2007). Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in *Saccharomyces cerevisiae*. PLoS Biol. *5*, e324.

Cao, L., Alani, E., and Kleckner, N. (1990). A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. Cell *61*, 1089-1101.

Chen, S.Y., Tsubouchi, T., Rockmill, B., Sandler, J.S., Richards, D.R., Vader, G., Hochwagen, A., Roeder, G.S., and Fung, J.C. (2008). Global analysis of the meiotic crossover landscape. Dev. Cell *15*, 401-415.

D'Ambrosio, C., Schmidt, C.K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K., and Uhlmann, F. (2008). Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. Genes. Dev. *22*, 2215-2227.

Diaz, R.L., Alcid, A.D., Berger, J.M., and Keeney, S. (2002). Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation. Mol. Cell. Biol. *22*, 1106-1115.

Friden, P., and Schimmel, P. (1988). LEU3 of *Saccharomyces cerevisiae* activates multiple genes for branched-chain amino acid biosynthesis by binding to a common decanucleotide core sequence. Mol. Cell. Biol. *8*, 2690-2697.

Friedel, M., Nikolajewa, S., Suhnel, J., and Wilhelm, T. (2009). DiProDB: a database for dinucleotide properties. Nucl. Acids Res. *37*, D37-40.

Goldfarb, T., and Lichten, M. (2010). Frequent and efficient use of the sister chromatid for DNA double-strand break repair during budding yeast meiosis. PLoS Biol. *8*, e1000520.

Goldway, M., Sherman, A., Zenvirth, D., Arbel, T., and Simchen, G. (1993). A short chromosomal region with major roles in yeast chromosome III meiotic disjunction, recombination and double strand breaks. Genetics *133*, 159-169.

Hartigan, J.A., and Wong, M.A. (1979). A K-means clustering algorithm. Applied Statistics *28*, 100-108.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S.*, et al.* (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods *6*, 283-289.

Horowitz, H., Thorburn, P., and Haber, J.E. (1984). Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. Mol. Cell. Biol. *4*, 2509-2517.

Hunter, N. (2007). Meiotic recombination. In Molecular Genetics of Recombination, A. Aguilera, and R. Rothstein, eds. (Berlin Heidelberg, Springer-Verlag), pp. 381-442.

Jiang, C., and Pugh, B.F. (2009). A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. Genome Biol. *10*, R109.

Kane, S., and Roth, R. (1974). Carbohydrate metabolism during ascospore development in yeast. J. Bacteriol. *118*, 8-14.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kugou, K., Fukuda, T., Yamada, S., Ito, M., Sasanuma, H., Mori, S., Katou, Y., Itoh, T., Matsumoto, K., Shibata, T.*, et al.* (2009). Rec8 guides canonical Spo11 distribution along yeast meiotic chromosomes. Mol. Biol. Cell *20*, 3064-3076.

Lindroos, H.B., Strom, L., Itoh, T., Katou, Y., Shirahige, K., and Sjogren, C. (2006). Chromosomal association of the Smc5/6 complex reveals that it functions in differently regulated pathways. Mol. Cell *22*, 755-767.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V.*, et al.* (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337-341.

Liu, J., Wu, T.-C., and Lichten, M. (1995). The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. EMBO J. *14*, 4599-4608.

Loney, E.R., Inglis, P.W., Sharp, S., Pryde, F.E., Kent, N.A., Mellor, J., and Louis, E.J. (2009). Repressive and non-repressive chromatin at native telomeres in *Saccharomyces cerevisiae*. Epigenetics Chromatin *2*, 18.

Louis, E.J. (1995). The chromosome ends of *Saccharomyces cerevisiae*. Yeast *11*, 1553-1573.

Louis, E.J., and Haber, J.E. (1992). The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. Genetics *131*, 559-574.

Louis, E.J., Naumova, E.S., Lee, A., Naumov, G., and Haber, J.E. (1994). The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. Genetics *136*, 789-802.

MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. BMC Bioinformatics *7*, 113.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., and Steinmetz, L.M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature *454*, 479-485.

Martini, E., Diaz, R.L., Hunter, N., and Keeney, S. (2006). Crossover homeostasis in yeast meiosis. Cell *126*, 285-295.

Miura, F., Kawaguchi, N., Sese, J., Toyoda, A., Hattori, M., Morishita, S., and Ito, T. (2006). A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc. Natl. Acad. Sci. USA *103*, 17846-17851.

Murakami, H., Borde, V., Nicolas, A., and Keeney, S. (2009). Gel electrophoresis assays for analyzing DNA double-strand breaks in *Saccharomyces cerevisiae* at various spatial resolutions. Methods Mol. Biol. *557*, 117-142.

Murakami, H., Borde, V., Shibata, T., Lichten, M., and Ohta, K. (2003). Correlation between premeiotic DNA replication and chromatin transition at yeast recombination initiation sites. Nucl. Acids Res. *31*, 4085-4090.

Murakami, H., and Nicolas, A. (2009). Locally, meiotic double-strand breaks targeted by Gal4BD-Spo11 occur at discrete sites with a sequence preference. Mol. Cell. Biol. *29*, 3500-3516.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344-1349.

Neale, M.J., Pan, J., and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. Nature *436*, 1053-1057.

Nichols, M.D., DeAngelis, K., Keck, J.L., and Berger, J.M. (1999). Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. EMBO J. *18*, 6177-6188.

Nishant, K.T., Wei, W., Mancera, E., Argueso, J.L., Schlattl, A., Delhomme, N., Ma, X., Bustamante, C.D., Korbel, J.O., Gu, Z.*, et al.* (2010). The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. PLoS Genet. *6*.

Oh, S.D., Lao, J.P., Hwang, P.Y., Taylor, A.F., Smith, G.R., and Hunter, N. (2007). BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. Cell *130*, 259-272.

Ohta, K., Nicolas, A., Furuse, M., Nabetani, A., Ogawa, H., and Shibata, T. (1998). Mutations in the *MRE11*, *RAD50*, *XRS2*, and *MRE2* genes alter chromatin configuration at meiotic DNA double-stranded break sites in premeiotic and meiotic cells. Proc. Natl. Acad. Sci. USA *95*, 646-651.

Ohta, K., Shibata, T., and Nicolas, A. (1994). Changes in chromatin structure at recombination initiation sites during yeast meiosis. EMBO J. *13*, 5754-5763.

Oshima, Y. (1997). The phosphatase system in *Saccharomyces cerevisiae*. Genes Genet. Syst. *72*, 323-334.

Perocchi, F., Xu, Z., Clauder-Munster, S., and Steinmetz, L.M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucl. Acids Res. *35*, e128.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605-1612.

Primig, M., Williams, R.M., Winzeler, E.A., Tevzadze, G.G., Conway, A.R., Hwang, S.Y., Davis, R.W., and Esposito, R.E. (2000). The core meiotic transcriptome in budding yeasts. Nat. Genet. *26*, 415-423.

Rockmill, B., Voelkel-Meiman, K., and Roeder, G.S. (2006). Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in *Saccharomyces cerevisiae*. Genetics *174*, 1745-1754.

Rumble, S.M., Lacroute, P., Dalca, A.V., Fiume, M., Sidow, A., and Brudno, M. (2009). SHRiMP: accurate mapping of short color-space reads. PLoS Comput. Biol. *5*, e1000386.

Schacherer, J., Ruderfer, D.M., Gresham, D., Dolinski, K., Botstein, D., and Kruglyak, L. (2007). Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. PLoS One *2*, e322.

Schwank, S., Ebbert, R., Rautenstrauss, K., Schweizer, E., and Schuller, H.J. (1995). Yeast transcriptional activator INO2 interacts as an Ino2p/Ino4p basic helix-loop-helix heteromeric complex with the inositol/choline-responsive element necessary for expression of phospholipid biosynthetic genes in *Saccharomyces cerevisiae*. Nucl. Acids Res. *23*, 230-237.

Smith, K.N., Penkner, A., Ohta, K., Klein, F., and Nicolas, A. (2001). B-type cyclins *CLB5* and *CLB6* control the initiation of recombination and synaptonemal complex formation in yeast meiosis. Curr. Biol. *11*, 88-97.

Sun, H., Treco, D., Schultes, N.P., and Szostak, J.W. (1989). Double-strand breaks at an initiation site for meiotic gene conversion. Nature *338*, 87-90.

Wu, T.-C., and Lichten, M. (1994). Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science *263*, 515-518.

Xu, L., and Kleckner, N. (1995). Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast. EMBO J. *14*, 5115-5128.

Yamashita, K., Shinohara, M., and Shinohara, A. (2004). Rad6-Bre1-mediated histone H2B ubiquitylation modulates the formation of double-strand breaks during meiosis. Proc. Natl. Acad. Sci. USA *101*, 11380-11385.

Zhang, L., and Hach, A. (1999). Molecular mechanism of heme signaling in yeast: the transcriptional activator Hap1 serves as the key mediator. Cell. Mol. Life Sci. *56*, 415-426.

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. Nucl. Acids Res. *33*, 2838-2851.