# Mismatch repair impedes meiotic crossover interference

Tim J. Cooper[1], Margaret R. Crawford[1], Laura J. Hunt[1], Marie-Claude Marsolier-Kergoat[2], Bertrand Llorente[3] and *Matthew J. Neale[1].

1. *Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, UK*

2. CEA/DRF, I2BC/UMR 9198, SBIGeM, Gif-sur-Yvette, France; *CNRS-UMR 7026, Éco-anthropologie et Ethnobiologie, Musée de l'Homme, 17, Place du Trocadéro et du 11 Novembre, Paris, France.*

3. *Cancer Research Centre of Marseille, CNRS UMR7258, INSERM U1068, Institut Paoli-Calmettes, Aix-Marseille Université UM105, Marseille, France.*

## Summary

Sequence divergence, mediated by the anti-recombinogenic activity of mismatch repair (MMR), forms a potent barrier to meiotic recombination and in turn the formation of viable gametes[1–5]. However, exactly how MMR jeopardizes meiotic success is unclear. Here we utilize a combination of *S. cerevisiae* genetics, genome-wide mapping of recombination and computational modelling to demonstrate that MMR unexpectedly influences the global distribution of recombination through preferential suppression of interfering crossovers (COs) at regions of greater sequence divergence. As a result, inactivation of MMR not only increases the rate of recombination, as previously observed, but also, paradoxically, the strength of CO interference. Our observations reveal a new mechanism by which MMR spatially sculpts the meiotic landscape—linking CO control to the mechanisms that can reproductively isolate a population, and highlighting how genomes may become meiotically incompatible at the molecular level, dependent upon interactions of the primary DNA sequence.

## Introduction & Results

Meiosis, a specialized two step nuclear division, is responsible for the generation of genetically diverse, haploid gametes. An integral feature of the meiotic program is the initiation of homologous recombination and the subsequent formation of reciprocal, interhomologue exchanges known as crossovers (COs) that are essential for faithful disjunction of meiotic chromosomes during anaphase I (reviewed in[6,7]). Failure to form at least one CO per homologue pair risks the formation of aneuploid gametes and thus the process of CO formation is highly regulated. Within many organisms, including *S. cerevisiae*, *M. musculus*, *H. sapiens* and *A. thaliana*, two subclasses of CO co-exist. Interfering ZMM (Zip2-Zip3-Zip4-Spo16, Msh4-Msh5, Mer3)-dependent class I COs, which account for

31 the majority of COs formed (~70-85% within S. cerevisiae), are dispersed evenly along each chromosome by
32 means of CO interference—a process that suppresses the formation of COs in proximity to one another. A
33 subpopulation of recombination events alternatively resolve as non-interfering class II COs dependent upon
34 Mus81-Mms4, Yen1 or Slx1-Slx4. Homologous recombination, the process responsible for CO formation,
35 requires a repair template with near-perfect homology[8]. By contrast, recombination between polymorphic,
36 homoeologous substrates is markedly inefficient, leading to reduced rates of meiotic CO formation, reduced spore
37 viability and increased chromosomal non-disjunction during meiosis I within hybrid strains of *S. cerevisiae*—
38 phenotypes linked to incipient speciation and which are largely reversed within MMR-deficient strains[1–4]. Despite
39 characterization of this anti-recombinogenic activity, a detailed analysis of how MMR alters meiotic recombination
40 on a genome-wide level has not yet been achieved.
41
42 In order to investigate the impact sequence divergence has upon CO formation, we mapped recombination (Fig.
43 1a, Methods) within six wild type and 13 MMR-defective *msh2*Δ meioses—obtained from a cross of two widely
44 utilized laboratory isolates: S288c and SK1[3,9] (~65,000 SNPs, ~4,000 high confidence INDELs, ~0.57%
45 divergence). Additionally, we re-analyzed datasets comprising 51 wild type[10,11] and four *msh2*Δ tetrads[12] from a
46 S96 x YJM789 cross of *S. cerevisiae* (~0.6% divergence). On average, we identified 74.3 ±5.4 and 105.9 ±7.8
47 COs per meiosis within our SK1 x S288c wild type and *msh2*Δ samples respectively, corresponding to a significant
48 1.4-fold increase in CO frequency ($p<0.01$; Two-sample T-test) (Fig. 1b). A significant *msh2*Δ-dependent increase
49 (~1.25-fold, $p<0.01$; Two-sample T-test) is also observed within S96 x YJM789 (Fig. 1b)—collectively reaffirming
50 the known anti-recombinogenic activity of Msh2[1–5]. Notably, CO frequencies are considerably higher within S96 x
51 YJM789 than S288c x SK1 (91.4 vs. 74.3 COs per wild type meiosis)—suggesting that cross-specific differences
52 may exist.
53
54 To investigate any possible effect of Msh2 on CO patterning, we calculated the distribution of inter-crossover
55 distances (ICDs)—the separation (in bp) between successive COs along every chromosome. To accommodate
56 comparisons between different sample sizes, ICDs are transformed (Methods and Supplementary Fig. 1) and
57 visualized in rank order as empirical cumulative distribution functions (eCDFs). CO interference—observed as a
58 statistically significant ($p<0.01$; Two-sample Kolmogorov-Smirnov (KS) test) right-ward skew away from conditions
59 of simulated independence—is present within all mapped strains (Fig. 1c—f) and is stronger within S96 x YJM789
60 than within S288c x SK1, further suggesting that the meiotic CO landscape is regulated in a cross-specific manner

(Supplementary Fig. 2 and Supplementary Discussion). Unexpectedly, inactivation of Msh2 within both hybrid crosses (Fig. 1c—d) further skews ICD curves to the right ($p < 0.01$; Two-sample KS test)—indicating increased CO interference. Mutation of a second MMR factor, *PMS1*, which acts downstream of Msh2 mismatch binding, caused similar increases in both CO frequency and CO interference relative to wild type (Supplementary Fig. 3a—b), but no additive effect in a double *msh2Δpms1Δ* strain was observed—suggesting that these phenotypes arise as a general consequence of MMR inactivation.

To account for any impact increased CO frequency may have upon CO distribution, we utilized an *ndt80AR* strain where meiotic prophase length is extended via temporary repression of the Ndt80 transcription factor[13,14]. On average, we identified 94.5 ±16.3 COs per meiosis within *ndt80AR*—a significant increase relative to wild type ($p<0.01$; Two-sample T-test) (Fig. 1b). However, no further increase occurs upon inactivation of Msh2 (*msh2Δndt80AR*, 97.5 ±15.4 COs, $p=0.87$; Two-sample T-test) (Fig. 1b). Importantly, despite the lack of change in CO frequency, the *msh2Δ*-dependent skew in CO distribution was still observed (Fig. 1e) whereas increased CO frequency alone (*ndt80AR*) does not alter CO distribution compared to wild type (Fig. 1f) ($p=0.91$; Two-sample KS test). Inactivation of MMR within hybrid *S. cerevisiae* strains therefore gives rise to two distinct phenotypes relative to wild type: (i) increased CO frequency, as previously observed[1–4] and (ii) a novel, paradoxical increase in CO interference—that can arise independently of changes to CO frequency.

A global increase in CO interference, as observed within *msh2Δ*, may arise when the balance between class I and class II CO formation is altered. CO distributions are typically modelled by the gamma (γ) distribution[15,16], where γ(α) values >1.0 indicate increasing strength of interference. However, we note that models based on a single (γ) distribution deviate substantially from experimental data, which contains significantly more ICDs <50kb than expected—indicative of a randomly distributed subpopulation of COs (i.e. class II) (Fig. 2a). To solve and investigate this problem, we developed computational methods based on (γ)-mixture modelling (Fig. 2b), to statistically deconvolute ICD data—obtaining estimates of the class I:class II ratio (Methods and Supplementary Fig. 4). A non-random ($\gamma(\alpha_I)$ >1.5), and a random ($\gamma(\alpha_{II})$ ~1.0) component was successfully identified for all genotypes. Importantly, simulations based on the obtained ratios improve model fit and eliminate the observed ICD deviations below 50kb (Fig. 2c). Consistent with prior estimates[17], *MSH2* wild type class I:class II ratios are estimated at ~2.0 and ~3.0 within S288c x SK1 and S96 x YJM789 respectively (Fig. 2d). By contrast, deletion of *MSH2* increases ratio estimates to ~5-6 (Fig. 2d), suggesting a ~1.7-fold increase in class I CO formation in the

91 absence of Msh2 (Fig. 2e). Corresponding *msh2*Δ-dependent decreases in class II CO formation are also
92 predicted (~0.5-0.7-fold) (Fig. 2f). Our modelling results therefore suggest that the formation of ZMM-dependent
93 class I COs is preferentially favored within Msh2-deficient cells and that the vast majority of additional COs
94 observed within *msh2*Δ relative to wild type are class I.
95
96 Zip3 foci specifically mark the sites of interfering, class I COs[18]. In order to verify the results of our modeling, we
97 counted Zip3 foci within spread, pachytene-arrested hybrid cells (Fig. 3a—b). On average, we identified 34.1-34.7
98 and 51.7-58.7 foci per meiosis (7-8h timepoints) within *ndt80AR* and *msh2*Δ*ndt80AR* respectively, corresponding
99 to a significant 1.5-fold increase in class I CO formation upon inactivation of MMR ($p<0.01$; Two-sample T-test).
100 Such an increase in Zip3 foci count is in agreement with the predictions of our model—strongly implicating Msh2
101 as a regulator of the class I:class II decision within *S. cerevisiae*.
102
103 MMR machinery forms a potent barrier to homoeologous recombination[1–5]. To understand the interplay between
104 DNA mismatches and CO formation, we calculated polymorphism densities (SNPs, INDELs) ±500 bp around
105 every mapped CO (Fig. 4a) and compared between genotypes. To generate a comparative reference, the
106 expected environment for meiotic recombination, as defined by the polymorphism density surrounding ~3600
107 recombination hotspot midpoints[19], was also calculated. Polymorphism density surrounding COs within *MSH2*
108 wild type strains (S288c x SK1: wild type, *ndt80AR*) is significantly different to expected in both distribution
109 ($p<0.01$; Two-sample KS test) and in mean variant density (5.32 vs. 6.18, $p<0.01$; Two-sample T-test)—
110 characterized by a skew towards COs arising within regions of lower genetic divergence (Fig. 4b). By contrast,
111 *msh2*Δ COs display statistical similarity to expected ($p>0.25$; Two-sample KS-test) (6.26 vs. 6.18, $p = 0.52$; Two-
112 sample T-test) (Fig. 4b), and such a disparity between wild type and *msh2*Δ is recaptured within the independent
113 S96 x YJM789 cross (Fig. 4c). A similar pattern was observed when considering polymorphism density arising
114 ±1000 bp around each CO but was diminished with increasing distance (±2000 bp)—revealing DNA mismatches
115 to exert a local effect on CO formation (Supplementary Fig. 5a—b and Supplementary Discussion).
116
117 Inactivation of *MSH2* increases CO interference by altering the class I:class II balance (Fig. 2, Fig. 3). To determine
118 whether the role that Msh2 plays in regulating CO outcome is related to its role in suppressing COs at sites where
119 mismatches will arise during strand invasion, we further calculated polymorphism densities within mutants that
120 disrupt class I (*zip3*Δ, *msh4*Δ) or class II (*mms4*Δ) CO formation[12]. Strikingly, mutants devoid of class I COs

121 phenocopy *msh2*Δ—that is, COs within these mutants are no longer skewed away from regions of higher
122 polymorphism density despite the presence of Msh2 and share mean densities that are not statistically dissimilar
123 to expected (p>0.5; Two-sample T-test) (Fig. 4b—c). Moreover, the impact of *zip3*Δ and *msh2*Δ appears to be
124 epistatic rather than additive (Fig. 4b), indicating a shared pathway. By contrast, removal of class II formation
125 (*mms4*Δ) has no impact on the interplay between CO formation and polymorphism density (Fig. 4c)—collectively
126 suggesting that mismatch-dependent repression of CO formation is class I specific.

127

128 **Discussion**

129 Sequence divergence suppresses recombination within a wide range of eukaryotes including *M. musculus* and
130 *H. sapiens*[20–22]. Findings presented here expand upon these observations and suggest that the anti-
131 recombinogenic activity of Msh2, exerted at homoeologous sites, does not mediate an indiscriminate suppression
132 of COs but rather acts preferentially at sites of ZMM-dependent, interfering COs—thereby altering the spatial
133 distribution of recombination across the genome by modulating the class I:class II balance (Fig. 4d) and
134 highlighting how even low rates of divergence, present within intra-specific hybrids of *S. cerevisiae*, can profoundly
135 impact upon meiosis. Nevertheless, COs frequently encounter heterologous regions (Fig. 4b—c) and yet, despite
136 this, class I COs still form at a high rate in wild type (67%-75%) (Fig. 2d). Mismatches are therefore unlikely to form
137 an absolute barrier to interfering COs, but rather influence the *probability* of formation. In our envisioned model
138 (Fig. 4d), rejected class I events are redirected toward the NCO pathway, via synthesis-dependent strand
139 annealing (SDSA)[9,23] and double Holliday-junction (dHJ) dissolution[9,24], or toward inter-sister recombination, which
140 remains invisible in our assay.

141

142 Inhibition of homoeologous recombination, by means of heteroduplex rejection, relies upon Msh2 and the RecQ-
143 family helicase, Sgs1[5,25,26]. An *sgs1*Δ mutant may therefore be expected to phenocopy *msh2*Δ if suppression of
144 class I COs occurs via this mechanism. However, the strength of CO interference is reduced in *sgs1*Δ relative to
145 wild type (Supplementary Fig. 3c) and the increase in CO interference, observed upon the inactivation of Msh2,
146 requires Sgs1 (Supplementary Fig. 3d)—suggesting that Msh2 and Sgs1 are not epistatic, but rather antagonistic
147 in the formation of class I COs. Thus, unexpectedly, Msh2 appears to mediate suppression of class I COs in a
148 pathway different to that of Sgs1-mediated heteroduplex rejection, instead relying upon the downstream
149 properties or factors of MMR, including Pms1 (Supplementary Fig. 3a), to achieve its effect.

150

Exactly how MMR-specificity for class I COs arises remains unclear. *In vitro* data suggests that Mlh1-Mlh3—essential class I CO factors—facilitate binding of Msh2 to heteroduplex DNA arising at sites where mismatches exist between parental strains and are in turn stimulated by Msh2[27]. The distinct genetics of each CO subclass may therefore be responsible for their differential sensitivity to sequence mismatch, through preferential recruitment or activation of Msh2 and other MMR components at class I sites. Interestingly, large-scale regions of heterozygosity promote CO formation within *A. thaliana*[28]—proposed to arise from either increased recombination due to mismatch-dependent delay in CO maturation, or increased sensitivity of mismatched events to crossover designation[28]. Given the evolutionary conservation of MMR and class I CO formation[6,29], our results in *S. cerevisiae* may distinguish between these mechanisms, suggesting the former is more likely.

Understanding the molecular mechanisms that contribute to speciation is fundamental to our understanding of biological diversity and evolution. As class I CO formation is specifically required for correct meiosis I disjunction[6], our observations provide a mechanistic insight into how hybrid sterility may arise. Specifically, the activity of Msh2 has the potential to jeopardize chromosome segregation, risking the formation of aneuploid gametes and rendering hybrids derived from distantly related individuals sub- or infertile—contributing to the sexual isolation of a population. MMR may therefore not only influence the rates of evolution, favoring more gradual changes to the gene pool by limiting the amount of genetic exchange that may occur between divergent homologues, but also serve to promote speciation over evolutionary time.

**Contributions**

T.J.C, M.C. and M.J.N. conceived of the project. T.J.C. analyzed and interpreted the data, performed in-silico simulations and designed the mixture modelling algorithm. M.C. performed all genome-wide mapping and event calling. L.J.H. performed all microscopy and foci analysis. M.M.K. and B.L. provided scripts, protocols and additional samples. T.J.C. and M.J.N. wrote the manuscript.

**Competing interests**

The authors declare no competing financial interests.

**Corresponding author**

Correspondence to Matthew J. Neale.

211 **References**


212 1. Greig, D., Travisano, M., Louis, E. J. & Borts, R. H. A role for the mismatch repair system during incipient
213 speciation in Saccharomyces. *J. Evol. Biol.* **16,** 429–37 (2003).
214 2. Hunter, N., Chambers, S. R., Louis, E. J. & Borts, R. H. The mismatch repair system contributes to
215 meiotic sterility in an interspecific yeast hybrid. *EMBO J.* **15,** 1726–33 (1996).
216 3. Martini, E. *et al.* Genome-wide analysis of heteroduplex DNA in mismatch repair-deficient yeast cells
217 reveals novel properties of meiotic recombination pathways. *PLoS Genet.* **7,** e1002305 (2011).
218 4. Chambers, S. R., Hunter, N., Louis, E. J. & Borts, R. H. The mismatch repair system reduces meiotic
219 homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol. Cell. Biol.*
220 **16,** 6110–20 (1996).
221 5. Spies, M. & Fishel, R. Mismatch Repair during Homologous and Homeologous Recombination. *Cold*
222 *Spring Harb. Perspect. Biol.* **7,** a022657 (2015).
223 6. Gray, S. & Cohen, P. E. Control of Meiotic Crossovers: From Double-Strand Break Formation to
224 Designation. *Annu. Rev. Genet.* **50,** 175–210 (2016).
225 7. Berchowitz, L. E. & Copenhaver, G. P. Genetic interference: don't stand so close to me. *Curr. Genomics*
226 **11,** 91–102 (2010).
227 8. Heyer, W.-D., Ehmsen, K. T. & Liu, J. Regulation of Homologous Recombination in Eukaryotes. *Annu.*
228 *Rev. Genet.* **44,** 113–139 (2010).
229 9. Marsolier-Kergoat, M.-C., Khan, M. M., Schott, J., Zhu, X. & Llorente, B. Mechanistic View and Genetic
230 Control of DNA Recombination during Meiosis. *Mol. Cell* **70,** 9–20.e6 (2018).
231 10. Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L. M. High-resolution mapping of meiotic
232 crossovers and non-crossovers in yeast. *Nature* **454,** 479–485 (2008).
233 11. Liu, Y. *et al.* Down-Regulation of Rad51 Activity during Meiosis in Yeast Prevents Competition with
234 Dmc1 for Repair of Double-Strand Breaks. *PLoS Genet.* **10,** e1004005 (2014).
235 12. Oke, A., Anderson, C. M., Yam, P., Fung, J. C. & Fishel, R. Controlling Meiotic Recombinational Repair –
236 Specifying the Roles of ZMMs, Sgs1 and Mus81/Mms4 in Crossover Formation. *PLoS Genet.* **10,**
237 e1004690 (2014).
238 13. Benjamin, K. R., Zhang, C., Shokat, K. M. & Herskowitz, I. Control of landmark events in meiosis by the
239 CDK Cdc28 and the meiosis-specific kinase Ime2. *Genes Dev.* **17,** 1524–39 (2003).
240 14. Xu, L., Ajimura, M., Padmore, R., Klein, C. & Kleckner, N. NDT80, a meiosis-specific gene required for

exit from pachytene in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **15,** 6572–81 (1995).

15. Chen, S. Y. *et al.* Global Analysis of the Meiotic Crossover Landscape. *Dev. Cell* **15,** 401–415 (2008).
16. Anderson, C. M., Oke, A., Yam, P., Zhuge, T. & Fung, J. C. Reduced Crossover Interference and Increased ZMM-Independent Recombination in the Absence of Tel1/ATM. *PLOS Genet.* **11,** e1005478 (2015).
17. de los Santos, T. *et al.* The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics* **164,** 81–94 (2003).
18. Zhang, L. *et al.* Topoisomerase II mediates meiotic crossover interference. *Nature* **511,** 551–556 (2014).
19. Pan, J. *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144,** 719–31 (2011).
20. Cole, F., Keeney, S. & Jasin, M. Comprehensive, Fine-Scale Dissection of Homologous Recombination Outcomes at a Hot Spot in Mouse Meiosis. *Mol. Cell* **39,** 700–710 (2010).
21. Baudat, F. & de Massy, B. Cis- and trans-acting elements regulate the mouse Psmb9 meiotic recombination hotspot. *PLoS Genet.* **3,** e100 (2007).
22. Jeffreys, A. J. & Neumann, R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* **14,** 2277–2287 (2005).
23. Allers, T. & Lichten, M. Intermediates of yeast meiotic recombination contain heteroduplex DNA. *Mol. Cell* **8,** 225–31 (2001).
24. Bizard, A. H. & Hickson, I. D. The dissolution of double Holliday junctions. *Cold Spring Harb. Perspect. Biol.* **6,** a016477 (2014).
25. Sugawara, N., Goldfarb, T., Studamire, B., Alani, E. & Haber, J. E. Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. *Proc. Natl. Acad. Sci.* **101,** 9315–9320 (2004).
26. Goldfarb, T. & Alani, E. Distinct roles for the Saccharomyces cerevisiae mismatch repair proteins in heteroduplex rejection, mismatch repair and nonhomologous tail removal. *Genetics* **169,** 563–74 (2005).
27. Rogacheva, M. V. *et al.* Mlh1-Mlh3, a Meiotic Crossover and DNA Mismatch Repair Factor, Is a Msh2-Msh3-stimulated Endonuclease. *J. Biol. Chem.* **289,** 5664–5673 (2014).
28. Ziolkowski, P. A. *et al.* Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. *Elife* **4,** (2015).

29. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18,** 85–98 (2008).

**Methods/Supplementary References**

30. Kane, S. M. & Roth, R. Carbohydrate metabolism during ascospore development in yeast. *J. Bacteriol.* **118,** 8–14 (1974).

31. Mortimer, R. K. & Johnston, J. R. Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113,** 35–43 (1986).

32. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14,** 953–961 (1998).

33. Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast* **15,** 1541–1553 (1999).

34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

35. Engel, S. R. & Cherry, J. M. The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple Saccharomyces cerevisiae strains at the Saccharomyces Genome Database. *Database (Oxford).* **2013,** bat012 (2013).

36. Van der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. in *Current Protocols in Bioinformatics* **43,** 11.10.1-11.10.33 (John Wiley & Sons, Inc., 2013).

37. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

38. Zhang, L., Liang, Z., Hutchinson, J. & Kleckner, N. Crossover Patterning by the Beam-Film Model: Analysis and Implications. *PLoS Genet.* **10,** e1004042 (2014).

39. Do, C. B. & Batzoglou, S. What is the expectation maximization algorithm? *Nat. Biotechnol.* **26,** 897–899 (2008).

40. Batten, L. M. & Beutelspacher, A. *The theory of finite linear spaces : combinatorics of points and lines*. (Cambridge University Press, 1993).

41. Massey, F. J. & Jr. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **46,** 68 (1951).

42. Miller, L. H. Table of Percentage Points of Kolmogorov Statistics. *J. Am. Stat. Assoc.* **51,** 111 (1956).

43. Shinohara, M., Oh, S. D., Hunter, N. & Shinohara, A. Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. *Nat. Genet.* **40,** 299–309 (2008).

44. Herbert, A. D., Carr, A. M. & Hoffmann, E. FindFoci: A Focus Detection Algorithm with Automated

Parameter Training That Closely Matches Human Assignments, Reduces Human Inconsistencies and Increases Speed of Analysis. *PLoS One* **9,** e114749 (2014).

45. Al-Sweel, N. *et al. Mlh3 Mutations in Baker'S Yeast Alter Meiotic Recombination Outcomes By Increasing Noncrossover Events Genome-Wide*. *PLoS Genetics* **13,** (2017).

361 **Methods**

362 **Yeast Strains**

363 All Saccharomyces cerevisiae strains used in this study are derivatives of SK1[30] and S288c[31]. Hybrid strains,
364 utilised in genome-wide mapping, were derived from a cross of haploid SK1 and S288c. Strain genotypes are
365 detailed in (Supplementary Table 1). Knockouts were performed and tested by standard transformation and PCR
366 techniques[32]. *msh2*Δ::*kanMX6* and *zip3*Δ::*HphMX* were generated by PCR mediated gene replacement using a
367 pFA6a-*kanMX6* or pFA6-*hphMX* plasmid[33]. The $P_{GAL}$-*NDT80::TRP1* allele has the natural *NDT80* promoter
368 replaced by the *GAL1-10* promoter, and strains include a *GAL4::ER* chimeric transactivator for β-estradiol-induced
369 expression[13]. S288c x SK1 hybrids create viable spores (91.98% WT, 72.99% *msh2*Δ spore viability, data not
370 shown), eliminating any observational bias that may arise from assaying a limited, surviving population *[Crawford*
371 *et al. 2018, DOI pending]*.

372

373 **Meiotic Timecourse (*ndt80AR* strains)**

374 Diploid strains were incubated at 30°C on YPD plates for 48h. For SK1 diploids, a single colony is inoculated into
375 4 mL YPD (1% yeast extract, 2% peptone, 2% glucose) and incubated at 30°C at 250 rpm for 24 h. For hybrid
376 crosses, haploid parental isolates are mated in 1 mL YPD for 8 h. An additional 3 mL of YPD is subsequently
377 added and the cells are grown for 16 h. Cells are inoculated to a density of (OD600) 0.2 into 30 mL YPA (1%
378 yeast extract, 2% peptone, 1% K-acetate) and incubated at 250 rpm at 30°C for 14h. Cells are collected by
379 centrifugation, washed in $H_2O$, and resuspended in 30mL pre-warmed sporulation media (2% potassium acetate,
380 5 µg/mL Adenine, 5 µg/mL Arginine, 5 µg/mL Histidine, 15 µg/mL Leucine, 5 µg/mL Tryptophan, 5 µg/mL Uracil).
381 The culture is then incubated at 30°C at 250 rpm for the duration of the time course. After 8h, 2mL of the
382 synchronised cultures were split and exposed to β-estradiol to a final concentration of 2mM, which induces the
383 transcription of NDT80 and thus sporulation. Cultures were then incubated to a total of 48 h at 30°C prior to
384 dissection.

385

386 **Tetrad Dissection**

387 In order to produce hybrid spores for sequencing, SK1 x S288c haploid parents were mated for 8-14 h on YPD
388 plates, with the exception of *ndt80AR* strains, which are mated and grown in liquid YPD for 24 h (see above).
389 Haploids are mated freshly on each occasion and not propagated as diploids, in order to reduce mitotic
390 recombination. Sporulation is induced, and tetrads were dissected after 72 h in 2% potassium acetate. For octads,

391 spores are additionally grown for 4-8 h on YPD plates until a single mitotic division is completed, after which the
392 mother-daughter pair are separated. Colonies were grown for 16 h within liquid YPD for genomic DNA extraction.
393 Only tetrads and octads producing four or eight viable spores/colonies, respectively, are considered for
394 genotyping by NGS.
395
396 **NGS Library Preparation**
397 Genomic DNA is purified from overnight, saturated YPD cultures using standard phenol-chloroform extraction
398 techniques. Samples of genomic DNA are diluted to 0.2-0.3 ng/µL. DNA concentration is measured using the
399 Qubit High Sensitivity dsDNA Assay. Genomic DNA is fragmented, indexed and amplified via the Nextera XT DNA
400 library Prep Kit according to the best practices recommended by Illumina. In order to check fragment length
401 distribution and concentration of purified libraries, 1 µL of undiluted library is run on an Agilent Technology 2100
402 Bioanalyzer using a High Sensitivity DNA chip. To pool samples for sequencing, 5 µL of each sample is combined
403 into a 1.5 mL tube and mixed. 24 µL of the mix is transferred to a tube containing 570 µL hybridisation buffer. The
404 mix is boiled at 96°C for 2 minutes and placed in ice water for 5 minutes. 6 µL of denatured PhiX control (prepared
405 according to Illumina protocol, final concentration 1%) is added to the library, mixed well and then loaded into a
406 MiSeq reagent cartridge. Sequencing was performed in-house using Illumina MiSeq instruments.
407
408 **Alignment, SNP and indel Detection**
409 Individual spores were sequenced to an average read-depth of ~45x. Initially, paired-end read FASTQ files are
410 aligned, via Bowtie2[34], to the SacCer3 reference genome (v. R64-2-1)[35] using the parameters: -X 1000 —local —
411 mp 5,1 -D 20 -R 3 -N 1 -L 20 -i S,1,0.50. In order to create a custom SK1 genome to facilitate more accurate
412 genotype-calling, SNP and indel polymorphisms were detected using the GATK (GenomeAnalysisToolkit) function
413 *HaplotypeCaller*[36]. An in-house script (*VariantCaller.pl*) subsequently parses the resulting VCF files from 72 spores
414 to calculate: (i) the call frequency (% of spores any given allele is present within), (ii) the cumulative allelic read
415 depth (% of reads that contain a specific allele at a specific loci), and (iii) the cumulative total read depth. To identify
416 legitimate SNPs and indels, variants were filtered for a call-frequency between 44-55%, a total read depth of >250
417 and an allelic read depth of 95%. Variants within repeat regions, long terminal repeats, retrotransposons and
418 telomeres were also discarded—yielding a final, robust list of 64,591 SNPs and 3972 indels amounting to ~0.57%
419 divergence. A custom SK1 genome (SK1_Mod) is then generated by modifying SacCer3 (v. R64-2-1) to include
420 all filtered/called SNPs and indels.

421 **Genotype-Calling**

422 Spore data from individual samples was aligned to the both the custom SK1_Mod genome and the SacCer3
423 reference (see below). Alignment produces a SAM file, which is converted into a sorted BAM file using the
424 Samtools function, *view*[37], for downstream processing. The PySamStats (v. 1.0.1, Miles & Mattioni) module,
425 *variation*, is used to process the sorted BAM file for each sequenced spore, producing a list of the number of
426 reads containing A/C/T/G, insertion or deletion for each genomic position specified in the S288c and SK1
427 references. Variant reads are isolated and genotyped using in-house, custom scripts as follows. Genotypes are
428 assigned according to the rules: (i) A minimum read-depth of 5; (ii) A SNP is called as having the variant genotype
429 if >=70% of the reads at that position match the called variant, or as reference if =>90% of the reads match the
430 reference; (iii) If the variant and reference reads are above 90% of all reads and within 70% of each other, the
431 position is called as heteroduplex; (iv) indels are called as having the variant genotype if >=30% of the reads at
432 that position match the variant. Such a low threshold is utilised because alignment of indel sequences is biased
433 towards the reference, which means that they are unlikely to be erroneously called as matching the variant
434 genotype. For an indel to be called as the reference genotype, >=95% of the reads must match the reference
435 sequence and there must be fewer than 2 reads matching the variant call. Any variants that fall below these
436 thresholds are discarded. Genotype calls are converted into a binary signal, either 1 for S288c or 0 for SK1.

437

438 **Event Calling**

439 Using the binarised input, chromosomes are split into segments with the same segregation pattern using scripts
440 available in[3,9]. Segment types (i.e. 1:7, 2:6, 2:6, 3:5, 4:4, 4:4*, 5:3, 6:2, 6:2* or 7:1 as described in[9] are also
441 recorded. Recombination events are subsequently called as being a set of segments located between two 4:4
442 segments longer than 1.5 kb[9]. A 4:4 segment corresponds to a Mendelian segregation profile, 5:3 and 3:5
443 segments to half-conversion tracts and 6:2 and 2:6 segments to full conversion tracts. Each recombination event
444 can contain between 0-2 COs or NCOs. Events are additionally classified by the number of chromatids involved
445 (i.e. 1, 2, either sister or non-sister, 3, 4).

446

447 **Event Position & Inter-Crossover Distances (ICDs)**

448 Crossover position, or "midpoint", is defined as the distance between the mid-points of the first and last SNP/indel
449 markers—an estimate of true event tract length. Inter-crossover distances (ICDs) are then calculated as the
450 distance (in bp) between successive CO midpoints.

451 **Simulating Meiosis**

452 Randomised or mixed (class I + class II) ICD simulations were performed using a novel simulation platform
453 (*RecombineSim*) built in MATLAB 2018a. A typical simulation run is depicted in (Supplementary Fig. 4a). In brief,
454 virtual chromosomes are constructed as binned, numerical arrays at a 100 bp resolution adjusted to reflect the
455 limit of experimental detection governed by the leftmost and rightmost genetic markers (SNPs/indels). Any given
456 100 bp bin possesses a numerical recombination potential (recom(P)), which governs the ability of an interfering
457 CO to successfully form at that site. Class I COs impose a zone of "interference", by altering recom(P) values in
458 adjacent bins in a distance-dependent manner—a similar principle to the beam-film model of CO interference[38].
459 The exact shape and width of interference imposed is determined by the best fit (γ) parameters (α,β) for the
460 genotype currently being simulated—and applied as a hazard function (EQN 1.1):

461 $$h(x) = \frac{PDF(x)}{1-CDF(x)} \quad (1.1)$$

462 A hazard function describes the probability that, given a pre-existing CO at position x(0), another CO will form at
463 any given distance (x) away[15]—and thus is a natural representation of interference. A fractional amount of class II
464 COs that remain insensitive to recom(P) are introduced via the $C_{PROB}$ parameter where necessary. All simulations
465 (N = 10,000 cells) precisely match the experimentally observed event count for any given genotype or cell. Virtual
466 event merging, at the 1.5 kb threshold, is included in order to closely match the in vivo data.

467

468 **(γ)-Mixture Modelling**

469 Distributional analysis of CO distributions is complicated by the existence of non-interfering, Mus81-Mms4 class
470 II COs—indistinguishable from interfering ZMM-dependent class I COs in our assay. In essence, meiotic ICDs
471 represent a heterogenous, mixed system (Fig. 2b) with unknown quantities of each subclass. Latent variables (e.g.
472 class II CO %) may, however, be inferred through probabilistic and statistical methods. Expanding upon the use
473 of the gamma (γ) distribution to model meiotic ICDs[15,16], experimental data was deconvoluted by fitting two (γ)
474 distributions—one for each subclass of CO—via an expectation maximisation (EM) algorithm (MATLAB 2018a).
475 EM is a commonly applied method for iterative clustering and parameter estimation in mixed models[39]. Briefly, any
476 given ICD is assigned a probability reflective of how likely it is to belong to one of the two sub-distributions.
477 Subsequently, sub-distributions are iteratively shifted and data point identity is reassigned until a maximum
478 likelihood (ML) solution is converged upon. One (γ) distribution is expected to yield a final (γ)(α) value of ~1.0 (class
479 II, random), while the other is expected to produce a (γ)(α) value of >1.5 (class I, non-random), with their relative
480 contributions to the overall mixture (i.e. the class I:class II ratio) is dependent upon genotype. In order to validate

481 this approach, simulated ICD datasets of two component mixtures with known parameters, at variable sample
482 sizes (S), were generated using *RecombineSim* and deconvoluted (Supplementary Fig. 4b). As a measure of
483 accuracy, the average % difference (N(%Δ)) between estimated and actual parameters was calculated. Accuracy
484 is dependent upon sample size (S) and to a lesser extent on the relative proportions of each subpopulation—and
485 thus how likely a subpopulation is to be readily observed within the mixed population. For example, (γ) mixtures
486 containing 10 or 25% class II COs exhibit average errors of 10.0% and 9.1% at (S) = 500 and 4.53% and 3.73%
487 at (S) = 2000 respectively (Supplementary Fig. 4c). Experimental datasets range from (S) values of 354 to 3365,
488 therefore reasonable error rates of ~<10% were expected. The ability to mathematically separate CO subclasses
489 may be of great use to assess, in more detail, the mixed composition of meiotic CO formation present in both
490 published and future datasets of this type—enabling further identification of those components that directly
491 influence the CO interference mechanism.
492

493 **ICD Transformation**

494 The formation of a variable number of events (N) within a finite space (*lim*) (i.e. a chromosome or genome length)
495 skews CDFs i.e. a higher CO frequency causes a downward shift in ICD size. An ICD distribution produced under
496 identical spatial rules but with a different event count would therefore generate significantly different CDFs—failing
497 or biasing statistical testing and undermining the ability to assess distributional agreement. This skew can be
498 readily observed using simulated data (Supplementary Fig. 1a). Notably, higher values of (N) cause a leftward
499 skew. The relationship between (N) and ICD size for a given *lim* is, however, linear[40]. Consequently, in order to
500 isolate the distributional identity of any given sample (i.e. isolate γ(α) from γ(β)), ICD data can be transformed by
501 calculating the product of ICD size (ICD x event count). Data transformation results in perfectly aligned CDFs
502 despite varying (N), validating this approach (Supplementary Fig. 1b).
503

504 **Statistical Analyses**

505 **Kolmogorov-Smirnov Test**

506 A *Kolmogorov-Smirnov* goodness-of-fit (GoF) test is a non-parametric test used to compare continuous
507 probability distributions in order to assess the null hypothesis that both samples derive from identical populations,
508 based on their maximal difference ($D_{KS}$)[41,42]. (P) values of the KS-test effectively describe the probability that, if the
509 null hypothesis is true, the observed CDFs would be as far apart as observed. (P) values may therefore constitute

510 an indirect measure of distributional agreement, as employed throughout this paper. KS-tests were performed
511 using the MATLAB 2018a packages: *kstest* and *kstest2*.
512
513 **Two-sample T-tests**
514 A two-sample T-test is utilised to determine whether a difference in mean value is significant or has arisen by
515 chance. Two-sample T-Tests were performed using the MATLAB 2018a package: *ttest2*.
516
517 **Microscopy & Cytological Analysis**
518 4.5 mL of meiotic culture is spun down on a bench centrifuge and resuspended to 500 µL with 1M pH 7.0 D-
519 Sorbitol. 12 µL of 1.0 M DTT and 7 µL of 10 mg/mL Zymolyase in 10% glucose solution is added and cells are
520 spheroplasted by incubation at 37°C for 35-50m with agitation. Spheroplasting success is determined by taking
521 2-3 µL of the aforementioned solution and adding an equivalent volume of 1.0% (w/v) Sodium N-Lauroylsarcosine
522 while under microscopic observation. Cells should immediately lyse as the exposed membrane is disrupted by
523 the detergent. 3.5 mL of Stop Solution (0.1M MES, 1 mM EDTA, 0.5 mM $MgCl_2$, 1M D-Sorbitol, pH 6.4) is
524 subsequently added and the cells are spun down to be resuspended in 100 µL Spread Solution (0.1M MES, 1
525 mM EDTA, 0.5 mM $MgCl_2$, pH 6.4) and distributed between 4 slides, which are soaked in 70% EtOH overnight
526 and wiped clean. To each slide, fixative (4.0% (w/v) Formaldehyde, 3.8% (w/v) Sucrose, pH7.5) is added dropwise,
527 followed by detergent (1% Lipsol, 0.1% Bibby Sterilin) to a ratio of 1:3:6 (suspension : fixative : detergent) before
528 lightly mixing and incubating for 1m at room temperature (RT). Further fixative is added dropwise to a final ratio of
529 1:9:6 and the mixture spread across the slide. Each spread is subsequently incubated at RT for 30m in damp
530 conditions, then allowed to air-dry at RT overnight. Once dry, slides are sequentially washed in 0.2% (v/v) PhotoFlo
531 Wetting Agent (Kodak) and $dH_2O$, and stored at 4°C.
532
533 Slides are washed once in 0.025% Triton X-100 for 10m at RT and twice in PBS for 5m at RT. Slides are blocked
534 in 5% skimmed milk with PBS for 3h at 37°C. Excess liquid is removed and slides laid horizontally in damp
535 conditions. 40 µL of primary antibody (anti-Zip3[43] from rat at 1:200 and anti-Red1 (Genecust, affinity purified,
536 raised against aa(426-827)) from rabbit at 1:200) in 1% skimmed milk with PBS is added under coverslips. Slides
537 are incubated at 4°C overnight (15.5h) and washed three times in PBS for 5m at RT. Excess liquid is then removed
538 and slides are returned to damp conditions. 40 µL of secondary antibody (anti-rat AlexaFluor555 at 1:200 and
539 anti-rabbit AlexaFluor488 at 1:500) in 1% skimmed milk with PBS is added under coverslips. Slides are incubated

at RT for 2.5h and then washed three times with PBS for 5m at RT. Cover slips are affixed using Vectashield mounting medium with DAPI, sealed with clear varnish and imaged on an Olympus IX71 (z = 0.2μM, Exposure times: TRITC-mCherry = 0.2s, eGFP = 1.0s, DAPI = 0.1s). Images were randomised, deconvoluted via Huygens (software) and foci were automatically counted using an in-house plugin for ImageJ (FindFoci) as previously described[44], with an appropriate mask to discard signals outside of nuclei.

**Data Availability**

Raw sequence data is deposited in the NIH Sequence Read Archive (SRA) under accession numbers SRP152953 (*zip3*Δ), SRP151982 (wild type, *msh2*Δ, *ndt80AR*) and SRP111430 (*msh2*Δ). Scripts, tools, software and additional data are available upon request or at [GitHub repository link].

570 **Supplementary Discussion**

571 **The SK1-ML3 allele has a reduced capacity to generate CO interference**

572 As previously noted, wild type CO frequencies are higher within S96 x YJM789 (74.3 vs 97.3 COs/meiosis) than
573 S288c x SK1 (Fig. 1b). Moreover, wild type CO interference is significantly stronger within S96 x YJM789 ($p<0.01$;
574 Two-sample KS test) (Supplementary Fig. 2a) and (γ)-mixture modelling results (Fig. 2d) suggest that class I CO
575 fraction is higher in S96 x YJM789 (75% vs. 67%). Molecular incompatibilities between certain alleles of the CO
576 formation or CO interference machinery may account for these cross-specific differences. To investigate this
577 hypothesis further, we analysed the frequency and distribution of COs within a *mlh3*Δ S288c x YJM789
578 background containing a functional copy of the SK1 *MLH3* allele (*SK1-MLH3*)[45]. Surprisingly, introduction of the
579 SK1 *MLH3* allele is sufficient to reduce the global strength of CO interference and produce a CO distribution
580 identical to that of S288c x SK1 wild type ($p = 0.91$; Two-sample KS Test) (Supplementary Fig. 2b). *SK1-MLH3*
581 also results in a modest but significant ($p < 0.01$; Two-sample T-test) reduction in CO frequency (97.2 vs. 90.9
582 COs per meiosis) (Supplementary Fig. 2c). Collectively, these results suggest that, in some manner, the SK1
583 *MLH3* allele is deficient in the generation of CO interference and/or CO-formation—however, the reduction in CO
584 frequency observed within S288c x SK1 cannot be fully ascribed to SK1 *MLH3* alone. In the context of this study,
585 such a defect improves the visibility of the *msh2*Δ phenotype in S288c x SK1 (Fig. 1c—d).

586

587 **Localised impact of polymorphism density upon CO formation**

588 S288c x SK1 variants have an average density of 1 per 175 bp and a median inter-variant distance of 81 bp
589 (93.12% of inter-variant distances are <500bp) and are therefore evenly spaced and present at high density across
590 each chromosome. Maps of polymorphism density are shown in (Supplementary Fig. 5c—d) for two example
591 chromosomes. In general, S288c x SK1 chromosomes are organised into local peaks and troughs of variant
592 density while maintaining overall uniformity. Therefore, it seems unlikely that inactivation of Msh2 would result in a
593 gross-redistribution of CO formation toward any particular region of the chromosome as it may do within
594 organisms with less uniform SNP/indel density, such as *A. thaliana*[28]. To further investigate the way in which
595 polymorphisms sculpt the meiotic landscape, we repeated the analysis shown in (Fig. 4b—c) using expanded
596 ±1000bp and ±2000bp windows (Supplementary Fig. 5a—b). A Msh2-dependent and statistically significant skew
597 toward regions of lower sequence divergence ($p<0.01$; Two-sample KS test) is retained at ±1000bp but is
598 significantly diminished at ±2000bp ($p=0.51$; Two-sample KS test), suggesting that mismatched sequences have
599 the greatest impact when present within the recombination intermediate structures.

## Figure Legends

**Fig. 1 | Inactivation of Msh2 increases CO frequency and the global strength of CO interference**

**a,** Genome-wide mapping of recombination. Meiosis is induced within hybrid S288c x SK1 *S. cerevisiae* diploid cells and genomic material is prepped from individual, isolated spores for paired-end Illumina sequencing in order to genotype SNP/INDEL patterns and therefore determine the parental origin of any given loci (Methods). Only a single chromosome is shown for clarity. Inter-crossover distances (ICDs), a measure of CO interference, are calculated as the distance (in bp) between successive COs along a given chromosome. **b,** Average number of COs per meiosis for each genotype. The number of individual meioses sequenced per genotype is indicated. Error bars: 95% confidence intervals (CI). *P values: Two-sample T-test.* **c—f,** Empirical cumulative distribution functions (eCDFs) showing the fraction of ICDs at or below a given size. The total number of experimental ICDs is indicated in brackets. ICDs are transformed as per the protocol described to correct for skews generated by differing CO frequencies (Methods). Untransformed ICD plots are available in (Supplementary Fig. 1c—h). Randomized datasets were generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were performed between genotypes as indicated (triangular legend). *P values: Two-sample KS-test.*

**Fig. 2 | Computational modelling predicts a Msh2-dependent shift in the class I:class II CO ratio**

**a,** Ratio of experimentally observed ICD sizes (OBS) versus the theoretical expectation based on a single, best-fit gamma (γ)-distribution (EXP). Ratio values were calculated at 5kb intervals. **b,** Example (γ) mixture model (α1.0-β1.0 + α3.0β5.0). I = class I. II = class II. Meiotic datasets, owing to the existence of two CO subclasses, are a heterogenous population of three ICD types (as shown). **c,** As in **(a)** but based on a mixed (γ)-model (no. of distributions fitted = 2). **d,** Best-fit (γ) mixture modelling results. N = sample size (no. of ICDs). $\alpha_{.}$ = Single-fit γ(α) value. $P_{.}$ = Fit quality of a single (γ)-distribution (one-sample KS-test). αI/II = Mixed model γ(α) values. Class I/II = estimated fraction of each CO subclass. Ratio = class I:class II. $P_{..}$ = Fit quality of a mixed (γ)-mode (Two-sample KS-test) **e—f,** Estimated class I and class II CO counts respectively. Estimates were obtained using the best-fit class I:class II ratios. Total CO frequencies are overlaid (grey bar). Error bars: 95% confidence intervals (CI). The number of individual meioses sequenced per genotype is indicated.

**Fig. 3 | Zip3 foci counts are significantly elevated within Msh2-deficient cells**

**a,** Box-and-whisker plot showing Zip3 foci counts from spread, pachytene arrested, Red1-positive S288c x SK1 cells. Midlines denote median values. *P values: Two-sample T-test*. **b,** Representative example for each genotype

630 (7h timepoint). Cells are fluorescently labelled for the meiosis-specific axis protein Red1 (green), the interfering CO
631 marker, Zip3 (red) and DNA (blue). Only cells with sufficient Red1 signal, as determined by the relative intensity of
632 Red1 foci above the background, were included in the analysis. The total number of nuclei counted is indicated
633 in brackets.
634
635 **Fig. 4 | Suppression of class I COs occurs at regions of higher sequence divergence**
636 **a,** SNP/INDEL count is assayed using a ±500bp window centered on CO or DSB hotspot midpoints. All contained
637 SNP/INDELs are then subsequently tallied, with equal weight. **b—c,** Empirical cumulative distribution functions
638 (eCDFs) showing the fraction of COs that reside within a region of a given SNP/INDEL count for S288c x SK1 and
639 S96 x YJM789 respectively. Expected is calculated using DSB hotspot midpoints[19]. Pairwise goodness-of-fit tests
640 were performed between pooled *msh2*Δ and *MSH2*$^+$ datasets as indicated (triangular legend). P values: Two-
641 sample KS-test. **d,** Model summarizing the impact Msh2-dependent regulation has upon CO formation. In the
642 presence of Msh2 or a functional MMR pathway, regions of higher sequence divergence redirect repair toward
643 the NCO or inter-sister pathways. Inactivation of MMR alleviates this repression, increasing the frequency of class
644 I CO formation and thus the global strength of CO interference.
645
646 **Supplementary Fig. 1 | CO interference is present to varying degrees within all mapped strains**
647 **a,** Empirical cumulative distribution function (eCDF) showing ICD data derived from interfering simulations (γ(α) =
648 3.0) at varying CO per cell frequencies (N). **b,** As in **(a)** but ICDs are transformed as per the protocol described to
649 correct for skews generated by differing CO frequencies (Methods). **c—h,** eCDFs showing the fraction of ICDs at
650 or below a given size. The total number of experimental ICDs is indicated in brackets. Randomized datasets were
651 generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were
652 performed between genotypes as indicated (triangular legend). P values: Two-sample KS-test.
653
654 **Supplementary Fig. 2 | Cross-specific differences—the SK1-MLH3 allele has reduced capacity to mediate**
655 **CO interference**
656 **a—b,** Empirical cumulative distribution functions (eCDFs) showing the fraction of ICDs at or below a given size.
657 The total number of experimental ICDs is indicated in brackets. ICDs are transformed as per the protocol
658 described to correct for skews generated by differing CO frequencies (Methods). Randomized datasets were
659 generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were

performed between genotypes as indicated (triangular legend). A schematic of the *SK1-MLH3* strain analyzed is shown. P values: Two-sample KS-test. **c,** Average number of COs per meiosis for each genotype. P values: Two-sample T-test.

**Supplementary Fig. 3 | Mechanistic details of MMR-dependent suppression of interfering COs**

**a,** Empirical cumulative distribution function (eCDF) showing the fraction of ICDs at or below a given size. The total number of experimental ICDs is indicated in brackets. ICDs are transformed as per the protocol described to correct for skews generated by differing CO frequencies. Randomized datasets were generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were performed between genotypes as indicated (triangular legend). P values: Two-sample KS-test. **b,** Average number of COs per meiosis for each genotype. The number of individual meioses sequenced per genotype is indicated. Error bars: 95% confidence intervals (CI*). P values: Two-sample T-test.* **c—d,** As in **(a)** but for differing genotypes.

**Supplementary Fig. 4 | Modelling CO distributions.**

**a,** *RecombineSim* overview. Virtual chromosomes are constructed at a 100bp resolution as binned, numerical arrays upon which meiotic CO formation is simulated (Methods). Any given 100bp contains a value in the range of [0.0-1.0], designating its recombination potential (Rec(P)). Prior to CO formation, bins are initially populated with [1.0]—denoting an equal probability of class I CO formation. During the formation of an interfering CO, *RecombineSim* imposes CO interference as a distance-dependent zone of repression by modifying Rec(P) values—influencing the position of all subsequent interfering COs. Non-interfering, class II COs are distributed randomly independently of Rec(P) and do not impose CO interference. Successive events falling within a set threshold of one another (e.g. 1.5kb) are merged into a single event residing at the midpoint position. These processes repeat until a pre-determined number of simulated ICDs are obtained. **b,** Gamma (γ) mixture modelling was utilised to resolve and estimate individual components of simulated two component mixtures with known parameters (α,β), at known weights (W)—generated via *RecombineSim*. A set of representative examples are shown. Percentage differences between actual and estimated parameters are calculated and averaged to estimate error rate (N(%Δ)) and algorithm accuracy. S = number of ICDs. **c,** Error rate (N(%Δ)) values for three (γ) mixtures calculated at varying sample size (S).

**Supplementary Fig. 5 | Localised impact of polymorphism density upon CO formation.**

**a—b,** Empirical cumulative distribution functions (eCDFs) showing the fraction of COs that reside within a region (±1000bp and ±2000bp respectively) of a given SNP/INDEL count (S288c x SK1 only). Expected is calculated using DSB hotspot midpoints[19]. Pairwise goodness-of-fit tests were performed between pooled *msh2*Δ and *MSH2*$^{+}$ datasets as indicated (triangular legend). P values: Two-sample KS-test. **c—d,** Smoothed SNP/INDEL density maps for S288c x SK1 ChrVI and ChrVII respectively.

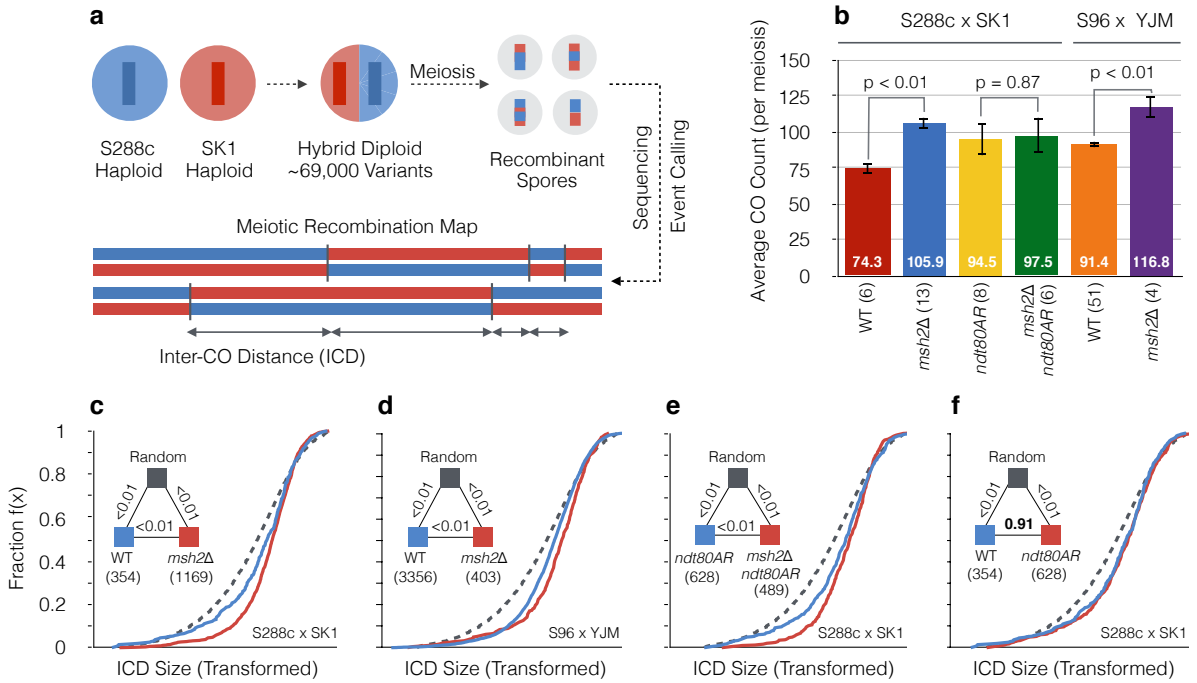



**Fig. 1 | Inactivation of Msh2 increases CO frequency and the global strength of CO interference**

**a,** Genome-wide mapping of recombination. Meiosis is induced within hybrid S288c x SK1 *S. cerevisiae* diploid cells and genomic material is prepped from individual, isolated spores for paired-end Illumina sequencing in order to genotype SNP/INDEL patterns and therefore determine the parental origin of any given loci (Methods). Only a single chromosome is shown for clarity. Inter-crossover distances (ICDs), a measure of CO interference, are calculated as the distance (in bp) between successive COs along a given chromosome. **b,** Average number of COs per meioses for each genotype. The number of individual meioses sequenced per genotype is indicated. Error bars: 95% confidence intervals (CI). *P values: Two-sample T-test.* **c—f,** Empirical cumulative distribution functions (eCDFs) showing the fraction of ICDs at or below a given size. The total number of experimental ICDs is indicated in brackets. ICDs are transformed as per the protocol described to correct for skews generated by differing CO frequencies (Methods). Untransformed ICD plots are available in (Supplementary Fig. 1c—h). Randomized datasets were generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were performed between genotypes as indicated (triangular legend). *P values: Two-sample KS-test.*

**FIGURE 1**

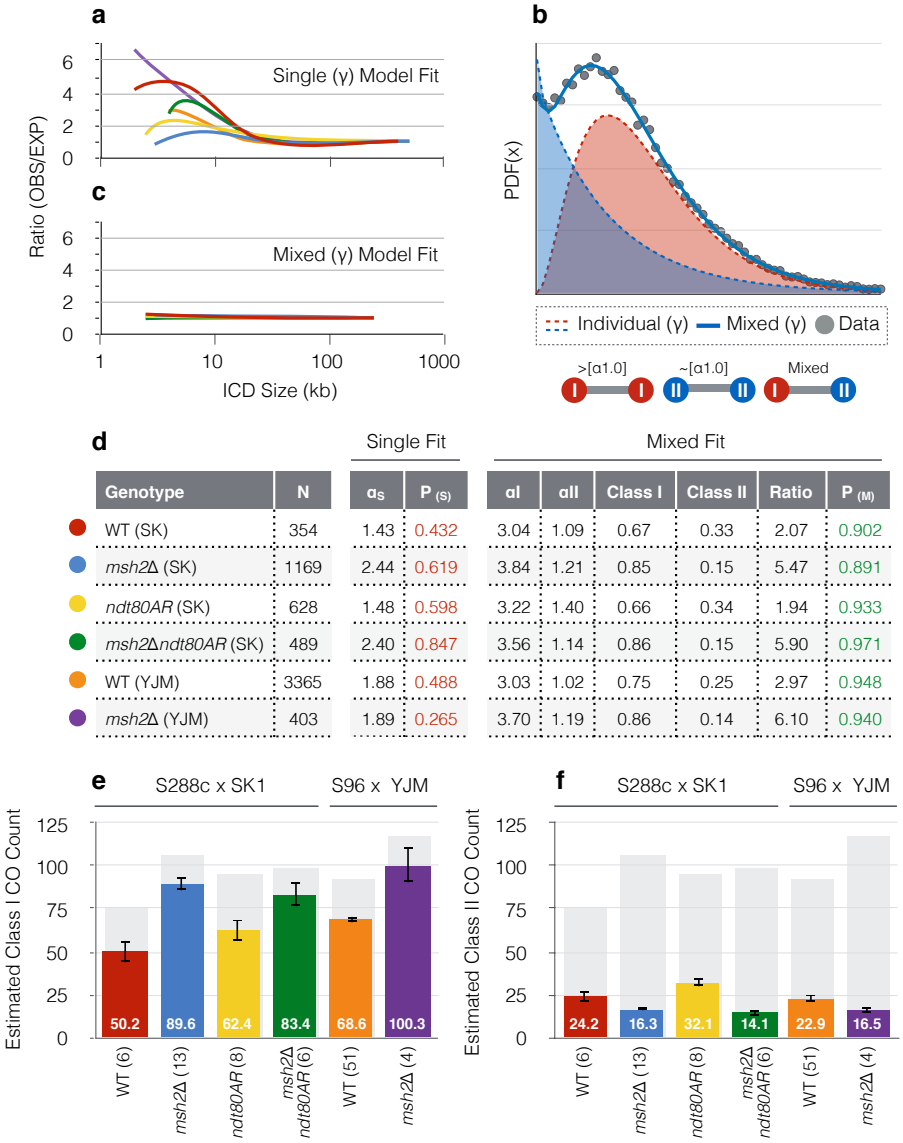



**Fig. 2 | Computational modelling predicts a Msh2-dependent shift in the class I:class II CO ratio**

**a,** Ratio of experimentally observed ICD sizes (OBS) versus the theoretical expectation based on a single, best-fit gamma (γ)-distribution (EXP). Ratio values were calculated at 5kb intervals. **b,** Example (γ) mixture model (α1.0-β1.0 + α3.0β5.0). I = class I. II = class II. Meiotic datasets, owing to the existence of two CO subclasses, are a heterogenous population of three ICD types (as shown). **c,** As in **(a)** but based on a mixed (γ)-model (no. of distributions fitted = 2). **d,** Best-fit (γ) mixture modelling results. N = sample size (total number of ICDs). $\alpha_s$ = Single-fit γ(α) value. $P_{(s)}$ = Fit quality of a single (γ)-distribution (one-sample KS-test). αI/II = Mixed model γ(α) values. Class I/II = estimated fraction of each CO subclass. Ratio = class I:class II. $P_{(M)}$ = Fit quality of a mixed (γ)-mode (Two-sample KS-test) **e—f,** Estimated class I and class II CO counts respectively. Estimates were obtained using the best-fit class I:class II ratios. Total CO frequencies are overlaid (grey bar). Error bars: 95% confidence intervals (CI). The number of individual meioses sequenced per genotype is indicated.

**FIGURE 2**

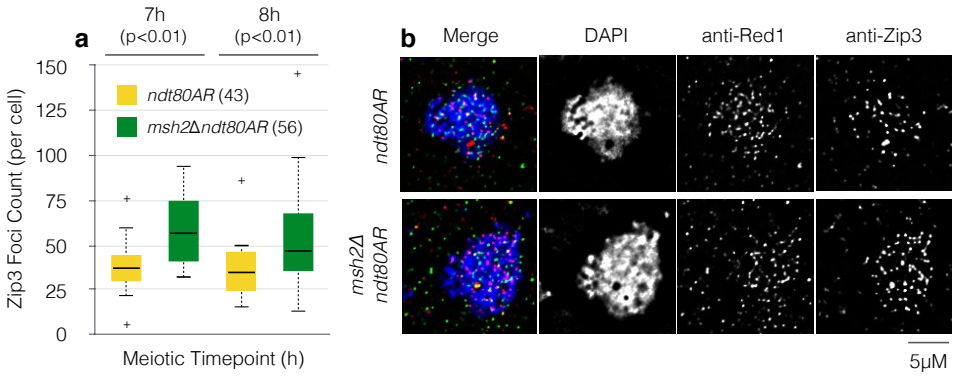



**Fig. 3 | Zip3 foci counts are significantly elevated within Msh2-deficient cells**

**a,** Box-and-whisker plot showing Zip3 foci counts from spread, pachytene arrested, Red1-positive S288c x SK1 cells. Midlines denote median values. *P values: Two-sample T-test.* **b,** Representative example for each genotype (7h timepoint). Cells are fluorescently labelled for the meiosis-specific axis protein Red1 (green), the interfering CO marker, Zip3 (red) and DNA (blue). Only cells with sufficient Red1 signal, as determined by the relative intensity of Red1 foci above the background, were included in the analysis. The total number of nuclei counted is indicated in brackets.

**FIGURE 3**

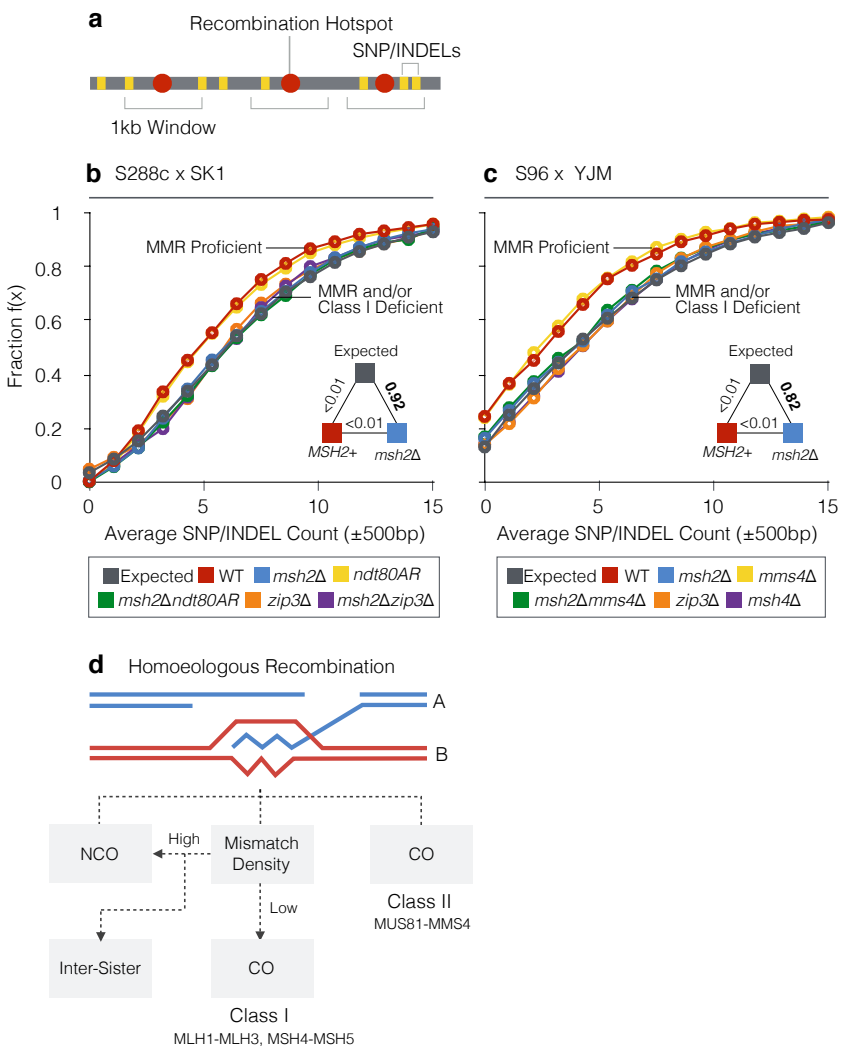
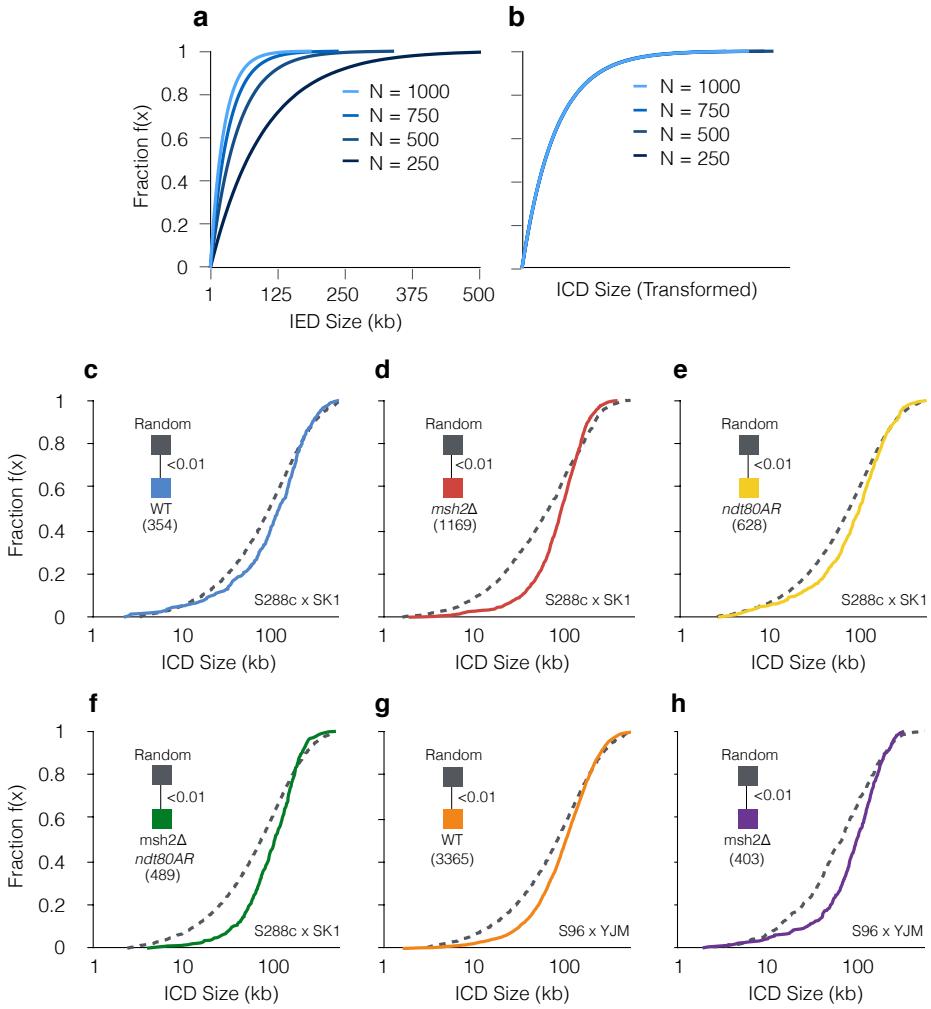



**Fig. 4 | Suppression of class I COs occurs at regions of higher sequence divergence**

**a,** SNP/INDEL count is assayed using a ±500bp window centered on CO or DSB hotspot midpoints. All contained SNP/INDELs are then subsequently tallied, with equal weight. **b—c,** Empirical cumulative distribution functions (eCDFs) showing the fraction of COs that reside within a region of a given SNP/INDEL count for S288c x SK1 and S96 x YJM789 respectively. Expected is calculated using DSB hotspot midpoints[18]. Pairwise goodness-of-fit tests were performed between pooled *msh2Δ* and *MSH2*$^{+}$ datasets as indicated (triangular legend). P values: Two-sample KS-test. **d,** Model summarizing the impact Msh2-dependent regulation has upon CO formation. In the presence of Msh2 or a functional MMR pathway, regions of higher sequence divergence redirect repair toward the NCO or inter-sister pathways. Inactivation of MMR alleviates this repression, increasing the frequency of class I CO formation and thus the global strength of CO interference.

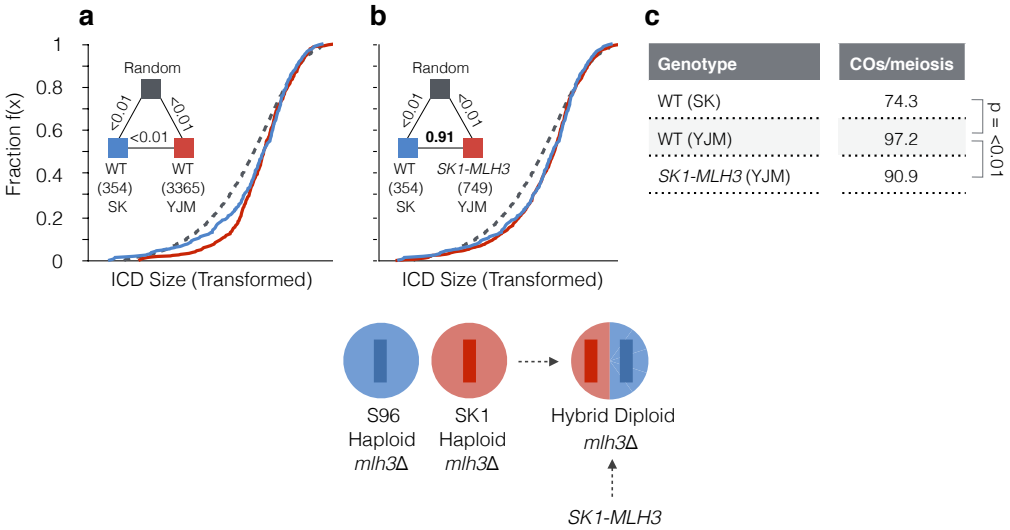
**FIGURE 4**




**Supplementary Fig. 1 | CO interference is present to varying degrees within all mapped strains**

**a,** Empirical cumulative distribution function (eCDF) showing ICD data derived from interfering simulations (γ(α) = 3.0) at varying CO per cell frequencies (N). **b,** As in **(a)** but ICDs are transformed as per the protocol described to correct for skews generated by differing CO frequencies (Methods). **c—h,** eCDFs showing the fraction of ICDs at or below a given size. The total number of experimental ICDs is indicated in brackets. Randomized datasets were generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were performed between genotypes as indicated (triangular legend). P values: Two-sample KS-test.

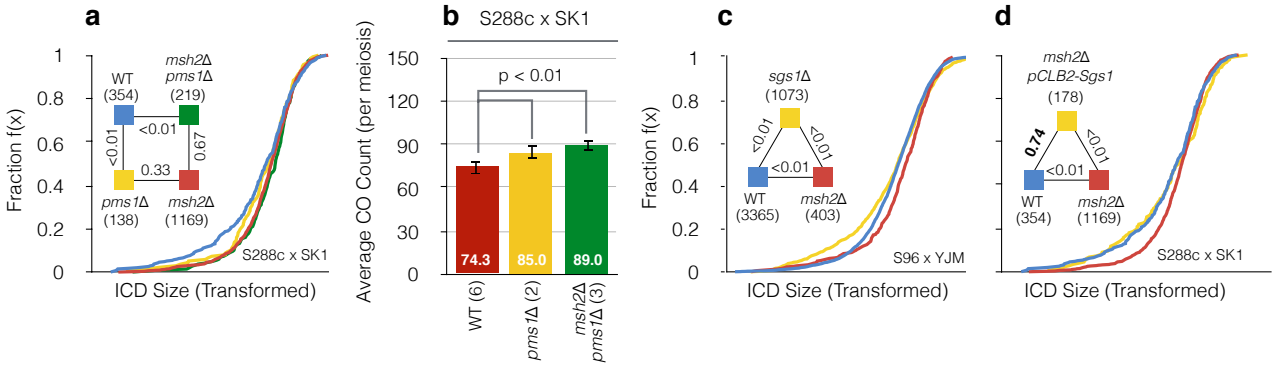
**FIGURE S1**




**Supplementary Fig. 2 | Cross-specific differences—the *SK1-MLH3* allele has reduced capacity to mediate CO interference**

**a—b,** Empirical cumulative distribution functions (eCDFs) showing the fraction of ICDs at or below a given size. The total number of experimental ICDs is indicated in brackets. ICDs are transformed as per the protocol described to correct for skews generated by differing CO frequencies (Methods). Randomized datasets were generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were performed between genotypes as indicated (triangular legend). A schematic of the *SK1-MLH3* strain analyzed is shown. P values: Two-sample KS-test. **c,** Average number of COs per meiosis for each genotype. P values: Two-sample T-test.

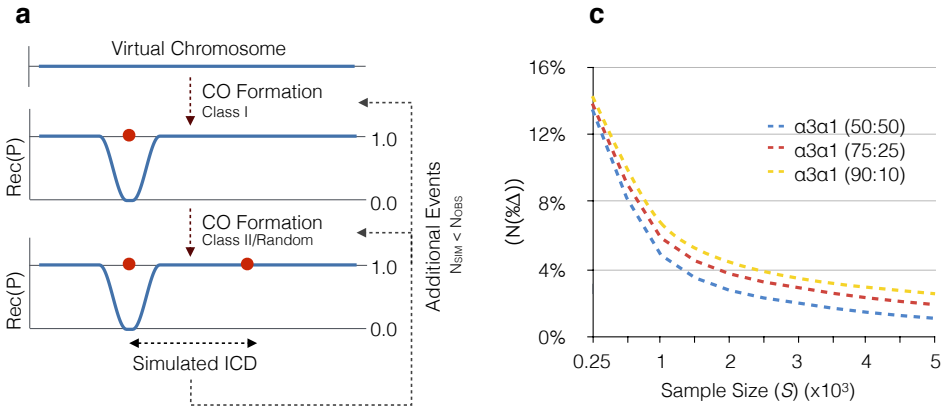
**FIGURE S2**




**Supplementary Fig. 3 | Mechanistic details of MMR-dependent suppression of interfering COs**

**a,** Empirical cumulative distribution function (eCDF) showing the fraction of ICDs at or below a given size. The total number of experimental ICDs is indicated in brackets. ICDs are transformed as per the protocol described to correct for skews generated by differing CO frequencies (Methods). Randomized datasets were generated via simulation to represent a state of no interference (Methods). Pairwise goodness-of-fit tests were performed between genotypes as indicated (triangular legend). P values: Two-sample KS-test. **b,** Average number of COs per meiosis for each genotype. The number of individual meioses sequenced per genotype is indicated. Error bars: 95% confidence intervals (CI*). P values: Two-sample T-test.* **c—d,** As in **(a)** but for differing genotypes.

**FIGURE S3**



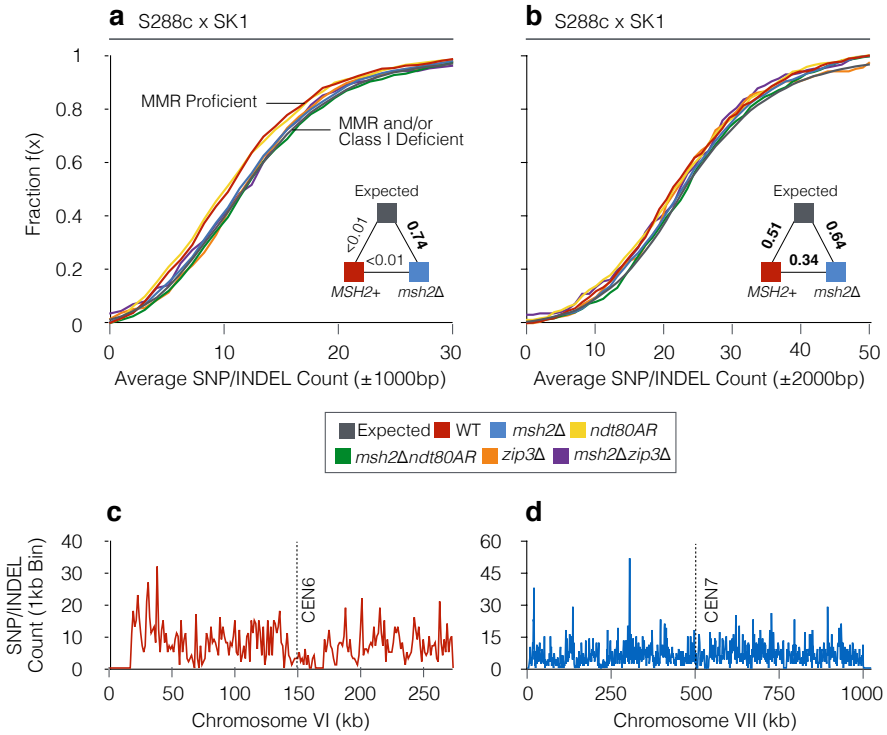

**b**

| | | Actual Mixture (Simulated) | | | | | | | Estimated Mixture (γ Mixture Model) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | S | α1 | β1 | α2 | β2 | $W_1$ | $W_2$ | α1 | β1 | α2 | β2 | $W_1$ | $W_2$ | N(%Δ) |
| Variable S | | 250 | 3.0 | 35000 | 1.0 | 75000 | 0.75 | 0.25 | 3.45 | 31139 | 0.89 | 84392 | 0.69 | 0.31 | 13.14 |
| | | 500 | 3.0 | 35000 | 1.0 | 75000 | 0.75 | 0.25 | 3.29 | 32889 | 1.11 | 69448 | 0.72 | 0.28 | 8.16 |
| | | 1000 | 3.0 | 35000 | 1.0 | 75000 | 0.75 | 0.25 | 2.91 | 35917 | 0.94 | 79650 | 0.78 | 0.22 | 5.79 |
| Variable W | | 1000 | 3.0 | 35000 | 1.0 | 75000 | 0.90 | 0.10 | 2.83 | 37110 | 0.92 | 78595 | 0.91 | 0.09 | 6.28 |
| | | 1000 | 3.0 | 35000 | 1.0 | 75000 | 0.50 | 0.50 | 3.08 | 36894 | 0.94 | 77998 | 0.48 | 0.52 | 4.24 |
| Higher α | | 1000 | 4.0 | 25000 | 2.0 | 50000 | 0.75 | 0.25 | 4.09 | 23976 | 1.95 | 51704 | 0.77 | 0.23 | 4.21 |
| | | 1000 | 5.0 | 25000 | 3.0 | 50000 | 0.75 | 0.25 | 5.10 | 25997 | 3.06 | 51874 | 0.78 | 0.22 | 4.55 |

**Supplementary Fig. 4 | Modelling CO distributions.**

**a,** *RecombineSim* overview. Virtual chromosomes are constructed at a 100bp resolution as binned, numerical arrays upon which meiotic CO formation is simulated (Methods). Any given 100bp contains a value in the range of [0.0-1.0], designating its recombination potential (Rec(P)). Prior to CO formation, bins are initially populated with [1.0]—denoting an equal probability of class I CO formation. During the formation of an interfering CO, *RecombineSim* imposes CO interference as a distance-dependent zone of repression by modifying Rec(P) values—influencing the position of all subsequent interfering COs. Non-interfering, class II COs are distributed randomly independently of Rec(P) and do not impose CO interference. Successive events falling within a set threshold of one another (e.g. 1.5kb) are merged into a single event residing at the midpoint position. These processes repeat until a pre-determined number of simulated ICDs are obtained. **b,** Gamma (γ) mixture modelling was utilised to resolve and estimate individual components of simulated two component mixtures with known parameters (α,β), at known weights (W)—generated via *RecombineSim*. A set of representative examples are shown. Percentage differences between actual and estimated parameters are calculated and averaged to estimate error rate (N(%Δ)) and algorithm accuracy. S = number of ICDs. **c,** Error rate (N(%Δ)) values for three (γ) mixtures calculated at varying sample size (S).

**FIGURE S4**




**Supplementary Fig. 5 | Localised impact of polymorphism density upon CO formation.**

**a—b,** Empirical cumulative distribution functions (eCDFs) showing the fraction of COs that reside within a region (±1000bp and ±2000bp respectively) of a given SNP/INDEL count (S288c x SK1 only). Expected is calculated using DSB hotspot midpoints[18]. Pairwise goodness-of-fit tests were performed between pooled *msh2*Δ and *MSH2*$^{+}$ datasets as indicated (triangular legend). P values: Two-sample KS-test. **c—d,** Smoothed SNP/INDEL density maps for S288c x SK1 ChrVI and ChrVII respectively.

**FIGURE S5**

| Genotype | Strain | Background | Mat | Genotype |
|---|---|---|---|---|
| WT | MJ513 | SK1 | a | *ho::LYS2 lys2Δ leu2Δ arg4Δ* |
| | MJ600 | S288c | α | *ade8Δ* |
| *msh2Δ* | MC26 | SK1 | α | *ho::LYS2 lys2Δ ura3Δ arg4 leu2 msh2Δ::Kan* |
| | MC49 | S288c | a | *ade8Δ msh2Δ::Kan* |
| *ndt80AR* | MJ43 | SK1 | α | *ho::LYS2 lys2Δ arg4Δ leu2Δ::hisG trp1Δ::hisG his4XΔ::LEU2 nuc1Δ::LEU2 PGAL1-NDT80::TRP1 ura3::pGPD1-GAL4(848)-ER::URA3* |
| | MC42 | S288c | a | *ade8Δ ndt80Δ::Kan* |
| *msh2Δndt80AR* | MC298 | SK1 | a | *ho::LYS2 lys2Δ ura3Δ arg4 leu2 trp1Δ::hisG ura3Δ::PGPD1-GAL4(848)-ER::URA3 PGAL1-NDT80::TRP1 msh2Δ::Kan* |
| | MC300 | S288c | α | *ade8Δ ndt80Δ::Kan msh2Δ::Kan* |
| *zip3Δ* | MC322 | SK1 | α | *ho::LYS2 lys2Δ ura3Δ arg4 leu2 zip3Δ::HphMX4* |
| | MC313 | S288c | a | *ade8Δ zip3Δ::HphMX4* |
| *msh2Δzip3Δ* | MC326 | SK1 | α | *ho::LYS2 lys2Δ ura3Δ arg4 leu2 msh2Δ::Kan zip3Δ::HphMX4* |
| | MC317 | S288c | a | *ade8Δ msh2Δ::Kan zip3Δ::HphMX4* |

**Supplementary Table 1 | Strain Table (S288c x SK1).**

**TABLE S1**