Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing

Sijia Lu,[1]*† Chenghang Zong,[1]* Wei Fan,[2]* Mingyu Yang,[2]* Jinsen Li,[2] Alec R. Chapman,[1,3] Ping Zhu,[2] Xuesong Hu,[2] Liya Xu,[2] Liying Yan,[4,5] Fan Bai,[2] Jie Qiao,[4,5] Fuchou Tang,[2] Ruiqiang Li,[2,6]‡ X. Sunney Xie[1,2]‡

Meiotic recombination creates genetic diversity and ensures segregation of homologous chromosomes. Previous population analyses yielded results averaged among individuals and affected by evolutionary pressures. We sequenced 99 sperm from an Asian male by using the newly developed amplification method—multiple annealing and looping-based amplification cycles—to phase the personal genome and map recombination events at high resolution, which are nonuniformly distributed across the genome in the absence of selection pressure. The paucity of recombination near transcription start sites observed in individual sperm indicates that such a phenomenon is intrinsic to the molecular mechanism of meiosis. Interestingly, a decreased crossover frequency combined with an increase of autosomal aneuploidy is observable on a global per-sperm basis.

Meiosis plays a crucial role in generating haploid gametes for sexual reproduction. In most organisms, the presence of crossovers between homologous chromosomes, in combination with connections between sister chromatids, creates a physical connection that ensures regular segregation of homologs at the first of the two meiotic divisions (*1*). Abnormality in generating crossovers is the leading cause of miscarriage and birth defects (*2*). Crossovers also create new combinations of alleles, thus contributing to genetic diversity and evolution (*3*).

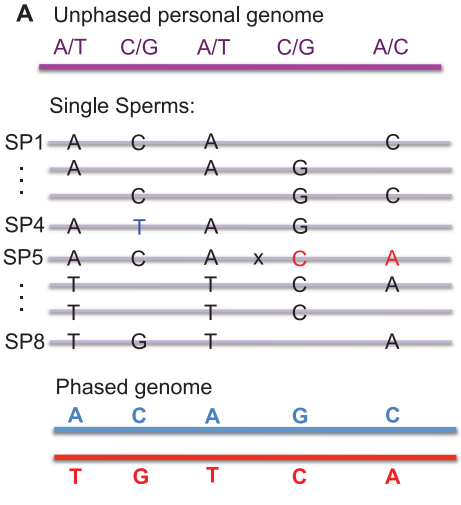
Recent linkage disequilibrium and pedigree studies have shown that the distribution of recombination is highly uneven across the human genome (*4*, *5*), as in all studied organisms. Substantial recombination active regions are not conserved between humans and chimpanzees (*6*–*8*) or among different human populations (*9*, *10*), suggesting that these regions are quickly evolving and might even be individual-specific (*11*). However, such variation in the human population would be masked by the population average, and resolution of this variation would require comparison of recombination genome-wide among many single genomes.

Whole-genome amplification (WGA) of single sperm cells was proposed decades ago to facilitate mapping recombination at the individual level (*12*). With the development of high-throughput genotyping technologies (*13*, *14*), whole-genome mapping of recombination events in single gametes of an individual is achievable and was recently demonstrated by performing WGA by multiple displacement amplification (MDA) (*15*) on single sperm cells, followed by genotyping with DNA microarrays recently demonstrated by Wang *et al*. (*16*). However, due to the amplification bias and, consequently, insufficient marker density, the resolution of crossover locations has been limited to ~150 kb thus far. In addition, in their recent work (*16*), Wang *et al*. relied on prior knowledge of the chromosome-level haplotype information of the analyzed individual, which is experimentally inconvenient to obtain and is currently available for only a few individuals (*17*–*19*).

Here, we demonstrate a more general approach for studying recombination in single sperm cells of an individual, without prior knowledge of the haplotype information. We isolated single sperm cells from a healthy Asian male donor in his late 40s. The donor has healthy offspring of both genders and normal clinical semen analysis results. We used a recently developed method, multiple annealing and looping-based amplification cycles (MALBAC) (*20*), to perform WGA on single cells. MALBAC provides substantially improved amplification evenness compared with the prevailing WGA methods, such as MDA. We sequenced 93 sperm at ~1× genome depth and 6 sperm at ~5× depth, achieving genome coverages of ~23 and ~43%, respectively (table S1). Three of the 99 sperm samples were found to contain more than one haploid cell and were filtered out in downstream analysis (fig. S1). Approximately 89% of the sequencing reads from single sperm can be aligned to the human genome, in agreement with that of a typical human resequencing project.

We further sequenced the diploid genome of the donor at ~70× depth and identified ~2.8 million single nucleotide polymorphisms (SNPs). About 1.4 million of them are heterozygous (hetSNPs) (table S2) (*21*). Among the hetSNP sites, ~500,000 (35%) and ~300,000 (20%) could be genotyped with a >99% accuracy (Phred quality score > 20) threshold for the high coverage (5×) and low coverage (1×) sperm cells, respectively (table S3).

Phase information is crucial for the correct description and interpretation of the human



B

Heterozygous SNPs	1,390,544
% covered by ≥ 2 sperm	88.0%
% phased by sperm linkage	82.4%
% consistent with family trio	99.5%
Phased length	Whole chromosome

Fig. 1. Principle of whole-genome phasing of an individual using the SNP linkage information from individual sperm cells. (**A**) We sequenced the diploid genome and identified five hetSNPs with unknown linkage information shown in purple. Individual sperm cells were sequenced after MALBAC amplification, from which SNP linkage information in each sperm was used to infer the phase information in the diploid genome. (**B**) Performance of whole-genome phasing by SNP linkage in sperm cells.

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [2]Biodynamic Optical Imaging Center, School of Life Sciences, Peking University, Beijing 100871, P. R. China. [3]Program in Biophysics, Harvard University, Cambridge, MA 02138, USA. [4]Center for Reproductive Medicine, Third Hospital, Peking University, Beijing 100191, P. R. China. [5]Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing 100191, P. R. China. [6]Peking-Tsinghua Center for Life Sciences, School of Life Sciences, Peking University, Beijing 100871, P. R. China.

*These authors contributed equally to this work.
†Present address: Yikon Genomics, 1 China Medical City Avenue, TQB Building 5th Floor, Taizhou, Jiangsu, China.
‡To whom correspondence should be addressed. E-mail: lirq@pku.edu.cn (R.L.); xie@chemistry.harvard.edu (X.S.X.)

genome (*22*) and is essential for mapping crossovers. We phased the hetSNPs into chromosome-level haplotypes by comparing the SNP linkage information across all sperm (Fig. 1A) (*21*). Because crossovers (such as the A-C link in SP5) and false SNP identification (such as the highlighted T in SP4) are low-probability events, most SNP linkage information identified in a sperm reflects the true SNP linkage in the somatic genome. These SNP linkages were calculated statistically by comparing across all sperm cells. In doing so, we were able to phase ~1.1 million (~82%) hetSNPs with high confidence into two sets of chromosome haplotypes. To verify the phasing result, we lightly sequenced the genomes from the donor's parents (~10× each) and used a pedigree approach to infer the phase information of the donor (tables S3 and S4) (*21*, *23*). We obtained ~99.5% consistency between the two methods, indicating the high accuracy of our approach in phasing hetSNPs into chromosome-level haplotypes (Fig. 1B and table S4) (*21*). The percentage of phased hetSNPs could be further improved with higher sequencing depths from each sperm (currently only ~1×).

Several methods for haplotyping individual humans have been described previously (*19*, *24*, *25*). However, these methods often involved labor-intensive sample preparations and had limited haplotype block size (<1 Mb). Our method enables whole-genome phasing into haplotypes of complete chromosomes, without requiring cell culture or devices for separating metaphase chromosomes (*18*, *26*).

With the diploid genome phased into haplotypes of complete chromosomes, we used SNPs as markers to map the positions of crossovers in each sperm. We used a hidden Markov model to accurately determine the positions for most crossovers, and we manually identified the crossovers for the low-confidence regions (Fig. 2A) (*21*). We identified 2368 autosomal crossover events in the sperm cells with a complete haploid genome. The average of ~26.0 crossovers per cell is consistent with reported pedigree studies (*27*, *28*). The amplification evenness of MALBAC allowed us to achieve high resolution in detecting crossovers with only ~1× sequencing depth from each sperm. About 93, 80, and 45% of the crossovers can be confidently assigned to intervals of 200, 100, and 30 kb, respectively (Fig. 2B), compared with 59, 37, and 13% from the recently reported single-sperm study (*16*). Of the crossovers unambiguously resolvable within a 10-kb interval, ~40% are found to overlap with the male-specific recombination hotspots inferred from the deCODE project (*9*). Also, ~45% of these crossovers are close to the PRDM9 binding motif CCnCCnTnnCCnC, which is consistent with previous population studies (*29*).

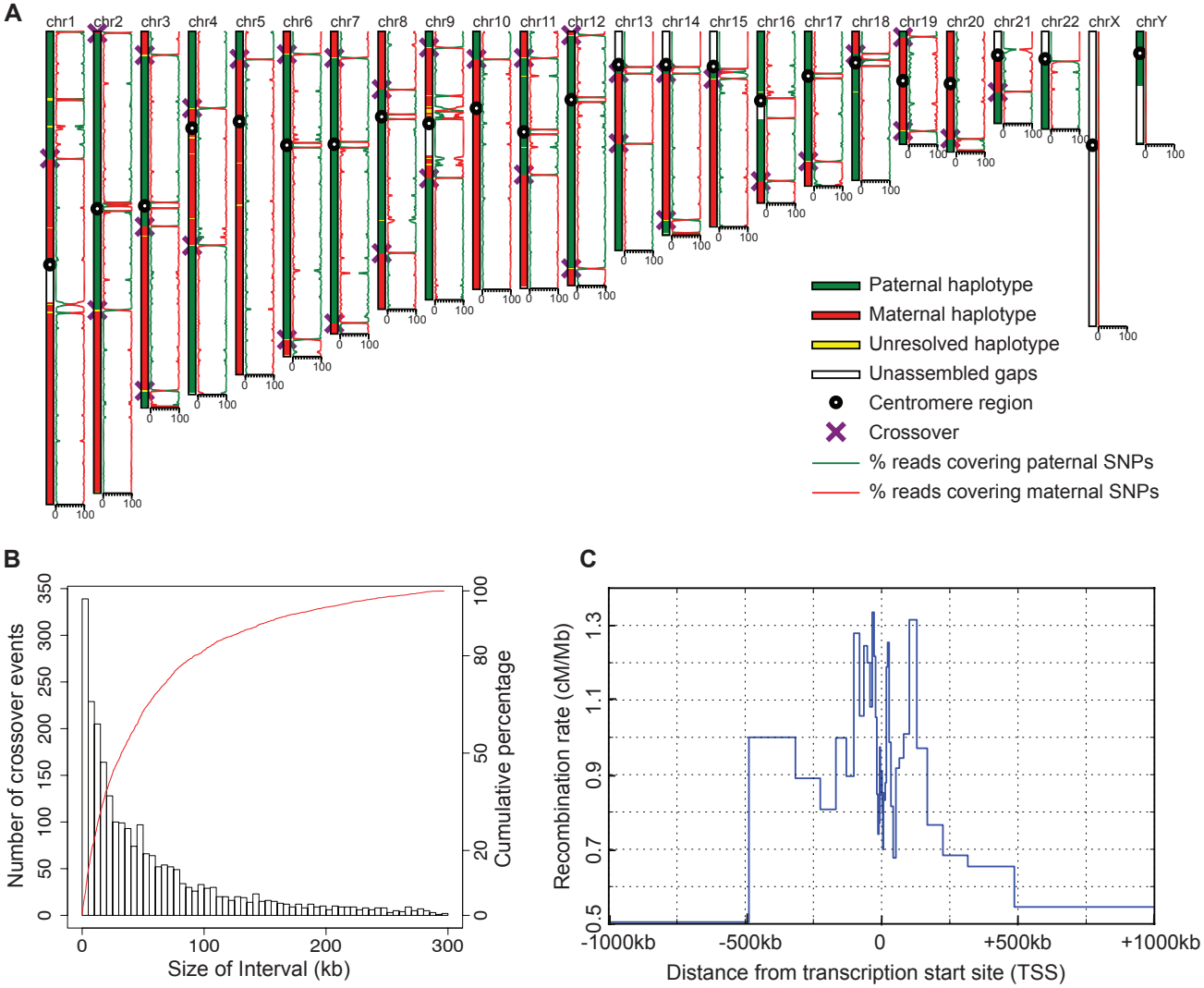


Fig. 2. Identifying crossover positions in individual sperm cells. (**A**) Parental haplotype contributions are determined by comparing the percentage of reads covering the paternal or maternal SNPs, and crossover positions are detected by identifying the crossing locations of the two parental haplotypes by a hidden Markov model. (**B**) Resolution of crossover determination. About 60% of the crossovers can be determined within intervals of 50 kb. (**C**) Distribution of recombination rate relative to TSSs. cM, centimorgans.

Recombination rates correlate positively with gene density in both yeast and humans (*27*, *30*). However, at a finer scale, recombination rates in human populations are lower very close to genes (within 20 kb) and higher tens or hundreds of kilobases away from the transcription start sites (TSSs) (*4*, *9*, *28*). This feature is an average of different individuals that reflects the cumulative effect of human evolutionary history, and it may also be complicated by selecting against the recombinations that compromise offspring viability. Our method detects recombination features based on single gametes, which are free of selection effects of population studies. We analyzed the crossovers resolvable within 30 kb and derived the recombination rate relative to the TSSs of the individual (Fig. 2C) (*21*). We observed lower recombination rates close to the TSSs and higher rates tens of kilobases away, consistent with the results of previous population studies (*4*, *9*, *28*), indicating that the reduced recombination rate close to TSSs is primarily due to the variation of recombination probability during meiosis rather than due to selection.

Previous population studies have shown that recombination events have a nonuniform distribution across the genome, which reflects the cumulative evolutionary history of recombination (*5*). By binning the crossover incidence into 3-Mb units in autosomes, we constructed a genetic map of recombination of the individual. We compared our map to a sex-averaged map based on population (HapMap) (*4*) and a male-specific map based on pedigree (deCODE) (*9*) (Fig. 3, A and B) and obtained correlation coefficients of 0.71 and 0.77, respectively. In some of the bins, we observed a significant difference between HapMap and the donor (table S6), which can be explained by sex-specific recombination variations.

A recent study reported that crossover active regions that are specific to an individual exist at a megabase scale (*16*). We also found nine bins showing significant differences between the donor and deCODE (table S7). However, we note that most of these regions are very close to the centromere or the ends of the chromosomes, where the estimation of the recombination rates was considered unreliable and excluded in deCODE (*9*). Therefore, we suspect that these differences mainly reflect the incompleteness of the deCODE database. Our results suggest that the distribution of recombination in

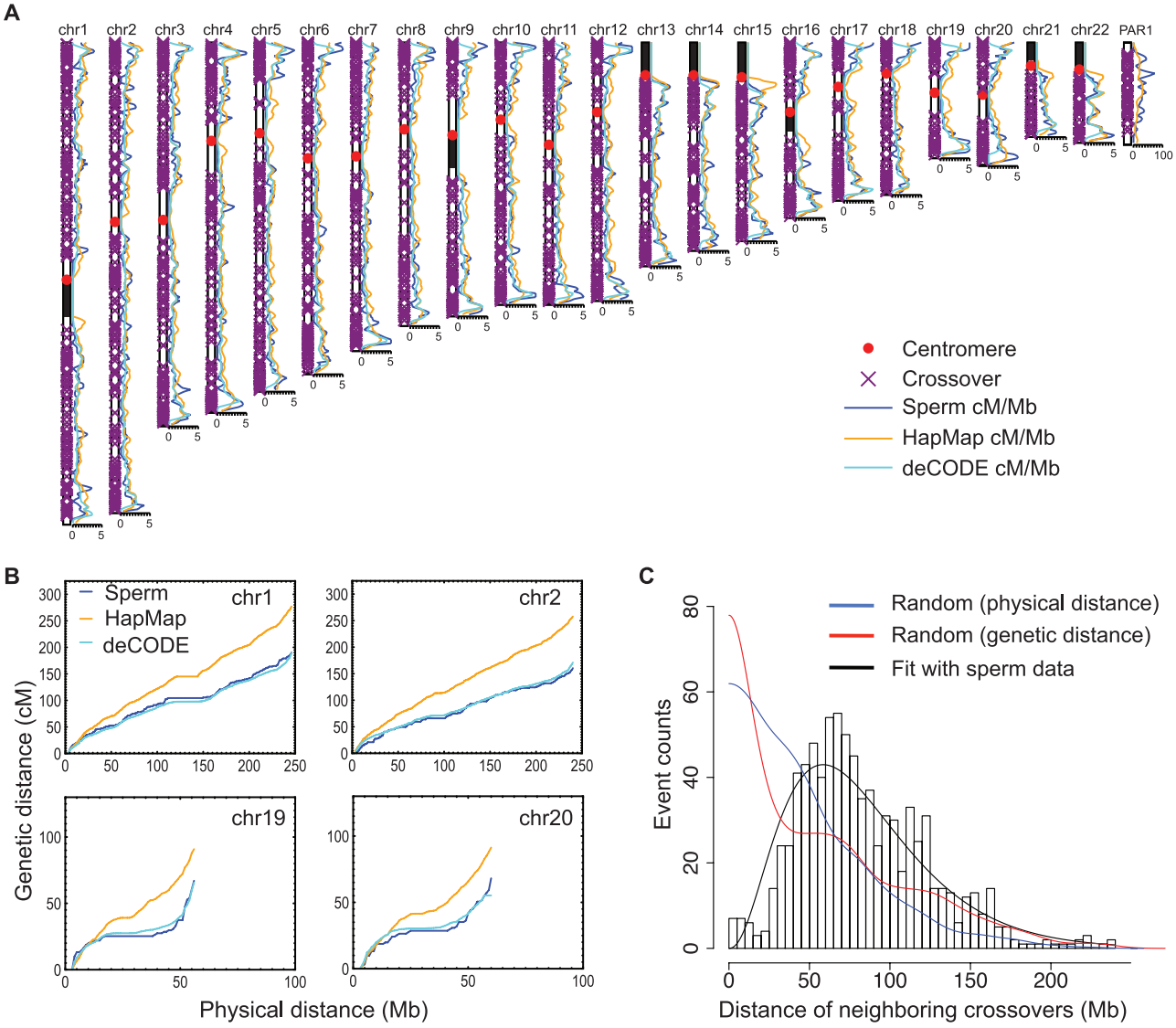


Fig. 3. Genome-wide distribution of recombination. (**A**) Comparison of the sperm recombination rates to the HapMap and deCODE (male-specific) genetic maps across the human genome. We used a 3-Mb statistical window size and a 1-Mb moving step. (**B**) A personal genetic map. Relations of physical and genetic length of selected chromosomes. (**C**) Distance distribution of crossovers co-occuring on the same chromosome. The experimental data are fit with a gamma distribution (α = 3.35), indicating a strong deviation from the random distribution. In comparison, we generated random crossovers based on physical and genetic distances.

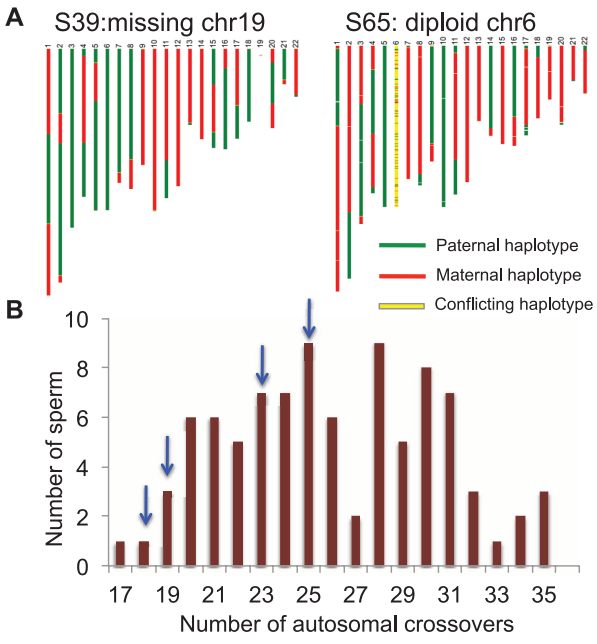


Fig. 4. Detecting aneuploidy and crossover in the same sperm. (**A**) Two of the four sperm cells that exhibit autosomal abnormality. Few reads are mapped to chr19 in S39, indicating a loss of chr19. Both parental haplotypes are found in chr6 of S65, indicating a disomy chr6 in the sperm. The detailed coverage analysis on all four aneuploid sperm is shown in fig. S2. (**B**) Distribution of the autosomal crossover number. Arrows indicate the number of crossovers in sperm cells with autosomal aneuploidy.

the individual generally agrees with the population average at the megabase scale, indicating a general consistency of large-scale recombination distribution in human evolution. With the rapid development of sequencing technologies, more sperm can be analyzed in the future from different individuals to look into fine-scale recombination differences. We estimated that at least 1000 sperm are required to identify personal recombination differences with statistical significance (fig. S4) (*21*).

Obtaining the genome sequence of each sperm also allowed us to examine the coexisting crossovers on the same chromosome. The adjacent crossovers exhibit longer distances than expected by random chance (Fig. 3C and figs. S5 and S6), which is consistent with the well-established phenomenon of crossover interference (*31*, *32*). Although we have higher resolution to detect crossovers than in a previous study, we did not see the reported phenomenon of substantial double crossovers occurring close together (e.g., 1 to 5 Mb) (*33*), which suggests that such phenomenon is likely not general and may only exist in certain populations.

Failure to form crossovers during meiosis gives rise to chromosome segregation errors that result in aneuploidy. Autosomal aneuploidy is often lethal to embryos, with the exception of a few chromosomes that result in severe health consequences early in development (e.g., trisomy 21, Down syndrome). Reduced recombination activity is often associated with male infertility and sperm aneuploidy (*34*). By comparing the coverage depth and SNPs along the genome of the sperm cells, we detected four cells either missing or having additional autosomes (Fig. 4A and fig. S2) (*21*). The rate of chromosome mis-segregation is consistent with the reported imaging studies on selected loci of human spermatocytes (*35*, *36*).

We next compared the crossover number of the aneuploid sperm to the normal sperm. Interestingly, sperm cells with aneuploid autosomes exhibit significantly fewer crossovers than normal cells, on average ($P = 0.01$). Our result suggests that autosomal segregation errors are not generated randomly during spermatogenesis. Instead, the error rate is higher in the spermatocytes with relatively repressed crossover activity. However, such a trend does not seem to be significant for sex chromosome aneuploidy, as we observed a sperm with 30 autosomal crossovers but no sex chromosome. Indeed, the crossover probability in the pseudoautosomal region of the sex chromosomes has no noticeable correlation with that of the autosomes (tables S8 and S9) (*21*), suggesting a different mechanism of crossover generation for autosomes and sex chromosomes, which is consistent with an earlier study in mice (*37*). We were unable to determine whether the chromosomes exhibiting aneuploidy underwent recombination, as recombination events are observable only when a single copy is present. MALBAC allows direct examination of meiotic crossovers and chromosome segregation errors on a per–meiotic nucleus basis at high resolution, enabling further applications for the study of genome instability and male infertility.

References and Notes

1. M. Petronczki, M. F. Siomos, K. Nasmyth, *Cell* **112**, 423 (2003).
2. C. J. Epstein, *The Consequences of Chromosome Imbalance: Principles, Mechanisms, and Models* (Cambridge Univ. Press, Cambridge, 2007).
3. A. J. Jeffreys, C. A. May, *Nat. Genet.* **36**, 151 (2004).
4. S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, *Science* **310**, 321 (2005).
5. K. Paigen, P. Petkov, *Nat. Rev. Genet.* **11**, 221 (2010).
6. W. Winckler *et al.*, *Science* **308**, 107 (2005).
7. S. E. Ptak *et al.*, *Nat. Genet.* **37**, 429 (2005).
8. A. Auton *et al.*, *Science* **336**, 193 (2012).
9. A. Kong *et al.*, *Nature* **467**, 1099 (2010).
10. A. G. Hinch *et al.*, *Nature* **476**, 170 (2011).
11. A. J. Jeffreys, R. Neumann, *Nat. Genet.* **31**, 267 (2002).
12. L. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5847 (1992).
13. D. J. Lockhart, E. A. Winzeler, *Nature* **405**, 827 (2000).
14. M. L. Metzker, *Nat. Rev. Genet.* **11**, 31 (2010).
15. F. B. Dean *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5261 (2002).
16. J. Wang, H. C. Fan, B. Behr, S. R. Quake, *Cell* **150**, 402 (2012).
17. S. Levy *et al.*, *PLoS Biol.* **5**, e254 (2007).
18. H. C. Fan, J. Wang, A. Potanina, S. R. Quake, *Nat. Biotechnol.* **29**, 51 (2011).
19. B. A. Peters *et al.*, *Nature* **487**, 190 (2012).
20. C. Zong, S. Lu, A. R. Chapman, X. S. Xie, *Science*, **338**, 1622 (2012).
21. Materials and methods, as well as other supplementary materials, are available on *Science* Online.
22. R. Tewhey, V. Bansal, A. Torkamani, E. J. Topol, N. J. Schork, *Nat. Rev. Genet.* **12**, 215 (2011).
23. S. R. Browning, B. L. Browning, *Nat. Rev. Genet.* **12**, 703 (2011).
24. J. O. Kitzman *et al.*, *Nat. Biotechnol.* **29**, 59 (2011).
25. E.-K. Suk *et al.*, *Genome Res.* **21**, 1672 (2011).
26. L. Ma *et al.*, *Nat. Methods* **7**, 299 (2010).
27. A. Kong *et al.*, *Nat. Genet.* **31**, 241 (2002).
28. G. Coop, X. Wen, C. Ober, J. K. Pritchard, M. Przeworski, *Science* **319**, 1395 (2008).
29. S. Myers, C. Freeman, A. Auton, P. Donnelly, G. McVean, *Nat. Genet.* **40**, 1124 (2008).
30. T. D. Petes, *Nat. Rev. Genet.* **2**, 360 (2001).
31. N. Kleckner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12592 (2004).
32. M. A. Handel, J. C. Schimenti, *Nat. Rev. Genet.* **11**, 124 (2010).
33. A. Fledel-Alon *et al.*, *PLoS Genet.* **5**, e1000658 (2009).
34. K. A. Ferguson, E. C. Wong, V. Chow, M. Nigro, S. Ma, *Hum. Mol. Genet.* **16**, 2870 (2007).
35. E. L. Spriggs, A. W. Rademaker, R. H. Martin, *Cytogenet. Cell Genet.* **71**, 47 (1995).
36. S. E. Downie, S. P. Flaherty, N. J. Swann, C. D. Matthews, *Mol. Hum. Reprod.* **3**, 815 (1997).
37. L. Kauppi *et al.*, *Science* **331**, 916 (2011).

Acknowledgments: This work was supported by U.S. NIH National Human Genome Research Institute grants (HG005097-1 and HG005613-01) to X.S.X. and by funding from Peking University to the Biodynamic Optical Imaging Center (BIOPIC). The computing was carried out at the National Supercomputer Center in Tianjin, China, and the calculations were performed on TianHe-1. We thank G. S. Young for assistance in sample prep and N. Kleckner and J. Zhang for critical comments on the manuscript. The sequencing data has been deposited with the National Center for Biotechnology Information; the accession number is SRA060945. C.Z., S.L., and X.S.X. are authors on a patent applied for by Harvard University that covers the MALBAC technology.

Supplementary Materials

www.sciencemag.org/cgi/content/full/338/6114/1627/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S9
References

21 August 2012; accepted 12 November 2012
10.1126/science.1229112