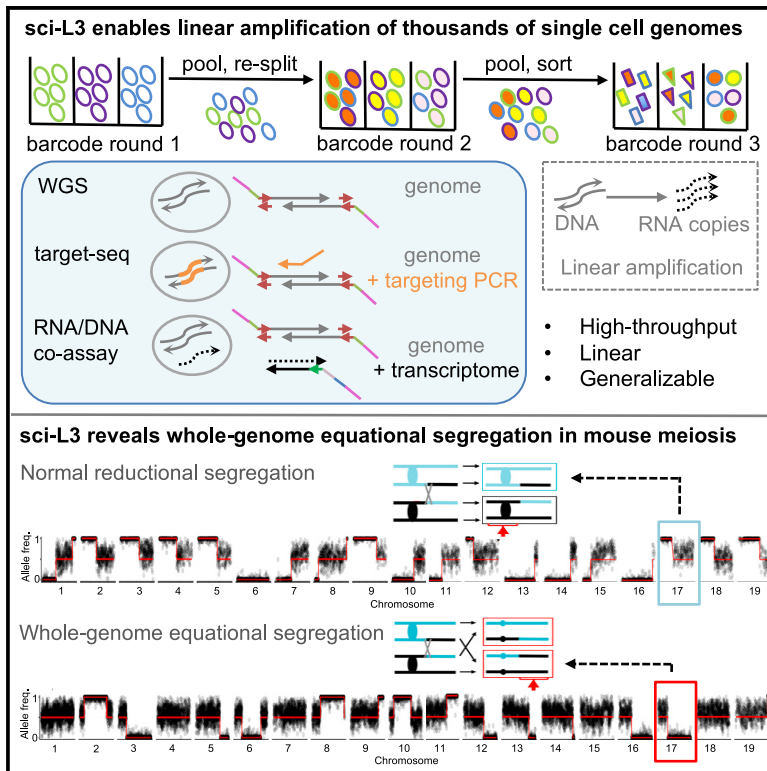


Molecular Cell

High-Throughput Single-Cell Sequencing with Linear Amplification

Graphical Abstract



Authors

Yi Yin, Yue Jiang,
Kwan-Wood Gabriel Lam, ...,
R. Daniel Camerini-Otero,
Andrew C. Adey, Jay Shendure

Correspondence

yy224@uw.edu (Y.Y.),
shendure@uw.edu (J.S.)

In Brief

Yin et al. developed sci-L3, which combines a scalable single-cell barcoding scheme with linear amplification. The method flexibly enables single-cell whole-genome sequencing, targeted DNA sequencing, or concurrent profiling of the genome and transcriptome. With sci-L3, the authors discovered mitotic-like whole-genome chromosome segregation in male mouse meiosis I.

Highlights

- Flexible method for single-cell whole genomes, targeted sequencing, and RNA/DNA co-assay
- Potential throughput of 1 million cells by 3-level single-cell combinatorial indexing
- Linear amplification by *in vitro* transcription for uniform coverage of the genome
- Discovery of whole-genome equational chromosome segregation in mouse meiosis I



High-Throughput Single-Cell Sequencing with Linear Amplification

Yi Yin,^{1,*} Yue Jiang,² Kwan-Wood Gabriel Lam,³ Joel B. Berletch,⁴ Christine M. Disteche,⁴ William S. Noble,¹ Frank J. Steemers,⁵ R. Daniel Camerini-Otero,³ Andrew C. Adey,^{6,7} and Jay Shendure^{1,8,9,10,11,*}

¹Department of Genome Sciences, University of Washington, Seattle, WA, USA

²Seattle, WA, USA

³Genetics and Biochemistry Branch, NIDDK, NIH, Bethesda, MD, USA

⁴Department of Pathology, University of Washington, Seattle, WA, USA

⁵Illumina, San Diego, CA, USA

⁶Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA

⁷Knight Cardiovascular Institute, Portland, OR, USA

⁸Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA

⁹Brotman Baty Institute for Precision Medicine, Seattle, WA, USA

¹⁰Howard Hughes Medical Institute, Seattle, WA, USA

¹¹Lead Contact

*Correspondence: yy224@uw.edu (Y.Y.), shendure@uw.edu (J.S.)

<https://doi.org/10.1016/j.molcel.2019.08.002>

SUMMARY

Conventional methods for single-cell genome sequencing are limited with respect to uniformity and throughput. Here, we describe sci-L3, a single-cell sequencing method that combines combinatorial indexing (sci-) and linear (L) amplification. The sci-L3 method adopts a 3-level (3) indexing scheme that minimizes amplification biases while enabling exponential gains in throughput. We demonstrate the generalizability of sci-L3 with proof-of-concept demonstrations of single-cell whole-genome sequencing (sci-L3-WGS), targeted sequencing (sci-L3-target-seq), and a co-assay of the genome and transcriptome (sci-L3-RNA/DNA). We apply sci-L3-WGS to profile the genomes of >10,000 sperm and sperm precursors from F1 hybrid mice, mapping 86,786 crossovers and characterizing rare chromosome mis-segregation events in meiosis, including instances of whole-genome equational chromosome segregation. We anticipate that sci-L3 assays can be applied to fully characterize recombination landscapes, to couple CRISPR perturbations and measurements of genome stability, and to other goals requiring high-throughput, high-coverage single-cell sequencing.

INTRODUCTION

Most contemporary single-cell genome-sequencing methods rely on compartmentalization of individual cells, which limits throughput, and/or PCR amplification, which skews uniformity. To address the former, we and colleagues developed single-cell combinatorial indexing (sci-), wherein one performs several

rounds of split-pool barcoding to uniquely index the nucleic acid contents of single cells, enabling exponential gains in throughput with each successive round of indexing. Sci-methods have been successfully developed to profile chromatin accessibility, transcriptomes, genomes, methylomes, and chromosome conformation in large numbers of single cells (Cao et al., 2017; Cusanovich et al., 2015; Mulqueen et al., 2018; Ramani et al., 2017; Vitak et al., 2017). To address the latter, linear amplification represents an alternative to PCR that has previously been used in single-cell assays (Eberwine et al., 1992; Hashimshony et al., 2012; Sos et al., 2016). For example, linear amplification via transposon insertion (LIANTI) uses the Tn5 transposon to fragment the genome and simultaneously insert a T7 RNA promoter for *in vitro* transcription (IVT) (Chen et al., 2017). By avoiding exponential amplification, LIANTI maintains uniformity and minimizes sequence errors. However, it remains low-throughput, requiring serial library preparation from each cell.

To address both limitations at once, we developed sci-L3, which integrates sci- and linear amplification. With three rounds of indexing, sci-L3 improves the throughput of LIANTI to at least thousands and potentially millions of cells per experiment, while retaining the advantages of linear amplification. We demonstrate the generalizability of sci-L3 by establishing methods for single-cell whole-genome sequencing (sci-L3-WGS), targeted genome sequencing (sci-L3-target-seq), and a co-assay of the genome and transcriptome (sci-L3-RNA/DNA). As a further demonstration, we apply sci-L3-WGS to map an unprecedented number of meiotic crossover and rare chromosome mis-segregation events in premature and mature male germ cells from both infertile, interspecific (B6 × Spretus) and fertile, intraspecific (B6 × Cast) F1 male mice.

DESIGN

The sci-L3 strategy has major advantages over current alternatives and any simple combination of sci- and LIANTI. First, its



potential throughput is >1 million cells per experiment at a low library preparation cost (Cao et al., 2019). Second, the unidirectional nature of the barcode structure of sci-L3 facilitates either whole-genome or targeted sequencing of single cells. Third, as a generalizable scheme for high-throughput cellular indexing coupled to linear amplification, sci-L3 can be adapted to additional goals with small modifications, as demonstrated here by our proof-of-concept of a single-cell RNA/DNA co-assay.

RESULTS

Proof-of-Concept of Sci-L3-WGS and Sci-L3-Target-Seq

The three-level combinatorial indexing and amplification schemes of sci-L3-WGS and sci-L3-target-seq are shown in Figure 1A: (1) Cells are fixed with formaldehyde and nucleosomes are depleted by SDS (Vitak et al., 2017); nuclei are distributed to a first round of wells. (2) A first round of barcodes is added by indexed Tn5 “tagmentation” within each well. A spacer sequence is included 5' to the barcodes as a “landing pad” for the subsequent ligation step (Figure 2; STAR Methods, “Methods and Molecular Design of Sci-L3-WGS and Sci-L3-Target-Seq”). (3) All nuclei are pooled and redistributed to a second round of wells; a second round of barcodes is added by ligation, together with a T7 promoter positioned outside both barcodes. (4) All nuclei are pooled and flow sorted to a final round of wells. Nuclei of different ploidies can be gated and enriched by DAPI staining. Also, simple dilution is an alternative to fluorescence-activated cell sorting (FACS), which can reduce loss. (5) Sorted nuclei are lysed and subjected to gap extension to form a duplex T7 promoter. This is followed by IVT, reverse transcription (RT), and second-strand synthesis (SSS). A third round of barcodes is added during SSS, along with unique molecular identifiers (UMIs) to tag individual IVT transcripts. (6) Duplex DNA molecules (Figure 1B, top), each containing three barcodes that define their cell of origin, are compatible with conventional library preparation for sci-L3-WGS (e.g., appending sequence adaptors by ligation [Figure 1B, center] or tagmentation) or slightly modified methods for sci-L3-target-seq (e.g., adding a PCR step with one target-specific primer [Figure 1B, bottom]).

As a proof-of-concept, we mixed mouse and human cells and performed sci-L3-WGS. For >95% of the resulting single-cell genomes, the vast majority of reads mapped either to the mouse or human genome (Figure 1C). The performance of sci-L3-WGS is compared to both LIANTI and our previous PCR-based sci-DNA-seq method (Vitak et al., 2017) in Table 1. The advantages of sci-L3-WGS include (1) we generally recover 90% of sorted cells as compared to 60% recovery with sci-DNA-seq. (2) With 40% fewer raw reads (329 million by sci-L3-WGS versus 549 million by sci-DNA-seq), sci-L3-WGS produced coverage at ~97,000 unique Tn5 insertions per cell, as compared to ~30,000 with sci-DNA-seq, a >3-fold improvement. Sequencing fewer cells to a higher depth, we observed ~660,000 unique Tn5 insertions per cell while maintaining higher library complexity, suggesting a further improvement of >20-fold. (3) The rate of mappable reads is improved from 61% with LIANTI to 86% with sci-L3-WGS. This is likely because LIANTI is entirely in-tube, making it hard to remove artifacts (e.g., secondary to self-insertion of Tn5), whereas with sci-L3-WGS, nuclei are pel-

leted several times to remove excess free DNA. (4) Unlike PCR-based methods wherein duplicate reads contain correlated errors, the “duplicate” reads of sci-L3-WGS almost always correspond to independent transcripts of the original template and are therefore useful for variant calling.

With sci-L3-WGS, Tn5 inserts on average every 0.5–1.5 kb of the human genome, and IVT yields ~1,000 transcripts. This corresponds to 2–6 million unique Tn5 insertions, and therefore 2–6 billion unique IVT transcripts per cell. It is obviously impractical to sequence these libraries to saturation. Here, we define depth as the ratio of unique transcripts sequenced to unique Tn5 insertions mapped. In this study, most libraries are sequenced at a depth of 1–2 times, resulting in 0.5%–5% coverage of the genome of each cell. The distribution of unique Tn5 insertions per cell in the human and mouse mixture experiment is shown in Figure 1D and for other experiments in Figure S1. The estimated relative chromosomal copy numbers for representative single cells is shown in Figure 1E and their distributions across all cells in Figure 1F. To extrapolate expected coverage per cell at higher depths, we fit the number of unique insertions as a function of depth (Figure S1G). We expect to observe 4.2 and 6.0 million unique insertions per cell at a depth of 5 and 10 times, respectively, which corresponds to 16% and 22% coverage of the genomes of individual cells.

For sci-L3-target-seq, after second strand synthesis, we added sequencing adaptors by PCR with one primer bearing the third barcode, but the other primer targeting a specific genomic region (Figure 1B, bottom). To quantify the efficiency of sci-L3-target-seq, we integrated a lentiviral CRISPR library at a low MOI (STAR Methods, “Methods and Molecular Design of Sci-L3-WGS and Sci-L3-Target-Seq”) and recovered the DNA sequences corresponding to single guide RNA (sgRNA) spacers by sci-L3-target-seq. For 97 of 1,003 single cells, we successfully recovered a single integrated sgRNA. This ~10% efficiency per haplotype is broadly consistent with the observed genome coverage of 22% with sci-L3-WGS (Figure S1G).

Note that at the molecular level, we have modified the sci- and LIANTI methods in several ways. Briefly, we (1) changed the design of the Tn5 transposon to be compatible with ligation, enabling a third round of indexing; (2) added a loop structure bearing the T7 promoter to facilitate intramolecular ligation, and (3) changed the RT scheme to only require successful ligation at one of the two ends of the first-round barcoded molecules. Supposing that a single ligation event has 50% efficiency, this modification renders a 75% success rate at the ligation step instead of 25% (Figure S1). We depict the structures of the molecules after each barcoding step in Figure 2 and discuss rationales, scalability, and costs for these designs in STAR Methods, “Methods and Molecular Design of Sci-L3-WGS and Sci-L3-Target-Seq” and Table S1. For libraries of 1,000, 10,000, and 1 million single cells, we estimate the cost of sci-L3-WGS to be 1.5%, 0.26%, and 0.014% of LIANTI. The use of 3, rather than 2, levels of combinatorial indexing can be leveraged either to increase throughput (e.g., the cost of constructing libraries for 1 million cells at a 5% collision rate with 3-level sci-L3-WGS is ~\$8,000) or to reduce the collision rate (e.g., the cost of constructing libraries for 10,000 cells at a 1% collision rate with 3-level sci-L3-WGS is ~\$1,500).

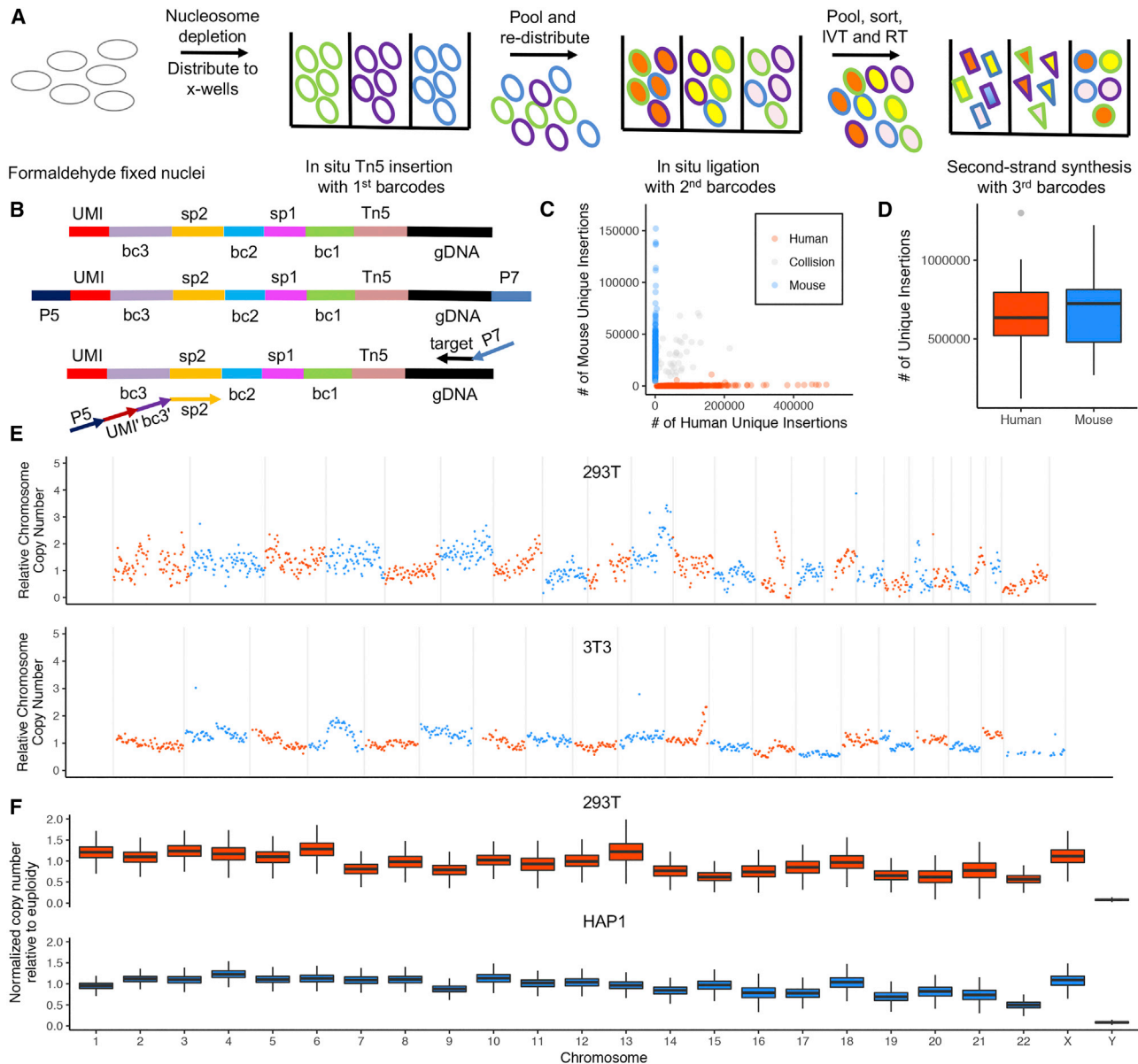


Figure 1. Sci-L3-WGS Enables High-Throughput Single-Cell Sequencing with Linear Amplification

(A) Sci-L3-WGS workflow.

(B) Top: barcode structure of resulting DNA duplexes. bc, barcode; sp, spacer; gDNA, genomic DNA. Center: example library structure for sci-L3-WGS. P5 and P7 sequencing adaptors are added by A-tailing and ligation. Note that having P7 on the UMI end and P5 on the gDNA end are equally possible due to symmetry of ligation. Bottom: example library structure for sci-L3-target-seq. P5 and P7 sequencing adaptors are added by priming from spacer 2 (sp2) and targeted loci of interest in the genome, respectively. Note that a new third round of barcode bc3' is also added by PCR corresponding to each bc3 in the WGS library, and new UMIs are added outside bc3'.

(C) Scatterplot of numbers of unique Tn5 insertions from human and mouse cells at low sequencing depth, 24 bc1 × 64 bc2 × 6 bc3 sci-L3-WGS, 100–300 cells sorted per well. Blue, inferred mouse cells (percentage of mouse reads >95%; median 98.7%; n = 315); red, inferred human cells (percentage of human reads >95%; median 99.8%; n = 719); gray, inferred collisions (n = 48; 4% of cells). “Contaminating” reads arise randomly throughout the genome (see Figure S1H). (D) Boxplots showing number of unique Tn5 insertions per cell at mean 2.4 million raw reads per cell and 1.78 times depth. Depth defined as ratio of unique IVT transcripts to unique Tn5 insertions. Thick horizontal lines, medians; upper and lower box edges, first and third quartiles, respectively; whiskers, 1.5 times the interquartile range; circles, outliers. See Figure S1 and STAR Methods, “Methods and Molecular Design of Sci-L3-WGS and Sci-L3-Target-Seq” for characterization of the libraries made with improved versions of the protocol.

(E) Example chromosome CNV plots for individual cells. Top, HEK293T cell, 2.6 million raw reads, 2.4 million unique molecules, 1.3 million unique Tn5 insertions with MAPQ >1. Bottom, 3T3 cell, 2.7 million raw reads, 2.4 million unique molecules, 1.2 million unique Tn5 insertions with MAPQ >1.

(F) Boxplots for copy number variation across 822 293T cells or 1,453 HAP1 cells. The y axis depicts reads fraction per chromosome normalized by chromosome length such that a euploid chromosome without segmental copy gain or loss is expected to have a value of 1.

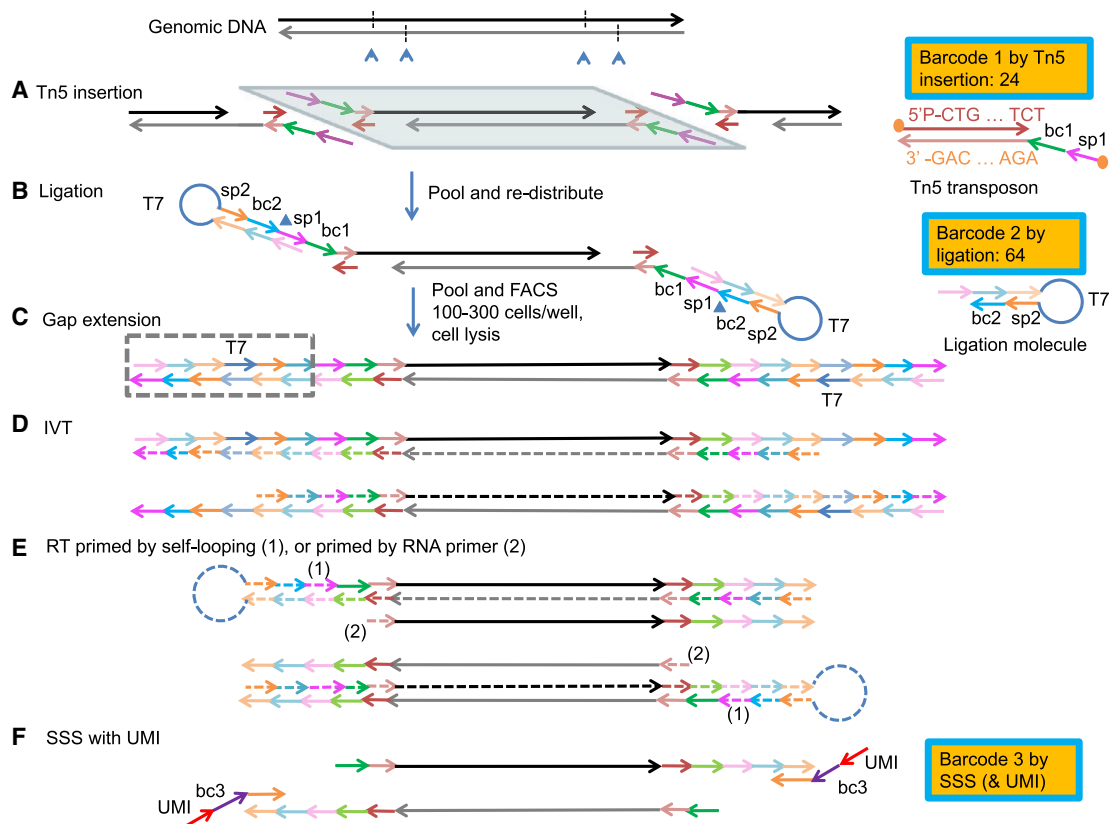


Figure 2. Molecular Structures for Sci-LIANTI at Each Step

(A) Tn5 adaptors have both 5' ends phosphorylated, one required for insertion and the other for ligation. The overhang of the annealed transposon contains first round barcodes (bc1) and a spacer (sp1) for ligation.

(B) The ligation molecule is pre-annealed as a hairpin loop, which reduces intermolecular ligation from three molecules to two molecules; the hairpin structure also helps improve RT efficiency in downstream steps. The hairpin contains (1) an overhang that anneals with sp1 for ligation, (2) the second round barcodes (bc2) and a spacer (sp2) that serves as a priming site in the stem for SSS in downstream steps, and (3) a T7 promoter in the loop for IVT.

(C) Gap extension converts the looped T7 promoter to a duplex. Note that if ligation is successful on both ends, then T7 promoters are present on both sides; however, if ligation is successful on one end, then the boxed portion will be missing. Nevertheless, both can be reverse transcribed in downstream steps with different RT primers.

(D) IVT generates single-stranded RNA amplicons downstream of the T7 promoter.

(E) If ligation is successful on both ends, then RT is preferably primed by self-looped RT primers, which are inherited from the looped ligation molecule; if ligation is successful on only one end, then RT is primed by additional RNA RT primers added in excess. Excess RNA primers are then removed before SSS to avoid interfering with SSS reaction.

(F) Double-stranded DNA molecules are produced by SSS, which primes off sp2 to simultaneously add the third barcode and to UMI tag each transcript.

Dashed line: RNA, solid line: DNA. For more details, see STAR Methods, "Methods and Design of Sci-L3-WGS and Sci-L3-Target-Seq."

Development of a Scalable Single-Cell RNA/DNA Co-assay

We realized that sci-L3 could be further adapted to other nucleic acid targets with small modifications. To illustrate this, we developed a sci-L3-RNA/DNA co-assay. In brief, the first round of DNA barcoding is performed by Tn5 insertion as in sci-L3-WGS, but we concurrently perform a first round of RNA barcoding, tagging mRNAs via RT with a barcode and UMI-bearing polyT primer (Figure 3A). Both the Tn5 insertion and RT primer bear overhangs that mediate the ligation of the second round of barcodes and a T7 promoter, effectively enabling three-level indexing and subsequent IVT-based linear amplification in a manner largely identical to sci-L3-WGS (Figures 3A and 3B; STAR Methods "Methods and Molecular

Design of Sci-L3-RNA/DNA Co-assay"). As a proof-of-concept, we mixed mouse cells with cells from two human cell lines and performed the sci-L3-RNA/DNA co-assay. For the vast majority of cells, reads mapped either to the mouse or human genome, both for RNA (5.2% collision rate) and DNA (6.6% collision rate) (Figures 3C and 3D). Furthermore, consistent with a successful co-assay, 100% of cells were assigned the same species label by their RNA and DNA profiles. As a further check, we performed t-Distributed Stochastic Neighbor Embedding (t-SNE) based on their RNA profiles, resulting in two clusters. Labeling each cell by the presence or absence of a Y chromosome in the DNA profiles coherently identified BJ (male) versus HEK293T cells (female) (Figure 3E) with 96.5% accuracy.

Table 1. Performance Comparison of Sci-DNA-Seq, Sci-L3-WGS, and LIANTI

	Ideal No. Reads/			Ideal No. Reads/			Median			Median		
	No. Raw Reads (Million)	No. Cells Sorted	Single Cell Read Cutoff	No. Cells Recovered	Cells Recovered (%)	Recovered Single Cell (Thousand)	Mapping Rate of All Raw Reads (%)	Unique Insertion/Single Cell (Thousand)	Median Tn5 Insertions' Complexity (%)	Median Library Complexity (%)	No. Cells >5e4 Reads (%)	Cells >5e4 Reads (%)
Sci-DNA-seq	549	6,336	6,052	3,123	49 ^a	175.8	85	29.5 ^b	53 ^b	53 ^c	1,056 ^d	17 ^d
Sci-L3-WGS (low depth)	329	2,400	18,945	2,200	92 ^a	149.5	87	97.3 ^b	95 ^b	99 ^c	1,828 ^d	83 ^d
Sci-L3-WGS (high depth)	256	200	159,012	191	96	1,340.3	86 ^e	660.7 ^b	73 ^b	98 ^c	191	96
LIANTI	3.84	3	1,280,000	3	100	1,280	61 ^e	789.5	98	99 ^c	3	100

Sci-DNA-seq data from xSDS method of Vitak et al. (2017). LIANTI from in-tube method of Chen et al. (2017). For sci-L3-WGS, we show results for libraries yi140 and yi141 (high sequencing depth) and yi144 and yi145 (low sequencing depth). These 4 libraries use an optimized protocol with concentrated Tn5 transposome (0.2 μM) and an improved RT reaction with additional RNA primers (see Figure S1 and STAR Methods, "Methods and Molecular Design of Sci-L3-WGS and Sci-L3-Target-Seq" for details). Superscripts indicate comparisons of interest.

^aPercentage of recovered single cells from sorting is improved 1.9-fold with sci-L3-WGS over sci-DNA-seq.

^bUnique insertions at varying sequencing depth; rows 1 and 2 are compared at similar number of raw reads, with 3.3-fold improvement with 40% fewer raw reads with sci-L3-WGS over sci-DNA-seq; rows 1 and 3 are compared at similar library complexity with 22.4-fold improvement at 20% better Tn5 insertion complexity with sci-L3-WGS over LIANTI.

^cMedian library complexity showing methods including both LIANTI and sci-L3-WGS have minimal PCR duplicates.

^dNumber of cells with >50,000 unique reads recovered are improved 1.8-fold with sci-L3-WGS over LIANTI.

^eMapping rate of raw reads is improved 1.4-fold with sci-L3-WGS over LIANTI.

Single-Cell DNA Profiling of Mouse Germ Cells with Sci-L3-WGS

In normal mitotic cell divisions, diploid chromosomes undergo replication to generate four copies of DNA, and sister chromatids segregate apart into reciprocal daughter cells. Daughter cells receive one copy of each maternally and paternally inherited DNA sequence and almost always maintain heterozygosity at the centromere-proximal sequences (Figure S2A). Rarely, chromosomes undergo mitotic crossover between chromosome homologs, which can sometimes result in diploid cells with loss-of-heterozygosity (LOH) at sequences centromere distal to the crossover if the two recombined chromatids segregate into different daughter cells (Figures S2B and S2C).

In meiosis, sister chromatids first co-segregate into the same daughter cell, and homologs segregate into reciprocal daughter cells in the meiosis I (MI) stage, also known as reductional segregation, resulting in 2C cells (DNA content of an unreplicated diploid cell) with LOH at the centromere proximal sequences (Figures S2D and S2E). For successful reductional segregation of chromosomes in MI (Figure S2D), crossovers initiated by Spo11-catalyzed double-strand breaks (DSBs) (Baudat et al., 2000; Keeney et al., 1997; Romanienko and Camerini-Otero, 2000) provide the link and necessary tension (Hong et al., 2013) between chromosome homologs. Rarely, chromosomes will segregate in a meiotic fashion without any inter-homolog crossover, resulting in uniparental disomy (UPD). After MI, these 2C cells undergo mitosis-like chromosome segregation in meiosis II (MII), also called equational segregation, such that sister chromatids segregate apart to form 1C gametes (Figure S2E). Below we refer to meiotic or reductional segregation during MI, in which sister chromatids segregate together, as reductional segregation, and mitosis-like or equational segregation during MII, in which sister chromatids segregate apart, as equational segregation.

To date, most work on the relation between crossover position and chromosome segregation has been performed by imaging (Wang et al., 2017a, 2017b), which fails to fully characterize the underlying genomic sequences that are prone to meiotic crossover. Several assays enable detailed mapping of meiotic DSB hotspots (Lange et al., 2016; Smagulova et al., 2011, 2016), but these assays do not directly map meiotic crossovers. Assays that do dissect crossover versus noncrossover at a fine scale are restricted to a few hotspots (Cole et al., 2014). Consequently, we know much less about the relation between crossovers and chromosome-scale features such as replication domains than we do about meiotic DSB hotspots (Baudat et al., 2013; Choi and Henderson, 2015; Yamada et al., 2017). Genome-wide meiotic crossover maps have been generated by mapping tetrads in yeast (Mancera et al., 2008; Zhang et al., 2017), single human sperm, and complete human female meioses (Hou et al., 2013; Lu et al., 2012; Ottolini et al., 2015; Wang et al., 2012). With the exception of the studies of human female meiosis, which analyzed 87 complete meioses, most crossover maps are limited in at least three respects: (1) mature 1C gametes are analyzed in which the cells have completed both rounds of meiotic division, which prevents direct observation of the more informative intermediate 2C cells to evaluate whether and how often chromosomes undergo reductional versus equational

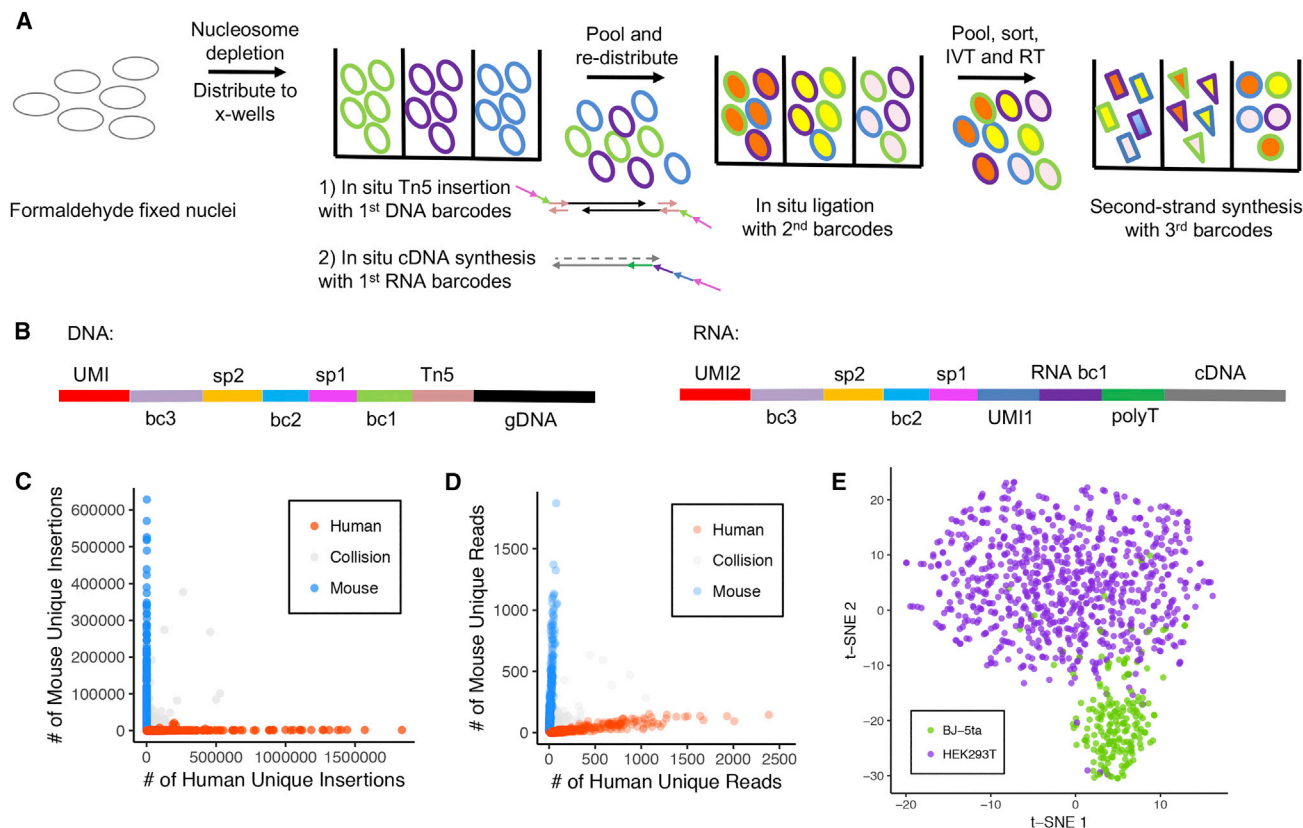


Figure 3. Sci-L3-Based RNA/DNA Co-assay Enables Scalable Joint Profiling of Single-Cell Genomes and Transcriptomes

(A) Schematic of sci-L3-RNA/DNA co-assay. Note that both Tn5 transposon and cDNA synthesis primer contain the same phosphorylated ligation landing pad (pink) at 5' overhang outside the first round barcodes.

(B) Barcode structures of resulting amplified duplexes corresponding to genome (left) and transcriptome (right).

(C) Scatterplot of numbers of unique Tn5 insertions from human and mouse cells at low and high sequencing depth plotted together. Blue, inferred mouse cells (percentage of mouse reads >95%, median of 99.5%; $n = 2002$); red, inferred human cells (percentage of human reads >95%; median of 99.8%; $n = 2,419$); gray, inferred collisions ($n = 149$; 6.6%).

(D) Same as in (C) for RNA. Blue, inferred mouse cells (median purity 95.1%); red, inferred human cells (median purity 91.5%); gray, inferred collisions ($n = 272$; 12%).

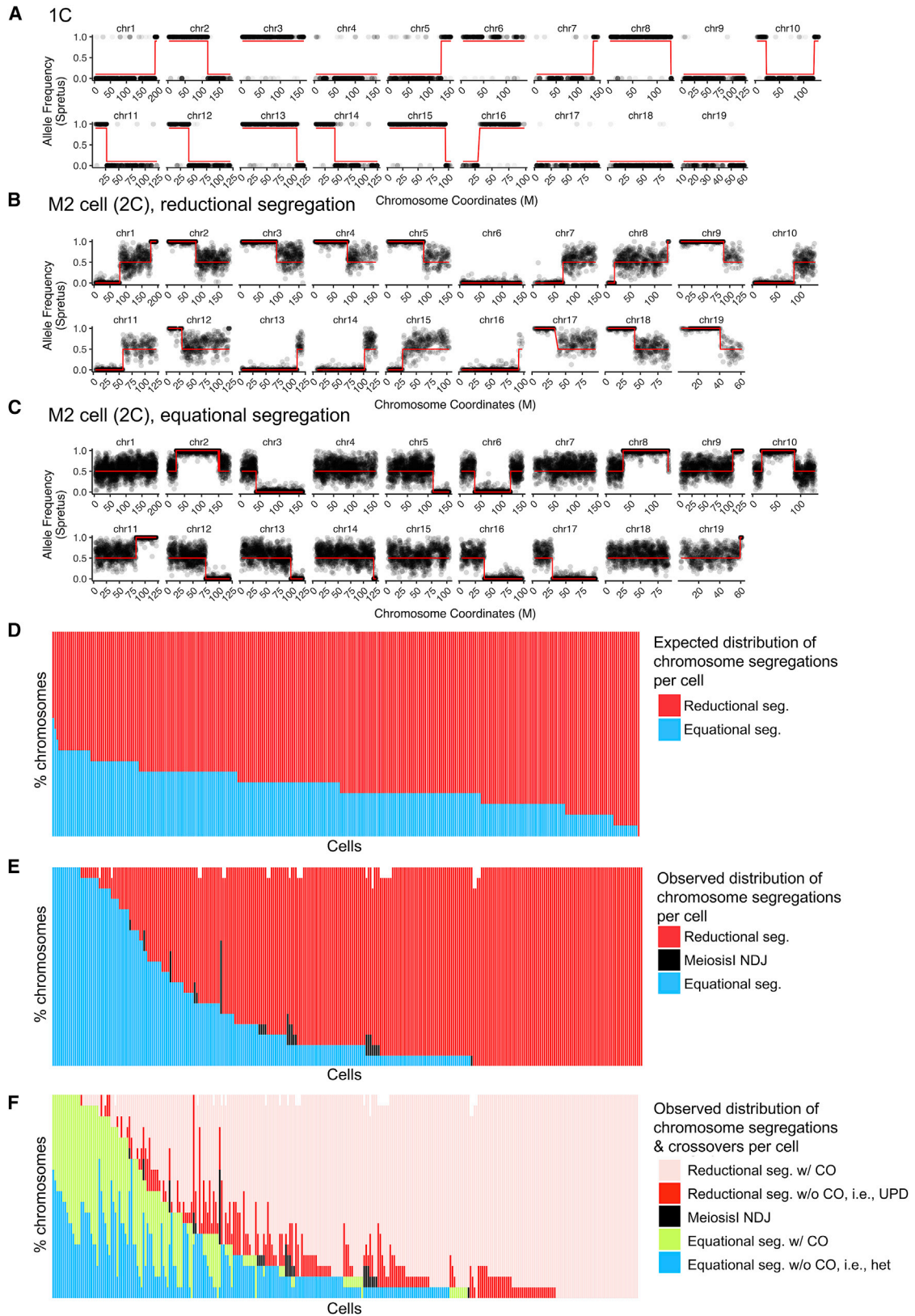
(E) t-SNE based on RNA profiles results in two clusters corresponding to BJ (male) and HEK293T (female) cells. Colors based on the presence or absence of Y chromosomes in DNA profiles.

segregation during MI (Figure S2); (2) abnormal cells are selected against due to their failure to proceed to the mature gametic state; and (3) analyses by single sperm or oocyte sequencing are limited in throughput and to a few hundred cells at the most, and as such could miss out on rare events. Even for fertile crosses, the number of offspring that can be reasonably generated and genotyped is quite limited (Liu et al., 2014).

To address all of these limitations, we applied sci-L3-WGS to infertile offspring of an interspecific cross (female *Mus musculus domesticus* C57BL/6 [B6] x male *Mus spretus* SPRET/Ei [Spret]) and fertile offspring of an intraspecific hybrid (female B6 x male *Mus musculus castaneus* CAST/Ei [Cast]). By sequencing sperm with a scalable technology, we are able to map an unprecedented number of crossover events for a mammalian system, in both infertile and fertile hybrids. Also, as this scale also enables us to recover profiles from rare 2C secondary spermatocytes, we can assess crossover and chromosome mis-segregation simultaneously from the same single cells.

Unlike inbred males and (B6 x Cast) F1 males, the epididymides of (B6 x Spret) F1 males (Berletch et al., 2015) contain extremely few morphologically mature sperm and limited numbers of round germ cells of unknown ploidy (Figures S3A and S3B). We observed a much higher fraction of 2C cells during FACS (Figures S3C and S3D; Table S2) than would be expected for a “normal” epididymis, which is dominated by 1C sperm. In contrast and as expected, the epididymides of (B6 x Cast) F1 males contained almost entirely 1C sperm (Figure S3E). For this cross, we therefore sorted 1C and 2C cells from dissociated testes (Figure S3F).

For cells from F1 males from both the (B6 x Spret) and (B6 x Cast) crosses, we performed sci-L3-WGS (details in STAR Methods, “Setup of Sci-L3-WGS Experiment in Two Crosses”). Although 1C and 2C cells can be distinguished informatically, their relative abundance still affects our analysis. Specifically, in the (B6 x Spret) cross, 1C cells are so rare that any doublets (e.g., two 1C cells stuck together or that incidentally receive the



(legend on next page)

same barcodes) do not substantially contribute to the 2C population. In contrast, in the (B6 × Cast) cross, the majority of cells are 1C despite enrichment (~85%; Figure S3G), such that there may be many 1C doublets that mimic 2C cells. We discuss how to informatically distinguish 1C doublets from bonafide 2C cells further below.

M2 Cells Exhibit Clustered Reductional or Equational Chromosome Segregation

Chromosome Segregation in M2 Cells from the Infertile (B6 × Spret) Cross

We first sought to analyze meiosis in cells from the epididymides of infertile (B6 × Spret) F1 males. Across 2 experiments, we profiled the genomes of 2,689 (92% of 2,919 sorted cells with >10,000 raw reads) and 4,239 (94% of 4,497 sorted cells with >30,000 raw reads) single cells (Figure S1F). At a depth of 1.6 and 1.4 times for the 2 libraries, we obtained a median of ~70,000 and ~144,000 unique Tn5 sites per cell, corresponding to 0.7% and 1.4% median genome coverage, respectively.

To identify crossover breakpoints, we implemented a hidden Markov model (HMM) that relied on high-quality reads that could clearly be assigned to B6 versus Spret (see STAR Methods, “Bioinformatic and Statistical Analyses”; Tables S3 and S4). We characterized crossovers in 1,663 1C cells (Figure 4A). Although the ~5,200 2C cells were expected to be overwhelmingly somatic, to our surprise, we identified 292 with a significant number of crossovers, which we called M2 cells (Figures 4B and 4C). Even more surprising, a substantial proportion of these exhibited equational, rather than reductional, segregation.

After an inter-homolog crossover occurs, if the chromosome segregates in a reductional fashion, then the region between the centromere and the position of crossover will become homozygous, whereas heterozygosity will be maintained downstream of the crossover (Figure S2D). However, if the chromosome segregates in an equational fashion, then LOH is observed centromere distal to the crossover if the recombined chromatids segregate apart (Figure S2B). An example of an M2 cell exhibiting the expected reductional segregation is shown in Figure 4B (note homozygosity between centromere and point of crossover), and an example of an M2 cell exhibiting the unexpected equational segregation is shown in Figure 4C (note consistent heterozygosity between centromere and point of crossover).

Within any given M2 cell, are the segregation patterns of individual chromosomes independent? If that were the case, across cells, then we would expect a binomial distribution of reductionally versus equationally segregated chromosomes, centered on the maximum likelihood estimate (MLE) of the probability, p , of reductional segregation ($p = 0.76$ from the data, 4,162/5,472; Figure 4D). However, of the 292 profiled M2 cells, we observe 202 cells with ≥ 15 reductionally segregated chromosomes (148 expected), and 38 cells with ≥ 15 equationally segregated chromosomes (0 expected) (Figure 4E; $p = 4e^{-23}$, Fisher’s exact test). This non-independence suggests the possibility of a cell-autonomous global sensing mechanism for deciding whether a cell proceeds with meiosis or returns to mitosis.

We can further classify cells by whether chromosomes in M2 cells have a crossover (Figure 4F). Reductionally segregated chromosomes appear to have more crossovers (pink in Figure 4F) than equationally segregated chromosomes (green in Figure 4F). Across the 292 M2 cells, we observed 4,162 examples of reductional segregation (90% with crossovers) and 1,310 examples of equational segregation (49% with crossovers). However, unlike in reductionally segregated chromosomes in which we can detect all of the crossovers as centromeric LOH, equationally segregated chromosomes have LOH only if the two recombined chromatids segregate apart into reciprocal daughter cells (Figure S2B). If instead recombined chromatids co-segregate, heterozygosity will be maintained throughout the chromosome, despite the undetectable linkage switch (Figure S2C). In Figure 4F, the ratio of having (shown in green) versus not having (shown in blue) an observable LOH in equationally segregated chromosomes is roughly 1:1. This could either mean that equationally segregated chromosomes have a 50% chance of segregating recombined chromatids together, if those completely heterozygous chromosomes (shown in blue) do have a linkage switch, or alternatively that equationally segregated chromosomes always segregate recombined chromatids apart, and the crossover frequency is reduced by half compared to reductionally segregated chromosomes.

Segmental or whole-chromosome LOH are known to be rare in mammalian mitotic cells. Nevertheless, to rule out a mitotic origin of such events, we examined such events in the Patski cell line, which is a spontaneously immortalized cell line derived from the female (B6 × Spret) F1 mouse. We analyzed 1,107 single cells

Figure 4. Sci-L3-WGS of Interspecific Hybrid Mouse Male Germline Reveals Numerous Examples of Non-independent Equational Segregation in MI

(A) Example crossover plot for 1C cell. The gray dot has a value of 1 for the Spret allele and 0 for the B6 allele.

(B) Example LOH plot for M2 cell with reductional segregation (see also Figure S2D): LOH centromere proximal to the crossover.

(C) Example LOH plot for M2 cell with equational segregation (see also Figure S2B): LOH centromere distal to the crossover, unlike in (B).

(D–F) Number of reductionally (red, pink, black) and equationally (blue, green) segregated chromosomes for each M2 cell. Each column represents 1 M2 cell (19 chromosomes per cell, distributed as indicated by colors).

(D) Expected distribution of reductional versus equational segregation based on binomial distribution, with $p = 0.76$ for reductional segregation.

(E) Observed data in M2 cells. In rare cases (27/5,548 chromosomes), we were not able to distinguish reductional versus equational segregation due to sparse SNP coverage (white space at the top of the panel). The black bar depicts MI nondisjunction (NDJ, 40 chromosomes in total), where we observed 0 or 4 copies of the chromatids. Note that NDJ is considered to be reductional segregation because the sister chromatids segregate together.

(F) Same as (E), but further broken down by the number of chromosomes with or without crossovers (abbreviated “CO”). Cells are sorted first by the number of equationally segregated chromosomes (light green and blue, in descending order) and then by the number of observed equationally segregated chromosomes without crossover (blue, in descending order).

In (A)–(C), the red line depicts fitted crossover transition via HMM. The centromere is located at left for a picture of each chromosome. In (B) and (C), the gray dot shows the allele frequency of Spret averaging 40 SNPs.

from Patski with sci-L3-WGS, among which we found an average of 0.36 UPD chromosomes and 0.098 segmental LOH events per cell (Table S3), a much reduced rate compared to M2 cells. Because such events are not necessarily independent (e.g., due to UPD during early passaging), the rate of independent LOH events is likely even lower. The distribution of these events is plotted in Figure S4M. The contrast between the low rate of mitotic LOH (expected) and the relatively high rate of 2C cells exhibiting equational segregation (unexpected) confirms that the latter are unlikely to correspond to somatic cells.

Chromosome Segregation in M2 Cells from the Fertile (B6 × Cast) Cross

We wondered whether equational segregation also occurs during MI in the fertile progeny of intraspecific (B6 × Cast) F1 males. As shown above, the epididymides from this cross consist almost entirely of 1C mature sperm; we therefore enriched for 2C secondary spermatocytes from whole testes. We then performed sci-L3-WGS on cells from both the epididymides and testes.

In a first quality control (QC) experiment, we distributed 1C round spermatids evenly and sorted only for 1C cells after two rounds of barcoding. The doublets, identified by virtue of being non-1C, allow us to quantify barcode collisions. Among 2,400 sorted cells (200/well), we recovered 2,127 (89%) with >7,000 reads per cell; 2,008 of these are 1Cs with meiotic crossovers, indicating a barcode collision rate of 5.5%. At a sequencing depth of 1.06 times, we obtained a median of ~60,000 unique Tn5 insertions per cell, corresponding to ~0.6% median genome coverage.

In a second experiment, we tagged 1C round spermatids from the testes (barcode group 1), 2C cells from the testes (barcode group 2; contaminated with large numbers of 1C spermatids, as shown in Figure S3F), and 1C mature sperm from the epididymis (barcode group 3; STAR Methods, “Setup of Sci-L3-WGS Experiment in Two Crosses”) in separate wells during the first round of barcoding. The rationale for separating barcode groups 1 and 2 was to test whether instances of whole-genome equational segregation were an artifact consequent to doublets (discussed further below). As a further enrichment, during the FACS step of sci-L3-WGS, for a subset of wells, we specifically gated for 2C cells (15.5% of all cells; Figure S3G). At a sequencing depth of 1.09 times, we obtained a median of ~94,000 unique Tn5 insertions per cell, corresponding to ~0.9% median genome coverage.

In total, we recovered 3,539 1C and 1,477 non-1C cells from this second experiment. More than 97% of the 1C cells derive from barcode groups 1 ($n = 1,853$) and 2 ($n = 1,598$) rather than 3 ($n = 88$), indicating that mature sperm from the epididymis are not well recovered by sci-L3-WGS. This suggests that the 1C cells recovered from (B6 × Spret) cross above are also likely not from mature sperm but rather from round spermatids, which is consistent with the low number of sperm with mature morphology (Figure S3B).

The 1,477 non-1C cells derived from both barcode group 1 ($n = 1,104$; presumably doublets of 1C round spermatids) and 2 ($n = 373$; presumably a mixture of bona fide M2 cells and 1C doublets). To identify a signature of 1C doublets, we examined the profiles of non-1C cells from barcode group 1 (which was

specifically pre-sorted for 1C content and unlikely to contain bona fide M2 cells). The centromere-proximal SNPs of 1C cells that have completed both rounds of meiotic divisions should either be B6 or Cast derived. For 1C doublets, these regions have an equal chance of appearing heterozygous or homozygous. Therefore, within any given 1C doublet, the number of chromosomes that appear to have segregated equationally, as well as the number that appear to have segregated reductionally, should follow a binomial distribution, with $n = 19$ and $p = 0.5$. This is what we observe for 1C doublets from barcode group 1 ($p = 0.53$, chi-square test; Figures 5A and 5B). In fact, there were only 11 1C doublet cells with at least 15 chromosomes that appear to segregate in a consistent fashion, whether equationally or reductionally (Tables S2 and S6).

Non-1C cells from barcode group 2 exhibited a very different distribution. Of 373 such cells, 258 are similar to the 1C doublets of barcode group 1 in having similar numbers of chromosomes with equational or reductional segregation patterns. The remaining 115 cells are biased, with at least 15 chromosomes segregating in a consistent fashion, whether equationally or reductionally (Figures 5C–5E; 115/373 for barcode group 2 versus 11/1,104 for barcode group 1; $p = 3e-70$, chi-square test; Table S6), with some exhibiting completely equational ($n = 6$) or completely reductional ($n = 91$) patterns.

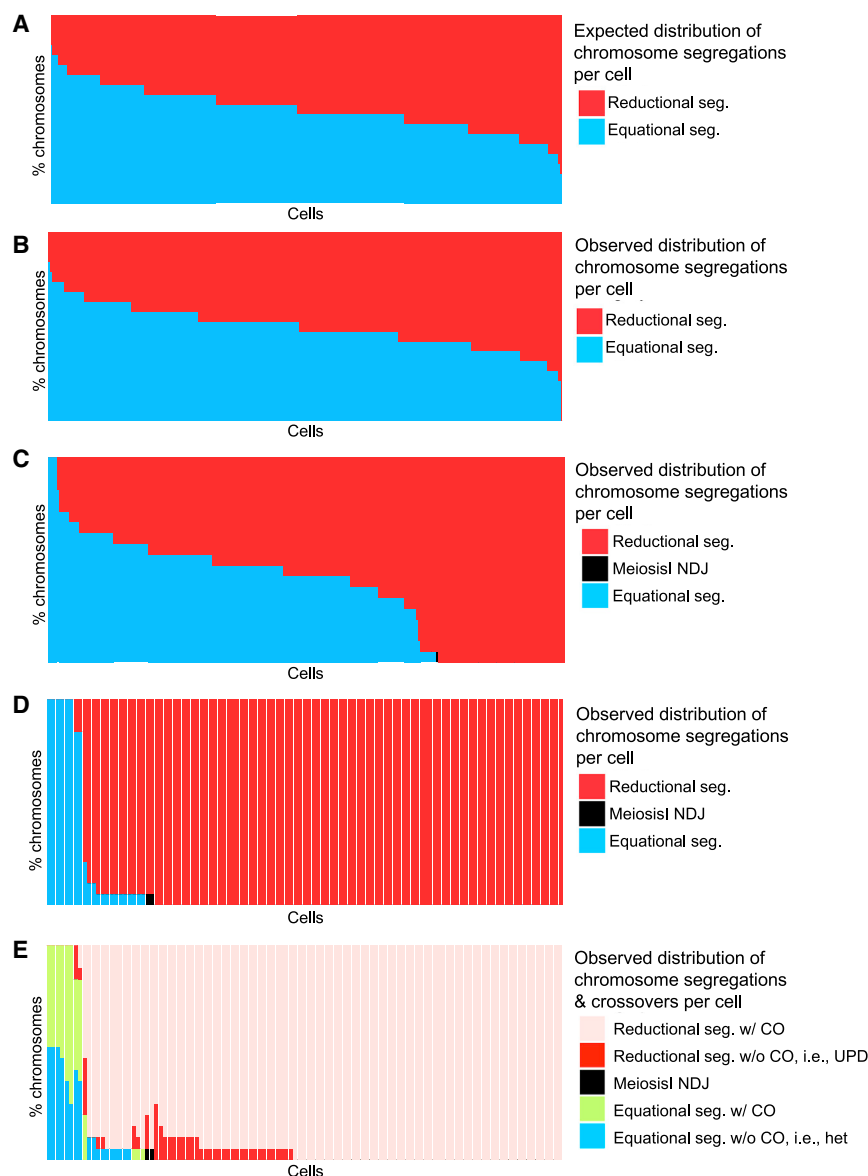
Finite-Mixture Model for Fitting the Three Populations of Non-1C Cells

To consider this more formally, we fit the data from each experiment to a Bayesian finite mixture of three binomial distributions (STAR Methods, “Finite Mixture Model for Fitting the Three Populations of Non-1C Cells”; Figure S3). The non-1C cells from the testes of intraspecific (B6 × Cast) F1 males (barcode group 2) are estimated to include subsets of cells segregating reductionally (28%) versus equationally (2%), as well as likely 1C doublets (69%) (Figure S3I). The proportions differ for M2 cells from the interspecific (B6 × Spret) F1 males, which are estimated to include subsets of cells segregating reductionally (66%) versus equationally (14%), as well as likely 1C doublets (20%) (Figure S3J). These analyses support the conclusion that the infertile (B6 × Spret) cross has a much higher proportion of cells biased toward equational rather than reductional segregation.

Distribution of Meiotic Crossovers at the Chromosomal Level

We next sought to investigate the genomic correlates of crossover events. We analyzed 1,663 1C cells harboring 19,601 crossover breakpoints and 240 M2 cells with 4,184 crossover breakpoints from the (B6 × Spret) cross, and 5,547 1C cells harboring 60,755 crossover breakpoints and 115 M2 cells with 2,246 crossover breakpoints from the (B6 × Cast) cross. To our knowledge, this is an unprecedented dataset with respect to the number of crossover events identified in association with mammalian meiosis.

The high-throughput nature of sci-L3-WGS allowed us to analyze large numbers of premature germ cells and identify the rare cell population that has completed MI but not MII, and thus to observe meiotic crossover and chromosome mis-segregation events in the same cell. In comparing an infertile,



interspecific (B6 × Spret) hybrid with a fertile, intraspecific (B6 × Cast) hybrid at a chromosomal level, we observe the following defects in MI: (1) the proportion of M2 cells that have at least 1 crossover on all 19 autosomes is reduced from $\sim 2/3$ in (B6 × Cast) to $\sim 1/2$ in (B6 × Spret); (2) the average number of crossovers per M2 cell is lower in (B6 × Spret), but the average number of crossovers per 1C cell is higher; (3) crossover interference is weaker in (B6 × Spret), in which the median distance between adjacent crossovers is reduced from 97 to 82 Mb; (4) in (B6 × Spret) M2 cells, crossovers tend to occur in the middle half of each chromosome arm, in contrast to 1Cs of both crosses as well as (B6 × Cast) M2 cells, where they favor the most centromere distal quartile; (5) among M2 cells with biased equational or reductional chromosome segregation, (B6 × Spret) exhibits a significantly higher proportion (38/240) of whole-genome equational segregation than

Figure 5. Sci-L3-WGS of the Intraspecific Hybrid Mouse Male Germline Also Reveals Numerous Examples of Non-independent Equational Segregation

(A and B) Number of reductionally and equationally segregated chromosomes for artificial “2C” cells from barcode group 1, which derive from doublets of 2 random 1C cells. Same depiction as in Figure 4.

(A) Expected distribution of reductional versus equational segregation based on the binomial distribution and assuming the probability of equational segregation; $p = 0.5$.

(B) Observed data in 2C cells, which matches the expected distribution shown in (A).

(C–E) Number of reductionally and equationally segregated chromosomes for non-1C cells from barcode group 2, which are a mixture of both artificial doublets of 2 random 1C nuclei and real 2C secondary spermatocytes.

(C) All non-1C cells from barcode group 2.

(D) Non-1C cells with biased chromosome segregation only (i.e., ≥ 15 chromosomes segregated either equationally or reductionally). Black bar depicts Meiosis I NDJ (2 of 2,185 chromosomes).

(E) Same as (D), but further broken down by the number of chromosomes with or without crossovers.

(B6 × Cast) (6/115); and (6) among M2 cells, the average number of sporadic equational segregations (also called reverse segregations [Ottolini et al., 2015]) is increased from 0.2 to 1.1. These findings suggest mechanisms that could contribute or reflect underlying factors that contribute to the infertility of (B6 × Spret) F1 males, including defects in crossover formation and positioning, compromised mechanisms for ensuring at least one crossover per chromosome, and an increase in both sporadic and whole-genome equational segregation. Details are presented in Figure S4 and

STAR Methods, “Distribution of Meiotic Crossovers at the Chromosomal Level.”

Distribution of Meiotic Crossover Events in Relation to the Landscape of the Genome

We next evaluated the distribution of crossovers at a finer scale in three ways (details in STAR Methods, “Distribution of Meiotic Crossover Events in Relation to Genomic Features”). First, we collapsed all of the crossover events to generate “hotness maps” along each chromosome and compared these to meiotic DSB maps (Brick et al., 2018; Smagulova et al., 2011, 2016; Lange et al., 2016), using Bayesian model averaging (BMA) to identify crossover-contributory features beyond Spo11 (Clyde et al., 2011; Figures 6A and 6B). Many, but not all, of the resulting features are consistent between the two crosses. For example, the positional biases of crossover formation, which can greatly

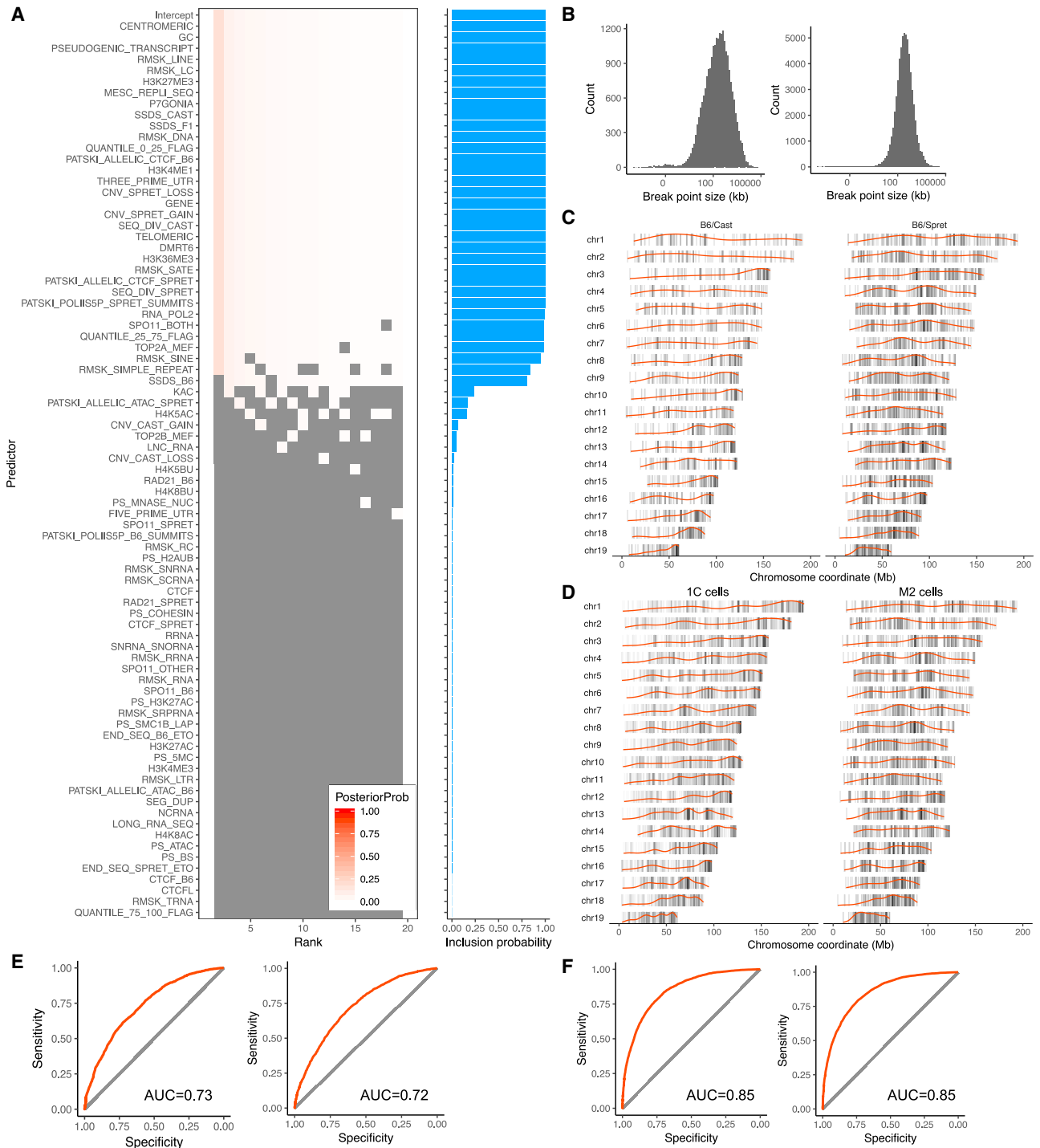


Figure 6. Meiotic Crossover Hotness and Explanatory Genomic Features

(A) Marginal inclusion probability (MIP) for features associated with crossover hotness by BMA. The x axis ranks models by posterior probability, where gray boxes depict features not included in each model (vertical line, 20 top models are shown) and orange color scale depicts posterior probability of the models. The combined dataset from both the (B6 × Spret) and (B6 × Cast) crosses is shown here. See Figure S5 for the two crosses analyzed separately.

(B) Log normal distribution of sizes for breakpoint resolution. Left: (B6 × Spret), median of 150 kb. Right: (B6 × Cast), median of 250 kb.

(C and D) Positions of the rightmost crossover of each chromosome.

(E) M2 cell. Crossovers in the (B6 × Cast) (left) cross prefer the centromere distal end of the chromosome, while crossovers in the (B6 × Spret) cross (right) prefer the center region of each chromosome arm. After accounting for inter-chromosome variability, we estimate that crossovers in the (B6 × Spret) cross are on average 5.5 Mb more centromere proximal. See Figure S7A, which is similar but for 1C cells.

(legend continued on next page)

affect the amount of tension enforced between chromosome homologs and consequently segregation, appear to be different (Figures 6C and 6D). Second, in both crosses, we found that 1C and M2 cells separated into 2 clusters upon principal-component analysis (PCA) on 78 aggregate crossover-related genomic features, suggesting cell-autonomous differences in terms of breakpoint patterns. Third, we constructed a predictive model of crossover locations and achieved an accuracy of 0.73 and 0.85 in distinguishing real crossover tracts from randomly sampled genomic tracts, in (B6 × Spret) and (B6 × Cast) crosses, respectively (Figures 6E and 6F).

DISCUSSION

Here, we describe sci-L3, a framework that combines three-level single-cell combinatorial indexing and linear amplification. We demonstrate a sci-L3-WGS, targeted DNA sequencing (sci-L3-target-seq), and a genome and transcriptome co-assay (sci-L3-RNA/DNA). With sci-L3-WGS, at least tens of thousands, and potentially millions, of single-cell genomes can be processed in a 2-day experiment, at a library construction cost of \$0.14 per cell for 10,000 cells and \$0.008 per cell for 1 million cells. The throughput of sci-L3-WGS is orders of magnitude higher than alternative sci-L3-WGS methods based on linear amplification, such as in-tube LIANTI (Chen et al., 2017). It furthermore improves on the number of unique molecules recovered from each single cell from the low thousands (Pellegrino et al., 2018) or low tens of thousands (Vitak et al., 2017) to the hundreds of thousands.

We applied sci-L3-WGS to study male mouse meiosis and identified an unexpected population of M2 cells. The single-cell nature of the data also allowed us to simultaneously characterize meiotic crossover and chromosome mis-segregation. Equational segregation events have previously been observed in complete analyses of human female meiosis (Ottolini et al., 2015), and we observe similar events here in the context of mouse male meiosis (i.e., equational segregation of one or several chromosomes). Among the 292 M2 cells we analyzed from the (B6 × Spret) cross, individual cells were biased toward equational or reductional chromosome segregation, suggesting a global sensing mechanism for deciding whether a cell proceeds with meiosis or returns to mitotic segregation of its chromosomes. Also, to our knowledge for the first time in mammalian meiosis, we observed multiple instances of whole-genome equational segregation during MI, suggesting a cell-autonomous rather than a chromosome-autonomous mode of equational segregation. We identified such events in both crosses, albeit more rarely in the fertile (B6 × Cast) cross.

The high incidence of whole-genome equational segregation, particularly in the interspecific (B6 × Spret) cross, raises more

questions than it answers. We depict the model and highlight several unresolved questions in Figure S7. In normal MI, centromere cohesion is maintained in reductional segregation and sister chromatids centromere proximal to the crossover do not split until MII (pattern 1 in Figure S7H). Equational segregation in MI indicates premature centromeric cohesin separation (pattern 2 and/or 3 in Figure S7H). Previous work has also shown that homolog pairing could be defective in these F1 crosses due to erosions of PRDM9 binding sites (Davies et al., 2016; Gregorova et al., 2018; Smagulova et al., 2016), and the pairing problem is probably more severe in the interspecific cross. In STAR Methods, “Speculations on the Causes and Consequences of Reverse Segregation,” we speculate on (1) what may cause premature centromeric cohesin separation, (2) whether one crossover is sufficient for proper reductional segregation, and (3) what consequences equational segregation in MI may have.

One key difference from simply combining the high-throughput single-cell combinatorial indexing (sci) scheme with linear amplification via transposon insertion (LIANTI) in the development of sci-L3 is that we introduced the T7 promoter by ligation, which not only enables more than two rounds of cell barcoding and further increased throughput at a much-reduced cost but also provides the flexibility to generalize the method to other single-cell assays with small tweaks of the protocol. As a first example, we demonstrate that sci-L3-WGS can be easily adapted to sci-L3-target-seq. Although single-cell targeted sequencing has been reported with the 10X Genomics platform, to our knowledge it is of RNA transcripts, rather than of DNA loci. Although the current 10% recovery rate per haplotype may not be ideal for targeted sequencing, it is mitigated by the large number of cells that can be analyzed. As a second example, we demonstrate that sci-L3-WGS can also be adapted to a sci-L3-RNA/DNA co-assay. We anticipate that it may be further possible to adapt sci-L3 to assay for transposase-accessible chromatin using sequencing (ATAC-seq), bisulfite-seq, and Hi-C for single-cell profiling of chromatin accessibility, the methylome, and chromatin conformation, respectively, which may have advantages over published sci-methods (Cusanovich et al., 2015; Mulqueen et al., 2018; Ramani et al., 2017) for these goals in terms of throughput and amplification uniformity.

Limitations

Sci-L3 has limitations, including genome coverage projected at 20% due to imperfect *in situ* nucleosome depletion, Tn5 insertion density, and ligation efficiency. In addition, the cost of WGS of large numbers of single cells is still prohibitive. Finally, while the scheme is largely generalizable to other single-cell assays and organisms, different assays and cell types

(D) Comparing 1C and M2 cells, (B6 × Spret) cross. After accounting for inter-chromosome variability, we estimate that crossovers in M2 cells (right) are on average 9.4 Mb more centromere proximal than in 1Cs (left) in the (B6 × Spret) cross. The same trend is observed to a lesser extent in the (B6 × Cast) cross (see Figure S7B).

(E) Area under the curve (AUC) of 0.73 quantifies expected accuracy in predicting if a region drawn from the mouse genome comes from (B6 × Spret) crossover tracts or an equal number of randomly sampled tracts. Left: all 76 features. Right: a subset of 25 features from BMA with MIP >0.5.

(F) AUC of 0.85 quantifies expected accuracy in predicting if a region drawn from the mouse genome comes from (B6 × Cast) crossover tracts or an equal number of randomly sampled tracts. Left: all 69 features. Right: a subset of 25 features from BMA with MIP >0.5.

may require additional optimization of the upstream nuclei preparation methods.

Conclusion

Sci-L3-WGS, sci-L3-target-seq, and the sci-L3-RNA/DNA co-assay substantially expand the toolset and potential throughput of single-cell sequencing. In this study, we furthermore show how sci-L3-WGS can provide a systematic and quantitative view of meiotic recombination and uncover rare whole-genome chromosome mis-segregation events. We anticipate that sci-L3 methods will be highly useful in other contexts in which single-cell genome sequencing is proving transformative (e.g., for studying rare inter-homolog mitotic crossovers, for dissecting the genetic heterogeneity and evolution of cancers).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Supplemental Results and Discussion
 - Sci-L3 Method
 - Setup of sci-L3-WGS experiment in two crosses
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Bioinformatic and statistical analyses
- DATA AND CODE AVAILABILITY
- ADDITIONAL RESOURCES
 - Detailed Protocol

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.molcel.2019.08.002>.

ACKNOWLEDGMENTS

The raw data are deposited with the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/PRJNA511715>). We thank G. Bonora, N. Kleckner, and members of the Shendure lab for helpful discussions. We thank C. Chen, D. Xing, J. Cao, and M. Spielmann for helpful technical suggestions; A. Leith for exceptional assistance in flow sorting; and T. Reh's lab for sharing the NIH/3T3 cell line. This work was funded by grants from the NIH (DP1HG007811 and R01HG006283 to J.S.; R35GM124704 to A.C.A.; DK107979 to J.S., W.S.N., and C.M.D.; and GM046883 to C.M.D.), NIDDK Intramural Research Program to R.D.C.-O., and the Paul G. Allen Frontiers Group (Allen Discovery Center grant to J.S.). Y.Y. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2248-16). J.S. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

Y.Y. developed techniques and performed the experiments; Y.Y. and Y.J. performed computational analyses; K.-W.G.L. and R.D.C.-O. provided (B6 × Cast) F1 mouse nuclei and helpful discussions; J.B.B. and C.M.D. provided (B6 × Spret) F1 mice; W.S.N. provided advice on the analyses; Y.Y. and J.S. wrote the paper with input from A.C.A. and all of the authors; and J.S. supervised the work.

DECLARATION OF INTERESTS

F.J.S. declares competing financial interests in the form of stock ownership and paid employment by Illumina. One or more embodiments of one or more patents and patent applications by the University of Washington and Illumina may encompass the methods, reagents, and data disclosed in this manuscript.

Received: September 7, 2018

Revised: May 22, 2019

Accepted: August 1, 2019

Published: September 5, 2019

REFERENCES

- Barral, S., Morozumi, Y., Tanaka, H., Montellier, E., Govin, J., de Dieuleveult, M., Charbonnier, G., Couté, Y., Puthier, D., Buchou, T., et al. (2017). Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Mol. Cell* 66, 89–101.e8.
- Baudat, F., Manova, K., Yuen, J.P., Jasin, M., and Keeney, S. (2000). Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol. Cell* 6, 989–998.
- Baudat, F., Imai, Y., and de Massy, B. (2013). Meiotic recombination in mammals: localization and regulation. *Nat. Rev. Genet.* 14, 794–806.
- Berletch, J.B., Ma, W., Yang, F., Shendure, J., Noble, W.S., Disteche, C.M., and Deng, X. (2015). Escape from X inactivation varies in mouse tissues. *PLoS Genet.* 11, e1005079.
- Bonora, G., Deng, X., Fang, H., Ramani, V., Qiu, R., Berletch, J.B., Filippova, G.N., Duan, Z., Shendure, J., Noble, W.S., and Disteche, C.M. (2018). Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. *Nat. Commun.* 9, 1445.
- Brick, K., Pratto, F., Sun, C.-Y., Camerini-Otero, R.D., and Petukhova, G. (2018). Analysis of Meiotic Double-Strand Break Initiation in Mammals. *Methods Enzymol.* 607, 391–418.
- Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.-R., Pekowska, A., Zhang, H., Rao, S.S.P., et al. (2017). Genome Organization Drives Chromosome Fragility. *Cell* 170, 507–521.e18.
- Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357, 661–667.
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566, 496–502.
- Chen, C., Xing, D., Tan, L., Li, H., Zhou, G., Huang, L., and Xie, X.S. (2017). Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* 356, 189–194.
- Choi, K., and Henderson, I.R. (2015). Meiotic recombination hotspots - a comparative view. *Plant J.* 83, 52–61.
- Clyde, M.A., Ghosh, J., and Littman, M.L. (2011). Bayesian Adaptive Sampling for Variable Selection and Model Averaging. *J. Comput. Graph. Stat.* 20, 80–101.
- Cole, F., Baudat, F., Grey, C., Keeney, S., de Massy, B., and Jasin, M. (2014). Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. *Nat. Genet.* 46, 1072–1080.
- Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348, 910–914.
- Davies, B., Hatton, E., Altemose, N., Hussin, J.G., Pratto, F., Zhang, G., Hinch, A.G., Moralli, D., Biggs, D., Diaz, R., et al. (2016). Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* 530, 171–176.

- Dayani, Y., Simchen, G., and Lichten, M. (2011). Meiotic recombination intermediates are resolved with minimal crossover formation during return-to-growth, an analogue of the mitotic cell cycle. *PLoS Genet.* 7, e1002083.
- de Boer, E., Jasin, M., and Keeney, S. (2015). Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hot spots in mice. *Genes Dev.* 29, 1721–1733.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA* 89, 3010–3014.
- Froenicke, L., Anderson, L.K., Wienberg, J., and Ashley, T. (2002). Male mouse recombination maps for each autosome identified by chromosome painting. *Am. J. Hum. Genet.* 71, 1353–1368.
- Gan, H., Wen, L., Liao, S., Lin, X., Ma, T., Liu, J., Song, C.-X., Wang, M., He, C., Han, C., and Tang, F. (2013). Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat. Commun.* 4, 1995.
- Goudarzi, A., Zhang, D., Huang, H., Barral, S., Kwon, O.K., Qi, S., Tang, Z., Buchou, T., Vitte, A.-L., He, T., et al. (2016). Dynamic Competing Histone H4 K5K8 Acetylation and Butyrylation Are Hallmarks of Highly Active Gene Promoters. *Mol. Cell* 62, 169–180.
- Gregorova, S., Gergelits, V., Chvatalova, I., Bhattacharyya, T., Valiskova, B., Fotopulosova, V., Jansa, P., Wiatrowska, D., and Forejt, J. (2018). Modulation of *Prdm9*-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. *eLife* 7, e34282.
- Hasegawa, K., Sin, H.-S., Maezawa, S., Broering, T.J., Kartashov, A.V., Alavattam, K.G., Ichijima, Y., Zhang, F., Bacon, W.C., Greis, K.D., et al. (2015). SCML2 establishes the male germline epigenome through regulation of histone H2A ubiquitination. *Dev. Cell* 32, 574–588.
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673.
- Hong, S., Sung, Y., Yu, M., Lee, M., Kleckner, N., and Kim, K.P. (2013). The logic and mechanism of homologous recombination partner choice. *Mol. Cell* 51, 440–453.
- Hou, Y., Fan, W., Yan, L., Li, R., Lian, Y., Huang, J., Li, J., Xu, L., Tang, F., Xie, X.S., and Qiao, J. (2013). Genome analyses of single human oocytes. *Cell* 155, 1492–1506.
- Inoue, K., Ichianagi, K., Fukuda, K., Glinka, M., and Sasaki, H. (2017). Switching of dominant retrotransposon silencing strategies from posttranscriptional to transcriptional mechanisms during male germ-cell development in mice. *PLoS Genet.* 13, e1006926.
- Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88, 375–384.
- Lange, J., Yamada, S., Tischfield, S.E., Pan, J., Kim, S., Zhu, X., Socci, N.D., Jasin, M., and Keeney, S. (2016). The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. *Cell* 167, 695–708.e16.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Lilue, J., Doran, A.G., Fiddes, I.T., Abrudan, M., Armstrong, J., Bennett, R., Chow, W., Collins, J., Czechanski, A., Danecek, P., et al. (2018). Multiple laboratory mouse reference genomes define strain specific haplotypes and novel functional loci. *bioRxiv*. <https://doi.org/10.1101/235838>.
- Liu, E.Y., Morgan, A.P., Chesler, E.J., Wang, W., Churchill, G.A., and Pardo-Manuel de Villena, F. (2014). High-resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male germline. *Genetics* 197, 91–106.
- Lu, S., Zong, C., Fan, W., Yang, M., Li, J., Chapman, A.R., Zhu, P., Hu, X., Xu, L., Yan, L., et al. (2012). Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* 338, 1627–1630.
- Maezawa, S., Yukawa, M., Alavattam, K.G., Barski, A., and Namekawa, S.H. (2018). Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res.* 46, 593–608.
- 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.
- Marchal, C., Sasaki, T., Vera, D., Wilson, K., Sima, J., Rivera-Mulia, J.C., Trevilla-García, C., Nogues, C., Nafie, E., and Gilbert, D.M. (2018). Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nat. Protoc.* 13, 819–839.
- Mu, W., Starmer, J., Shibata, Y., Yee, D., and Magnuson, T. (2017). EZH1 in germ cells safeguards the function of PRC2 during spermatogenesis. *Dev. Biol.* 424, 198–207.
- Mulqueen, R.M., Pokholok, D., Norberg, S.J., Torkency, K.A., Fields, A.J., Sun, D., Sinnamon, J.R., Shendure, J., Trapnell, C., O’Roak, B.J., et al. (2018). Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* 36, 428–431.
- Ottolini, C.S., Newnham, L., Capalbo, A., Natesan, S.A., Joshi, H.A., Cimadomo, D., Griffin, D.K., Sage, K., Summers, M.C., Thornhill, A.R., et al. (2015). Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. *Nat. Genet.* 47, 727–735.
- Pellegrino, M., Sciambi, A., Treusch, S., Durruthy-Durruthy, R., Gokhale, K., Jacob, J., Chen, T.X., Geis, J.A., Oldham, W., Matthews, J., et al. (2018). High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res.* 28, 1345–1352.
- Petes, T.D. (2001). Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* 2, 360–369.
- Petes, T.D., and Botstein, D. (1977). Simple Mendelian inheritance of the reiterated ribosomal DNA of yeast. *Proc. Natl. Acad. Sci. USA* 74, 5091–5095.
- Petes, T.D., and Merker, J.D. (2002). Context dependence of meiotic recombination hotspots in yeast: the relationship between recombination activity of a reporter construct and base composition. *Genetics* 162, 2049–2052.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ramani, V., Deng, X., Qiu, R., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan, Z., and Shendure, J. (2017). Massively multiplex single-cell Hi-C. *Nat. Methods* 14, 263–266.
- Romanienko, P.J., and Camerini-Otero, R.D. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol. Cell* 6, 975–987.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87.
- Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., van de Nobelen, S., Caesar, L., et al. (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics Chromatin* 5, 8.
- Smagulova, F., Gregoretti, I.V., Brick, K., Khil, P., Camerini-Otero, R.D., and Petukhova, G.V. (2011). Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* 472, 375–378.
- Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R.D., and Petukhova, G.V. (2016). The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes Dev.* 30, 266–280.
- Sos, B.C., Fung, H.-L., Gao, D.R., Osothprarop, T.F., Kia, A., He, M.M., and Zhang, K. (2016). Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. *Genome Biol.* 17, 20.
- Storzazzi, A., Tesse, S., Ruprich-Robert, G., Gargano, S., Pöggeler, S., Kleckner, N., and Zickler, D. (2008). Coupling meiotic chromosome axis integrity to recombination. *Genes Dev.* 22, 796–809.
- Vitak, S.A., Torkency, K.A., Rosenkrantz, J.L., Fields, A.J., Christiansen, L., Wong, M.H., Carbone, L., Steemers, F.J., and Adey, A. (2017). Sequencing

- thousands of single-cell genomes with combinatorial indexing. *Nat. Methods* **14**, 302–308.
- Wang, J., Fan, H.C., Behr, B., and Quake, S.R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150**, 402–412.
- Wang, S., Kleckner, N., and Zhang, L. (2017a). Crossover maturation inefficiency and aneuploidy in human female meiosis. *Cell Cycle* **16**, 1017–1019.
- Wang, S., Hassold, T., Hunt, P., White, M.A., Zickler, D., Kleckner, N., and Zhang, L. (2017b). Inefficient Crossover Maturation Underlies Elevated Aneuploidy in Human Female Meiosis. *Cell* **168**, 977–989.e17.
- Yamada, S., Kim, S., Tischfield, S.E., Jasin, M., Lange, J., and Keeney, S. (2017). Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice. *Cell Cycle* **16**, 1870–1884.
- Zhang, T., Murphy, M.W., Gearhart, M.D., Bardwell, V.J., and Zarkower, D. (2014). The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis. *Development* **141**, 3662–3671.
- Zhang, K., Wu, X.-C., Zheng, D.-Q., and Petes, T.D. (2017). Effects of Temperature on the Meiotic Recombination Landscape of the Yeast *Saccharomyces cerevisiae*. *MBio* **8**, e02099-17.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and analyzed data	This paper	SRA: PRJNA511715
Experimental Models: Cell Lines		
Patski	ENCODE	Christine Disteche lab
HEK293T	ATCC	CRL-3216
3T3	ATCC	CRL-1658
Experimental Models: Organisms/Strains		
(B6 × Spretus) F1 male	Charles River	Christine Disteche lab
(B6 × Cast) F1 male	The Jackson Laboratory	R. Daniel Camerini-Otero lab
Oligonucleotides		
Barcoding oligos for sci-L3, see Table S5	This paper	N/A
Software and Algorithms		
Samtools	Li and Durbin, 2009	http://samtools.sourceforge.net/
BWA	Li and Durbin, 2009	http://bio-bwa.sourceforge.net
Other		
Sci-L3 Methods	This paper	STAR Methods
R package for sci-L3 analyses	This paper	https://github.com/Yue-Jiang/scilantifig

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information should be directed to and will be fulfilled by the Lead Contact, Jay Shendure (shendure@uw.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice were euthanized according to University of Washington approved IACUC protocols (Christine Disteche lab) using CO₂ gas for at least 5 min of exposure, followed by a second method of euthanasia such as cervical dislocation in accordance with the Guiding Principles for the Care and Use of Laboratory Animals.

METHOD DETAILS

Supplemental Results and Discussion

Finite mixture model for fitting the three populations of non-1C cells

The non-1C cells recovered from (B6 × Cast) hybrid from barcode group 2 include 1C doublets, cells that appear biased toward equational segregation, and cells that appear biased toward reductional segregation. To quantify their relative proportions, we fit the data to a mixture of three binomial distributions, with probabilities of chromosomes segregating equationally of 0.01, 0.48 and 0.95, and mixing proportions of 0.28, 0.69 and 0.02 ([Figure S3H](#)). In contrast, when we attempt to similarly fit non-1C cells from barcode group 1 to a mixture of three binomial distributions, we obtain probabilities of chromosomes segregating equationally of 0.46, 0.5 and 0.53 (all close to 0.5), and mixing proportions of 0.24, 0.44 and 0.31 ([Figure S3I](#)).

Toward asking whether the proportion of M2 cells that are biased toward equational versus reductional segregation differs between the fertile and infertile crosses, we can similarly fit the chromosomal data from the (B6 × Spret) cross ([Figure 4E](#)), which yields probabilities of chromosomes segregating equationally of 0.05, 0.39 and 0.91, and mixing proportions of 0.66, 0.2 and 0.14 ([Figure S3J](#)). These proportions suggest that the infertile (B6 × Spret) cross has higher proportion of cells that are biased toward equational rather than reductional segregation.

Distribution of meiotic crossovers at the chromosomal level

Basing on 1,663 1C cells harboring 19,601 crossover breakpoints and 240 M2 cells with 4,184 crossover breakpoints from the (B6 × Spret) cross, and 5,547 1C cells harboring 60,755 crossover breakpoints and 115 M2 cells with 2,246 crossover breakpoints from the

(B6 × Cast) cross, we first considered the distribution of meiotic crossovers across chromosomes. Crossover density is defined here as the average number of crossovers per cell per division per Mb multiplied by 2 (in 1C cells) or 1 (in M2 cells). In the (B6 × Spret) cross, we observed a strong negative correlation between chromosome size and crossover density in 1C cells (Figure S4A, $r = -0.66$, $p = 0.002$). Consistent with previous findings (Lange et al., 2016), this correlation is only partly explained by Spo11 oligonucleotide complex density ($r = -0.46$, $p < 0.05$), suggesting that smaller chromosomes sustain more DSBs and those DSBs are more likely to give rise to crossovers. This negative correlation is even stronger in M2 cells (Figure S4B, $r = -0.83$, $p = 1e-5$). These observations suggest that smaller chromosomes are hotter for crossovers. The same trend is observed in the (B6 × Cast) cross. 1C cells had an average of 0.62 and 0.58 crossovers per chromosome per cell for inter- and intra-specific crosses, respectively, while M2 cells had an average of 0.92 and 1.03 per chromosome per cell (Figures S4C–S4F). The crossover rate in interspecific M2 cells is only 9% lower than crossover counts measured by Mlh1 foci in 4C spermatocytes in B6 inbred mice (Froenicke et al., 2002), despite a sequence divergence of 2%. The crossover rate in 1C cells is 45% lower than observed in single human sperm sequencing (Lu et al., 2012; Wang et al., 2012). The latter difference could largely be due to the telocentric nature of mouse chromosomes. Although the interspecific (B6 × Spret) cross has higher average number of crossovers detected in 1Cs compared to the (B6 × Cast) cross ($p = 7e-26$, Mann-Whitney test), the average number of crossovers in M2 cells are lower ($p = 2e-10$). We note that the proportion of M2 cells that segregated all 19 autosomes reductionally that have a crossover on every chromosome is higher for the (B6 × Cast) cross (60/91 of 66%) than the (B6 × Spret) cross (41/80 or 51%) ($p = 0.06$, Fisher's exact test), such that it could potentially contribute to the infertility of the latter.

To examine crossover interference, we took chromosomes with at least two crossovers and plotted the distance between adjacent crossovers, and compared this distribution to expectation based on random simulation (Figure S4G). The median observed distance between crossovers was 82 Mb for (B6 × Spret) and 97 Mb for (B6 × Cast); both are much larger than the expectation of 39 and 42 Mb ($p = 1e-267$ and $p < 2e-308$, respectively, Mann-Whitney test). This is consistent with the repulsion of crossovers in close proximity. Note that crossover interference is stronger in the (B6 × Cast) than the (B6 × Spret) cross, with longer distances between adjacent crossovers ($p = 5e-91$).

We also analyzed the distribution of uniparental chromosomes (i.e., no observed crossovers) in each single cell (Figure S4H; Table S4) and for each chromosome (Figure S4I) in (B6 × Spret) cross (the same trends hold for the (B6 × Cast) cross). Although shorter chromosomes exhibit elevated crossover rates when normalized by length, the rate of uniparental chromosomes (collapsed across all classes of cells) still negatively correlated with chromosome size (Figure S4I; $r = -0.91$, $p = 4.6e-8$).

While we have shown that M2 cells are strongly biased toward either equational or reductional segregation of their chromosomes, we also observed hundreds of sporadic equational segregation events among cells that have at least 15 chromosomes with reductional segregation. This phenomenon has previously been observed and termed as “reverse segregation” (Ottolini et al., 2015). In Figure S4J, we show chromosome distribution of these reverse segregation events. Note that although the rate of reverse segregation is significantly higher in the (B6 × Spret) cross (mean = 1.1) than the (B6 × Cast) cross (mean = 0.2, $p = 2e-14$, Mann-Whitney test), chromosomes 7 and 11 have the highest rates of reverse segregation in both crosses.

We then examined the normalized proportion of reads per cell that map to the mitochondrial genome (Figure S4K). The 1C cells exhibit a bimodal distribution in terms of the “copy number” of mitochondria DNA, an observation for which we lack a satisfactory explanation. We observed a modest negative correlation between the mitochondrial read proportion and the number of crossovers ($\rho = -0.11$, $p = 3e-6$). Interestingly, although of limited number, M2 cells that segregated at least 15 of their chromosomes either equationally versus reductionally had very different distributions of mitochondrial read proportions. Consistent with this, the mitochondrial read proportion positively correlated with the number of reductionally segregated chromosomes in M2 cells ($r = 0.18$, $p = 0.005$). Note that we are not able to evaluate this in the (B6 × Cast) cross because more than 90% of the single cells sequenced do not have any reads mapping to the mitochondrial genome. It is possible that the different methods used for nuclei isolation from the testes (B6 × Cast) versus the epididymis (B6 × Spret), coupled with pre-sorting of the nuclei from the testes, fractionated the mitochondria away from the bulk nuclei.

Distribution of meiotic crossover events in relation to genomic features

Genomic Features Regulating Crossover Hotness. To evaluate the distribution of crossovers at a finer scale, we collapsed all crossover events to generate “hotness maps” along each murine chromosome. We first compared these maps with the single-stranded DNA sequencing (SSDS) map (Brick et al., 2018; Smagulova et al., 2011, 2016) and the Spo11 oligonucleotide-complex map (Lange et al., 2016), which identify meiotic DSB hotspots at the highest resolution. DSB maps in the B6 strain from these two mapping methods strongly correlate with each other along 100 kb windows ($\rho = 0.87$, $p < 2e-308$). Although our 1C and M2 cell crossover pileups correlate with one another ($\rho = 0.67$ for (B6 × Spret) cross and $\rho = 0.55$ for (B6 × Cast) cross, $p < 2e-308$ for both), both deviate from the DSB maps. Of relevance, the PRDM9 gene, a major player for hotspot specification, has evolved to bind different motifs between diverged mouse strains, even between subspecies of mice (Davies et al., 2016; Gregorova et al., 2018). We found that in the intraspecific (B6 × Cast) cross, crossover hotness correlates better with DSB hot domains mapped in the Cast male than the B6 male ($\rho = 0.28$ and 0.12 , $p < 2e-308$ and $p = 1e-83$, respectively), possibly as a result of Cast PRDM9 allele being semi-dominant in the F1 hybrid. The correlation is stronger with DSB hot domains mapped in (B6 × Cast) F1 animals ($\rho = 0.3$, $p < 2e-308$). For the (B6 × Spret) cross, the erosion of PRDM9 consensus binding site results in four types of DSB hotspots defined by the Spo11 oligonucleotide-complex map: those that are conserved between B6 and Spret, termed as “symmetric” hotspots, those that are only present in B6 or Spret, termed as “asymmetric” hotspots, and those do not contain a PRDM9 binding site in either species. All four types

of DSB hot domains correlate poorly with crossovers from the (B6 × Spret) cross ($\rho = 0.13$, $p = 4e-87$ for using all Spo11 hotspots mapped in B6; $\rho = 0.11$, $p = 3e-63$ if we only use “symmetric hotspots”). One possibility is that the DSB sites in the (B6 × Spret) cross are strongly dominated by the Spret PRDM9 allele, such that the DSB hotspots mapped in the B6 strain background do not predict sites of crossovers.

Only 10% of meiotic-specific DSBs are repaired as crossovers. We next looked at what factors beyond Spo11 breaks contribute to crossover formation by building a linear model with Bayesian Model Averaging (BMA) (Clyde et al., 2011). As applied here, BMA takes a weighted average of the more than 15,000 variable selection models explored and weights them by the posterior probability of each model, which accounts for uncertainty in model selection, unlike some other variable selection techniques like Lasso regression. We quantified a marginal inclusion probability (MIP) for ~80 potentially explanatory variables. Features that are known to be relevant to meiotic crossovers such as Spo11 break sites, GC content, etc. are included in almost all the models with high probabilities (Figures 6A and S5); for example, regions with high GC content are hotter for crossover formation. We also found a few more features that have not previously been implicated in meiotic crossovers, such as specific families of repeats and chromatin marks, and particularly early replication domains. Correlation matrices between crossover hotness and all the features are plotted in Figure S6 for each cross. Features used and summaries of the simple linear models and BMA are included in Table S7. The breakpoint resolution (median ~150 kb for (B6 × Spret) and ~250 kb for (B6 × Cast); Figure 6B) is on par with previous efforts to map meiotic crossovers by single cell sequencing (150 - 500 kb) (Lu et al., 2012; Ottolini et al., 2015; Wang et al., 2012); however, the greater library complexity afforded by sci-L3-WGS enabled us to achieve this with a much lower sequencing depth.

Many of the features that correlate with crossover formation are consistent between the (B6 × Spret) and (B6 × Cast) crosses, but some are not. For example, the positional biases of crossover formation appear to be different. In 1C cells of both crosses, as well as in M2 cells in the (B6 × Cast) cross, crossovers are underrepresented within 10 Mb from the centromere and rather tend to occur near the telomere in the rightmost positional ‘quartile’. However, in M2 cells in the (B6 × Spret) cross, crossovers are underrepresented near the centromere as well as near the telomere, and rather tend to occur in the middle quartiles (Figure S6). This trend holds in the linear models where we account for contributions from all other features.

The position of a crossover can greatly affect the amount of tension enforced between chromosome homologs, which in turn facilitates proper chromosome segregation. We therefore explored this in more detail by taking only the rightmost crossover for each chromosome in each cell and examining its position along the chromosome arm in each cross (de Boer et al., 2015). Accounting for inter-chromosome variability with a linear mixed effect model, we estimate that the positions of the rightmost crossovers in the (B6 × Spret) cross are on average 1.6 Mb more centromere-proximal than those in the (B6 × Cast) cross in 1C cells (Figure S7A, $p = 1e-13$, F test), but are 5.5 Mb more centromere-proximal in the M2 cells (Figure 6C, $p = 2.2e-15$). Note that the rightmost crossovers in the M2 cells tend to be more centromere-proximal than those in the 1C cells in both crosses, but to a greater extent in the (B6 × Spret) cross (Figure 6D) than in the (B6 × Cast) cross (Figure S7B). These differences suggest that a subset of M2 cells in the (B6 × Spret) cross whose crossovers occur too close to the centromere may fail to mature into 1C cells, possibly due to defects in MI segregation. Similarly, although of limited number of events, we have also compared the positions of crossovers in M2 cells that have biased chromosome segregation and found that in both crosses, crossovers in cells with biased equational segregation are more centromere-distal than those in cells with biased reductional segregation, with differences of 13.7 Mb in the (B6 × Cast) cross ($p = 4e-15$) and of 8.7 Mb in the (B6 × Spret) cross ($p = 6e-14$) (Figures S7C and S7D). This suggests possible MI segregation defects in cells that have crossovers too close to the telomere. We propose a tentative model to explain this observation in Figure S7E.

Cell Heterogeneity in Terms of Crossover Break Points. Although 1C and M2 cells appear broadly similar in the crossover pileups, we wondered whether there was any structure to the features that influence crossover distributions in subsets of single cells. To explore this, we aggregated crossover-related information for each single cell for each of 78 features (See also “Bioinformatic and Statistical Analyses” section below). We then used principal component analysis (PCA) on a matrix with each row as one single cell and each column as one summarized feature value. For the (B6 × Spret) cross, the first two principal components (PCs) capture 26% of the variance, and for the (B6 × Cast) cross, PC1 and PC3 capture 17% of the variance. In both crosses, the 1C and M2 cells are separated into two clusters by these PCs (Figures S7F and S7G). The chromosomal distribution of crossovers, uniparental chromosomes and positions of crossovers in chromosome quartiles are the features that appear to drive the separation of 1C and M2 cells.

Predicting Crossover Tracts from Genomic Features. Finally, we sought to exploit the large number of events observed here to construct a predictive model of crossover locations. Specifically, we built a linear model of binary response with 1 being crossover tracts and 0 being a random tract sampled from the genome from the same tract length distribution (details in “Bioinformatic and Statistical Analyses” section below). Using the same 76 features as in the BMA analyses, we can predict crossover tracts on held-out data with an average Receiver Operator Curve (ROC) Area Under Curve (AUC) of 0.73 for (B6 × Spret) cross. With a subset of 25 variables of high inclusion probability (MIP > 0.5) identified by BMA, we achieve a similar average AUC of 0.72 (Figure 6E). Similarly, for the (B6 × Cast) cross, we achieve an average AUC of 0.85 when all features or a subset of 25 features with MIP > 0.5 are used (Figure 6F).

In sum, the improved genome coverage enabled high-resolution mapping of crossover break points compared to other single-cell sequencing methods, and the throughput for mapping a total of ~87,000 crossovers allowed us to better characterize genomic and epigenomic features associated with crossover hotness with pileup data (further discussion in “Crossover Hotness and Associated (Epi)genomic Factors” section below).

Speculations on the causes and consequences of reverse segregation

We have observed high incidence of reverse segregation, particularly in the interspecific (B6 × Spret) cross. Below we speculate on: 1) what might cause premature centromeric cohesin separation, 2) whether one crossover is sufficient for proper reductional segregation, and 3) what consequences equational segregation in MI may have.

First, it is possible that due to insufficient homolog pairing between B6 and Spret chromosomes, DSBs that should have been normally repaired off the homolog during meiosis are instead frequently repaired using sister chromatids as template. This could cause disruption of cohesins (Storlazzi et al., 2008) and lead to premature centromere cohesin separation.

Second, the current model suggests that one inter-homolog crossover and proper sister chromatid cohesion are sufficient for forming chiasmata (Figure S7H) despite initial insufficient homolog pairing in the interspecific cross. Once a crossover is successfully formed, chromosome segregation should not be impaired. In our study, on the individual chromosome level, the large numbers of equationally segregated chromosomes observed do have normal crossovers as evidenced by centromere-distal LOH, which could indicate that defects in the initial homolog pairing impact the ultimate outcome. On the genome level, however, we cannot confidently assess whether those cells with biased equational segregation have similar numbers of crossovers as their reductionally biased counterparts, because we can detect all crossovers for chromosomes that segregate reductionally, but we can only detect crossovers in equationally segregated chromosomes when the two recombinant chromatids segregate apart (Figures S2B, S2C, and S7H, patterns 2 and 3). Assuming recombinant chromatids are equally likely to segregate together or apart, the number of crossovers is not smaller in those genome-level equational segregation cases, although we cannot exclude the possibility that segregation is biased away from 50/50 due to unresolved recombination intermediates (Figure S7H, pattern 3).

Third, what are the consequences of these equationally segregated chromosomes? Do they return to mitosis, bearing extensive LOH, or do they proceed to MII, and if so, contributing to forming 1C gametes? In yeast, a phenomenon called “return-to-growth” has been characterized wherein cells that initiate the meiosis program can revert to normal mitotic divisions in the presence of proper nutrients, resulting in large numbers of LOH events (Dayani et al., 2011). In human female meiosis, chromosomes with reverse segregation proceed to MII, leading to one euploid oocyte and one euploid polar body 2, consistent with normal MII segregation; the authors suggest that unresolved recombination intermediates may have both caused the reverse segregation in MI and facilitated proper MII segregation by linking the otherwise unrelated homolog chromatids (Figure S7H, pattern 3) (Ottolini et al., 2015). Mlh1 is important in both mismatch repair (MMR) and for resolving Holliday junction intermediates in meiosis. Given the 2% sequence divergence between B6 and Spret, it is possible that Mlh1 is limiting due to intensive MMR and there may not be enough Mlh1 for resolving recombination intermediates. However, we emphasize that if recombinant homolog chromatids co-segregate, this would not lead to LOH (Figure S2C). Therefore, M2 cells with LOH and equational segregation cannot be explained by co-segregation of unresolved intermediates.

Lastly, in Figure S7H, we also show possible contributions to forming gametes from chromosomes without any inter-homolog crossover, probably due to insufficient homolog pairing, because one of the patterns (pattern 4) is not distinguishable from cells that have a crossover but co-segregate recombinant chromatids (pattern 3). However, if these cells without crossover contribute significantly to the 1C cells, we should observe a higher number of crossover-free chromosomes among the 1C cells. Of the 1C cells we observed in both crosses, the number of chromosomes with and without crossovers is roughly 50-50, indicating that they predominantly derive from some combination of patterns 1-3 in Figure S7H, and 2C cells without inter-homolog crossovers (patterns 4 and 5) do not substantially contribute to 1C cells that successfully complete MII.

Crossover hotness and associated (epi)genomic factors

Crossover hotness is a continuum and shaped by many factors. Crossovers in the (B6 × Cast) cross correlate more strongly with meiotic DSB hotspots mapped in the F1 cross than in individual maps for the two parental strains, which is expected based on the previous finding that novel meiotic hotspots can form in F1 hybrids (Smagulova et al., 2016). In the (B6 × Spret) cross, crossovers are weakly but positively correlated with Spo11 breaks. Note that the Spo11 map only accounts for the PRDM9 sites bound by PRDM9 protein of the B6 allele, and it is likely that the Spret copy of PRDM9 binds different sites and creates new meiotic DSB hotspots, not accounted for in our analyses. Genomic features that we observe to be positively correlated with meiotic crossovers include GC-rich regions (also the case in yeast meiosis (Petes, 2001; Petes and Merker, 2002)), CNV gains between the strains (Lilue et al., 2018), gene bodies, pseudogenic transcripts, CTCF binding sites, replication domains (Marchal et al., 2018), DNA transposons, satellite DNA and a subset of histone modifications including H3K4me1, H3K27me3 and H3K36me3 (Mu et al., 2017). Intriguingly, the binding sites of Dmrt6, involved in regulating the switch from mitotic to meiotic divisions in male germ cells (Zhang et al., 2014) are strongly correlated with meiotic crossover hotness. Genomic features that are notably negatively correlated with meiotic crossovers include 3' UTRs, LINEs, and low complexity DNA. Unlike in yeast, where rDNA is extremely cold for meiotic crossovers (Petes and Botstein, 1977), mouse rDNA does not appear to suppress crossovers. With these genomic features, we are able to distinguish real meiotic crossover initiation sites from randomly sampled tracts in the mouse genome, with 0.73 and 0.85 accuracy in (B6 × Spret) and (B6 × Cast), respectively, and the 0.85 prediction accuracy in the (B6 × Cast) cross holds with a subset of 25 genome features. We emphasize that although the various features behave largely consistently between modeling approaches, we cannot assign any causality without further experiments.

Sci-L3 Method

Methods and molecular design of sci-L3-WGS and sci-L3-target-seq

Single Cell Preparation and Nucleosome Depletion. Cell suspension is prepared by trypsinizing from a Petri dish or homogenizing from tissues. Male F1 mice were euthanized by CO₂ followed by cervical dislocation according to University of Washington IACUC approved protocols. For isolation of male germ cells, we dissected the epididymis by slicing the tubes within and incubating the tissue in 1 ml of 1xPBS supplemented with 10% FBS at room temperature for 15 min. After incubation the cell suspension was collected by pipetting. Cells isolated from the epididymis were used for experiments of the (B6 × Spret) cross and also as a source of mature sperm (“barcode group 3”) in the (B6 × Cast) cross. For isolation of nuclei from whole testis as an enrichment method for 2C cells for the (B6 × Cast) cross, we first crosslinked testicular cells with 1% formaldehyde and extracted nuclei using hypotonic buffer. We then FACS-sorted 1C and 2C nuclei by DNA content primarily based on DAPI signal. Cultured human and mouse cells are pelleted at 550 g for 5 min at 4°C and male germ cells are pelleted at 2400 g for 10 min at 4°C.

Nucleosome depletion largely follows xSDS methods in sci-DNA-seq (Vitak et al., 2017) except that the lysis buffer is modified to be compatible with downstream LIANTI protocol (Chen et al., 2017). Cells are crosslinked in 10 mL DMEM complete media with 406 μL 37% formaldehyde (final conc. 1.5%) at r.t. for 10 min (gently inverting the tubes). We then add 800 μL 2.5 M Glycine and incubate on ice for 5 min. Cells are pelleted and washed with 1 mL lysis buffer (60 mM Tris-Ac pH 8.3, 2 mM EDTA pH 8.0, 15 mM DTT). The pellet is resuspended in 1 mL lysis buffer with 0.1% IGEPAL (I8896, SIGMA) and incubated on ice for 20 min. Nuclei are then pelleted, washed with 1xNEBuffer2.1, and resuspended in 800 μL 1xNEBuffer2.1 with 0.3% SDS for nucleosome depletion at 42°C (vigorous shaking for 30 min, 500 rpm). We then add 180 μL 10% Triton-X and vigorous shaking for 30 min at 42°C (500 rpm). Permeabilized nuclei are then washed in 1 mL lysis buffer twice and resuspended in lysis buffer at 20,000 nuclei per μL.

Transposome Design and Assembly. Transposon DNA oligo is synthesized with both 5' of the two strands phosphorylated, one required for Tn5 insertion (5'/Phos/CTGTCTCTTATACACATCT, IDT, PAGE purification) similar as in LIANTI and Nextera, the other required for ligation (5'/Phos/GTCTTG XXXXXXXX [1st round barcode] AGATGTGTATAAGAG

ACAG, IDT, standard desalting). After annealing 1:1 with gradual cooling (95°C 5 min, −0.1°C/cycle, 9 s/cycle, 700 cycles to 25°C) in annealing buffer (10mM Tris-HCl pH 8.0, 50mM NaCl, 1mM EDTA, pH 8.0), Tn5 duplex with 5' overhang is diluted to 1.5 μM. We then add 7.2 μL storage buffer (1xTE with 50% Glycerol) to 12 μL ~1 μM Tn5 transposase (Lucigen, TNP92110) and incubate 0.79 μL diluted transposase with 0.4 μL 1.5 μM Tn5 duplex at r.t. for 30 min. The transposome dimerize to a final concentration of 0.2 μM. The transposome complex can be stably stored at −20°C for up to one year. We set up 24 reactions for barcoding 24 wells in the first round but more wells could be desirable depending the application. For each new biological application, we first further dilute the transposome to 0.1 μM for a test experiment. The number of unique reads and library complexity is less optimal (Figure S1) but usable for mapping at low resolution.

In Figure 2, we show molecular structures of sci-L3-WGS at each step. In commercial Nextera library preparation, one loses at least half of the sequenceable DNA material due to: 1) Tn5 insertion introduces symmetric transposon sequence at the two ends of fragmented genomic DNA, which can result in formation of hairpin loop when denatured and prevent PCR amplification; and 2) if the two ends are tagmented with both i5 or i7 with 50% chance, the molecule cannot be sequenced. One key advantage of LIANTI over Nextera-based library preparation, is that the looped Tn5 design breaks the symmetry introduced by transposome dimer and facilitates reverse transcription (RT) by using an intramolecular RT primer, also characteristic of the looped transposon. However, looped transposon is not compatible with more than two rounds of barcoding, which limits throughput and significantly increase library cost (see Table S1 for comparison). In the changes we made for sci-L3-WGS, we maintain advantages brought by looped Tn5 during the ligation step.

Tagmentation (first-round barcodes) and ligation (second-round barcodes)

We then distribute 1.5 μL of nuclei at 20,000/μL concentration into each well in a lo-bind 96-well plate, add 6.5 μL H₂O and 0.7 μL 50 mM MgCl₂ (final conc. of 3.24 mM accounting for the EDTA in the lysis buffer). The 1.2 μL transposome prepared above is added into each well and the plate is then incubated at 55°C for 20 min (thermomixer is recommended but not required). We then add 5 μL of stop solution (40 mM EDTA and 1 mM spermidine) and pool nuclei in a trough. An additional 1 mL of lysis buffer is added to the nuclei suspension before pelleting. After carefully removing the supernatant, we resuspend the nuclei in 312 μL resuspension buffer (24 μL 10mM dNTP, 48 μL 10x tagmentation buffer [50 mM MgCl₂, 100 mM Tris-HCl pH 8.0], 96 μL H₂O, 144 μL lysis buffer), and distribute 4.7 μL nuclei mix into each well of a new lo-bind 96-well plate. Hairpin ligation duplex (1. CAAGAC 2. Y'Y'Y'Y'Y'Y'Y' [reverse complement of 2nd round barcode] 3. CAGGAGCGAGCTGCATCCC 4. AATTTAATACGACTCACTATA 5. GGGATGCAGCTCGCTCCTG 6. YYYYYY [2nd round barcode]) is pre-annealed similarly as the Tn5 transposon duplex and diluted to 1.5 μM. Note that the ligation duplex contains five elements: 1) reverse complement of ligation adaptor on Tn5; 2) reverse complement of 2nd round barcode; 3) reverse complement of second-strand synthesis (SSS) primer; 4) T7 promoter, note that this is the loop region of the hairpin; 5) second-strand synthesis (SSS) primer region starting with GGG for enhancing T7 transcription (“sp2” in Figure 2B); 6) 2nd round barcode (“bc2” in Figure 2B). We add 0.8 μL of these duplex to each of the 64 wells with nuclei suspension and add 1.18 μL ligation mix (0.6 μL 10x NEB T4 ligase buffer, 0.48 μL PEG-4000, 0.1 μL T4 DNA ligase [Thermo EL0011]) into each well and incubate at 20°C for 30 min. Note that after ligation, the looped structure mimics that of LIANTI and facilitates efficiency at the RT step (discussed below), and that both rounds of barcodes are present at the 3' of the T7 promoter and thus will be included in the amplified molecule. Ligation reaction is stopped by adding 4 μL stop solution. Cells are then pooled in a new trough (~630 μL), stained with DAPI at a final conc. of 5 μg/mL and sorted 100-300 into each new well with 3 μL lysis buffer added prior to cell sorting. Note that each sorting event with FACS is

associated with ~3-5 nL FACS buffer depending on the size of the nozzle, we recommend keeping the total volume of liquid added into each well < 1 μ L to keep the salt concentration low.

Cell lysis, gap extension and linear amplification by in vitro transcription

We then proceed with a total of 3.5-4 μ L sorted nuclei in each well for cell lysis by incubating at 75°C for 45 min, cooling to 4°C and treating with freshly diluted QIAGEN Protease (final conc. 2mg/mL) at 55°C for 8 hr. Protease is then heat-inactivated by incubating at 75°C for 30 min. Cell lysate can be stored at -80°C. We recommend processing no more than 32 wells of samples (~9600 single cells) for each experiment because subsequent amplification step involves RNA and is time-sensitive. For gap extension (Figure 2C), polymerase with strand displacement activity is used by adding a mixture of 2 μ L H₂O, 0.7 μ L 10x tagmentation buffer, 0.35 μ L 10mM dNTP and 0.35 μ L Bst WarmStart 2.0 polymerase with strand displacement activity, and incubate at 68°C for 5 min. Note that if ligation is successful on both ends, the duplex is symmetric with T7 promoter on both sides, but if ligation is only successful on one end, the region in the dashed box is missing on one side. Inter-molecular ligation is generally inefficient. Although we have included pre-annealed hairpin loop to minimize the necessity of inter-molecular ligation, two molecules (instead of three without the hairpin loop) still need to find each other. If the ligation efficiency is 50%, having ligation on both ends has 25% rate, but having ligation on either end has 75% rate. Later in the RT step, we show that successful ligation is required for only one end. After gap extension, a 20 μ L T7 *in vitro* transcription system is assembled by adding 2 μ L H₂O, 2 μ L T7 Pol mix and 10 μ L rNMP mix (NEB, HiScribe T7 Quick High Yield RNA Synthesis Kit). The mixture is incubated at 37°C for 10-16 hr.

RNA purification, RT and SSS (or targeted sequencing)

Transcription is terminated by adding 2.2 μ L 0.5M EDTA. Amplified RNA molecules are then purified with RCC-5 (Zymo Research, R1016) and eluted with 18 μ L 0.1x TE. A 30 μ L RT system is assembled by first adding 0.6 μ L RNA RT primer (rArGrArUrGrUrGrUrArUrArArGrArGrArCrArG, IDT), 2 μ L 10 mM dNTP and 0.5 μ L SUPERase[•] In RNase Inhibitor (20 U/ μ L, Thermo Fisher AM2696). We then incubate at 70°C for 1 min and 90°C for 20 s for denaturing and removing secondary structures and sudden cool on ice. SuperScript IV Reverse Transcriptase (SSIV, Thermo Fisher 18090050) is used for RT with 6 μ L 5x RT buffer, 1.5 μ L 0.1M DTT, 1 μ L SUPERase[•] In and 1 μ L SSIV. The RT reaction is incubated at 55°C for 15 min, 60°C for 10 min, 65°C for 12 min, 70°C for 8 min, 75°C for 5 min, and 80°C for 10 min. The reaction is cooled to r.t. before adding 0.5 μ L RNaseH (NEB) and 0.3 μ L RNaseA (Life Technologies, AM2270) and incubating at 37°C for 30 min. Note that Figure 2E depicts two scenarios during the RT step: 1) if both ends have successful ligation, RT is likely primed by fold-back loop as in LIANTI; 2) if only one end has successful ligation, RT is likely primed by the RNA RT primer added before the denaturing step. Excessive RNA primers and RNA transcripts are degraded after cDNA synthesis. Lastly, we synthesize the second strand with Q5 DNA polymerase by adding 27 μ L H₂O, 20 μ L 5x Q5 buffer, 20 μ L Q5 GC enhancer, 1 μ L Q5 polymerase and 1 μ L SSS primer (NNNN [UMI] ZZZZZ [3rd round barcode] GGGATG CAGCTCGCTCCTG, IDT, standard desalting). Resulting double stranded DNA can be purified with DCC-5 (Zymo Research, D4014) and proceed with library preparation kit such as NEBNext Ultra II with the minimal 3 cycles of PCR for adding the sequencing adaptor.

It is worth noting that the SSS step can be easily modified to enable targeted sequencing by using a single cell barcode primer with P5 end (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC

GCTCTCCGATCT NNNNNNN ZZZZZ [3rd round barcode] GGGATGCAGCTCGCTCCTG) together with a targeting primer for one region in the genome (Figure 1B). For example, in applications where one integrates lentivirus-based CRISPR library (Shalem et al., 2014), the guide RNA sequence in each single cell could be read off using P7 end with lentivirus-integrated CRISPR library primer, CAAGCAGAAGACGGCATACGAGAT TCGCCTTG [index 1] GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTCCGAC TCGGTGCCACTTTT TCAA), thus bypassing the need to sequence the whole genome and enrich for a specific region of interest. In this case, the library preparation step can be omitted and replaced by gel or bead purification to remove primer dimers.

Methods and molecular design of sci-L3-RNA/DNA co-assay

Single Cell Preparation and Nucleosome Depletion. Cell suspensions are prepared with the same protocol as in sci-L3-WGS other than differences indicated below. HEK293T, BJ-5ta and 3T3 cells were trypsinized from a Petri dish and fixed with 2% PFA in 1x PBS at room temperature for 10 min at 1M/mL cell concentration. Subsequent quenching (with Glycine), washing, nuclei isolation (with 0.1% IGEPAL), nucleosome depletion (xSDS method) steps are identical with sci-L3-WGS except that we add 1% Superase-In to all the lysis buffer and 1xNEBuffer2.1. Nuclei are resuspended in lysis buffer with 1% Superase-In at 20,000 nuclei per μ L.

Transposome and Reverse Transcription (RT) Primer Design. For the single cell genome amplification component, transposome design and assembly are identical to sci-L3-WGS.

For single cell transcriptome profiling component, reverse transcription primers share similar structure with sci-RNA-seq in (Cao et al., 2017; Cusanovich et al., 2015; Mulqueen et al., 2018; Ramani et al., 2017; Vitak et al., 2017) for the reverse transcription aspect, i.e., polyT priming part of the oligo, but contain a different barcode structure and landing pad for the subsequent ligation step (/5Phos/GTCTTG [same landing pad sequence as in sci-L3-WGS] NNNNNN [UMI1 for tagging unique transcripts] X'X'X'X'X'X'X' [1st round barcode for transcriptome, which are different sequences from Tn5 transposon barcodes] TTTTTTTTTTTTTTTTTTTT TTT TTTTTVN, IDT, standard desalting).

RT and Tagmentation (First-Round Barcodes), Ligation (Second-Round Barcodes), FACS, and Cell Lysis. We then distribute 1.5 μ L of nuclei at 20,000/ μ L concentration into each well in a lo-bind 96-well plate, add 0.2 μ L H₂O, 0.3 μ L 50 mM MgCl₂ (to neutralize EDTA in the lysis buffer), 0.25 μ L 10mM dNTP and 1 μ L 25 μ M RT primer described above to prepare for the RT step. The nuclei mixture is then incubated at 55°C for 5 min to remove secondary structures and quickly quench on ice. We then add 1 μ L 5x RT buffer, 0.03 μ L 100 mM DTT (note that there is DTT from lysis buffer, final conc. 5 mM), 0.25 μ L SSIV, 0.25 μ L RNaseOUT (Thermo Fisher Cat. No. 10777019), incubate

for RT reaction at 25°C 1 min, 37°C 1 min, 42°C 1 min, 50°C 1 min, 55°C 15 min. Then add 0.4 μ L MgCl₂ and 3.52 μ L H₂O and the 1.2 μ L transposome prepared above into each well. All subsequent steps until after cell lysis are identical to sci-L3-WGS.

Gap Extension and Linear Amplification by *In Vitro* Transcription. We use random heptamer for gap extension with partial NEBNext Read 1 primer as the 5' overhang (CACGACGCTCTCCGATCT NNNNNNN). We add 1 μ L of 20 μ M oligo, incubate at 95°C for 3 min to denature the DNA, and gradually cool to r.t. (~5 min) for the oligos to anneal. We then add 2 μ L H₂O, 0.8 μ L 10x NEBuffer2, 0.4 μ L 10mM dNTP, 0.4 μ L Klenow Fragment (3' → 5' exo-, NEB M0212S) and incubate at 30°C for 8 min and 75°C for 10 min. After gap extension, a 20 μ L T7 *in vitro* transcription system is assembled by the same sci-L3-WGS protocol.

RNA Purification, RT, and SSS. All the steps are identical to sci-L3-WGS except for different oligo sequences. At the RT step after IVT, instead using 0.6 μ L RNA RT primer, we use 0.6 μ L NEBNext Read 1 primer (AATGATACGGCGACCACCG AGATCTACA CTCTTCCCTACACGACGCTCTCCGATCT, P5 end of Illumina sequencing, IDT). For SSS primer, we use AAGCAGAAGACGGC ATACGAGAT [P7 end] NNNN [UMI2] Z'Z'Z'Z'Z'Z' [3rd round barcode] CGTCTCTAC GGGATGCAGCTCGCTCCTG to add the sequencing adaptor. Note that the resulting double stranded DNA now contains both the P5 and P7 end for Illumina sequencing and can be purified with 1.1x AmpureXP beads and proceed with sequencing. The library preparation step and the minimal 3 cycles of PCR in sci-L3-WGS for adding the sequencing adaptor are unnecessary for the co-assay.

Setup of sci-L3-WGS experiment in two crosses

(B6 × Spret) cross

We pooled cells isolated from 6 and 3 epididymides from (B6 × Spret) F1 males aged 70 days and 88 days, respectively, in two separate experiments, and fixed with 1% formaldehyde. For each experiment, after nucleosome depletion, we distributed 30,000 cells per well and performed *in situ* indexed Tn5 insertion across 24 wells to add the first-round barcodes. We then pooled all cells and redistributed these to 64 wells to add the second-round barcodes and T7 promoter by ligation. After again pooling all cells, we split the cell mixture 1:6, FACS-sorted the majority of cells (6/7), and diluted the rest (1/7). The resulting wells contained 100 to 360 cells per well with an estimated collision rate of 4%–11%.

(B6 × Cast) cross

From 6 testes, we recovered ~12M 1C round spermatids and ~0.5M 2C cells. However, due to the > 20-fold higher number of 1C cells, we still found many 1C cells in the population sorted for 2C cells (Figure S3F). In one of the sci-L3-WGS experiments where we tried to enrich for 2C cells, we estimate that we tagged ~160k sperm from the epididymis, ~160k 1C round spermatids and ~70k 2C cells, and further enriched for 2C cells during the FACS step of sci-L3-WGS (Figure S3G). However, despite two rounds of enrichment, 1C cells still dominated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic and statistical analyses

Read processing, alignment and SNV calling

Base calls were converted to fastq file by bcl2fastq with 1 mismatch allowed for errors in the index. We then used customized shell script “sci_lianti_v2.sh” for de-multiplexing (python scripts and the R Markdown file can be separately downloaded from the “inst” folder of the R package; the R package containing intermediate data files for generating all the main and supplemental figures can be downloaded and installed via the following link: <https://github.com/Yue-Jiang/sciliantifig>), which calls python scripts or NGS tools for the following steps: 1) order read pairs such that all single-cell combinatorial barcodes are in read 1 (R1); 2) de-multiplex 3rd round (SSS, 6nt, no error allowed) barcodes and attach both the barcodes and UMI for transcripts to the read names, and split library by 3rd round barcodes. Note that all subsequent steps are done in parallel for individual libraries split up by 3rd round barcodes, which contain 100-300 single cells; 3) using cutadapt to split 1st (Tn5, 8nt, 1 error allowed) and 2nd rounds (ligation, 7nt, 1 error allowed) of barcodes in R1, errors being calculated by Levenshtein distance, and attach both rounds of barcodes to the read names. This step is done in paired-end mode, i.e., if R1 does not have the correct barcode and spacer structure, the paired read 2 (R2) is discarded; 4) using cutadapt to clean up R2; 5) align in paired-end mode to hg19 or mm10 genome with bwa mem (Li and Durbin, 2009). For experiments where we assess barcode collision, we use concatenated reference of hg19 and mm10 and use uniquely aligned reads to determine relative mapping rate to human or mouse genomes; 6) split bam files into single cell bam files using 1st and 2nd rounds of barcodes attached in the read name; 7) convert bam file to bed files with bedtools (Quinlan and Hall, 2010), and determine unique insertion sites if either R1 or R2 shares the same end points. Unique Tn5 insertion site is defined as fragments where both ends of the read pair need to be different; 8) using the “pileup” function in the “lianti” package (<https://github.com/lh3/lianti/blob/master/pileup.c>) (Chen et al., 2017) to call variants in a allele-aware mode. Note that we include the combined bulk bam file (generated by samtools merge (Chen et al., 2017; Li and Durbin, 2009) of all the ~6900 single cells, more than 30x) with each single cell bam file at this step such that the threshold of depth at each SNP location only needs to be exceeded in the bulk file for a SNP call to be included in the final vcf. file, therefore raw counts of the REF and ALT alleles are included in the single cell column as long as the variant is present as a heterozygous SNP in the bulk file. This circumvents the problem of high false negative rate due to low-depth sequencing in single cells by converting the *de novo* SNP calling question to a genotyping question; 9) annotate SNV called in terms of SNP quality in each single cell by the reference SNP vcf. file for Spret (SPRET_EiJ.mgp.v5.snps.dbSNP142.vcf.gz downloaded from the Mouse Genome Project). The annotated SNP file is then used as input for subsequent crossover break point analyses.

HMM for calling breakpoints

The genotype at a given SNP site is determined by comparing the number of reads supporting reference and alternative alleles. For 1C cells, the crossover position is determined by fitting a hidden Markov model with three states: reference, alternative and heterozygous.

The transition matrix is specified as below:

From\To	reference	alternative	heterozygous
reference	1 - transprob	transprob * 0.3	transprob * 0.7
alternative	transprob * 0.3	1 - transprob	transprob * 0.7
heterozygous	transprob * 0.5	transprob * 0.5	1 - transprob

We selected the parameters manually based on visual assessment of how well the HMM captures the apparent structure in the data and that the results do not change appreciably when we vary the primary parameter by two orders of magnitude. The transprob takes a very small number [$1e-10$ / (total number of SNPs on the given chromosome) in this case] to reflect the belief that state transitioning at any individual SNP site should be a very rare event. The further breakdown of transprob by fractions of 0.3 and 0.7 aims to suppress rapid successive transitions of the form reference-alternative-reference or alternative-reference-alternative.

The emission matrix is specified as below:

State\Emission	reference	alternative
reference	0.9	0.1
alternative	0.1	0.9
heterozygous	0.5	0.5

After hidden states are called for each individual SNP, continuous long state blocks are called by removing state blocks shorter than 50kb. The crossover position is then determined by where the long state block switches to a different state, where the breakpoint tract start position is the last SNP position of the previous state block and the tract end position is the first SNP position of the following state block.

For M2 cells, an average allele frequency is first obtained by averaging over alleles within a window of 40 SNPs. The binned allele frequencies are then used to infer underlying chromosome states from a hidden Markov model with single Gaussian probability distributions.

The transition matrix is specified as below:

From\To	reference	alternative	heterozygous
reference	1 - transprob	0	transprob
alternative	0	1 - transprob	transprob
heterozygous	transprob * 0.5	transprob * 0.5	1 - transprob

The emission matrix is specified as below:

State	Emission
reference	Normal(0.05, 0.1)
alternative	Normal(0.5, 0.1)
heterozygous	Normal(0.95, 0.1)

Continuous long state blocks are called by removing state blocks shorter than 50kb, then approximate break point position is determined by where the long state blocks switches to a different state. The approximate break point position is then refined by a likelihood ratio test aiming to find the likely break point within the upstream 20 and downstream 20 SNPs around the approximate break point. For each SNP, the probability of observing the observed genotype is specified as:

State\Observed	reference	alternative
reference	1 - error_prob	error_prob
alternative	error_prob	1 - error_prob
heterozygous	0.5	0.5

The `error_prob` is specified as $1e-3$ which reflects the probability that a SNP is called incorrectly. For each SNP around the approximate break point, the likelihood of it being the actual break point is calculated by the above distribution. All SNPs with likelihood greater than $0.01 \times$ maximum likelihood are considered to be within the break point range. The start of the break tract is determined as the leftmost SNP within these SNPs, while the end of the break tract as the rightmost SNP. As in the 1C case, all M2 cell breakpoint tracts are further manually examined to remove artifacts, e.g., where two immediately adjacent switches are present within 50kb. We also performed the same breakpoint calling in mitotically dividing Patski cells. For M2 cells and Patski cells, we also manually examined breakpoint tracts by comparing bin sizes of 10 and 40 SNPs for cells with sparse genome coverage.

This step generate crossover break points summarized in [Table S3](#). We postprocess to add the chromosome segregation information based on whether the centromeric region, i.e., the starting region of each chromosome, is heterozygous (“mt,” mitotic segregation) or homozygous (“me,” meiotic segregation).

Analyses of uniparental chromosomes

This step takes the `rds` file from the HMM output and generates uniparental chromosome calls summarized in [Table S4](#) (See `call_uniparental_chrom.R` code and annotations).

Analyses of meiotic crossover and chromosome segregation at the chromosomal level

This step generates chromosomal level characteristics of meiotic crossovers shown in [Figure S4](#) (see sections [Figure S4](#) in `sci-L3-WGS-figures.Rmd` for R code and annotations).

Fitting a finite mixture model to the 2C cells in barcode group 2 in the (B6 × Cast) cross

We fit the data to a mixture of three binomial distributions parameterized by p_1 , p_2 , p_3 , respectively, denoting their probabilities of chromosomes segregating equationally. The relative contribution of these three binomial distributions are denoted by a length 3 vector θ . We estimate p_1 , p_2 , p_3 as well as θ by drawing samples from their posterior distributions using the R package `rstan` (<https://mc-stan.org/users/interfaces/rstan>) with uniform Dirichlet prior for θ : $\theta \sim \text{Dir}(K = 3, \alpha = 1)$, and beta prior for p : $p \sim \text{Beta}(a = 5, b = 5)$. For further details on the model specification, see the Stan file `mt_mixture_model.stan`.

Preprocessing of datasets from other genomic studies for building linear models of crossover hotness and cell clustering

We processed datasets from previous genomic studies and from downloaded mouse annotation file in `gff3` format and RepeatMasker from UCSC Genome Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>) in terms of various genome elements. Datasets based on mm9 are first lifted over to mm10. These datasets roughly fall into two categories: count data in `bed` format or signal of various genetic or epigenetic marks in `bedGraph` format. For cell clustering and predictive modeling, crossover tracts have different lengths. We normalize count data by dividing the total amount of sequences summed up from all the crossover in each single cell for the cell clustering analyses and we normalize by dividing tract lengths plus 1 kb for each crossover tracts or randomly sampled tracts such that extremely short tracts will not be overly weighted. Note that the median tract length is 150 kb such that adding the 1 kb do not include much extra sequence. For dataset with continuous signal of various marks, we take the average signal of marks that intersect with crossover or random tracts. For the crossover pileup dataset, since we used evenly-sized 100 kb windows, we did not normalize for tract lengths when using count data.

In addition to datasets mentioned in the [Discussion](#) section, where features have statistically significant association with crossover occurrence, we also used the following datasets: 1) sequence divergence ([Lilue et al., 2018](#)); 2) ATAC-seq and H3K27ac mapped from purified pachytene spermatocytes ([Maetzawa et al., 2018](#)); 3) bisulfite sequencing from spermatogonia ([Inoue et al., 2017](#)); 4) MNase-based nucleosome positioning in spermatocytes ([Barral et al., 2017](#)); 5) H4K5 and H4K8 butyrylation and acetylation in spermatocytes ([Goudarzi et al., 2016](#)); 6) H2A ubiquitination in spermatocytes ([Hasegawa et al., 2015](#)); 7) binding sites of CTCFL, the testis-specific paralog of CTCF binding sites ([Sleutels et al., 2012](#)); 8) 5-hmC map in pachytene spermatocytes ([Gan et al., 2013](#)); 9) End-seq after etoposide treatment and CTCF and RAD21 ChIP-seq in activated B cells, TOP2A and TOP2B ChIP-seq in MEFs ([Canela et al., 2017](#)); 10) Patski allelic ATAC-seq data ([Bonora et al., 2018](#)).

PCA for cell clustering, BMA for linear models of crossover hotness and random forest for predictive models of crossover and random tracts

Principal component analysis is used to visualize in 2D the separation of 1C and M2 cells based on their break point features. We aggregated crossover-related information for each single cell a total of 78 features corresponding to three types. As a first type, we simply calculated the number of crossover or whole-chromosome LOH events for each chromosome in each cell. As a second type, for features such as GC content, sequence divergence, intensity of chromatin marks, etc., we calculated median values for the crossover breakpoints in each cell. As a third type, we calculated normalized counts of genomic elements such as genes bodies, long terminal repeats (LTR), LINE elements that overlapped with crossover breakpoints in each cell.

Bayesian model averaging using the “`bas`” package ([Clyde et al., 2011](#)) is used to construct linear models predicting crossover hotness (function `bas.lm` sampling 2^{14} models with default settings), and variables important for predicting hotness are identified based on their marginal inclusion probabilities. Random forests are trained to distinguish true crossover tracts from tracts randomly sampled from the genome resembling the “null” distribution. Model accuracy is determined by full nested 5-fold cross validation, with 5 external folds and 5 folds within each training set (see section called “Models” in `sci-L3-WGS-figures.Rmd` for R code and annotations).

To estimate the strain (or cell type) effect on the positioning of the rightmost crossovers along chromosomes, we use a linear mixed effect model with fixed effect for strain (or cell type) and random intercept for chromosome to account for inter-chromosome variability (see section called “Karyotype Plots” in sci-L3-WGS-figures.Rmd for R code and annotations).

DATA AND CODE AVAILABILITY

Customized shell script “sci_lianti_v2.sh” for de-multiplexing (python scripts and the R Markdown file are uploaded separately as “sci_lianti_inst.tar.gz”; the R package containing intermediate data files for generating all the main and supplemental figures can be downloaded and installed via the following link: <https://drive.google.com/file/d/19NFubouHrahZ8WoblL-tcDrrTIIzEpJh/view?usp=sharing>).

ADDITIONAL RESOURCES

Detailed Protocol

See [Sci-L3 Method](#) section above.