
# Supplementary Materials for

## 3D Genome Structures of Single Diploid Human Cells

Longzhi Tan*, Dong Xing*, Chi-Han Chang, Heng Li, X. Sunney Xie.

correspondence to: xie@chemistry.harvard.edu

**This PDF file includes:**

Materials and Methods
Figs. S1 to S20
Tables S1 to S2
References (*42-44*)

## Materials and Methods

Human Subjects

The collection and analysis of peripheral blood from the male donor (the same as in (*16*)) were approved by the Institutional Review Board (IRB) at Harvard.

Published Data

Numbers of contacts were taken from Extended Data Table 1 (column "final contacts") of (*11*) ($n$ = 8), the GEO accession (GSE94489, column "total_contacts" from files "GSE94489_2i_diploids_features_table.txt.gz" and "GSE94489_serum_diploids_features_table.txt.gz", keeping only cells whose "group" column is "G1") of (*10*) ($n$ = 750), Table S1 (column "total number of contacts", keeping only cells whose "cell type" column is "K562", "Intermediate", "Intermediate-Hoechst", "SN", "SN-Hoechst", "NSN", or "NSN-Hoechst") of (*8*) ($n$ = 34 for K562 and $n$ = 120 for mOocyte).

Centromere coordinates were downloaded from the Table Browser ("human" -> "Feb. 2009 (GRCh37/hg19)" -> "mapping and sequencing" -> "gap", and "mouse" -> "Dec. 2011 (GRCm38/mm10)" -> "mapping and sequencing" -> "gap") of the UCSC Genome Browser.

Raw data of mESCs were downloaded from the "Diploid_20" sample from the GEO accession (GSE94489) of (*10*). Samples were demultiplexed by fastq-multx (*42*) (version 1.3.1). A subset of 10 diploid mESCs were chosen so that they were both flow-sorted and inferred *in silico* to be G1 phase, harbored no large chromosomal aberrations, and were among the top 2i-cultured cells in terms of the number of contacts. These cells (**Table S1**) yielded a median of 0.309 million author-defined contacts ($n$ = 10, min = 0.298 million, max = 0.371 million) (*10*), or 0.226 million by our definition (min = 0.201 million, max = 0.284 million).

Raw data of the best mouse cell ("1_oocyte_NSN", tetraploid) and the best human cell ("54_K562-B", presumably triploid) were downloaded from the GEO accession (GSE80006) of (*8*). These two cells yielded 1.91 (87% intrachromosomal) and 0.36 (86% intrachromosomal) million author-defined contacts, respectively. With our "dip-c" package, the two cells yielded 2.43 (84% intrachromosomal) and 0.76 (54% intrachromosomal) million contacts, respectively, corresponding to 1.22 and 0.38 million per 2N.

SNP data were taken from the Illumina Platinum Genomes FTP site (the file "2016-1.0/hg19/small_variants/NA12878/NA12878.vcf.gz") for GM12878, Mingyu Yang (*16*) (the folder "Sperm_project_released_data/04.haplotypes/combined_haplotypes") for the blood donor, and the Sanger Institute Mouse Genomes Project (the file "mgp.v5.merged.snps_all.dbSNP142.vcf.gz") for mESCs.

Promoter capture Hi-C interactions were taken from "PCHiC_peak_matrix_cutoff5.tsv" in Data S1 of (*35*). Only intra-chromosomal interactions with a minimal genomic distance of 100 kb were kept ($n$ = 632,986). For each interaction, interaction strength in B lymphocytes, T lymphocytes, or monocytes/neutrophils, respectively, was calculated by averaging corresponding columns ("nB" and "tB" for B lymphocytes, "nCD4", "tCD4", "nCD8" and "tCD8" for T lymphocytes, "Mon" and "Neu" for monocytes/neutrophils). Cell-type-specific interactions were defined as strength $\geq 6$ in the cell type of interest and $\leq 3$ in the other 2

cell types ($n$ = 12,338 for B lymphocytes and $n$ = 8,538 for monocytes/neutrophils). In each single cell, the level of cell-type-specific promoter proximity was defined as the percentage of interacting pairs that were within a 3D distance of 3 particle radii.

Bulk Hi-C on GM12878 (*19*) and on T lymphocytes (*40*) was visualized by Juicebox.js (*41*) with balanced normalization.

Generation of a List of Phased SNPs

For GM12878, all 2.15 million heterozygous SNPs were extracted from the VCF file, assuming a genotype format of "paternal | maternal".

For PBMCs, raw sequencing reads of the family trio were downloaded from the SRA (SRX205465 for the blood donor, SRX205467 for the donor's mother, SRX205466 for the donor's father) (*16*) and mapped without pre-processing. SNPs were jointly called by GATK (version 3.8.0) HaplotypeCaller with default parameters and filtered with recommended parameters ("-selectType SNP" for "SelectVariants", "--filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"" for "VariantFiltration", and "--excludeFiltered -restrictAllelesTo BIALLELIC" for "SelectVariants" again). Population-based phasing on autosomes were performed by shapeit (version v2.r837) with its genetic map ("genetic_map_b37"), reference panel ("1000GP_Phase3"), and recommend effective population size for Asians ("--effective-size 14269"), yielding 1.92 million phased SNPs. Population-based phasing was then combined with published sperm-based phasing (1.23 million SNPs), yielding 1.98 million phased SNPs. When in conflict (0.03 million out of 1.17 million), sperm-based results were used.

For mESCs, the parental lines were 129S4/SvJaeJ × Castaneus (*11*). As an approximation, 129S1/SvImJ and CAST/EiJ were extracted from the dbSNP VCF file, yielding 21.50 million sites that were different between the two strains.

Identification of the Active and the Inactive X Chromosomes by RNA Expression

Although the GM12878 cell line that we obtained (Coriell Institute) was not a single-cell clone, it anecdotally preferred the maternal X chromosome as the active one. To confirm this by RT-PCR, we extracted total RNA with RNeasy Mini with DNase (Qiagen) with a centrifugal homogenizer (Invitrogen), and synthesized cDNA with ProtoScript II (NEB) with random primers. Regions harboring the following heterozygous SNPs (according to ENCODE) were amplified by Q5 Hot Start High-Fidelity Master Mix (NEB) and analyzed by Sanger sequencing:

1. XIST (from the inactive X): rs1620574, paternal = C, maternal = T, primers = ACTGGATGGAAGACCACAAC + GTGTCTTGGGTAGCAGAAGAA.
2. EBP (from the active X): rs3048, paternal = T, maternal = G, primers = TATACACACGCAGCCATCAG + CTTCACAGCATCAAGCACAAG.
3. TBL1X (from the active X): rs16985675, paternal = A, maternal = G, primers = TGTGATGGCTGAATGGAAAGA + GAAAGGTACAGAGGGAGAGAGA.
4. SLC25A53 (also known as MCART6, from the active X): rs5916825, paternal = A, maternal = G, primers = GCACTGCAGGTGGAAAGAATA + GCTGTGGCTGGAAATCCTAAA.
5. ATRX (from the active X): rs3088074, paternal = G, maternal = C, primers = TCTCCATCAGTTGTTCCATTCT + CTTCCACTGATGGTGTCGATAA.

Isolation of PBMCs from Blood

Blood was drawn into K2EDTA-coated tubes (BD) and placed on ice immediately. PBMCs were isolated according to the manual of Ficoll-Paque PLUS (GE), with 1 X PBS + 2 mM EDTA as the salt solution.

Chromatin Conformation Capture in Dip-C

We improved the sensitivity from three aspects. First, all biotin-related steps were omitted. This modification avoided inefficient procedures of single-cell biotin-pulldown and blunt-end DNA ligation, baring more similarities to the original 3C (*4*) than to its Hi-C derivatives (*5, 19*), and was independently adopted by others (*8, 9*). Second, the ligation product was amplified by a sensitive and uniform whole-genome amplification method, META (Xing *et al*. U.S. provisional patent 62/509,981), analogous to our recently published LIANTI (*14*). Third, META added sequencing adapters through PCR, rather than through ligation in traditional library preparations. This modification reduced artefactual paired-end chimera, and was independently adopted by others (*10, 11*). Below is the detailed procedure.

Cells were fixed by 2% PFA (EMS) in PBS or culture media without serum at room temperature for 10 min with rotation. PFA was quenched by the addition of 2 M glycine (0.2-um filtered) to a final concentration of 0.127 M and incubation on ice for 5 min. Cells were washed by ice-cold PBS (centrifugation: 600 g, 5 min) and pellets were stored at −80 C.

Permeabilization and digestion were performed by removing biotin-related steps from published Hi-C protocols. Results were comparable between different protocols.

In one variant (based on (*10*); for replicate 1 of GM12878), cells were first permeabilized in 1 mL ice-cold Permeabilization Buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% Igepal CA 630 (Sigma), cOmplete Mini EDTA-free (Roche)) on ice for 30 min with occasional inversion, washed by 800 uL 1.24 X NEBuffer 3 (NEB) (centrifugation: 600 g, 6 min), and further permeabilized in 400 uL 1.24 X NEBuffer 3 + 0.3% SDS at 37 C for 1 h with 950 RPM shaking. SDS was quenched by the addition of 40 uL 20% Triton X-100 and incubation at 37 C for 1 h with 950 RPM shaking. Cells were then digested by the addition of 50 uL 25 U/uL MboI (NEB R0147M) and incubation at 37 C overnight with 950 RPM shaking.

In another variant (based on (*19*); for replicate 2 of GM12878 and PBMCs), cells were first permeabilized in 500 uL ice-cold Hi-C Lysis Buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% Igepal CA 630) and 100 uL protease inhibitors (Sigma P8340) on ice for ≥ 15 min, washed by ice-cold Hi-C Lysis Buffer (centrifugation: 2500 g, 5 min), and further permeabilized in 50 uL 0.5% SDS at 62 C for 10 min. SDS was quenched by the addition of 145 uL water and 25 uL 10% Triton X-100 and incubation at 37 C for 15 min with rotation. Cells were then digested by the addition of 25 uL 10 X NEBuffer 2 (NEB) and 20 uL 25 U/uL MboI (NEB R0147M) and incubation at 37 C overnight with rotation. This variant might reduce clumping and disappearance of some cells.

On the next day, cells were washed by 1 mL Ligation Buffer (1 X T4 DNA ligase buffer (NEB B0202S), 0.1 mg/mL BSA (NEB B9000S)), and ligated in 1 mL Ligation Buffer and 10 uL 1 U/uL T4 DNA ligase (Life Tech 15224-025) at 16 C for 4 h.

Single-cell Isolation by Flow Cytometry

For GM12878 and PBMCs, ligated cells (in ligation buffer) were filtered by a 40-um cell strainer (Falcon) and sorted into 0.2-mL UV-irradiated DNA low-bind tubes (MAXYMum Recovery, Axygen) containing lysis buffer with a FACSJazz flow cytometer (BD; 100-um nozzle). Events were first gated on FSC and SSC as "cells", and then on FSC and trigger pulse width as "singlets". The sorting mode was "1.0 drop single".

Whole-genome Amplification in Dip-C

*Design of META*

In Nextera kits (Illumina), two different tags — each harboring one side of the Illumina adapter — were randomly inserted into the input DNA. The two tags would then function as two PCR primers to amplify the resulting fragments. Fragments that ended with two different tags would be amplified, while fragments that ended with two same tags would be lost. As a result, at least 50% of input DNA would be lost.

In META (Xing *et al*. U.S. provisional patent 62/509,981), such loss was greatly reduced by inserting $n$ different tags. As a result, only $1/n$ of input DNA would be lost. Illumina adapters were added later by two short PCR steps.

In this work, we used META with $n = 20$ tags:

1. AGAAGCCGTGTGCCGGTCTA
2. ATCGTGCGGACGAGACAGCA
3. AATCCTAGCACCGGTTCGCC
4. ACGTGTTGCAGGTGCACTCG
5. ACACCACACGGCCTAGAGTC
6. TGGACAATCACGCGACCAGC
7. TCATCTAACGCGCACCGTGC
8. TTCGTCGGCTCTCTCGAACC
9. TGGTGGAGCGTGCAGACTCT
10. TATCTTCCTGCGCAGCGGAC
11. CTGACGTGTGAGGCGCTAGA
12. CCATCATCCAACCGGCTTCG
13. CACGAGAAGCCGTCCGCTTA
14. CGTACGTGCAACACTCCGCT
15. CTTGGTCAGGCGAGAAGCAC
16. GGCGTGATCAGTGCGTGGAT
17. GAGCGTTTGGTGACCGCCAT
18. GCCTGCGGTCCATTGACCTA
19. GTAAGCCACTCCAGCGTCAC
20. GATCTGTTGCGCGTCTGGTG

*Preparation of META reagents*

Carrier ssDNA (for use in Lysis Buffer) could be either the same as in LIANTI (5′-TCAGGTTTTCCTGAA-3′) (*14*) or the same as the META 20-primer Mix. Store at −20 C.

Transposome (for use in Transposition Mix) was partly similar to LIANTI (*14*), but with two modifications. First, one strand of the transposon was 5′-/Phos/-

CTGTCTCTTATACACATCT-3′, while the other strand was in the form of 5′-[META tag]-AGATGTGTATAAGAGACAG-3′. Each of the oligos (IDT, purification: PAGE) was dissolved in 0.1 X TE to a final concentration of 100 uM. For each of the $n = 20$ META tags, two strands were annealed at a final concentration of 5 uM each. The 20 annealed transposons were then pooled with equal volumes. Second, the transposase was purified after expression from the pTXB1-Tn5 plasmid (Addgene). Transposome was assembled at a final concentration of 1.25 uM dimer (2.5 uM monomer), 1:10 diluted (125 nM dimer, or 250 nM monomer), and aliquoted for single uses and store at −80 C.

20-primer Mix (for use in PCR Mix 1) was in the form of 5′-[META tag]-AGATGTGTATAAG-3′. Each of the oligos (IDT, purification: standard desalting) was dissolved in 0.1 X TE to a final concentration of 100 uM, and combined with equal volumes (100 uM total, or 5 uM each). Store at −20 C.

40-primer Mix (for use in PCR Mix 2) was in the form of 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-[META tag]-AGATGTGTATAAG-3′ for one side of the Illumina adapter, and 5′-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-[META tag]-AGATGTGTATAAG-3′ for the other. Each of the oligos (IDT, purification: PAGE) was dissolved in 0.1 X TE to a final concentration of 50 uM, and combined with equal volumes (50 uM total, or 1.25 uM each). Store at −20 C.

*Cell lysis*

Single cells were lysed in 3 uL META Lysis Buffer (20 mM Tris pH 8.0, 20 mM NaCl, 0.1% Triton X-100, 15 mM DTT, 1 mM EDTA, 1.5 mg/mL Qiagen protease, 0.5 uM carrier ssDNA) at 50 C for 6 h, 65 C for 12 h, 70 C for 30 min. For some cell types and recent lots of Qiagen protease, lysis may need to be shortened (for example 50 C or 65 C for 1 h, 70 C for 15 min). Lysed cells could be stored at −80 C for a few months.

Alternatively, single cells might be directly placed in empty tubes and stored at −80 C for longer times before addition of META Lysis Buffer, although we have not tested this extensively.

*Whole-genome amplification*

Lysate was transposed by the addition of 5 uL Transposition Mix (leading to a final concentration of 10 mM TAPS pH 8.5, 5 mM MgCl2, 8% PEG 8000, 1:2,640 (0.5 nM dimer) META transposome) and incubation at 55 C for 10 min. Transposases were removed by the addition of 2 uL Stop Mix (1 uL 2 mg/mL Qiagen protease diluted in water, and 1 uL 0.5 M NaCl, 75 mM EDTA) and incubation at 50 C for 40 min, 70 C for 20 min.

Whole-genome amplification was performed by the addition of 10 uL PCR Mix 1 (4 uL Q5 reaction buffer (NEB), 4 uL Q5 high GC enhancer (NEB), 0.5 uL 100 mM MgCl2, 0.5 uL 100 uM (total) META 20-primer Mix, 0.4 uL 10 mM (each) dNTP mix, 0.2 uL water, 0.2 uL 20 mg/mL BSA (NEB B9000S), 0.2 uL Q5 (NEB M0491S)) and incubation at 72 C for 3 min, 98 C for 20 s, 12 cycles of [98 C for 10 s, 65 C for 1 min, 72 C for 2 min], and 65 C for 5 min.

Optionally, the amplification product could be purified at this step and analyzed with a High Sensitivity DNA Kit on a Bioanalyzer (Agilent) for quality control.

*Library preparation*

Sequencing libraries were prepared by two additional PCR steps. In the first PCR step, previous primers were removed by the addition of 0.5 uL 20 U/uL ExoI (NEB M0293S) and incubation at 37 C for 30 min, 72 C for 20 min. White precipitates might form at this step or at the following steps. PCR was performed by the addition of 9.5 uL PCR Mix 2 (2 uL Q5 reaction buffer (NEB), 2 uL Q5 high GC enhancer (NEB), 3 uL 50 uM (total) META 40-primer Mix, 0.2 uL 10 mM (each) dNTP mix, 2.2 uL water, 0.1 uL Q5 (NEB M0491S)) and incubation at 98 C for 30 s, 2 cycles of 98 C for 10 s + 65 C for 1 min + 72 C for 2 min, and 65 C for 5 min. In the second PCR step, primers were similarly removed by the addition of 0.5 uL 20 U/uL ExoI (NEB M0293S) and incubation at 37 C for 30 min, 72 C for 20 min. PCR was similarly performed by the addition of 2.5 uL NEB Index Primer (NEB E7335S, E7500S, E7710S, E7730S) and 7 uL PCR Mix 3 (2 uL Q5 reaction buffer (NEB), 2 uL Q5 high GC enhancer (NEB), 2.5 uL NEB Universal Primer, 0.2 uL 10 mM (each) dNTP mix, 0.2 uL water, 0.1 uL Q5 (NEB M0491S)) and incubation at 98 C for 30 s, 2 or more cycles of 98 C for 10 s + 65 C for 1 min + 72 C for 2 min, and 65 C for 5 min. Libraries could be pooled at this step or at any step afterwards.

Libraries were purified by a DNA Clean and Concentrator-5 column (Zymo D4013) with 200 uL DNA Binding Buffer (a ratio of 1:5) and eluted in 25 uL 0.1 X TE. A representative Bioanalyzer trace was shown in **Fig. S1A**. Size selection was performed with Ampure XP beads (Beckman Coulter, typically 0.65 X) (**Table S2**). A representative Bioanalyzer trace was shown in **Fig. S1B**. Note that only a fraction of molecules were Illumina libraries (harboring genomic DNA + 78 bp or more META adaptors + ~ 130 bp Illumina adaptors), while others had incompatible or partial Illumina adapters. Therefore, libraries must be quantified by qPCR.

*Sequencing*

Libraries were quantified by qPCR and sequenced with paired-end 250-bp reads on a HiSeq 2500 (Illumina). To avoid diversity issues (especially at the 19 bp right after the META tag), 20% PhiX was added. A list of all libraries was shown in **Table S2**. Raw sequencing outputs were 10–47 Gb per cell, corresponding to raw sequencing depths of 3–16 X. Similar to LIANTI (*14*), mapped outputs were lower because of transposed sequences and overlapping reads 1 and 2.

Statistical Property of Interchromosomal and Long-range Intrachromosomal Contacts

In bulk Hi-C, the probability of intrachromosomal contacts is well known to systematically decrease over genomic separation (bp) — in particular, the probability of a contact joining coordinate $x$ (bp) and $x + \Delta x$ on the same chromosome is approximately proportional to $\Delta x^{-1}$ (the "fractal globule" model) (*5, 10, 34*).

We wondered if a similar rule governed interchromosomal contacts. In particular, given that an interchromosomal contact joined coordinate $x$ on one chromosome and $y$ on another, the two contacting chromosomes might be seen as "tethered" at ($x$, $y$) and thus formed more contacts nearby. Such conditional properties were hidden in bulk Hi-C because interchromosomal contacts were highly stochastic.

We began by considering the conditional probability, given a contact joining coordinate $x$ (bp, called a "leg") on one chromosome and $y$ (bp, the other "leg" of the

same contact) on another, that another contact joined $x + \Delta x$ and $y + \Delta y$. Naively, if the two contacting chromosomes were "concatenated" at ($x$, $y$) and intermingled as if they were a single chromosome, the conditional probability density would be proportional to $(\Delta x + \Delta y)^{-1}$ — in other words, the inverse of the $L^1$ norm of ($\Delta x$, $\Delta y$). In contrast, we found the conditional probability density to be approximately proportional to the inverse of the $L^{0.5}$ distance (not a norm; also known as the Minkoski distance of order 0.5) of ($\Delta x$, $\Delta y$); in other words, $p(\Delta x, \Delta y) \propto (\sqrt{\Delta x} + \sqrt{\Delta y})^{-2}$ (**Fig. S7A**). Therefore, simultaneously large Δx and Δy were relatively disfavored, suggesting that two contacting chromosomes tended not to fully intermingle but rather protrude into each other. This empirical formula held across different genomic scales and for long-range intrachromosomal contacts (**Fig. S7C**). Similar to the unconditional $\Delta x^{-1}$ density of intrachromosomal contacts (*5, 10, 34*), the conditional density also varied systematically across the cell cycle (**Fig. S7D**). Utilizing the $L^{0.5}$ property, we defined a contact "neighborhood" as a superellipse with radius = 10 Mb and exponent = 0.5, where haplotypes of nearby contacts could vote to impute the haplotypes of each contact of interest (**Fig. S7E**).

Haplotype Imputation and 3D Reconstruction in Dip-C

*Overview*

Code will be available on GitHub as a "dip-c" package (https://github.com/tanlongzhi/dip-c). Starting from FASTQ files, 3D reconstruction consisted of the following steps: preprocessing → alignment → contact identification → artifact removal → haplotype imputation (2D) → [with replicates from here on] 3 rounds of 3D reconstruction at 100-kb resolution + haplotype imputation (3D) → 2 rounds of 3D reconstruction at 20-kb resolution + haplotype imputation (3D).

An alternative, faster implementation of the Dip-C algorithm will be available as part of a "hickit" package (https://github.com/lh3/hickit). Haplotype imputation in "hickit" does not involve 3D modeling.

Details of the procedures and file formats will be found on the GitHub page; and we welcome suggestions and efforts to make them better and easier to use for the single-cell 3C/Hi-C community.

Below is a brief description of each analysis step in the "dip-c" package:

*Preprocessing*

Most reads followed a format of [META tag]-AGATGTGTATAAGAGACAG-[genomic DNA]-CTGTCTCTTATACACATCT-[reverse complement of another META tag], although a small fraction harbored extra META adaptors (and very rarely, genomic DNA in between, which was discarded). Similar to (*14*), META and Illumina adapters were removed, and the two ends (read 1 and read 2) were merged if they overlapped.

*Alignment*

Similar to (*14*), reads were mapped by BWA-MEM (*43*) (version 0.7.15) with default parameters to the human reference genome GRCh37 (for GM12878 and PBMCs) or to the mouse reference genome GRCm38.p5 (GENCODE, for mESCs). A representative distribution of template (genomic DNA) lengths was shown in **Fig. S1C**.

*Contact identification*

From each read or read pair, all high-quality (mapping quality ≥ 20, edit distance per bp alignment ≤ 0.05) primary and supplementary alignments (in BWA-MEM, different parts of a single read can be locally aligned to different regions of the genome, producing one primary alignment and one or more supplementary alignments) were extracted as "segments". If a segment overlapped with a phased SNP, a haplotype would be assigned if base quality ≥ 20. Chromatin contacts were identified as all pairs of segment end points — each end point (a genomic coordinate, with a haplotype if applicable) defined as a "leg" (for example, "chr14, 23,882,391 bp, paternal") — that were separated by > 1 kb in each read or read pair. Each contact would thus contain two legs (for example, "chr14, 23,882,391 bp, paternal — chr14, 23,448,197 bp, unknown"). An example of this procedure was shown in **Fig. S14**. PCR duplicates were removed by iteratively merging near-identical (both legs differed by ≤ 1 kb) contacts. Depending on size selection and cells, 4–15% of sequencing reads yielded contacts (**Table S2**).

*Artifact removal*

Similar to (*11*), "promiscuous" legs (alignment artifacts) were removed if > 10 other legs fell within 1 kb; subsequently, "isolated" contacts (3C/Hi-C artifacts) were removed if < 5 other contacts fell within 10 Mb in $L^{0.5}$ distance (instead of $L^{\infty}$ distance in (*11*)).

*Haplotype imputation (2D)*

In each round of imputation, contacts in an "evidence" set voted to impute unknown haplotypes of contacts in a "target" set. For each target contact, a list of compatible haplotype tuples was first enumerated. For example, a contact joining the maternal chromosome 1 and an unknown haplotype of chromosome 2 would be compatible with two possible haplotype tuples, (Chr 1♀, Chr 2♂) and (Chr 1♀, Chr 2♀). Each evidence contact would then vote for haplotype tuples from this list, if such contact fell within 10 Mb in $L^{0.5}$ distance from the target contact and was compatible with one and only one haplotype tuple from the list. Imputation would occur if the winning haplotype tuple gathered ≥ 3 votes and ≥ 90% of all votes. An example of the above procedure was shown in (**Fig. S7E**).

Special care was taken for intrachromosomal contacts because intrahomologous contacts were far more frequent than interhomologous contacts, especially at short ranges (small genomic separation). A target contact would be assumed intrahomologous without voting, if its two legs were separated by ≤ 10 Mb; otherwise, voting still occurred but a winning interhomologous vote would only be accepted if two legs were separated by ≥ 100 Mb. In addition, intrachromosomal contacts that had unknown haplotypes on both legs were not imputed.

The imputation procedure began with all contacts that had known haplotypes on at least one leg as both the target and the evidence sets. Such imputation was repeated two more times, each time with previous results as the new evidence set. Results were subsequently cleaned by removal of isolated contacts (< 2 other contacts that had the same haplotypes within 10 Mb in $L^{0.5}$ distance). Finally, cleaned results were used as the evidence set to impute a target set of all interchromosomal contacts that had unknown haplotypes on both legs.

For males, pseudoautosomal regions (PARs) were excluded from this step.

The above imputation procedure is essentially semi-supervised classification with genomic coordiantes as the input and haplotypes as the labels. The first pass is even simpler, resembling a nearest neighbor classifier in supervised classification.

*3D reconstruction*

Simulated annealing was performed by nuc_dynamics (*11*) (parameters: "-temps 20 -s 8 4 2 0.4 0.2 0.1" for 100-kb structures or "-temps 20 -s 8 4 2 0.4 0.2 0.1 0.04 0.02" for 20-kb structures) with minor modifications. First, the backbone energy function remained harmonic for large distances to reduce imputation errors. Second, removal of isolated contacts was skipped because it was already performed. Third, the output was in a simple "3D genome (3DG)" format (tab delimited: chromosome name, genomic coordinate (bp), $x$, $y$, $z$) because the original PDB format did not allow > 99,999 atoms. An example code was provided to convert 3DG to mmCIF for visualization in PyMol (run "set connect_mode, 4" before loading).

An alternative method for 3D reconstruction, force-directed graph (FDG), provided a simpler algorithm for fast 3D visualization and potentially for the generation of final 3D structures. A proof-of-concept implementation could be found in the "force" command of the "dip-c" package; a more comprehensive implementation could be found in the "hickit" package.

*Haplotype imputation (3D)*

Partly similar to (*11*), unknown haplotypes of each contact were imputed by comparing 3D distances in a draft structure. For each contact, a list of compatible haplotype tuples was first enumerated (same as 2D imputation). For each possible haplotype tuple, 3D positions of the two leg were calculated by linear interpolation along the polymer of particles, between which the 3D distance was recorded. Imputation would occur if the winning haplotype tuple (the shortest 3D distance) yielded a 3D distance $\leq 20$ particle radii and $\leq 0.5$ times the second shortest 3D distance. Intrachromosomal contacts whose legs were separated by < draft structure resolution (bp) were not imputed. Finally, results were cleaned by removal of isolated contacts (< 2 other contacts that had the same haplotypes within 10 Mb in $L^{0.5}$ distance). For males, PARs were included in this step.

*Removal of repetitive regions*

Similar to (*11*), particles that harbored few contacts, such as centromeres and heterochromatic repeats, were removed from the final 3D structure. For each particle, the number of contact legs within 0.5 Mb was recorded. The bottom 6% of all particles were removed.

*Cross-validation*

For each GM12878 cell, 10% of all SNPs were randomly held out from the list of phased SNPs. The imputed haplotype of each leg was compared to original leg (ground truth). Imputation accuracy was estimated by the fraction of correctly imputed legs given that the ground truth was known (~ 5 k such legs per cell) (**Table S1**).

Analysis of 3D Structures

*Estimation of reconstruction uncertainty*

Similar to (*11*), three replicate structures were generated with different random seeds. After removal of repetitive regions (see above), shared genomic particles were extracted from the replicates and aligned with the Kabsch algorithm in a pairwise manner (**Fig. S9**). For each particle, r.m.s. deviation was calculated between all pairs of replicates. A lower bound for reconstruction uncertain of each cell was estimated by the median r.m.s. deviation across all particles.

For mESCs, five replicates were generated, because their smaller numbers of contacts and limited chromosome intermingling occasionally led to suboptimal structures. These suboptimal structures satisfied fewer contacts than other replicates, and were thus excluded (**Table S1**). Typical median r.m.s. deviation was ~ 1.3 particle radii (~ 130 nm).

*Chromosome intermingling*

The extent of chromosome intermingling of each 20-kb particle was quantified according to the chromosomes of nearby particles within 3 particle radii. Basic intermingling was defined as the fraction of nearby particles (excluding itself) that were not from the same chromosome. Multi-chromosome intermingling was defined as Shannon's diversity index of chromosomes ($-\sum p_i \ln p_i$, where $p_i$ denoted the fraction of nearby particles from chromosome $i$). Another measure of diversity — species richness (the number of nearby chromosomes) — yielded similar results. The two homologs of the same chromosome were counted as two different chromosomes. Note that quantification could be unreliable near repetitive regions such as the centromere.

*Single-cell chromatin compartments*

In a previous study (*10*), the single-cell chromatin compartment ("compartment association score") of each genomic bin was defined as the average A/B compartment (as measured by bulk Hi-C) of other bins that it contacted. Partly similar to this definition, we defined the single-cell chromatin compartment of each 20-kb particle as the average CpG frequency (a proxy of A/B compartments (*21*)) of nearby particles (including itself) within 3 particle radii. This definition was equivalent to 3D smoothing/diffusion of CpG frequencies in each structure.

In some analysis, compartments were rank-normalized to 0–1 in each cell because CpG frequencies were more variable in highly euchromatic regions.

Note that in addition to simple A/B compartments (with CpG frequency (*21*) or GC content (*8*) as a proxy), the above calculation could also be performed on other genomic vectors such as sub-compartments (*19*) (such as the polycomb), DNA methylation, and ChIP-Seq.

*Relationship between the single-cell chromatin compartments of the two alleles*

Single-cell compartment of each genomic locus has been reported to vary between single cells (*10*); however, the maternal and paternal alleles could not be distinguished, and it remained unclear whether the two varied in a coordinated manner. Across the genome, single-cell compartment varied both between different cells ("between cells") and between the two alleles of the same cell ("within cells"), and had near-identical averages for the maternal and paternal alleles (**Fig. S12A**). Difference between cells concentrated in regions whose average compartment was neither extremely euchromatic nor extremely heterochromatic (left panel of **Fig. S12C**), consistent with a previous

report on mice (*10*). Interestingly, difference within cells followed a near-identical pattern (**Fig. S12A**, right panel of **Fig. S12C**), suggesting that cell-to-cell heterogeneity of chromatin compartments was dominated by allele-to-allele heterogeneity. Supporting this idea, we found on average near-zero Spearman's correlation (median = 0.02, 0.03, and 0.05 for GM12878, presumable T lymphocytes, and combined, respectively) between the maternal and paternal compartments (**Fig. S12D**). Therefore, compartments of the two alleles behaved almost independently in a single cell.

*Potential somatic pairing*

We captured potential somatic pairing — spatial proximity of the maternal and paternal alleles — in our 3D models. We found a higher degree of somatic pairing in the PBMCs, with up to 0.4% of all 20-kb particles residing within 5 particle radii (~ 500 nm) from their homologs, while mESCs exhibited the lowest degree of pairing, consistent with their lowest levels of chromosome intermingling (**Fig. S12F**). **Fig. S12G** visualized two pairs of intermingling homologs: (1) the two homologs of chromosome 11 in a GM12878 cell lightly intermingled on their surfaces, with only a few loci paired — including the imprinted H19/IGF2 locus (consistent with (*44*)); and (2) the two homologs of chromosome 19 in a T lymphocyte aligned and extensively intermingled, with multiple loci paired across the chromosome. Note that imputation was less effective when homologous loci were in close proximity, leading to higher reconstruction uncertainty.

*Identification of single-cell domains through matrices of radii of gyration*

In a previous study (*11*), 3D sizes — as measured by radii of gyration — were calculated in each cell for bulk-defined domains. We extended this calculation to all $n(n + 1)/2$ possible subchains of each chromosome (a polymer of $n$ point-mass particles), yielding a matrix whose rows, columns, and values represented the starts, ends, and radii of subchains, respectively. Note that radii were less accurate near repetitive regions because those particles were excluded from calculation.

In a matrix of radii of gyration, single-cell domains were identified as squares that had relatively small radii (partly similar to (*8*)). Each 20-kb particle was initialized as a single domain (radius = 0). In each round, all possible ways of merging two adjacent domain into one domain were enumerated; and the one leading to the smallest radius was performed. Merging repeated until the entire chromosome became a single domain.

Note that in addition to radii of gyration, the above calculation could also be performed with other measures of 3D sizes, such as higher moments to further disfavor a few protruding particles, or relative sizes normalized by numbers of particles to favor larger but denser domain.

*Configuration of centromeres and telomeres*

A previous study (*11*) noticed a tendency for Rabl configuration in mESCs. We further quantified the extent of Rabl configuration by the length of summed centromere-to-telomere vectors normalized by the total particle number. This definition was equivalent to the length (in particle radii) of the average projection of vectors connecting each 20-kb particle to its centromeric neighbor.

We quantified the radial difference between centromeres and telomeres by the summed centromere-to-telomere difference in distances from the nuclear center of mass

normalized by the total particle number. This definition was equivalent to the average change in distance (in particle radii) from the nuclear center of mass between each 20-kb particle and its centromeric neighbor.

Contact-based Analysis

*Overview*

To rule out artifacts from the 3D modeling procedure, some of our conclusions were confirmed by contact-based, instead of 3D-structure-based, analysis.

*Single-cell chromatin compartments*

Similar to a previous study (*10*), the single-cell chromatin compartment of each genomic bin was defined as the average CpG frequency (instead of A/B compartment in the previous study) of other bins (excluding self) that it contacted ("contacting CpG"), weighted by the number of contacts. Haplotype-resolved contacts (raw contacts whose haplotypes were assigned through imputation) were used for diploid features, while both haplotype-resolved and raw contacts were used for cell-type features.

*Promoter-enhancer looping*

The formation of a promoter-enhancer loop was defined as having at least one raw contact (not haplotype-resolved) within ± 20 kb of its predefined (by bulk promoter capture Hi-C (*35*)) endpoints.

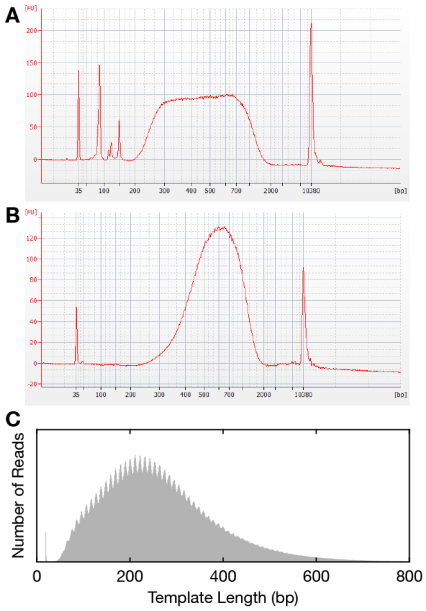



**Fig. S1.**

**Fragment length distributions after whole-genome amplification. (A)** Bioanalyzer trace of a representative single cell after whole-genome amplification and library preparation but before size selection. Between the two Bioanalyzer DNA markers (leftmost and rightmost peaks), desired products formed a smear of 200–2,000 bp, while some undesired products (left peaks) remained after Zymo purification. **(B)** Bioanalyzer trace of a representative pooled library after size selection by 0.65 X Ampure beads. Only desired products > 300 bp remained. **(C)** Histogram of template (genomic DNA) lengths of concordantly mapped reads from a representative cell.

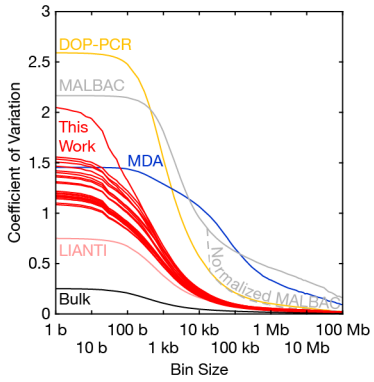
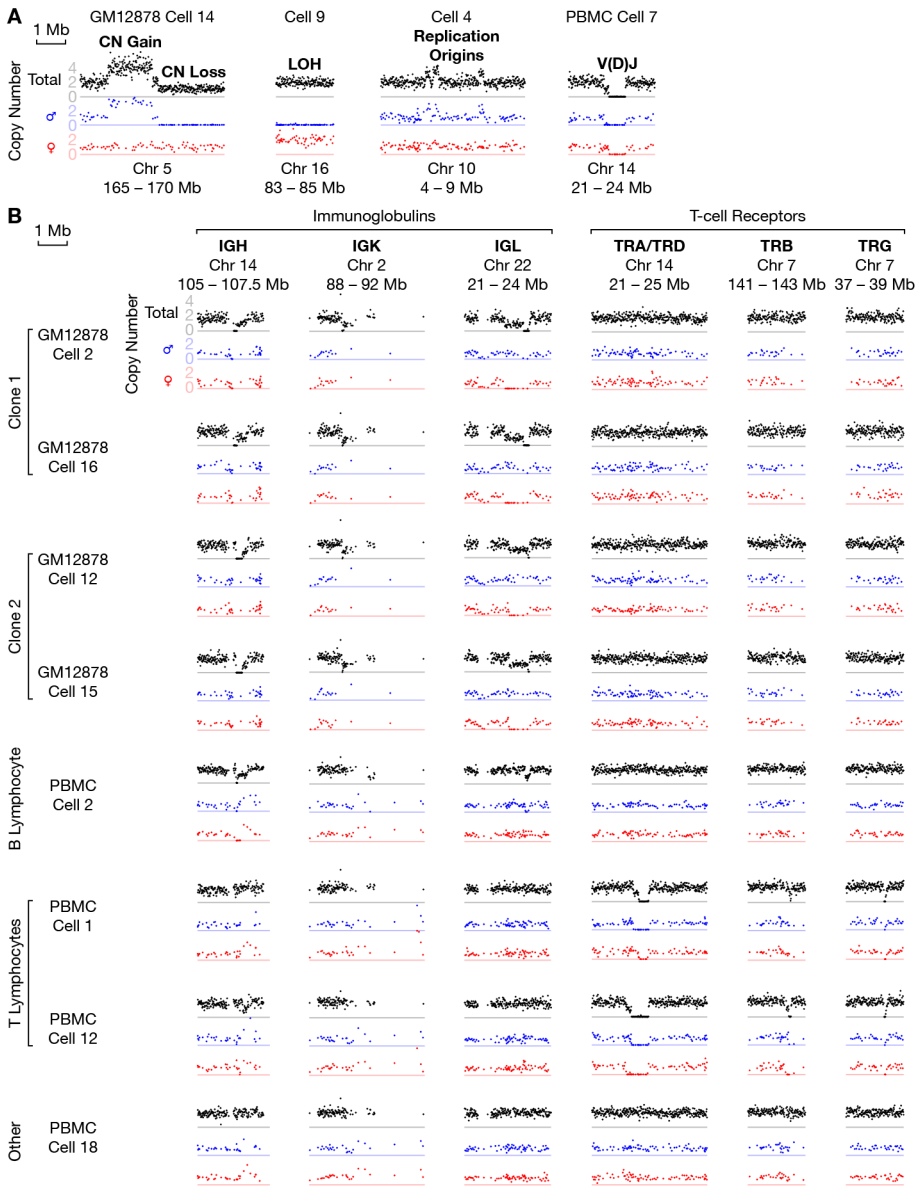



**Fig. S2.**

**Amplification uniformity was similar to that of LIANTI for bin sizes ≥ 10 kb, as measured by the coefficient of variation (CV) of binned read depths.** The CV plot of each all single cells (red) overlaid on those of bulk and other whole-genome amplification methods (*14*). S-phase cells (GM12878 Cells 4 and 13), cells with copy-number variations (CNVs) (GM12878 Cells 1, 10, 11, 12, 14, and 15), and the damaged cell (GM12878 Cell 8) were excluded.




## Fig. S3.

**Simultaneous detection of copy-number variations (CNVs), including V(D)J recombination, at high resolution. (A)** Binned read depths around representative regions of copy-number (CN) changes. Each dot represented the average depth of a 10-kb bin after repeat masking (*14*) for the total CN (black), or of 50 consecutive SNPs for the haplotype CNs (blue and red). Depths were normalized by setting the genome-wide median to 2 (black) or 1 (blue and red). **(B)** Binned read depths around immunoglobulin (IGH: heavy; IGK: light κ; IGL: light λ) and T-cell receptor (TRA/TRD: α/δ; TRB: β; TRG: γ) loci in representative cells. GM12878 exhibited clear clonal structure at immunoglobulin loci.

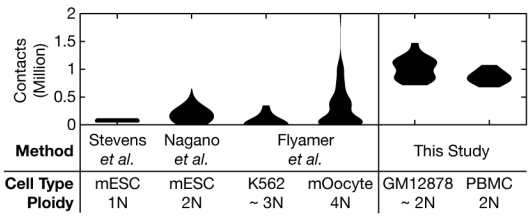



**Fig. S4.**

**Dip-C detected more chromatin contacts than previous methods.** Distribution of the number of contacts per cell for each method, as shown by a violin plot. Dip-C (right) detected more contacts than previous methods (left) (*8, 10, 11*), with medians at least 5 times as high. Kernel size was 0.1 million. Only G1-phase (according to *in sillico* phasing) cells were shown for (*10*). From left to right, median = 0.15, 0.17, 0.00094, 0.11, 1.04, and 0.84 million per 2N; and $n$ = 8, 750, 34, 120, 17, and 18.

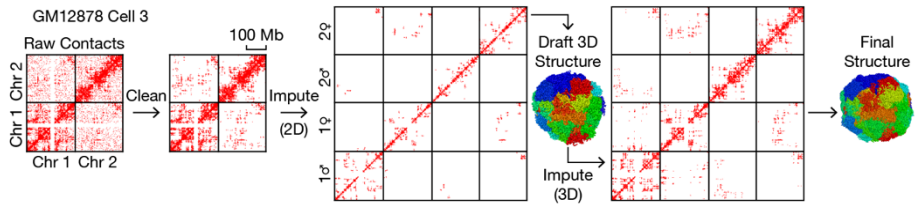



**Fig. S5.**

**Schematics of the Dip-C algorithm.** In the "dip-c" package, after removal of 3C/Hi-C artefacts ("clean") and initial imputation of chromosome haplotypes connected by each contact ("impute 2D"), haplotypes were further imputed iteratively through a series of draft 3D models ("impute 3D") before generating a final structure. For visual clarity, only the first two chromosomes are shown. In the "hickit" package, haplotype imputation does not involve draft 3D models.

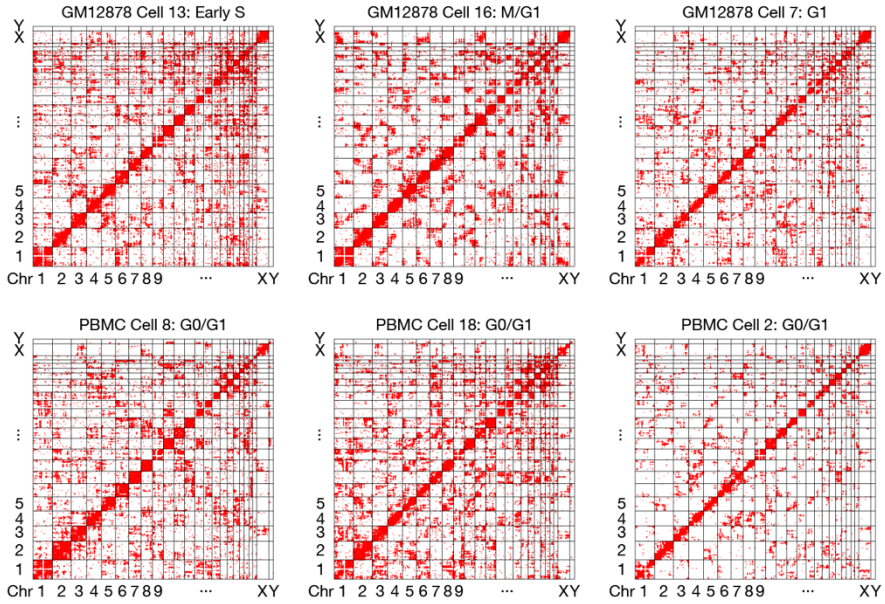



**Fig. S6.**

**Chromatin contacts in representative cells after artifact removal.** Similar to the left panel of **Fig. 1B** but with different cells. “G1”, “G0/G1”, “M/G1”, and “S” denoted cell-cycle phases.

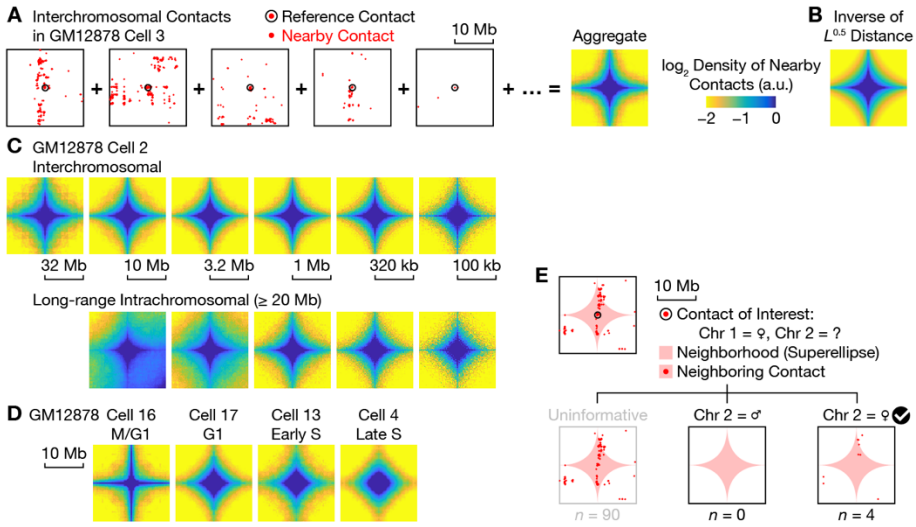



**Fig. S7.**

**Unknown haplotypes of chromatin contacts could be imputed based on a novel statistical property of interchromosomal and long-range intrachromosomal contacts. (A)** In a single cell, around each interchromosomal contact ("reference"; circled), nearby contacts (red dots) were more likely to occur in a superelliptical region. Horizontal and vertical axes represented genomic coordinates along the two contacting chromosomes. Density of nearby contacts (heatmap) was calculated by overlaying all interchromosomal contacts ($n$ = 201 k for the cell shown) at the origin and binning nearby contacts into 200-kb squares, and colored by setting the 0.95-quantile to blue. **(B)** An empirical formula where the density of nearby contacts was inversely proportional to the $L^{0.5}$ distance (not a norm) of the genomic distances from the origin. The contours of this distribution were superellipses with exponent = 0.5. **(C)** Similar density profiles were observed across different genomic scales for both interchromosomal (top row) and long-range intrachromosomal (bottom row) contacts. Bin sizes were scaled in proportion. For intrachromosomal contacts, the bottom right corner was closer to the diagonal and thus had more nearby contacts. Thin crosshairs in the rightmost panels might be an artifact caused by the fact that each reference contact (ligation of two sticky ends) also corresponded to two other sticky ends, which were more likely for form other contacts than uncut or unligated regions. **(D)** Density profiles varied systematically across the cell cycle. **(E)** Schematic of haplotype imputation. Missing haplotype of a given contact (circled) could be imputed from haplotypes of other contacts (red dots) in its superelliptical neighborhood (red shade).

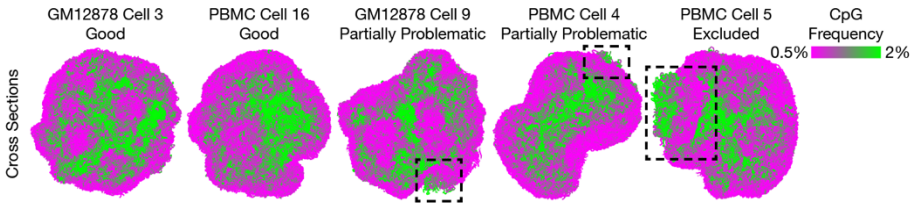



**Fig. S8.**

**Problematic regions in 3D structures were manually removed.** Cross sections of representative cells with euchromatin (green) and heterochromatin (magenta) visualized by CpG frequencies (*21*), similar to **Fig. 2C**. Some regions (dashed boxes) incorrectly positioned CpG-rich, euchromatic particles on the outside and were thus manually removed. Two PBMCs (Cells 5 and 15) harbored large problematic regions and were excluded.

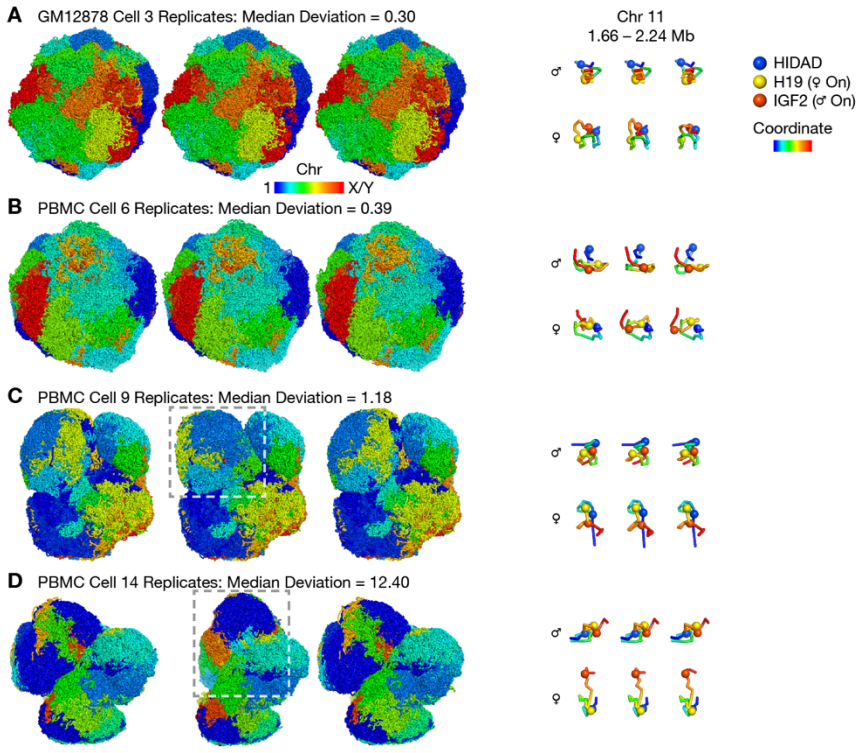



**Fig. S9.**

**Replicate structures estimated a lower bound for reconstruction uncertainty.** Each panel showed three replicate structures (left: whole genome; right: a 0.58-Mb region around the imprinted H19/IGF2 locus) in a representative cell. Deviations between replicates were apparently large for two multi-lobed presumable monocytes/neutrophils, as shown in (**C**) and (**D**), because some lobes (dashed boxes) could be rotated or mirrored with minimal perturbation on local contacts. Unit of deviations was particle radii.

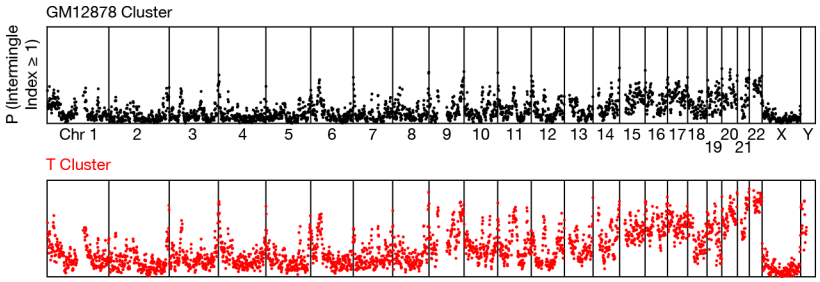



**Fig. S10.**

**Genomic regions that frequently contacted multiple chromosomes were similar between different cell types.** Similar to **Fig. S20**, but separately for two cell-type clusters. Despite a higher overall extent of chromosome intermingling in PBMCs, the genome-wide pattern remained generally similar to that of GM12878.

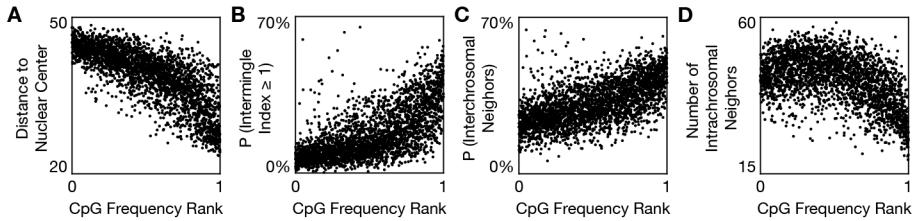



**Fig. S11.**

**Different chromatin compartments exhibited different radial preferences and different extents of chromosome intermingling.** For each 1-Mb genomic region, four properties (averaged across all cells) — **(A)** radial preference (same as **Fig. 2D**), **(B)** extent of multi-chromosome intermingling (same as **Fig. 2I**), **(C)** extent of chromosome intermingling (as quantified by the percentage of nearby particles that were not from the same chromosome), and **(D)** preference for the interior of a chromosome territory (as quantified by the number of nearby particles that were from the same chromosome) — were plotted against its CpG frequency (rank normalized to 0–1 across the genome). The CpG-rich, euchromatic compartment A tended to reside in the nuclear interior and on the surface of chromosome territories, leading to more frequent chromosome intermingling.

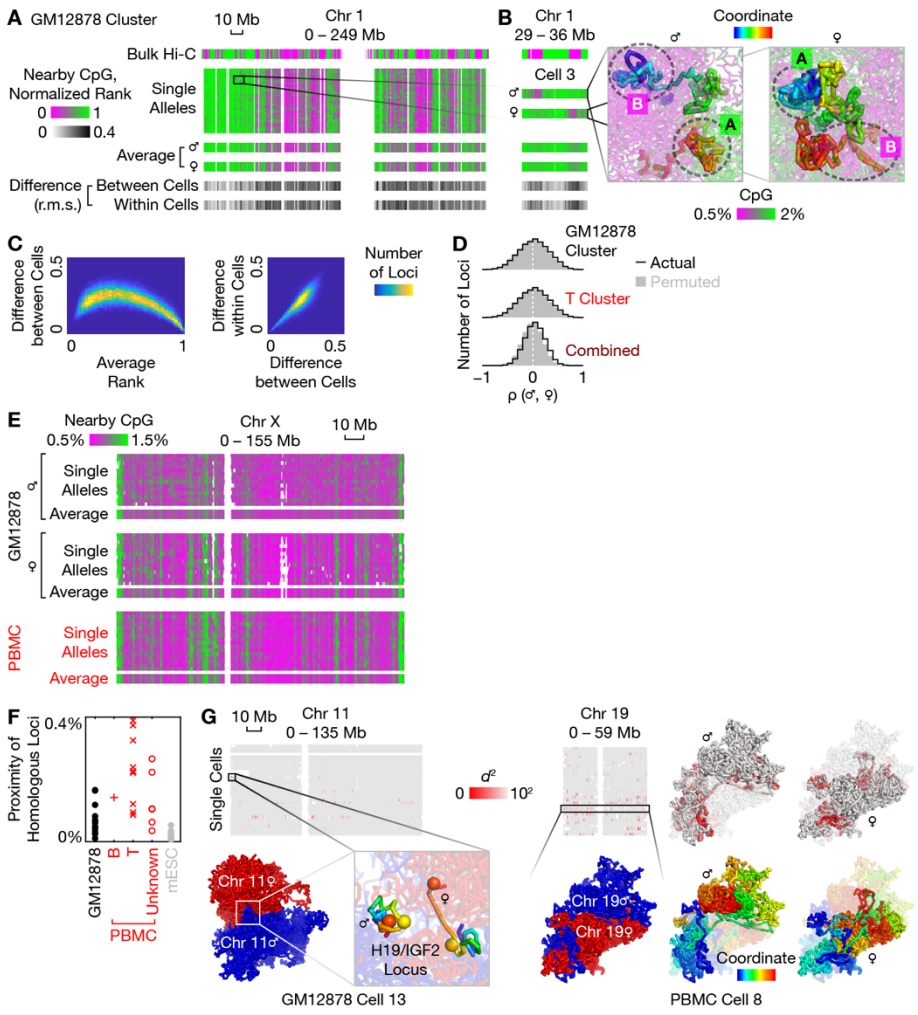



**Fig. S12.**

**Single-cell chromatin compartments of the two alleles varied almost independently in each cell, with the exceptions of X chromosomes and of rare somatic pairing. (A)** Decomposition of heterogeneity of single-cell compartments revealed a near-identical pattern between cells and between the two alleles in each cell ("within cells"), exemplified by chromosome 1. Single-cell compartment was defined for each 20-kb particle as the average CpG frequency of nearby particles (rank normalized to 0–1 in each cell). **(B)** An example region (thick rainbow sticks) that was differentially compartmentalized between the two homologs in a single cell. **(C)** 2D histograms between the average compartment and the average between-cell difference (left panel) and between the average between-cell and within-cell differences (right panel) for all 20-kb genomic loci. **(D)** Histogram (black lines) of Spearman's correlation between the maternal and paternal compartments for all 20-kb genomic loci. Permuted histograms (gray shade) were generated by randomly permuting cell labels for each locus. White dashed lines indicated zero correlation. The two cell-type clusters were defined as in **Fig. 4**. **(E)** Single-cell chromatin compartments of the female and male X chromosomes (not rank normalized). Active X chromosomes featured clear chromatin compartmentalization, while the average compartments of inactive X chromosomes were

more uniform along the chromosome. **(I)** The fraction of potential somatic pairing of homologous loci (distance ≤ 5 particle radii) among all 20-kb genomic bins. Each marker represented a single cell. **(J)** Two example pairs of intermingling homologs. Heatmaps and the top two structures were colored by distances from the homologous locus. Inset showed potential pairing of the imprinted H19/IGF2 locus, with the same spheres and colors as in **Fig. 3A**. GM12878 Cells 4 and 16 were excluded.

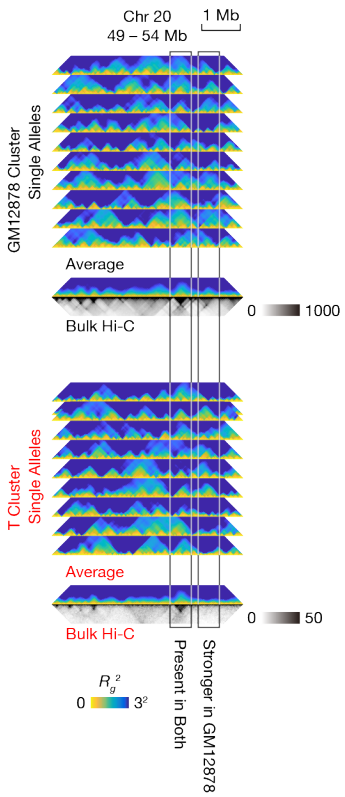
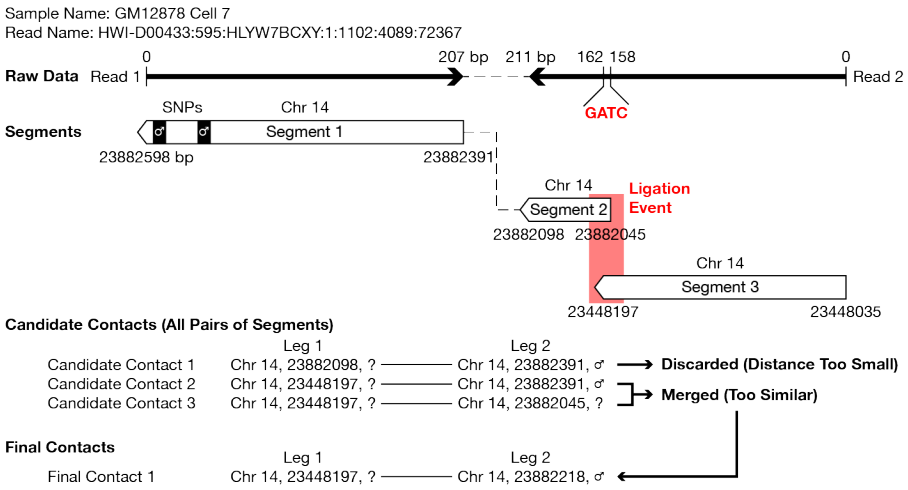



**Fig. S13.**

**The average domains of each cell type agreed well with bulk Hi-C from the same cell type.** Similar to **Fig. S19**, but calculated separately for two cell-type clusters. This 5-Mb region (chosen according to Fig. 2A of (*19*)) consisted of bulk domains that were common to both cell types, and a domain specific to GM12878. For visual clarity, only a random subset of 5 cells (10 alleles) were shown per cell-type cluster. Averages were r.m.s. of radii of gyration. GM12878 Cells 4 and 16 were excluded. Bulk Hi-C (black heatmap, with 25-kb bins) was from (*19, 40*).




## Fig. S14.

**Schematic of contact identification from sequencing reads.** After adapter trimming, each sequencing read or read pair ("raw data") was aligned to the human genome by BWA-MEM, which supports local alignment. All high-quality primary (segment 1 for read 1, and segment 3 for read 2) and supplementary (segment 2) alignments were extracted as "segments". If a segment overlapped with a phased SNP (black vertical lines), a haplotype (paternal for segment 1) would be assigned if base quality is high. Each end point of a segment is defined as a "leg". Chromatin contacts were identified as all pairs of segment end points (each pair being a "candidate contact") — that were separated by > 1 kb (thus excluding candidate contact 1). In this example, candidate contacts 2 and 3 were merged because of the similarity between their legs, yielding one "final contact".

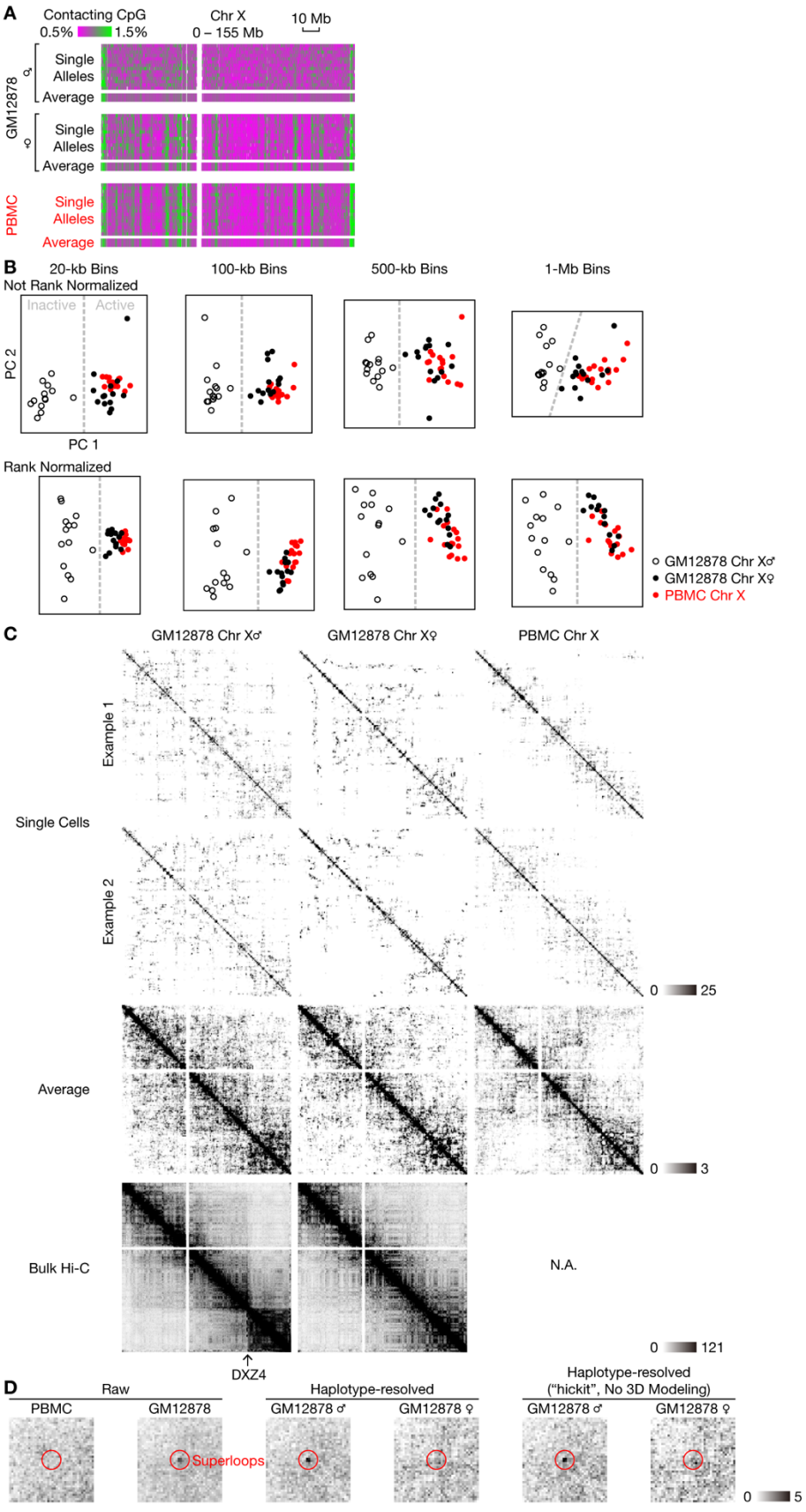

A
Contacting CpG
0.5%
1.5%
Chr X
0 – 155 Mb
10 Mb
GM12878
♂
Single Alleles
Average
♀
Single Alleles
Average
PBMC
Single Alleles
Average
B
20-kb Bins
100-kb Bins
500-kb Bins
1-Mb Bins
Not Rank Normalized
Inactive
Active
PC 2
PC 1
Rank Normalized
GM12878 Chr X♂
GM12878 Chr X♀
PBMC Chr X
C
GM12878 Chr X♂
GM12878 Chr X♀
PBMC Chr X
Example 1
Single Cells
Example 2
0
25
Average
0
3
Bulk Hi-C
N.A.
0
121
DXZ4
D
Raw
Haplotype-resolved
Haplotype-resolved
("hickit", No 3D Modeling)
PBMC
GM12878
GM12878 ♂
GM12878 ♀
GM12878 ♂
GM12878 ♀
Superloops
0
5

**Fig. S15.**

**Contact-based analysis of X chromosomes confirmed our 3D-structure-based results. (A)** Same as **Fig. S12E**, but with a contact-based definition of single-cell chromatin compartments ("contacting CpG"). Bin size was 100 kb. **(B)** Same as **Fig. 3C**, but with a contact-based definition of single-cell chromatin compartments. Bottom panels rank-normalized compartments to 0–1 for each chromosome before PCA. **(C)** Average haplotype-resolved contact maps of the inactive and active X chromosomes agreed well with bulk Hi-C. Each contact map was binned every 1 Mb and normalized by the mean (whose value was defined as one) of all bins. Bulk Hi-C (500-kb bins) was from Fig. 7D of (*19*). **(D)** Average contact maps aggregated (similar to **Fig. S8A**) from all "superloop" (*19*) confirmed that superloops only formed in the inactive X chromosome. Each map was binned every 100 kb and covered ± 2 Mb around the loop centers.

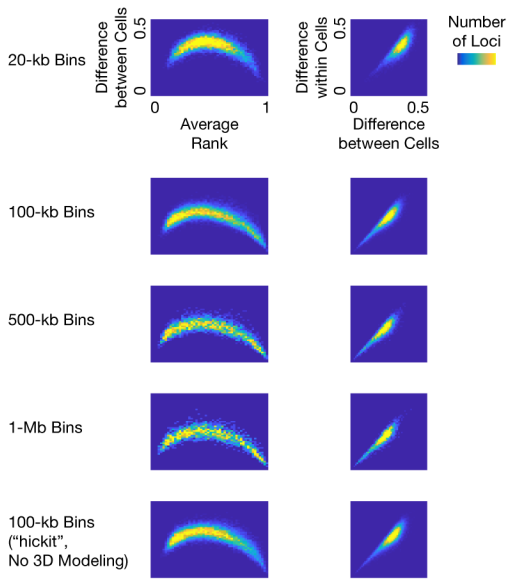



**Fig. S16.**

**Contact-based analysis of single-cell chromatin compartments of the two alleles confirmed our 3D-structure-based results.** Same as **Fig. S12C**, but with a contact-based definition of single-cell chromatin compartments and a series of bin sizes.

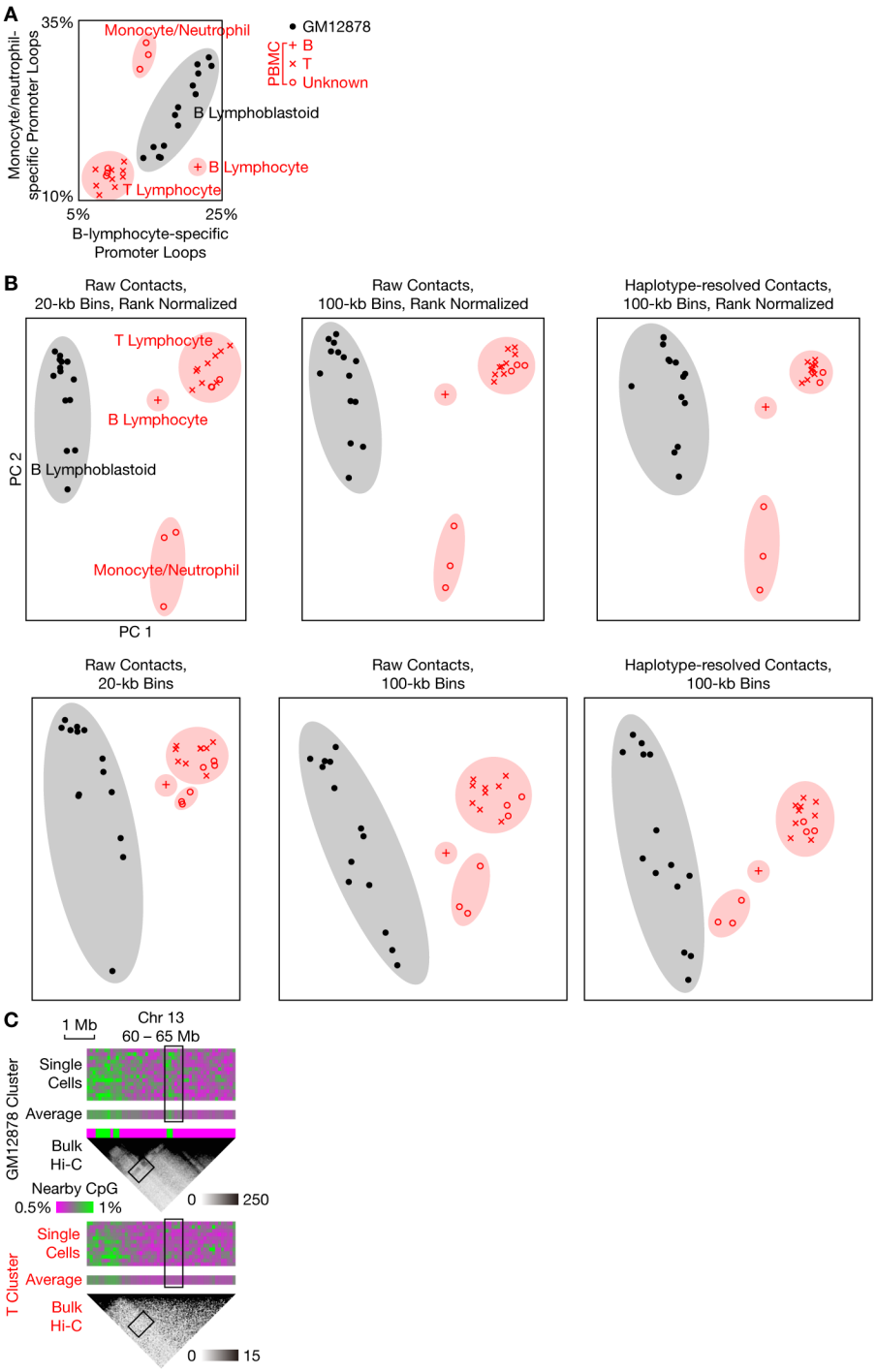



**Fig. S17.**

**Contact-based analysis of cell types confirmed our 3D-structure-based results. (A)** Same as **Fig. 4D**, but with a contact-based definition of promoter-enhancer loop formation. **(B)** Same as **Fig. 4E**, but with a contact-based definition of single-cell chromatin compartments. Top panels rank-normalized compartments to 0–1 for each cell before PCA. Rightmost panels used haplotype-resolved contacts instead of raw contacts.

**(C)** Same as **Fig. 4F**, but with a contact-based definition of single-cell chromatin compartments. Bin size was 100 kb.

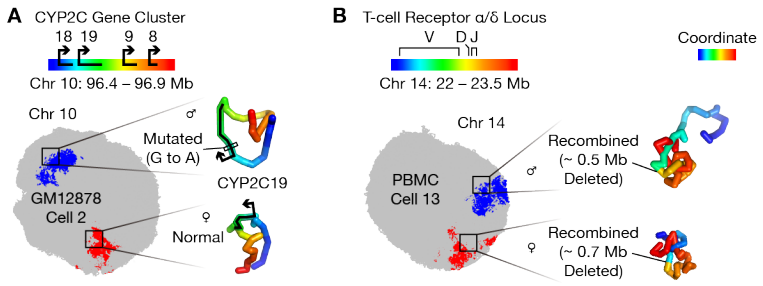



**Fig. S18.**

**Genomic changes, such as SNPs and CNVs, can be pinpointed to their precise spatial locations in the cell nucleus. (A)** 3D localization of a paternally inherited drug-response single-nucleotide mutation (rs4244285, G to A) in the CYP2C19 gene in a representative GM12878 cell. Arrows represented the directions of transcription of CYP2C genes. **(B)** 3D localization of two different somatic DNA deletions — results of V(D)J recombination — in the two alleles of the T-cell receptor α/δ locus in a representative T lymphocyte.

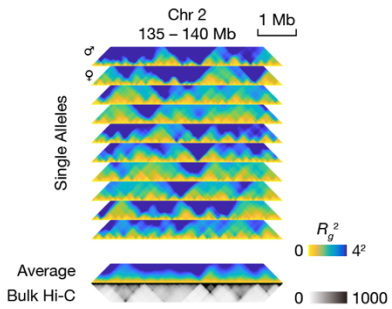



**Fig. S19.**

**Single-cell domains were highly heterogeneous between cells.** Matrices of radii of 5 representative cells (10 alleles in total) in a 5-Mb region (chosen according to Figure 2e of (*24*)). Averages were root-mean-square (r.m.s.). Bulk Hi-C (black heatmap, with 25-kb bins) was from (*19*). GM12878 Cells 4 and 16 were excluded.

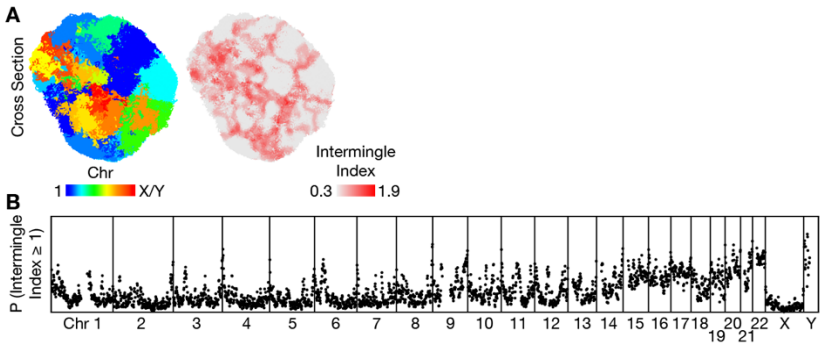



**Fig. S20.**

**Some genomic regions were enriched in multi-chromosome intermingling. (A)** The extent of multi-chromosome intermingling was quantified by the diversity of chromosomes, as measured by Shannon's index, with 3 particle radii of each 20-kb particle. **(B)** Probability of extensive multi-chromosome intermingling (smoothed by 1-Mb windows) across the human genome. Axis limits were 0 and 0.8. GM12878 Cells 4 and 16 were excluded.

**Table S1.**
Information about each single cell.

| cell | sex | sample | cell cycle phase | chromosomal abnormality | TRA/TRD | TRB | TRG | IGH |
|---|---|---|---|---|---|---|---|---|
| **gm12878_01** | female | GM12878 rep 1 | G1 | +10(pat), +16(pat) partial, +19(pat) | none | none | none | both (pattern 1) |
| **gm12878_02** | female | GM12878 rep 1 | G1 | none | none | none | none | both (pattern 1) |
| **gm12878_03** | female | GM12878 rep 1 | G1 | none | none | none | none | both (pattern 1) |
| **gm12878_04** | female | GM12878 rep 1 | late S | LOH to 1(mat), LOH to 2(pat), +4(pat), +11(mat) | none | none | none | both (pattern 2) |
| **gm12878_05** | female | GM12878 rep 1 | G1 | none | none | none | none | both (pattern 1) |
| **gm12878_06** | female | GM12878 rep 1 | G1 | none | none | none | none | both (pattern 1) |
| **gm12878_07** | female | GM12878 rep 1 | G1 | none | none | none | none | both (pattern 1) |
| **gm12878_08** | female | GM12878 rep 1 | damaged | damaged | damaged | damaged | damaged | both (pattern 1) |
| **gm12878_09** | female | GM12878 rep 2 | G1 | LOH to 16(mat) partial | none | none | none | both (pattern 1) |
| **gm12878_10** | female | GM12878 rep 2 | G1 | -2(pat) partial, -6(pat) | none | none | none | both (pattern 3) |
| **gm12878_11** | female | GM12878 rep 2 | G1 | +11(mat) partial, +15(pat) partial | none | none | none | both (pattern 3) |
| **gm12878_12** | female | GM12878 rep 2 | G1 | +11(mat) partial, +15(pat) partial | none | none | none | both (pattern 3) |
| **gm12878_13** | female | GM12878 rep 2 | early S | none | none | none | none | both (pattern 3) |
| **gm12878_14** | female | GM12878 rep 2 | G1 | +5(pat) partial, -5(pat) partial | none | none | none | both (pattern 3) |
| **gm12878_15** | female | GM12878 rep 2 | G1 | +2(pat) partial, -2(pat) partial | none | none | none | both (pattern 3) |
| **gm12878_16** | female | GM12878 rep 2 | M/G1 | none | none | none | none | both (pattern 1) |
| **gm12878_17** | female | GM12878 rep 2 | G1 | none | none | none | none | both (pattern 3) |
| **pbmc_01** | male | PBMC | G0/G1 | none | both | uncertain | both | none |
| **pbmc_02** | male | PBMC | G0/G1 | none | none | none | none | both |
| **pbmc_03** | male | PBMC | G0/G1 | none | none | none | none | none |
| **pbmc_04** | male | PBMC | G0/G1 | none | pat | none | both or pat | none |
| **pbmc_05** | male | PBMC | G0/G1 | none | both | both | both | none |
| **pbmc_06** | male | PBMC | G0/G1 | none | mat | mat | both or pat | none |
| **pbmc_07** | male | PBMC | G0/G1 | none | both | mat | both | none |
| **pbmc_08** | male | PBMC | G0/G1 | none | both | mat | both or mat | none |
| **pbmc_09** | male | PBMC | G0/G1 | none | none | none | none | none |
| **pbmc_10** | male | PBMC | G0/G1 | none | both | mat | both or pat | none |
| **pbmc_11** | male | PBMC | G0/G1 | none | none | none | none | none |
| **pbmc_12** | male | PBMC | G0/G1 | none | both | both | both | none |
| **pbmc_13** | male | PBMC | G0/G1 | none | both | mat | both or pat | none |
| **pbmc_14** | male | PBMC | G0/G1 | none | none | none | none | none |
| **pbmc_15** | male | PBMC | G0/G1 | none | both | both | both | none |
| **pbmc_16** | male | PBMC | G0/G1 | none | both | none | both or mat | none |
| **pbmc_17** | male | PBMC | G0/G1 | none | none | none | none | none |
| **pbmc_18** | male | PBMC | G0/G1 | none | none | none | none | none |
| **mESC_1CDX1-257** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-273** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-285** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-286** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-345** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-347** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-391** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-413** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-422** | female | mESC | G1 | none | none | none | none | none |
| **mESC_1CDX1-424** | female | mESC | G1 | none | none | none | none | none |

| IGK | IGL | IGTR summary | PCA-based cell type | total contacts | intra (%) | leg phased (%) | excluded regions for contacts |
|---|---|---|---|---|---|---|---|
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 838182 | 77.21 | 8.65 | 10, 16: 40M-end, 19 |
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 708428 | 78.92 | 8.41 | none |
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 859770 | 76.32 | 9.01 | none |
| uncertain (pattern 2) | none or mat or pat (pattern 2) | B | B lymphoblastoid | 1476944 | 78.35 | 9.61 | 1, 2, 4, 11 |
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 795998 | 76.59 | 9.42 | none |
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 809401 | 77.97 | 8.94 | none |
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 968514 | 74.13 | 9.42 | none |
| both (pattern 1) | both (pattern 1, damaged) | B | B lymphoblastoid | 812217 | 43.54 | 8.96 | damaged |
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 1038774 | 75.7 | 9.07 | 16: 40M-end |
| both (pattern 1) | both (pattern 3) | B | B lymphoblastoid | 907061 | 77.54 | 9.11 | none |
| both (pattern 1) | both (pattern 3) | B | B lymphoblastoid | 1261430 | 78.99 | 8.9 | 11: start-13.2M, 15: 62.2M-end |
| both (pattern 1) | both (pattern 3) | B | B lymphoblastoid | 1173076 | 75.73 | 8.98 | 11: start-13.2M, 15: 62.2M-end |
| both (pattern 1) | both (pattern 3) | B | B lymphoblastoid | 1171749 | 78.19 | 9.21 | none |
| both (pattern 1) | both (pattern 3) | B | B lymphoblastoid | 1126990 | 77.37 | 9.1 | 5: 165.8M-167.8M |
| both (pattern 1) | both (pattern 3) | B | B lymphoblastoid | 1154785 | 72.12 | 8.77 | 2: 238.5M-end |
| both (pattern 1) | both (pattern 1) | B | B lymphoblastoid | 1081805 | 80.53 | 9.11 | none |
| both (pattern 1) | both (pattern 3) | B | B lymphoblastoid | 1120600 | 73.86 | 8.73 | none |
| none | none | T | T lymphocyte | 839010 | 71.75 | 8.9 | none |
| pat | pat | B | B lymphocyte | 1021315 | 66.43 | 8.78 | none |
| none | none | other | T lymphocyte | 843536 | 60.57 | 8.94 | none |
| none | none | T | T lymphocyte | 881132 | 56.46 | 8.89 | none |
| none | none | T | T lymphocyte | 766964 | 67.4 | 9.17 | none |
| none | none | T | T lymphocyte | 872713 | 69.55 | 8.9 | none |
| none | none | T | T lymphocyte | 925643 | 65.08 | 8.83 | none |
| none | none | T | T lymphocyte | 831457 | 66.91 | 8.83 | none |
| none | none | other | monocyte/neutrophil | 991167 | 66.25 | 9.01 | none |
| none | none | T | T lymphocyte | 814616 | 66.81 | 8.74 | none |
| none | none | other | T lymphocyte | 761342 | 64.21 | 8.67 | none |
| none | none | T | T lymphocyte | 765000 | 69.06 | 8.62 | none |
| none | none | T | T lymphocyte | 785919 | 62.16 | 8.66 | none |
| none | none | other | monocyte/neutrophil | 1080640 | 68.63 | 8.47 | none |
| none | none | T | T lymphocyte | 792928 | 70.72 | 8.71 | none |
| none | none | T | T lymphocyte | 670789 | 66.73 | 9.06 | none |
| none | none | other | T lymphocyte | 874453 | 58.95 | 8.8 | none |
| none | none | other | monocyte/neutrophil | 997363 | 61.71 | 8.67 | none |
| none | none | none | none | 200993 | 93.24 | 52.03 | none |
| none | none | none | none | 248989 | 91.59 | 51.54 | none |
| none | none | none | none | 229453 | 90.11 | 51.42 | none |
| none | none | none | none | 221295 | 78.38 | 51.7 | none |
| none | none | none | none | 242679 | 91.81 | 51.67 | none |
| none | none | none | none | 220431 | 90.75 | 51.58 | none |
| none | none | none | none | 225701 | 87.88 | 51.46 | none |
| none | none | none | none | 284482 | 92.27 | 51.7 | none |
| none | none | none | none | 225828 | 91.17 | 51.51 | none |
| none | none | none | none | 224672 | 94.34 | 51.34 | none |

| haploid regions | parameters | total contacts after exclusion | cleaned contacts | cleaned contacts (%) | imputed contacts (initial) |
|---|---|---|---|---|---|
| none | default | 700073 | 657377 | 93.90 | 154223 |
| none | default | 708428 | 662037 | 93.45 | 145332 |
| none | default | 859770 | 802576 | 93.35 | 203672 |
| none | default | 953266 | 910625 | 95.53 | 241717 |
| none | default | 795998 | 748508 | 94.03 | 194899 |
| none | default | 809401 | 753158 | 93.05 | 179472 |
| none | default | 968514 | 899189 | 92.84 | 254875 |
| damaged | damaged | damaged | damaged | damaged | damaged |
| none | default | 1016132 | 943049 | 92.81 | 246638 |
| 2(mat): 232.8M-end, 6(mat) | default | 907061 | 825166 | 90.97 | 209778 |
| none | default | 1212817 | 1126475 | 92.88 | 249817 |
| none | default | 1126871 | 1029942 | 91.40 | 252743 |
| none | default | 1171749 | 1065539 | 90.94 | 252611 |
| 5(mat): 167.8M-end | default | 1125570 | 1026524 | 91.20 | 249700 |
| none | default | 1151010 | 1055182 | 91.67 | 292978 |
| none | default | 1081805 | 992536 | 91.75 | 179077 |
| none | default | 1120600 | 1026364 | 91.59 | 271214 |
| none | default | 839010 | 749242 | 89.30 | 200591 |
| none | default | 1021315 | 925232 | 90.59 | 295445 |
| none | default | 843536 | 759194 | 90.00 | 259802 |
| none | default | 881132 | 791556 | 89.83 | 291237 |
| none | default | 766964 | 685878 | 89.43 | 209674 |
| none | default | 872713 | 778976 | 89.26 | 214875 |
| none | default | 925643 | 823505 | 88.97 | 253974 |
| none | default | 831457 | 742170 | 89.26 | 224766 |
| none | default | 991167 | 910034 | 91.81 | 310795 |
| none | default | 814616 | 733695 | 90.07 | 220700 |
| none | default | 761342 | 683784 | 89.81 | 213646 |
| none | default | 765000 | 680583 | 88.97 | 187373 |
| none | default | 785919 | 703694 | 89.54 | 228671 |
| none | default | 1080640 | 995614 | 92.13 | 317701 |
| none | default | 792928 | 713076 | 89.93 | 197087 |
| none | default | 670789 | 596832 | 88.97 | 179069 |
| none | default | 874453 | 796327 | 91.07 | 294657 |
| none | default | 997363 | 909649 | 91.21 | 317057 |
| none | clean -c1; impute -C1; impute3 -C1 | 200993 | 198791 | 98.90 | 139449 |
| none | clean -c1; impute -C1; impute3 -C1 | 248989 | 246875 | 99.15 | 181170 |
| none | clean -c1; impute -C1; impute3 -C1 | 229453 | 226964 | 98.92 | 163831 |
| none | clean -c1; impute -C1; impute3 -C1 | 221295 | 218671 | 98.81 | 159879 |
| none | clean -c1; impute -C1; impute3 -C1 | 242679 | 240327 | 99.03 | 168863 |
| none | clean -c1; impute -C1; impute3 -C1 | 220431 | 217994 | 98.89 | 157004 |
| none | clean -c1; impute -C1; impute3 -C1 | 225701 | 223479 | 99.02 | 163133 |
| none | clean -c1; impute -C1; impute3 -C1 | 284482 | 282173 | 99.19 | 206144 |
| none | clean -c1; impute -C1; impute3 -C1 | 225828 | 223698 | 99.06 | 161741 |
| none | clean -c1; impute -C1; impute3 -C1 | 224672 | 222410 | 98.99 | 155895 |

| imputed contacts (initial, %) | CV accuracy (initial, %) | imputed contacts (final) | imputed contacts (final, %) | CV accuracy (final, %) | median deviation | excluded replicates |
|---|---|---|---|---|---|---|
| 22.03 | 98.78 | 430374 | 61.48 | 96.08 | 0.39 | none |
| 20.51 | 98.44 | 444720 | 62.78 | 96.72 | 0.51 | none |
| 23.69 | 98.63 | 554347 | 64.48 | 96.94 | 0.30 | none |
| 25.36 | 98.65 | 496285 | 52.06 | 89.31 | 0.65 | none |
| 24.48 | 98.75 | 522713 | 65.67 | 97.00 | 0.34 | none |
| 22.17 | 98.91 | 499581 | 61.72 | 96.89 | 0.43 | none |
| 26.32 | 99.02 | 592398 | 61.17 | 96.38 | 0.32 | none |
| damaged | damaged | damaged | damaged | damaged | damaged | damaged |
| 24.27 | 98.57 | 625981 | 61.60 | 96.41 | 0.53 | none |
| 23.13 | 98.84 | 499853 | 55.11 | 95.67 | 0.42 | none |
| 20.60 | 98.66 | 743628 | 61.31 | 96.31 | 0.39 | none |
| 22.43 | 98.76 | 599673 | 53.22 | 94.90 | 0.45 | none |
| 21.56 | 98.40 | 576953 | 49.24 | 93.26 | 0.58 | none |
| 22.18 | 98.29 | 564282 | 50.13 | 94.20 | 0.60 | none |
| 25.45 | 98.83 | 658848 | 57.24 | 96.02 | 0.26 | none |
| 16.55 | 98.76 | 707742 | 65.42 | 96.95 | 0.29 | none |
| 24.20 | 98.69 | 630322 | 56.25 | 95.57 | 0.35 | none |
| 23.91 | not tested | 510331 | 60.83 | not tested | 0.43 | none |
| 28.93 | not tested | 689371 | 67.50 | not tested | 0.33 | none |
| 30.80 | not tested | 563589 | 66.81 | not tested | 0.24 | none |
| 33.05 | not tested | 577826 | 65.58 | not tested | 0.32 | none |
| 27.34 | not tested | 490298 | 63.93 | not tested | 1.35 | none |
| 24.62 | not tested | 538701 | 61.73 | not tested | 0.39 | none |
| 27.44 | not tested | 574846 | 62.10 | not tested | 0.33 | none |
| 27.03 | not tested | 531281 | 63.90 | not tested | 0.64 | none |
| 31.36 | not tested | 685376 | 69.15 | not tested | 1.18 | none |
| 27.09 | not tested | 517702 | 63.55 | not tested | 0.31 | none |
| 28.06 | not tested | 494200 | 64.91 | not tested | 0.41 | none |
| 24.49 | not tested | 479411 | 62.67 | not tested | 0.45 | none |
| 29.10 | not tested | 517966 | 65.91 | not tested | 0.32 | none |
| 29.40 | not tested | 745126 | 68.95 | not tested | 12.40 | none |
| 24.86 | not tested | 497934 | 62.80 | not tested | 2.59 | none |
| 26.70 | not tested | 424817 | 63.33 | not tested | 0.45 | none |
| 33.70 | not tested | 593079 | 67.82 | not tested | 0.24 | none |
| 31.79 | not tested | 680158 | 68.20 | not tested | 0.30 | none |
| 69.38 | not tested | 172916 | 86.03 | not tested | 1.35 | none |
| 72.76 | not tested | 214758 | 86.25 | not tested | 11.13 | replicate 0 |
| 71.40 | not tested | 197087 | 85.89 | not tested | 0.92 | none |
| 72.25 | not tested | 194350 | 87.82 | not tested | 1.17 | none |
| 69.58 | not tested | 209399 | 86.29 | not tested | 1.07 | none |
| 71.23 | not tested | 188888 | 85.69 | not tested | 5.77 | replicate 0 |
| 72.28 | not tested | 195150 | 86.46 | not tested | 4.69 | replicate 1 |
| 72.46 | not tested | 246023 | 86.48 | not tested | 1.64 | none |
| 71.62 | not tested | 193496 | 85.68 | not tested | 18.95 | replicates 0, 1, 4 |
| 69.39 | not tested | 193403 | 86.08 | not tested | 1.34 | none |

| median deviation after exclusion | problematic regions in structures |
|---|---|
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| damaged | damaged |
| none | 16(mat), 18(mat) |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | 12(mat): 56M-64M |
| none | 4(mat): start-4M, 6(pat): 24M-44M, 6(mat): 26M-27M, 16(mat): 11M-13M and 29M-31M, 20(mat): 30M-37M |
| none | none |
| none | 16(mat): 69M-76M |
| none | poor structure |
| none | none |
| none | none |
| none | 3(mat): start-75M |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | none |
| none | poor structure |
| none | none |
| none | none |
| none | none |
| none | none |
| 2.20 | none |
| none | none |
| none | none |
| none | none |
| 1.23 | none |
| 1.00 | none |
| none | none |
| 1.26 | none |
| none | none |

**Table S2.**
Information about each library.

| sample | cell | beads | reads | read length (bp) | raw output (Gb) | raw contacts | contacts per read |
|---|---|---|---|---|---|---|---|
| PBMC | pbmc_01 | 0.55X | 29516698 | 2 x 250 | 14.76 | 2353096 | 8.0% |
| PBMC | pbmc_02 | 0.55X | 30304558 | 2 x 250 | 15.15 | 2922553 | 9.6% |
| PBMC | pbmc_03 | 0.55X | 29591906 | 2 x 250 | 14.80 | 2361087 | 8.0% |
| PBMC | pbmc_04 | 0.55X | 30790567 | 2 x 250 | 15.40 | 2579121 | 8.4% |
| PBMC | pbmc_05 | 0.55X | 27628023 | 2 x 250 | 13.81 | 2344667 | 8.5% |
| PBMC | pbmc_06 | 0.55X | 31475715 | 2 x 250 | 15.74 | 2587478 | 8.2% |
| PBMC | pbmc_07 | 0.55X | 32898011 | 2 x 250 | 16.45 | 2753001 | 8.4% |
| PBMC | pbmc_08 | 0.55X | 29187722 | 2 x 250 | 14.59 | 2322152 | 8.0% |
| PBMC | pbmc_09 | 0.55X | 24508234 | 2 x 250 | 12.25 | 2781493 | 11.3% |
| PBMC | pbmc_10 | 0.55X | 27392087 | 2 x 250 | 13.70 | 2112805 | 7.7% |
| PBMC | pbmc_11 | 0.55X | 25315629 | 2 x 250 | 12.66 | 1914559 | 7.6% |
| PBMC | pbmc_12 | 0.55X | 26353795 | 2 x 250 | 13.18 | 1957398 | 7.4% |
| PBMC | pbmc_13 | 0.55X | 26540576 | 2 x 250 | 13.27 | 2038172 | 7.7% |
| PBMC | pbmc_14 | 0.55X | 25808744 | 2 x 250 | 12.90 | 2719614 | 10.5% |
| PBMC | pbmc_15 | 0.55X | 25792691 | 2 x 250 | 12.90 | 2057137 | 8.0% |
| PBMC | pbmc_16 | 0.55X | 24152154 | 2 x 250 | 12.08 | 1922204 | 8.0% |
| PBMC | pbmc_17 | 0.55X | 26331787 | 2 x 250 | 13.17 | 2311835 | 8.8% |
| PBMC | pbmc_18 | 0.55X | 26954134 | 2 x 250 | 13.48 | 2650581 | 9.8% |
| PBMC | pbmc_01 | 0.70X | 56432294 | 2 x 250 | 28.22 | 2045767 | 3.6% |
| PBMC | pbmc_02 | 0.70X | 65667842 | 2 x 250 | 32.83 | 2841207 | 4.3% |
| PBMC | pbmc_03 | 0.70X | 56166561 | 2 x 250 | 28.08 | 2048560 | 3.6% |
| PBMC | pbmc_04 | 0.70X | 59684323 | 2 x 250 | 29.84 | 2279598 | 3.8% |
| PBMC | pbmc_05 | 0.70X | 41868827 | 2 x 250 | 20.93 | 1704111 | 4.1% |
| PBMC | pbmc_06 | 0.70X | 61265416 | 2 x 250 | 30.63 | 2303785 | 3.8% |
| PBMC | pbmc_07 | 0.70X | 66734534 | 2 x 250 | 33.37 | 2541488 | 3.8% |
| PBMC | pbmc_08 | 0.70X | 56744464 | 2 x 250 | 28.37 | 2054706 | 3.6% |
| PBMC | pbmc_09 | 0.70X | 44122759 | 2 x 250 | 22.06 | 2257049 | 5.1% |
| PBMC | pbmc_10 | 0.70X | 66402649 | 2 x 250 | 33.20 | 2173000 | 3.3% |
| PBMC | pbmc_11 | 0.70X | 58375843 | 2 x 250 | 29.19 | 2008037 | 3.4% |
| PBMC | pbmc_12 | 0.70X | 61526305 | 2 x 250 | 30.76 | 2098748 | 3.4% |
| PBMC | pbmc_13 | 0.70X | 62449294 | 2 x 250 | 31.22 | 2211650 | 3.5% |
| PBMC | pbmc_14 | 0.70X | 68370847 | 2 x 250 | 34.19 | 3233052 | 4.7% |
| PBMC | pbmc_15 | 0.70X | 57855786 | 2 x 250 | 28.93 | 2106934 | 3.6% |
| PBMC | pbmc_16 | 0.70X | 41953169 | 2 x 250 | 20.98 | 1573140 | 3.7% |
| PBMC | pbmc_17 | 0.70X | 56894178 | 2 x 250 | 28.45 | 2268202 | 4.0% |
| PBMC | pbmc_18 | 0.70X | 65057025 | 2 x 250 | 32.53 | 2906330 | 4.5% |
| GM12878 rep2 | gm12878_10 | 0.55X | 30645718 | 2 x 250 | 15.32 | 3465868 | 11.3% |
| GM12878 rep2 | gm12878_11 | 0.55X | 30466991 | 2 x 250 | 15.23 | 4461519 | 14.6% |
| GM12878 rep2 | gm12878_12 | 0.55X | 36904284 | 2 x 250 | 18.45 | 4908798 | 13.3% |
| GM12878 rep2 | gm12878_13 | 0.55X | 36720085 | 2 x 250 | 18.36 | 4985112 | 13.6% |
| GM12878 rep2 | gm12878_14 | 0.55X | 34506334 | 2 x 250 | 17.25 | 4679830 | 13.6% |
| GM12878 rep2 | gm12878_15 | 0.55X | 36009660 | 2 x 250 | 18.00 | 4419475 | 12.3% |
| GM12878 rep2 | gm12878_16 | 0.55X | 34028329 | 2 x 250 | 17.01 | 4629410 | 13.6% |
| GM12878 rep2 | gm12878_17 | 0.55X | 33847717 | 2 x 250 | 16.92 | 4110830 | 12.1% |
| GM12878 rep2 | gm12878_10 | 0.80-0.55X | 36613453 | 2 x 250 | 18.31 | 1849343 | 5.1% |
| GM12878 rep2 | gm12878_11 | 0.80-0.55X | 39661308 | 2 x 250 | 19.83 | 2542454 | 6.4% |
| GM12878 rep2 | gm12878_12 | 0.80-0.55X | 45827485 | 2 x 250 | 22.91 | 2652325 | 5.8% |
| GM12878 rep2 | gm12878_13 | 0.80-0.55X | 38657611 | 2 x 250 | 19.33 | 2320584 | 6.0% |
| GM12878 rep2 | gm12878_14 | 0.80-0.55X | 38478646 | 2 x 250 | 19.24 | 2300650 | 6.0% |
| GM12878 rep2 | gm12878_15 | 0.80-0.55X | 49779727 | 2 x 250 | 24.89 | 2690969 | 5.4% |
| GM12878 rep2 | gm12878_16 | 0.80-0.55X | 40286934 | 2 x 250 | 20.14 | 2351932 | 5.8% |
| GM12878 rep2 | gm12878_17 | 0.80-0.55X | 50279929 | 2 x 250 | 25.14 | 2597339 | 5.2% |
| GM12878 rep1 | gm12878_01 | 0.65X | 20077063 | 2 x 250 | 10.04 | 1723312 | 8.6% |
| GM12878 rep1 | gm12878_02 | 0.65X | 21243002 | 2 x 250 | 10.62 | 1446834 | 6.8% |
| GM12878 rep1 | gm12878_03 | 0.65X | 20596574 | 2 x 250 | 10.30 | 1788802 | 8.7% |
| GM12878 rep1 | gm12878_04 | 0.65X | 31292591 | 2 x 250 | 15.65 | 3262654 | 10.4% |
| GM12878 rep1 | gm12878_05 | 0.65X | 19329405 | 2 x 250 | 9.66 | 1796686 | 9.3% |
| GM12878 rep1 | gm12878_06 | 0.65X | 20472346 | 2 x 250 | 10.24 | 1781561 | 8.7% |
| GM12878 rep1 | gm12878_07 | 0.65X | 31995455 | 2 x 250 | 16.00 | 3081657 | 9.6% |
| GM12878 rep1 | gm12878_08 | 0.65X | 37797562 | 2 x 250 | 18.90 | 3544659 | 9.4% |
| GM12878 rep1 | gm12878_09 | 0.65X | 38396878 | 2 x 250 | 19.20 | 3388005 | 8.8% |
| mESC | 1CDX1-257 | unknown | 1658048 | 2 x 151 | 0.50 | 718661 | 43.3% |
| mESC | 1CDX1-273 | unknown | 1956038 | 2 x 151 | 0.59 | 919642 | 47.0% |
| mESC | 1CDX1-285 | unknown | 1517119 | 2 x 151 | 0.46 | 712780 | 47.0% |
| mESC | 1CDX1-286 | unknown | 1462610 | 2 x 151 | 0.44 | 700254 | 47.9% |
| mESC | 1CDX1-345 | unknown | 1854034 | 2 x 151 | 0.56 | 877727 | 47.3% |
| mESC | 1CDX1-347 | unknown | 1679331 | 2 x 151 | 0.51 | 791139 | 47.1% |
| mESC | 1CDX1-391 | unknown | 1421212 | 2 x 151 | 0.43 | 660049 | 46.4% |
| mESC | 1CDX1-413 | unknown | 1740091 | 2 x 151 | 0.53 | 830237 | 47.7% |
| mESC | 1CDX1-422 | unknown | 1296957 | 2 x 151 | 0.39 | 606752 | 46.8% |
| mESC | 1CDX1-424 | unknown | 1415864 | 2 x 151 | 0.43 | 660682 | 46.7% |