

Supplementary Materials for

Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI)

Chongyi Chen, Dong Xing, Longzhi Tan, Heng Li, Guangyu Zhou, Lei Huang, X. Sunney Xie*

*Corresponding author. Email: xie@chemistry.harvard.edu

Published 14 April 2017, *Science* **356**, 189 (2017)
DOI: 10.1126/science.aak9787

This PDF file includes:

Materials and Methods
Figs. S1 to S16
Table S1
References

Materials and Methods

Single cell preparation

We chose a diploid somatic cell line, BJ human skin fibroblast (ATCC CRL-2522). The lack of aneuploidy in this cell line makes it an ideal system to characterize the performance of whole genome amplification methods. BJ cells were cultivated in Eagle's minimum essential medium (ATCC) supplemented with 10% fetal bovine serum, 100 I.U./mL penicillin and 100 µg/mL streptomycin in 5% CO₂ 37°C incubator. Upon harvest, BJ cells were trypsinized and washed in 1X PBS several times, followed by resuspension into 1X PBS. Single BJ cells were then mouth-pipetted or FACS-sorted into PCR tubes (Maximum Recovery, Axygen) containing 3 µL lysis buffer (60 mM Tris-Ac pH 8.3, 2 mM EDTA pH 8.0, 15 mM DTT, 0.5 µM carrier ssDNA (5'-TCAGGTTTCCTGAA-3', IDT oligo with standard desalting)), spun down and incubated at 75°C for 30 min. 0.5 µL 4 mg/mL QIAGEN protease (dissolved in water and stored at 4°C) was then added, spun down, and incubated at 55°C for 4 hrs followed by 75°C for 30 min. The resulting single-cell lysate in PCR tubes could be subject to LIANTI amplification immediately, or stored in a freezer for later use.

LIAINTI transposome preparation

LIAINTI transposon DNA (5'/Phos/CTGTCTCTTACACATCTAACAGAACATTAAATCGACTCACTATA

GGGAGATGTGTATAAGAGACAG-3', IDT oligo with PAGE purification) was annealed by gradual cooling in annealing buffer (20 mM Tris-Ac pH 8.3, 50 mM NaCl, 2 mM EDTA pH 8.0) into a 1.5 µM self-looping structure. The 1.5 µM annealed LIAINTI transposon DNA was then mixed with an equal volume of ~1 µM Tn5 transposase (Epicentre, transfer to Lucigen starting Jan 1, 2017, catalog number TNP92110) and incubated at room temperature for 30 min, dimerizing into LIAINTI transposome with a final concentration of ~0.25 µM. The LIAINTI transposome can be stored at -20°C for a long time, and each single-cell LIAINTI amplification needs ~0.5 µL transposome.

Alternatively, to save the cost of the expensive Tn5 transposase purchased from Epicentre or Lucigen (~\$10 of transposase per cell for LIAINTI at the current price offered by Lucigen), one can also express and purify the Tn5 transposase in house from the plasmid pTXB1-Tn5 (Addgene), following the published protocol (48).

It is worth to note that a transposon typically means a single piece of DNA that can loop and insert itself into the genome with the help of transposase. Here our LIAINTI transposon is an engineered form of "pre-cut transposon" containing two separate DNA pieces. After the binding

of LIANTI transposon with Tn5 transposase, a LIANTI transposome dimer forms containing two transposase monomers and two DNA pieces (Fig. 1B).

Single-cell WGA by Linear Amplification via Transposon Insertion (LIANTI)

Starting with the single-cell lysate, a 20 μ L transposition mixture was assembled in the buffer containing 2.5 mM MgCl₂ and 6.25 nM LIANTI transpososome. The transposition reaction was carried out at 55°C for 12 min, followed by transposase removal by 68°C for 30 min in the presence of 2.5 mM EDTA in a total volume of ~21 μ L. Both ends of the transposed fragments were then filled and extended by 0.02 U/ μ L Q5 DNA polymerase (New England Biolabs) at 73°C for 30 seconds in the presence of 2 mM MgCl and 200 μ M dNTPs, in a total volume of ~23 μ L without the supplement of specific Q5 reaction buffer. Q5 DNA polymerase was then inactivated by 0.1 mg/mL QIAGEN protease 50°C incubation for 1 hour in the presence of 2.5 mM EDTA, followed by protease heat inactivation by 77°C for 20 min in the presence of 450 mM NaCl in a total volume of ~25 μ L. The transposed genomic DNA fragments were then diluted by nuclease-free water (Life Technologies) to assemble a 90 μ L T7 in vitro transcription reaction (1X RNAPol Reaction Buffer, New England Biolabs), together with 1.8 mM each NTPs, 10% DMSO, 10 mM DTT, 0.4 μ M carrier ssDNA, 0.67 U/ μ L SUPERase In RNase Inhibitor (Life Technologies) and 2 U/ μ L T7 RNA polymerase (Epicentre or New England Biolabs), and incubated at 37°C overnight for 10 - 16 hours.

After overnight incubation, RNAs transcribed from the whole genome were column purified (Zymo Research, R1015) and eluted into 18 μ L 0.1X TE. Following denaturation and ice quench, a 30 μ L reverse transcription reaction was assembled including 0.67 mM each dNTPs, 0.6 U/ μ L SUPERase In RNase Inhibitor (Life Technologies) and 6 U/ μ L SuperScript IV Reserve Transcriptase (Life Technologies) in SuperScript IV buffer provided by the manufacturer, and the incubation program was 55°C 15 min, 60°C 10 min, 65°C 12 min, 70°C 8 min, 75°C 5 min, 80°C 10 min. RNA was then removed by incubation at 37°C for 30 min with 10 ng/ μ L affinity-purified RNase A (Life Technologies, AM2270) and 0.08 U/ μ L RNase H (New England Biolabs). Second strand synthesis was carried out in a 100 μ L Q5 DNA polymerase system (New England Biolabs, 1X Q5 reaction buffer, 1X Q5 High GC enhancer, 200 μ M dNTPs, 0.5 μ M primer, 0.02 U/ μ L Q5 DNA polymerase), with the primer (5'-NNNNNNNNGGGAGATGTGTATAAGAGACAG-3', IDT oligo with PAGE purification). The incubation program was 98°C 40 s, 58°C 30 s, 60°C 30 s, 65°C 30 s, 70°C 30 s, 72°C 6 min. The resulted LIANTI amplicons were column purified into 20 μ L elution buffer (Zymo Research), and can be stored at -20°C for a long time.

From a single human cell with ~ 6 pg of genomic DNA, we routinely acquire ~20 ng DNA of LIANTI amplicons. For the negative control assay where LIANTI was performed without any

cell/DNA to start with, no amplification product was detected, suggesting negligible contamination from LIANTI reagents and procedure.

Upon library preparation, LIANTI amplicons were subject to conventional sonication process (Covaris S2) to the length required by sequencing platforms. The parameters used for sonication were duty cycle 10%, intensity 5, burst 200, and applied 2 cycles with a total of 80 s. The final insert size was suitable for 2X125 bp or 2X150 bp pair-end Illumina sequencing. Library preparation was performed by NEBNext Ultra or NEBNext Ultra II DNA Library Prep Kit for Illumimna (New England Biolabs), following the instructions of the manufacturer and skipping the optional size selection step.

Molecular features of LIANTI

Due to the intrinsic symmetry of transposome dimer, genomic DNA fragments created by transposon insertion are always attached by the same sequence on both ends, leading to the failure of PCR amplification due to fragment self-looping. In previous applications involving Tn5 transposition, a mixture of two types of transposon DNA were used to circumvent this problem. By chance, half of the DNA fragments will be tagged by different sequences on both ends, enabling PCR amplification, while the other half are still tagged by the same sequence on both ends, preventing PCR amplification. As a result, 50% of the originating DNA fragments are intrinsically lost due to transposome symmetry, making previous methods (49, 50) difficult for single-cell whole genome amplification.

To break the symmetry without losing half of the DNA fragments before amplification, we designed a single-stranded loop in the transposon DNA for LIANTI (Fig. 1B). Genomic DNA fragments after transposon insertion are symmetric with the same sequence on both ends containing T7 promoters. Unlike PCR, in vitro transcription is not affected by the intrinsic symmetry of double-stranded DNA templates, and instead creates asymmetry in the linearly amplified single-stranded RNAs by only transcribing regions downstream of the T7 promoter. However, the 19-bp transposase binding sites remain on both ends, capable of self-looping thus preventing efficient reverse transcription. This issue is solved by taking advantage of the looped structure designed in LIANTI transposon, which has been converted to another 19-bp sequence on the 3' end of each RNA molecule. As shown in Fig. 1C, instead of self-looping (~500-nt loop), linearly amplified RNAs form self-priming structures (27-nt loop) much more efficiently, enabling subsequent reverse transcription.

Interestingly, during in vitro transcription linear amplification of DNA fragments attached by T7 promoter on both ends, we found that the transcription is often dominated by one random direction. This is likely due to the suppression of transcription initiation and/or elongation by

head-on collision with the transcription complex from the opposite direction. As a result, the majority of DNA fragments are transcribed in a unidirectional way despite the attachment of T7 promoters on both ends, and we don't have strand-specific information of the original double-stranded DNA after LIANTI amplification and sequencing. Since we can't distinguish and sequence the amplicons from both strands of the duplex DNA simultaneously, kindred cells have to be used to correct SNV false positives generated due to chemical stability of DNA bases.

It is worth to note that T7 RNA polymerase, besides its transcriptional activity on DNA template to synthesize single-stranded RNA, has another enzymatic activity called RNA-dependent RNA polymerase activity, where it extends from RNA primers along a single-stranded RNA template to generate double-stranded RNA (51-53). As a result, during in vitro transcription some transcribed RNAs adopting the self-priming structure may end up with a dsRNA region of varied lengths on the 3' end. After reverse transcription and RNase treatment, this may lead to varied mapping coordinates of the 3' end of the LIANTI amplicons from the same original gDNA fragment. Furthermore, during the sonication process in library preparation new 3' ends were created in sequencing reads, further complicating the situation. To deal with such complication, in the data analysis pipeline the sequencing reads were grouped only based on the same 5'-ends.

In addition, some RNAs without the self-priming structure on their 3' ends due to early transcription termination may nonspecifically anneal with each other for cDNA synthesis during reverse transcription. However, such cDNAs will share the same 5' ends with other amplicons from the same original DNA fragment, without complicating downstream data analysis.

Finally, Tn5 transposition has an intrinsic mechanism to leave a 9-bp overlap between the adjacent genomic fragments it separates by insertion. Considering this feature, amplicons mapped to adjacent genomic positions containing such an overlap were merged in the data analysis pipeline.

Single replicating cells in early S-phase

BJ cells on a dish with ~10% confluence were incubated for 21 hrs in growth medium supplemented with 2 mM thymidine. Thymidine was then completely removed by 1X PBS wash for several times, followed by incubation for 10 hrs in growth medium. After that, cells were gain incubated for 17 hrs in growth medium supplemented with 2 mM thymidine. The double thymidine block protocol generated synchronized BJ cells at the beginning of S-phase.

Upon harvest, thymidine was completely removed by 1X PBS wash for several times, and synchronized BJ cells were released into early S-phase by incubation in growth medium. Within

the first hour of S-phase release, the cells were trypsinized, 1X PBS washed and 4% paraformaldehyde fixed for 10 min under room temperature. The fixed cells were washed several times with 1X PBS and FACS sorted into PCR tubes containing the lysis buffer. An additional 65C 3 hrs incubation for reverse crosslink was carried out, followed by single cell lysis, LIANTI assay, and library preparation using the same protocol as previous experiments.

Ultraviolet (UV) radiation treatment

BJ cells on a dish with ~50% confluence were washed by 1X PBS several times. After final wash and complete PBS removal, the cells on the dish were exposed to UV-C radiation (UV crosslinker, VWR) at given doses (5, 15, and 30 J/m²). Immediately after UV-C radiation, the cells were trypsinized and replated into new dishes for incubation. After several rounds of cell division, single BJ cells were trypsinized, mouth pipetted and seeded into poly-D-lysine bio-coated 96-well tissue culture plate (Corning), and incubated in the presence of condition medium and recombinant human fibroblast growth factor (R&D Systems). After several generations of cell division from single cells to ~16 cells, pairs of kindred cells divided from the original seeding cell were trypsinized and mouth pipetted into PCR tubes containing the lysis buffer, followed by single cell lysis, LIANTI assay, and library preparation using the same protocol as previous experiments.

Uracil-DNA glycosylase (UDG) treatment

UDG treatment was performed within the LIANTI assay, by adding 1 µL USER enzyme (a mixture of uracil-DNA glycosylase and DNA glycosylase-lyase endonuclease VIII, New England Biolabs) into the mixture right before the in vitro transcription assembly, and incubated at 37°C for 30 min. 3 µL UGI (uracil glycosylase inhibitor, New England Biolabs) was then added and incubated at 37°C for 10 min. In vitro transcription assay was then assembled the same as the LIANTI assay.

Illumina sequencing

The BJ bulk sample, single BJ cells, UDG-treated cells, and UV-induced cells were sequenced on Illumina HiSeq 2500 platform, using the "standard run" mode (8-lane-per-flowcell) with 2X125 bp pair-end sequencing, or the "rapid run" mode (2-lane-per-flowcell) with 2X150 bp pair-end sequencing. Each sample was sequenced by 2 HiSeq lanes generating ~90 Gb raw data. The raw

data of single BJ cells using previous WGA methods other than LIANTI were downloaded from (1).

Single replicating BJ cells in early S-phase were sequenced on Illumina HiSeq 4000 platform, using the "standard run" mode (8-lane-per-flowcell) with 2X150 bp pair-end sequencing. 4 HiSeq lanes were shared by 24 samples with NEBNext Indexes.

Sequencing data processing and alignment

Before read mapping, we identified and trimmed off Illumina sequencing adapters, inline barcodes, binding sequences and T7 promoter sequences, and then merged the two ends in a read pair if there is an unambiguous overlap no shorter than 8bp. We mapped the processed reads to the human reference genome GRCh37 with BWA-MEM under the default setting (54). These steps are implemented as the "trim" command of the "lianti" toolkit (<https://gitlab.com/lh3/lianti>). After read mapping, we mark PCR duplicate with adna-ldup from <https://github.com/DReichlab/adna>. Unlike typical markDuplicators, this tool is aware of inline barcodes and thus can accurately identify PCR duplicates. Here is the command line:

```
seqtk mergepe reads.R1.fq.gz reads.R2.fq.gz | lianti trim - | bwa mem -Cpt8 hs37.fa - \  
| samtools view -uS - | sambamba sort /dev/stdin | adna-ldup -T - > mapping.bam
```

It is worth to note that during the alignment of LIANTI sequencing reads, unlike bulk sequencing and other single-cell WGA data, the de-duplication process is based on the same mapping coordinate plus the same molecular barcode (UMI) on the 5' end. After the attachment of UMI during second strand synthesis, LIANTI amplicons linearly amplified from the same original gDNA fragment would contain the same mapping coordinate but different 5'-end UMI, thus won't be removed during de-duplication process. The duplicates generated in library preparation PCR would contain the same mapping coordinate plus the same 5'-end UMI, and will be removed correctly during de-duplication.

Amplification uniformity characterization

Plots of GC-corrected read depths (Figure 1D) were generated by first calculating the average read depth with SAMTOOLS (version 1.3.1-34-g26e1ea5 of the develop branch) and AWK in each of the consecutive, non-overlapping 10-Kb or 1-Mb bin across the human genome that is not masked by a 75-bp universal mask (um75-hs37d5). A bin was discarded if it spanned two chromosomes, or if more than 20% of its physical length was masked. A standard curve of GC bias was then created by categorizing each bin into a GC-percentage bin (sized 2%, from 0% to

100%), and fitting a polynomial of degree 4 (55) in MATLAB (version R2016a, using the function “polyfit”) through the median average depth in each GC-percentage bin versus the center of each GC-percentage bin. Finally, GC biases were corrected by normalizing each average read depth according to the standard curve. Estimated copy numbers were calculated by dividing the normalized average read depth by their median and then multiplying it by two. For MALBAC, instead of GC correction, estimated copy numbers were calculated by dividing the average read depth by the average of the two other cells, each of which was first normalized by its own median average depth.

Plots of coefficients of variation (Figure 1E) were generated by calculating the average read depth with in each of the consecutive, non-overlapping bin of various sizes across the human genome that is not masked, and then calculating their coefficients of variation. GC bias was not corrected for this analysis. For MALBAC, normalization by two other cells only worked for bin sizes ≥ 20 Kb; therefore, a straight line was used to connect normalized and raw MALBAC.

We believe that the CV plot is a better measure of amplification uniformity than both the Lorenz curve and the power spectrum. First, the value of the CV at a 1-bp bin size contains information similar to the Lorenz curve, because the CV equals to the relative root mean square difference divided by root 2, while the Gini coefficient equals to the relative mean absolute difference divided by 2. Therefore, both the 1-bp CV and the Lorenz curve provide estimates for the variation of per-base read depths. Second, CVs at >1 -bp bin sizes conveniently extends the idea of per-base uniformity to uniformity on larger scales, which naturally reflects capabilities of calling CNVs from read depths. Therefore, the CV plot can be seen as an extension of the Lorenz curve to binned, rather than per-base, read depths. Finally, on larger scales, CVs are more reproducible than the power spectrum. Discrete Fourier transform decomposes a serial of read depths into sinusoidal waves, which cannot realistically represent the usually aperiodic amplification biases and noises, and is thus sensitive to details of handling regions with zero depths (such as the centromeres) or very high depths (such as repetitive regions). In reality, amplitudes are usually concentrated exclusively at the two ends of the spectrum, with intermediate frequencies containing only a minority of the actual read-depth variation.

The theoretical lower limit of CVs (“Poisson”) reflects the Poisson sampling error of sequencing to a finite depth. However, this limit depends on many details, such as the length of each chromosome, how the genome is masked, the type and length of reads, and the insert size distribution. In the simplest case, a continuous genome with boundaries ignored is sequenced by single-end reads with a length $l = 125$ to a depth of $d = 30$. At a bin size L , the theoretical lower limit is:

$$CV(L) = \begin{cases} \sqrt{\frac{1}{dL} - \frac{2-1}{3dL^2}} & (L \geq) \\ \sqrt{\frac{1}{d} - \frac{L^2-1}{3dL}} & (L \leq - 1) \end{cases}$$

We compared the performance of LIANTI with other WGA methods. MALBAC WGA was done with Yikon Genomics Single Cell Whole Genome Amplification Kit, the MALBAC-like WGA with Rubicon Genomics PicoPLEX WGA Kit, MDA with Qiagen REPLI-g Single Cell Kit as well as General Electric illustra Single Cell GenomiPhi DNA Amplification Kit, and DOP-PCR was done with Sigma-Aldrich GenomePlex Single Cell Whole Genome Amplification Kit. Genome coverage was determined as one minus the horizontal intercept of the Lorenz curve.

Copy number variation (CNV) detection

The CNVs of BJ bulk sample were called by Control-FREEC (55) with 500bp bin size.

For Lianti cells, we first identify and collapse fragments by grouping reads that have the same positions on the 5'-end of reads. We also merge fragments having a 7–9bp overlap as this may be created by Tn5 transposition and discard fragments consisting of 3 reads or less as noises. This process is called digital counting. It ideally gives the absolute copy number, though in practice it is noisy. The noisiness may be attributed, at least partially, to the inevitable random loss of DNA fragments leading to allele dropout, the misalignment from repetitive and low-complexity regions across the human genome, and the potential nonspecific annealing and enzymatic activity during LIANTI and library preparation.

We use a max-scoring approach to identify CNVs. For gains, let g_i be a score at position i . We let $g_i=1$ if the number of alleles is above 2, or -4 otherwise. We then try to find all intervals $[a,b]$ such that the sum of g_i in the interval is a positive local maximum in that extending or shrinking the interval would lead to a smaller sum. High-scoring intervals are potential CNV gains. To find a statistically significant scoring threshold, we randomly shuffle the allele counts for 200 times, find the top score and fit an Extreme Value Distribution (EVD), which approximates the null distribution of max scores. We report all intervals with P-value below 1e-10, according to the estimated EVD, as CNV gains. CNV losses are called in a similar way except the scoring scheme. These steps are all implemented in the companion “lianti” toolkit (<https://gitlab.com/lh3/lianti>).

Structural variation (SV) detection

The ability to detect structural variations (SVs) in a single cell has been confounded by the chimera artifact commonly associated with WGA (56). By eliminating nonspecific priming commonly employed in many other WGA methods, the chimera rate of LIANTI is much reduced, allowing us to detect micro-deletion SVs based on split reads from a single BJ cell with a 22% false discovery rate (FDR), which is the lowest among all WGA methods (Fig. S5A-B).

Micro-deletion SVs were called by LUMPY (57) (version 0.2.13) with the following parameters “-b -mw 4 -tt 0 -sr id:<sample_name>, bam_file:<bam_file>, back_distance:10, weight:1, min_mapping_threshold:20” and only deletions between 1 Kb and 1 Mb were analyzed. True positives were the “pairstopair -type both” intersect with calls from the standard (“Bulk1”) by BEDTOOLS (58) (version 2.26.0). False positives were the “pairstopair -type notboth” intersect with less stringent calls (same parameters except for “-mw 1”) from the standard (“Bulk1”).

Single nucleotide variation (SNV) detection

We call a heterozygous SNP from the bulk sample if (i) the total read depth is above 15; (ii) both alleles are supported by 5 reads or more; (iii) the fraction of reads supporting each allele is at least 30%; (iv) the root-mean-square mapping quality is no less than 40. For a non-Lianti single-cell sample, we call a non-reference (NR) allele if (1) the NR allele is supported by at least 5 reads in the single-cell sample; (2) the fraction of the NR reads in the single-cell sample is at least 20%; (3) the root-mean-square mapping quality is no less than 40. For Lianti samples, we further (4) discard a NR allele if fragments with the same 5'-end mapping position (i.e. they are likely to come from the same haplotype) have both the NR allele and a reference allele. We count a false negative if a heterozygote is called in the bulk but the NR allele is uncalled in the single-cell sample with the criteria above. The false negative rate (FNR) equals the number of false negatives as a fraction of total number of called heterozygotes in the bulk. We count a NR allele dropout if at a heterozygote in the bulk sample, no reads in the single-cell sample have the NR allele in the heterozygote. The NR allele dropout rate (NR-ADO) equals the number of NR dropout out of the total number of called heterozygotes in the bulk. Similarly, we may define reference-ADO. It is slightly smaller the NR-ADO due to the reference bias during mapping. For example, for sample BJ1, the NR-ADO is 17.6% and the reference-ADO is 17.4%. We only report the average as the difference is small enough.

We call an SNV if there is a called NR allele and: (a) the total read depth in the bulk sample is above 15; (b) no bulk read has the NR allele; (c) if two SNVs are within 100bp from each other, both are discarded. With this procedure, we called 15,940 SNVs on the autosome of sample BJ1.

1.5% of them overlap with SNPs from the 1000 Genomes Project over 0.1% frequency. This suggests most SNVs are not contaminations from other human samples.

Calling SNVs from a pair of kindred Lianti cells is similar to calling from one Lianti cell, except rule (1): we require the NR allele to be supported by at least 4 reads from each single cell, i.e. the NR allele should be evident in both cells in the kindred pair. We have also tried the same procedure to call UV-induced short insertions and deletions (INDELs). We called 294 putative INDELs from the BJ1-BJ2 pair, the negative control, but less than 100 INDELs in all the paired kindred cells. We do not have enough power to separate false positives and true UV-induced INDELs, if there are any.

The command lines used in this section are as follows:

```
lianti pileup -P20 -ycf hs37.fa bulk1.bam MDA.bam | k8 plp-diff.js /dev/stdin > MDA.diff  
lianti pileup -P20 -ycf hs37.fa -L1 bulk1.bam lianti.bam | k8 plp-diff.js /dev/stdin > lianti.diff  
lianti pileup -P20 -ycf hs37.fa -L2 bulk1.bam lianti-1.bam lianti-2.bam \  
| k8 plp-diff.js -p /dev/stdin > lianti-pair.diff
```

Program “lianti” and “plp-diff.js” can be found at <https://gitlab.com/lh3/lianti>; “k8” can be downloaded at <http://bit.ly/k8-js>. Note that we also discarded all SNVs overlapping the um-hd37d5 mask created by the SGDP project (available at <https://github.com/lh3/sgdp-fermi>). This mask filters 13% of human genome that is enriched with misassemblies, common CNVs, low-complexity regions and hard-to-map regions.

The numbers of SNVs called from UV-radiated kindred cell pairs were 1649, 3168, 2709, 1761, for 5 J/m² pair, 15 J/m² pair, 30 J/m² pair 1, 30 J/m² pair 2, respectively. The FNR of SNVs called from UV-radiated kindred cell pairs were 67.3%, 45.0%, 66.2%, 57.3%, for 5 J/m² pair, 15 J/m² pair, 30 J/m² pair 1, 30 J/m² pair 2, respectively.

Comparison between different single-cell WGA methods

Multiple droplet MDA methods have been developed demonstrating enhanced performance than traditional in-tube MDA (9, 12, 59-62). However, the reported parameters are not directly comparable due to the differences in single cells to start with (bacteria, normal diploid human cells, cancer cells, dividing cells, etc), sequencing depth used, and the criteria applied in the data analysis pipeline, each of which matters a lot in the final performance.

To evaluate and compare different methods, here we provided a fair comparison starting with the same normal diploid human single cells, sequencing to the same amount of raw data, analyzing using the same pipeline and criteria, and clearly showed that LIANTI outperforms

previous WGA methods in genome coverage, amplification uniformity, accuracy for single nucleotide variation (SNV) detection, and spatial resolution for copy number variation (CNV) measurement.

It is worth to note that several recent publications achieved high throughput profiling of single-cell genomes or single-cell Hi-C assay based on combinatorial barcoding (63, 64). While the current LIANTI assay does not have such a high scalability, it offers much higher genome coverage, amplification uniformity, and the accuracy for SNV and CNV detection.

Analysis of single replicating cells in early S-phase

Read depths were plotted in a similar manner as non-replicating cells with two exceptions. First, GC bias was not corrected, because GC percentage correlates strongly with replication timing. This causes the read depth values for 4 copies to be slightly less than 2 times the values for 2 copies. Second, in each cell the normalization factor, namely the read depth value corresponding to 2 copies, was determined by visually inspecting the masked, autosome-wide histogram of read depths, instead of simply taking the median. This particular value was 3.90 for BJ-S1, 3.75 for BJ-S2, 3.95 for BJ-S3, 3.85 for BJ-S4, 4.20 for BJ-S5, 4.40 for BJ-S6, 3.40 for BJ-S7, 3.50 for BJ-S8, 3.10 for BJ-S9, 2.50 for BJ-S10, and 4.15 for BJ-S11.

These early S-phase cells harbor more “dips” than usual in read depths on the scale of tens of Kbs, presumably caused by DNA loss as a result of fixation; however, these “dips” have limited impact on bin sizes greater than or equal to 100 Kb. In addition, one cell (BJ-S2) harbors a large copy-number gain; therefore, the corresponding region (chr7, from 80 Mb onwards) was excluded from subsequent analysis.

The 2-d histogram for each cell pairs was created from normalized and masked read depths binned every 100 Kb. Similar to the amplification uniformity plots, a bin was discarded if it spanned two chromosomes, or if more than 20% of its physical length was masked by a 75-bp universal mask (um75-hs37d5). The 2-d density was smoothed in MATLAB by “ksdensity” with a Normal kernel of a 0.1 bandwidth.

The 2-d histogram between each cell and public data was created from normalized and masked read depths or public data values binned every 100 Kb. A bin was only discarded if it contains any missing values from the public data. For this part only, the mask was a more stringent, 35-bp universal mask (um35-hs37d5). The 2-d density was smoothed in MATLAB by “ksdensity” with a Normal kernel of a 0.1 bandwidth for read depths and DNase-Seq signals and 2 for Repli-Seq.

Bulk replication timings (Repli-Seq) (23) were the same as those used for UV SNV analysis.

DNase-Seq signals (-log₁₀ p-values) were downloaded from the DS18224 sample of the Roadmap Epigenomics Project (24):

http://egg2.wustl.edu/roadmap/data/byFileType/signal/unconsolidated/pval/UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin01.DS18224.pval.signal.bigwig

Analysis of UV-induced SNVs

For each set of genomic regions of interest, the fraction of *de novo* SNVs in the regions was calculated by intersecting the locations of SNVs with a BED file containing the regions using BEDTOOLS. The expected fraction of *de novo* SNVs in the regions, namely the null model, was calculated by dividing the total length of the regions that are not masked by the total length of the human genome that are not masked (2,342,671,754 bp). Single-cell SNVs that do not intersect with kindred cells (presumably false positive SNVs) served as an additional control for the SNV-calling pipeline.

DNase-Seq peaks were downloaded from the E055 sample of the Roadmap Epigenomics Project (24): <http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/E055-DNase.macs2.narrowPeak.gz>

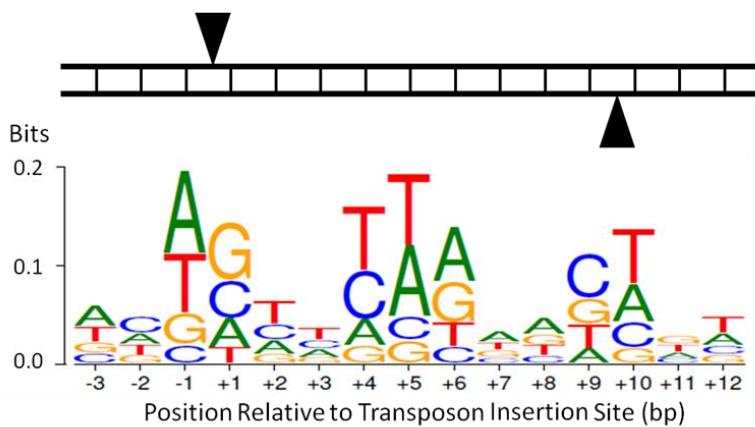
Early-replicating regions were generated by thresholding values more than 50 using BWTOOL (version 1.0) on the following data (23) deposited on the UCSC Genome Browser: <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwRepliSeq/wgEncodeUwRepliSeqBjWaveSignalRep1.bigWig>

Transcribed regions (UCSC Genes) were downloaded from the UCSC Genome Browser: UCSC Table Browser --> hg19 --> Genes and Gene Predictions --> UCSC Genes --> knownGene --> BED

Stranded mutation spectrum was calculated according to (26) using the same data as transcribed regions. For example, if a C-to-T SNV overlaps with plus-strand genes only, it will be counted as one C-to-T SNV on the non-template strand and one G-to-A SNV on the template strand. The expected non-template-to-template ratio of mutation rates were calculated from the percentages of A, C, T, and G in all template strands and in all non-template strands.

Supporting Figures and Table

A



B

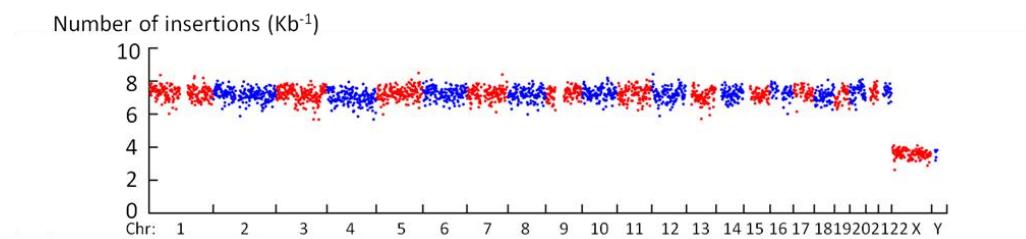


Fig. S1. LIANTI transposon insertion sites throughout the human genome. Insertion sites were identified by first grouping reads from single cell BJ1 into fragments in a “do-not-merge” mode (command: “lianti group -Mn3”), then only keeping fragments with a root mean square mapping quality of at least 20, and finally extracting insertion sites and insertion orientation from closed ends of these fragments. **(A)** Sequence logo. Sequences near insertion sites were extracted from the reference genome with BEDTOOLS, and the logo was generated with WEBLOGO (version 3.4). The y axis represents sequence conservation measured in bits. The maximal sequence conservation (having 100% one kind of nucleotide) is 2 bits. **The maximum is only 0.2 bits on the y-axis, suggesting the insertion and fragmentation is largely random.** **(B)** Insertion density throughout the genome with 1-Mb bin size. Insertion sites were filtered by the 75-bp universal mask, and then counted in consecutive, non-overlapping 1-Mb (physical length) bins across each chromosome. Bins that were more than 20% masked were discarded. In each bin, insertion density was calculated by dividing the number of insertion sites by the length of the bin after masking.

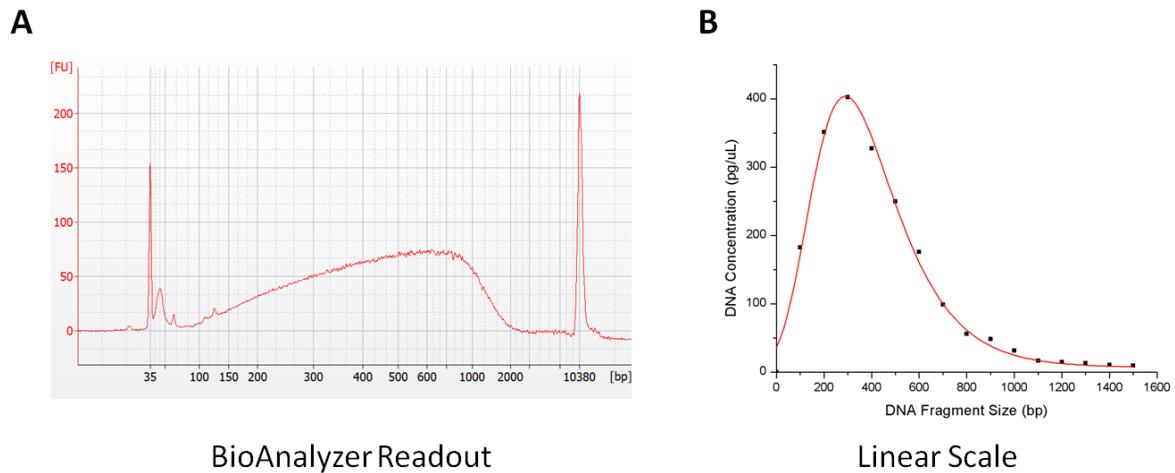


Fig. S2. Fragment size distribution of LIANTI amplicons of the genomic DNA in a single cell. **(A)** BioAnalyzer readout. **(B)** Linear scale plot generated based on BioAnalyzer readout.

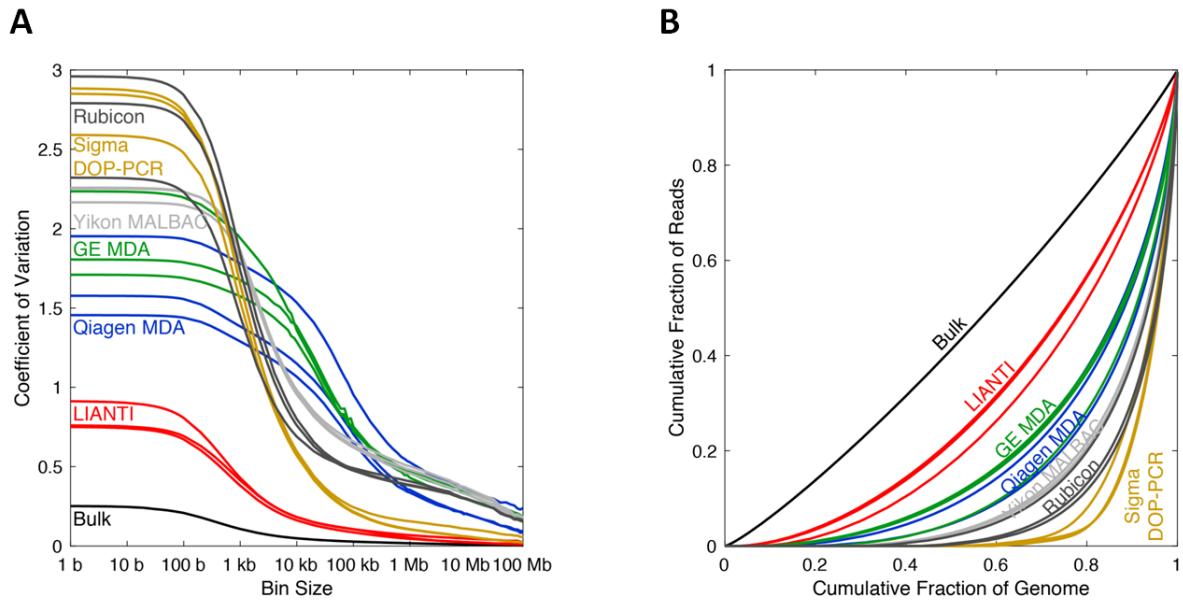


Fig. S3. Amplification uniformity comparison of various single-cell WGA methods, with 3 individual cells per WGA method plotted separately. **(A)** Amplification uniformity by the coefficient of variation for read depths along the genome with different bin sizes between 1 bp and 100 Mb, representing amplification uniformity on all scales. **(B)** Amplification uniformity by Lorenz curve of cumulative fraction of reads versus cumulative fractions of genome. Perfectly uniform coverage leads to the diagonal line, and deviation from the diagonal line represents amplification bias.

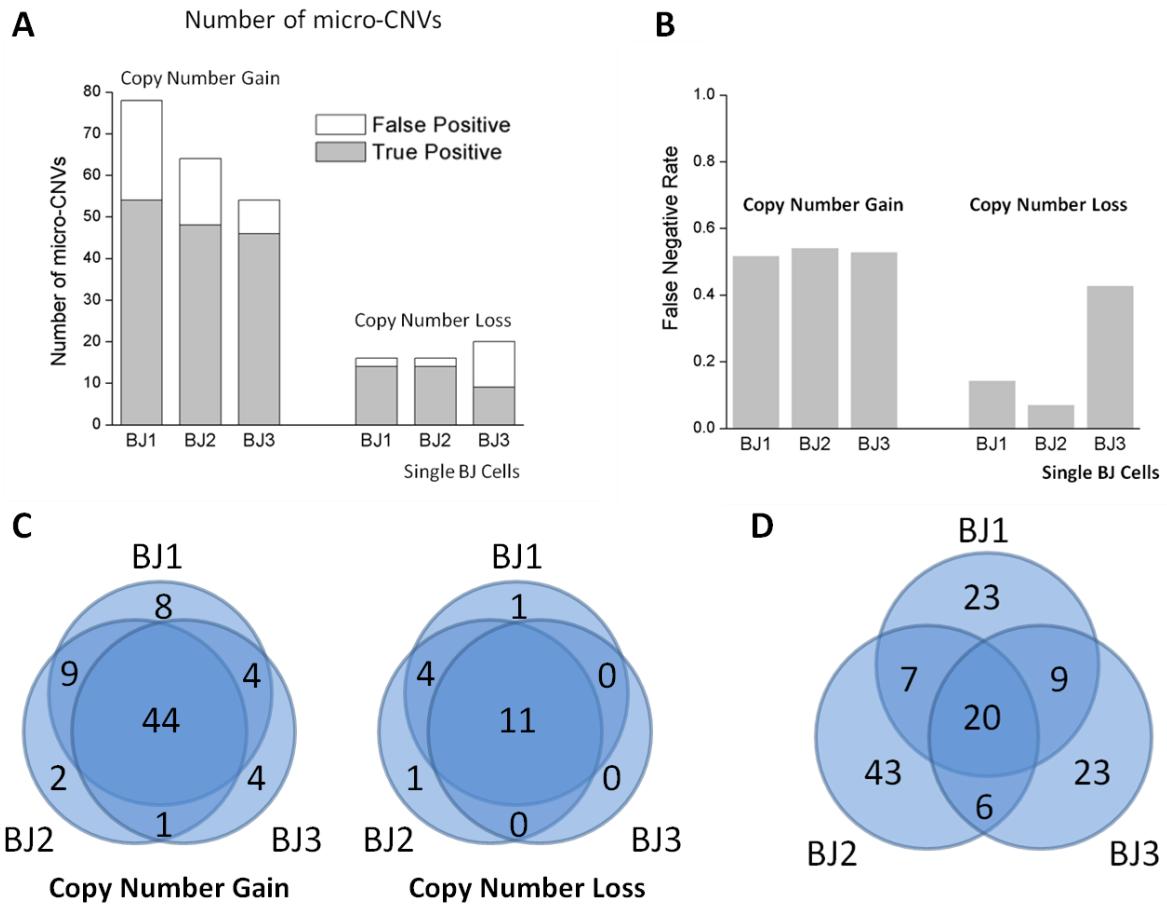


Fig. S4. Detection of micro-CNVs in single BJ cells. **(A)** True positives and false positives of the genome-wide micro-CNV detection in a single BJ cell, including copy number gain and 2-to-0 copy number loss. “True Positive” represent micro-CNVs confirmed in the unamplified bulk. “False Positive” represent micro-CNVs not detected in the unamplified bulk. **(B)** False negative rates (FNR) of micro-CNV detection in single BJ cells, including copy number gain and 2-to-0 copy number loss. **(C)** Consistency and reproducibility of micro-CNV detection in 3 individual BJ cells with identical genome, including copy number gain and 2-to-0 copy number loss. **(D)** Number of false discoveries of 2-to-1 copy number loss in 3 individual BJ cells, caused by random allele dropouts and systematic error within repetitive regions.

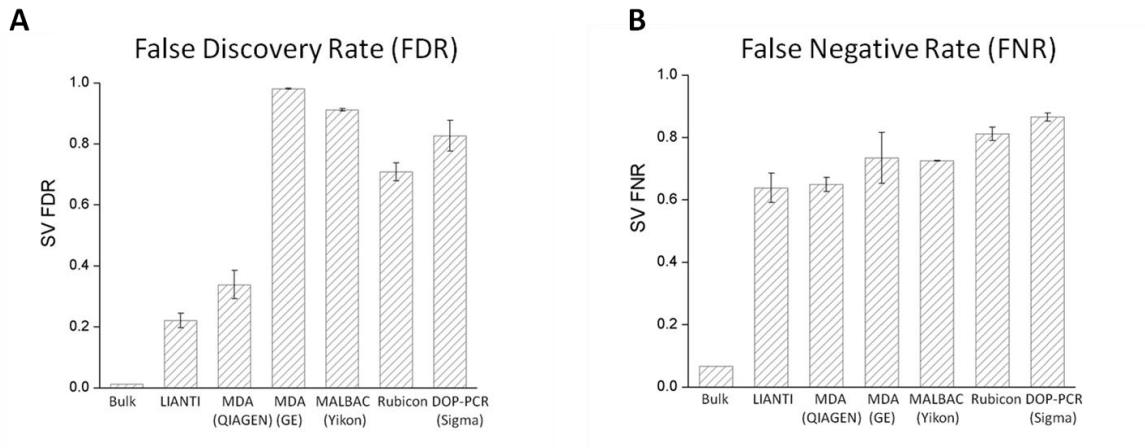
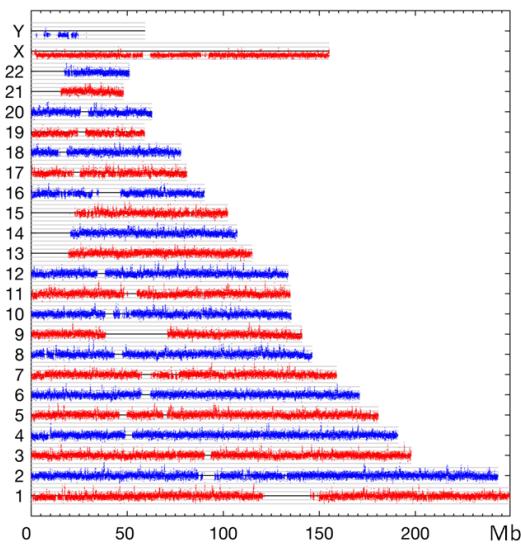
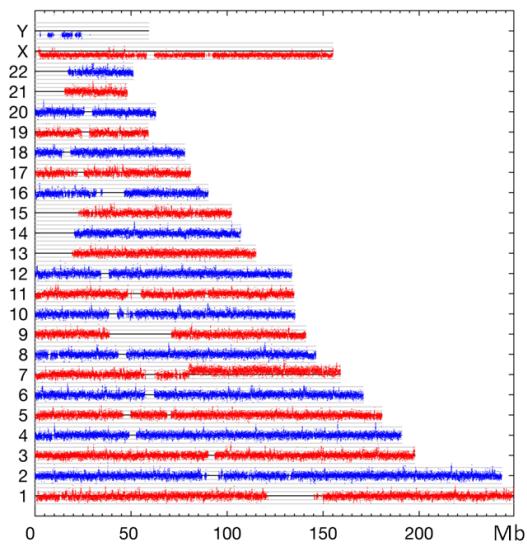


Fig. S5. Detection of SVs in single BJ cells. **(A)** False discovery rate (FDR). **(B)** False negative rate (FNR). The standard error is calculated from 3 individual BJ cells.

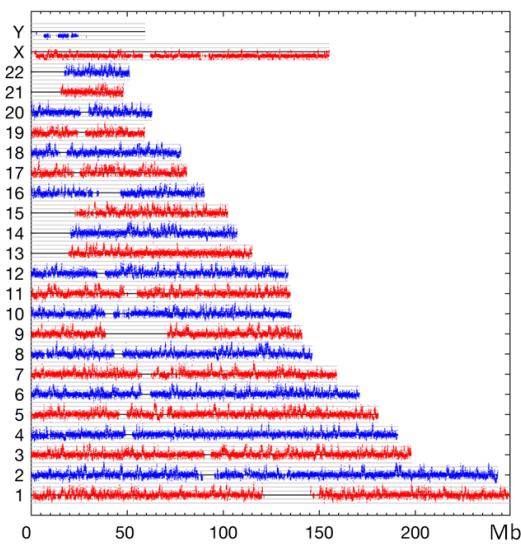
BJ-S1



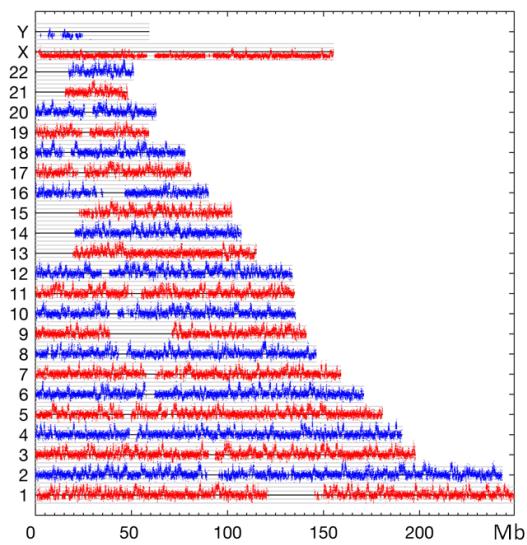
BJ-S2



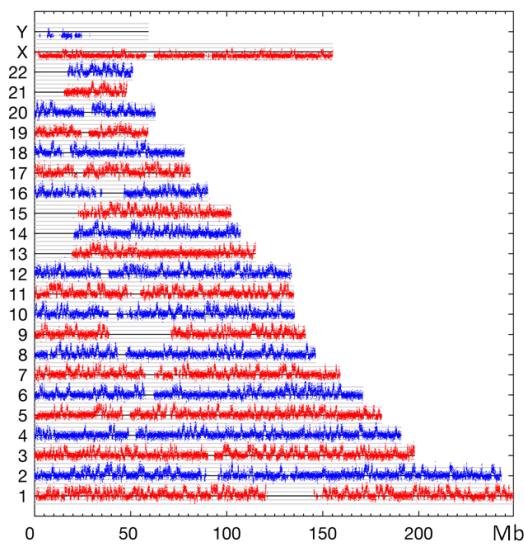
BJ-S3



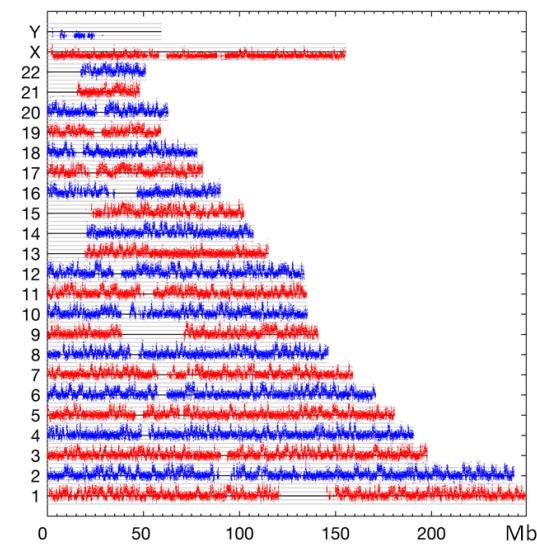
BJ-S4



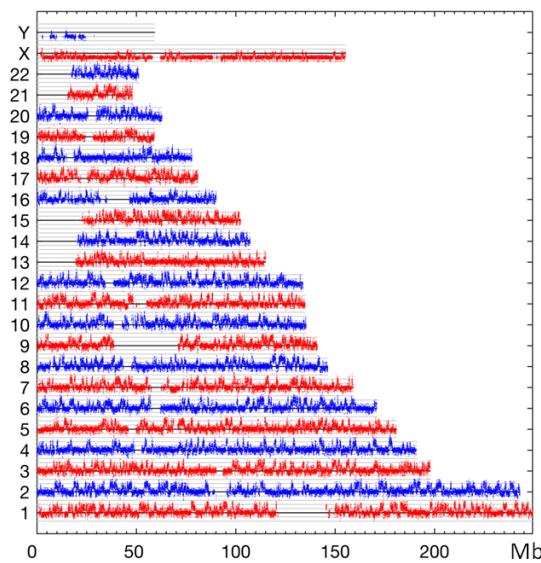
BJ-S5



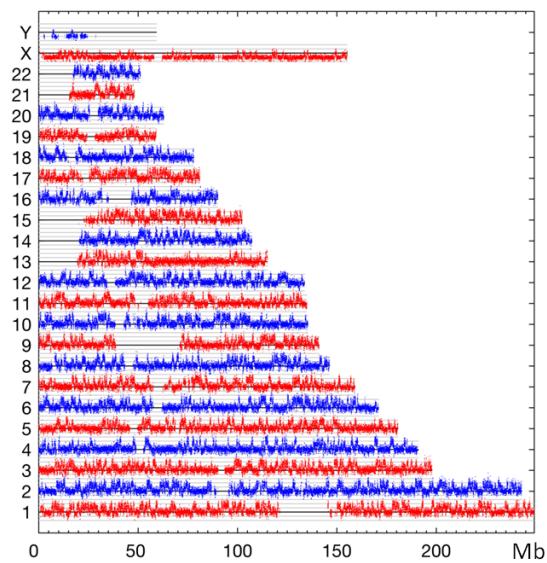
BJ-S6



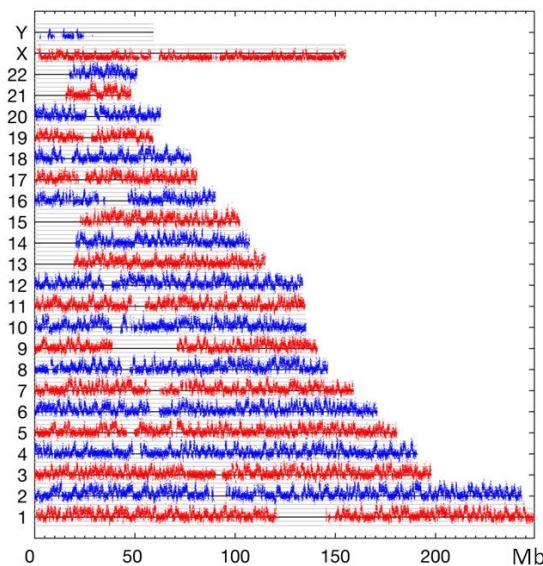
BJ-S7



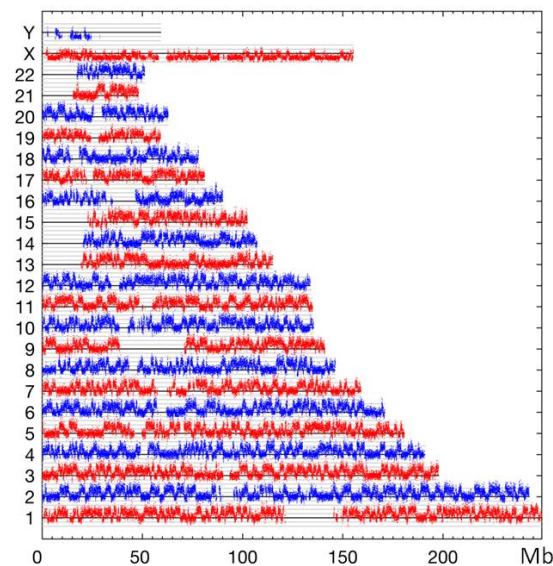
BJ-S8



BJ-S9



BJ-S10



BJ-S11

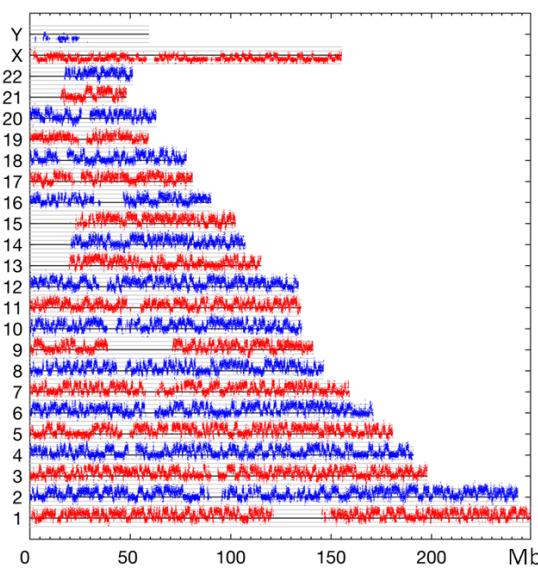


Fig. S6. Genome-wide detection of replication origin firing and replicon formation events based on the copy number gain with 10-Kb bin size in 11 single cells in early S-phase.

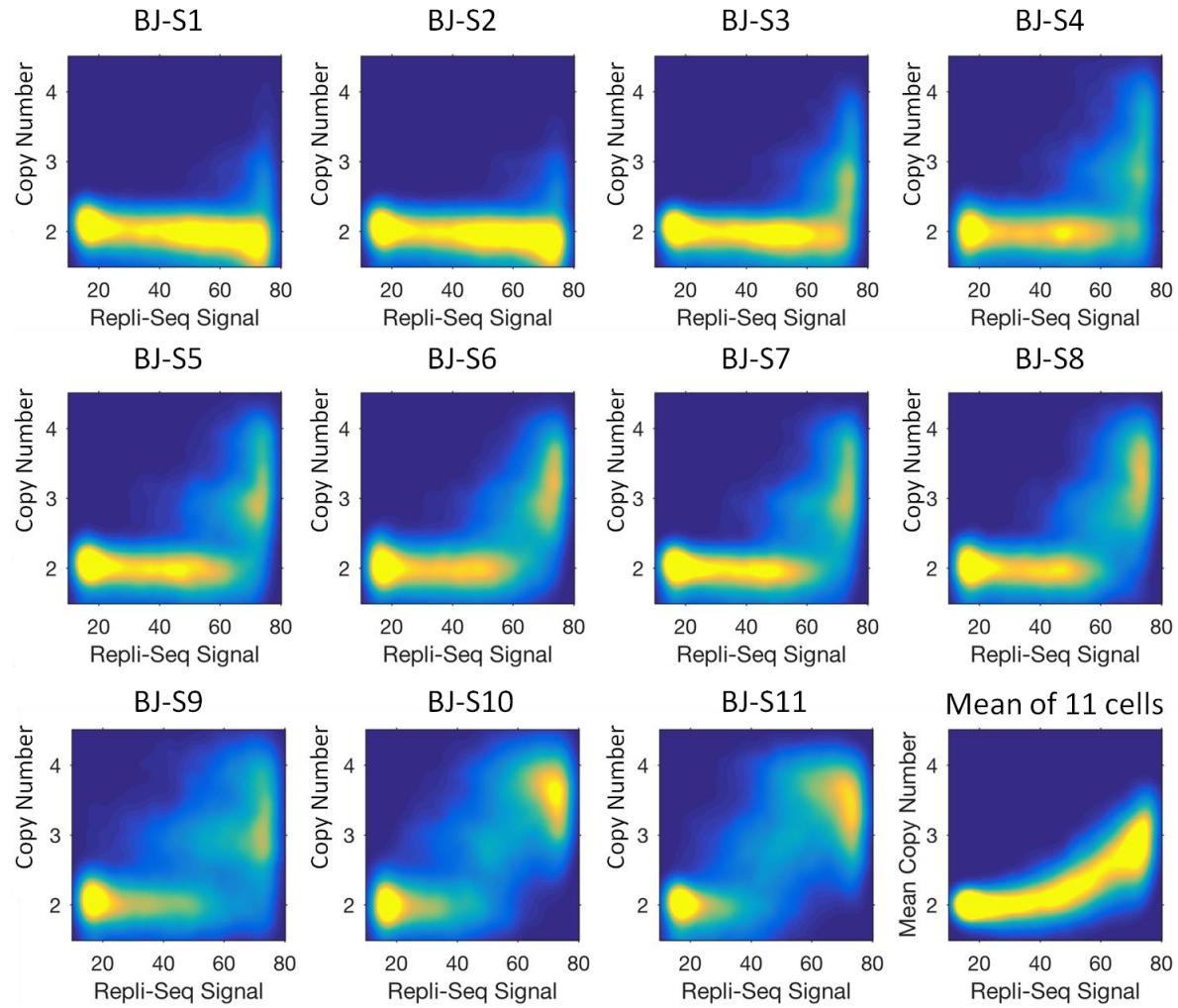


Fig. S7. Correlation plots of replicon copy numbers of 11 single cells and the mean of these 11 cells with the bulk readout of the Repli-Seq assay using 100-Kb bin size.

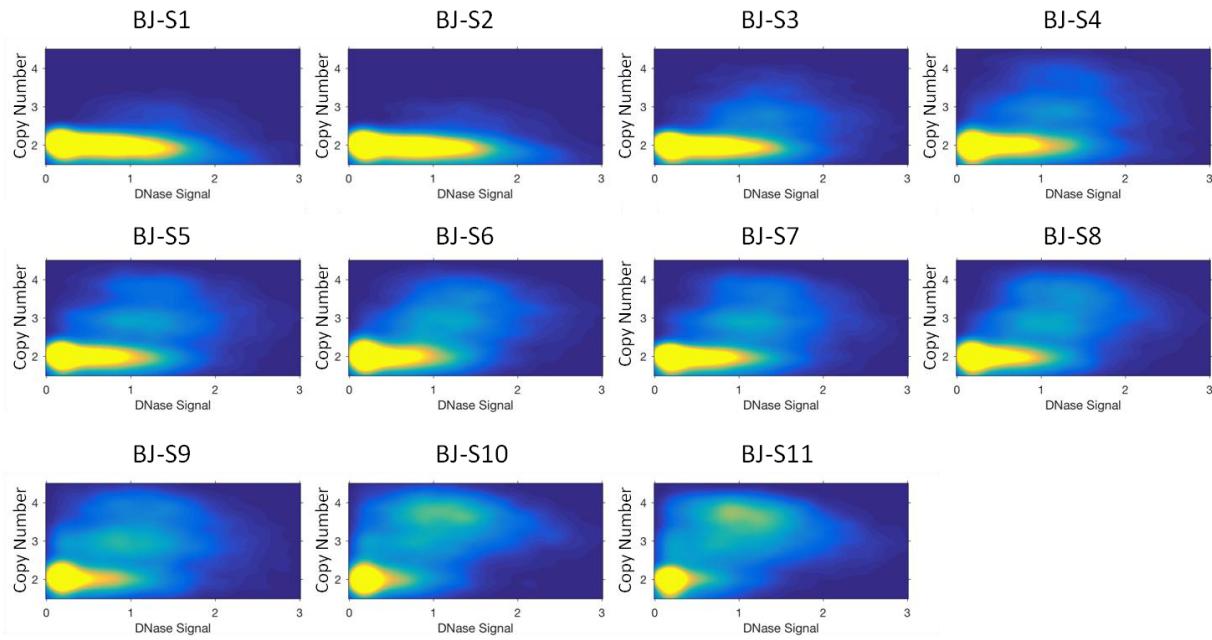


Fig. S8. Correlation plots of replicon copy numbers of 11 single cells with the bulk readout of the DNase I hypersensitive assay using 100-Kb bin size.

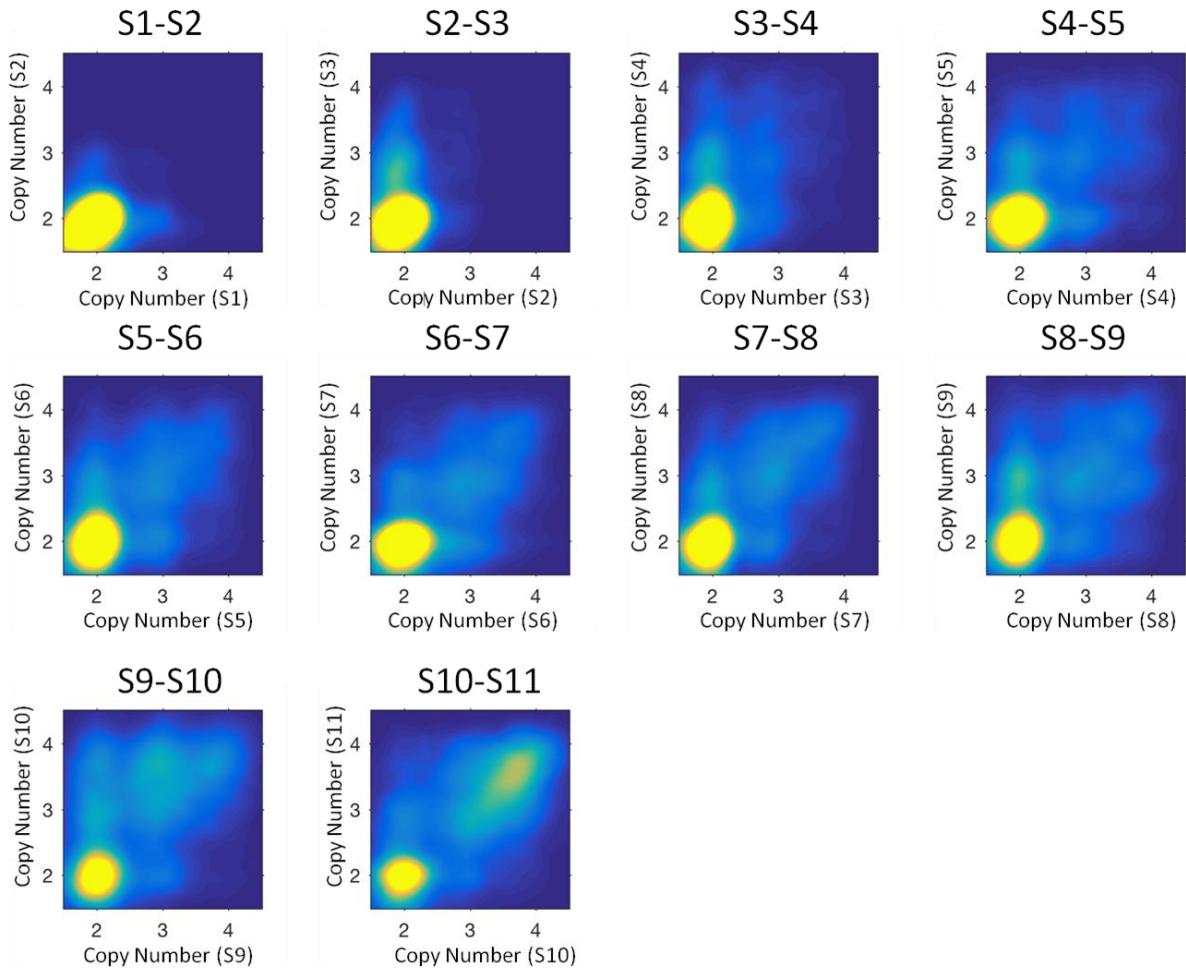


Fig. S9. Correlation plots of replicon copy numbers between 10 pairs of single cells close in replication progress in S-phase with 100-Kb bin size.

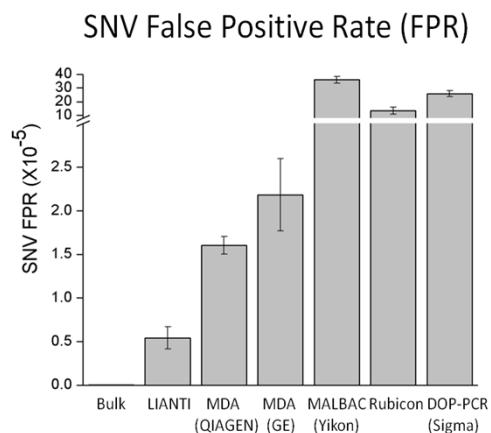
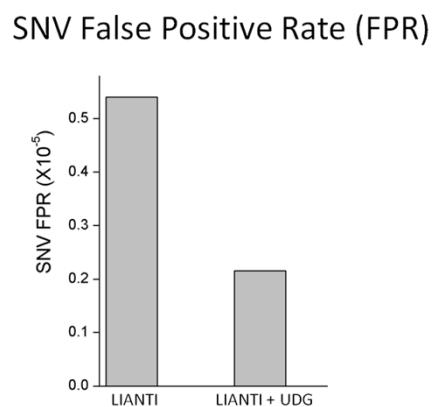
A**B**

Fig. S10. Detection of SNVs in single BJ cells. **(A)** False positive rate (FPR) of various WGA methods. The standard error is calculated from 3 individual BJ cells. **(B)** False positive rate (FPR) of LIANTI with UDG treatment.

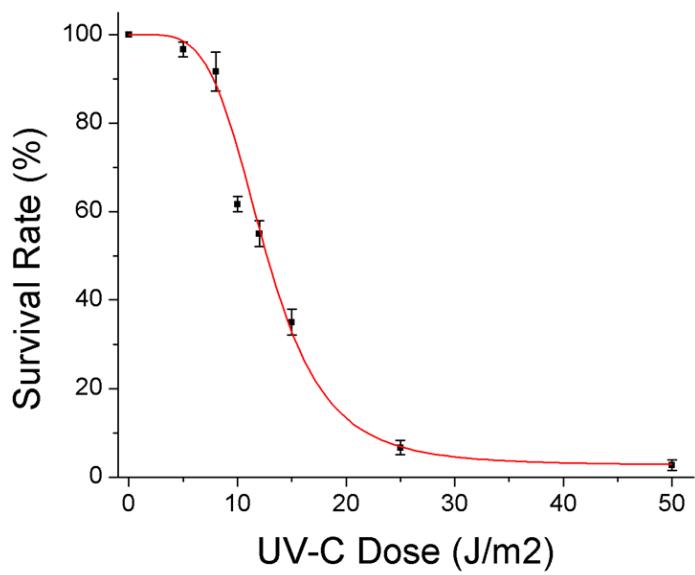


Fig. S11. Survival curve of BJ cells after UV-C radiation with various doses. The standard error is calculated from experimental triplicates. The curve is fitted by sigmoid function.

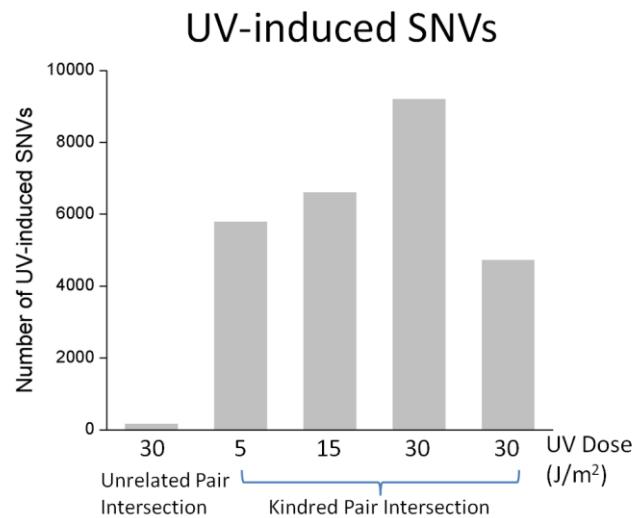


Fig. S12. Number of UV-induced SNVs throughout the genome as a function of UV dose.

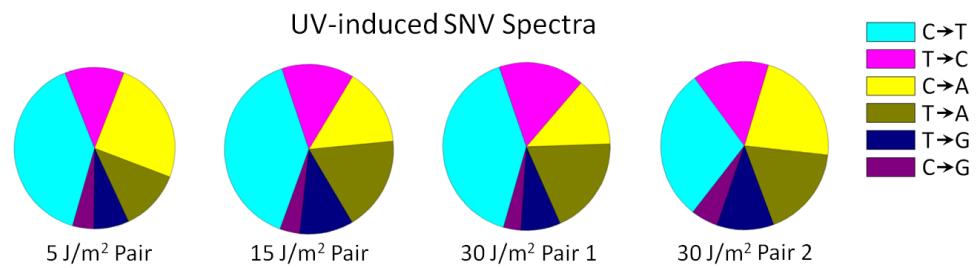


Fig. S13. Spectra of UV-induced SNVs throughout the genome.

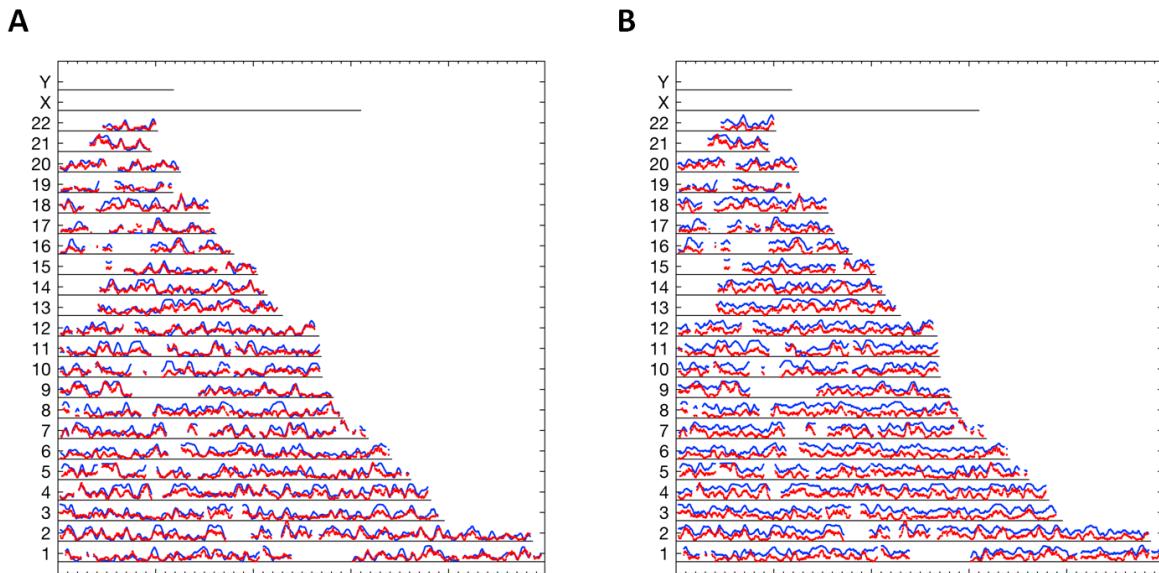


Fig. S14. Overlay of the density of UV-induced SNVs (red) throughout the genome and (A) the minus Repli-Seq signal (blue) reflecting the replicated genomic regions, as well as (B) the minus DNase I hypersensitive signal (blue). Both signals were calculated in 2-Mb moving windows with 100-kb increments. SNVs from all four kindred pairs were pooled ($n = 9403$).

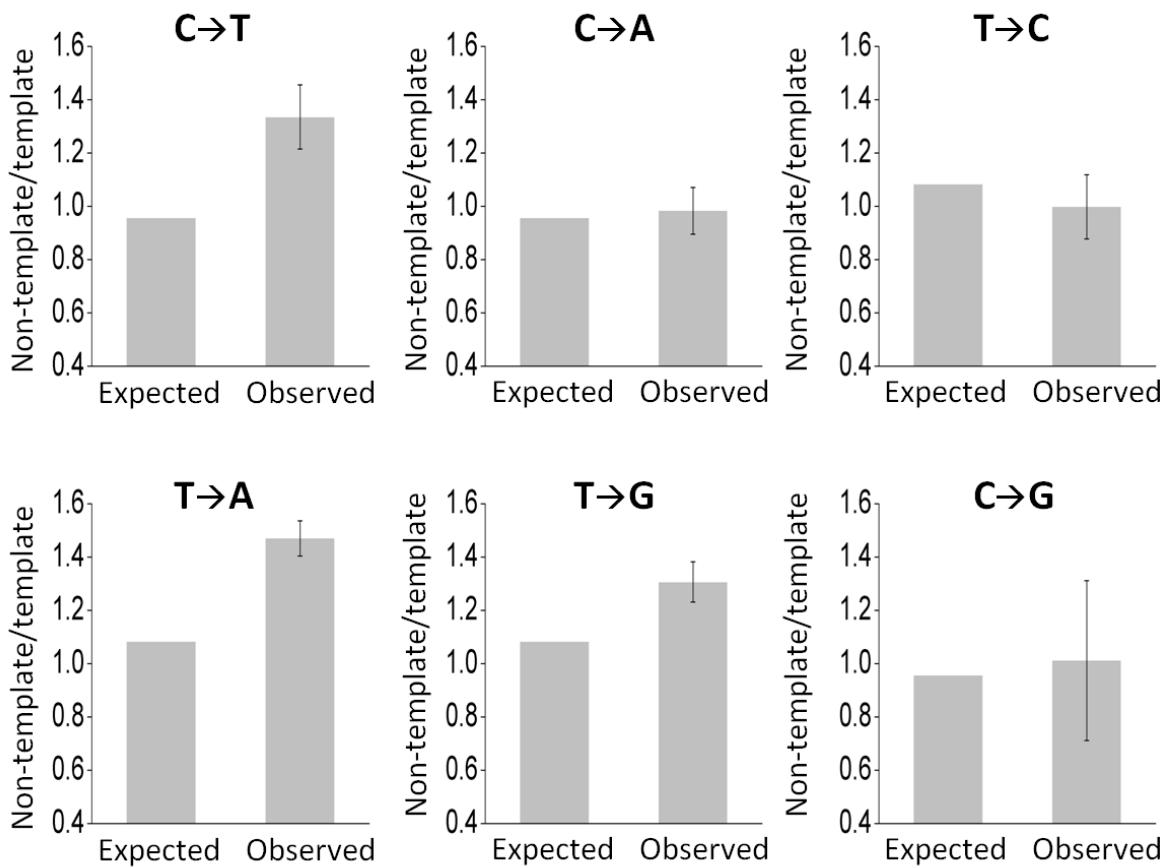


Fig. S15. Non-template-to-template ratio of all types of UV-induced mutations within transcribed regions. "Expected" column is the ratio simulated assuming random distribution of SNVs on both strands. "Observed" column is the ratio observed in UV-irradiated samples, with the standard error calculated from 4 kindred pairs.

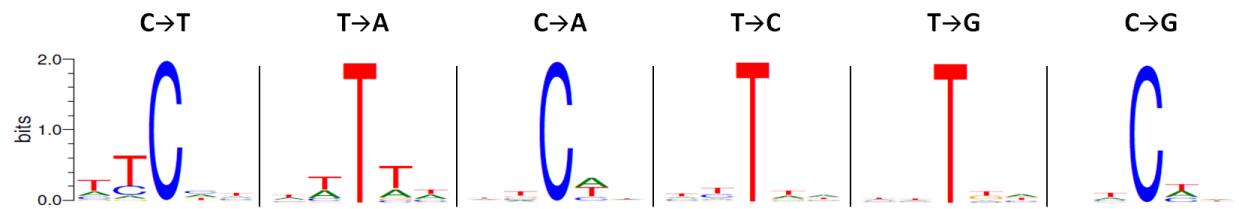


Fig. S16. Sequence context of each type of UV-induced mutation.

Method	Sample	Raw Data (Gb)	Coverage (%)	Allele Dropout Rate (ADO)	False Negative Rate (FNR)
Bulk	Bulk1*	119	*Bulk1 is the standard for comparison		
	Bulk2	138	99.9	0	0.003
LIANTI	BJ1	86	97.0	0.175	0.459
	BJ2	99	96.6	0.186	0.463
	BJ3	65	91.4	0.299	0.620
MDA (Qiagen)	Q1	82	88.1	0.329	0.609
	Q5	84	80.0	0.438	0.673
	Q9	90	91.8	0.249	0.540
MALBAC (Yikon)	YK1	91	72.0	0.477	0.708
	YK2	96	72.6	0.443	0.673
	YK5	96	73.4	0.436	0.659
DOP-PCR (Sigma)	S3	83	47.9	0.765	0.871
	S4	82	45.0	0.801	0.910
	S5	86	41.1	0.816	0.901
MDA (GE)	GE2	96	90.6	0.304	0.615
	GE4	129	92.0	0.281	0.588
	GE10	90	76.5	0.444	0.706
MALBAC-like (Rubicon)	R3	82	56.9	0.646	0.821
	R7	82	68.4	0.494	0.758
	R9	86	54.1	0.684	0.843

Table S1. Comparison of various single-cell WGA methods. For each WGA method, 3 single BJ cell samples are subject to whole genome amplification and sequencing.

References and Notes

1. L. Huang, F. Ma, A. Chapman, S. Lu, X. S. Xie, Single-cell whole-genome amplification and sequencing: Methodology and applications. *Annu. Rev. Genomics Hum. Genet.* **16**, 79–102 (2015). [doi:10.1146/annurev-genom-090413-025352](https://doi.org/10.1146/annurev-genom-090413-025352) Medline
2. C. Gawad, W. Koh, S. R. Quake, Single-cell genome sequencing: Current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016). [doi:10.1038/nrg.2015.16](https://doi.org/10.1038/nrg.2015.16) Medline
3. L. Yan, L. Huang, L. Xu, J. Huang, F. Ma, X. Zhu, Y. Tang, M. Liu, Y. Lian, P. Liu, R. Li, S. Lu, F. Tang, J. Qiao, X. S. Xie, Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by next-generation sequencing with linkage analyses. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15964–15969 (2015). [doi:10.1073/pnas.1523297113](https://doi.org/10.1073/pnas.1523297113) Medline
4. X. Ni, M. Zhuo, Z. Su, J. Duan, Y. Gao, Z. Wang, C. Zong, H. Bai, A. R. Chapman, J. Zhao, L. Xu, T. An, Q. Ma, Y. Wang, M. Wu, Y. Sun, S. Wang, Z. Li, X. Yang, J. Yong, X.-D. Su, Y. Lu, F. Bai, X. S. Xie, J. Wang, Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 21083–21088 (2013). [doi:10.1073/pnas.1320659110](https://doi.org/10.1073/pnas.1320659110) Medline
5. C. Zong, S. Lu, A. R. Chapman, X. S. Xie, Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012). [doi:10.1126/science.1229164](https://doi.org/10.1126/science.1229164) Medline
6. S. Lu, C. Zong, W. Fan, M. Yang, J. Li, A. R. Chapman, P. Zhu, X. Hu, L. Xu, L. Yan, F. Bai, J. Qiao, F. Tang, R. Li, X. S. Xie, Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* **338**, 1627–1630 (2012). [doi:10.1126/science.1229112](https://doi.org/10.1126/science.1229112) Medline
7. Y. Hou, W. Fan, L. Yan, R. Li, Y. Lian, J. Huang, J. Li, L. Xu, F. Tang, X. S. Xie, J. Qiao, Genome analyses of single human oocytes. *Cell* **155**, 1492–1506 (2013). [doi:10.1016/j.cell.2013.11.040](https://doi.org/10.1016/j.cell.2013.11.040) Medline
8. N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W. R. McCombie, J. Hicks, M. Wigler, Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011). [doi:10.1038/nature09807](https://doi.org/10.1038/nature09807) Medline
9. Y. Wang, J. Waters, M. L. Leung, A. Unruh, W. Roh, X. Shi, K. Chen, P. Scheet, S. Vattathil, H. Liang, A. Multani, H. Zhang, R. Zhao, F. Michor, F. Meric-Bernstam, N. E. Navin, Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014). [doi:10.1038/nature13600](https://doi.org/10.1038/nature13600) Medline
10. H. Telenius, N. P. Carter, C. E. Bebb, M. Nordenskjöld, B. A. J. Ponder, A. Tunnacliffe, Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics* **13**, 718–725 (1992). [doi:10.1016/0888-7543\(92\)90147-K](https://doi.org/10.1016/0888-7543(92)90147-K) Medline
11. F. B. Dean, S. Hosono, L. Fang, X. Wu, A. F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S. F. Kingsmore, M. Egholm, R. S. Lasken, Comprehensive

- human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5261–5266 (2002). [doi:10.1073/pnas.082089499](https://doi.org/10.1073/pnas.082089499) [Medline](#)
12. K. Leung, A. Klaus, B. K. Lin, E. Laks, J. Biele, D. Lai, A. Bashashati, Y.-F. Huang, R. Aniba, M. Moksa, A. Steif, A.-M. Mes-Masson, M. Hirst, S. P. Shah, S. Aparicio, C. L. Hansen, Robust high-performance nanoliter-volume single-cell multiple displacement amplification on planar substrates. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8484–8489 (2016). [doi:10.1073/pnas.1520964113](https://doi.org/10.1073/pnas.1520964113) [Medline](#)
13. I. Y. Goryshin, W. S. Reznikoff, Tn5 in vitro transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998). [doi:10.1074/jbc.273.13.7367](https://doi.org/10.1074/jbc.273.13.7367) [Medline](#)
14. R. N. Van Gelder, M. E. von Zastrow, A. Yool, W. C. Dement, J. D. Barchas, J. H. Eberwine, Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1663–1667 (1990). [doi:10.1073/pnas.87.5.1663](https://doi.org/10.1073/pnas.87.5.1663) [Medline](#)
15. A. Adey, H. G. Morrison, Asan, X. Xun, J. O. Kitzman, E. H. Turner, B. Stackhouse, A. P. MacKenzie, N. C. Caruccio, X. Zhang, J. Shendure, Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010). [doi:10.1186/gb-2010-11-12-r119](https://doi.org/10.1186/gb-2010-11-12-r119) [Medline](#)
16. J. A. Reuter, D. V. Spacek, R. K. Pai, M. P. Snyder, Simul-seq: Combined DNA and RNA sequencing for whole-genome and transcriptome profiling. *Nat. Methods* **13**, 953–958 (2016). [doi:10.1038/nmeth.4028](https://doi.org/10.1038/nmeth.4028) [Medline](#)
17. H. Zahn, A. Steif, E. Laks, P. Eirew, M. VanInsberghe, S. P. Shah, S. Aparicio, C. L. Hansen, Scalable whole-genome single-cell library preparation without preamplification. *Nat. Methods* **14**, 167–173 (2017). [doi:10.1038/nmeth.4140](https://doi.org/10.1038/nmeth.4140) [Medline](#)
18. N. Rhind, D. M. Gilbert, DNA replication timing. *Cold Spring Harb. Perspect. Biol.* **5**, a010132 (2013). [doi:10.1101/cshperspect.a010132](https://doi.org/10.1101/cshperspect.a010132) [Medline](#)
19. M. Frangkos, O. Ganier, P. Coulombe, M. Méchali, DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.* **16**, 360–374 (2015). [doi:10.1038/nrm4002](https://doi.org/10.1038/nrm4002) [Medline](#)
20. R. Lebofsky, R. Heilig, M. Sonnleitner, J. Weissenbach, A. Bensimon, DNA replication origin interference increases the spacing between initiation events in human cells. *Mol. Biol. Cell* **17**, 5337–5345 (2006). [doi:10.1091/mbc.E06-04-0298](https://doi.org/10.1091/mbc.E06-04-0298) [Medline](#)
21. B. Miotto, Z. Ji, K. Struhl, Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4810–E4819 (2016). [doi:10.1073/pnas.1609060113](https://doi.org/10.1073/pnas.1609060113) [Medline](#)
22. N. Van der Aa, J. Cheng, L. Mateiu, M. Zamani Esteki, P. Kumar, E. Dimitriadou, E. Vanneste, Y. Moreau, J. R. Vermeesch, T. Voet, Genome-wide copy number profiling of single cells in S-phase reveals DNA-replication domains. *Nucleic Acids Res.* **41**, e66 (2013). [doi:10.1093/nar/gks1352](https://doi.org/10.1093/nar/gks1352) [Medline](#)
23. R. S. Hansen, S. Thomas, R. Sandstrom, T. K. Canfield, R. E. Thurman, M. Weaver, M. O. Dorschner, S. M. Gartler, J. A. Stamatoyannopoulos, Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 139–144 (2010). [doi:10.1073/pnas.0912402107](https://doi.org/10.1073/pnas.0912402107) [Medline](#)

24. S. John, P. J. Sabo, T. K. Canfield, K. Lee, S. Vong, M. Weaver, H. Wang, J. Vierstra, A. P. Reynolds, R. E. Thurman, J. A. Stamatoyannopoulos, Genome-scale mapping of DNase I hypersensitivity. *Curr. Protoc. Mol. Biol.* **103**, 21.27.1–21.27.20 (2013). [doi:10.1002/0471142727.mb2127s103](https://doi.org/10.1002/0471142727.mb2127s103)
25. J. Wang, H. C. Fan, B. Behr, S. R. Quake, Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150**, 402–412 (2012). [doi:10.1016/j.cell.2012.06.030](https://doi.org/10.1016/j.cell.2012.06.030) [Medline](#)
26. M. A. Lodato, M. B. Woodworth, S. Lee, G. D. Evrony, B. K. Mehta, A. Karger, S. Lee, T. W. Chittenden, A. M. D’Gama, X. Cai, L. J. Luquette, E. Lee, P. J. Park, C. A. Walsh, Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015). [doi:10.1126/science.aab1785](https://doi.org/10.1126/science.aab1785) [Medline](#)
27. T. Strachan, J. Goodship, P. Chinnery, *Genetics and Genomics in Medicine* (Garland Science, New York, 2014).
28. A. Beletskii, A. S. Bhagwat, Transcription-induced mutations: Increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13919–13924 (1996). [doi:10.1073/pnas.93.24.13919](https://doi.org/10.1073/pnas.93.24.13919) [Medline](#)
29. M. Hofreiter, V. Jaenicke, D. Serre, A. von Haeseler, S. Pääbo, DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793–4799 (2001). [doi:10.1093/nar/29.23.4793](https://doi.org/10.1093/nar/29.23.4793) [Medline](#)
30. J. C. Shen, W. M. Rideout III, P. A. Jones, The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972–976 (1994). [doi:10.1093/nar/22.6.972](https://doi.org/10.1093/nar/22.6.972) [Medline](#)
31. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015). [doi:10.1098/rstb.2013.0624](https://doi.org/10.1098/rstb.2013.0624) [Medline](#)
32. K. C. Cheng, D. S. Cahill, H. Kasai, S. Nishimura, L. A. Loeb, 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G→T and A→C substitutions. *J. Biol. Chem.* **267**, 166–172 (1992). [Medline](#)
33. L. Chen, P. Liu, T. C. Evans Jr., L. M. Ettwiller, DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science* **355**, 752–756 (2017). [doi:10.1126/science.aai8690](https://doi.org/10.1126/science.aai8690) [Medline](#)
34. R. P. Rastogi, A. Richa, A. Kumar, M. B. Tyagi, R. P. Sinha, Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J. Nucleic Acids* **2010**, 592980 (2010). [doi:10.4061/2010/592980](https://doi.org/10.4061/2010/592980) [Medline](#)
35. J. O. Fuss, P. K. Cooper, DNA repair: Dynamic defenders against cancer and aging. *PLOS Biol.* **4**, e203 (2006). [doi:10.1371/journal.pbio.0040203](https://doi.org/10.1371/journal.pbio.0040203) [Medline](#)
36. A. Stary, P. Kannouche, A. R. Lehmann, A. Sarasin, Role of DNA polymerase η in the UV mutation spectrum in human cells. *J. Biol. Chem.* **278**, 18767–18775 (2003). [doi:10.1074/jbc.M211838200](https://doi.org/10.1074/jbc.M211838200) [Medline](#)
37. M. Krauthammer, Y. Kong, B. H. Ha, P. Evans, A. Bacchicocchi, J. P. McCusker, E. Cheng, M. J. Davis, G. Goh, M. Choi, S. Ariyan, D. Narayan, K. Dutton-Regester, A. Capatana,

- E. C. Holman, M. Bosenberg, M. Sznol, H. M. Kluger, D. E. Brash, D. F. Stern, M. A. Materin, R. S. Lo, S. Mane, S. Ma, K. K. Kidd, N. K. Hayward, R. P. Lifton, J. Schlessinger, T. J. Boggon, R. Halaban, Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat. Genet.* **44**, 1006–1014 (2012). [doi:10.1038/ng.2359](https://doi.org/10.1038/ng.2359) [Medline](#)
38. I. Martincorena, A. Roshan, M. Gerstung, P. Ellis, P. Van Loo, S. McLaren, D. C. Wedge, A. Fullam, L. B. Alexandrov, J. M. Tubio, L. Stebbings, A. Menzies, S. Widaa, M. R. Stratton, P. H. Jones, P. J. Campbell, High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015). [doi:10.1126/science.aaa6806](https://doi.org/10.1126/science.aaa6806) [Medline](#)
39. N. Saini, S. A. Roberts, L. J. Klimczak, K. Chan, S. A. Grimm, S. Dai, D. C. Fargo, J. C. Boyer, W. K. Kaufmann, J. A. Taylor, E. Lee, I. Cortes-Ciriano, P. J. Park, S. H. Schurman, E. P. Malc, P. A. Mieczkowski, D. A. Gordenin, The impact of environmental and endogenous damage on somatic mutation load in human skin fibroblasts. *PLOS Genet.* **12**, e1006385 (2016). [doi:10.1371/journal.pgen.1006385](https://doi.org/10.1371/journal.pgen.1006385) [Medline](#)
40. J. Hu, S. Adar, C. P. Selby, J. D. Lieb, A. Sancar, Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev.* **29**, 948–960 (2015). [doi:10.1101/gad.261271.115](https://doi.org/10.1101/gad.261271.115) [Medline](#)
41. G. Hendriks, F. Calléja, A. Besaratinia, H. Vrieling, G. P. Pfeifer, L. H. F. Mullenders, J. G. Jansen, N. de Wind, Transcription-dependent cytosine deamination is a novel mechanism in ultraviolet light-induced mutagenesis. *Curr. Biol.* **20**, 170–175 (2010). [doi:10.1016/j.cub.2009.11.061](https://doi.org/10.1016/j.cub.2009.11.061) [Medline](#)
42. P. Polak, R. Karlić, A. Koren, R. Thurman, R. Sandstrom, M. S. Lawrence, A. Reynolds, E. Rynes, K. Vlahoviček, J. A. Stamatoyannopoulos, S. R. Sunyaev, Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364 (2015). [doi:10.1038/nature14221](https://doi.org/10.1038/nature14221) [Medline](#)
43. P. Polak, M. S. Lawrence, E. Haugen, N. Stoletzki, P. Stojanov, R. E. Thurman, L. A. Garraway, S. Mirkin, G. Getz, J. A. Stamatoyannopoulos, S. R. Sunyaev, Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair. *Nat. Biotechnol.* **32**, 71–75 (2014). [doi:10.1038/nbt.2778](https://doi.org/10.1038/nbt.2778) [Medline](#)
44. B. Schuster-Böckler, B. Lehner, Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012). [doi:10.1038/nature11273](https://doi.org/10.1038/nature11273) [Medline](#)
45. R. Sabarinathan, L. Mularoni, J. Deu-Pons, A. Gonzalez-Perez, N. López-Bigas, Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016). [doi:10.1038/nature17661](https://doi.org/10.1038/nature17661) [Medline](#)
46. N. J. Haradhvala, P. Polak, P. Stojanov, K. R. Covington, E. Shinbrot, J. M. Hess, E. Rheinbay, J. Kim, Y. E. Maruvka, L. Z. Braunstein, A. Kamburov, P. C. Hanawalt, D. A. Wheeler, A. Koren, M. S. Lawrence, G. Getz, Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell* **164**, 538–549 (2016). [doi:10.1016/j.cell.2015.12.050](https://doi.org/10.1016/j.cell.2015.12.050) [Medline](#)

47. H. Vrieling, J. Venema, M.-L. van Rooyen, A. van Hoffen, P. Menichini, M. Z. Zdzienicka, J. W. I. M. Simons, L. H. F. Mullenders, A. A. van Zeeland, Strand specificity for UV-induced DNA repair and mutations in the Chinese hamster HPRT gene. *Nucleic Acids Res.* **19**, 2411–2415 (1991). [doi:10.1093/nar/19.9.2411](https://doi.org/10.1093/nar/19.9.2411) [Medline](#)
48. S. Picelli, Å. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014). [doi:10.1101/gr.177881.114](https://doi.org/10.1101/gr.177881.114) [Medline](#)
49. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013). [doi:10.1038/nmeth.2688](https://doi.org/10.1038/nmeth.2688) [Medline](#)
50. A. Adey, J. O. Kitzman, J. N. Burton, R. Daza, A. Kumar, L. Christiansen, M. Ronaghi, S. Amini, K. L. Gunderson, F. J. Steemers, J. Shendure, In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Res.* **24**, 2041–2049 (2014). [doi:10.1101/gr.178319.114](https://doi.org/10.1101/gr.178319.114) [Medline](#)
51. N. Arnaud-Barbe, V. Cheynet-Sauvion, G. Oriol, B. Mandrand, F. Mallet, Transcription of RNA templates by T7 RNA polymerase. *Nucleic Acids Res.* **26**, 3550–3554 (1998). [doi:10.1093/nar/26.15.3550](https://doi.org/10.1093/nar/26.15.3550) [Medline](#)
52. C. Cazenave, O. C. Uhlenbeck, RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6972–6976 (1994). [doi:10.1073/pnas.91.15.6972](https://doi.org/10.1073/pnas.91.15.6972) [Medline](#)
53. H. S. Zaher, P. J. Unrau, T7 RNA polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes. *Biochemistry* **43**, 7873–7880 (2004). [doi:10.1021/bi0497300](https://doi.org/10.1021/bi0497300) [Medline](#)
54. H. Li, <https://arxiv.org/abs/1303.3997> (2013).
55. V. Boeva, A. Zinovyev, K. Bleakley, J.-P. Vert, I. Janoueix-Lerosey, O. Delattre, E. Barillot, Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics* **27**, 268–269 (2011). [doi:10.1093/bioinformatics/btq635](https://doi.org/10.1093/bioinformatics/btq635) [Medline](#)
56. R. S. Lasken, T. B. Stockwell, Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol.* **7**, 19 (2007). [doi:10.1186/1472-6750-7-19](https://doi.org/10.1186/1472-6750-7-19) [Medline](#)
57. R. M. Layer, C. Chiang, A. R. Quinlan, I. M. Hall, LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014). [doi:10.1186/gb-2014-15-6-r84](https://doi.org/10.1186/gb-2014-15-6-r84) [Medline](#)
58. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). [doi:10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) [Medline](#)
59. A. M. Sidore, F. Lan, S. W. Lim, A. R. Abate, Enhanced sequencing coverage with digital droplet multiple displacement amplification. *Nucleic Acids Res.* **44**, e66 (2016). [doi:10.1093/nar/gkv1493](https://doi.org/10.1093/nar/gkv1493) [Medline](#)

60. L. Xu, I. L. Brito, E. J. Alm, P. C. Blainey, Virtual microfluidics for digital quantification and single-cell sequencing. *Nat. Methods* **13**, 759–762 (2016). [doi:10.1038/nmeth.3955](https://doi.org/10.1038/nmeth.3955) [Medline](#)
61. Y. Fu, C. Li, S. Lu, W. Zhou, F. Tang, X. S. Xie, Y. Huang, Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11923–11928 (2015). [doi:10.1073/pnas.1513988112](https://doi.org/10.1073/pnas.1513988112) [Medline](#)
62. J. Gole, A. Gore, A. Richards, Y.-J. Chiu, H.-L. Fung, D. Bushman, H.-I. Chiang, J. Chun, Y.-H. Lo, K. Zhang, Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat. Biotechnol.* **31**, 1126–1132 (2013). [doi:10.1038/nbt.2720](https://doi.org/10.1038/nbt.2720) [Medline](#)
63. S. A. Vitak, K. A. Torkenczy, J. L. Rosenkrantz, A. J. Fields, L. Christiansen, M. H. Wong, L. Carbone, F. J. Steemers, A. Adey, Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat. Methods* **14**, 302–308 (2017). [doi:10.1038/nmeth.4154](https://doi.org/10.1038/nmeth.4154) [Medline](#)
64. V. Ramani, X. Deng, R. Qiu, K. L. Gunderson, F. J. Steemers, C. M. Disteche, W. S. Noble, Z. Duan, J. Shendure, Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017). [doi:10.1038/nmeth.4155](https://doi.org/10.1038/nmeth.4155) [Medline](#)