HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2017 August 01.

Published in final edited form as:

Science. 2017 April 14; 356(6334): 189–194. doi:10.1126/science.aak9787.

Single-Cell Whole Genome Analyses by Linear Amplification via Transposon Insertion (LIANTI)

Chongyi Chen[1,*], **Dong Xing**[1,*], **Longzhi Tan**[1,*], **Heng Li**[2,*], **Guangyu Zhou**[1], **Lei Huang**[3,4], and **X. Sunney Xie**[1,3,4,†]

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[3]Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China

[4]Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing 100871, China

Abstract

Single-cell genomics is important for biology and medicine. However, current whole genome amplification (WGA) methods are limited by low accuracy of copy number variation (CNV) detection and low amplification fidelity. Here we report an improved single-cell WGA method, Linear Amplification via Transposon Insertion (LIANTI), which outperforms existing methods, enabling micro-CNV detection with kilobase resolution. This allowed direct observation of stochastic firing of DNA replication origins, different from cell to cell. We also show that the predominant cytosine-to-thymine mutations observed in single-cell genomics often arise from the artifact of cytosine deamination upon cell lysis. However, calling single nucleotide variations (SNVs) can be accomplished by sequencing kindred cells. We determined the spectrum of SNVs in a single human cell after ultraviolet radiation, revealing their nonrandom genome-wide distribution.

Rapid advances in DNA sequencing have led to a wealth of knowledge about genomes of various species including human, most of which have been derived from bulk measurements from a large number of cells. However, a single cell, has a unique genome even within an individual human being. For example, each germ cell is distinct, carrying different combinations of paternal and maternal genes. Somatic cells have spontaneous genomic changes that take place stochastically in time and genomic position. These include single-nucleotide variations (SNVs), copy number variations (CNVs) and structural variations (SVs). Such genomic changes can lead to cancer and other diseases. As such, characterization of single cell genomes has attracted increasing attention in recent years (1, 2). The importance of single-cell genomics becomes more apparent in the case of precious

[†]To whom correspondence should be addressed. xie@chemistry.harvard.edu.
[*]These authors contributed equally to the work.

and rare samples, such as embryonic cells and circulating tumor cells (3, 4), or when probing stochastic changes and cell-to-cell heterogeneity (5–9).

Due to the trace amount of genomic DNA, single-cell genome sequencing has relied on whole genome amplification (WGA). Among previous WGA methods, degenerate oligonucleotide-primed PCR (DOP-PCR) is an exponential PCR reaction with degenerate priming (10). Multiple displacement amplification (MDA) uses a strand-displacing DNA polymerase to exponentially amplify single-stranded DNA into a hyperbranched structure. (11, 12). Multiple annealing and looping-based amplification cycles (MALBAC) employs quasi-linear amplification through looping-based amplicon protection followed by PCR (5). All these methods involve nonspecific priming and exponential amplification that create amplification bias and errors.

To reduce such bias and errors, we have developed a new WGA method, Linear Amplification via Transposon Insertion (LIANTI), which combines Tn5 transposition (13) and T7 in vitro transcription (14) for single-cell genomic analyses. Random fragmentation and tagging of genomic DNA by Tn5 transposition has been used to prepare DNA sequencing libraries by introducing priming sites for PCR amplification (15). However, such exponential amplification is associated with amplification bias and errors, limiting its applications in single-cell genomics (16, 17). Here we demonstrate linear amplification, whose advantage over exponential amplification is illustrated in Fig. 1A.

In LIANTI, genomic DNA from a single cell is randomly fragmented by Tn5 transposition of a specially designed LIANTI transposon that includes a T7 promoter(Fig. 1B). Genomic DNA fragments tagged by T7 promoters are linearly amplified into thousands of copies of RNAs through in vitro transcription, followed by reverse transcription and second strand synthesis into double-stranded LIANTI amplicons ready for DNA library preparation (Fig. 1C). LIANTI eliminates nonspecific priming and exponential amplification used in other single-cell WGA methods, greatly reducing amplification bias and errors.

We used LIANTI to amplify genomic DNA from single BJ cells, a human diploid cell line from skin fibroblasts chosen for no aneuploidy. Single-cell genomic DNA was randomly fragmented by Tn5 transposition to give an average fragment size of ~400 bp ((Fig. S1A-B and Fig. S2A-B). Following overnight in vitro transcription, we routinely acquire ~20 nanograms of LIANTI amplicons for DNA library preparation. We sequenced BJ cells at ~30x depth, and performed a systematic comparison between LIANTI and previous WGA methods (data from (1)). LIANTI achieves 97% genome coverage and a 17% allele dropout rate (ADO), outperforming other WGA methods (Table S1).

To evaluate amplification uniformity, we plotted the average read depths in 1-Mb bins across the genome for LIANTI, MDA, MALBAC and DOP-PCR, together with a zoom-in to a 10-Mb region on chromosome 1 with 10-Kb bins (Fig. 1D). On both scales, LIANTI exhibits the highest amplification uniformity compared to the other methods. To better quantify amplification bias on all scales, we plotted the coefficient of variation (CV) of the read depth along the genome as a function of the bin size (Fig. 1E, Fig. S3A), which is more reproducible and informative than power spectra and Lorenz curves (Fig. S3B) used

previously (see supplementary materials). LIANTI achieves the lowest CV values with respect to all bin sizes, offering the highest accuracy for CNV detection.

The spatial resolution of CNV detection in a single cell has been limited to ~1 Mb due to the amplification noise of previous WGA methods. In LIANTI, amplification noise, though much reduced, still exists due to different amplification factors for each fragment, preventing accurate detection of micro-CNVs (< 100 Kb CNVs). To further reduce this noise, instead of relying on read depths, we carry out digital counting of the inferred fragment numbers, as shown in Fig. 2A. This is done by taking advantage of the fact that LIANTI amplicons mapped to the reference genome with the same ends should originate from the same genomic DNA fragment, hence allowing more accurate inference of the fragment numbers at each genomic position. For example, in Fig. 2B, the unamplified bulk (top panel) shows a 2-to-1 copy number loss. However, the LIANTI single-cell read depth raw data (middle panel) obscures this micro-CNV. The inferred fragment number by the digital-counting analysis (bottom panel) better resolved the micro-CNV. Digital counting improves the resolution of micro-CNV detection to ~10 Kb. We characterized the false positives and false negatives for micro-CNV detection in a single BJ cell (Fig. S4A-B). Results differ for copy number gains, 2-to-0 copy number losses, and 2-to-1 copy number losses (Fig. S4C-D). none of which were possible by previous WGA methods at this resolution.

We took advantage of LIANTI's capability to detect micro-CNVs to probe DNA replication, an important question in biology. In particular, whether the firing of replication origins and replicon formation (~50-120 Kb) are stochastic has been a subject of intensive investigation (18–21), which can be best answered by single-cell measurements. Recently, MDA was used to probe single-cell DNA replication (22), but was unable to resolve individual replicons due to its poor spatial resolution.

Here we show in Figure. 2C whole genome sequencing with LIANTI for 11 BJ cells picked from a synchronized population in early S-phase. The genome-wide replication origin firing and replicon formation events were detected by the copy number gain of 2→3 and 3→4 with kibobase resolution (Fig. 2C, Fig. S6). The genome-wide replicon copy numbers of a single cell correlate well with the conventional bulk readouts of the Repli-Seq assay (23) (Fig. 2D, Fig. S7) and the DNase I hypersensitive assay (24) (Fig. 2D, Fig. S8), suggesting a subset of replication origins are used in individual cells. Figure 2E shows the correlation plots of replicon copy numbers between pairs of single cells close in replication progress in S-phase (Fig. 2E, Fig. S9). While the diagonal signal represents replicons shared by both cells (deterministic), the strong off-diagonal signal suggests a large degree of stochasticity in terms of replication origin firing, which is different from cell to cell.

In terms of SNV detection accuracy, among all WGA methods, LIANTI gives the lowest false positive rate (FPR) of 5.4×10^{-6} for single-BJ-cell SNV detection (Fig. 3A, Fig. S10A), which is still higher than the anticipation from linear amplification. We further characterized the mutation spectra of false positives, and found both LIANTI and MDA exhibit a C→T false positive predominance, which is not seen in the unamplified bulk (Fig. 3B). We note that such “de novo” C→T mutations have reported in many previous single-cell genomic studies (25, 26) and most recently in non-dividing neurons (26).

We instead attribute this predominant C→T observation to the experimental artifact of C→U deamination after cell lysis, which is well known as the most common cause of point mutations (27, 28) and is especially prominent in ancient DNA (29). C→U deamination is a natural process that occurs at a low rate randomly in the genome (30), hence would be difficult to see in bulk sequencing due to the extremely low allele frequency. To test whether C→T false positive predominance in LIANTI is caused by C→U deamination, we treated genomic DNA from a lysed cell before LIANTI amplification with uracil-DNA glycosylase (UDG), which functions as part of the DNA repair system in live cells, in order to eliminate cytosine-deaminated uracil bases (31). Indeed, a significant reduction of C→T SNVs were observed (Fig. 3B, Fig. S10B), proving that the commonly observed C→T SNV predominance in the field of single-cell genomics is caused by in vitro cytosine deamination artifact and is a false positive.

Likewise, the second most frequent false positive is A→G (Fig. 3B), which happens to be the second most common spontaneous mutation of DNA bases due to adenine deamination (27). Another common type of false positive is G→T (Fig. 3B), which is likely caused by guanine oxidation to 8-hydroxyguanine (32, 33). We concluded that the accuracy of single-cell SNV detection for any WGA methods is fundamentally limited by chemical instability of DNA bases in the absence of cellular DNA repair systems. As a result, sequencing two kindred cells (5), which are a pair of cells derived from the division of a single cell, is necessary to filter out such false positives occurring randomly in the genome.

We further demonstrate the use of LIANTI for the study of mutations generated by ultraviolet (UV) radiation. It is well known that exposure to UV radiation in sunlight leads to DNA damage and potential skin cancer, attracting many mechanistic studies. UV radiation generates cyclobutane pyrimidine dimers (CPDs) and 6,4 photoproducts (PPs) on genomic DNA (34), which are subject to nucleotide excision repair (NER) (35). If the damage is not repaired before DNA replication, error-prone translesion synthesis DNA polymerase is recruited to the damaged region, giving rise to de novo SNVs (36). However, these mutations are different from cell to cell due to the randomness of UV damage along the genome, which necessitates single-cell whole genome amplification and sequencing.

To characterize UV-induced genome-wide mutations, we exposed human skin fibroblast BJ cells to different UV doses. After propagating several cell cyles without UV, a single cell under investigation was cultured to generate a pair of kindred cells, which were subject to LIANTI and sequencing in order to eliminate false positive SNVs (Fig. 4A).

We detected 4700-9300 UV-induced SNVs throughout the genome from each pair of kindred cells (Fig. S12). The SNV spectra show a C→T predominance (Fig. 4B, Fig. S13), in good agreement with the previously reported SNV spectra of sun-exposed normal human skin and melanomas (37–39). While examining the point mutation distribution along the genome, we discovered a depletion of mutations within transcribed regions (Fig. 4C), which can be explained by the involvement of transcription-coupled NER (40, 41). We also observed a significant depletion within DNase I hypersensitive sites and early-replicating regions (Fig. 4C). When plotting throughout the genome, we observed a strong anti-correlation between the density of UV-induced SNVs and Repli-Seq signal reflecting the

replicated genomic regions, as well as the DNase I hypersensitive signal (Fig. 4D, Fig. S14). Similar phenomena have also been observed in cancer genomes without UV radiation (42–44), which was attributed to NER impairment by DNA-bound proteins (43, 45).

We further examined the propensity of mutations for the two strands within transcribed regions, and observed a C→T enrichment in the non-template strand (Fig. 4E). The same enrichment was also observed in UV-associated cancer genomes (46), which can be explained by the preferred CPD/PP removal by transcription-coupled NER on the template strand (40, 47). When plotting the sequence context of C→T mutations, the adjacent base is mostly T on the 5' side (Fig. 4E), consistent with the well-known mechanism of UV-induced CPD/PP formation of T:C, followed by error-prone translesion synthesis (36). Interestingly, we also observed an enrichment of T→A in the non-template strand (Fig. 4E, Fig. S15), suggesting the involvement of transcription-coupled NER as well. We further plotted the sequence context of T→A mutations, and found the adjacent base is mostly T on both sides (Fig. 4E, Fig. S16), suggesting that T→A may be caused by UV-induced CPD/PP of T:T, followed by a different kind of error-prone translesion synthesis.

We note high throughput sequencing of many single cells can be easily achieved by adding combinatorial cellular barcodes in the LIANTI transposon and primer. In addition to fundamental investigations illustrated, the high precision of micro-CNVs detection and the ability to call individual SNVs in a single cell will allow better genetic screening in reproductive medicine and provide unprecedented information of how genome variation takes place in cancer and other diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The LIANTI development was supported by U.S. National Institutes of Health (NIH) Director's Pioneer Award (5DP1CA186693), and the study of DNA damage by UV radiation was supported by a National Cancer Institute grant (5R33CA174560). The comparison with other methods and the study of micro CNVs were supported by Beijing Municipal Science & Technology Commission Grants (D1511000024150002 to X.S.X.), National Key Technologies R&D program (2016YFC0900100 to L.H.), and the funding from Beijing Advanced Innovation Center for Genomics at Peking University. L.T. was supported by an HHMI International Student Research Fellowship. We thank Yi Yin, Pin Cui, Alec Chapman, Yaqiong Tang and other members in the group for their assistance and helpful discussions. Raw sequencing data were deposited at the National Center for Biotechnology Information with accession number SRP102259 at the link: http://www.ncbi.nlm.nih.gov/sra/SRP102259. X.S.X., C.C, X.D. are inventors on a patent application filed by Harvard University that covers the single cell whole genome sequencing by Linear Amplification via Transposon Insertion (LIANTI) technology.

References

1. Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. Annu Rev Genomics Hum Genet. 2015; 16:79–102. [PubMed: 26077818]
2. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet. 2016
3. Yan L, et al. 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. 2015; 112:15964–15969. [PubMed: 26712022]

4. Ni X, et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc Natl Acad Sci U S A. 2013; 110:21083–21088. [PubMed: 24324171]
5. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2012; 338:1622–1626. [PubMed: 23258894]
6. Lu S, et al. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. Science. 2012; 338:1627–1630. [PubMed: 23258895]
7. Hou Y, et al. Genome analyses of single human oocytes. Cell. 2013; 155:1492–1506. [PubMed: 24360273]
8. Navin N, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011; 472:90–94. [PubMed: 21399628]
9. Wang Y, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature. 2014; 512:155–160. [PubMed: 25079324]
10. Telenius H, et al. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics. 1992; 13:718–725. [PubMed: 1639399]
11. Dean FB, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A. 2002; 99:5261–5266. [PubMed: 11959976]
12. Leung K, et al. Robust high-performance nanoliter-volume single-cell multiple displacement amplification on planar substrates. Proc Natl Acad Sci U S A. 2016; 113:8484–8489. [PubMed: 27412862]
13. Goryshin IY, Reznikoff WS. Tn5 in vitro transposition. J Biol Chem. 1998; 273:7367–7374. [PubMed: 9516433]
14. Van Gelder RN, et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A. 1990; 87:1663–1667. [PubMed: 1689846]
15. Adey A, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. 2010; 11:R119. [PubMed: 21143862]
16. Reuter JA, Spacek DV, Pai RK, Snyder MP. Simul-seq: combined DNA and RNA sequencing for whole-genome and transcriptome profiling. Nat Methods. 2016; 13:953–958. [PubMed: 27723755]
17. Zahn H, et al. Scalable whole-genome single-cell library preparation without preamplification. Nat Methods. 2017
18. Rhind N, Gilbert DM. DNA replication timing. Cold Spring Harb Perspect Biol. 2013; 5:a010132. [PubMed: 23838440]
19. Fragkos M, Ganier O, Coulombe P, Mechali M. DNA replication origin activation in space and time. Nat Rev Mol Cell Biol. 2015; 16:360–374. [PubMed: 25999062]
20. Lebofsky R, Heilig R, Sonnleitner M, Weissenbach J, Bensimon A. DNA replication origin interference increases the spacing between initiation events in human cells. Mol Biol Cell. 2006; 17:5337–5345. [PubMed: 17005913]
21. Miotto B, Ji Z, Struhl K. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc Natl Acad Sci U S A. 2016; 113:E4810–4819. [PubMed: 27436900]
22. Van der Aa N, et al. Genome-wide copy number profiling of single cells in S-phase reveals DNA-replication domains. Nucleic Acids Res. 2013; 41:e66. [PubMed: 23295674]
23. Hansen RS, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A. 2010; 107:139–144. [PubMed: 19966280]
24. John S, et al. Genome-scale mapping of DNase I hypersensitivity. Curr Protoc Mol Biol. 2013 Chapter 27, Unit 21 27.
25. Wang J, Fan HC, Behr B, Quake SR. Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm. Cell. 2012; 150:402–412. [PubMed: 22817899]
26. Lodato MA, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. Science. 2015; 350:94–98. [PubMed: 26430121]
27. Tom Strachan JG, Chinnery Patrick. Genetics and Genomics in Medicine. 2014

28. Beletskii A, Bhagwat AS. 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. 1996; 93:13919–13924. [PubMed: 8943036]
29. Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Paabo S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Res. 2001; 29:4793–4799. [PubMed: 11726688]
30. Shen JC, Rideout WM 3rd, Jones PA. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res. 1994; 22:972–976. [PubMed: 8152929]
31. Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos Trans R Soc Lond B Biol Sci. 2015; 370:20130624. [PubMed: 25487342]
32. Cheng KC, Cahill DS, Kasai H, Nishimura S, Loeb LA. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. J Biol Chem. 1992; 267:166–172. [PubMed: 1730583]
33. Chen L, Liu P, Evans TC Jr, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science. 2017; 355:752–756. [PubMed: 28209900]
34. Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. J Nucleic Acids. 2010; 2010:592980. [PubMed: 21209706]
35. Fuss JO, Cooper PK. DNA repair: dynamic defenders against cancer and aging. PLoS Biol. 2006; 4:e203. [PubMed: 16752948]
36. Stary A, Kannouche P, Lehmann AR, Sarasin A. Role of DNA polymerase eta in the UV mutation spectrum in human cells. J Biol Chem. 2003; 278:18767–18775. [PubMed: 12644471]
37. Krauthammer M, et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. Nat Genet. 2012; 44:1006–1014. [PubMed: 22842228]
38. Martincorena I, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015; 348:880–886. [PubMed: 25999502]
39. Saini N, et al. The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. PLoS Genet. 2016; 12:e1006385. [PubMed: 27788131]
40. Hu J, Adar S, Selby CP, Lieb JD, Sancar A. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. Genes Dev. 2015; 29:948–960. [PubMed: 25934506]
41. Hendriks G, et al. Transcription-dependent cytosine deamination is a novel mechanism in ultraviolet light-induced mutagenesis. Curr Biol. 2010; 20:170–175. [PubMed: 20045328]
42. Polak P, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature. 2015; 518:360–364. [PubMed: 25693567]
43. Polak P, et al. Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair. Nat Biotechnol. 2014; 32:71–75. [PubMed: 24336318]
44. Schuster-Bockler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature. 2012; 488:504–507. [PubMed: 22820252]
45. Sabarinathan R, Mularoni L, Deu-Pons J, Gonzalez-Perez A, Lopez-Bigas N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature. 2016; 532:264–267. [PubMed: 27075101]
46. Haradhvala NJ, et al. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. Cell. 2016; 164:538–549. [PubMed: 26806129]
47. Vrieling H, et al. Strand specificity for UV-induced DNA repair and mutations in the Chinese hamster HPRT gene. Nucleic Acids Res. 1991; 19:2411–2415. [PubMed: 1674998]
48. Picelli S, et al. Tn5 transposase and tagmentation procedures for massively-scaled sequencing projects. Genome Res. 2014
49. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013
50. Adey A, et al. In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. Genome Res. 2014; 24:2041–2049. [PubMed: 25327137]

51. Arnaud-Barbe N, Cheynet-Sauvion V, Oriol G, Mandrand B, Mallet F. Transcription of RNA templates by T7 RNA polymerase. Nucleic Acids Res. 1998; 26:3550–3554. [PubMed: 9671817]
52. Cazenave C, Uhlenbeck OC. RNA template-directed RNA synthesis by T7 RNA polymerase. Proc Natl Acad Sci U S A. 1994; 91:6972–6976. [PubMed: 7518923]
53. Zaher HS, Unrau PJ. T7 RNA polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes. Biochemistry. 2004; 43:7873–7880. [PubMed: 15196031]
54. Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. Annu Rev Genomics Hum Genet. 2015; 16:79–102. [PubMed: 26077818]
55. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv. 2013; 1303.3997v2
56. Boeva V, et al. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. Bioinformatics. 2011; 27:268–269. [PubMed: 21081509]
57. Lasken RS, Stockwell TB. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnol. 2007; 7:19. [PubMed: 17430586]
58. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014; 15:R84. [PubMed: 24970577]
59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
60. Sidore AM, Lan F, Lim SW, Abate AR. Enhanced sequencing coverage with digital droplet multiple displacement amplification. Nucleic Acids Res. 2016; 44:e66. [PubMed: 26704978]
61. Xu L, Brito IL, Alm EJ, Blainey PC. Virtual microfluidics for digital quantification and single-cell sequencing. Nat Methods. 2016; 13:759–762. [PubMed: 27479330]
62. Leung K, et al. Robust high-performance nanoliter-volume single-cell multiple displacement amplification on planar substrates. Proc Natl Acad Sci U S A. 2016; 113:8484–8489. [PubMed: 27412862]
63. Fu Y, et al. Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. Proc Natl Acad Sci U S A. 2015; 112:11923–11928. [PubMed: 26340991]
64. Gole J, et al. Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. Nat Biotechnol. 2013; 31:1126–1132. [PubMed: 24213699]
65. Wang Y, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature. 2014; 512:155–160. [PubMed: 25079324]
66. Vitak SA, et al. Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods. 2017; 14:302–308. [PubMed: 28135258]
67. Ramani V, et al. Massively multiplex single-cell Hi-C. Nat Methods. 2017; 14:263–266. [PubMed: 28135255]
68. Hansen RS, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A. 2010; 107:139–144. [PubMed: 19966280]
69. John S, et al. Genome-scale mapping of DNase I hypersensitivity. Curr Protoc Mol Biol. 2013 Chapter 27, Unit 21 27.
70. Lodato MA, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. Science. 2015; 350:94–98. [PubMed: 26430121]

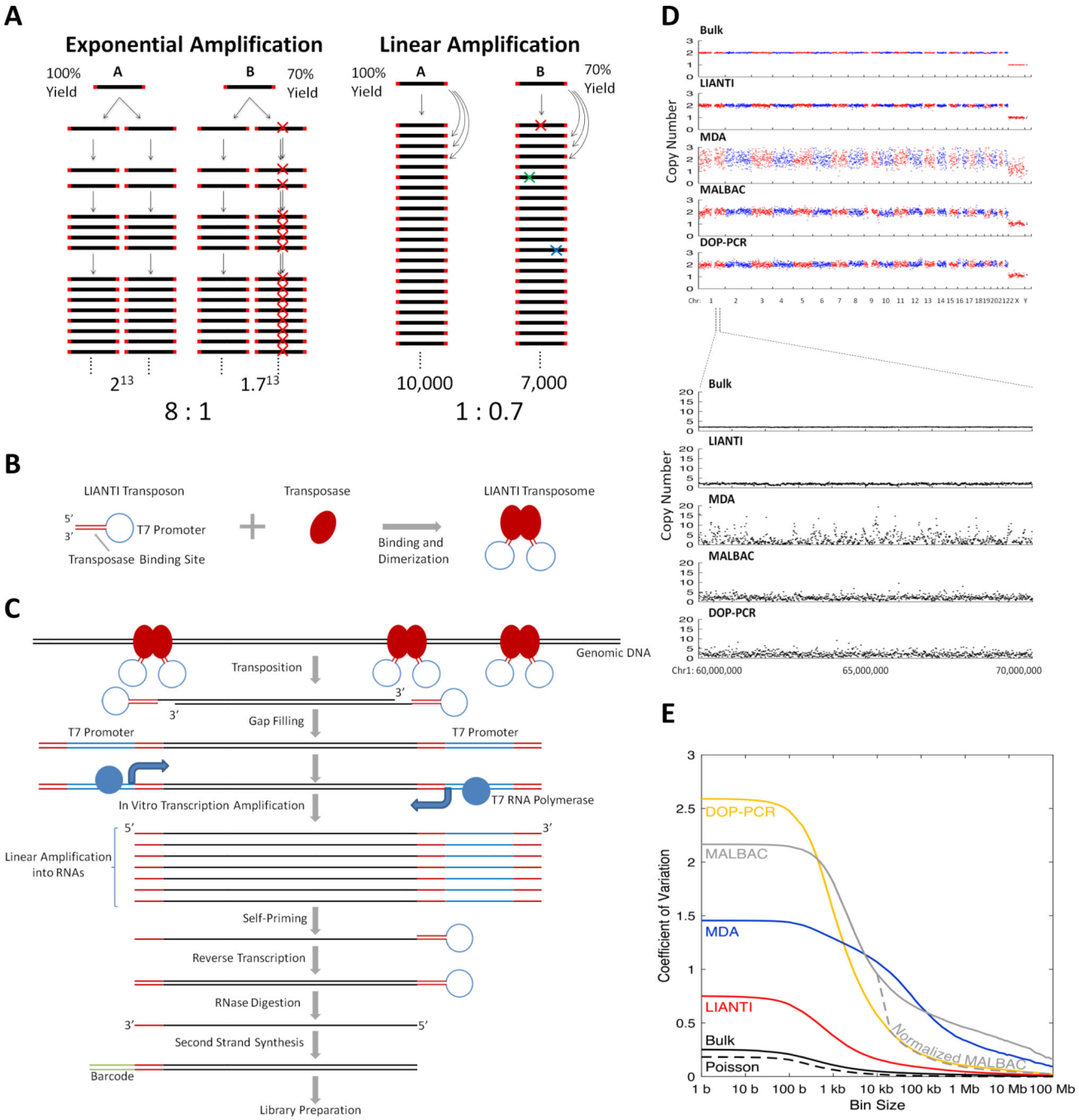

Fig. 1.

LIANTI single-cell whole genome amplification scheme and amplification uniformity. (**A**) Comparison of exponential and linear amplification. Assuming the DNA fragments A and B have replication yields of 100% and 70% per round, respectively. For a final amplification factor of ~10,000 of fragment A, exponential amplification results in a ratio of 8 : 1, hampering the accuracy of CNV detection. In contrast, linear amplification exhibits a much smaller ratio of 1 : 0.7. Linear amplification is also superior to exponential amplification in fidelity. In exponential amplification, a polymerase of the highest fidelity (10^{-7}) replicating

the human genome (3×10^9 bp) in the first cycle would give ~300 errors, which will be propagated permanently in the next replication cycles, leading to false positive SNVs. In contrast, in linear amplification, the errors would appear randomly at different locations in the amplicons and can be easily filtered out. (**B**) LIANTI transposon and transposome. LIANTI transposon consists of a 19-bp double-stranded transposase binding site and a single-stranded T7 promoter loop. Equal molar of LIANTI transposon and Tn5 transposase are mixed and dimerized to form LIANTI transposome. (**C**) LIANTI scheme. Genomic DNA from a single cell is randomly fragmented and tagged by LIANTI transposon, followed by DNA polymerase gap extension to convert single-stranded T7 promoter loops into double-stranded T7 promoters on both ends of each fragment. In vitro transcription overnight is performed to linearly amplify the genomic DNA fragments into genomic RNAs which are capable of self-priming on the 3' end. After reverse transcription, RNase digestion and second strand synthesis, double-stranded LIANTI amplicons tagged with unique molecular barcodes are formed, representing the amplified product of the original genomic DNA from a single cell, and ready for DNA library preparation and next generation sequencing. (**D**) Read depths across the genome with 1-Mb bin size, and a zoom in to a 10-Mb region (Chr1:60,000,000-70,000,000) with 10-Kb bin size. The MALBAC data is normalized by the average of two other MALBAC cells to remove the sequence-dependent bias reproducible from cell to cell. (**E**) Coefficient of variation for read depths along the genome as a function of bin sizes from 1 bp to 100 Mb, showing amplification noise on all scales for single-cell WGA methods, including DOP-PCR, MDA, MALBAC, and LIANTI. The normalized MALBAC data (dashed) is shown together with the unnormalized MALBAC data. Only the unnormalized data of the other methods are shown as no significant improvement by normalization were observed. Poisson curve is the expected coefficient of variation for read depth assuming only Poisson noise. LINATI exhibits a much improved amplification uniformity over the previous methods on all scales.

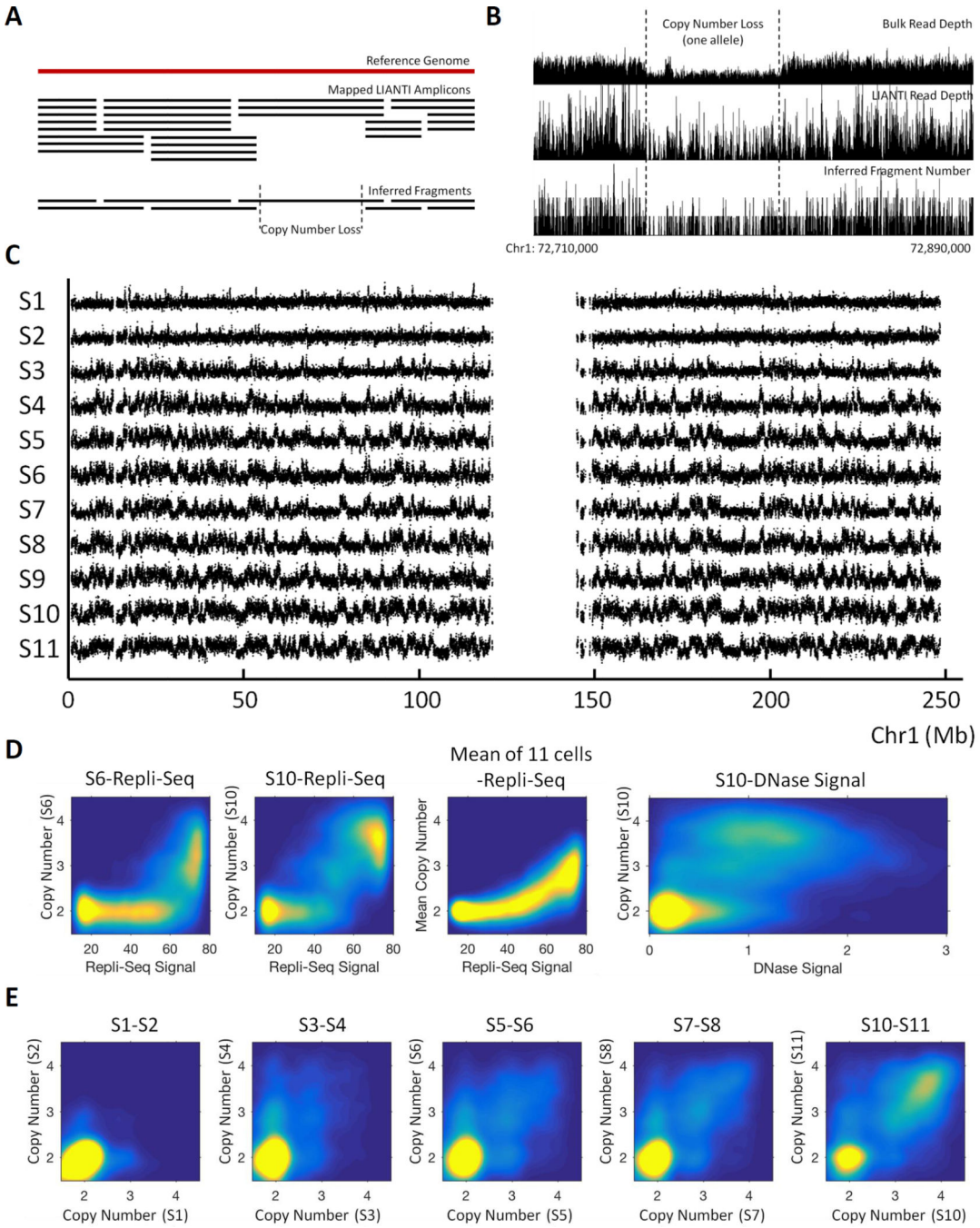

Fig. 2.

Genome-wide detection of micro-CNVs and replication origin firing events in single BJ cells. (**A**) Principle for the inference of fragment numbers by LIANTI. Single-cell LIANTI amplicons mapped to the same starting and ending coordinates on the reference genome are grouped as originated from one fragment of the genomic DNA. This allows for the correction of the different amplification efficiency, often size dependent, for each fragment. The digital counting of the inferred fragment number across the genome is shown here for 2-to-1 copy number loss. (**B**) Example of a 57-Kb 2-to-1 micro-CNV detected in a single BJ

cell, plotted with 100-bp bin size. Top panel is the read depth from unamplified bulk sequencing showing the existence of the micro-CNV. Middle panel is the read depth of the single-cell LIANTI amplicons, which obscures the micro-CNV due to amplification noise at this resolution. Bottom panel shows the inferred fragment number by LIANTI digital-counting analysis, which recovers the micro-CNV in the single cell. (**C**) Genome-wide detection of replication origin firing and replicon formation based on the copy number gain in 11 single cells with 10-Kb bin size (~250 Mb Chr1 shown in the plot). (**D**) Correlation plots of single-cell replicon copy numbers with the bulk readouts of the Repli-Seq assay and the DNase I hypersensitive assay using 100-Kb bin size. (**E**) Correlation plots of replicon copy numbers between pairs of single cells close in replication progress in S-phase using 100-Kb bin size. The diagonal signal represents replicon copy numbers shared by both cells, and the off-diagonal signal suggests stochastic origin firing and replicon formation, which is different from cell to cell.

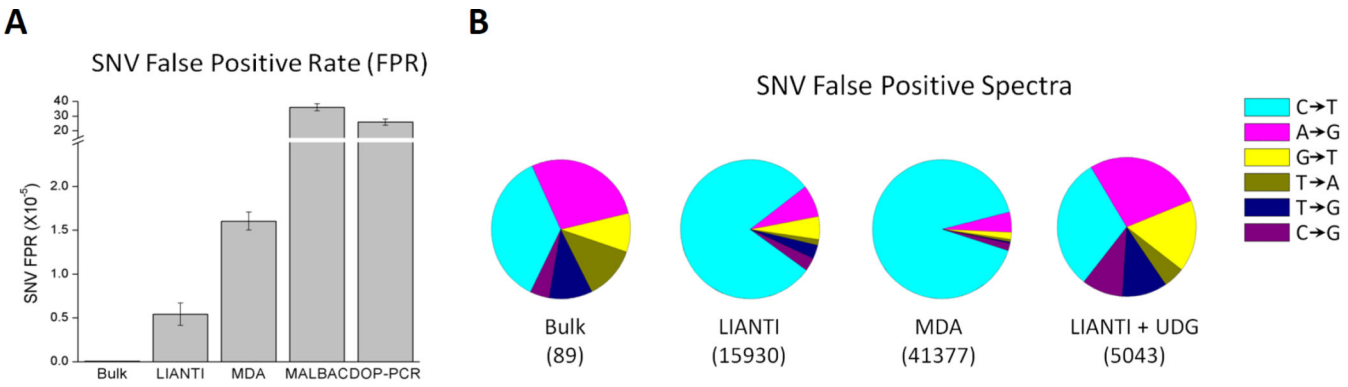

Fig. 3.

Detection of SNVs in single BJ cells. (**A**) False positive rates of SNV detection in a single BJ cell. The error bars were calculated from three different BJ cells. (**B**) Spectra of SNV false positives in unamplified bulk, single-cell LIANTI, single-cell MDA and single-cell UDG-treated LIANTI samples. The number of false positives is shown in the bracket for each sample. Both LIANTI and MDA results exhibit predominant C→T false positives not seen in the unamplified bulk. Similar C→T SNVs have been reported in previous single-cell MDA studies and attributed to de novo mutations (26). We attribute the phenomenon to the spontaneous C→U deamination upon cell lysis, which is often seen in ancient DNA bulk samples. We prove that such C→T deamination accounts for the observed SNV false positives by WGA of the cell lysate treated with uracil-DNA glycosylase (UDG), which eliminates cytosine-deaminated uracil bases and hence recovers the reduced C→T false positive fraction in the bulk.

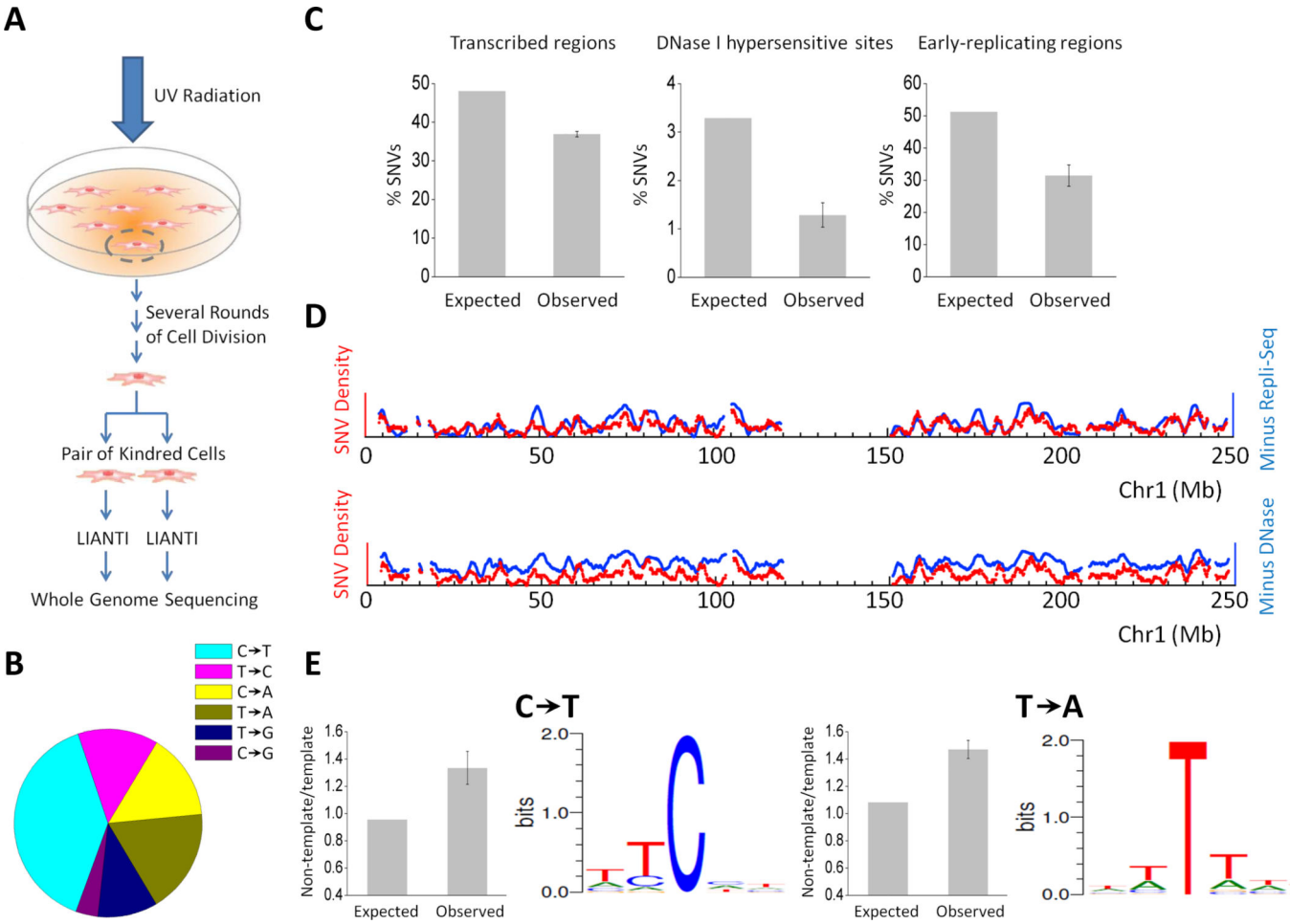

Fig. 4.

Genome-wide profiling of UV-induced mutations in single BJ cells. (**A**) Experimental design. BJ cells cultivated in dishes are exposed to UV radiation at a dose of 5, 15 and 30 J/m^2, respectively. Single cells that survived cell cycle arrest and apoptosis were picked and allowed to divide into multiple kindred cells (Fig. S11), among which a pair of kindred cells are picked for LIANTI. (**B**) Spectra of UV-induced SNVs in a representative cell exposed to 15 J/m^2 UV radiation. (**C**) Depletion of UV-induced SNVs within transcribed regions, DNase I hypersensitive sites and early-replicating regions. "Expected" column is the percentage of SNVs simulated assuming random distribution along the genome. "Observed" column is the percentage of SNVs observed in UV-radiated samples, with the error bars calculated from four kindred pairs. (**D**) Overlay of the density of UV-induced SNVs (red) and the minus Repli-Seq signal (blue) reflecting the replicated genomic regions, as well as the minus DNase I hypersensitive signal (blue) throughout the genome (~250 Mb Chr1 shown in the plot). Both signals were calculated in 2-Mb moving windows with 100-kb increments. (**E**) Non-template-to-template ratio of UV-induced C→T and T→A mutations within transcribed regions, and the sequence context of such mutations. "Expected" column is the ratio simulated assuming random distribution of SNVs on both strands. "Observed" column is the ratio observed in UV-radiated samples, with the error bars calculated from

four kindred pairs. Sequence context is plotted based on the frequency of each base next to the corresponding type of mutation.