Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell

Chenghang Zong *et al.*

Science **338**, 1622 (2012);
DOI: 10.1126/science.1229164

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

genes in extremely large cohorts, as may be required for the definitive implication of rare variants or de novo mutations in any genetically complex disorder.

References and Notes

1. G. V. Kryukov, A. Shpunt, J. A. Stamatoyannopoulos, S. R. Sunyaev, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3871 (2009).
2. B. J. O'Roak *et al.*, *Nat. Genet.* **43**, 585 (2011).
3. B. J. O'Roak *et al.*, *Nature* **485**, 246 (2012).
4. S. J. Sanders *et al.*, *Nature* **485**, 237 (2012).
5. B. M. Neale *et al.*, *Nature* **485**, 242 (2012).
6. I. Iossifov *et al.*, *Neuron* **74**, 285 (2012).
7. E. H. Turner, C. Lee, S. B. Ng, D. A. Nickerson, J. Shendure, *Nat. Methods* **6**, 315 (2009).
8. G. J. Porreca *et al.*, *Nat. Methods* **4**, 931 (2007).
9. S. Krishnakumar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9296 (2008).
10. See supplementary text on *Science* Online.
11. G. D. Fischbach, C. Lord, *Neuron* **68**, 192 (2010).
12. Materials and methods are available as supplementary materials on *Science* Online.
13. C. Betancur, *Brain Res.* **1380**, 42 (2011).
14. G. M. Cooper *et al.*, *Nat. Genet.* **43**, 838 (2011).
15. M. Lynch, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 961 (2010).
16. A. Kong *et al.*, *Nature* **488**, 471 (2012).
17. C. A. Williams, A. Dagli, A. Battaglia, *Am. J. Med. Genet. A.* **146A**, 2023 (2008).
18. R. S. Møller *et al.*, *Am. J. Hum. Genet.* **82**, 1165 (2008).
19. B. W. van Bon *et al.*, *Clin. Genet.* **79**, 296 (2011).
20. F. Guedj *et al.*, *Neurobiol. Dis.* **46**, 190 (2012).
21. M. E. Talkowski *et al.*, *Cell* **149**, 525 (2012).
22. J. Zhou, L. F. Parada, *Curr. Opin. Neurobiol.* **22**, 873 (2012).
23. I. Letunic, T. Doerks, P. Bork, *Nucleic Acids Res.* **40** (Database issue), D302 (2012).

Acknowledgments: We thank the National Heart, Lung, and Blood Institute, NIH Grand Opportunity (GO) Exome Sequencing Project and its ongoing studies, which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the Women's Health Initiative Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926), and the Heart GO Sequencing Project (HL-103010); we also thank B. Vernot, M. Dennis, T. Brown, and other members of the Eichler and Shendure labs for helpful discussions. We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic data on the Simons Foundation Autism Research Initiative (SFARI) Base. Approved researchers can obtain the SSC population dataset described in this study (https://ordering.base.sfari.org/~browse_collection/archive[ssc_v13]/ui:view) by applying at https://base.sfari.org. This work was supported by grants from the Simons Foundation (SFARI 137578, 191889 to E.E.E., J.S., and R.B.), NIH HD065285 (E.E.E. and J.S.), NIH NS069605 (H.C.M.), and R01 NS064077 (D.D.). E.B. is an Alfred P. Sloan Research Fellow. E.E.E. is an Investigator of the Howard Hughes Medical Institute. Scientific advisory boards or consulting affiliations: Ariosa Diagnostics (J.S.), Stratos Genomics (J.S.), Good Start Genetics (J.S.), Adaptive Biotechnologies (J.S.), Pacific Biosciences (E.E.E.), SynapDx (E.E.E.), DNAnexus (E.E.E.), and SFARI GENE (H.C.M.). B.J.O. is an inventor on patent PCT/US2009/30620: Mutations in contactin associated protein 2 are associated with increased risk for idiopathic autism. Raw sequencing data available at the National Database for Autism Research, NDARCOL1878.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1227764/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S18
References (*24–100*)

23 July 2012; accepted 1 November 2012
Published online 15 November 2012;
10.1126/science.1227764

Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell

Chenghang Zong,[1]* Sijia Lu,[1]*† Alec R. Chapman,[1,2]* X. Sunney Xie[1]‡

Kindred cells can have different genomes because of dynamic changes in DNA. Single-cell sequencing is needed to characterize these genomic differences but has been hindered by whole-genome amplification bias, resulting in low genome coverage. Here, we report on a new amplification method—multiple annealing and looping-based amplification cycles (MALBAC)—that offers high uniformity across the genome. Sequencing MALBAC-amplified DNA achieves 93% genome coverage ≥1x for a single human cell at 25x mean sequencing depth. We detected digitized copy-number variations (CNVs) of a single cancer cell. By sequencing three kindred cells, we were able to identify individual single-nucleotide variations (SNVs), with no false positives detected. We directly measured the genome-wide mutation rate of a cancer cell line and found that purine-pyrimidine exchanges occurred unusually frequently among the newly acquired SNVs.

Single-molecule and single-cell studies reveal behaviors that are hidden in bulk measurements (*1*, *2*). In a human cell, the genetic information is encoded in 46 chromosomes. The variations occurring in these chromosomes, such as single-nucleotide variations (SNVs) and copy-number variations (CNVs) (*3*), are the driving forces in biological processes such as evolution and cancer. Such dynamic variations are reflected in the genomic heterogeneity among a population of cells, which demands characterization of genomes at the single-cell level (*4–6*). Single-cell genomics analysis is also necessary when the number of cells available is limited to few or one, such as prenatal testing samples (*7*, *8*), circulating tumor cells (*9*), and forensic specimens (*10*).

Prompted by rapid progress in next-generation sequencing techniques (*11*), there have been several reports on whole-genome sequencing of single cells (*12–16*). These methods have relied on whole-genome amplification (WGA) of an individual cell to generate enough DNA for sequencing (*17–21*). However, WGA methods in general are prone to amplification bias, which results in low genome coverage. Polymerase chain reaction (PCR)–based WGA introduces sequence-dependent bias because of the exponential amplification with random primers (*17*, *18*, *22*). Multiple displacement amplification (MDA), which uses random priming and the strand-displacing ϕ29 polymerase under isothermal conditions (*19*), has provided improvements over PCR-based methods but still exhibits considerable bias, again due to nonlinear amplification.

Here we report a new WGA method, multiple annealing and looping-based amplification cycles (MALBAC), which introduces quasilinear preamplification to reduce the bias associated with nonlinear amplification. Picograms of DNA fragments (~10 to 100 kb) from a single human cell serve as templates for amplification with MALBAC (Fig. 1). The amplification is initiated with a pool of random primers, each having a common 27-nucleotide sequence and 8 variable nucleotides that can evenly hybridize to the templates at 0°C. At an elevated temperature of 65°C, DNA polymerases with strand-displacement activity are used to generate semiamplicons with variable lengths (0.5 to 1.5 kb), which are then melted off from the template at 94°C. Amplification of the semiamplicons gives full amplicons that have complementary ends. The temperature is cycled to 58°C to allow the looping of full amplicons, which prevents further amplification and cross-hybridizations. Five cycles of preamplification are followed by exponential amplification of the full amplicons by PCR to generate micrograms of DNA required for next-generation sequencing (Fig. 1). In the PCR, oligonucleotides with the common 27-nucleotide sequence are used as the primers.

We used MALBAC to amplify the DNA of single SW480 cancer cells. With ~25x mean

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [2]Program in Biophysics, Harvard University, Cambridge, MA 02138, USA.

*These authors contributed equally to the work.
†Present address: Yikon Genomics, 1 China Medical City Avenue, TQB Building, 5th floor, Taizhou, Jiangsu, China.
‡To whom correspondence should be addressed. E-mail: xie@chemistry.harvard.edu

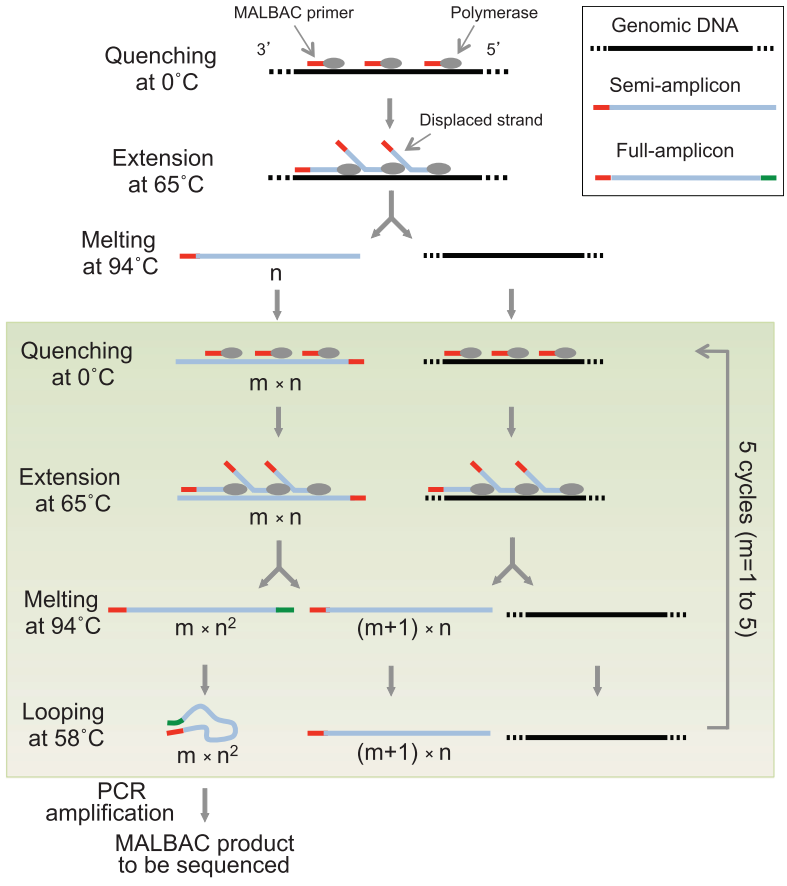

Fig. 1. MALBAC single-cell whole-genome amplification. A single cell is picked and lysed. First, genomic DNA of the single cell is melted into single-stranded DNA molecules at 94°C. MALBAC primers then anneal randomly to single-stranded DNA molecules at 0°C and are extended by a polymerase with displacement activity at elevated temperatures, creating semiamplicons. In the following five temperature cycles, after the step of looping the full amplicons, single-stranded amplicons and the genomic DNA are used as a template to produce full amplicons and additional semiamplicons, respectively. For full amplicons, the 3′ end is complementary to the sequence on the 5′ end. The two ends hybridize to form looped DNA, which can efficiently prevent the full amplicon from being used as a template, therefore warranting a close-to-linear amplification. After the five cycles of linear preamplification, only the full amplicons can be exponentially amplified in the following PCR using the common 27-nucleotide sequence as the primer. PCR reaction will generate microgram level of DNA material for sequencing experiments.

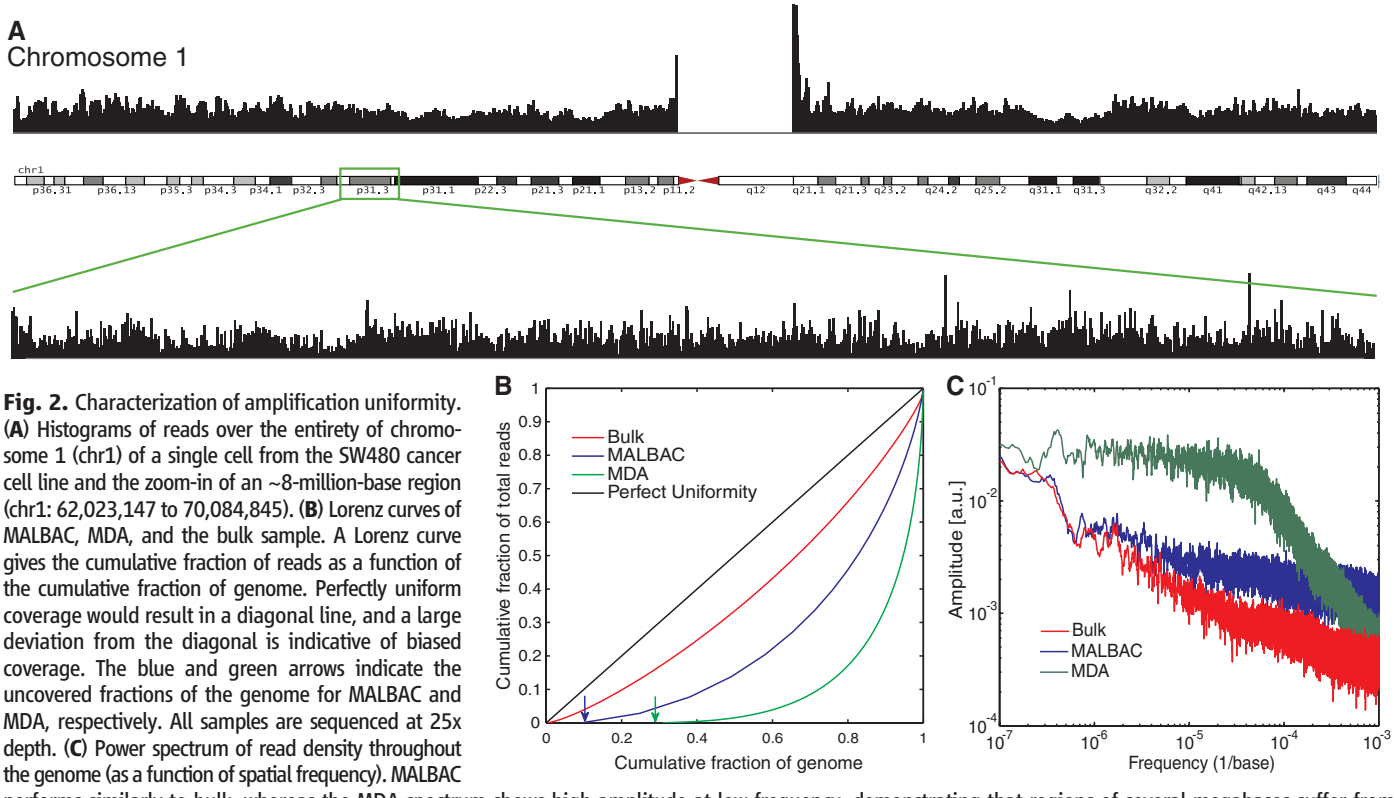

Fig. 2. Characterization of amplification uniformity. (**A**) Histograms of reads over the entirety of chromosome 1 (chr1) of a single cell from the SW480 cancer cell line and the zoom-in of an ~8-million-base region (chr1: 62,023,147 to 70,084,845). (**B**) Lorenz curves of MALBAC, MDA, and the bulk sample. A Lorenz curve gives the cumulative fraction of reads as a function of the cumulative fraction of genome. Perfectly uniform coverage would result in a diagonal line, and a large deviation from the diagonal is indicative of biased coverage. The blue and green arrows indicate the uncovered fractions of the genome for MALBAC and MDA, respectively. All samples are sequenced at 25x depth. (**C**) Power spectrum of read density throughout the genome (as a function of spatial frequency). MALBAC performs similarly to bulk, whereas the MDA spectrum shows high amplitude at low frequency, demonstrating that regions of several megabases suffer from under- and overamplification. This observation is consistent with the variations of read depth in fig. S3.

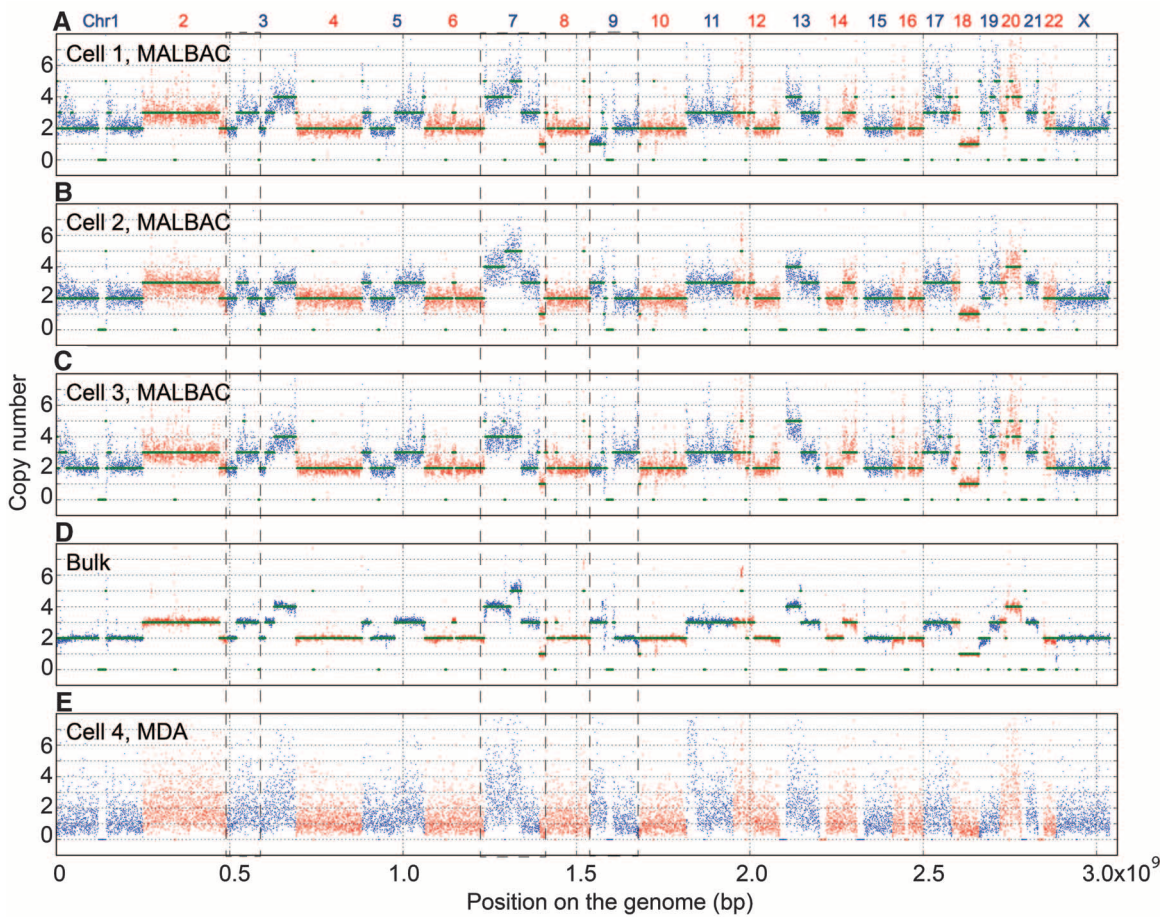

Fig. 3. CNVs of single cancer cells. Digitized copy numbers across the genome are plotted for three single cells (**A** to **C**) as well as the bulk sample (**D**) from the SW480 cancer cell line. The bottom panel shows the result based on MDA amplification (**E**). Green lines are fitted CNV numbers obtained from the hidden Markov model (see supplementary materials). The single cells are sequenced at only 0.8x depth, whereas the bulk and MDA are done at 25x. More single-cell CNV analyses are included in the supplementary materials (fig. S4). The regions within the dashed box exhibit the CNV differences among single cells and the bulk, which cannot be resolved by MDA. The binning window is 200 kb.

sequencing depth, we consistently achieved ~85% and up to 93% genome coverage at ≥1x depth on either strand (Fig. 2A). As a comparison, we performed MDA on a single cell from the same cancer cell line. At 25x mean sequencing depth, MDA covered 72% of the genome at ≥1x coverage. Although substantial variations of the coverage have been reported for MDA (*15*, *16*, *20*, *23*), MALBAC coverage is reproducible.

We used Lorenz curves to evaluate coverage uniformity along the genome. We plotted the cumulative fraction of the total reads that cover a given cumulative fraction of genome (Fig. 2B). The diagonal line indicates a perfectly uniform distribution of reads, and deviation from the diagonal line indicates an uneven distribution of reads. We compared the Lorenz curves for bulk sequencing, MALBAC, and MDA at ~25x mean sequencing depth (Fig. 2B). It is evident that MALBAC outperforms MDA in uniformity of genome coverage. We also plotted the power spectrum of read density variations to show the spatial scale at which the variations take place. For MDA, large amplitudes at low frequencies (inverse genome distance) were observed, indicating that large contiguous regions of the genome are over- or underamplified. In contrast, MALBAC has a power spectrum similar to that of the unamplified bulk.

Table 1. Comparison of single-cell SNVs for bulk, MDA, and MALBAC.

	Heterozygous SNVs	Homozygous SNVs	Total SNVs
	Bulk		
SNVs	911,958	1,930,204	2,842,162
	Single-cell MDA		
SNVs	93,140 (2,828)*	1,238,286 (1,973)	1,331,426 (4,801)
Detection efficiency	10%	63%	41%
	Single-cell MALBAC		
SNVs	756,812 (108,481)	1,539,326 (6,821)	2,296,138 (115,302)
Detection efficiency	71%	80%	76%

*The number in parentheses indicates the number of false positives.

Table 2. MALBAC identification of total SNVs and newly acquired SNVs using two and three kindred cells.

	Heterozygous SNVs	Homozygous SNVs	Total SNVs
	Two kindred cells		
SNVs	615,387	1,322,555	1,937,942
Detection efficiency	67%	68%	68%
Newly acquired SNVs	145 (~100)*	3 (~0)	148 (~100)
	Three kindred cells		
SNVs	660,246	1,577,798	2,238,044
Detection efficiency	72%	81%	80%
Newly acquired SNVs	30 (~0)	5 (~0)	35 (~0)

*The number in parentheses indicates the number of false positives. "~0" indicates undetected in Sanger sequencing when PCR primers can be readily designed.

CNVs due to insertions, deletions, or multiplications of genome segments are frequently observed in almost all categories of human tumors (*13*, *24*, *25*). MALBAC's lack of large-scale bias makes it amenable to probing CNVs in single cells. We determined the digitized CNVs across the whole genomes of three individual cells from the SW480 cancer cell line (Fig. 3, A to C). CNVs of five cells are included in the supplementary materials. The chromosomes exhibit distinct CNV differences among the three individual cancer cells and in the bulk result (Fig. 3D), which are difficult to resolve by MDA (Fig. 3E). For the MALBAC data, we used a hidden Markov model to quantify CNVs (see supplementary materials). We confirmed the gross features of CNVs detected by MALBAC with a previously published karyotyping study (*26*).

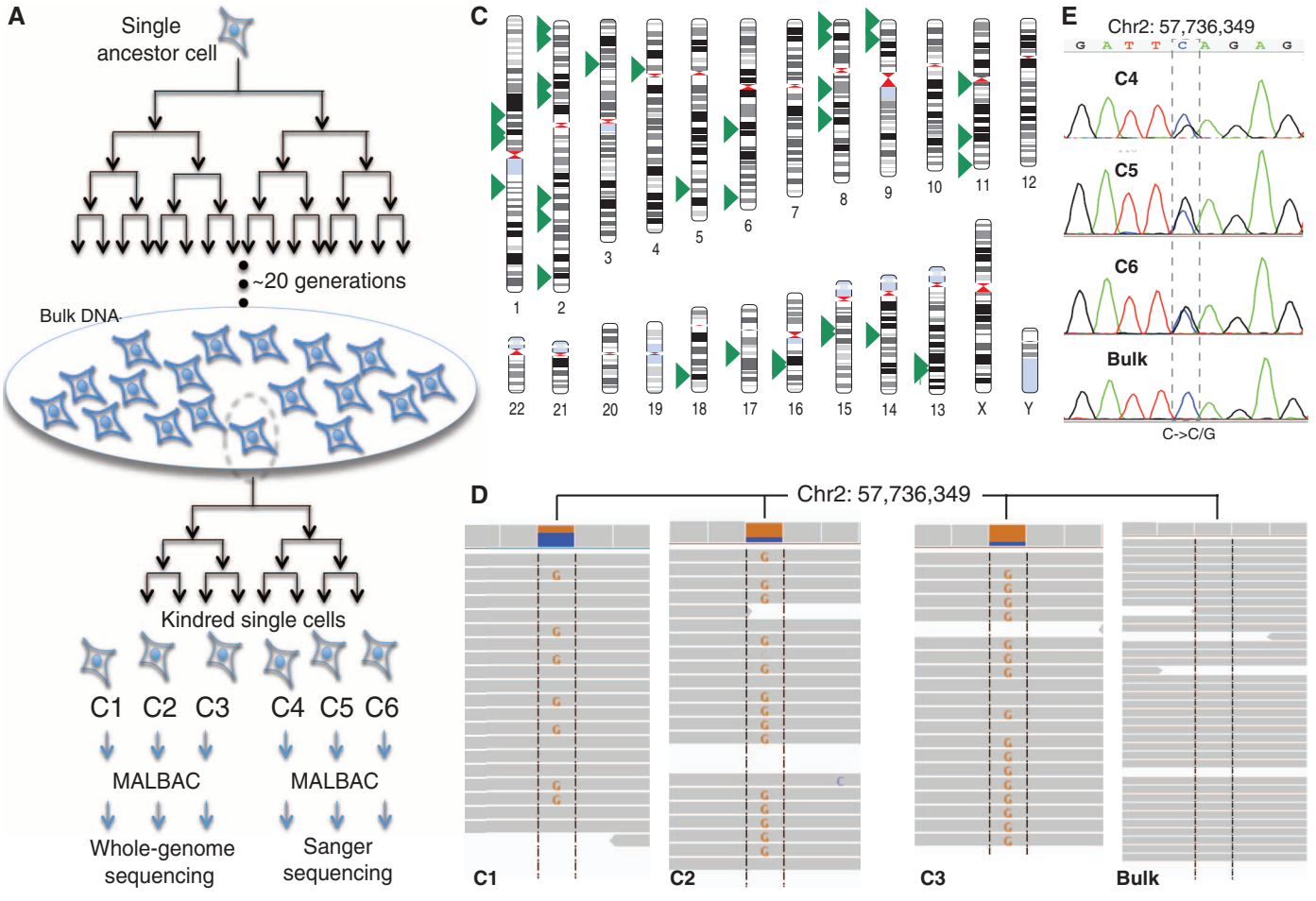

Fig. 4. Identifying newly acquired SNVs and estimation of mutation rate of a cancer cell line (SW480). (**A**) Experiment design. A single ancestor cell is chosen and cultured for ~20 generations. The vast majority of cells are used to extract DNA for bulk sequencing to represent the ancestor cell's genome. A single cell from this culture is chosen for another expansion of four generations. The kindred cells are isolated for single-cell whole-genome amplification. Single-cell samples C1, C2, and C3 are used for high-throughput sequencing. Samples C4, C5, and C6 are used for verifying SNVs with Sanger sequencing. (**B**) Three-dimensional *P*-value plot of a one-sided binomial test for SNV candidates from the three kindred cells. The black dots are the false positives due to uncorrelated amplification errors; all of them are on the *x*-*y*-*z* axes and *x*-*y*, *y*-*z*, *x*-*z* planes. Outside of the three planes, the 166 green dots are the residual false positives due to correlated errors from homopolymers, tandem repeats, high-GC content, and high-density SNV regions, and the 35 red dots are the newly acquired SNVs during the 20 generations of clonal expansion (see supplementary materials). The homozygous SNVs are located at the (1,1,1) position. (**C**) Locations of the 35 newly acquired SNVs on the chromosomes of a single cell (see supplementary materials). (**D**) Next-generation sequencing data of a newly acquired SNV. The SNV (C→G) exists in the high-throughput data of all three kindred cells but not in the bulk data. (**E**) Sanger sequencing data of single cells C4, C5, and C6 confirm that this SNV is not a false positive, whereas the Sanger sequencing of the bulk confirms that this SNV is not a false negative of next-generation sequencing of the bulk (i.e., this SNV is indeed absent in the bulk).

For example, both MALBAC-based quantification of CNVs and spectral karyotyping show one copy of chromosome 18 and three copies of chromosome 17 in the SW480 cancer cell line. Although the majority of copy numbers are consistent between single cells, we also observe cell-to-cell variations as labeled by the dashed boxes in Fig. 3.

Attempts have been made recently to identify SNVs from a single cell by MDA (*15*, *16*, *23*). The first challenge in accurate SNV identification from a single cell is substantial human contamination from the environment and the operators, given picograms of DNA from a single human cell. The second challenge is low detection yield (high false negative rates), particularly where alleles drop out due to amplification bias. The third challenge is false positives associated with amplification and sequencing errors, either random or systematic (*27*).

To meet the first challenge, we took special precautions to decontaminate with ultraviolet radiation before each experiment was conducted in a restricted clean room. An alternative approach to reduce contamination is microfluidics (*28*).

With regard to the second challenge, MALBAC allowed us to identify 2.2×10^6 single-cell SNVs compared with 2.8×10^6 detected SNVs in bulk, yielding a 76% detection efficiency, in contrast to 41% with MDA (Table 1). This improvement resulted from improved uniformity by MALBAC (fig. S6). Listed separately in Table 1 are heterozygous and homozygous SNVs. Next, we calculated the allele dropout rate. Comparison of single-cell and bulk SNVs showed that 7288 of the SNVs genotyped as homozygous mutations by MALBAC are actually heterozygous in bulk, which corresponds to a ~1% allele dropout rate in MALBAC (see supplementary materials). In contrast, with MDA we found 172,563 incorrect homozygous identifications, corresponding to an allele dropout rate of ~65% (see supplementary materials).

Compared to the bulk data, the MALBAC data contains 1.1×10^5 false positives (Table 1) out of 3×10^9 bases in the genome. This corresponds to a $\sim 4 \times 10^{-5}$ false-positive rate, which is due to the errors made by the polymerases in the semiamplicons generated in the first MALBAC cycle and propagated through the later amplification. Although improving the polymerase's error rate is possible, our strategy to reduce the false-positive rate was to sequence two or three kindred cells derived from the same cell. The simultaneous appearance of an SNV in the kindred cells would indicate a true SNV. The false-positive rate due to uncorrelated random errors can be reduced to $\sim 10^{-8}$ with two kindred cells and $\sim 10^{-12}$ with three kindred cells.

However, there are false positives due to correlated errors—that is, systematic sequencing and amplification errors. We filtered out these errors by comparing two unrelated single cells that are not from the same lineage (fig. S5) and additional sceening. After this procedure, we can identify true SNVs of a single cell with no false positives detected by Sanger sequencing, as described below (Table 2).

To gain insight into the mutation process in the cancer cells, we clonally expanded a single ancestor cell picked from a heterogeneous population of the SW480 cancer cell line for 20 generations (Fig. 4A). We extracted DNA from this single-cell clonal expansion for bulk sequencing, which reflects the genome of the ancestor cell. We then picked a single cell from this clone. To detect SNVs acquired by the cell during expansion, we grew another four generations to obtain the kindred cells denoted C1 to C16. We individually sequenced three kindred cells—C1, C2, and C3—after MALBAC amplification. After filtering correlated and uncorrelated errors (Fig. 4B), we detected 35 unique SNVs shown in Fig. 4C.

We took 24 out of 35 unique SNVs for which we can readily design PCR primers and confirmed that they are neither false positives by Sanger sequencing C4 to C6 nor false negatives by Sanger sequencing the bulk. (See the supplementary materials for Sanger sequencing data.) As an example, Fig. 4, D and E shows the MALBAC and Sanger sequencing result of one such SNV.

These 35 unique SNVs are newly acquired during the 20 cell divisions. Adjusting for a detection efficiency of 72% for heterozygous SNVs, we estimate that ~49 mutations occurred in the 20 generations, yielding a mutation rate of ~2.5 nucleotides per cell generation, consistent with our estimation based on the bulk data (see supplementary materials). The mutation rate of this cancer cell line is about 10 times as high as the mutation rate estimated based on germline studies (*29*–*31*).

Mutations can be transitions (purine↔purine exchange, i.e., A↔G, or pyrimidine↔pyrimidine exchange, i.e., C↔T) or transversions (purine↔pyrimidine exchanges, i.e., A/G↔C/T). Transitions are more common. Unexpectedly, we found that the transition/transversion (tstv) ratio for the 35 newly acquired SNVs detected is only 0.30, whereas the ratio for the total SNVs of this cell line is 2.01, as expected for common human mutations (*32*). To further confirm that this observation is not due to single-cell amplification, we sequenced the bulk DNA of the original heterogeneous culture (see supplementary materials). The tstv ratio for SNVs detected in the single-cell expanded bulk but not in the original heterogeneous bulk was 0.75. Both significantly low tstv ratios indicate that transitions are not favored over transversion for newly acquired SNVs in this cancer cell line (see supplementary materials). Although understanding the underlying mechanism of this phenomenon will require similar measurements in other systems, it is evident that, by allowing precise characterization of CNVs and SNVs, MALBAC can shed light on the individuality, heterogeneity, and dynamics of the genomes of single cells.

References and Notes

1. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
2. G. W. Li, X. S. Xie, *Nature* **475**, 308 (2011).
3. S. Negrini, V. G. Gorgoulis, T. D. Halazonetis, *Nat. Rev. Mol. Cell Biol.* **11**, 220 (2010).
4. C. Lengauer, K. W. Kinzler, B. Vogelstein, *Nature* **396**, 643 (1998).
5. S. Yachida *et al.*, *Nature* **467**, 1114 (2010).
6. P. J. Campbell *et al.*, *Nature* **467**, 1109 (2010).
7. Y. M. Lo *et al.*, *Sci. Transl. Med.* **2**, 61ra91 (2010).
8. J. O. Kitzman *et al.*, *Sci. Transl. Med.* **4**, 137ra76 (2012).
9. S. Nagrath *et al.*, *Nature* **450**, 1235 (2007).
10. E. K. Hanson, J. Ballantyne, *Anal. Biochem.* **346**, 246 (2005).
11. M. L. Metzker, *Nat. Rev. Genet.* **11**, 31 (2010).
12. H. C. Fan, J. Wang, A. Potanina, S. R. Quake, *Nat. Biotechnol.* **29**, 51 (2011).
13. N. Navin *et al.*, *Nature* **472**, 90 (2011).
14. M. Gundry, W. G. Li, S. B. Maqbool, J. Vijg, *Nucleic Acids Res.* **40**, 2032 (2012).
15. Y. Hou *et al.*, *Cell* **148**, 873 (2012).
16. J. Wang, H. C. Fan, B. Behr, S. R. Quake, *Cell* **150**, 402 (2012).
17. L. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5847 (1992).
18. H. Telenius *et al.*, *Genomics* **13**, 718 (1992).
19. F. B. Dean, J. R. Nelson, T. L. Giesler, R. S. Lasken, *Genome Res.* **11**, 1095 (2001).
20. K. Zhang *et al.*, *Nat. Biotechnol.* **24**, 680 (2006).
21. K. Lao, N. L. Xu, N. A. Straus, *Biotechnol. J.* **3**, 378 (2008).
22. W. Dietmaier *et al.*, *Am. J. Pathol.* **154**, 83 (1999).
23. X. Xu *et al.*, *Cell* **148**, 886 (2012).
24. R. Beroukhim *et al.*, *Nature* **463**, 899 (2010).
25. P. J. Stephens *et al.*, *Cell* **144**, 27 (2011).
26. P. J. Rochette, N. Bastien, J. Lavoie, S. L. Guérin, R. Drouin, *J. Mol. Biol.* **352**, 44 (2005).
27. D. MacArthur, *Nature* **487**, 427 (2012).
28. P. C. Blainey, S. R. Quake, *Nucleic Acids Res.* **39**, e19 (2011).
29. J. W. Drake, B. Charlesworth, D. Charlesworth, J. F. Crow, *Genetics* **148**, 1667 (1998).
30. J. C. Roach *et al.*, *Science* **328**, 636 (2010).
31. D. F. Conrad *et al.*, 1000 Genomes Project, *Nat. Genet.* **43**, 712 (2011).
32. D. L. Altshuler *et al.*, 1000 Genomes Project Consortium, *Nature* **467**, 1061 (2010).

Acknowledgments: This work was supported by U.S. National Institutes of Health National Human Genome Research Institute grants (HG005097-1 and HG005613-01) and in part by Bill & Melinda Gates Foundation OPP42867 to X.S.X. A.R.C. was supported by an NIH Molecular Biophysics Training grant (NIH/NIGMS T32 GM008313). We thank P. Choi for his involvement in the early stage of the project and J. Lu and L. Song for their assistance on the experiments. We thank J. Yong for his help on single-cell expansion and isolation and Y. Zhang at Biodynamic and Optical Imaging Center (BIOPIC) at Peking University for assistance on sequencing. The sequencing data are deposited at the National Center for Biotechnology Information with accession no. SRA060929. C.Z., S.L., and X.S.X. are authors on a patent applied for by Harvard University that covers the MALBAC technology.

Supplementary Materials

www.sciencemag.org/cgi/content/full/338/6114/1622/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S3
References (*33*–*37*)

22 August 2012; accepted 12 November 2012
10.1126/science.1229164