Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification

Yusi Fu*a, Chunmei Li*b, Sijia Lu*b,1, Wenxiong Zhou*a, Fuchou Tang*c, X. Sunney Xie*a,b,2, and Yanyi Huang*a,c,d,2

aBiodynamic Optical Imaging Center, School of Life Sciences, Peking University, Beijing 100871, China; bDepartment of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; cPeking–Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China; and dCollege of Engineering, Peking University, Beijing 100871, China

Contributed by X. Sunney Xie, July 28, 2015 (sent for review June 7, 2015; reviewed by Luke P. Lee)

Whole-genome amplification (WGA) for next-generation sequencing has seen wide applications in biology and medicine when characterization of the genome of a single cell is required. High uniformity and fidelity of WGA is needed to accurately determine genomic variations, such as copy number variations (CNVs) and single-nucleotide variations (SNVs). Prevailing WGA methods have been limited by fluctuation of the amplification yield along the genome, as well as false-positive and -negative errors for SNV identification. Here, we report emulsion WGA (eWGA) to overcome these problems. We divide single-cell genomic DNA into a large number (105) of picoliter aqueous droplets in oil. Containing only a few DNA fragments, each droplet is led to reach saturation of DNA amplification before demulsification such that the differences in amplification gain among the fragments are minimized. We demonstrate the proof-of-principle of eWGA with multiple displacement amplification (MDA), a popular WGA method. This easy-to-operate approach enables simultaneous detection of CNVs and SNVs in an individual human cell, exhibiting significantly improved amplification evenness and accuracy.

Significance

Uniform and accurate single-cell whole-genome amplification is important when starting material is limited and precious. We develop an emulsion-based amplification method that can suppress the amplification bias to detect high-resolution copy number variations of a single cell, and to simultaneously detect the single-nucleotide variations with high accuracy. This approach is compatible with various amplification protocols including the widely used multiple displacement amplification, which has been demonstrated in this paper.

*Author contributions: Y.H. designed research; C.L. prepared sequencing libraries; Y.F., C.L., and W.Z. performed research; Y.F., S.L., F.T., X.S.X., and Y.H. analyzed data; and Y.F., S.L., F.T., X.S.X., and Y.H. wrote the paper.

Reviewers included: L.P.L., University of California, Berkeley.

Conflict of interest statement: S.L. and X.S.X. are cofounders and shareholders of Yikon Genomics.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the NCBI Sequence Read Archive database (accession no. SRX952908).

1Present address: Yikon Genomics, Co., Ltd., Taizhou, Jiangsu 225300, China.

2To whom correspondence may be addressed. Email: xie@chemistry.harvard.edu or yanyil@psu.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1513988112/-/DCSupplemental.
cell’s genome. By distributing single-cell genomic DNA fragments into a large number (10^2) of picoliter droplets, a few DNA fragments in each droplet is allowed to reach saturation of DNA amplification. After merging the droplets by demulsification, the differences in amplification gain among the DNA fragments are significantly minimized.

Although this approach can be used for any chemistry of WGA, we take MDA as an example to greatly reduce the random bias of amplification by separating the reactions into a large amount of emulsion droplets. We carried out detailed comparison with MDA, MALBAC, and DOP-PCR performed in tube using single cells from normal diploid human cells and a monoclonal human cancer cell line with inherited CNVs. Our results indicate that eWGA not only offers higher coverage but also enables simultaneous detection of SNVs and CNVs with higher accuracy and finer resolution, outperforming the prevailing single-cell amplification methods in many aspects.

Results and Discussion

**eMDA Sequencing Library Preparation.** MDA, an easy-operating and widely used single-cell WGA protocol, is used for the proof-of-concept of eWGA. We lysed individual cells to release the genomic DNA (gDNA) fragments and dehybridized them to single strands by heating. After adding the MDA reaction buffer, the solution (10 μL) was distributed into ~7×10^3 droplets, 14 pL each, using a microfluidic chip (Fig. 1 A and B, and SI Appendix, Fig. S1). This process is carried out at 4 °C to keep the amplification from starting. Under this lysing condition, the estimated mean size of DNA fragments is ~10 kb (18). Thus, for a single diploid cell, each droplet contains one fragment on average. We have tested different dilutions of DNA and observed the decline of mapping rate with further dilution, especially when the average fragment is far less than one per droplet. This is because a large number of empty droplets increases the ratio of nonspecific product of amplification. On the other hand, more DNA fragments in one droplet (>10 per droplet) impairs the evenness of WGA. The aneuploidy of the cell will affect the actual number of fragments per droplet. We found eMDA performance is stable when each droplet has one to two fragments. We collected all of the droplets in a microcentrifuge tube (Fig. 1 C and D). In contrast to the conventional single-tube MDA reaction, which exhibits more serious amplification bias with longer reaction time, in the emulsion MDA (eMDA) reaction each droplet produces similar amount of amplification products due to the eventual saturation of the polymerization reaction in each droplet (SI Appendix, Fig. S2). After heat inactivation of the enzyme and demulsification, the amplification uniformity is accomplished in the aqueous solution, and the amplification products are used to construct sequencing libraries.

**eMDA Amplify Normal Diploid Single Cells Evenly and Completely.** We chose human umbilical vein endothelial cell (HUVEC), a normal human diploid cell line, to validate the amplification evenness of eMDA using bulk (200 ng) genomic DNA from HUVECs as a reference. We carried out 10 single-cell eMDA experiments and compared the sequencing results with those of single-cell MALBAC or conventional MDA reactions. We divided the human genome into bins with mean size of 52.4 kb using dynamic binning method (5) and applied shallowly sequenced data, 3M uniquely mapped reads for each single cell, to calculate the copy number in each bin (Fig. 2A and SI Appendix, Figs. S3 and S4). eMDA showed the most uniform amplification across the whole genome, with coefficient of variation (CV) of 0.36, which is significantly lower than the conventional MDA (CV = 2.23) (Fig. 2A, Table 1, and SI Appendix, Fig. S5). From the reads covering autosomes and sex chromosomes of these single HUVECs (Fig. 2B), we found eMDA providing the smallest deviation from a priori expectation using bulk DNA as a reference.

![Fig. 1. The experimental process of eWGA-seq and emulsion generation.](image)

We sequenced a few single HUVECs to a greater depth (>14x) using eMDA, MALBAC, and conventional MDA, and plotted the Lorenz curves of coverage to further validate the evenness of eMDA (Fig. 2C). As perfectly uniform coverage would result in a diagonal line, eMDA shows the best uniformity across the whole genome, compared with MALBAC and conventional MDA-amplified single cells, and is closest to the unamplified bulk sample. In contrast to the previously reported nanoliter MDA reaction in which the amplification gain is reduced (20, 22), our eMDA yields a similar gain as the conventional MDA to ensure a high coverage breadth of the genome. We showed that emulsion would not result in losing fragments of DNA as eMDA exhibits slightly higher coverage breadth (72.3% at 10x, for a human diploid cell) than MALBAC (67.5%) or conventional MDA (68.5%) at the same sequencing depth (SI Appendix, Figs. S6 and S7).

We also plotted the power spectra of read density as a function of the spatial frequency (Fig. 2D) based on the sequencing result using different protocols. The analysis confirmed that, for single-cell sequencing, eMDA provides the best uniformity among the three methods by offering smaller copy number fluctuation at all frequencies due to the effectively suppressed amplification bias through compartmentation. Because the intrinsic amplification randomness still exists within each droplet, the uniformity improvement is more significant in the lower frequency (large bin size) region than in the higher frequency domain.
To estimate the accuracy of CNV identification of these methods, we carried out a simulation by calling the artificial CNVs with both copy number gain (2 to 3) and loss (2 to 1), in silico generated within diploid autosomes (SI Appendix, Fig. S8). The accuracy is the ratio of simulated CNVs that could be detected at the 52.4-kb resolution. eMDA shows much higher accuracy to identify the CNVs at the range from 300 kb to 2 Mb. We also performed an intersample correction (24) for MALBAC to eliminate the sequence-dependent bias (SI Appendix, Fig. S9), whereas for eMDA such normalization is unnecessary. This feature is very important in various medical applications such as in vitro fertilization preimplantation screening because a standard normalization sample and the expertise of performing complicated cross-sample normalization are often not available. eMDA was
eMDA Amplifies Normal Diploid Single Cells with Higher Accuracy.

Exome Coverage Breadth and SNV Detection in Single Cancer Cells. We then investigated the accuracy of SNV identification from single HT-29 cells using eMDA. We performed exome enrichment and sequencing for all samples and used bulk HT-29 exome as a reference. eMDA shows highest coverage (≥1x depth, 90 ± 5%), followed by MALBAC (79 ± 4%), conventional MDA (74 ± 11%), and DOP-PCR (44 ± 4%) (Fig. 3E). eMDA also exhibits high accuracy of replication events in single cells, with highest true-positive ratio and lowest false-positive rate among all methods we tested (Fig. 3F). As expected, eMDA also noticeably reduce the ADO to 24% from 43% of conventional MDA for these nondiploid single cells (SI Appendix, Fig. S12B).

Conclusion

Our method, eWGA, applies emulsion to divide the DNA fragments from a single cell to a large number of aqueous droplets in oil and drives the amplification to saturation in each droplet. Using MDA protocol as a demonstration, this approach can dramatically reduce the amplification bias while retaining the high accuracy of replication. Unlike other microfluidics-based WGA methods (20, 23), which improved the uniformity by reducing
Fig. 3. The comparison of WGA methods for sequencing single HT-29 cells. (A) The circos plot (27) showing the copy number profiles from unamplified genomic DNA and from a single cell amplified by eMDA. (B) The zoomed-in copy number distribution of chr3 and chrX with a binning size of 52.4 kb. The smallest CNV detected is 5 bins. (C) Heat map showing copy number gains and losses of single cells with different amplification methods, with unamplified genomic DNA as reference. The correlation efficiencies between single-cell WGA methods and bulk reference are also listed. (D) The CNV detection sensitivity under different bin size threshold of single-cell WGA methods. The filled area represents the SD of each method. (E) The coverage ratio of exome captured single-cell WGA samples using unamplified sample as reference. (F) The homozygous SNVs detected in single cells using different WGA methods. The blue line shows the number of homozygous SNVs identified in the unamplified sample. The blue bars show the SNVs that matched bulk reference, whereas the red bars show the discordant SNVs.
the gain compared with conventional MDA, eMDA has the gain of \( \sim 2 \times 10^7 \), which is comparable to the conventional MDA in tube with single human cells as starting material. With the high coverage breadth across the whole genome, eMDA also enables us to detect more SNVs than existing methods and the pollution rate is alleviated with the use of emulsion. eMDA is compatible with targeted enrichment methods such as exome capture, which is useful when only certain regions are of interest in genetic analyses. By using eMDA, the first method (to our knowledge) that enables simultaneous identification of both small CNVs and high-confidence SNVs from a single human cell, we are able to detect CNVs at 250-kb size with 50-kb resolution, and SNVs with error rate \(< 2 \times 10^{-5}\). We envision that such emulsion approach will also improve the amplification performance of other WGA methods, for example MALBAC, for single-cell genomic studies.

Materials and Methods

Device Fabrication. Microfluidic emulsion-generating chips were made of polydimethylsiloxane (PDMS). The mold used to cast the chips was made by etching photore sist on a silicon wafer using photolithography. In brief, SU-8 2025 (MicroChem) was spin coated onto the wafer at 1700 rpm for 60 s on a spin coater (KW-4A, SETCAS Electronics Co., Ltd), resulting in a thickness of 50 μm of photore sist. Then the wafer was baked at 95 °C for 5 min. The wafer was exposed to UV light for 30 s through a mask defining the channel geometry and then the wafer was baked again at 95 °C for 10 min. The unexposed photore sist was removed with solvent and the wafer was hard-baked at 150 °C for 3 h. The mold was treated with trimethyl chlorosilane vapor for 10 min before use. Then 30 g degased and well-mixed 5:1 (base:curing agent) PDMS (Sylgard 184, Dow Corning) was poured on the wafer, and baked together at 80 °C for 15 min before peeled off. Then we punched the holes for the inlets of reagent/oil and the outlet for connecting a micro-tubing that transferred the emulsion droplets to a 200 μl micro-centrifuge tube. Then the patterned PDMS slab was bonded with a piece of cover glass precoated with 20:1 (base:curing agent) PDMS through baking at 80 °C for 3 h. The resulting chip is shown as SI Appendix, Fig. S1.

Conventional Single-Cell MDA Reaction. The gDNA was fragmented by heating (4 min at 98 °C, and 2 min at 95 °C) in 4 μl lysis buffer [30 mM Tris·HCl (pH = 8.0), 10 mM NaCl, 1 mg/mL proteinase (Qiagen), 5 mM EDTA and 0.5% Triton X-100]. Then 6 μl MDA reaction buffer was added to reach 10 μl total volume with a final concentration of 1X Phi-29 buffer (NEB), 50 μM N6 primer with two phosphoro Thornton bonds at the 3′-side (Invitrogen), 1 mM dNTP (NEB). We heated the tube at 95 °C for 5 min, and then immediately put it on ice for at least 20 min to anneal the random hexmers to fragmented gDNA. We then added 8 units of Phi-29 polymerase (NEB) and briefly centrifuged. Then MDA reactions were carried out at 30 °C. Reactions were terminated at 65 °C for 10 min after 10 h amplification.

Single-Cell eMDA Reaction. The reaction buffer preparation is identical to MDA reactions. However, to prevent the reaction from initiating prior to droplet generation, the Phi-29 polymerase was added to the reaction mix immediately before emulsion generation. The reaction buffer was kept at 4 °C to prevent the amplification from starting before being dispersed into droplets. The emulsion droplets were collected into a tube and then incubated at 30 °C for 8–10 h before termination at 65 °C for 10 min.

A detailed description of remaining material and methods can be found in SI Appendix.

ACKNOWLEDGMENTS. We thank Dr. Ruiqiang Li, Dr. Zhilong Yu, Xiannian Zhang, Wei Fan, Ang Li, Zitian Chen, Tao Chen, and Zhe Su for their help on experiments and data analysis, Dr. Yun Zhang and the High-Throughput Sequencing Center of Peking University for assistance with Illumina sequencing, and the National Center for Protein Sciences Beijing (Peking University) for assistance with FACS analysis. This work was supported by National Natural Science Foundation of China Grants 21327808, 91313302, and 21222501 (to Y.H.).