CellPress

Molecular Cell
Technology Preview

Down the Rabbit Hole of Single-Cell Genome Analysis

James Eberwine[1,*]
[1]Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
*Correspondence: eberwine@mail.med.upenn.edu
http://dx.doi.org/10.1016/j.molcel.2017.04.015

Chen et al. (2017) demonstrate whole-genome amplification of single-cell genomic DNA using linear nucleic acid amplification. This provides reliable single-nucleotide variation (SNV) detection across the single-cell genome, facilitating an understanding of cell-to-cell similarities and distinctions.

In *Alice's Adventures in Wonderland* (Carroll, 1866), Alice exclaimed, ''Let me think: was I the same when I got up this morning? I almost think I can remember feeling a little different. But if I'm not the same, the next question is, Who in the world am I? Ah, that's the great puzzle.''

This is the current state of single-cell genomic DNA analysis; it is often difficult to unequivocally characterize the cognate DNA sequence of a cell due to technical limitations. The ultimate sensitivity for whole-genome sequence analysis would be determination of the high-confidence identity of each individual nucleotide in the genome so that single-nucleotide variations (SNVs) can be identified. Single-cell genomic DNA nucleotide sequence identification has been complicated by the need for amplification of the DNA to large enough levels to be sequenced, usually using exponential amplification procedure, which is prone to systematic error. The newest technique in single-cell genomic DNA amplification procedures is described in a paper by Chen et al. (2017), where they report on a novel approach for genomic DNA sequencing and SNV detection called Linear Amplification via Transposon Insertion (LIANTI). This procedure uses linear RNA amplification of genomic DNA sequence to generate sequencing libraries that consequently exhibit significantly less systematic experimental error than PCR-based methods.

Detection of single-nucleotide differences between genomes has been experimentally assessed since the time of restriction fragment polymorphism analysis. Detection of sequence variation became much easier with the advent of DNA sequencing technologies. Initially, large amounts of DNA isolated from tissues or cell lines permitted detection of SNPs that were present as germline differences, with SNV detection dependent upon having a large number of expressing cells. SNV detection becomes problematic when small amounts of genomic DNA are used and requires nucleic acid amplification. The amplification procedures that have been used comprise at least some exponential component and include multiple strand displacement (MDA), degenerate oligonucleotide-primed PCR (DOP-PCR), and multiple annealing and looping-based amplification cycles (MALBAC).

The LIANTI procedure takes advantage of two general use molecular techniques that have been employed to analyze single cells: (1) the linear RNA amplification afforded by T7 RNA polymerase and (2) the transposon method for making sequencing libraries, which employs the in vitro DNA cutting and insertion activity of transposase (Goryshin and Reznikoff, 1998). Linear amplification was developed nearly 25 years ago for RNA (Van Gelder et al., 1990) (called amplified RNA, or aRNA). The aRNA procedure has been shown to be linear, with fewer errors propagated through the amplification while being more precise than PCR (Dueck et al., 2016; Eberwine et al., 1992). Chen and colleagues adapted aRNA for use in analysis of genomic DNA by cleverly melding the T7 promoter site with a transposase binding site stem to create the LIANTI transposon (Figure 1). This transposon is mixed with transposase to generate the LIANTI transposome. Upon mixing the LIANTI transposome with DNA isolated from the single cell, the transposase mediates the random integration of the LIANTI transposon into the DNA and subsequent excision of genomic DNA between two transposon integration sites. The excised transposon-flanked DNA is strand-end repaired, producing a double-stranded piece of genomic DNA with a T7 promoter situated at the ends of the DNA. Upon addition of T7 RNA polymerase, single-stranded aRNA is generated that contains one strand of transposase double-stranded binding site on the 5′ end and the full LIANTI transposon site on the 3′ end. Due to complementarity of the transposase-binding site of the full LIANTI transposon, the 3′ end transposase site acts to prime DNA synthesis from the aRNA. Upon RNase digestion of the RNA from the RNA-DNA hybrid, the remaining single-stranded DNA has a single-stranded transposase binding site that can be used to prime second-strand DNA synthesis, resulting in a barcoded double-stranded DNA that is used to create the sequencing library.

The authors report that when the LIANTI procedure is used to sequence a single BJ cell's genome to 30× coverage, it produces 97% read coverage with only 17% allele drop out. There is also a remarkably low 0.7 coefficient of variation with the LIANTI procedure for even small sequence ranges over 100 bases and less. Such uniformity makes it possible to have high confidence in the single-nucleotide resolvability of the procedure. As with any new methodology, there are outstanding questions that future papers should address, including the following: (1) How does the known GpC island sequence preference for Tn5 transposon insertion affect sequence coverage? (2) Since both sense and antisense

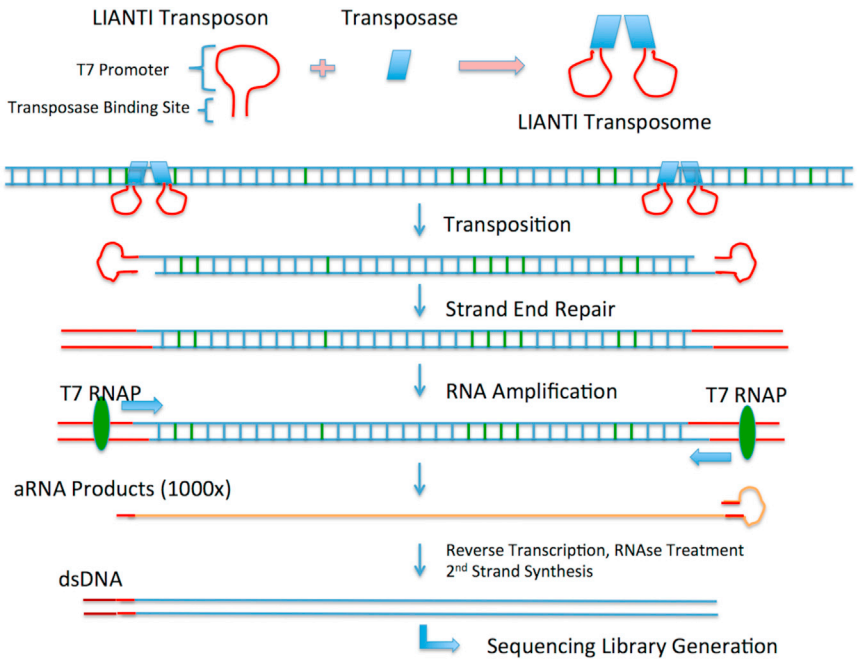

Figure 1. LIANTI Methodology
Schematic showing the steps in the LIANTI procedure for genomic DNA sequence analysis. Adapted from Chen et al. (2017).

sequences should be equally synthesized during the aRNA amplification, is there an issue with sequence fall out due to interference of processing T7 RNA polymerases from each of the opposite strands? (3) Is there a sequence bias or fallout issue with the resultant "equally abundant" sense/antisense because of annealing of these complementary RNAs?

Among the biologies that were looked at with LIANTI was the frequency of particular types of SNVs in the genome. One of the most common SNVs in single-cell analysis is the C-T transitions observed in neurons, sperm, and cancer cells. The authors reasoned that chemical deamination of C during DNA isolation could artificially create U bases, which is known to occur (Krokan et al., 2002), making the reported numbers aberrantly high. Using the DNA repair enzyme uracil DNA glycosylase, which eliminates uracil bases that form from deamination of C, the authors report that the number of such single-cell C-T SNVs is reduced by nearly 60% to levels comparable to what is reported for bulk DNA. While this result has wide-ranging implications, it is curious that the bulk signature of essentially a single cell type, the BJ cells, shows few C-T transitions, while each of the individual cells showed significant levels upon either LIANTI or MDA amplification. It is unclear how generalizable these results will be to other cell types.

The design of the LIANTI procedure suggests that it might be useful in examining other cellular biologies. For example, since LIANTI uses transposons to introduce the T7 promoter sequence into genomic DNA, it may be possible to use LIANTI to assess open chromatin structure from single cells akin to ATAC-seq (Buenrostro et al., 2015). Whether LIANTI would be better than ATAC-seq, which currently can assess ~3% of the available sites in single-cell chromatin, is unclear, but as a single LIANTI transposon can provide an ability to amplify associated sequences (aRNA requirement) rather than needing two transposon integrations (PCR requirement of standard ATAC-seq protocol), there may be advantages to LIANTI for chromatin analysis.

The foundational structure of a cell's identity is the palette of expression possibilities presented by its genomic DNA. Indeed, LIANTI promises to overcome one of the major issues in understanding this foundation by providing a high-confidence, high-resolution, and high-coverage primary sequence of the single cell's genomic DNA. In answer to Alice's question, "Who in the world am I?" while LIANTI analysis of genomic DNA sequence won't tell us cell type, it can tell us which cells are distinct and outliers. This information is essential in understanding the genetic basis of cellular variability that underlies cellular identity.

REFERENCES

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Nature *523*, 486–490.

Carroll, L. (1866). Alice's Adventures in Wonderland / by Lewis Carroll; with Forty-Two Illustrations by John Tenniel (Macmillan).

Chen, C., Xing, D., Tan, L., Li, H., Zhou, G., Huang, L., and Xie, X.S. (2017). Science *356*, 189–194.

Dueck, H.R., Ai, R., Camarena, A., Ding, B., Dominguez, R., Evgrafov, O.V., Fan, J.B., Fisher, S.A., Herstein, J.S., Kim, T.K., et al. (2016). BMC Genomics *17*, 966–980.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Proc. Natl. Acad. Sci. USA *89*, 3010–3014.

Goryshin, I.Y., and Reznikoff, W.S. (1998). J. Biol. Chem. *273*, 7367–7374.

Krokan, H.E., Drabløs, F., and Slupphaug, G. (2002). Oncogene *21*, 8935–8948.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. (1990). Proc. Natl. Acad. Sci. USA *87*, 1663–1667.