

Review

Single-Cell DNA Sequencing: From Analog to Digital

Larry Xi

Digenomix Corp, South San Francisco, California, USA.

***Corresponding author:** Larry Xi. 7000 Shoreline Court #100, South San Francisco, CA 94080. Email: lxixi@digenomix.net

Citation: Larry Xi. Single-Cell DNA Sequencing: From Analog to Digital. *Cancer Research Frontiers*. 2017; 3(1): 161-169. doi: 10.17980/2017.161

Copyright: © 2017 Larry Xi. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Competing Interests: Larry Xi is CEO and Founder of Digenomix, which commercializes transposon barcoded (TnBC) library methodology.

Received Jan 30, 2018; Revised Mar 23, 2018; Accepted Mar 27, 2018. Published Apr 10, 2018

Abstract:

Single-cell DNA sequencing is emerging as a powerful and essential tool for interrogating heterogeneous tissues such as cancer samples. Its full potential in cancer research and diagnostics will only be realized when digital methodology is implemented and the cost is significantly reduced.

Keywords: single-cell, NGS, TnBC, heterogeneity, digital, UFI

Introduction

Single-cell DNA sequencing refers to sequencing of genomic DNA isolated from single cells. This DNA source is in contrast to that used for bulk DNA sequencing, in which genomic DNA is isolated from hundreds to millions of cells. Bulk DNA sequencing includes Sanger sequencing and next-generation sequencing (NGS) methods and has been widely used in genomic studies on human diseases. Bulk DNA sequencing has been effective in studies of homogeneous systems, but has been proven inadequate in analysis of solid tumors, which contains cancer cells with various degrees in heterogeneity as well as noncancerous fibroblasts, endothelial cells, lymphocytes, and macrophages. Noncancerous cells can contribute more than 50% of the total DNA extracted from tumors, potentially masking important genetic aberrations⁽¹⁾. Even when normal cells are removed, bulk sequencing still averages out both the heterogeneity of cancerous cells in a tumor tissue and genomic instability over time^(2, 3). Single-cell DNA sequencing can reveal the dynamics of mutations of

tumor cell subpopulations in unprecedented detail⁽³⁻⁹⁾

The mutation spectrum of cancer and requirements for single-cell DNA sequencing

The challenge of tumor biology and tumor diagnostics is the tracking of mutations and their association with cancer. Mutations in tumors consist of a full spectrum of genetic changes, including single nucleotide variations (SNVs) and structural changes: inversions; translocations; copy number variations (CNVs), which involve hundreds to millions of bases; and changes in ploidy, which are copy number changes at the chromosomal level or sub-chromosomal level⁽¹⁰⁾. Most bulk sequencing studies have focused on SNVs, as they have been regarded as the most frequent variations in genomes⁽¹¹⁾. In contrast, CNVs have been underreported in tumorigenesis⁽¹⁰⁾. However, in almost all relevant single-cell studies, CNVs have been reported to be involved in the largest portion of the mutated genome and associated with both early and late stages of tumor development^(8, 9). Therefore,

accurate and sensitive CNV detection should be an important aspect of single-cell DNA sequencing.

Another overlooked structural change is the ploidy change. In a cancer case study, Gray et al. reported that the entire cancer genome of a patient was amplified uniformly to 4n without significant SNVs in cancer-related genes, resulting in an almost identical NGS result to that of normal 2n cells. Gray concluded that a NGS-based approach alone is insufficient for cancer diagnostics⁽¹²⁾.

Given that cancer mutations are comprised of a full range of mutations, and that any one or any combination of mutations can lead to cancer, an ideal single-cell DNA sequencing method should be capable of capturing mutations of all kinds. This seems to be one of the most demanding requirements for single-cell approaches. For example, in Gray's case, they split the sample, with one portion subject to NGS interrogation for SNVs and the other portion investigated by FISH to examine ploidy. However, these two methods cannot be carried out on the same single cell.

A lot of single-cell interrogations require preparation of single-cell suspension.

A recent review covers singulating, sorting, and enrichment of targeted populations⁽¹³⁾. As the singulating process destroys spatial information of single cells in the tissue environment, ideally, each single cell in the suspension carries this information⁽¹⁴⁾, (WO/2017/075293), which assists in understanding and interpretation of the roles of various cells and their interactions in a 3D space.

Once a single-cell suspension is prepared, assays for single-cell RNA detection⁽¹⁵⁾¹⁵, single-cell protein detection⁽¹⁶⁾, single-cell DNA detection⁽¹⁷⁾, or co-detections⁽¹⁸⁾ can be carried out to address the needs of investigators. This review deals with only one detection method, DNA sequencing.

Single-cell DNA sequencing requires the construction of DNA libraries from single cells. The quantity of genomic DNA in a regular single cell is about 6 pg, which is inadequate for a complete NGS library for full coverage. The single-cell genome requires amplification by many thousand-fold before sequencing. This amplification process must generate a sufficient quantity of DNA in an unbiased and uniform way with minimal dropout. Because

amplification is the critical step of single-cell DNA library preparation, in most cases, the amplification method is used to represent the single-cell DNA library preparation method. DOP-PCR^(19, 20), MDA⁽²¹⁻²⁴⁾, and MALBAC⁽⁶⁾ are cases in the point. The same convention is followed in this article. Likewise, what are described as single-cell sequencing methods are actually library preparation methods.

Shallow sequencing libraries

In the past several years, several labs reported various single-cell library preparation methods (See Table 1). One method that has gained favor is degenerate oligonucleotide-primed PCR (DOP-PCR)⁽¹⁹⁾. As the name implies, the primer in DOP-PCR is a random hexamer sequence flanked by a defined sequence at its 5' end and by another defined sequence at the 3' end. In the first exponential PCR cycles, annealing and extension are done at a low temperature to ensure random priming at low stringency, and then PCR amplification shifts to a higher annealing temperature to achieve greater stringency. DOP-PCR generates a pool of 200–1000 bp fragments that can be converted to a NGS library.

By shallowly sequencing the single-cell libraries generated by DOP-PCR⁽²⁰⁾, the Hicks lab demonstrated robust determination of genome-wide copy number profiles in single cells, which was in high concordance with results of bulk analysis. They also examined breast cancer cells from human tissues and discovered extensive heterogeneity. The DOP-PCR single-cell method has been an economic choice to study tumor clonality in terms of CNVs. The lack of SNV information hinders its wider application.

An alternative way to detect CNVs by shallow sequencing was reported by Zahn et al., who constructed Nextera[®] libraries directly on the unamplified single-cell genomes⁽⁷⁾ and sequenced the libraries without further amplification. The quantity of DNA from a single cell is not sufficient for high coverage, but 2%–7% coverage was sufficient for their intended detection of CNVs. They demonstrated the presence of minor subpopulations of cells in cancer tissues, which were missed by bulk sequencing. Interestingly, to compensate for the lack of comprehensive SNV information, the authors generated a “bulk-equivalent” SNV map by aggregating the shallow reads of single-cell genomes

Table 1: Summary of single-cell DNA sequencing methods.

Method	Analog Mode						Digital Mode
	DLP	DOP-PCR	MDA	MALBAC	LIANTI	Targeted	TnBC
Reference	22	19, 20	21-24	6, 7	29	28	27
Genome amplification?	No	Yes	Yes	Yes	Yes	Yes	Yes
Random primers used in amplification?	-	Yes	Yes	Yes	No	No	No
Generation of library before amplification?	-	No	No	No	Yes	-	Yes
Absolute copy number detection ¹	No	No	No	No	No	No	Yes
CNV detection method	Abundance	Abundance	Abundance	Abundance	Abundance	No	Abundance or UFI ²
SNVs ³	Partial	Partial	Yes, analog	Yes, analog	Yes, analog	Yes, analog	Yes, digital
Intended for full coverage?	No	No	Yes	Yes	Yes	No	Yes
Amendable for targeted Sequencing?	No	No	Unreliable ⁴	Unreliable ⁴	Unreliable ⁴	Yes	Yes

Notes:

¹ Whether or not the absolute number of copies is determined dictates whether the SNV call is digital or analog.

² CNVs can be detected in the TnBC library based on abundance in shallow sequencing, or detected in absolute number based on UFI in high coverage.

³ When the copy number of the mutant is determined in digital and absolute number, the mutation call is digital. Otherwise it is analog.

⁴ The targeted region can be detected, but with unreliable results.

in silico. This SNV detection method cannot replace SNV detection at single-cell resolution.

The copy number call employed by both methods is based on the assumption that the number of reads mapping to a region is proportional to the number of times the region appears in the genome. A diploid region is expected to generate roughly twice as many reads as a haploid region, and two alleles are expected to generate roughly equal numbers of reads.

However, several unexplored factors may complicate the seemingly simple and straightforward copy number call. First, random primers may not be

completely random, so some regions may be primed more frequently than other regions; second, random annealing is sometimes haphazard, making the method inconsistent; third, the defined sequence at the 3' end is not uniformly distributed along the genome, and therefore some regions will be favored over other regions; fourth, similar to the priming issue in DOP-PCR, formation of clusters on the NGS platform may not be random, but skewed in a different way, which could add noise in read counts; fifth, the chance to generate fragments of different sizes may not be equal, while each fragment has an equal weight in

copy number call; sixth, the chance to generate fragments of different GC content may not be equal in a given experiment, but the fragments have equal weights in copy number calls; seventh, it is not certain how fragment size or GC content will affect the efficiency of cluster formation in different types of instruments or sequencing chemistry.

All these concerns dissipate when a digital sequencing method is used (see TnBC library below).

Deep analog sequencing libraries

Detecting SNVs has been the mainstay for many tumor studies and diagnostics. Two single cell preparation methods, multiple displacement amplification (MDA)⁽²¹⁻²⁴⁾ and multiple annealing and looping-based amplification cycles (MALBAC)⁽⁶⁾ are two frequently used methods to detect SNVs along with CNVs.

MDA uses a pool of degenerate nucleotides to initiate amplification by a highly processive Phi29 DNA polymerase. The process generates copious random genomic DNA⁽²⁵⁾. The amplification proceeds so fast that a large quantity of amplified long fragments from single-cell DNA are generated in a few hours. An NGS library from each amplified single-cell genome is then constructed using Nextera⁽²⁵⁾ or by ligation⁽²⁶⁾. Navin's lab at MD Anderson Cancer Center reported use of MDA libraries to study estrogen-receptor-positive (ER1) breast cancer and a triple-negative (TN) ductal carcinoma⁽⁸⁾. They showed aneuploid rearrangements occurring early in tumor evolution and remaining stable as tumor masses clonally expanded. SNVs were evolving gradually and gaining extensive clonal diversity. TN tumor cells had an accelerated mutation rate when compared with the ER1 tumor cells.

Another method, MALBAC, requires use of a pool of primers, each primer having eight variable nucleotides connected to a common 27-nucleotide tail at its 5' end. The first amplification cycles are quasi-linear, followed by an exponential phase, when the common 27mer is used as the primer at a higher annealing temperature⁽⁶⁾. Zong et al. used MALBAC to identify a total of 2.2×10^6 SNVs among all single cells by the MALBAC approach, comparable to the number (2.8×10^6) detected by the bulk method⁽⁶⁾.

Because cancer genomes often experience copy number changes and SNV detection is influenced by copy number changes, we decided to analyze how

reliably the reported CNV changes are detected using published data. We found that the abundance of reads per bin from the same assumed copy number regions in chromosomes in MALBAC and DOP-PCR data sets follows normal distributions, and their data are just slightly noisier than the bulk data set, the majority of the bins falling between 50% and 150% of the mean. In contrast, the read abundance per bin from a single-cell MDA library clearly did not follow normal distribution. The deviations were astonishingly large. A significant fraction of bins recovered 10 times more reads than the majority of bins⁽²⁷⁾. Unfortunately, neither the authors nor manufacturers of the commercial kits providing DOP-PCR, MDA, or MALBAC suggested any statistic tools to address the issue. It is desirable to have a confidence score, similar to Phred scores used in sequencing, attached to every CNV call and associated SNV determinations.

Targeted analog sequencing libraries

Pellegrino et al. reported a single-cell targeted analogy sequencing library construction method based on droplets. The workflow consists of two steps: first, individual cells are encapsulated in droplets and lysed, then the droplets containing the genomes of individual cells are paired with molecular barcodes and PCR amplification reagents with a panel of primers targeting genes for amplification⁽²⁸⁾.

By this single-cell approach, the authors identified TP53 as the more plausible founding, mutation, while bulk sequencing mistakenly inferred DNMT3A as the founding clone. The biggest advantage of this method is that many cells can be interrogated at low sequencing costs. However, as no CNV information is yielded, this method will miss a significant portion of oncogenic mutations⁽²⁸⁾.

Digital sequencing libraries

Last year, we and two other labs reported constructing single-cell DNA libraries by using transposition directly on single-cell genomes^(7, 27, 29). However, similarity stops there. We took a further step by turning the methodology into a digital method, making it able to detect absolute ploidy, the absolute copy number of any gene, and point mutations at the single-cell level.

Transposon barcoded (TnBC) library prep involves six steps: (1) Isolation of single cells in a

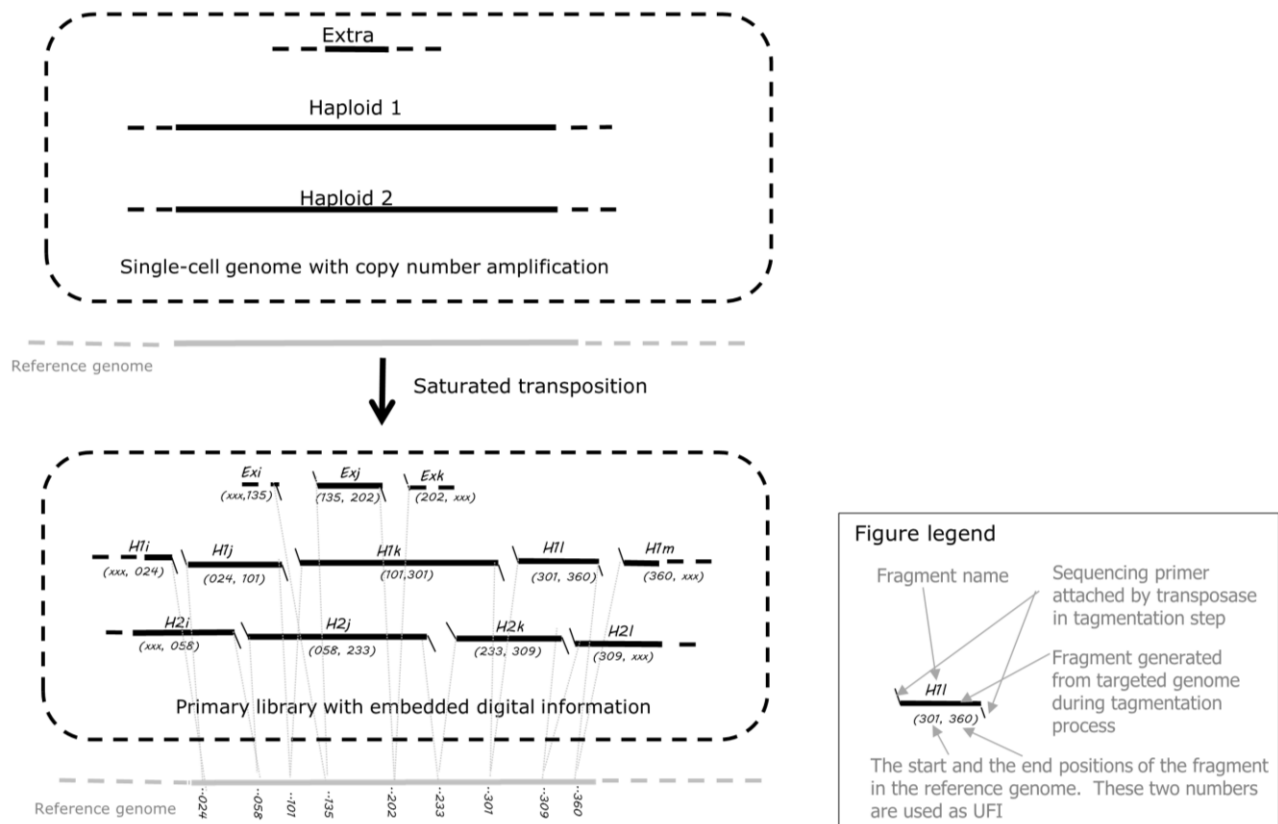


Figure 1. Generating primary TnBC library with UFI for each fragment by saturated transposition of single-cell genome. All numbers are fictional. “xxx” is a number that is not defined in the figure.

reaction vessel. The reaction vessel can be a PCR tube or a chamber in a microfluidic device, such as the Fluidigm® C1™ integrated fluidic circuit (IFC); (2) Lysing cells and removing histones from chromosome DNA. This was accomplished using proteinase K digestion. After digestion, proteinase K needs to be denatured, and the residue activity is further minimized by adding proteinase K inhibitor so that the transposase activity used in the next step will not be compromised; (3) Tagmentation of the single-cell genomic DNA with a saturating concentration of loaded transposase to make a primary library. The loaded transposases can be either Thermo Scientific™ MuSeek™ or Illumina® Nextera®. These loaded transposases are bioengineered to make NGS libraries. (4) Removal of transposase from tagmented DNA; (5) Amplification of primary library; (6) Addition of a sample barcode for each single-cell library.

Figure 1 explains how digitization was achieved during tagmentation to generate a primary TnBC library. Tagmentation starts with single-cell genomic DNA. A normal cell typically has two identical haploids, Haploid 1 and Haploid 2, with their coordination represented by the reference genome shown by the gray line under the box. A cancer cell may have an extra copy of the gene, which is represented by the dark line marked “Extra”. Tagmentation would break Haploid 1 into an array of contiguous fragments, ... H1i, H1j, H1k, H1l, H1m ..., Haploid 2 into another array of contiguous fragments, ... H2i, H2j, H2k, H2l ..., and the extra copy into the third array with ... Exi, Exj, Exk Because the tagmentation is a random process, even if the Haploid 1, Haploid 2 and Extra copy have the exact same sequence, they generate different sets of fragments. Each fragment is amplified, sequenced, and mapped to the reference genome. The combined start and end

points of each fragment is used as a unique fragment index (UFI), which is represented as two numbers in parentheses in Figure 1. Although these two numbers were not known before sequencing, they are embedded. As demonstrated by the figure, each haploid generates a unique array of fragments, and they can be represented mathematically. For example, Haploid 1 is made of [..., (xxx, 024), (024, 101), (101, 301), (301, 360), (360, xxx), ...]; Haploid 2 consists of [..., (xxx, 058), (058, 233), (233, 309), (309, xxx), ...]; the extra copy has [..., (xxx, 135), (135, 202), (202, xxx) ...]. It is clear that the absolute number of gene copies in a cell can be back-calculated by determining the number of unique arrays of fragments. This is why a TnBC library can be called a digital library.

The preparation process for a primary TnBC library is very similar to the regular NGS library prep method practiced by thousands of labs around the world ⁽²⁶⁾, except that a TnBC library is made from a single-cell genome. Because the DNA comes from a defined single cell, the resulting primary TnBC library carries the digital property. If the starting DNA comes from exactly 100 normal cells, saturated tagmentation would lead to 200 distinct arrays of fragments for each haploid on paper. Since recovering all 200 arrays of fragments by sequencing is impractical, the resulting library is practically non-digital. In common practice, DNA samples used for regular tagmentation libraries are not from a finite number of cells, so the libraries can hardly retain the digital features.

One of the great potentials of a TnBC library is that a sub-library of targeted regions can be enriched using pull-down probes ⁽³⁰⁾ from an amplified TnBC library. Because sub-libraries retain the digital features, high-quality data of the targeted area can be obtained, paving the way to reduce the cost of single-cell analysis by magnitudes.

The ability to detect changes in absolute copy number and ploidy in the early stages of tumorigenesis will be beneficial because these changes could be harbingers of important states in cancer development ^(31, 32). This will overcome the drawback of NGS noted by Gray ⁽¹²⁾. Liu's lab reported that polyploid giant cancer cells (PGCCs) are quite universally observed in cancer biopsies, especially from late stage tissues ⁽³³⁻³⁵⁾. Given that PGCCs exhibit high drug resistance and extremely high tumor-forming tendencies in mouse

models, single-cell DNA digital sequencing will give researchers a tool for molecular characterizations that no other method alone can offer.

Bioinformatics on single-cell digital sequencing libraries

Most existing bioinformatics tools have been developed for bulk sequencing, which is based on sampling from thousands to millions of identical copies. In addition, most of the tools are based on the hidden assumption that each cell has two copies of each gene ⁽³⁶⁾. As such, these tools can be good for cells that are homogeneous and diploid, but are of limited use for heterogeneous samples like cancer cells ⁽¹⁾. Mutations in genes of elevated copy numbers in subclones in a tumor are usually oversampled, and current bioinformatic tools do not differentiate the CNVs of all cells from CNVs that happen in a percentage of cells. Point mutation calls by most current bioinformatics tools do not consider copy number changes of the target genes. These shortcomings can explain why bulk sequencing may misinterpret the clonal structure of cancer and misidentify driver genes ⁽²⁸⁾.

For most single-cell DNA sequencing methods, DNA from single cells is amplified to a "bulk" level, and then libraries constructed from the amplified DNA is sequenced. Bioinformatics tools to analyze amplified single-cell genomes inherited similar issues because the assumptions do not match the realities of the genomic makeup of cancer cells.

Mutation calls on bulk libraries remain a difficult issue bioinformatically ⁽²⁾. In contrast, mutation calls on a TnBC library will be easier, more accurate, more precise, and quantitative, as each fragment has a UFI and the copy number of any region is an absolute digital. This benefit is not shared by other single-cell approaches.

Since amplification of DNA inevitably introduces errors, the detected variations in single-cell sequencing consist of biological mutations of cancer tissues and amplification errors. While no current bioinformatic tools were designed to tell the difference, UFI can help to differentiate the two.

For every two neighboring fragments from one haploid, the 3' part of the UFI of the upstream fragment shares with the 5' part of the UFI of the downstream fragment. Two neighboring fragments

can be haplotyped. Since a cancerous cell may carry no copy or multiple copies of a gene, single-cell digital sequencing will give a full picture of all haploids within a functioning unit, information that is lost in other approaches^(21, 37, 38).

Promises and challenges

Single-cell digital sequencing bears promises and elicits challenges. By using TnBC libraries, genomes of cancer tissue will be delineated in unprecedented detail. Even for homogeneous normal tissues, we expect that new findings may challenge old theories. We may find that the homogeneous cells are not as homogeneous as we believed to be⁽³⁹⁾. The fluidity of mammalian genome structure will be revealed with new details.

Although single-cell DNA sequencing has yielded more insightful information on cancer tissues than bulk sequencing, the cost associated with the single-cell approach is a major obstacle to wider application. Even as the sequencing cost of one genome may drop to \$100, the cost for hundreds of single-cell genomes would be tens of thousands of dollars. By using pull down probes to extract targeted regions of TnBC libraries, the sequencing cost for the majority of useful and actionable information may drop by magnitudes, making the single-cell approach affordable in routine cancer diagnostics.

Conducting single-cell digital sequencing also brings challenges. For example, TnBC library construction will require molecular biological tools such as transposons, primers, proteases, and polymerases to meet very stringent requirements. All reagents should be nuclease-free, since any nuclease contamination will destroy tiny, unique genetic material from single cells.

Secondary and tertiary bioinformatic tools will be needed to handle the simple but unique data structure of TnBC libraries. Developing ways to combine single-cell analysis with pathological

observation and clinical practice will require personnel armed with new knowledge, as well as new infrastructure.

Single-cell DNA sequencing will provide a powerful tool for diagnosing cancer and for guiding treatment. A sample obtained by biopsy provides genetic data on only one point in space and time. As surgery or therapy starts, ctDNA detection using NGS or digital PCR based on the mutation spectrum obtained from single-cell DNA sequencing will be a good complementary tool to monitor treatment progress and extend the use of mutation data obtained by single-cell DNA sequencing. One can foresee that single-cell DNA sequencing will be at center stage in the fight against cancers. I am optimistic that we will see wider and wider applications of single-cell digital sequencing in cancer research and diagnostics. The method will have numerous translational applications in the clinic, particularly in the area of molecular pathology and guidance of therapies.

Abbreviations:

bp: base pair
 CNV: copy number variation
 DLP: direct library preparation
 DNA: deoxyribonucleic acid
 DOP-PCR: degenerate oligonucleotide-primed PCR
 ER1: estrogen-receptor-positive
 FISH: fluorescence in situ hybridization
 MALBAC: multiple annealing and looping-based amplification cycles
 MDA: multiple displacement amplification
 NGS: next-generation sequencing
 pg: picogram
 PGCC: polyploid giant cancer cell
 SNV: single nucleotide variation
 TN: triple-negative
 TnBC: transposon barcoded
 UFI: unique fragment index

References

1. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat Commun.* 2015 Dec 9;6:10001. DOI: 10.1038/ncomms10001.
2. Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, et al. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell.* 2013 May 9;153(4):919-29. DOI: 10.1016/j.cell.2013.04.010.

3. Francis JM, Zhang CZ, Maire CL, Jung J, Manzo VE, Adalsteinsson VA, et al. EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.* 2014 Aug;4(8):956-71. DOI: 10.1158/2159-8290.CD-13-0879.
4. Navin NE. Delineating cancer evolution with single-cell sequencing. *Sci Transl Med.* 2015 Jul 15;7(296):296fs29. DOI: 10.1126/scitranslmed.aac8319.
5. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature.* 2011 Apr 7;472(7341):90-4. DOI: 10.1038/nature09807.
6. 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 Dec 21;338(6114):1622-6. DOI: 10.1126/science.1229164.
7. Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, Shah SP, et al. Scalable whole-genome single-cell library preparation without preamplification. *Nat Methods.* 2017 Feb;14(2):167-73. DOI: 10.1038/nmeth.4140.
8. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature.* 2014 Aug 14;512(7513):155-60. DOI: 10.1038/nature13600.
9. Gawad C, Koh W, Quake SR. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc Natl Acad Sci U S A.* 2014 Dec 16;111(50):17947-52. DOI: 10.1073/pnas.1420822111.
10. Talseth-Palmer BA, Scott RJ. Genetic variation and its role in malignancy. *Int J Biomed Sci.* 2011 Sep;7(3):158-71.
11. Brookes AJ. The essence of SNPs. *Gene.* 1999 Jul 8;234(2):177-86.
12. Gray PN, Dunlop CL, Elliott AM. Not All Next Generation Sequencing Diagnostics are Created Equal: Understanding the Nuances of Solid Tumor Assay Design for Somatic Mutation Detection. *Cancers (Basel).* 2015 Jul 17;7(3):1313-32. DOI: 10.3390/cancers7030837.
13. Hu P, Zhang W, Xin H, Deng G. Single Cell Isolation and Analysis. *Front Cell Dev Biol.* 2016;4:116. DOI: 10.3389/fcell.2016.00116.
14. Kuo CT, Thompson AM, Gallina ME, Ye F, Johnson ES, Sun W, et al. Optical painting and fluorescence activated sorting of single adherent cells labelled with photoswitchable Pdots. *Nat Commun.* 2016 Apr 27;7:11468. DOI: 10.1038/ncomms11468.
15. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature.* 2014 Jun 19;510(7505):363-9. DOI: 10.1038/nature13437.
16. Gong H, Holcomb I, Ooi A, Wang X, Majonis D, Unger MA, et al. Simple Method To Prepare Oligonucleotide-Conjugated Antibodies and Its Application in Multiplex Protein Detection in Single Cells. *Bioconjug Chem.* 2016 Jan 20;27(1):217-25. DOI: 10.1021/acs.bioconjchem.5b00613.
17. Szulwach KE, Chen P, Wang X, Wang J, Weaver LS, Gonzales ML, et al. Single-Cell Genetic Analysis Using Automated Microfluidics to Resolve Somatic Mosaicism. *PLoS One.* 2015;10(8):e0135007. DOI: 10.1371/journal.pone.0135007.
18. Genshaft AS, Li S, Gallant CJ, Darmanis S, Prakadan SM, Ziegler CG, et al. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* 2016 Sep 19;17(1):188. DOI: 10.1186/s13059-016-1045-6.
19. Arneson N, Hughes S, Houlston R, Done S. Whole-Genome Amplification by Degenerate Oligonucleotide Primed PCR (DOP-PCR). *CSH Protoc.* 2008 Jan 1;2008:pdb prot4919. DOI: 10.1101/pdb.prot4919.
20. Baslan T, Hicks J. Single cell sequencing approaches for complex biological systems. *Curr Opin Genet Dev.* 2014 Jun;26:59-65. DOI: 10.1016/j.gde.2014.06.004.
21. Fan HC, Wang J, Potanina A, Quake SR. Whole-genome molecular haplotyping of single cells. *Nat Biotechnol.* 2011 Jan;29(1):51-7. DOI: 10.1038/nbt.1739.
22. Gundry M, Li W, Maqbool SB, Vijg J. Direct, genome-wide assessment of DNA mutations in single cells. *Nucleic Acids Res.* 2012 Mar;40(5):2032-40. DOI: 10.1093/nar/gkr949.
23. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell.* 2012 Mar 2;148(5):873-85. DOI: 10.1016/j.cell.2012.02.028.

24. 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 Jul 20;150(2):402-12. DOI: 10.1016/j.cell.2012.06.030.
25. Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res*. 2001 Jun;11(6):1095-9. DOI: 10.1101/gr.180501.
26. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*. 2014;56(2):61-4, 6, 8, passim. DOI: 10.2144/000114133.
27. Xi L, Belyaev A, Spurgeon S, Wang X, Gong H, Aboukhalil R, et al. New library construction method for single-cell genomes. *PLoS One*. 2017;12(7):e0181163. DOI: 10.1371/journal.pone.0181163.
28. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet*. 2016 Mar;17(3):175-88. DOI: 10.1038/nrg.2015.16.
29. Chen C, Xing D, Tan L, Li H, Zhou G, Huang L, et al. Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science*. 2017 Apr 14;356(6334):189-94. DOI: 10.1126/science.aak9787.
30. Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, et al. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics*. 2011 Nov;10(6):374-86. DOI: 10.1093/bfpg/elr033.
31. Mutter GL, Pomponio RJ. Molecular diagnosis of sex chromosome aneuploidy using quantitative PCR. *Nucleic Acids Res*. 1991 Aug 11;19(15):4203-7.
32. Storchova Z, Kuffer C. The consequences of tetraploidy and aneuploidy. *J Cell Sci*. 2008 Dec 1;121(Pt 23):3859-66. DOI: 10.1242/jcs.039537.
33. Zhang S, Mercado-Uribe I, Xing Z, Sun B, Kuang J, Liu J. Generation of cancer stem-like cells through the formation of polyploid giant cancer cells. *Oncogene*. 2014 Jan 2;33(1):116-28. DOI: 10.1038/onc.2013.96.
34. Niu N, Zhang J, Zhang N, Mercado-Uribe I, Tao F, Han Z, et al. Linking genomic reorganization to tumor initiation via the giant cell cycle. *Oncogenesis*. 2016 Dec 19;5(12):e281. DOI: 10.1038/oncsis.2016.75.
35. Niu N, Mercado-Uribe I, Liu J. Dedifferentiation into blastomere-like cancer stem cells via formation of polyploid giant cancer cells. *Oncogene*. 2017 Aug 24;36(34):4887-900. DOI: 10.1038/onc.2017.72.
36. Hajirasouliha I, Mahmoodi A, Raphael BJ. A combinatorial approach for analyzing intra-tumor heterogeneity from high-throughput sequencing data. *Bioinformatics*. 2014 Jun 15;30(12):i78-86. DOI: 10.1093/bioinformatics/btu284.
37. Zheng GX, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, et al. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotechnol*. 2016 Mar;34(3):303-11. DOI: 10.1038/nbt.3432.
38. Zhang F, Christiansen L, Thomas J, Pokholok D, Jackson R, Morrell N, et al. Haplotype phasing of whole human genomes using bead-based barcode partitioning in a single tube. *Nat Biotechnol*. 2017 Sep;35(9):852-7. DOI: 10.1038/nbt.3897.
39. Cordaux R. The human genome in the LINE of fire. *Proc Natl Acad Sci U S A*. 2008 Dec 9;105(49):19033-4. DOI: 10.1073/pnas.0810202105.