Single-Cell Next-Generation Sequencing and Its Applications in Cancer Biology

Biaoru Li, Xiaomeng Zhang, and Jie Zheng

Abstract A complete set of DNA with its transcripts is defined as genome, which includes both the genes and the noncoding sequences of the DNA/RNA. After making advances in decoding different genomes across species, genomic techniques such as SNP microarrays and gene expression microarray have been synchronously developed to analyze the genomic functions. Now, scientists are able to take the study of genomics into deep consideration of biological evolution and mechanism of different diseases. However, there are still challenges with the genomic technology. Some tissues of human and animals, such as tumor tissues, contain multiple heterogeneous cells, making analysis extremely difficult. Additionally, some specimens have very few cells, such as circulating tumor cells. To fully study DNA genomic changes and its expression changes in cancer, single-cell genomic techniques have been broadly applied to fields such as cytogenomic diagnosis for specimens on glass slides, tumor cells in circulating blood, measurement of sensitivity and specificity of genomic analysis at tumor tissue level, mechanism of differentiation of cancer stem cell, etc. Recently, next-generation sequencing (NGS) has become an important tool in single-cell genomic analysis. Here, we systemically introduce single-cell NGS from single-cell sampling, single-cell NGS, and single-cell NGS-related bioinformatics into its application for tumor biology. This chapter also describes some advantages of single-cell NGS and addresses some challenges of single-cell NGS for genomics analysis due to the specimen features.

B. Li, M.D., Ph.D. (✉)
Department of Pediatrics, Medical College at GA, Augusta, GA 30912, USA
e-mail: BLI@gru.edu; brli1@juno.com

X. Zhang • J. Zheng
School of Computer Engineering, Nanyang Technological University,
Singapore 639798, Singapore

W. Wu, H. Choudhry (eds.), *Next Generation Sequencing in Cancer Research, Volume 2*, DOI 10.1007/978-3-319-15811-2_1

1 Introduction

DNA (deoxyribonucleic acid) composed of four bases and its double helical strand structure was first demonstrated by James D. Watson and Francis Crick in 1953 [1]. Since then, genes at the DNA and mRNA levels were broadly studied for their functions such as normal evolution of species and mechanism of human diseases as described by Drs. Er and Chang in 2012 [2]. During the early period of research, DNA sequencing techniques played important roles for studying gene structures and gene expression. In 1977, Frederick Sanger launched DNA sequencing technology that relied on DNA chain-termination method (Sanger sequencing) [3] and Walter Gilbert studied chemical modification and cleavage at specific bases of DNA as an early sequencing technology [4]. Sanger sequencing is described as the first-generation DNA sequencing due to its high efficiency and low radioactivity as delineated by Dr. Pareek in 2011 [5]. Following great accomplishments from the human genome project in 2002–2003, massively parallel sequencing systems called as next-generation sequencing (NGS) were brought about the world. In 2005, the 454 sequencing system provided massively parallel sequencing reading platform as reported by Margulies et al. in 2005 [6]; Solexa developed Genome Analyzer system as portrayed by Warren et al. in 2006 [7]; and Agencourt supplied SOLiD platform as explained by Mardis in 2008 [8]. All three NGS systems have similar features including high throughput and accuracy although there are differences such as the read lengths. Recently, the founder companies were bought by other companies. For instance, SOLiD system was purchased by Applied Biosystems in 2006; in 2007, 454 sequencing system was bought by Roche and Solexa system was picked up by Illumina as reviewed by Dr. Liu in 2012 [9]. The three systems exhibit their advantages including their read length, accuracy, and applications as presented in Table 1. NGS system has also been developed into compact model for small size of sample, such as Ion Personal Genome Machine (PGM) and MiSEQ. These two systems were extended by Ion Torrent and Illumina for their advantages in fast running and its cheap costs as shown in Table 2. Moreover, accompanied with increasing new modifications in NGS, a third-generation sequencing such as Single-Molecule

Table 1 NGS system comparison

Systems	454 GS FLX	HiSeq system	SOLiD system
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp
Reads	1 M	3 G	1200–1400 M
Output data/Run	0.7 Gb	600 Gb	120 Gb
Time/Run	24 h	3–10 days	7 days for SE
Advantage	Read length, fast	High throughput	Accuracy
Accuracy	99.90 %	98 %	99.94 %
Disadvantage	Low throughput	Short read assembly	Short read assembly
Cost/million bases	$10	$0.07	$0.13

Table 2 Compact NGS

Compact NGS	MiSeq	PGM
Sequencing method	Sequencing by synthesis	Semiconductor technology with a simple sequencing chemistry
Read length	Up to 2×300 bp	200–400 bp
Output	540 MB to 15 GB	30 MB to 2 GB
Sequencing time	4 h for 1×36 single read, 27 h for 2×300 bp end read	2.3–4.4 h for 200 bp reads 3.7–7.3 h for 400 bp reads
Sample preparation time	About 2 h	8 samples in parallel, less than 6 h
Input amount	Nanogram (Nextera)	μg

Real-Time (SMRT) has increased appliances in genomic studies. SMRT does not need PCR amplification and the nucleotides real-time signal of SMRT in enzymatic reaction can be captured by fluorescent (Pacbio) or electric current (Nanopore). Technically, NGS of whole-genomic DNA is called as DNA-Seq consisting of whole-genomics sequencing (WGS) and whole-exome sequencing (WES); NGS of whole mRNA is named as RNA-Seq; NGS of whole microRNA is said as miRNA-Seq and so on.

The research and development (R&D) of single-cell genomics in tumor biology has the advantage of requiring few cells and a cellular environment of mixed cells. For instance, development of clonal cell (such as cancer stem cell in cancer) occurred with subtle heterogeneity at an early period including few mutations and chromosomal rearrangements finally leading to massive cell proliferation and differentiation in a mixed tissue due to switch of the tumor cell program with enriched genomic changes. In the tumorigenesis, very few cells are available in the early period while mixed-cell tumor tissues arise in late tumor development according to Dr. Breivik's study in 2005 [10]. All these reasons require R&D of single-cell genomic techniques to study the tumorigenesis. In addition, genomic diagnosis for a given type of cells in mixed-cell tumor tissue can only adopt very small numbers of cells such as clinic biopsy specimens or single-cell isolated from laser capture microscopy of tumor tissues. The single-cell technique with downstream genomics needs to be applied itself from cells on slides in molecular pathology and cytogenetic. Moreover, it is necessary for biomarker discovery of tumor cells in circulating blood as described by Liberko et al. in 2013 [11]. Several years earlier, genomics of identification and quantification have been developed into single-cell genomic level including Array-CGH and SNP-microarray for DNA genomics and mRNA microarrays, subtractive cloning and differential display (DD) for mRNA genomic profiles as illustrated by Ning et al. in 2014 [12]. Technically, each single-cell genomic analysis and diagnosis has its own disadvantages and advantages. After NGS was applied in 2007 and developed into single-cell genomic technique in 2010, single-cell NGS techniques have allowed physicians and scientists to use the important tools for single-cell diagnosis as explained by Ebenezer et al. in 2012 [13].

In order to distinctly advocate single-cell NGS, here we will first introduce the single-cell techniques and then present single-cell NGS techniques with downstream

single-cell NGS bioinformatics. Finally, we will briefly review applications to the study of cancer biology by using the single-cell NGS techniques. In conclusion part, we will also discuss advantages and disadvantages of applying different single-cell genomic techniques.

2 Single-Cell Technique

Tumor specimens of animal and human tissue often contain multiple cells. Different DNA changes and different gene expression profiles in a given type of cells coexist in the same specimen of animal and human tissue. Theoretically, important findings of genomic-DNA SNP profile or mRNA expression profiles will be unclear for a certain type of cells if we make use of tissue-level genomic profile. Therefore, pure or representative single cells will provide the most precise analysis possible of these subtle gene expression patterns in the given type of cells. Here, in order to explicitly discuss single-cell NGS, two techniques, or single-cell sampling and DNA/mRNA amplification from a single cell will be first introduced.

2.1 Single-Cell Sampling

As shown in Table 3, flow-cytometric cell sorting (FACS) and laser-based microdissection of tumor tissues provide ways to isolate single cells for DNA genomics change and gene expression profiling in a given type of cells. In FACS system, cells labeled with fluorescent signals in solution can be isolated based on a specific biomarker such as a tumor antigen attached to an antibody labeled by a fluorescent signal. At present, FACS can specifically separate targeted cells and collect the single cell into 96 wells for downstream genomics (AmpliGrid by Advalytix) as reported by Brück et al. in 2010 [14]. Although FACS and multicolor FACS can

Table 3 Single-cell sampling

Methods	Advantages	Disadvantages
Laser-capture microdissection	Microenvironment and local data	Theoretical damage to the target cell
Laser-assisted mechanical microdissection	Microenvironment and local data	Laborious
Laser-catapult microdissection	Very little contamination with microenvironment and local data	Special slides
Flow-cytometric cell sorting	Auto- and rapid separation into 96-well plate	Limit in some cells such as neuron without microenvironment data

isolate and sort homogeneous cells, even single cell, three challenges limit their applications: (a) FACS cannot be subject to some types of cells such as neurons; (b) intracellular biomarker cannot be well defined and sorted by FACS; (c) the tumor microenvironment of a cell cannot be evaluated by FACS. The microdissection technique can avoid the aforementioned three limitations. In 1976, the use of lasers in tissue microdissection has been reported by Meier-Ruge et al. [15]. In contrast to single-cell FACS, microdissection allows both rapid in vivo localization and ability to analyze the cellular microenvironment as depicted by Schutze et al. in 1998 [16]. At present, three microdissection systems have been broadly developed as reported by Li in 2005 [17]: (1) laser-assisted mechanical tissue microdissection, (2) laser pressure catapult microdissection, and (3) laser capture microdissection (LCM). Laser-assisted mechanical tissue microdissection can focus on small target cell areas, reducing the chance of contamination with neighboring cells as portrayed by Emmert-Buck et al. in 1996 [18]. Although the concept of using a laser to dissect out individual cells is quite simple, the technique is laborious. Laser pressure catapult microdissection concentrates on an interesting region with a high-energy cutting laser. Following a low-power laser sets the depth of the tissue section, a pressure wave then separates the targeted tissue from the slide and catapults it into a receptacle. The high precision of the thin beam laser is sufficient to isolate subcellular targets such as chromosomes. The absence of physical contact between the surrounding tissues and the collection apparatus results in a much lower incidence of contamination. In laser capture microdissection, a thin ethylene vinyl acetate film is mounted on the tissue section. After an infrared laser heats and melts a cell of interest, the resolidified plastic film binds directly to this cell and catches it as reported by Fend et al. in 1999 [19]. Now, all of three systems are commercially available for laboratory studies in animals, plants, and human beings.

2.2 Genomic Amplification from Single Cells

In a human diploid cell, the quantity of DNA is a constant or 6.6 pg of each diploid single cell (three billion base pairs multiply two for diploid and multiply 660 for molecular weight of each base pair), although about 5 pg per human cell is harvested in real experiment. After more than 10 years effort, genomics DNA isolation and amplification from single cells are very mature called as whole-genome amplification (WGA) (Table 4). Now, three companies [Genomeplex kit (Sigma), Picoplex kit (Rubicon), and Genomiphi kit (GE)] are commercially available for genomics DNA isolation and amplification for single cells as, respectively, delineated by Fiegler et al. in 2007 [20], Kurihara et al. in 2011 [21], and Pan et al. in 2008 [22]. All three products work very well for genomic DNA amplification although there are some subtle differences such as base pair length and PCR amplification techniques (see Table 4) and although some scientists prefer to perform MDA (Multiple Displacement Amplification) from the product to process DNA of single cell.

Table 4 Single-cell genomic amplification

Genomic types	Methods	Advantages	Primers	Amplification products	Commercial available
DNA amplification	Primer-extension preamplification (PEP)	Stable genomic amplification	Degenerate oligonucleotide primer (DOP)	400–1,500 bp	Sigma (Genomeplex)
	Two cycling	Stable genomic amplification	Degenerate oligonucleotide primer (DOP)	400–500 bp	Rubicon (picoplex)
	Phi	>1 kb	Multiple strand displacement (MSP)	Variable for repeat	GE (Genomiphi)
mRNA amplification	Tang's	Longer fragment	UP1 for olig-T and UP2 with OligoA	Longer until 3 kb	Combined kit with design
	Smart-Seq	Easy and commercial available	UP1 for olig-T and UP2 with CCC switch	Nanogram start	SMARTer® Ultra™ Low RNA Kit
	STRT	Longer fragment with barcoding sequencing	UP1 for olig-T and UP2 with CCC switch and barcoding seq	Longer until 2 kb	Combined kit with design
	Cell-Seq	Linear amplification	UP1 for olig-T and barcoding seq and UP2 with multiplex cell seq	Sensitivity	Combined kit with design

The quantity of mRNA in a single cell is greatly different, 1.0–20 pg (about 5×10^5–10×10^7 molecules) based on the cell size, cell function, and cell differentiating stage as described by Ambion in 2004 [23]. Although some scientists try to isolate RNA from single cells, most of scientists prefer to use a crude cell lysate without purifying procedures as reported by Klebe et al. in 1996 [24]. This protocol has two important advantages. First, it ruptures the cells and releases the RNA directly into a cell lysis buffer without loss of RNA. Moreover, the heating step to rupture cells inactivates endogenous RNase for protecting RNA from degradation. Theoretically, mRNA amplification should be applied in single-cell genomic technique. Now, four mRNA amplifications strategies have been developed into single-cell RNA-Seq. Their performances with their amplification mechanism, primers design, and PCR product sizes are listed in Table 4: Smart-Seq (switching mechanism at the 5′ end of the RNA transcript), STRT techniques (single-cell tagged reverse transcription), CEL-seq, and Tang's method, as respectively reported by Ramsköld et al. in 2012 [25], Lobo et al. in 2009 [26], Hashimshony et al. in 2012 [27], and Tang et al. in 2009 [28]. Here, two basic mRNA amplification principles will be first launched: mRNA amplification (aRNA) and PCR-based cDNA amplification. The aRNA procedure begins with total RNA or poly(A)+RNA that is reversely transcribed using an oligo (dT) primer containing a T7 RNA polymerase promoter sequence. After first-strand synthesis, the reaction is treated with RNase H to fragment the mRNA. These fragments serve as primers during a second-strand synthesis reaction that produces a double-stranded DNA template for transcription. rRNA, mRNA fragments, and primers are removed before using the cDNA template to produce linearly amplified aRNA. The amplification yields can reach 1,000- to 5,000-fold following two rounds of in vitro transcription. RNA amplification is commercially available and has been increasingly reported in gene expression studies as described by Eberwine in 1996 [29]. PCR-based amplification has two protocols: specific profile and global profile applications. Specific profile methods such as RT-PCR or multiplex RT-PCR reactions are sensitive at the single-cell level, especially in nested PCR. Because the genes studied using these methods are preselected, it can only be applied to known genes. Global PCR-based approaches have been developed in genomic analysis. Two approaches are commercially available, homomeric tailings and 3′-(3-primer-end) amplification (TPEA). The homomeric tailings as designed by Toellner et al. in 1996 [30] use terminal deoxynucleotide transferase-generated homomeric 3′ tails to the first-strand cDNA. After RT-PCR and 3′ tailing addition and PCR amplification, it has been applied to the analysis of single-cell global gene expression. Even though homomeric tailings can be used effectively in global profile analysis, many of the cDNA copies are not full length and shorter cDNAs are preferentially amplified. 3′-end-amplification (TPEA) as reported by Dixon et al. in 1998 [31] is a randomized amplification of mRNA using an oligo-dT primer together with a 5′ primer containing a random pentamer. It can enable the detection of both high- and low-abundance mRNA transcripts from single cells.

3 Single-Cell Next-Generation Sequencing

3.1 Single-Cell DNA-Seq

Routine genomic DNA performance including next-generation sequencing and SNP microarrays requires sufficient and high-quality DNA. For single-cell genome analysis, as previously discussed, a special process termed whole-genome amplification (WGA) is added as illustrated in Fig. 1a. The WGA process can amplify the whole DNA population producing large amounts of DNA from a single cell whose quantity is comparable to routine genomic DNA. Due to the exponential amplification, three challenges will be created in the amplification process: amplified sequence bias during WGA, genetic material contamination caused by heterogeneous amplification, and genomic dropouts caused by tiny DNA materials as described by Gole and Gore in 2013 [32]. To overcome the three obstacles, each step of the process must be performed under quality control (QC) with Good Management Practice (GMP) compliance.

After performing single-cell sampling and WGA from target cells, high-throughput sequencing using WGA DNA is carried out using routine genomic DNA-Seq, which is briefly elaborated as follows. After the genomic DNA library is prepared, genomic DNA is fragmented and purified for enzymatic processes such as DNA end repair, A-tailing, adaptor ligation, DNA fragment size selection, and DNA fragment amplification. Following library amplification, the library is quantified

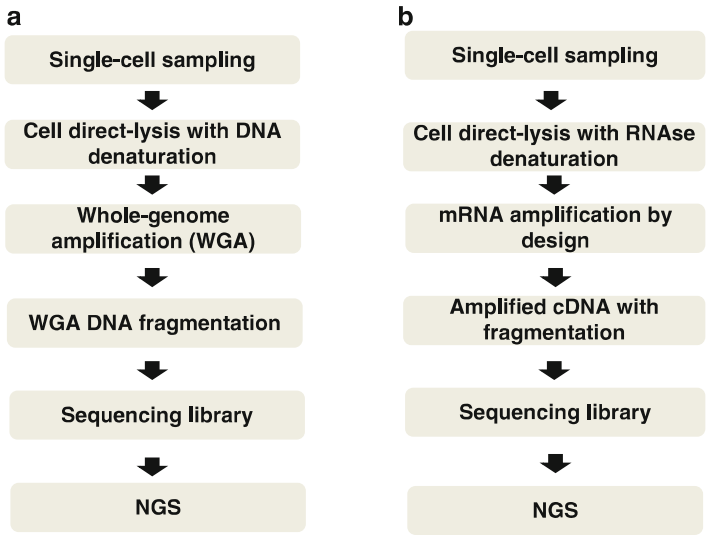

Fig. 1 The diagram of next-generation sequencing procedure: (**a**) Single-cell DNA-Seq workflow: after single-cell sampling and whole-genome amplification and fragmentation, library is quantified and is submitted to the sequencer; (**b**) single-cell RNA-Seq workflow: after single-cell sampling and whole-genome amplification by design and fragmented, library is quantified and is submitted to the sequencer

using real-time PCR and a predetermined amount of DNA library is submitted to the sequencer. The exact protocol of all of these steps is described in different NGS systems by Landau et al. in 2014 [33].

3.2 Single-Cell RNA-Seq

Routine RNA-Seq provides accurate quantification of mRNA expression levels with entire transcript lengths. Routine sampling for RNA-Seq is largely based on traditional molecular biological protocols including the basic steps of poly-(A)+RNA isolation, fragmentation, reverse transcription, and amplification before the actual sequencing takes place. The selection of poly-(A)+RNA is usually performed in order to suppress rRNA and tRNA. The fragmentation step is carried out in order to produce many short RNA or DNA fragments that represent the original transcript.

Following the basic principles of transcripts amplification discussed previously, four strategies of mRNA amplification for single-cell RNA-Seq have been developed into the single-cell level: Smart-Seq (switching mechanism at the 5′ end of the RNA transcript), STRT (single-cell tagged reverse transcription), CEL-seq, and Tang's amplification. After performing single-cell sampling and transcripts amplification and fragmentation, high-throughput RNA-Seq is performed using routine RNA-Seq. As DNA-seq performance, after a genomic RNA library is prepared, genomic fragment is purified for enzymatic processes such as end repair, A-tailing, adaptor ligation, library fragment size selection, and library fragment amplification. The library of accurate quantity is also measured by real-time PCR and then accurate amount of library is submitted to the sequencer as shown in Fig. 1b. The detailed protocol of NGS is described in different NGS platforms by Panagopoulos et al. in 2014 [34].

4 Single-Cell NGS-Related Bioinformatics

Next-generation sequencing (NGS) is a radical breakthrough at whole-genome level, offering unprecedented data depth not found in previous Sanger sequencing technology. A number of NGS platforms are developed based on different sequencing technologies, the details of which are beyond the scope of the work. Here, we simply highlight that all NGS platforms perform a common task that is to sequence millions of small fragments of DNA in parallel. Consequently, each of several billion bases in the target species or disease genome is sequenced multiple times, leading to a high level of data depth and accuracy. By making use of appropriate bioinformatics analysis tools, these fragments of data can be pieced together whereby individual reads are mapped to a species-specific reference genome. The mapped genome is highly sought after as it may shed light on the unexpected DNA variation or the quantity of RNA expression. In this section, we will focus our discussion on bioinformatics analysis related to single-cell NGS, to be more limited, single-cell DNA and RNA sequencing.

4.1 Single-Cell DNA-Seq Bioinformatics

As discussed in Sect. 3.1, single-cell DNA-Seq faces three obstacles (amplified sequence bias, genetic material contamination, and genomic dropouts). These obstacles make the single-cell DNA sequencing data inaccuracy. The impaired single-cell sequence data cannot be analyzed by most bioinformatics tools developed for bulk cell sequencing. To tackle this problem, some new bioinformatics tools have been designed for analyzing single-cell sequencing following the eruption of single-cell sequencing data. In this section, we will describe the applications of single-cell bioinformatics in analyzing single-cell WGS or WES.

Theoretically, single-cell sequencing data open up an opportunity to study genealogy of an individual tumor cell. The genealogy of the tumor cell unveils the complete picture from the earliest signs of mutation until accumulated heterogeneous tumor. If the mutation pedigree is constructed in a systematic manner, any unrelated lineage can be easily identified. In early model, Navin and his colleagues performed copy number variation analysis on breast tumors using low coverage single nucleus sequencing as reported in 2011 [35]. Their study aimed to explain clonal evolution of the tumors. They constructed a phylogenetic tree based on sample cell numbers and subpopulations based on the distances in the tree between the samples. Following Navin's analysis, Hou et al. used exome sequencing data from 58 single cells of an essential thrombocythemia (ET) tumor and Li et al. utilized exome sequencing data from 66 single cell samples of a bladder transitional cell carcinoma, respectively, to perform mutation to study subgroup of the samples in 2012 [36, 37]. All of these studies clearly illustrated clonal evolution using single-cell sequencing. In 2014, Kim and his colleagues continued working on the model of mutation pedigree including temporal and lineage relationships among DNA sequence mutation sites. They applied their algorithm in an 18-sites map as a lineage [38] which Dr. Hou had previously identified as lineage family in their single-cell sequencing dataset so that Dr. Kim proposed a new method to construct evolutionary mutation tree, which could indicate the temporal order relationship between mutation sites. They also proposed a method for estimating the proportion of time starting from the earliest mutation event and from the emergence of most recent common ancestor, respectively, toward the end of mutation. In conclusion, many new bioinformatics tools designed can be used to analyze single-cell DNA sequencing data as illustrated in Fig. 2.

4.2 Single-Cell RNA-Seq Bioinformatics

On top of the usual RNA-seq processes, single-cell RNA-Seq performance requires two additional processes: single-cell sampling and RNA amplification. Although four techniques of RNA amplification available from company products have been developed for RNA-seq, sensitivity and specificity of genomic expression after

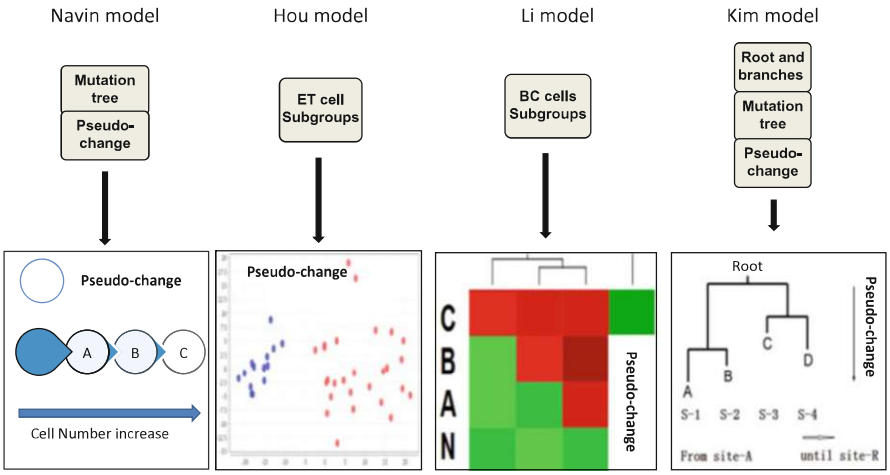

Fig. 2 Four models of single-cell NGS to detect DNA informative change and pseudo-change. Navin model is to study clonal evolution of mutation tree from A to C according to enhancement of tumor cell number from A to C related with mutation order pattern and pseudo-change based on unrelated information; Hou model is to use Principle Component Analysis (PCA) as model: PCA-1 as *x*-axis and PCA-2 as *y*-axis, *blue spots* from normal cells and *red spots* from tumor cells, all exome sequencing data from 58 single cells of an essential thrombocythemia (ET) tumor to study ET cell mutation subgroups; Li model is to utilize mutation pattern in heat map to study mutation subgroup from 66 single cell samples of a bladder transitional cell carcinoma, in which N is control from normal cell, A is mutation pattern explained as earliest cell, B is second, and C is third described as continuance pattern with column indicating different genes, *red* is higher frequency mutation and *green* is lower mutation frequency; Kim model is working on mutation pedigree among DNA sequence mutation sites in which they have 18 sites defined as branches (total 18 sites from site-A to site-R although the figure show only from S-1 to S-4). They can map root to branches including mutation tree and pseudo-change

RNA amplification have been carefully analyzed by our scientists. Our single-cell RNA-Seq technique from CD8 cell of tumor-infiltrating lymphocytes (TILs) demonstrated that fragments were 250–450 bp after fragmentation, amplification, and adapter addition. There were 11.6 million reads mapped in raw sequencing reads (19.6 million). The numbers of mapped genes, mapped transcripts, and mapped exons were 31,332, 41,210, and 85,786, respectively. All QC results illuminated that RNA-seq techniques could be used for single-cell genomic performance. Analysis of the mapped genes showed that the number of genes mapped by RNA-seq (6,767 genes) was much higher than that of differential display (288 libraries) among similar specimens which we had previously developed. The single-cell RNA-Seq can detect gene splicing using different subtype by using TGF-beta analysis. The results using Q-RT-PCR assays demonstrated that sensitivity was 76 % and specificity was 55 % from the single-cell RNA-Seq technique although some gene expression was still missing (2/8 genes). Therefore, the results support that RNA-Seq technique is feasible to analyze single-cell mRNA specimens as described by Xu et al. in 2013 [39].

5 Single-Cell NGS Application

5.1 Pathological Diagnosis

The pathologic diagnosis of tumors relies on cell morphology, tumor cell arrangement, and its infiltrating into normal tissue. The diagnosis of cytogenetics depends on chromosome structure with its number and arrangement change. Following the development of single-cell techniques, a new term, "Single-Cell Diagnosis," has arisen in disease diagnosis in which single-cell genomic diagnosis is involved in molecular pathology and genetics, especially for tumor diagnosis. Now, single-cell genomic diagnosis can be applied for many clinical specimens, such as surgical specimens, biopsy specimen, and tumor cell from circulating blood. Single-cell genomic analysis and diagnosis have much more advantages than other diagnosis. For instances, along with genomic analysis from tumor cells, genomic data can convert pathological changes of tumors into biomarker discovery; in pace with genomic analysis, single-cell genomic analysis can link tumor diagnosis into targeted therapeutics so single-cell genomic diagnosis can be used for personalized therapy; in addition, single-cell genomic analysis and diagnosis can be developed for several other applications such as study of mechanism of tumorigenesis as explained by Macaulay et al. in 2014 [40]. Single-cell NGS of cancer diseases is one of earliest applications for next-generation sequencing. Because single-cell NGS plays a very important role in cancer biomarker discovery and personalized therapy, now, Genomeplex kit from Sigma Inc, Picoplex kit from Rubicon Inc., and Genomiphi from GE all participate in the research and development of single-cell NGS related to genomic analysis and diagnosis such as single cell from slides or single cell from circulating blood of tumor disease.

5.2 Biomarker Discovery

Early diagnosis and treatment is an important impact to reduce the mortality of tumor disease. Currently, some of screening tools (CT, X-ray, mammography, and invasive needle or surgical evaluation for cancer disease) are not sensitive enough for early detection of the diseases, thus some of tumors cannot be treated at an early stage. Theoretically, if some special biomarkers can define each type of tumor cells, it should be the best way for early diagnosis although it is difficult to define now. Genomic technologies have allowed scientists to discover some special biomarkers from thousands of gene expression profiles and evaluate functions of special biomarkers to obtain a global view of tumor cells. After tumor cells are defined on slides or after tumor cells are harvested from circulating blood, single-cell genomic diagnosis is a rational module to define the tumor biomarkers. Single-cell RNA-Seq has been begun to apply for biomarker discovery including their therapeutic targeting. Now, Single-Molecule Real-Time (SMRT), third-generation sequencing has been successfully applied for biomarker discovery from glioblastomas as delineated by Meldrum et al. in 2011 [41].

5.3 Therapeutic Targeting Identification

Recent development of cancer research has enabled scientists to understand the difference of certain type of cancers to respond to chemotherapy analyzed by single-nucleotide polymorphisms (SNP) and genome-wide association studies (GWAS). GWAS analysis, one of genomic medicine, emphasizes different responses of drugs in a certain SNP, called as pharmacogenetics. Because SNP is the information archive but most of the FDA compounds and drugs are directed at phenotype alteration (such as RNA or proteins), not direct to DNA archives, the phenotype products of genotype change have also a great impact on the genomic medicine. Now gene expression profiles related network are used to uncover genomic expression signature (GES, previously called as therapeutics targeting identification, TI) to discover sensitive drugs, broadly called as pharmacogenomics. According to both concepts, drug discovery based on either GWAS or genomic expression signature related network is increasingly developed in treatment of drug-resistant tumor diseases as discussed below.

5.3.1 GWAS Related with Therapeutic Targeting and Personalized Therapy

Cancer stem cells (CSCs) and drug-resistant tumor cells mixed in tumor tissues play an important function in the tumor development and progression. CSCs drive the metastatic spread of cancer and are able to resist conventional therapies so that the disease is difficult to be completely eradicated. If a specific mutant or fusion protein, which results in tumor development or resistance of conventional therapy, can be uncovered by GWAS analysis, a specific targeting compound or Ab to target this mutant or fusion protein will offer a new therapeutic tool to treat drug-resistant tumor cells. According to this concept, several special antibodies and compounds to these mutant or fusion proteins have been routinely used to treat drug-resistant tumors called as molecular therapy or targeted therapy (or one kind of personalized therapy) as illustrated by Guan et al. in 2012 [42]. Single-cell DNA genomics can definitely uncover mutant and fusion proteins by GWAS analysis. Now single-cell NGS-related GWAS analysis is being developed in the tumor cells from slides or from circulating blood of tumor disease.

5.3.2 Network Related with Personalized Therapy

In clinical fields, besides GWAS-related personalized therapy as discussed in Sect. 5.3.1, genomic (or proteomics) expression profile, a second module of personalized medicine of special therapeutic strategies, is going to extend into different diseases. The personalized medicine is directly tailored for physicians to prevent and care individual patient relying on personal genomic expression profiles. It is often called as “the right treatment for the right person at the right time.” All examples of successful personalized treatments require a rational clinical genomic

expression analysis based on R&D of clinical genomic expression diagnosis, and we have successfully established a bioinformatics module from genomic expression profile for personalized therapy in 2008 [43]. The module included mRNA genomic expression profile mined from a specimen, genomic expression signature discovered by quantitative network, and sensitive drugs uncovered from drug bank. Now, after Single-Molecule Real-Time (SMRT), a third-generation sequencing, is brought into the new fields, single-cell genomic diagnosis (such as single cell from slides or single cell from circulating blood) related with discovery of genomic expression signature will make great contribution for the personalized therapy.

5.3.3 Network Related with Personalized Immunotherapy

Personalized immunotherapy is a major breakthrough in cancer immunotherapy including genetically engineered T cells by chimeric-antigen-receptor to kill own tumor cells, using own tumor cells to develop a personalized vaccine to kill own tumors, and activating T-cells quiescent network using own T-cells to kill own cancer cells. CD8 cells from tumor infiltrating lymphocytes (TILs) can directly and specifically recognize and kill own tumor cells after they are activated and expanded ex vivo. If the cells, which have been attached to tumor cells and will recognize own specific tumor antigen, are harvested by single-cell technique, the single-cell genomic profiles will play an important role in a personalized immunotherapy. We have studied single-cell genomic profiles from TILs for more than 10 years. According to concepts of immunology and tumor immunotherapy, CD8 cell of TILs has two obvious advantages: (a) the CD8 T-cells have function of MHC class I to access tumor cells; (b) the CD8 T-cell is specifically recognizing tumor antigen to kill tumor cells. If we uncover genomic profiles related to the specific CD8 cell from TILs which has been specifically accessing tumor cells obtained by single-cell technique, the genomic profiles can decode CD8 cell quiescence. Under culturing the TILs ex vivo combined with dequiescence and with specific function activity by network analysis in silico, the cultured TILs have much stronger function to kill tumor cells. As we all know, CD8 T-cell is an earliest cell model to be developed by single-cell genomic technique. In order to develop personalized immunotherapy to treat tumor diseases, we have developed single-cell genomic techniques from single-cell differential display, single-cell microarray until single-cell NGS as reported by Zhang et al. in 2009 [44]. Now, single-cell NGS-related quantitative network is being developed in personalized immunotherapy to treat advanced tumor diseases.

5.4 Tumorigenesis Related to Cancer Stem Cell

As discussed earlier, single-cell NGS can identify the earliest mutations and set phylogenetic tree of tumor cells. All of these pedigree trees can address clonal evolution. The earliest mutation site is located at the root and then gradual extension

from the root to other sites in the tree. Eventually, the trees can be used to estimate the earliest mutation event of the tumor to the most recent common ancestor (MRCA) of the cells. Because very early CSCs have very few cells, the CSCs definitely require single-cell NGS to mine genomic change related with pedigree tree as described by Jiao et al. in 2014 [45]. If genomic profiles are discovered to the CSCs tumorigenesis, a new generation of therapeutic strategies including GWAS-based molecular therapy, personalized therapy, and personalized immunotherapy as all discussed earlier will appear in the treatment of tumor diseases.

6 Conclusion

Single-cell techniques with downstream genomic analysis have emerged in application of single-cell specimens from glass slides or circulating tumor cell and mixed cells tumor tissue. Recently, next-generation sequencing (NGS) has become an important tool in single-cell level. According to current R&D of single-cell NGS as given in Table 5, single-cell RNA-Seq has same significant advantages as routine RNA-Seq. As single-cell RNA-Seq is adopted to analyze transcriptome profiles, reported results include quantitative mRNA expression, RNA splicing, and new transcripts. Moreover, if RNA-Seq data using BWA platform mapping with GATK/Samtool analysis, which compare DNA reference and SNP reference of genome, are utilized to analyze the genomic profiles, they also can uncover SNP, deletion, and insertion in the exome region, so that results of single-cell RNA-Seq are much better

Table 5 Comparison of single-cell genomic techniques

Genomic types	Methods	Advantages	Disadvantages
mRNA transcriptome	Single-cell NGS	Genomic expression with splicing and exome SNP, deletion, and insertion	Bias, dropouts, and contamination
	Single-cell microarray	Genomic expression with good model for normalization	Bias, dropouts, and contamination
	Single-cell differential display	Genomic expression with very good specificity	Bias, dropouts, and contamination with lower sensitivity
DNA genomic change	Single-cell NGS	DNA level change with genetic tree discovery and new SNP discovery	Bias, dropouts, and contamination
	Single-cell SNP microarray	DNA level change with good bioinformatics support	Bias, dropouts, and contamination with limiting known SNP
	Single-cell ACGH	Chromosome level change with good SOP for clinical diagnosis	Resolution level only for chromosome and large deletion and insertion

than those from single-cell microarray. As most of single-cell genomic techniques, single-cell NGS still has three challenges: bias produced by amplification, genetic material contamination caused by heterogeneous amplification, and genomic drop-outs caused by tiny DNA materials. In order to avoid the three problems, single-cell performance definitely requires a GMP regulation with QC monitor. Technically, several single cells such as 5–10 are minimal cell numbers for DNA-Seq and several single cells are optimal selection for RNA-Seq due to cell dropout from single-cell sampling process. The new genomic technique and its analysis will be developed into diagnosis of molecular pathology and cytogenetics of cancer diseases, discovery of differentiation biomarkers of cancer stem cells, and inducing therapy for cancer stem cells; furthermore, clinical application of molecular therapy, personalized therapy, and personalized immunotherapy for cancer patients.

Acknowledgments Under the support of Dr. H.D. Preisler, we have set up the method to analyze single-cell genomic profiles of CD3, CD4, and CD8 from TIL and tumor cell from solid tumors. The work is supported by both National Cancer Institute IRG-91-022-09, USA, for Dr. Biaoru Li and AcRF Tier 2 grant ARC39/13 (MOE2013-T2-1-079) and AcRF Tier 1 seed fund on Complexity RGC2/13 (M4011101), Ministry of Education, Singapore for Dr. Jie Zheng. During the 10-year effort, Qianqing Ding, Hongliang Hu, Yunbo Xu et al. gave the work great contributions in culture TIL cell and establishment of local galaxy analysis system. Nancy S. Debry and Shen Li contribute the chapter modification. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation.

Competing interests statement: The authors declare competing financial interests.

References

1. Watson JD, Crick FH. The structure of DNA. Cold Spring Harb Symp Quant Biol. 1953;18:123–31.
2. Er TK, Chang JG. High-resolution melting: applications in genetic disorders. Clin Chim Acta. 2012;414:197–201. doi:10.1016/j.cca.2012.09.012.
3. Sanger F, Nicklen S, et al. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463–7.
4. Maxam AM. Gilbert W (1992) A new method for sequencing DNA. Biotechnology. 1977;24:99–103.
5. Pareek CS, Smoczynski R, et al. Sequencing technologies and genome sequencing. J Appl Genet. 2011;52(4):413–35. doi:10.1007/s13353-011-0057.
6. Margulies M, Egholm M, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80.
7. Warren RL, Sutton GG, et al. Assembling millions of short DNA sequences using SSAKE. Bioinformatics. 2006;23(4):500–1.
8. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24(3):133–41.
9. Liu L, Li Y, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;25:1364–75. doi:10.1155/2012/251364.
10. Breivik J. The evolutionary origin of genetic instability in cancer development. Semin Cancer Biol. 2005;15(1):51–60.

11. Liberko M, Kolostova K, et al. Essentials of circulating tumor cells for clinical research and practice. Rev Oncol Hematol. 2013;88(2):338–56. doi:10.1016/j.critrevonc.2013.05.002.
12. Ning L, Liu G, et al. Current challenges in the bioinformatics of single cell genomics. Front Oncol. 2014;4:7. doi:10.3389/fonc.2014.00007.
13. Ebenezer V, Medlin LK, et al. Molecular detection, quantification, and diversity evaluation of microalgae. Biotechnology (NY). 2012;14(2):129–42. doi:10.1007/s10126-011-9427-y.
14. Brück S, Evers H, et al. Single cells for forensic DNA analysis–from evidence material to test tube. J Forensic Sci. 2010;56(1):176–80. doi:10.1111/j.1556-4029.2010.01553.
15. Meier-Ruge W, Bielser W, et al. The laser in the Lowry technique for microdissection of freeze-dried tissue slices. Histochem J. 1976;8(4):387–401.
16. Schutze K, LAHR G, et al. Identification of expressed genes by laser-mediated manipulation of single cells. Nat Biotechnol. 1998;16(8):737–42.
17. Li B. A strategy to identify genomic expression at single-cell level or a small number of cells. J Biotechnol. 2005;8(1):71–81. doi:10.2225/vol8-issue1-fulltext-3.
18. Emmert-Buck MR, Bonner RF, et al. Laser capture microdissection. Science. 1996;274(5289): 998–1001.
19. Fend F, Emmert-Buck MR, et al. Immuno-LCM: laser capture microdissection of immunostained frozen sections for mRNA analysis. Am J Pathol. 1999;154(6):1857–66.
20. Fiegler H, Geigl JB, et al. High resolution array-CGH analysis of single cells. Nucleic Acids Res. 2007;35(3):e15.
21. Kurihara T, Kamberov E, et al. Rubicon PicoPlex-NGS Kits available for sequencing single cells using the Illumina Genome Analyzer. J Biomol Tech. 2011;22(Suppl):S51.
22. Pan X, Urban AE, et al. A procedure for highly specific, sensitive, and unbiased whole-genome amplification. Proc Natl Acad Sci U S A. 2008;105(40):15499–504. doi:10.1073/pnas.0808028105.
23. Ambion Catalog. Macromolecular components of *E. coli* and Hela cells. 2004. p. 192
24. Klebe RJ, Rodriguez SA, et al. RT-PCR without RNA isolation. Biotechniques. 1996;21(6): 1094–100.
25. Ramsköld D, Luo S, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30(8):777–82.
26. Lobo MK. Molecular profiling of striatonigral and striatopallidal medium spiny neurons past, present, and future. Int Rev Neurobiol. 2009;89:1–35. doi:10.1016/S0074-7742(09)89001-6.
27. Hashimshony T, Wagner F, et al. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2012;2(3):666–73. doi:10.1016/j.celrep.2012.08.003.
28. Tang F, Barbacioru C, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6(5):377–82. doi:10.1038/NMETH.1315.
29. Eberwine J. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. Biotechniques. 1996;20(4):584–91.
30. Toellner KM, Gulbranson-Judge A, et al. Immunoglobulin switch transcript production in vivo related to the site and time of antigen-specific B cell activation. J Exp Med. 1996;183(5): 2303–12.
31. Dixon AK, Richardson PJ. Expression profiling of single cells using 3 primer end amplification (TPEA) PCR. Nucleic Acids Res. 1998;26(19):4426–31.
32. Gole J, Gore A. Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. Nat Biotechnol. 2013;31(12):1126–32. doi:10.1038/nbt.2720.
33. Landau YE, Lichter-Konecki U, et al. Genomics in newborn screening. J Pediatr. 2014;164(1):14–9. doi:10.1016/j.jpeds.2013.07.028.
34. Panagopoulos I, Thorsen J, et al. Sequential combination of karyotyping and RNA-sequencing in the search for cancer-specific fusion genes. J Biochem Cell Biol. 2014;S1357–2725(14):00176–9. doi:10.1016/j.biocel.2014.05.018.
35. Navin N, Kendall J, et al. Tumor evolution inferred by single-cell sequencing. Nature. 2011;472(7341):90–4. doi:10.1038/nature09807.
36. Hou Y, Song L, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell. 2012;148(5):873–85. doi:10.1016/j.cell.2012.02.028.

37. Li Y, Xu X, et al. Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. Gigascience. 2012;1(1):12. doi:10.1186/2047-217X-1-12.
38. Kim KI, Simon R. Using single cell sequencing data to model the evolutionary history of a tumor. BMC Bioinformatics. 2014;15:27. doi:10.1186/1471-2105-15-27.
39. Xu YB, Hu HL, et al. Feasibility of whole RNA sequencing from single-cell mRNA amplification. Genet Res Int. 2013;2013:724124. doi:10.1155/2013/724124.
40. Macaulay IC, Voet T. Single cell genomics: advances and future perspectives. PLoS Genet. 2014;10(1):e1004126. doi:10.1371/journal.pgen.1004126.
41. Meldrum C, Doyle MA, et al. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011;32(4):177–95.
42. Guan YF, Li GR, et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. Chin J Cancer. 2012;31(10):463–70. doi:10.5732/cjc.012.10216.
43. Li B, Senzer N, et al. Approach to individual cancer target identification. In: 11th annual meeting of the American Society of Gene Therapy. 2008;8(45):1004
44. Zhang W, Ding JQ, et al. Genomic expression analysis by single-cell mRNA differential display of quiescent CD8 T cells from tumour-infiltrating lymphocytes obtained from in vivo liver tumors. Immunology. 2009;127(1):83–90. doi:10.1111/j.1365-2567.2008.02926.
45. Jiao W, Vembus S, et al. Inferring clonal evolution of tumors from single nucleotide somatic mutations. BMC Bioinformatics. 2014;15:35. doi:10.1186/1471-2105-15-35.