

# Applications of Single Cell Sequencing in Cancer

Kuo Ping Chiu

**Abstract** Single cell sequencing (SCS) was chosen as Method of the Year in 2013 by *Nature Methods*. It refers to the deep sequencing of DNA or RNA molecules of single cells for the study of genome, transcriptome, noncoding RNA, and protein sequences encompassed within the single cells of interest. SCS possesses a strong potential to exert a great impact on biological research due to its unprecedented opportunity and capability of resolving sequence composition and genetic variations on single cell basis, which can be further integrated to explain the biological phenomena at the population or organismal level. During the past few years, SCS has produced promising results in a number of fields, especially cancer research. Prior to SCS, tissue sequencing could only analyze a tumor mass or a cancer cell population as a whole. With methods for single cell isolation, further powered by molecular level resolution made possible by minute DNA amplification and nucleotide level resolution made possible by next-generation sequencing, SCS technologies are becoming robust approaches for the study of cancer genomics and transcriptomics, and the cancer–normal cell interactions in the microenvironment. SCS analyses of cancer evolution, for example, have shown the capability of being able to empirically infer the driver mutations and map the sequential mutation events during cancer development. Undoubtedly, these novel approaches will produce profound health benefits. SCS will continue to attract more attention, and further expansion of its applications in many biological fields is inevitable. This chapter focuses on SCS methodologies, its existing applications, especially its capability in reconstructing the evolutionary history of cancer progression and in profiling cancer transcriptome, and the potential applications expected to come.

**Keywords** Single cell sequencing (SCS) · Single cell transcriptome sequencing (SCT)

---

K.P. Chiu (✉)  
Academia Sinica, Taipei, Taiwan  
e-mail: chiukp@gate.sinica.edu.tw

## 1 Introduction

A normal diploid human genome is composed of ~6 billion bp of nuclear DNA distributed in 23 pairs of linear chromosomes, forming DNA-protein complexes, or chromatin, during the interphase of cell cycle. Besides, there are thousands of copies of mitochondrial genomes (mitogenomes) per cell, each composed of 16,569 bp of deoxynucleotides forming a circular DNA structure. Together, the nuclear genome and the mitochondrial genome form the basis of the information reservoir. Through differential expression of ~26,000 protein-coding genes [1] and hundreds, or more, noncoding RNAs, they drive all the biological activities of human life. As the doctrine of the Central Dogma teaches [2], gene expression is mediated by transcription of genomic sequences to produce RNA molecules (mRNA, tRNA, rRNA, and other noncoding RNA) and translation of the protein-coding mRNA molecules.

All matter in the universe is subject to dynamic environmental impacts. There is no exception for DNA and RNA molecules. Environmental impacts, including toxic chemical compounds, radiations, magnetism, reactive oxygen species (ROS), and many others, may cause DNA/RNA damages. Aberrations in DNA/RNA sequences are then subject to selection by evolution which constantly occurs in the context of the environment. Some aberrations, either as small as single base alterations or as large as deletions or insertions of chromosomal fragments and chromosomal breakage and rearrangements, are likely to result in cell death. Some aberrations, on the other hand, may lead to tumorigenesis [3].

One of the most remarkable features of cancer is genome instability, which acts as a driving force for genetic mutation and plays a key role in wide spectrum reshuffling of the genomic material [4, 5], eventually causing a tumor mass to diversify and form a mixture of heterogeneous, aneuploid genomes. During cancer progression, genomic alterations are recorded as “mutational fingerprints,” which, in theory, can be traced to reveal the cancer evolutionary history. Through gene expression, the genomic alterations are carried onto transcriptome and then proteomes, showing abnormal gene expression profile and altered proteomic state, respectively. Besides cancerous cells, tumors are also composed of infiltrating normal cells (such as fibroblasts) and immune cells (such as macrophages and lymphocytes) [6, 7]. Multiple lines of evidence indicate that the complexity of cancer cannot be fully appreciated by conventional approaches which study cancer cell population as a whole. With the advances of sequencing technologies, aberrations in nucleic acids can now be revealed and analyzed by next-generation sequencing (NGS) at a large scale [8, 9], and single cell sequencing (SCS) technologies are becoming the right choice. Then, may cancer genomes in a tumor be subdivided into subpopulations or subclones? Is SCS able to reveal the developmental history of a tumor? Moreover, since solid tumors also contain a number of noncancerous cells and the gene expression of these “normal” cells is known to be influenced by cancer cells, can SCS provide further insight into how these cells interact to foster cancer progression? From the technical point of view,

can SCS compete against other ensemble methods for early detection of cancer? How far can SCS go to straighten out these issues and help answer the questions?

To maintain cell integrity and help define the boundary of cellular activities, each single eukaryotic cell has a well-defined cell membrane, which can be directly observed under an electron microscope. Imaging-based analysis of gene expression at the single cell level by *in situ* hybridization followed by imaging had long preceded PCR and NGS approaches [10, 11]. Automation of Sanger sequencing together with the invention of PCR by Kary Mullis in 1980s stimulated the invention of NGS platforms during 2005–2007, which have created a sustainable state of sequencing-based biological research [12], and are now, by studying single cells [13], revolutionizing cancer research [14].

Prior to the invention of NGS technologies, tumor evolutionary histories could be studied by dissecting tumors into subregions based on phenotypical characteristics or topographical locations, followed by sequencing library preparation using tissues from subregions, sequencing, and sequence data analysis [15]. With single cell sequencing technologies, the process can be significantly simplified. Recent studies have shown that the mutational fingerprints in single cells of the same tumor, when revealed by SCS, can be interconnected to form lineages of clonal expansion and the history of cancer evolution [14, 16]. Furthermore, the temporal order of the mutational fingerprints may also indicate the driver mutations and their role in cancer progression. Recent researches have also shown that the expression of cancer-associated genes (e.g., oncogenes, tumor suppressor genes, oncogenic miRNA, hormone receptor genes, etc.) can now be better understood with single cell resolution [13, 17]. How they coordinate and interplay with one another will provide us in-depth understanding of intercellular molecular activities in a cell population.

## 2 Single Cell Sequencing and Its Challenges

One cannot fully appreciate the beauty of a technology without knowing the challenges it has overcome. Indeed, it is much more difficult to conduct single cell sequencing than tissue sequencing [18], mainly because of the limitation in the quantity of genetic materials a single cell can provide. Before single molecule sequencing becomes sophisticated and available for minute DNA sequencing, SCS technologies have to rely on NGS technologies for sequencing the genetic materials in single cells. Conventional NGS technologies require bulk genetic materials to sequence. However, each single cell, even for cancer cells, contains only a minute quantity (normally at picogram level), with at least three orders of magnitude less than that required for conventional NGS protocols. This gap can only be filled by DNA and/or RNA amplification using PCR-based and/or RNA polymerase-based strategies, especially the former. PCR amplification is crucial for SCS. PCR, however, may result in bias due to imbalanced amplification efficiency between amplicons and thus needs to be carefully designed. In fact, this was one of the major

barriers that hindered the progress of SCS during the past. Furthermore, because most molecules in a single cell are of low abundance and low copy number, SCS is extremely vulnerable to fluctuations which may result from genetic material loss, poor personal skill, or contamination. Thus, the experimental procedure of a SCS protocol, especially from single cell or nucleus isolation up to DNA/RNA amplification, has to be well formulated and every step needs to be carefully characterized and optimized when dealing with trace amount of DNA or RNA. Moreover, sequencing error is also an issue that needs to be taken care of. While heterozygous alleles have equal chance of being sequenced, sequencing errors are shown at relatively much lower frequency and can be corrected by sufficient fold-of-coverage, or, as demonstrated by Kim and Simon, by incorporating the probability of sequencing error into Bayesian probability test (see below) [19].

## ***2.1 Cells Suitable for Single Cell Sequencing***

SCS is suitable for the study of a number of normal and diseased cell types of both prokaryotes and eukaryotes [20, 21]. These include early-stage embryonic cells, stem cells, immune cells, rare cells, microbes that cannot be easily cultured, differentiated cells, bacteria or virus infected cells, and cancer cells at various stages. The application of SCS in the study of cancer and normal cells in the tumor-surrounding microenvironment is of particular interest because of their complexity and tight association with human disease. Cultured cancer cells are probably the most accessible cancer cells for researchers, while circulating tumor cells (CTCs) and intratumor cancer cells require reliable experimental procedure to isolate. SCS analyses will benefit the diagnosis and help in guiding chemotherapy and monitoring/following up the progress of treatment.

## ***2.2 Methods for Cancer Cell Isolation***

Depending on the conditions of the single cells, a number of methods are available for single cell isolation. The most commonly used methods are micromanipulation, flow cytometry, microdissection, single cell labeling, and cell trapping [21]. Recently, an automated system (C1 System by Fluidigm<sup>TM</sup>) for single cell genome and transcriptome analyses has been made commercially available. It is relatively easy to isolate single cells from a cell culture or from the blood, because most of these cells are either already separated or can be separated easily. In this situation, single cells can be washed, diluted, and isolated by manipulation such as mouth pipetting under a microscope. Compared to flow-sorting, micromanipulation is more tedious and may not be suitable for collecting a large number of single cells. However, micromanipulation is probably the mildest and gentlest approach that can minimize the impact of harsh conditions such as high pressure produced by flow

cytometer and thus can preserve the cell integrity. As such, this method is useful for transcriptome analysis. On the other hand, flow cytometry is probably the most efficient approach [21]. As mentioned above, since high flow pressure can easily damage the cell membrane and cause leakage of the cytoplasm and cross-contamination, precaution has to be taken before using flow cytometer for whole cell isolation. A short period of cell wash and/or cell culture may be needed. Single cell trapping which uses a matrix coated with specific molecule(s) is also for isolating specific cells. It is important to pick single cells that are representative, healthy, with well-maintained cellular integrity, and free of contamination. The automated C1 System facilitates single cell genome and transcriptome analyses by increasing throughput, reducing technical variations, and easing control and comparison [22, 23]. However, its potential problem with primer dimers may result in false positive signals [22]. Moreover, its sensitivity may not be high enough to detect low abundance transcripts. How to improve these drawbacks seems to be critical issues for future improvement for the automated single cell instruments.

### ***2.3 Methods for Single Cell Whole Genome Amplification***

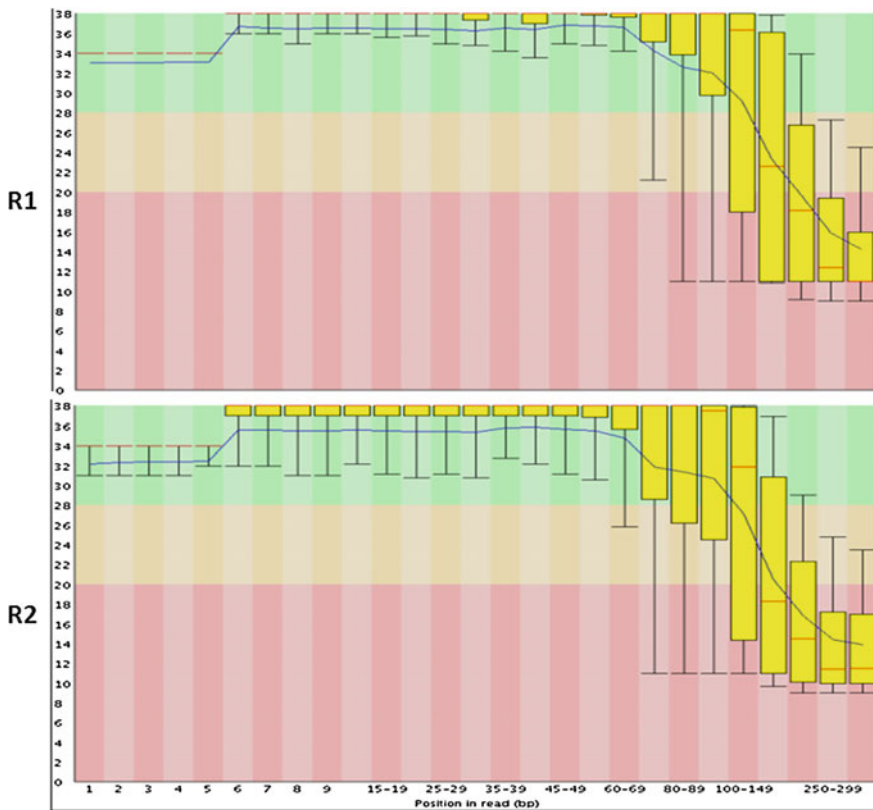
Whole genome amplification (WGA) is the key for single cell sequencing, no matter whether it is for exome sequencing or whole genome sequencing of the single cells. To preserve the original state of the genome by minimizing uneven amplification is essential for whole genome amplification. This is frequently done by reducing the number of PCR cycles, preventing PCR by-products, or using barcodes. Barcodes can not only minimize bias resulted from the differences in personal skills and experimental procedure, but also allow multiple small samples to be combined into a larger sample. A number of methods for DNA amplification, together with their limitations, have been reviewed previously [24]. Here, we quickly skim through the methods employed by Navin and Hou.

In 2011, Navin and colleagues adopted degenerated oligos to prime single cell whole genome amplification (DOP-WGA) [14]. Although efficient enough for the authors to generate reliable datasets for copy number variation (CNV) analysis, this approach provided only a low coverage ( $\sim 6\times$ ), presumably due to the limitation in the size range of DOP-WGA products. In 2012, Hou and colleagues published the Multiple Displacement Amplification (MDA) method for single cell whole genome amplification. In the method, they used the  $\Phi 29$  (Phi29) enzyme to amplify DNA in linear fashion [16]. The products were subsequently subjected to a fluorometry-based quantitation procedure which selected quality sequences for further analysis (see below). This seems to be an efficient approach for single cell genomic amplification for NGS. However, as drawbacks of the approach, allelic dropout and imbalanced amplification have been reported [24]. Imbalanced amplification is a common phenomenon for multiplex PCR. Previous studies have indicated that %GC is responsible for these drawbacks. GC pair is stronger than AT pair because, while AT base pairing is mediated by two hydrogen bonds, GC pairing is mediated by three

hydrogen bonds. Besides, high GC content favors Z-form conformation [25]. These properties influence the efficiency of DNA amplification, causing imbalanced PCR amplifications (especially when a significant number of primer pairs are deployed across the entire genome [26]), with allelic dropouts being the most severe cases. Kim and Simon proposed a computational approach to correct potential sequence errors which may be introduced by multiple displacement amplification to cause false discovery or allelic dropout, or by sequencing—see Fig. 3.

## 2.4 NGS DNA Sequencing

Besides genome amplification, next-generation sequencing (NGS), which is currently the only method able to provide sufficient coverage (sequencing depth) for



**Fig. 1** Comparison between R1 and R2 QV profiles. The *M. cyclopi*s mitochondrial DNA library was sequenced by MiSeq with  $2 \times 300$  bp paired-end sequencing. The R1 and R2 QV profiles were then displayed in parallel to indicate the faster decrease in R2 quality

reliable analysis, is also essential for single cell sequencing. After years of competition since 2005, the Illumina platform has prevailed over the others (e.g., 454 and SOLiD systems) [9]. In terms of single cell sequencing, long reads would be better for at least most cases. However, R2 quality drops faster than R1, making 250 bp an evident barrier for the current Illumina sequencing technology (Fig. 1). The faster quality drop for R2 may result from cluster regeneration. Right before R2 sequencing, R1 clusters (i.e., clusters used as the templates for R1 sequencing) are replaced by their complementary clusters (i.e., R2 clusters made from the complementary strands of the R1 clusters). During the process, some templates in each cluster may get damaged or degraded, making the cluster less sufficient and thus more vulnerable to sequencing reactions. Limited read length is a natural phenomenon. By nature, reading of some templates in a cluster may go wrong in any step of the sequencing reaction. The mistakes accumulate over time, causing the quality value to drop gradually.

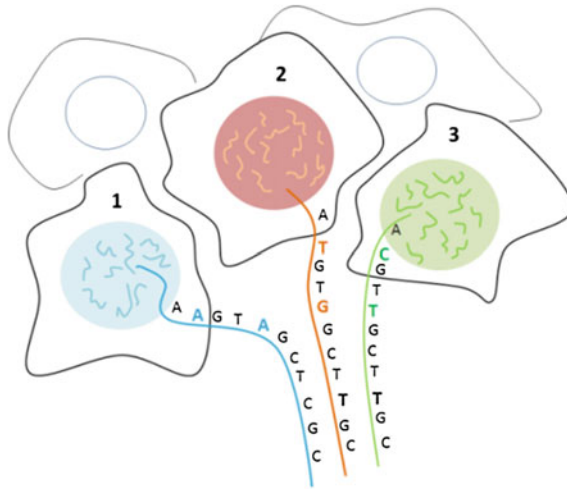
### 3 Existing Applications

#### 3.1 *Genome Sequencing of Individual Cancer Cells*

Depending on the original cell type and the developmental stage, cancer cells may exist as various forms. During cancer progression, cancer cells further diversify in genomic makeup and function. The heterogeneity of cancer genome presents a challenge for cancer research. However, at the same time it also provides an opportunity for the study of intratumor substructure, cancer progression, and cancer evolution. This is made possible by single cell sequencing, a breakthrough in next-generation sequencing, of the genomes in the same tumor mass.

The report by Navin and colleagues in 2011 using single nucleus sequencing (SNS) to study the evolutionary history of human breast cancer marked a breakthrough in cancer research [14]. The experimental procedure can be outlined to include three major steps: isolation of single cancer nuclei by flow-sorting, whole genome amplification by random priming with degenerated oligonucleotides, and next-generation sequencing. The sequence reads were then analyzed to resolve genomic differences in copy number among individual cancer cells (Fig. 2).

The experimental approach was first validated by using single nuclei isolated from SK-BR-3 cell line, together with a million-cell population control from the same cell line. After genomic amplification using degenerated oligos in random priming, they obtained only a low coverage ( $\sim 6\times$ ) of the single cell genomes. However, such level of coverage is sufficient for CNV analysis. In terms of bioinformatics, the authors designed unique analytical approaches. For example, instead of using fixed intervals to calculate integer copy number, they used variable length bins but with uniform expected unique counts, which would correct for biases that have been reported in WGA. Pileups (over-replicated loci) were found to



**Fig. 2** Single cell genome sequencing reveals the differences in genomic sequences between single cells. Individual cells, or nuclei, are first collected in separate tubes. Genomic sequences (chromosomes) are amplified and sequenced. Sequence reads are then mapped against human genome assembly to identify their locations and sequence variations (e.g., copy number variations, insertions, deletions, single base alterations, etc.). Cross-comparison on sequence variation of single cell libraries, if isolated from the same cancer, allows us to reconstitute the evolutionary history of that cancer

be randomly distributed and sparse so that would not affect the results. In both single cells and million-cell population control, they found major amplifications in genes encoding MET, TPD52, ERBB2, and BCAS1 proteins. Deletion in DCC (deletion in colorectal cancer) gene was also detected in both single and 1-million-cell population of SK-BR-3 cells. These results generated from the SK-BR-3 cell line allowed them to move forward to study single nuclei isolated from different sections of breast tumors.

They divided a high-grade (grade III) triple negative (ER<sup>-</sup>, PR<sup>-</sup>, and Her2<sup>-</sup>) carcinoma (labeled as T10) into six sections and analyzed 100 single cell nuclei isolated from these sections. From the study, they identified distinct clonal subpopulations in the genetically heterogeneous ductal carcinoma. The integer copy number profiles were built and analyzed to contain 63 % of normal cells and 37 % tumor cells and infiltrated with leukocytes. By calculating pair-wise distances between the 100 profiles followed by building a phylogenetic tree using neighbor joining, four subpopulations were identified, one with flat diploid profile and three with complex (advanced) genomic structures, suggesting three clonal expansions. Moreover, their method was able to detect diverse chromosome gains and losses and discern ‘pseudodiploid’ nuclei in diploid nuclei. Further clonal analysis allowed the authors to trace the evolutionary history of the cancer from the primary stage to the metastatic stage. The data further suggested that, differing from gradual models



of tumor progression, tumors grow by “punctuated” clonal expansion, without discernible intermediate branching.

The application of single nucleus sequencing in reconstitution of cancer evolutionary history, which suggested cancer progression mediated by punctuated clonal expansion, was reviewed [21, 27]. The drawbacks of such copy number-based approach include low coverage and being unable to reach down to the nucleotide resolution.

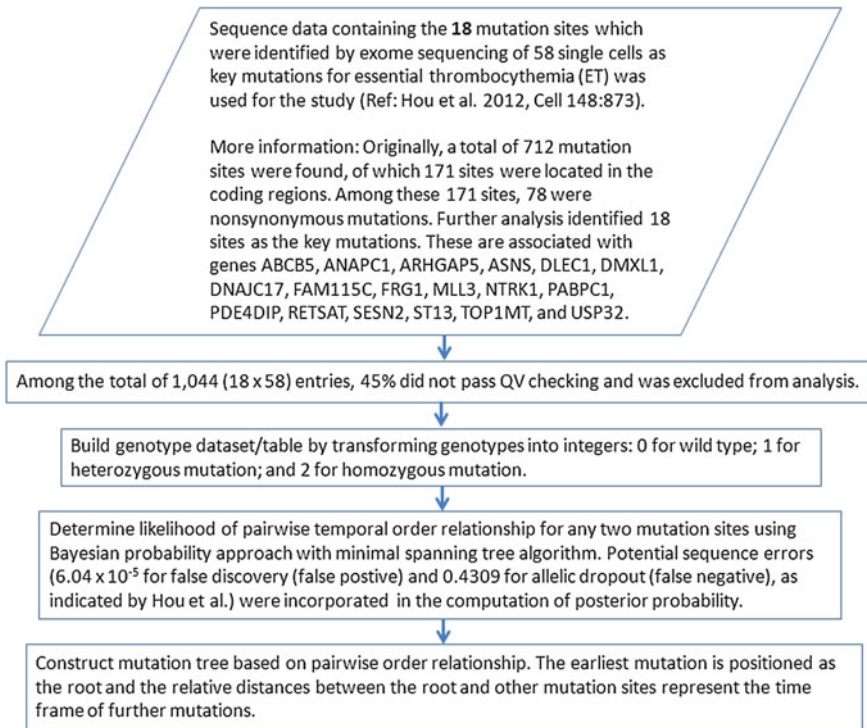
Following the report by Navin et al., Hou and colleagues reported an MDA-based exome sequencing of single cancer cells [16]. Deviating from copy number-based SCS analysis, this report presented a pilot study at the single cell nucleotide level. The single cell nucleotide sequencing is made possible by using multiple displacement amplification of the whole genome. MDA products are mostly of high molecular weight (>10 Kb), being able to boost genome coverage.

The MDA method was first tested with two single cells under multicell control and hg18 was used as the human genome reference. The coverage was found to be  $\geq 15\times$  (mean fold coverage =  $18\times$ ), and sequences of both single cells covered more than 90 % of the reference genome, while more than 95 % of the bases in hg18 were recovered with  $\geq 15\times$  sequencing depth. WGA failure was found to be associated with GC content, with failed regions containing higher GC% than the average 41 % GC content in human genome. The ratio of allele dropout (ADO), which indicated whether non-amplification occurred in one of the alleles present in a heterozygous sample and would lead to false negative, was maintained at  $\sim 11$  %. ADO showed no bias relative to genomic location, and errors of MDA also showed no preferences on genes or functions.

After testing, the authors applied MDA procedure to study the genes involved in essential thrombocythemia (ET) evolution. For certain reasons, they conducted exome sequencing instead of whole genome sequencing. A total of 90 single cells from an ET patient were sequenced to a mean depth of  $30\times$ . After filtering out the single cells with <70 % coverage, 58 single cells were chosen for further analysis. These single cells have an average of  $\sim 70$  % of target bases at  $\geq 5$  depth. They considered this coverage as sufficient for population variant calling when multiple single cells have the same variant. Exome sequencing generated SMAFS (somatic mutant allele frequency spectrum) for evolutionary study. Results indicate that ET patient carries a distinct set of mutations and a monoclonal origin of ET cancer cells.

In a parallel study, Xu et al. used single cell exome sequencing to investigate clear cell renal cell carcinoma (ccRCC) by which they revealed kidney tumor-specific single nucleotide mutations [28]. Unlike reports by Navin and colleagues, no significant clonal subpopulations were identified in their ccRCC cases, presumably due to the difference in cancer type and origin. This single cell exome sequencing also revealed single nucleotide mutation characteristic of the kidney tumor.

Besides the above-mentioned reports, a computational approach for inferring the evolutionary mutation history of a cancer using single cell sequencing data has also been reported by Kim and Simon (Fig. 3) [19]. Although the quality of a library is

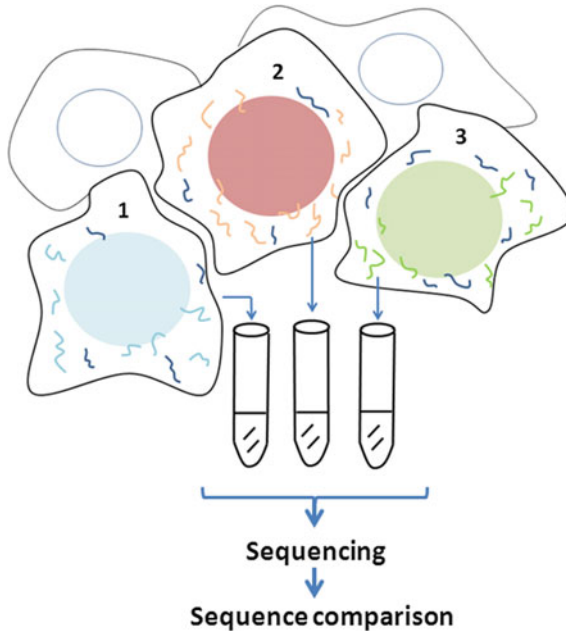


**Fig. 3** Workflow for the construction of evolutionary mutation tree presented by Kim and Simon

influenced by a number of factors such as the make or design of a sequencer, personal skill, and the quality or the preparation of the material, the forefront quality control would remove the questionable reads, keeping the sequencing errors in the qualified reads at low rate, while on the other hand, heterozygous alleles remain having equal chance to be detected in the sequence reads, making these two types of sequence variations readily distinguishable from each other. The sequencing errors can be more easily detected by sufficient coverage (normally set at 30-fold or above) and then removed by programs, or by incorporating the probability of sequencing error into computation, as demonstrated by the authors.

### 3.2 Transcriptome Sequencing

Sometimes transcriptional information of individual cells, instead of a cell population, is desired. Obtaining such information relies on single cell transcriptome (SCT) sequencing and analysis (<http://genomebiology.com/2010/11/S1/P8>), which remains a great challenge for current technologies [17].



**Fig. 4** Single cell transcriptome sequencing. The profiles of gene expression in single cells can now be studied by single cell transcriptome sequencing. Basically, the procedure is similar to that used for transcriptome analysis of a cell population, except that each cell has to be collected separately and then goes through cell lysis, cDNA synthesis, cDNA amplification, sequencing, sequence data analysis, and cross-library comparison. To prevent material loss, all reactions are conducted at very low volume (e.g., a few microliters or less) and wash is also minimized until the cDNA molecules have been amplified

Indeed, several lines of evidence show that stochastic gene expression is likely to be a natural phenomenon and thus the gene expression of a tumor or a tissue should be considered as a combinatorial phenomenon summarized from its constituent single cell transcriptomes. SCT sequencing is becoming the most advanced approach for studying gene expression and regulation (Fig. 4), and this line of application heavily relies on cDNA synthesis. In fact, there were methods for single cell cDNA synthesis published before NGS became a popular technology for sequencing. These include the first single cell transcriptome analysis reported by Eberwine and colleagues in 1992 [12]. Later in 2006, Kurimoto et al. published another method for microarray-mediated SCT analysis [29]. The first NGS-based single cell transcriptome sequencing was published in 2009 [13]. In 2012, Ramskold and colleagues published an elegant method for cDNA synthesis and amplification [30]. This approach has been commercialized by Clontech to make a kit called “the SMARTer Ultra Low RNA Kit.” Instead of using oligo-dT as employed by Tang et al., Ramskold and colleagues used CDS primer, which carries a VN tail (V stands for ‘non-T’ and N stands for ‘any base’) in the 3’ end of the

oligo-dT sequence, to prime the first strand cDNA synthesis. The VN tail significantly enhances the specificity of priming because the VN tail allows the prime to “hook” to the last two bases right in front of the polyA tail in the mRNA molecule. Without the VN tail, the primer would ‘slip’ within the polyA region, resulting in a significant amount of imprecise priming. The above-mentioned automated single cell analysis system is an efficient approach, but solely for the study of a limited number of genes or genomic regions. Similarly, primer dimer is an issue of concern [22].

Single cells are delicate entities, and thus concerns about the accuracy of SCT analysis are inevitable. Various potential factors that may cause transcriptional variations have been, and will continue to be, examined. Indeed, it can be difficult to identify the factors causing variations in single cell transcriptomes. Some variations between individual cells may be real, but some may result from differences in personal technical skills and thus need to be minimized. To minimize the influence of variability in personal technical skill, it is recommended to increase the number of SCTs and use internal controls such as housekeeping genes and previously studied expression patterns of certain genes. For example, in our study of single cell transcriptomes of MCF-7 breast cancer [31], we used the expression the *LDHB* (lactate dehydrogenase) gene, which is known to be completely shutdown in MCF-7 cells, as an internal control. As expected, its expression was not detected in all libraries (data not shown).

## 4 SCT Protocols Can Be Modified for Various Reasons

Protocols may need to be modified for certain reasons [9]. For example, a protocol might not have been optimized when it is published. This frequently occurs when it is published in a hurry by companies trying to catch up market demand. Besides, taking advantage of its low input requirement, one can adopt single cell sequencing protocol to generate sufficient amount of input material for a regular sequencer. As shown in the following section, we can use total RNA or mRNA, instead of single cells, as the input material for a single cell protocol. By so doing, we bypass the conventional protocol and use the single cell protocol to rescue the situation when material is not sufficient for a regular sequencer.

## 5 Using Different Types of Materials as the Input

Since polyA<sup>+</sup>-RNA molecules in single cells are the only molecules required for double-stranded cDNA synthesis, it is reasonable to use either total RNA or mRNA to replace single cells.

The majority of RNA molecules are ribosomal RNA (rRNA) and transfer RNA (tRNA), while mRNA species constitute only  $\sim 2\%$  of the total RNA. The presence of rRNA and tRNA may reduce the efficiency of cDNA synthesis because of a number of reactions, including oligo-dT priming, reverse transcription, and PCR amplification. As such, it is strongly recommended to use mRNA as the starting material, if the amount of total RNA is sufficient for mRNA isolation. In fact, it has been empirically demonstrated that mRNA works better than total RNA.

Then, how do we correlate the results produced from mRNA or total RNA with cell number? Using MCF-7 as an example, each MCF-7 cell expresses about 10 pg of total RNA. Accordingly, 50 ng of total RNA is equivalent to about 5000 cells, and 50 ng of mRNA is equivalent to 250,000 (=0.25 million) single cells. One can calculate and use a certain amount of total RNA or mRNA based on the number of cells he/she wants to use in the study. Since the total amount of RNA expressed from a single cell varies across different cell types, it is strongly recommended to empirically fine-tune this value based on the cell type being used.

Completion of the second PCR amplification marks the junction where the SCT protocol and other protocols meet. Now the retrieval of mRNA molecular information from a single cell is completed and preparation of a sequencing library can be initiated. Here, one can determine what sequencing libraries to make: shotgun fragment sequencing, paired-end (PE) sequencing, or pair-end ditag (PED) sequencing.

## 6 Future Potential Applications of Single Cell Sequencing

We can expect many more SCS applications to be developed for cancer research in the near future. Using transcription factor binding site (TFBS) analysis as an example, there is a strong potential for us to conduct TFBS analysis at the single cell level. Currently, most cancer researches have been focused on the study of mutations in cancer-associated genes such as *KRAS*, *TP53*, *cMYC*, etc. Less attention was paid to in vivo study of how alterations in DNA motifs interact with transcription factors (TFs) and/or other intracellular proteins, and how a mutation in TF affects its DNA binding. Will it form complexes with other unexpected proteins? Or, will it bind to different locations in the genome? It would be interesting to further understand how an altered motif influence TF binding at the single cell level and how the effects at the single cell level exert a combinatorial effect at the population level. In theory, one would expect an alteration in DNA motif to result in a corresponding switch in the interacting protein(s), which may in turn play a role in tumorigenesis, angiogenesis, and/or metastasis. Empirical SCS data will help to either prove or disapprove the speculation.

## 7 Further Improvements

To enhance SCS data analysis and protocol design, it can be helpful to produce a virtual population profile through the integration of SCS profiles by statistical approach. By comparing the virtual population profile with the empirically produced population profile, we can evaluate and improve the SCS procedure. Conceivably, protocols for single cell experiments have to be reproducible, straightforward, and adaptable. However, many protocols do not yet meet these criteria and need to be optimized.

The progress of biological research heavily relies on the advance of biotechnologies. There is no doubt that more innovative SCS approaches will be created in the near future.

## References

1. Venter JC et al (2001) The sequence of the human genome. *Science* 291:1304–1351. doi:[10.1126/science.1058040](https://doi.org/10.1126/science.1058040)
2. Crick F (1970) Central dogma of molecular biology. *Nature* 227:561–563
3. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646–674. doi:[10.1016/j.cell.2011.02.013](https://doi.org/10.1016/j.cell.2011.02.013)
4. Negrini S, Gorgoulis VG, Halazonetis TD (2010) Genomic instability—an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 11:220–228. doi:[10.1038/nrm2858](https://doi.org/10.1038/nrm2858)
5. Aguilera A, Gomez-Gonzalez B (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet* 9:204–217. doi:[10.1038/nrg2268](https://doi.org/10.1038/nrg2268)
6. Gerdes MJ et al (2014) Emerging understanding of multiscale tumor heterogeneity. *Front Oncol* 4:366. doi:[10.3389/fonc.2014.00366](https://doi.org/10.3389/fonc.2014.00366)
7. Wels J, Kaplan RN, Rafii S, Lyden D (2008) Migratory neighbors and distant invaders: tumor-associated niche cells. *Genes Dev* 22:559–574. doi:[10.1101/gad.1636908](https://doi.org/10.1101/gad.1636908)
8. Metzker ML (2010) Sequencing technologies—the next generation. *Nat Rev Genet* 11:31–46. doi:[10.1038/nrg2626](https://doi.org/10.1038/nrg2626)
9. Chiu KP (2015) Next-generation sequencing and sequence data analysis, 1st edn. Bentham Science Publishers
10. Gall JG, Pardue ML (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A* 63:378–383
11. Capodiceci P et al (2005) Gene expression profiling in single cells within tissue. *Nat Methods* 2:663–665. doi:[10.1038/nmeth786](https://doi.org/10.1038/nmeth786)
12. Eberwine J et al (1992) Analysis of gene expression in single live neurons. *Proc Natl Acad Sci U S A* 89:3010–3014
13. Tang F et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6:377–382. doi:[10.1038/nmeth.1315](https://doi.org/10.1038/nmeth.1315)
14. Navin N et al (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472:90–94. doi:[10.1038/nature09807](https://doi.org/10.1038/nature09807)
15. Tsao JL et al (2000) Genetic reconstruction of individual colorectal tumor histories. *Proc Natl Acad Sci U S A* 97:1236–1241
16. Hou Y et al (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 148:873–885. doi:[10.1016/j.cell.2012.02.028](https://doi.org/10.1016/j.cell.2012.02.028)

17. Tang F, Lao K, Surani MA (2011) Development and applications of single-cell transcriptome analysis. *Nat Methods* 8:S6–11. doi:[10.1038/nmeth.1557](https://doi.org/10.1038/nmeth.1557)
18. Eberwine J, Sul JY, Bartfai T, Kim J (2014) The promise of single-cell sequencing. *Nat Methods* 11:25–27
19. Kim KI, Simon R (2014) Using single cell sequencing data to model the evolutionary history of a tumor. *BMC Bioinformatics* 15:27. doi:[10.1186/1471-2105-15-27](https://doi.org/10.1186/1471-2105-15-27)
20. Shapiro E, Biezuner T, Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 14:618–630. doi:[10.1038/nrg3542](https://doi.org/10.1038/nrg3542)
21. Navin N, Hicks J (2011) Future medical applications of single-cell sequencing in cancer. *Genome Med* 3:31. doi:[10.1186/gm247](https://doi.org/10.1186/gm247)
22. Guo G et al (2013) Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* 13:492–505. doi:[10.1016/j.stem.2013.07.017](https://doi.org/10.1016/j.stem.2013.07.017)
23. Gawad C, Koh W, Quake SR (2014) Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc Natl Acad Sci U S A* 111:17947–17952. doi:[10.1073/pnas.1420822111](https://doi.org/10.1073/pnas.1420822111)
24. Macaulay IC, Voet T (2014) Single cell genomics: advances and future perspectives. *PLoS Genet* 10:e1004126. doi:[10.1371/journal.pgen.1004126](https://doi.org/10.1371/journal.pgen.1004126)
25. Foloppe N, MacKerell AD Jr (1999) Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys J* 76:3206–3218. doi:[10.1016/S0006-3495\(99\)77472-2](https://doi.org/10.1016/S0006-3495(99)77472-2)
26. Mamedov TG et al (2008) A fundamental study of the PCR amplification of GC-rich DNA templates. *Comput Biol Chem* 32:452–457. doi:[10.1016/j.compbiolchem.2008.07.021](https://doi.org/10.1016/j.compbiolchem.2008.07.021)
27. Navin NE (2014) Cancer genomics: one cell at a time. *Genome Biol* 15:452. doi:[10.1186/s13059-014-0452-9](https://doi.org/10.1186/s13059-014-0452-9)
28. Xu X et al (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 148:886–895. doi:[10.1016/j.cell.2012.02.025](https://doi.org/10.1016/j.cell.2012.02.025)
29. Kurimoto K et al (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res* 34:e42. doi:[10.1093/nar/gkl050](https://doi.org/10.1093/nar/gkl050)
30. Ramskold D et al (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* 30:777–782. doi:[10.1038/nbt.2282](https://doi.org/10.1038/nbt.2282)
31. Chiang Y-S et al (2010) Single cell transcriptome analysis upon MCF-7 breast cancer. *Genome Biol* 11:P8