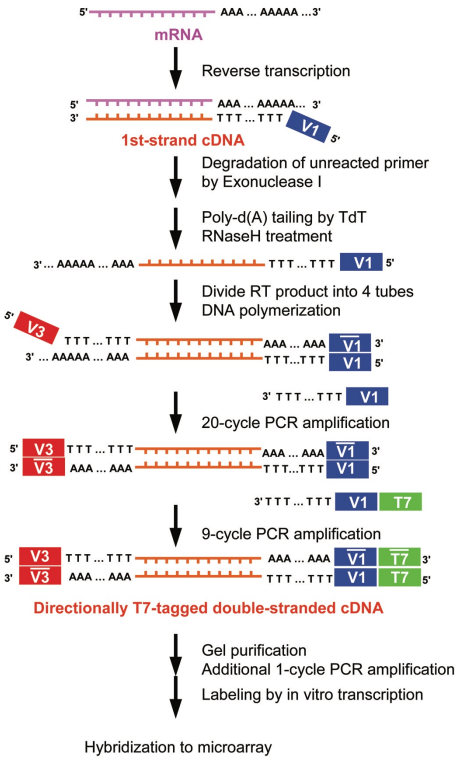
Single cell RNA-seq principle of past and present

Guo Siyuan 11611118

For a long time, the biological research method is carried out at the multicellular level, which means the results we get from experiments are the average values of the character of numerous cells. However, as we know, there is not a cell which is the same as another in biology examples. In another word, every cell is different, which is called the genetic heterogeneity. Usually, the traditional method loses information of heterogeneity, which makes researchers cannot comprehend phenomena at single-cell level. Of course, the flow cytometry (FCM) can find the heterogeneity in cells. But compared with single-cell sequencing, FCM provide fewer data. In the same way, single-cell sequencing can detect heterogeneous information that cannot be obtained by sequencing hybrid samples, which will lead to the new dimension of the whole field of biology.

It has been 10 years since the *Prof. Fuchou Tang* published the first real single-cell transcriptome study[2]. In this essay, I will talk about several changes experienced by the single-cell transcriptome principle. I believe that understanding the evolution of the single-cell transcriptome sequencing principle from a vertical development perspective will give us a deeper understanding of this technology.

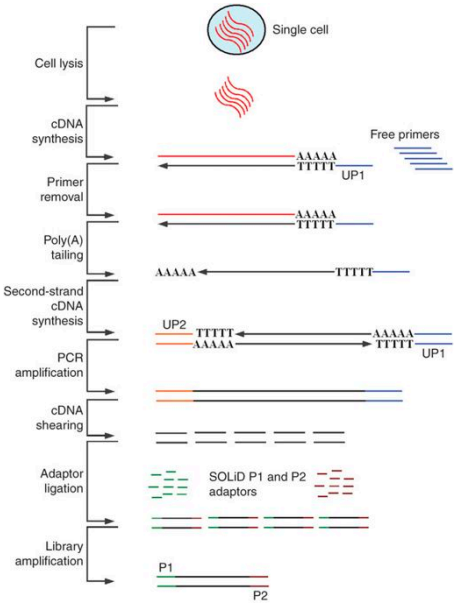
I think it should start from 2006. The reason why I didn't start from 2009 is that I think that this article by *Kurimoto* published in *Nucleic Acids Research*[1] has a great influence on the development of the principle of single-cell transcriptome sequencing. It is worth noting that the single cell transcriptome has been used to study the cell population.



1. The first strand cDNA is obtained by inverting the mRNA by the V1 primer with polyT, and then the excess primer is cleared by exonuclease I.
2. Adding polyA to the end of the reversed first strand cDNA by TdT (terminal deoxynucleotidyl transferase), and digest the RNA with RNaseH enzyme.
3. The cDNA with polyA tail is used as a template, and the double-stranded cDNA is formed by the V3 primer with polyT, and then the V1 primer is added to amplify, and finally the amplified double-stranded cDNA is formed.
4. Continuing to amplify with T7&V1 and V3 primers.

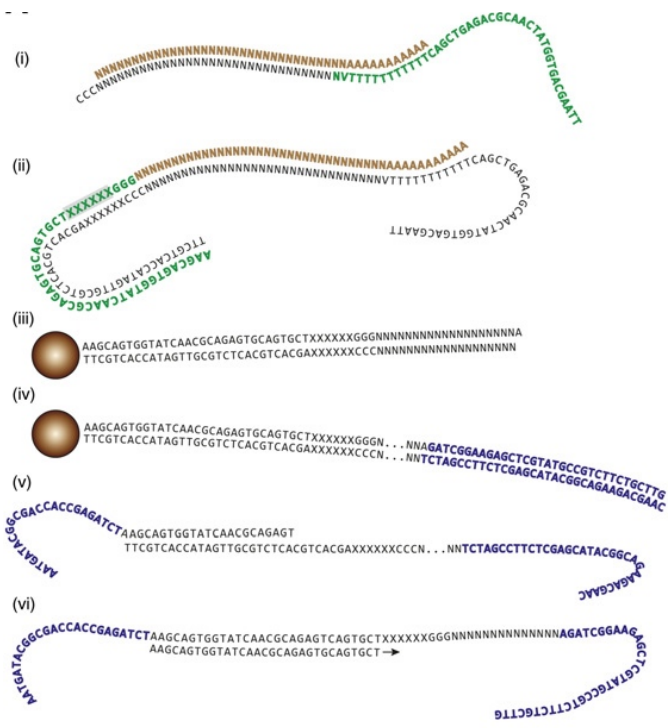
There are two main considerations for joining the T7 promoter:

1. The cDNA undergoes two amplifications of 20 cycles and 9 cycles, respectively, which will reduce the bias of PCR amplification compared to direct amplification of 29 cycles.
2. The latter part of the experiment was not directly sequenced, but the cDNA was reverse transcribed into RNA by the T7 promoter, and transcript information was obtained by hybridization with the microarray.



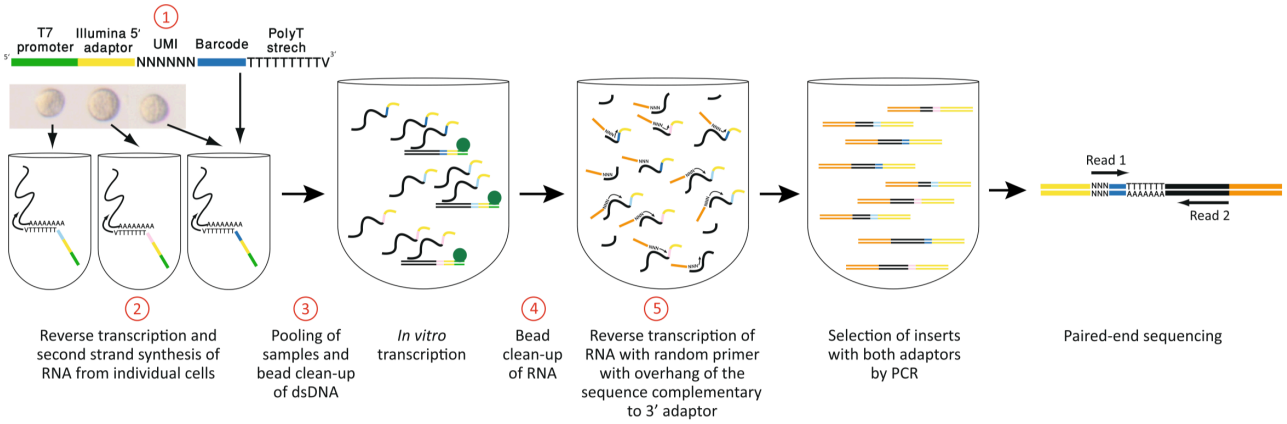
In 2009, *mRNA-Seq whole-transcriptome analysis of a single cell*[2] was published. This article officially opened the door to the single-cell transcriptome. Here, the article in 2006 was extended with the idea of adding polyA to the end of cDNA, but in the final reading of cDNA information, Applied Biosystem next-generation sequencing SOLiD system was used, which replaced the chip reading method. The step of adding the T7 promoter is reduced. (In 2005, the 454 system came out, SOLiD system came out in 2007) It is worth noting that at that time, there was already a low-input to the single cell level.

The next is the most widely used template conversion method. Perhaps the best one is SMART-seq. In fact, the START-seq[3] published in 2011 has been applied to the template conversion method, and the Barcode tagging method is used to achieve a relatively high-throughput single-cell transcriptome sequencing method. (The full-length cDNA was obtained by template conversion using the relevant reverse transcriptase in 1999[4])



1. Reversing the RNA by oligo-dT primer to obtain the first strand cDNA, and form 3-6 C bases at its end.
2. The template-switching oligonucleotide (TSO) primer added later, because it carries three RNA bases G during synthesis, which will hybridize with the C bases added to the ends of the first strand cDNA, and the template is converted into the first strand cDNA. In this way, we have a double-stranded cDNA.
3. Then use beads to grasp the 5' end, then fragment, and introduce the 3'-end sequencing adaptor by adding polyA to the end, and finally introduce the other end sequencing linked by PCR, and sequence it on the machine. Because the label of the cell is added at the 5' end, when fragmentation, only the fragment linked to Barcode has cell source information. Barcode can also be added to the 3' end, which enriches the 3' end sequencing. At the same time, if the sequencing linker is designed into the primers at the outset, it will not be needed to be introduced later. This stratagem can be applied in high-throughput sequencing. The scientist can also build a library of cDNA from a single cell without Barcode, so that they can get full-length cDNA information, that is, the SMART-seq1&2.

In the principle published in 2006, the T7 promoter is used. And the principle described below is related to it, here I directly introduce the advanced version of the method which was published in 2016, CEL-seq2[5]. CEL-seq2 uses an in vitro reverse transcription amplification method and introduces UMI (unique molecular identifiers). The approach using UMI to count the number of transcripts of RNA was published in *Nature Method* in 2011[6]. UMI was also used by almost all subsequent single-cell high-throughput transcriptome sequencing methods.



1. Starting from the mRNA in the cell, the oligo-dT primer used for reversal is rich in components this time, with Barcode, UMI, sequencing linker, and T7 promoter.

2. Reverse the RNA to form a strand cDNA, then purify the cDNA, forming ssRNA in vitro through the T7 promoter, and then reverse the RNA by random primers with the other end of the sequencing linker. In this way, we can obtain cDNAs with known sequencing linkers at both ends, and expand the library by using these two linkers.

In this essay, the main principles of obtaining cDNA libraries from mRNA in single-cell transcriptome sequencing are introduced. Of course, there are also single-cell methods involving some articles. I want to look at the development of the single-cell principle through a global vertical perspective, and then analyze it with our present God's vision, and we will find changes in our thinking.

Reference:

1. Kurimoto K , Yabuta Y , Ohinata Y , et al. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis[J]. Nucleic Acids Research, 2006, 34(5):17-0.
2. Tang F , Barbacioru C , Wang Y , et al. mRNA-Seq whole-transcriptome analysis of a single cell.[J]. Nature Methods, 2009, 6(5):377-382.
3. Islam S , Kjallquist U , Moliner A , et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq[J]. Genome Research, 2011, 21(7):1160-1167.
4. Schmidt W M , Mueller M W . CapSelect: A highly sensitive method for 5′ CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs[J]. Nucleic Acids Research, 1999, 27(21):e31.
5. Hashimshony T , Senderovich N , Avital G , et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq[J]. Genome Biology, 2016, 17(1):77.
6. Kivioja T , V?H?Rautio A , Karlsson K , et al. Counting absolute numbers of molecules using unique molecular identifiers[J]. NATURE METHODS, 2011, 9(1):72-74.