

Revealing cell fate decisions during reprogramming by scRNA-seq

Yu Liang ^{1,*}

¹Human Aging Research Institute, School of Life Science, Nanchang University, Nanchang, China.

Abstract. Single-cell RNA sequencing (scRNA-seq) technologies serve as powerful tools to dissect cellular heterogeneity comprehensively. With the rapid development of scRNA-seq, many previously unsolved questions were answered by using scRNA-seq. Cell reprogramming allows to reprogram the somatic cell into pluripotent stem cells by specific transcription factors or small molecules. However, the underlying mechanism for the reprogramming progress remains unclear in some aspects for it is a highly heterogeneous process. By using scRNA-seq, it is of great value for better understanding the mechanism of reprogramming process by analyzing cell fate conversion at single-cell level. In this review, we will introduce the methods of scRNA-seq and generation of iPSCs by reprogramming, and summarize the main researches that revealing reprogramming mechanism with the use scRNA-seq.

1 INTRODUCTION OF SINGLE-CELL RNA SEQUENCING TECHNOLOGIES

With the development of next-generation sequencing (NGS) technology, people have had a further understanding of genomics, transcriptomes, epigenetics. RNA-seq provides a far more accurate measurement of levels of transcripts and their isoforms using deep-sequencing technologies [1]. However, RNA-seq is originally used for a large population of cells, without distinguishing the differences of single cells from the population. In 2009, Tang et al described a method of single-cell RNA sequencing (scRNA-seq) using optimized NGS [2, 3]. With the establishment of scRNA-seq, these single-cell analyses will allow researchers to uncover new and potentially unexpected biological discoveries relative to traditional profiling methods that assess bulk populations [4].

The scRNA-seq protocols are usually divided into following steps: 1) single-cell isolation, 2) RNA capture and transcriptome amplification, 3) sequencing library preparation, 4) sequencing and data analysis [4].

An individual cell could be isolated in different approaches: limiting dilution [5, 6], micromanipulation, laser capture microdissection(LCM) [7, 8], flow-activated cell sorting (FACS) [9] or microfluidic technology. Limiting dilution with pipette is a commonly used approach to obtain single cells, which is low cost but with low efficiency due to the statistical distribution of cells. Micromanipulation is the classical method to isolate cells from embryos, however it is time-consuming and low-throughput [10]. LCM utilizes a laser system to isolate cells from solid samples which protect the cells but expensive. FACS is the most

commonly used method to isolate single cells in a high throughput manner. It is especially preferred when a particular cell population with specific cell surface markers need to be analyzed [11]. Recently, microfluidic technology has become a popular strategy for single cell isolation, for it enables precise fluid control, requiring low sample consumption meanwhile achieving high throughput analysis [12, 13]. Based on microfluidic systems, some widely used commercial platforms have been developed such as Fluidigm C1 and Chromium system from 10x Genomics, which provide a convenient and high-throughput processing method to study rare cell types in a heterogeneous population.

After isolation, single cells are lysed in a hypotonic buffer where mRNA are captured and bind by a poly(dT) primer. Single-cell transcriptome amplification requires first reverse transcription of mRNA into cDNA by PCR. Currently, both PCR and in vitro transcription (IVT) are widely used [14]. PCR takes the advantage of fast and exponential amplification of cDNA, but resulting primer dimers [14]. IVT utilizes T7 RNA polymerase to linearly amplify RNA from the first strand of cDNA, followed by an additional round of reverse transcription of the amplified RNA. Since both strategies generates some biases, a challenge is to develop sensitive, precise, and reliable technologies to generate amplified transcripts for sequencing from individual cells [15].

Prior to sequencing, sequencing library is prepared followed the protocol of commercial kits such as Nextera kit. Currently, cell barcoding is widely used to increase throughput for sequencing by addition of a specific tag for each cells. Sequencing for single cell is compatible with most bulk RNA-Seq platform such as illumina platform. ScRNA-seq data are generated and interpreted by appropriate computational and statistical methods,

* Corresponding author: yuliang_ly@163.com

which is indeed challenging to generate unbiased results from a mass of single-cell data.

2 METHODS OF SINGLE-CELL RNA-SEQ

A multitude of scRNA-seq methods have been developed in recent years, with dramatic advances in throughput and accuracy, and enabling major discoveries. These newly developed scRNA-seq methods differ in how single cells are sorted, and how they tag transcripts for the cell-of-origin and generate library for sequencing. Here, we will discuss major scRNA-seq methods (Table 1).

In 2009, Tang et al described a method of scRNA-seq for the first time. They modified the widely used method to achieve single-cell whole transcriptome amplification, generating cDNA as long as 3 kilobases (kb) efficiently and without biases [2, 16, 17]. Poly A was added to the 3' end of first-strand cDNA up to 3kb, as the binding site of poly T in the second-strand cDNA. They found the genes and splicing sites that could not be figured out previously in mouse blastomeres. However, the limitation for Tang's method was that they could not determine the length and sequence of mRNA, or the strand location of transcripts.

Islam et al introduced single-cell tagged reverse transcription (STRT) to specifically select and sequence RNA 5' end in 2011 [18]. The key of this method is adding an additional CCC sequence in the 3' end of first-strand cDNA. Furthermore, a helper oligo with distinct barcodes and a universal primer sequence was used. Based on the 5' end selection, STRT is suitable for identification of transcription start site (TSS) which is often lost in methods that show 3' bias like Tang's method.

Ramsköld et al developed Smart-seq in 2012. Smart-Seq generates and amplifies full-length cDNA from a single cell by exploiting the template-switching capacity of the reverse transcriptase from M-MLV. Template-switching and the terminal transferase activity of the enzyme are critical to the success of Smart-seq. As the enzyme arrives at the 5'-end of the mRNA, the reverse transcriptase adds a few non-templated C nucleotides and switches templates to reverse-transcribes to the end of the oligonucleotides. Smart-seq improved read coverage across transcripts, which significantly enhances detailed analyses of alternative transcript isoforms and identification of SNPs [19]. On the basis of Smart-seq, in 2013 they developed Smart-seq2 which improved coverages and accuracy compared to smart-seq libraries and generated with off-the-shelf reagents at lower cost [20]. In Smart-seq2, template-switching oligo (TSOs) results a two-fold increase in cDNA yield compared to SMARTer IIA oligo in Smart-seq. In addition, each detailed procedures were optimized in smart-seq2 to enable higher sensitivity and fewer technical biases.

CEL-Seq, a method for overcoming inefficient application of single cells was developed by barcoding and pooling samples before linearly amplifying mRNA with the use of in vitro transcription [21]. The reverse-

transcription primer was designed with an anchored polyT, a unique barcode, the 5' Illumina adaptor, and a T7 promoter. Comparing with PCR-based amplification method, it gives more reproducible, linear, and sensitive results. CEL-seq2 was developed in 2016, with higher sensitivity, lower costs, and less hands-on time [22]. In CEL-seq2, shorter primers integrated a unique molecular identifier (UMI), more effective reverse transcriptases and optimized clean-up method were used. By directly comparing with Smart-seq, CEL-seq2 showed better sensitivity and reproducibility [23].

In 2015, two different groups developed Drop-seq [12] and inDrop [13], by combining scRNA-seq and fluidic control of nanoliter-sized droplets. These drop-based methods resolved the limitations of ease and scale for broad application of scRNA-seq. In Drop-seq, thousands of cells can be analysis in a single reaction for each bead barcoded uniquely. In contrast to Drop-seq, which has 16 million barcodes in its bead library, the inDrop method generated only about 150,000 unique barcodes, which means it can process fewer cells in a single run.

In order to properly choose among available scRNA-Seq methods, a research systematically compared different scRNA-seq methods on their sensitivity, accuracy, precision and cost-efficiency. The results showed that Smart-seq on microfluidic platform is most sensitive, CEL-seq is the most precise method, and Drop-seq is the most efficient methods. Generally, accuracy is similar for all methods [24].

3 APPLICATION OF SCRNA-SEQ TO EXPLORE HETEROGENEITY

Recently, scRNA-seq has been widely used in tumor, developmental biology, neuroscience and other fields. It can reveal the differential expression of genes in single-cell level, which is helpful to study the cellular heterogeneity and randomness of gene regulation.

Cell heterogeneity is an important characteristic of tumor, which are divided into a diversity of genotypes and phenotypes. Kim et al [25] applied scRNA-seq to analyze the clonal evolution of breast cancer patients with chemotherapy. Heterogeneity within the tumor was found at the single-cell level, and different targeted drug pathway activation was found in refractory patients. Importantly, therapeutic strategies were developed to target two independent pathways in metastatic cancer cells.

Since only few cells in the early embryonic development of mammals, applying scRNA-seq to analyze embryonic development is important. Xue et al reported a comprehensive analysis of transcriptome dynamics from oocyte to morula in both human and mouse embryos using scRNA-seq as Tang described [3, 26]. They found that each development stage can be delineated concisely by a small number of functional modules of co-expressed genes, indicating a sequential order of transcriptional changes in several key pathways.

Also, scRNA-seq was widely used in neuroscience. Recently, Spaethling et al developed a culture system,

for culturing resected adult human brain tissue removed during neurosurgery [27]. By scRNA-seq, they identified

Table 1. Comparison of major scRNA-seq methods.

	Tang2009	STRT	Smart-Seq/ Smart-Seq2	CEL-Seq/ CEL-Seq2	Drop- Seq/inDrop
PolyA mRNA capture	PCR primer with an amine at the 5' end	Universal primer and oligo with six-base barcodes	Oligo-dT primer	primer with an anchored polyT, a unique barcode, the 5' Illumina adaptor, and a T7 promoter	Barcodes and UMI
cDNA amplification	PCR, full length	Single-primer PCR	PCR, full length	T7 promoter transcribed in vitro	PCR
Fragment and libraries preparation	Fragmented, then adding SOLiD library adapters	Fragmented by DNaseI without Mn ²⁺ , then ligating with an adapter containing the Illumina P2 sequence	Using Tn5 transposase, cDNA tagged with Nextera	Fragmented, then adding Illumina adapters	cDNA Tagmentation using NexteraXT kit
Transcription coverage	Full length	5' end	Full length	3' end	3' end

oligodendrocytes, microglia, neurons, endothelial cells, and astrocytes after 3 weeks in culture and detected more than 12,000 expressed genes including hundreds of cell-type enriched mRNAs, lncRNAs and pri-miRNAs.

4 GENERATION OF IPSCS BY REPROGRAMMING

Stem cells, a kind of cells that can differentiate into cell types, have the potential of self-renewal and directional differentiation. During development, uncommitted stem cells differentiate to various cell types. The lineage commitment and differentiation was considered to be unidirectional and irreversible traditionally, represented by a “Waddington’s landscape model”. However, this classic view of cell fate hierarchy was soon challenged by the study of somatic cell nuclear transfer (SCNT), which showed that somatic epigenome can be reprogrammed to pluripotency. Thereafter, numerous studies of cell fusion and transdifferentiation supported that cell fates could be manipulated in vitro. In 2006, Shinya Yamanaka demonstrated that mature somatic cells can be reprogrammed to a pluripotent state by defined factors, generating induced pluripotent stem cells (iPSCs). Since then, remarkable advances have been achieved in optimization of the methodology, and the application of iPSCs technologies to regenerative medicine and disease modeling.

Yamanaka’s team introduced four key factors, *Oc3/4*, *Sox2*, *Klf4*, and *c-Myc* (referred as OSKM), to generate ESC-like (embryonic stem cell) iPSCs from mouse embryonic and adult fibroblast cultures. Following the exciting research, three groups successfully achieved to generate iPSCs that were indistinguishable to ESCs [28-30]. Thereafter, reprogramming of human cells to iPSCs was achieved in 2007 [31, 32]. Yamanaka’s group demonstrated the induction of pluripotent cells from human fibroblasts by OSKM using a retroviral system, while Thomson’s group developed four different factors, *Oct4*, *Sox2*, *Nanog*, and *Lin28*, using a lentiviral

system. Since the integrating viral approaches for introduction of OSKM may lead to insertion mutations, interfering with the expression regulation of endogenous genes or being abnormally reactivated in terminally differentiated cells. Therefore, non-integrating viral approaches such as using Sendai virus or adenovirus to generate iPSCs was developed [33, 34]. Non-transgene reprogramming techniques such as delivering reprogramming factors in the form of mRNA or protein is a safer approach, avoided the possible infections in patients and the lack of a need to handle viral particles.

Besides OSKM, a unique set of miRNAs have been found specific in human embryonic stem cells (hESCs) and played important roles in reprogramming [35-38]. There were some findings provided the first proof-concept that miRNAs have the capability of directly converting fibroblasts to a cardiomyocyte-like phenotype in vitro [39]. Additionally, some antibodies were identified that can catalyze cellular dedifferentiation and nuclear reprogramming by acting at the cell surface [40]. According to this study, generating iPSCs by manipulating some pathway can efficiently eliminate the influence of exogenous induction and avoid the use of oncogenes.

5 UNDERSTANDING REPROGRAMMING TO THE INDUCED PLURIPOTENT STATE

Numerous studies have been made in understanding of the mechanism of iPSC reprogramming particularly in terms of transcriptional events. Many reports now suggest that successful reprogramming requires stepwise transition through key intermediate stages, and a few cells reach the final stage [41]. A study of mouse iPSC reprogramming found a sequential expression of pluripotency markers during the reprogramming process. The found that alkaline phosphatase was activated first, followed by silencing of somatic-specific genes and expression of SSEA1 in the specific stage [42]. In the

last, progressive silencing of exogenous genes with concomitant upregulation of endogenous Oct4 and Nanog marked fully reprogrammed cells [42]. Usually, induction of proliferation and acquisition of epithelial characteristics are accompanied in the early stage [41]. Nowadays, several models, such as elite model and stochastic/deterministic model, have been proposed to represent how cell fates and cell transitions are regulated during the reprogramming process [43]. However, these models could only partially explain the reprogramming mechanisms.

Interestingly, somatic reprogramming through forced expression of the same exogenous OSKM transcription factors in human and mouse cells generate pluripotent stem cells equivalent to different embryonic counterparts [44]. Mouse iPSCs are reprogrammed into a 'naïve' state similar to the state of mouse embryonic stem cells (mESCs), which are derived from inner cell mass (ICM) of developing blastocysts [45, 46]. Human iPSCs are in a 'primed' state similar to the state of human ESCs, and mouse epiblast stem cells (mEpiSCs) which are derived from the post-implantation epiblast of murine embryos [47]. Efforts have been made to realize the reprogramming of human cells to a 'naïve' state. Currently, several 'naïve'-specific, but not 'primed'-specific, cell surface marker proteins were demonstrated in 'naïve' and 'primed' human pluripotent stem cells [48]. More mechanistic insights of the difference in mouse and human somatic cell reprogramming would be beneficial for generation of human 'naïve' iPSCs.

Although our understanding of somatic cell reprogramming have largely improved since its initial discovery, it is possible that still only a small part of the complete picture has been revealed. Understanding the mechanism of reprogramming to pluripotency will have important implications not only to improve the method, but will help to propel their therapeutic applications. Based on the great improvement on scRNA-seq technology, the mechanism of reprogramming to the pluripotent state would be better understood.

6 REVEALING REPROGRAMMING PATH BY SCRNA-SEQ

Despite progresses made through bulk analyses as outlined above, very little is known at the single-cell level in reprogramming. Dissection of sophisticated biological processes at the single-cell resolution will facilitate the capture of transcriptional snapshots of infrequent occurrences. More researchers start to implement scRNA-seq technology to deconstruct the cellular heterogeneity and reprogramming trajectory (Figure 1).

It is widely accepted that all cells convert with different speed and through a variety of paths during reprogramming. In previous research, a single-cell transcriptional analysis of mouse fibroblasts reprogramming suggested that the reprogramming process had two phases: a prolonged stochastic phase followed by a rapid deterministic phase [49]. Following the research, Chung et al demonstrated that the

stochastic phase of reprogramming to pluripotency is an ordered probabilistic process through the single-cell transcript analysis of MRC5 human lung fibroblasts undergoing reprogramming by OSKM [50]. In addition, they found that cells followed two trajectories: one toward an ESC-like state (the "productive" trajectory) and the other away from both ESC and fibroblasts (the "alternative" trajectory) [50].

To find out the process of reprogramming, Francesconi et al employed MARS-Seq to analysis two types of reprogramming, the transdifferentiation of pre-B cells into macrophages induced by the TF C/EBP α [51] and the reprogramming of pre-B cells into iPSCs, based on the transient expression of C/EBP α followed by the induction of OSKM [52]. They found different activity of Myc have influences on efficiency of transdifferentiation and reprogramming [53]. Furthermore, these results illustrated the advantages of scRNA-seq on characterizing heterogeneity in cell fate conversion processes and identifying its underlying causes.

Several studies developed new mathematical algorithms for single-cell reprogramming analysis. Lin et al developed Single-cell Orientation Tracing (SOT) to analyze cell fate continuum based on single-cell RNA sequencing data generated in two distinct reprogramming system, the OSK-mediated reprogramming and the chemical reprogramming. They analyzed more than 150,000 single cells and found that cells bifurcate into two categories, reprogramming potential (RP) or non-reprogramming (NR). [54]. Another study introduced Waddington-OT to reconstruct the landscape of reprogramming from 315,000 scRNA-seq profiles, which revealed a wider range of developmental programs than previously characterized [55]. They found that cells adopt a mesenchymal-to-epithelial transition state will give rise to populations related to pluripotent, extra-embryonic, and neural cells, with each harboring multiple finer subpopulations. This study also predicts transcription factors and paracrine signals that affect cell fates during reprogramming.

Chemical reprogramming uncovered an intermediate extraembryonic endoderm (XEN)-like state, which is not exist in OSKM reprogramming [56]. ScRNA-seq analysis of 36,199 cells at multiple time points throughout the chemical reprogramming process reveal that a dynamic early two-cell (2C) embryonic-like programs are key aspects of successful reprogramming from XEN-like state to pluripotency [57]. Moreover, the reprogramming process is greatly accelerated via enhancing the 2C-like program by fine-tuning chemical treatment.

Collectively, these studies using scRNA-seq to analyze reprogramming shed light on mechanistic insights into the nature of induced pluripotency previously unknown, and provided great value for optimization of the method of reprogramming.

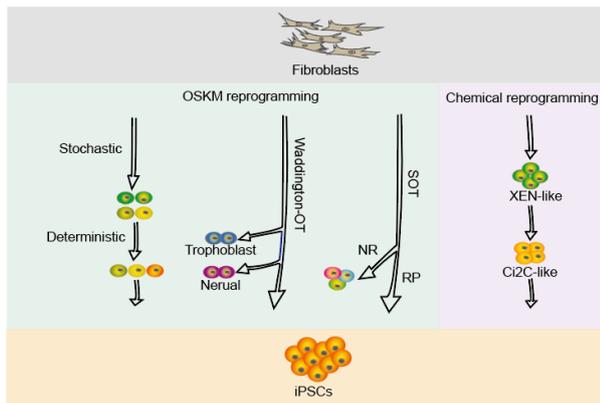


Fig. 1. The schematic diagram revealing the reprogramming path using scRNA-seq

7 SUMMARY

scRNA-seq technology have provided great insight into our understanding of biological diversity and rare cells that have previously been difficult to resolve from bulk tissue samples. Great achievement has been made in embryonic development, cancer, stem cells, immune system and tracing cell types in organs/tissues by using scRNA-seq. Since the advent of reprogramming to generate iPSCs, numerous studies have focused on revealing the mechanism of reprogramming. ScRNA-seq served as a powerful instrument to uncover previously unclear intermediate stages masked in the mixed cell population. The mechanisms of reprogramming proved by scRNA-seq provide more information on how single cell fate converted, indicating further researches on reprogramming with the use of scRNA-seq. More progress in understanding reprogramming mechanism or improving reprogramming method can be made with the use of scRNA-seq technology. It helps to figure out differences between single-cell fates, and the key factors devoted to reprogramming. Furthermore, developing appropriate algorithm to analyze the high-throughput data is crucial for obtaining the desired answers. Different algorithms have been developed to solve some single-cell biological questions, however more accurate, universal, and fast one is still needed.

REFERENCES

1. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nature Reviews Genetics, 2009. **10**(1): p. 57-63.
2. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell*. Nat Methods, 2009. **6**(5): p. 377-82.
3. Tang, F., et al., *RNA-Seq analysis to capture the transcriptome landscape of a single cell*. Nat Protoc, 2010. **5**(3): p. 516-35.
4. Hwang, B., J.H. Lee, and D. Bang, *Single-cell RNA sequencing technologies and bioinformatics pipelines*. Exp Mol Med, 2018. **50**(8): p. 96.
5. Reizel, Y., et al., *Colon stem cell and crypt dynamics exposed by cell lineage reconstruction*. PLoS Genet, 2011. **7**(7): p. e1002192.
6. Zhang, K., et al., *Sequencing genomes from single cells by polymerase cloning*. Nat Biotechnol, 2006. **24**(6): p. 680-6.
7. Frumkin, D., et al., *Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues*. BMC Biotechnology, 2008. **8**(1).
8. Yachida, S., et al., *Distant metastasis occurs late during the genetic evolution of pancreatic cancer*. Nature, 2010. **467**(7319): p. 1114-7.
9. Stepanauskas, R. and M.E. Sieracki, *Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time*. Proc Natl Acad Sci U S A, 2007. **104**(21): p. 9052-7.
10. Brehm-Stecher, B.F. and E.A. Johnson, *Single-cell microbiology: tools, technologies, and applications*. Microbiol Mol Biol Rev, 2004. **68**(3): p. 538-59, table of contents.
11. Julius, M.H., T. Masuda, and L.A. Herzenberg, *Demonstration that antigen-binding cells are precursors of antibody-producing cells after purification with a fluorescence-activated cell sorter*. Proc Natl Acad Sci U S A, 1972. **69**(7): p. 1934-8.
12. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets*. Cell, 2015. **161**(5): p. 1202-1214.
13. Klein, Allon M., et al., *Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells*. Cell, 2015. **161**(5): p. 1187-1201.
14. Kolodziejczyk, A.A., et al., *The technology and biology of single-cell RNA sequencing*. Mol Cell, 2015. **58**(4): p. 610-20.
15. Kanter, I. and T. Kalisky, *Single cell transcriptomics: methods and applications*. Front Oncol, 2015. **5**: p. 53.
16. Kurimoto, K., et al., *An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis*. Nucleic Acids Res, 2006. **34**(5): p. e42.
17. Kurimoto, K., et al., *Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis*. Nat Protoc, 2007. **2**(3): p. 739-52.
18. Islam, S., et al., *Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq*. Genome Res, 2011. **21**(7): p. 1160-7.
19. Ramsköld, D., et al., *Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells*. Nature Biotechnology, 2012. **30**(8): p. 777-782.
20. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells*. Nature Methods, 2013. **10**(11): p. 1096-1098.

21. Hashimshony, T., et al., *CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification*. Cell Rep, 2012. **2**(3): p. 666-73.
22. Hashimshony, T., et al., *CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq*. Genome Biology, 2016. **17**(1).
23. Hashimshony, T., et al., *CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq*. Genome Biol, 2016. **17**: p. 77.
24. Ziegenhain, C., et al., *Comparative Analysis of Single-Cell RNA Sequencing Methods*. Molecular Cell, 2017. **65**(4): p. 631-643.e4.
25. Kim, K.T., et al., *Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells*. Genome Biol, 2015. **16**: p. 127.
26. Xue, Z., et al., *Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing*. Nature, 2013. **500**(7464): p. 593-7.
27. Spaethling, J.M., et al., *Primary Cell Culture of Live Neurosurgically Resected Aged Adult Human Brain Cells and Single Cell Transcriptomics*. Cell Rep, 2017. **18**(3): p. 791-803.
28. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells*. Nature, 2007. **448**(7151): p. 313-7.
29. Wernig, M., et al., *In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state*. Nature, 2007. **448**(7151): p. 318-24.
30. Maherali, N., et al., *Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution*. Cell Stem Cell, 2007. **1**(1): p. 55-70.
31. Takahashi, K., et al., *Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors*. Cell, 2007. **131**(5): p. 861-872.
32. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells*. Science, 2007. **318**(5858): p. 1917-20.
33. Liang, P., et al., *Patient-Specific and Genome-Edited Induced Pluripotent Stem Cell-Derived Cardiomyocytes Elucidate Single-Cell Phenotype of Brugada Syndrome*. J Am Coll Cardiol, 2016. **68**(19): p. 2086-2096.
34. Al-Gharaibeh, A., et al., *Induced Pluripotent Stem Cell-Derived Neural Stem Cell Transplantations Reduced Behavioral Deficits and Ameliorated Neuropathological Changes in YAC128 Mouse Model of Huntington's Disease*. Front Neurosci, 2017. **11**: p. 628.
35. Suh, M.R., et al., *Human embryonic stem cells express a unique set of microRNAs*. Dev Biol, 2004. **270**(2): p. 488-98.
36. Bar, M., et al., *MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries*. Stem Cells, 2008. **26**(10): p. 2496-505.
37. Morin, R.D., et al., *Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells*. Genome Res, 2008. **18**(4): p. 610-21.
38. Laurent, L.C., et al., *Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence*. Stem Cells, 2008. **26**(6): p. 1506-16.
39. Jayawardena, T.M., et al., *MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes*. Circ Res, 2012. **110**(11): p. 1465-73.
40. Blanchard, J.W., et al., *Replacing reprogramming factors with antibodies selected from combinatorial antibody libraries*. Nat Biotechnol, 2017. **35**(10): p. 960-968.
41. Plath, K. and W.E. Lowry, *Progress in understanding reprogramming to the induced pluripotent state*. Nat Rev Genet, 2011. **12**(4): p. 253-65.
42. Brambrink, T., et al., *Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells*. Cell Stem Cell, 2008. **2**(2): p. 151-9.
43. Takahashi, K. and S. Yamanaka, *A decade of transcription factor-mediated reprogramming to pluripotency*. Nat Rev Mol Cell Biol, 2016. **17**(3): p. 183-93.
44. Teshigawara, R., et al., *Mechanism of human somatic reprogramming to iPS cell*. Lab Invest, 2017. **97**(10): p. 1152-1157.
45. Gafni, O., et al., *Derivation of novel human ground state naive pluripotent stem cells*. Nature, 2013. **504**(7479): p. 282-6.
46. Tesar, P.J., et al., *New cell lines from mouse epiblast share defining features with human embryonic stem cells*. Nature, 2007. **448**(7150): p. 196-9.
47. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos*. Nature, 1981. **292**(5819): p. 154-6.
48. Collier, A.J., et al., *Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States*. Cell Stem Cell, 2017. **20**(6): p. 874-890 e7.
49. Buganim, Y., D.A. Faddah, and R. Jaenisch, *Mechanisms and models of somatic cell reprogramming*. Nat Rev Genet, 2013. **14**(6): p. 427-39.
50. Chung, K.M., et al., *Single cell analysis reveals the stochastic phase of reprogramming to pluripotency is an ordered probabilistic process*. PLoS One, 2014. **9**(4): p. e95304.
51. Xie, H., et al., *Stepwise Reprogramming of B Cells into Macrophages*. Cell, 2004. **117**(5): p. 663-676.
52. Di Stefano, B., et al., *C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells*. Nature, 2014. **506**(7487): p. 235-9.

53. Francesconi, M., et al., *Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming*. Elife, 2019. **8**.
54. Guo, L., et al., *Resolving Cell Fate Decisions during Somatic Cell Reprogramming by Single-Cell RNA-Seq*. Mol Cell, 2019. **73**(4): p. 815-829 e7.
55. Schiebinger, G., et al., *Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming*. Cell, 2019. **176**(4): p. 928-943 e22.
56. Hou, P., et al., *Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds*. Science, 2013. **341**(6146): p. 651-4.
57. Zhao, T., et al., *Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming*. Cell Stem Cell, 2018. **23**(1): p. 31-45 e7.