ISSN No:-2456-2165

Bioinformatics Tools and Methods to Analyze Single-Cell RNA Sequencing Data

Sudheer Menon[1], Vincent Chi Hang Lui [1,2], Paul Kwong Hang Tam [1,2]
[1]Department of Surgery, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong.
[2]Dr. Li Dak-Sum Research Centre, The University of Hong Kong – Karolinska Institutet Collaboration in Regenerative Medicine, the University of Hong Kong, Hong Kong.

Abstract:- The rapid development of next generation sequencing (NGS) tool and technologies over the past few years, valuable insights have been gained into the complex and diverse biological systems with diverse range from microbial communities to cancerous genome. The NGS base technologies based of genomics, epigenomics and transcriptomics, are preoperationally concentrated on individual cells characterization. As an example, ScRNA sequencing can sort out complex and rare cellular populations, revealing the gene regulatory inter and intra relationships, and trace the progression of individual cell lineages during cell development. Researchers have previously undiscovered molecular details in single-nuclei and single-cell sequencing experiments. Since recent years, analytical approaches and sequencing methods have increased rapidly. In this review, our prime focus will be on challenges related to single cell isolation (SCI) and library preparation, analysis of ScRNA seq data by computational pipelines. These sequencing technologies will be greatly facilitating molecular biological improvement and bioinformatics due to the availability of bioinformatics tools.

Keywords:- *Single-cell RNA sequencing, alternative splicing, Next generation sequencing.*

I. INTRODUCTION

In a typical single-cell sequencing workflow, tissue and cells are prepared, cells are captured and a library is prepared, cells are sequenced, raw data is processed, and visualization and downstream analyses are performed. Since each tissue and cell type to be isolated may require a different protocol for preparing single cell suspensions there are multiple ways for their preparation. Genomics and transcriptome analysis are both powerful tools for tackling the long-standing problem of mapping genotypes to phenotypes in biology and medicine. The sequencing entire transcriptomes were equal to single cell level that was pioneered by Is cove and colleagues (Brady et al., 1990) and James Eber wine (James Eber wine *et al*., 1992) Through linear amplification by in vitro transcription and PCR base exponential amplification to expand the complementary DNA (cDNA) of the individual cells and cellular lines. Initially developed to detect DNA base microarray chips, and these recent technologies have been adapted for single cell RNA (ScRNA) sequencing (Tietjen *et al*., 2003). The first report on single cell transcriptome (Sc-trans) analysis was published that described the cell characterization at initial developmental stages (Tang *et al.,* 2009). Recent technical developments, and bioinformatics tools and techniques advanced researchers' ability to Insilco evaluation of diversified populations of immunological cells in health and disease (Shalek *et al.,* 2013). Further ScRNA sequencing is variously used to visualize the relationships of cell lineage in early developmental stages like, differentiation of myoblasts and the determination of lymphocyte fates (Stubbington et al., 2016). Similarly, some analyses such as detection of alternative splicing, exploration of allelic expression, and the identification of edited RNA, do not apply to ScRNA-seq protocols, which generate data for whole-transcript ScRNA-seq.

There are number of sequencing techniques, platforms which can affect sequencing data and also have been developed different techniques. As a result, choosing an appropriate analytical approach is essential to efficiently deal with the highly variable ScRNA sequencing data (Bacher and Kendziorski, 2016). A search tag cloud on SEQ answers shows that RNA-Seq is one of the top subjects in NGS (www.seqanswers.com/forums). Serial Gene expression analysis and Microarrays base studies are being replaced by RNA sequencing methodologies to quantify gene expressions. Because of the great depth of sequencing techniques, RNA sequencing can yield a repertoire of all the transcripts present in the tissue site at any specific point in time, including rare transcripts. By this, it produces almost complete portrait of the transcriptomic events occurred in any living cell. The multipurpose data obtained from this can be utilized in gene characterize to reveal information on novel transcripts, single-nucleotide polymorphisms, alternative splicing that have to measure gene expression level, single nucleotide polymorphisms and structural variations (Novales *et al*., 2008; Dissanayake *et al.,* 2009; Brautigan et al., 2011; Alagna *et al*., 2009)

The RNA sequencing about non-model species when the genetic tools and transcriptomic sequencing data may scarce. Furthermore, RNA is only sequenced in coding regions, instead of the whole genome, resources may be limited. RNA sequencing offers a few advantages over the whole-genome assembly, including lack of repetitive sequences and high GC participation. Various transcriptomic approaches have been used in plant species such as alga, moss non-model plant species (Trick *et al*., 2009; Franssen *et al.*, 2011).

➢ *Analyzing networks*

RNA-seq provides the unprecedented potential to uncover transcriptional regulatory networks' structure and function. Many studies are devoted to this goal. Genes or transcripts are regarded as nodes in a regulatory network reconstruction problem, and their edges represent the interactions or dependencies between them. The expression like weighed gene network analysis (WGCNA) is a technique that mainly used in bulk RNA sequencing (Zang 2005). Additionally, WGCNA has already contributed to new insights and into embryonic development through sequencing ScRNA (Xue *et al.,* 2003). WGCNA was executed on cells at the same developmental stage, thereby minimizing errors induced by high correlation coefficients due to subpopulations. WGCNA and other methods for reconstructing association networks have enabled significant progress in bulk studies. This technique was expected to do well in sequencing of single RNA molecules, but they do not show how nodes connect. Typically experiments involving temporal or perturbation are needed. In any case information provided by pseudo temporal ordering approaches, may enable partial order inference from snapshot cRNA-seq experiments. In two recent methods combine this information's with approaches familiar from traditional regulatory network reconstruction to arrive at inferences of regulatory relationships among genes (Ocone *et al.,* 2015). To enhanced innovative methods to do so in a very effective way to developed upon the type of information that can now be obtained from snapshot that have used for single sequencing experiments.

➢ *Single-cell isolation*

SC RNA-seq begins with the isolation of individual cells that despite the challenge of capturing them. Single cells, isolation is a procedure separating different living cells that are currently isolated by several different ways such as by read mapping, microfluidics, flow-activated cell sorting and laser capture microdissection (Kolodziejski *et al.,* 2015; Hwang *et al.,* 2018). This method is not well efficient by using pipettes to dilution cells and limiting dilution technique to isolates the cells. A micromanipulator is a classical technique which is used to recover cells from micro-sized samples, because this technique is a time-consuming and low-throughput technique. In flow-activated cell sorting single cells are isolated by requiring large numbers of cells (>10,000 in suspension) by the help of a laser system and computer. An advanced technique that has been used through microscope separating individual cells from solid tissues by laser capture microdissection. Because it offers small sample consumption, precise fluid control and small analysis costs. Single -cell isolation has many advantages and also protocols related capturing the target cells that ensure their purity (Hu *et al.,*2016). Last but not least microfluidics is increasingly popular.

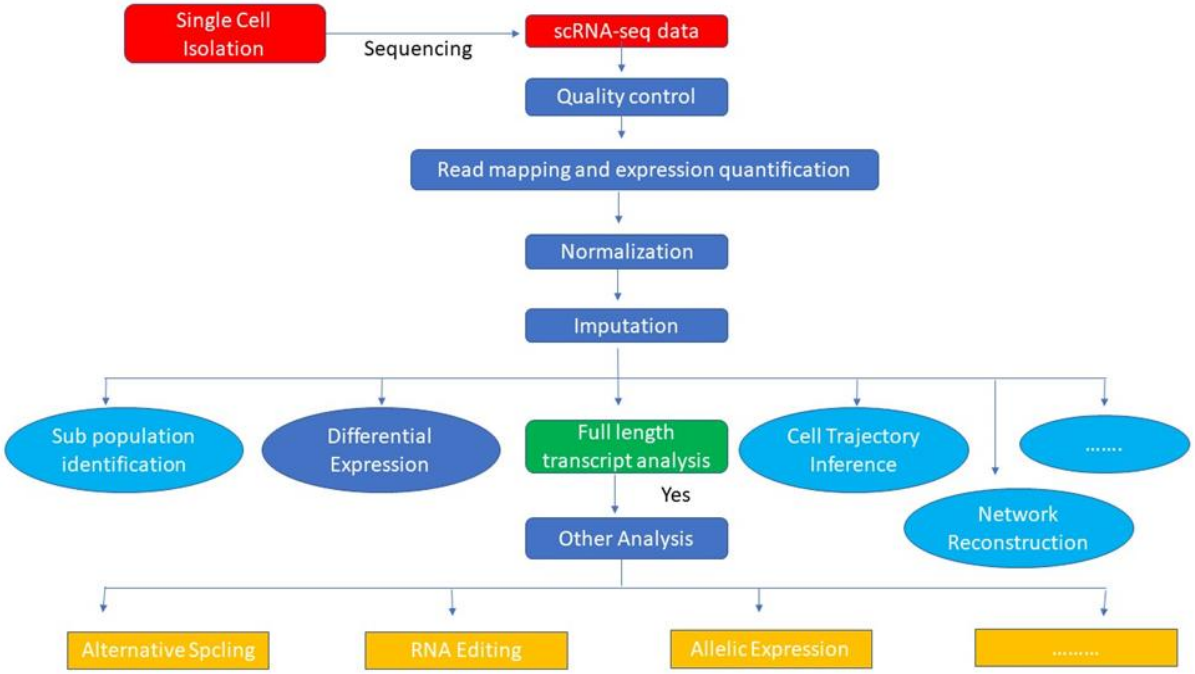

Fig 1:- This figure shows the all-different SC RNA-seq methods.

➢ *Technologies for isolating single cells*

Cell isolation may typically characterize by three parameters: efficiency, purity, and recovery. Currently, each of the three parameters has its advantages. Based on physical factors existing cell isolation technique such as size, density, electric changes and deformability they have fall into two such groups. Physically the isolation of single cells without labeling is the most advantageous feature. The second group of the affinity methods are based on biological proteins expression properties fluorescence-activated cell sorting, and magnetic-activated cell sorting (Dainiak*et al.,* 2007). Following are brief summaries of each method, highlighting their advantages and limitations (Table 1).

ISSN No:-2456-2165

Techniques	Throughput	Advantage	Disadvantage	References
Fluroesence - activated cell sorting	HIGH	High specificity multiple parameters	High skill needed	Gross et al., 2015
Magnetic - activated cell sorting	HIGH	Cost effective	Dissociated cells	Welzel et al., 2015
Laser capture microdissection	LOW	Intact fixed and live tissue	Contaminated by neighbouring cells	Espina et al., 2007
Manual cell picking	LOW	Intact live tissue	Low throughput	Citri et al., 2012
Microfluidic	HIGH	Integrated with amplification	High skill needed	Lecault et al., 2012

Table 1:- Techniques to isolate single cells

➢ *The genomics of individual cells*

By sequencing chromosomal variations like copy numbers and single-nucleotide variations we can identify chromosomal varieties. It showed to study different tumor evolution, gametic generation, causing genomic heterogeneity and mutations in a population. As a result, the amount of DNA in human genomes is often inadequate. E.g., the weightiness of one genomic DNA is just 6kg, currently in NGS applications there are copies in a normal cell which is not enough. Through using a traditional PCR method that causes simple difficulties and allelic dropout across the genome when applied to single cells. For genome sequencing unbiased amplification of DNA is crucial. There are different types of PCR such as linker adapter, primer extension and pre- amplification. A degenerate oligonucleotide-primed PCR was described by Telenius *et al.*, (1992) (Hubert *et al.*, 1992).

Due to its simplicity and high accuracy, procedures used in genomic analysis multiple displacement amplification techniques are one of the most popular methods. DNA can be amplified isothermally at 30˚C using random hexamer primers and phi29 DNA polymerase. During strand synthesis phi29 DNA polymerase can displace the strands within powerful strand displacement ability (Dean *et al.*, 2002).

An additional method, multiple annealing and looping-based amplification cycles (MALBAC) is used to detect accurate copy number of variations allowing for efficient genotyping of individual cell's uniformity (Zong*et al.*, 2012). Thus, MALBAC provides a significant advantage over other methods. Since it reduces the amplification errors and biases by copying the amplifier separately from the original template for starting material in exponential amplification (Wu *et al.*, 2014).

➢ *Technologies Available for ScRNA-Seq*

Single-cell transcriptomic research has incorporated a variety of ScRNA-seq technologies (Table2). Sequence-by-sequence approaches were developed and published firstly (Tang *et al.* 2009). Due to development of many other ScRNA-seq approaches were followed. There are different aspects of ScRNA-seq technologies such as cell breakdown, amplification, reverse transcription, unique molecular identifiers and transcript coverage. The major difference between different ScRNA-seq methods is that to capture full-length transcripts and sequence them while other firstly capture and sequence the 3′-end (Rosenberg *et al.*, 2018).

ISSN No:-2456-2165

Systems	Transcript coverage	UMI possibility	Strand specific	References
CEL-sequence	3 0 -one	Yes	Yes	Hashimshony *et al*., 2012
CEL-sequence2	3-one	Yes	Yes	Hashimshony et al., 2016
MARS-sequence	3 0 - one	Yes	Yes	Jaitin et al., 2014
Cyto Sequence	3 0 one	Yes	Yes	Fan H.C. et al., 2015
Drop-sequence	3 0 - one	Yes	Yes	Macosko et al., 2015
In Drop	3 0 -one	Yes	Yes	Klein et al., 2015
Chromium	3 0 -only	Yes	Yes	Zheng et al., 2017
Split-sequence	0 3 -only	Yes	Yes	Rosenberg et al., 2018
sci-RNA-sequence	3 0 -only	Yes	Yes	Cao et al., 2017
Seq-Well	0 3 -only	Yes	Yes	Gierahn et al., 2017
DroNC-sequence	0 3 -only	Yes	Yes	Habib et al., 2017
Quartz-Seq2	0 3 -only	Yes	Yes	Sasagawa et al., 2018
STRT-sequence and STRT/C1	50 -only	Yes	Yes	Islam et al., 2011, 2012
Tang method	Nearly full-length	No	No	Tang et al., 2009
Quartz-Sequence	Long	No	No	Sasagawa et al., 2013
Super-sequence	Long	No	No	Fan X. et al., 2015
Smart-sequence	Long	No	No	Ramskold et al., 2012
Smart-seq2	Long	No	No	Picelli et al., 2013
MATQ-sequence	Long	Yes	Yes	Sheng et al., 2017

Table 2:- Single-cell trajectory inference methods

There are many different Single cell RNA-seq procedures that may have disparate strengths and weaknesses (Kolodziejczyk *et al.,* 2015; Haque *et al.,* 2017; Picelli, 2017; Ziegenhain*et al.,* 2017). A study showed Smart-seq can detect a greater expressed number of genes than CEL-sequencing, MARS-seq, smart sequencing and drop sequencing tool (Hashimshony *et al.,* 2016; Jaitin *et al.,* 2014; Ramskold *et al.,* 2012; Ziegenhainet *al.,* 2017). The SCRNA-seq procedure has the highest technical efficiency than traditional bulk RNA-seq. Such as External RNA Control Consortium controls can be used to estimate technical variances between cells (External, 2005). The RNA spike-ins are RNA transcripts that hybridize and analyzes such as RNA-Seq, that can be calibrated through using unique molecular identifiers to estimate absolute molecular counts. Spike-ins is used by methods similar to higher protocol sequencing but not by droplet-based methods while unique molecular identifiers are typically employed by 3′-end sequencing technologies such as Drop-seq (Macosko *et al.,* 2015).

➢ *Uses of Bioinformatics tools*

A next-generation sequencing technique (NGS) is a high-throughput method that enables the identification of nucleotide sequences within DNA, and RNA molecules. Mathematical and statistical methods implemented in various programming paradigms and dedicated software tools can analyze and explain biological, molecular, cellular and genomic information (Metzker, M. L. 2010).

- KEGG Genes and proteins are organized into Ortholog groups and stored in the KEGG Orth ology Catalogue (Kanehisa *et al.*,2000)
- Green genes and Silva are ribosomal RNA gene databases for taxonomic annotation (McDonald *et al.*, 2012).
- Gene Bank built by National Center for Biotechnology Information, contains genome sequences over 250,000 species and data can be viewed by NCBI's retrieval system called Entrez. Included Coding and non-translated regions, promoters, terminators, axons, introns and repeats (Benson *et al*.,2002)

➢ *Data arranging and quantifying for ScRNA-Seq*

By Reading ratios remains important indicator for overall quality of single sequencing data and nucleotide - sequencing technologies, sequence transcripts into reads and then to generate raw, RNA-seq data in fastq format. Read alignment there have no difference between the RNA-seq have been discussed before, the mapping tools developed for bulk RNA-seq can also be used with ScRNA-seq data. (Li and Homer, 2010; Chen *et al.,* 2011). There are two basically read mapping algorithms and the spaced-seed indexing and Burrows-Wheeler transform based (Li and Homer, 2010). Aligners such as TopHat2 (Kim *et al.,* 2013). Read mapping and expression quantification is an array-based and suffix-based method that is more efficient than read mapping which is an important bioinformatic diverse tool, sequencing data for alignment of complementary deoxy ribose nucleic acid, but it requires considerable memory. There are currently few genome-guided assembly

tools that can achieve the exact same or better accuracy such as hierarchical indexing for spliced alignment of transcripts, RSEM software used for estimating gene, expression level and String tie which are commonly used to estimate the expression of genes or transcripts in ScRNA-seq studies by looking at the number of reads and transcripts per million (Pertea *et al.*, 2015)

Diverse Tools	Group	URL	References
TopHat2	**Read mapping**	https://ccb.jhu.edu/software/tophat/ index.shtml	**Kim *et al.*, 2013**
RSEM	**Expression quantification**	https://github.com/deweylab/RSEM	**Li and Dewey, 2011**
String Tie	**Expression quantification**	https://github.com/gpertea/stringtie	**Pertea *et al.*, 2015**
Cufflinks	**Expression quantification**	https://github.com/cole-trapnell-lab/cufflinks	**Trap Nell *et al.*, 2010**
RNA-seq Star	**Read mapping**	https://github.com/alexdobin/STAR	**Dobbin and Ginger as, 2015**
HISAT2	**Read mapping**	https://ccb.jhu.edu/software/hisat2/index.shtml	**Kim *et al.*, 2015**

Table 3:- Explanations of the major Single cell RNA-seq analyzing tools.

➢ *Study of RNA editing and alternative splicing techniques in ScRNA-Seq data*

Mainly in alternative splicing technology basically five modes that are commonly recognized for example, exon-skipping, mutually exclusive exons, alternative donor site, alternative acceptor site and intron retention. Some mechanisms played an important role in a variety of biological procedures and abnormal alternative splicing, which has been associated to cancer (Sven *et al.*, 2016). A lack of precision of SC RNA-seq data, splicing quantification led to the development of raw RNA-sequencing that is not suitable for SC RNA-seq data. It considers that expression dynamics influenced by important role in cell characterization through different parameters. It's promising to study AS at single-cell resolution to gain insights into isoform usage at the cellular level. There are different tools e.g., Single Splice, Expedition, BRIE, and census. Single Splice practices a statistical model to detect genes with significant isoform usage. Using a linear model, Census assigns a Dirichlet-multinomial distribution to the isoform counts in each gene. Differential isoform quantification based on different hierarchical models. Expedition contains a suite of algorithms for identifying AS, assigning splicing modes and visualizing modality changes (Table 3) (Welch *et al.*, 2016; Qui *et al.*, 2017; Huang and Sanguinetti, 2017; Song *et al.*, 2017). Currently, the RNA-editing detection have been difficult limited by SCRNA-seq and prevented the use of individual cells until now. By developing both single-cell editing detection algorithms and ScRNA-seq technologies will be possible to explore editing dynamics among single cells(Gott *et al.*, 2000)

Smart tools	URL	Reference
BRIE	https://github.com/huangyh09/brie	Huang and Sanguinetti, 2017
Census	http://cole-trapnell-lab.github.io/monocle-release/	Qiu *et al.*, 2017
Single Splice	https: //github.com/jw156605/Single Splice	Welch *et al.*, 2016
Expedition	https://github.com/YeoLab/Expedition	Song *et al.*, 2017

II. CONCLUSION

It is concluded that development and application of ScRNA-seq, information on the variability and dynamics of cell expression has been well gathered. Such single-cell analysis is positioned to provide a deeper understanding of biological complexity in disorders as well as at normal development. The issues mentioned above will solved early in future by the rapid progress in single cell isolation, new technologies are going to emerge as powerful way to clear these time-honored questions in biological research as well as clinical studies in the distant future. There have been proposed Single cell -sequencing methods are utilized for the frozen samples and their fixation are also proposed, by which heterogeneous clinical study will be benefited. Other benefits of developing protocol to capture not only poly A^+ but also polyA RNAs, the researchers will be able to obtain comprehensive pictures of both the protein-coding as well as the non-coding gene expressions in real time bases at single cell resolution level.

➢ *Abbreviations*

(NGS) next-generation sequencing, single-cell RNA (sc RNAsingle-cell suspensions(scS), serial analysis of gene expression (SAGE), single-nucleotide polymorphisms (SNPs), flow-activated cell sorting (FACS) laser capture microdissection (LCM)linker-adapter PCR (LA-PCR), interspersed repetitive sequence amplification PCR (IRS-PCR), multiple displacement amplification (MDA), multiple annealing and looping-based amplification cycles (MALBAC), transcripts per million reads (TPM)

ACKNOWLEDGEMENTS

This study receive funding from "The Hong Kong University Seed Funding for Strategic Interdisciplinary Research Scheme 2019/20"

REFERENCES

[1]. Alagna, F., & Nunzio, D. 'Agostino N, Torchia L, Servili M, Rao R et al.(2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. BMC Genomics, 10, 399.
[2]. Brady, G., Barbara, M., &Iscove, N. N. (1990). Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. Methods Mol Cell Biol, 2(1), 17-25.
[3]. Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., ... & Weber, A. P. (2011). An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. Plant physiology, 155(1), 142-156.
[4]. Chen, G., Wang, C., & Shi, T. (2011). Overview of available methods for diverse RNA-Seq data analyses. Science China Life Sciences, 54(12), 1121-1128.
[5]. Dainiak, M. B., Kumar, A., Galaev, I. Y., &Mattiasson, B. (2007). Methods in cell separations. Cell Separation, 1-18.
[6]. Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., ... & Coleman, P. (1992). Analysis of gene expression in single live neurons. Proceedings of the National Academy of Sciences, 89(7), 3010-3014.
[7]. Franssen, S. U., Shrestha, R. P., Bräutigam, A., Bornberg-Bauer, E., & Weber, A. P. (2011). Comprehensive transcriptome analysis of the highly complex Pisum sativum genome using next generation sequencing. BMC genomics, 12(1), 1-16.
[8]. Gott, J. M., and Emeson, R. B. (2000). Functions and mechanisms of RNA editing. Annu. Rev. Genet. 34, 499–531.
[9]. Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., &Koltay, P. (2015). Technologies for single-cell isolation. International journal of molecular sciences, 16(8), 16897-16919.
[10]. Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., De Leeuw, Y., Anavy, L., ... &Yanai, I. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome biology, 17(1), 1-7.
[11]. Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., ... & Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science, 343(6172), 776-779.
[12]. Kanehisa, M., &Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27-30.
[13]. Kim, D., Langmead, B., &Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature methods, 12(4), 357-360.
[14]. Langmore, J. P. (2002). Rubicon Genomics, Inc. Pharmacogenomics, 3(4), 557-560.
[15]. Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323
[16]. Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., ... &McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 161(5), 1202-1214.
[17]. Novaes, E., Drost, D., Farmerie, W. G., Pappas, G. J., Grattapaglia, D., Ronald, R., &Sederoff, R. R. Kirst (2008) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. BMC Genom, 9, 312.
[18]. Ocone, A., Haghverdi, L., Mueller, N. S., & Theis, F. J. (2015). Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. Bioinformatics, 31(12), i89-i96.
[19]. Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295.
[20]. Picelli, S. (2017). Single-cell RNA-sequencing: the future of genome biology is now. RNA biology, 14(5), 637-650.
[21]. Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y. A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods 14, 309–315.
[22]. Ramsköld, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., ... & Sandberg, R. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature biotechnology, 30(8), 777-782.
[23]. Sudheer Menon (2020) "Preparation and computational analysis of Bisulphite sequencing in Germfree Mice" *International Journal for Science and Advance Research In Technology*, 6(9) PP (557-565).
[24]. Sudheer Menon, Shanmughavel Piramanayakam and Gopal Agarwal (2021) "Computational identification of promoter regions in prokaryotes and Eukaryotes" EPRA International Journal of Agriculture and Rural Economic Research (ARER), Vol (9) Issue (7) July 2021, PP (21-28).
[25]. Sudheer Menon (2021) "Bioinformatics approaches to understand gene looping in human genome" EPRA International Journal of Research & Development (IJRD), Vol (6) Issue (7) July 2021, PP (170-173).
[26]. Sudheer Menon (2021) "Insilico analysis of terpenoids in Saccharomyces Cerevisiae"international Journal of Engineering Applied Sciences and Technology, 2021 Vol. 6, Issue1, ISSN No. 2455-2143, PP(43-52).
[27]. Sudheer Menon (2021) "Computational analysis of Histone modification and TFBs that mediates gene looping" Bioinformatics, Pharmaceutical, and Chemical Sciences (RJLBPCS), June 2021, 7(3) PP (53-70).

[28]. Sudheer Menon Shanmughavel piramanayakam, Gopal Prasad Agarwal (2021) "FPMD-Fungal promoter motif database: A database for the Promoter motifs regions in fungal genomes" EPRA International Journal of Multidisciplinary research,7(7) PP (620-623).
[29]. Sudheer Menon, Shanmughavel Piramanayakam and Gopal Agarwal (2021) Computational Identification of promoter regions in fungal genomes, International Journal of Advance Research, Ideas and Innovations in Technology, 7(4) PP (908-914).
[30]. Sudheer Menon, Vincent Chi Hang Lui and Paul Kwong Hang Tam (2021) Bioinformatics methods for identifying hirschsprung disease genes, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 9 Issue VII July, PP (2974-2978).
[31]. Stubbington, M. J., Lönnberg, T., Proserpio, V., Clare, S., Speak, A. O., Dougan, G., &Teichmann, S. A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nature methods, 13(4), 329-332.
[32]. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., ... &Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nature methods, 6(5), 377-382.
[33]. Welch, J. D., Hu, Y., and Prins, J. F. (2016). Robust detection of alternative splicing in a population of single cells. Nucleic Acids Res. 44:e73.
[34]. Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., ... & Quake, S. R. (2014). Quantitative assessment of single-cell RNA-sequencing methods. Nature methods, 11(1), 41-46.
[35]. Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., ... & Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature, 500(7464), 593-597.
[36]. Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology, 4(1).
[37]. Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., ... &Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. Molecular cell, 65(4), 631-643.
[38]. Zong, C., Lu, S., Chapman, A. R., &Xie, X. S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science, 338(6114), 1622-1626.