Swayam Course - Analytical Techniques

Week 15, Tutorial 40 - Single Cell Biology

Content Writer - Dr Subhradip Karmakar, Assistant Professor, Department of Biochemistry, AIIMS, New Delhi

1. Introduction:

The cell is a natural unit of all biological organisms. In multicellular organisms, diverse cell types are derived from zygote through cellular proliferation and differentiation. Deregulation of proliferation and differentiation lead to several human diseases such as cancers, neurological disorders and developmental disorders. Genomics, epigenetics, transcriptomic and proteomic profiling of single cells from heterogeneous tissue environment provide information the development of complex disorders. Single-cell sequencing provides genotypic or phenotypic fingerprint which is not possible from tissue due to its heterogeneous natures. Sequencing of biological molecules such as RNA or methylated DNA in a single cell provide phenotypic and genotypic characteristic of cell. Single Cell Biology focuses on new technologies and their applications in medical and biological analyses at single-cell resolution and often at a genome-wide scale enabling a new understanding of complex biological phenomena. Single-cell biology is a new field combining multiple disciplines. The technologies developed in genomics, transcriptomics, epigenomics, proteomics, metabolomics and other fields are often used in this new discipline. Such multiomics approaches will provide detailed exploration of cellular behaviour in health and disease. Some tools in traditional physics, chemistry and novel methods in mathematics are also contributing to the ability to study biology at the single-cell level. Single Cell Biology is an integrative science with focus on areas like Cell Metabolism, Cell signalling, Cell Physiology, Stem Cell , Cancer Cell biology, Structural biology, Cell Movement, Cell Senescence, Cellular compartments etc.

2. Purpose to study Single Cell Biology

It was the convention for scientists to investigate large numbers of cells to gain significant insights into the many different cell populations that make up an organism. RNA sequencing (RNA-seq) of bulk tissue populations led to the discovery of subpopulations of cells within tissues, but it was impossible to discern the details of how these populations arise and how they interact with one another. Single-cell omics analyses is the solution to the problem. Similarly, it was easy to look at how genes are expressed within an entire tissue, but difficult

to find out if all cells were expressing these genes at the same level. Genomic studies rely on studying collective averages obtained from pooling thousands to millions of cells, precluding genome-wide analysis of cell to cell variability. Therefore, single cell sequencing was developed alongside. Its necessity in research made it "method of the year" by Nature Methods in 2013.

3. Why Single Cell Biology

Due to the heterogeneity seen in both eukaryotic and prokaryotic cell populations, analysing a single cell makes it possible to discover mechanisms not seen when studying a bulk population of cells. Sequences of single-cell genome provide information about cell lineage that reflects DNA mutations and transcriptomic alterations in the cells acquiring genetic alterations over time. Single-cell multiomics approaches can be used to identify genomic variations that drive cellular function and transcriptional alterations in cancer. These approaches are essential to understand cellular mechanisms that develop resistance against chemotherapy in cancer. Single-cell genome sequencing is used to establish DNA-based cell lineage trees and transcriptomic state of a cancer cell. This genomic–transcriptomic signature will inform about the development of cellular heterogeneity in cancer cell during treatment. Single-cell genomics also able to inform low-frequency aneuploidies to high-frequency copy number variants in neuronal cell in the brain.

4. Single Cell Biology: It's Scope

Single Cell Biology is a fairly new arena: It evolved primarily due to the following lacunae: Inconsistency to study the genetic evolution of cancer. Since cancer cells are constantly mutating it is of great interest to see how cancers evolve at the genetic level. Conventional cell biology fails to address this inter-cell variations. These patterns of somatic mutations and copy number aberration can be observed using single cell sequencing. A key problem that has emerged from recent cancer research has been how to deal with the enormous heterogeneity found among the millions of cells that make up an individual tumor. Scientists now know that not all tumor cells are the same, even within an individual, and that these cells diversify into subpopulations, each of which has unique properties, or phenotypes. Of particular interest are cancer stem cells, which are typically resistant to existing cancer therapies and lead to relapse and recurrence of cancer following treatment.

Humans are composed of approximately 3.72×10^{13} single cells that live harmoniously in tissues among their neighbours. Despite the complexity of tissues, most genomic studies to

date have focused on analysing bulk tissue samples which are composed of millions of cells. In these averaged data sets, it is difficult to resolve cell-to-cell variations and identify rare cells that may play an important role in disease progression. The recent development of single-cell sequencing (SCS) methods has led to a paradigm shift in the field of genomics, away from bulk tissue analysis and toward detailed and comprehensive studies of individual cells. Single-cell technology is attractive to biologists and doctors. Because most of the samples that we study, including cancers, neurons, immune cells, stem cells, and many other types of tissues are heterogeneous and with some stochastic nature. Single cells are free from such noises. Characterize cancer stem cells from other subpopulations of cancer cells has therefore become an important goal, for once these cells are identified, their vulnerabilities could be studied with the aim of developing better, long lasting cancer therapies. Creating smart interactive algorithm that can robustly identify such subpopulations in a completely automatic and unsupervised way, based purely on high-dimensional single-cell data. This new method makes it possible to discover many new cell subpopulations that we have never seen before.

5. Global Initiative:

The Center for Topology of Cancer Evolution and Heterogeneity at Columbia University develop and utilize innovative mathematical and experimental techniques to explore how genetic diversity emerges in the cells that make up solid tumors. In this way it will address a key challenge facing cancer research in the age of precision medicine — how to identify the clonal variants within a tumor that are responsible for its growth, spread, and resistance to therapy. Ultimately, the strategies could be used to identify more effective biomarkers of disease and new therapeutic modules.

6. Advancement of Technology in Single Cell Biology

Two complementary methods are applied to genome and transcriptome sequencing from single cells. In gDNA–mRNA sequencing (DR-seq) approach, Genomic DNA (gDNA) and mRNA of single cell lysate are sequenced by modified multiple annealing, looping-based amplification cycles (MALBAC), mRNA library preparation and subsequent sequencing. In DR-seq, cell lysate is used to synthesise single-stranded cDNA from RNA through reverse transcription by using 5’ T7 promoter. The gDNA and single-stranded cDNA are amplified together by using MALBAC. gDNA is further amplified by PCR and second-strand of cDNA is amplified by in vitro transcription (IVT).

In another approach called genome and transcriptome sequencing (G&T-seq) method, polyadenylated mRNA of fully lysed single cell is isolated and separated from gDNA using oligo-dT-coated magnetic beads. mRNA is converted to cDNA and Smart-seq2 protocol is used to amplify the cDNA. Different methods are used to amplify the gDNA. For Separation of genome and transcriptome from cell lysates, the intact nucleus is separated from the cytoplasmic lysate. For genomic and epigenomic analysis, the nucleus can be used as a substrate. To perform mRNA profiling of the single cell, the cytoplasmic lysate can be used. A microfluidic platform method is used to separates cytoplasmic mRNA from nuclear gDNA. G&T-seq was applied to breast cancer and pluripotent stem cell of neurons from patients with chromosome 21 trisomy.

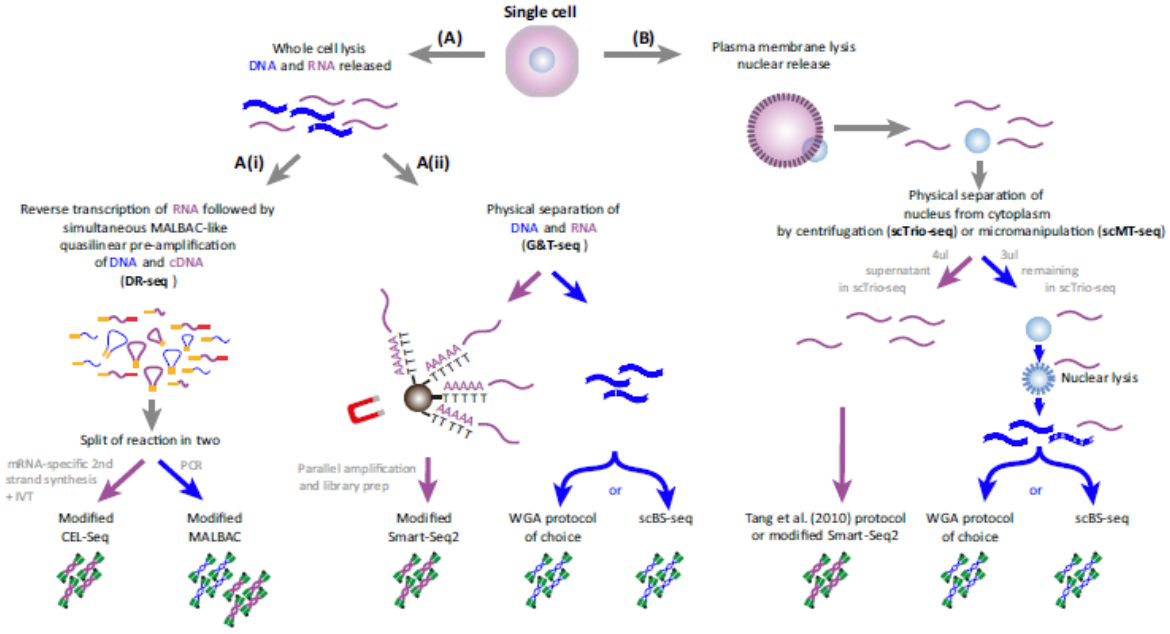

Figure 1: Protocol for DNA- and RNA-Sequencing of the Same Single Cell. (*Ref: Trends in Genetics, February 2017, Vol. 33, No. 2* *http://dx.doi.org/10.1016/j.tig.2016.12.003*)

Table: Multi-omics methods used in Single-Cell Sequencing:

Multi-omics method	Application
G&T-Seq	Transcriptional alterations of chromosomal aneuploidies and interchromosomal fusions
DR-Seq	copy number variations and gene expression alterations in among individual cells.

scMT-Seq	Identification of transcriptome and methylome alteration among single cells
scM&T-Seq	Identification of methylated distal regulatory elements and transcription of key pluripotency genes
scTrio-Seq	Identification of CNVs, DNA methylome, or transcriptome of individual cells
REAP-Seq (RNA expression and protein sequencing)	Protein expression and mRNA transcriptome identification. It is used for sub clustering of natural killer cells.
scNMT-Seq (single-cell nucleosome, methylation and transcription sequencing)	Identification of Nucleosome status, DNA methylome and mRNA transcriptome status

7. Single Cell Isolation Methods:

Several approaches such as mouth pipetting, serial dilution, robotic micromanipulation, flow-assisted cell sorting (FACS), and microfluidic platforms are used to isolate a single cell from an abundant population. Many commercial platforms are used to isolate circulating tumor cells (CTCs) from plasma of cancer patients. The CTCs are present in low frequencies (1 in 1 million). The FDA approved clinical system such as Cell-Search system uses ferrofluid nanoparticles conjugated to antibodies for EpCAM and CD45 to capture CTCs. Other technology such as nanopost microchip technology with EpCAM antibodies is used to isolate CTCs. Magsweeper (Illumina Inc.) is a technology which used to isolate CTCs by using a rotating magnet with EpCAM antibodies. The DEP-Array system (Silicon Biosciences) isolate single cell by using charge based microchip with dielectrophoretic cages. The CellCelector (Automated Lab Solutions) is used to isolate single cell by using robotic micromanipulation capillary system. Nanofilters are used to isolate rare cells by size exclusion method.

FACS or microfluidics platforms, such as the Fluidigm C1 robot or microdroplet methods are used to sort suspended cells into individual wells. FACS is used sort single cells into microtiter plates very fast. In Fluidigm C1 microfluidic robotic platform, cells are captured by using integrated fluidic circuits. The drawback of this method is that it sorts for cells

which are homogeneous in size. Another limitation is that the capture efficiency is low for sticky or non-spherical cells.

Table 1: List of Single-Cell Isolation Methods from abundant cell populations

Methodology	Principle	Advantage	Disadvantage
Flow sorting	Electric charge at high pressure	fluorescent markers are used to isolate subpopulations	Expensive
Robotic micromanipulation	Robotic-controlled micropipettes are used	Use of Fluorescence	Low throughput
Microfluid platforms	Single cells are used by flow channels of microfluidic chips	Chip based reactions are performed	Expensive
Serial dilution	One cell per microliter by serial dilution	Inexpensive	High probability of mixing multiple cells

Table 2: List of Single-Cell Isolation Methods from rare cell populations

Methodology	Principle	Advantage	Disadvantage
Laser-capture microdissection	Cells are isolated from a tissue on section slide by using lasers under a microscope	Preservation of Cellular protein	DNA/RNA will be damaged by UV

CellSearch	Magnetic nanoparticles conjugated to antibodies are used to separate cell on the basis of surface markers	Highly efficient	Bias towards cell surface marker
CellCelector	Robotic capillary based micromanipulator is used	Highly efficient	Expensive
MagSweeper	Rotating magnet coated EpCAM antibodies are used	Enrichment is very high	Bias towards cell surface marker
DEP-Array	Microchip with dielectropheretic cages	Highly sensitive	Time consuming
Nanofilters	Nanofabricated filters are used to separate cell on the basis size	Size-exclusion	Adherence of cells to filters

Flow-assisted cell sorting (FACS):

Advances have been made on the FACS technology including the instrumentation and the availability of a large number of highly specific antibodies and also with multiplexing. The capability of FACS technology has improved significantly from a technique limited to measuring 1–2 fluorescent species per cell to 10–15 species. The maximum number of proteins that can be simultaneously measured has progressively increased. Due to this progress, our understanding of immunology and stem cell biology has improved tremendously alongside the discovery of scores of functionally diverse cell populations. It has also been reported that using the next generation cytometry, “post-fluorescence” single cell technology termed mass cytometry is theoretically capable of measuring 70–100 parameters.

Laser Capture Microdissection (LCM):

Laser Capture Microdissection (LCM) is an advanced technology for isolating pure cell populations or a single cell from mostly solid tissue samples on a microscope slide. LCM is an automated sample preparation technique that enables isolation of specific cells from a mixed population under microscopic visualization. This technique of isolating a pure sample from a heterogeneous mixture allows for more efficient and accurate results with downstream micro-genomics applications such as next-generation sequencing, Sanger sequencing, PCR, and proteomics. It can accurately and efficiently target and capture the cells of interest to fully exploit emerging molecular analytical technologies, including PCR, microarrays and proteomics. There are two general classes of laser capture microdissection systems: infrared (IR LCM) and ultraviolet (UV LCM). The LCM system consists of an inverted microscope, a solid state near infrared laser diode, a laser control unit, a joy stick controlled microscope stage with a vacuum chuck for slide immobilization, a CCD camera, and a color monitor. The isolation and characterization of specific cells from a heterogeneous cell population is important for the analysis of gene expression because the differences between various cell types present within normal tissues have significant effects on gene expression analysis. LCM allows one to obtain answers in a single cell that may otherwise have been missed in a whole tissue scrape that contains a heterogeneous mixture of healthy, stromal and cancer cells. The LCM technique can be used to isolate just a few cells or even a single cell from tissue samples, blood, or even semen samples, depending on the application. The cells can be chosen on the basis of their morphology, immuno histochemical phenotype, or even their genotype, using in situ hybridization methods.

Magnetic-Activated Cell Sorting (MACS):

Magnetic-Activated Cell Sorting (MACS) is another commonly used passive separation technique to isolate different types of cells depending on their cluster of differentiation. It has been reported that MACS is capable of isolating specific cell populations with a purity >90% purification (Miltenyi et al., 1990). MACS is based on antibodies, enzymes, lectins, or strepavidins conjugated to magnetic beads to bind specific proteins on the target cells. When a mixed population of cells is placed in an external magnetic field, the magnetic beads will activate and the labeled cells will polarize while other cells are washed out. The remaining cells can be acquired by elution after the magnetic field is turned off. With this technique, the cells can be separated by charge with respect to the particular antigens. Positive separation techniques use coated magnetic beads and attract cells.

Manual cell picking/micromanipulation:

Manual cell picking is a simple, convenient, and efficient method for isolating single cells. Similar to LCM, manual cell picking micromanipulators also consists of an inverted microscope combined with micro-pipettes that are movable through motorized mechanical stages. Each isolated single cell can be observed and photographed under the microscope, thus enabling unbiased isolation. Unlike LCM that mainly isolates single cells from sections of fixed tissue, micromanipulation plays an important role in isolating live culture cells or embryo cells. Micromanipulation can be easily performed in an electrophysiology lab equipped with a patch clamp system. For example, after investigating neuronal function in brain slices preparations after standard whole-cell patch-clamp electrophysiological recordings, scientists would apply negative pressure through the patch pipette so that the cytosolic material containing cellular mRNA can be aspirated for further analysis. However, the throughput is limited and it requires highly skilled professionals to perform, it has the utility limitation when detecting complex changes.

Microfluidics:

Microfluidics is recognized as a powerful technology for investigating the inherent complexity of cellular systems. It provides precise fluid control, low sample consumption, device miniaturization, low analysis cost, and easy handling of nanoliters-volumes. Cell Sorting by a microfluidic chip can be divided into four categories: cell-affinity chromatography based microfluidic, physical characteristics of cell based microfluidic separation, immunomagnetic beads based microfluidic separation, and separation methods based on differences between dielectric properties of various cell types. Microfluidics can be combined with numerous separation methods, such as filtration and sedimentation or affinity-based technologies, FACS and MACS. In the recent years, numerous investigations and applications in microfluidic devices have been reported, including cancer research, microbiology, single-cell analysis, stem cell research, drug discovery, and screening

8. Single Cell Analysis: Challenges

The recent advent of highly parallelizable single-cell RNA sequencing (RNA Seq) technologies has opened a new window into the study of cell differentiation, commitment, and diversity. Rapid advances in the development of these technologies are being accompanied by the design of computational methods tailored to address the challenges presented by the analysis of single-cell RNA-sequencing data. While many methods have

been successfully used for the analysis of genomic data from bulk samples, the relatively small number of sequencing reads, the sparsity of data, and cell population heterogeneity present significant analytical challenges in effective data analysis. Recent advances in computational biology have greatly enhanced the quality of data analyses and provided important new biological insights in this field.

9. Getting Information out of Single Cell Biology Experiments

a. Single Cell Sequencing (DNA AND RNA):

During the last decade, high-throughput sequencing methods have revolutionized the entire field of biology. The opportunity to study entire transcriptomes in great detail using RNA sequencing (RNA-seq) has fuelled many important discoveries and is now a routine method in biomedical research. However, RNA-seq is typically performed in "bulk," and the data represent an average of gene expression patterns across thousands to millions of cells; this might obscure biologically relevant differences between cells. Single-cell RNA-seq (scRNA-seq) represents an approach to overcome this problem. By isolating single cells, capturing their transcripts, and generating sequencing libraries in which the transcripts are mapped to individual cells, scRNA-seq allows assessment of fundamental biological properties of cell populations and biological systems at unprecedented resolution. Single-cell sequencing (SCS) has emerged as a powerful new set of technologies for studying rare cells and delineating complex populations. Over the past few years, SCS methods for DNA and RNA have had a broad impact on many diverse fields of biology, including microbiology, neurobiology, development, and tissue mosaicism, immunology, and cancer research. Single neuron RNA sequencing of hippocampus and neocortical neurons of mouse embryo was used to determine transcriptomic signatures that drive embryonic development of brain. Single-cell RNA-seq was used to identify cell-to-cell transcriptional variation in hippocampal neurons. scRNA sequencing in single cells of brain was used to identify Notch signalling that play key role in brain development. RNA SCS of embryonic stem cells was used to identify 1,000 heterogeneous transcripts in same blastomere. RNA SCS was used to determine transcriptomic signatures that drive from oocyte to morula development in human and mouse embryos. RNA SCS was also addressed the gene expression patterns of single cells in development of kidney in mice. Transcriptomic profiling of 4,000 single spleen cells from

mouse in response to antigen activation with LPS by RNA SCS determined alterations of 1,575 genes. RNA SCS has been applied to study the mechanism of intratumor heterogeneity, clonal evolution, and metastatic dissemination in human cancers. RNA SCS was used to identify the transcriptomic profiling of CTCs in the blood of melanoma and metastatic breast cancer patients. DNA SCS in CTCs from patients with metastatic colon cancer was to identify the driver gene mutations which were also reported in primary colorectal tumor. Single cell exome sequencing and copy number profiling of CTCs from metastatic lung adenocarcinoma and metastatic prostate cancer patients were performed to identify mutations by MALBAC. SCS methods are used for early detection of tumor cells in bodily fluids such as urine, sputum, blood and fine-needle-aspirates samples. scRNA-seq are used to characterize the rare residual cells that survive after treatments and determine mechanism of tumor recurrence. scRNA-Seq is used to address the genetic alterations in rare tumor subpopulations that drive recurrent tumor. scRNA-seq is used to characterize intra-tumoral heterogeneity and to classify tumor subpopulation. It can also detect variation among cell states within homogenous cell population such as transcriptomic alterations of same cell type in different cell cycle stage.

b. Single Cell Analysis: RNA, DNA and PROTEINS

a) Single cell analysis tools can be divided into three groups: genomics, transcriptomics, and proteomics.
b) Due to next generation sequencing (NGS) technologies as well as whole genome/transcriptome amplification (WGA/WTA) approaches, a new scientific field of single cell genome studies have been established.
c) A combination of high-throughput and multiparameter approaches is used in single cell analysis which can reflect cell to cell variability and heterogeneous differences in the individual cells.
d) Transcriptome analysis focuses on the profiling of the complete set of RNA molecules in a given experiment and is mainly used to study gene expression.
e) The three pre-dominating technologies in the field are quantitative reverse transcription polymerase chain reaction (qRT-PCR), microarrays and RNA sequencing (RNA-seq), have now all been extended to single cell applications.

10. Single Cell RNA Analysis: scRNA-Seq

a) Single-cell RNA sequencing has emerged as an indispensable tool to dissect the cellular heterogeneity and decompose tissues into cell types and/or cell states, which offers enormous potential for de novo discovery.
b) Single-cell transcriptomic maps provide great resolution to reveal complex cellular events
c) scRNA-seq methods are composed of four steps: isolation and lysis of single cells or single nuclei, reverse transcription, cDNA amplification and sequencing library preparation.
d) The first three steps are often finished in one-tube reactions to minimize loss of material. At this time, the last step is commonly conducted using transposase Tn5-based fragmentation for library preparation.
e) Traditional microarray and bulk RNA-sequencing (RNA-seq) technologies profile the average gene expression level of all cells in the population.
f) In contrast, recent single-cell RNA-seq (scRNA-seq) methods enable the quantification of a much richer set of properties of the gene expression distribution across cells.

11. Summary of scRNA-Seq methods:

In single-cell RNA sequencing methods, RNA is amplified by whole-transcriptome amplification (WTA). In mammalian cell, 10 picograms of total RNA and 0.1 picograms of mRNA are present. In WTA methods, T7 RNA polymerase is used to amplify cDNA linearly by in vitro transcription (IVT) method. Oligo d(T) primers conjugated to adaptor sequences are used for reverse transcription. SMART-Seq (switching mechanism at 5 end of RNA template) is used to amplify the full length mRNA transcripts by using a Moloney Murine Leukemia Virus (MMLV) reverse transcriptase which has both template-switching and terminal transferase activity. This results in the addition of non-templated cytosine residues at the 5’ end of the cDNA. Single-cell RNA sequencing results in amplification bias. To reduce the technical bias, unique molecular indexes (UMIs) are used to label the original pool of RNA molecules prior to WTA. Barcoded cDNA libraries are prepared by WTA and multiplexed NGS are performed by using pooled libraries. Different approaches such as poly (A) tailing in the Tang protocol and QuartzSeq or template-switching mechanism are used for second-strand synthesis in the reverse transcription reaction.

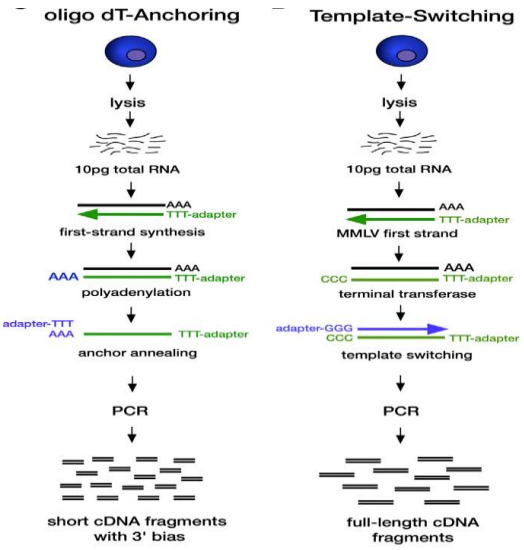

Figure 2: WTA Methods for Single-Cell RNA Sequencing (Ref. No. Molecular Cell 58, May 21, 2015; http://dx.doi.org/10.1016/j.molcel.2015.05.005)

In Tang method, manually picked cells under the microscope are lysed and polyadenylated RNA was reverse transcribed into 3 kb cDNA by using an oligo-dT primer carrying a specific anchor sequence (UP1). A poly (A) tail was added to the 3' -end of the first strand cDNA by using a terminal deoxynucleotidyl transferase. The second cDNA strand was synthesized by using a complementary poly (T) primer. Double-stranded cDNA was sheared and adaptor ligated libraries were then sequenced using a SOLiD sequencer (Applied Biosystems).

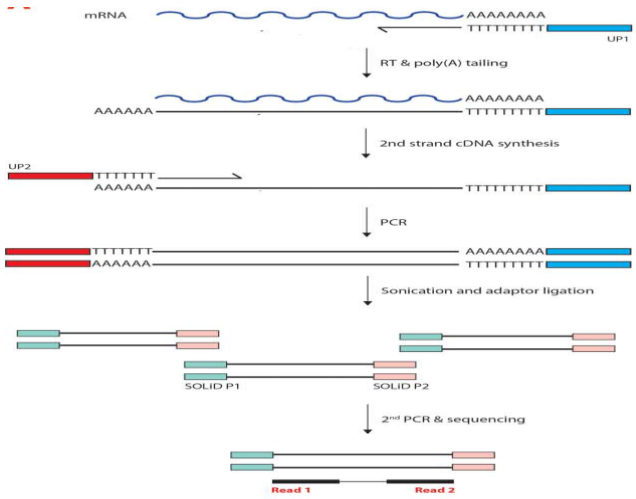

Figure: Schematic representation of Tang method. (Ref. No. RNA BIOLOGY 2017, VOL. 14, NO. 5, 637–650. http://dx.doi.org/10.1080/15476286.2016.1201618)

In the CEL-seq (Cell Expression by Linear amplification and Sequencing) method, RT reaction was performed by using a primer carrying a T7 promoter sequence at the 5'-end,

unique barcode attached adaptor sequence and an anchored poly(T) at the 3'-end. IVT reaction was performed after RT-PCR and second strand cDNA synthesis. The antisense RNA (aRNA) was fragmented, adaptor was ligated and reverse transcribed to DNA. Paired-end sequencing was performed. CEL-seq has less chance to identify of lowly expressed transcripts. Introduction of a Unique Molecular Identifier (UMI) in the oligo-dT primer becomes able to count unique RNA molecules. In Single-cell Tagged Reverse Transcription Sequencing (STRT-seq), anchored oligodT primers and a MMLV-based enzyme are used for reverse transcription. Biotinylated Template Switching Oligonucleotide (TSO) which has 6-base random barcode just upstream of the 3' end riboguanosines used in STRT-seq. Two uracils residues included in the TSO are used to block the oligonucleotide from cross-reacting as primer during the PCR. CytoSeq is used to determine the expression profiles of large and heterogeneous cell populations and identify the rare cell types in a large background population.

PCR or in vitro transcription (IVT) are used to amplify the cDNA after reverse transcription. These methods result in biases. In SmartSeq2, STRT, Tang protocol, and SC3-seq, PCR is sequence dependent nonlinear amplification. IVT in CEL-Seq (cell expression by linear amplification and sequencing) and MARSSeq (massively parallel RNA single-cell sequencing) is linear amplification which requires an additional round of reverse transcription of the amplified RNA. SmartSeq and SmartSeq2 are used to identify the alternative splicing, allele-specific expression (ASE) and novel exons by using template switching along with amplification of full-length cDNA transcript. SC3-seq was used in sequencing at SOLiD system.

12. Challenges of scRNA-seq:

Like any other technique, there are some challenges that we need to keep in mind when performing this method, and specially, when analysing the data provided by it. As we know, RNA has many functions in our cells, and there are different RNAs doing different functions. It might be regulatory functions. Either way, RNA losses could cause us to miss a low abundant transcript altogether! And, it is still very difficult to differentiate between technical noise and low abundant transcript. Therefore, it is essential to improve sensitivity. So, we can see all different RNA and be confident of its abundance in cell, because it gives us great information about the small regulatory (and often low abundance) RNA. Single-cell RNA-seq protocols are complex, involving multiple steps each contributing to the substantially increased noise level of scRNA-seq relative to bulk RNA-seq. Unique molecular identifiers

(UMI) were introduced as a barcoding technique to reduce amplification noise, but the observed expression distribution computed from observed UMI counts is, for most genes, still a poor representation of their true expression distribution. scRNA-seq enables the quantification of each gene's expression distribution across cells, thus allowing the assessment of the dispersion, nonzero fraction, and other aspects of its distribution beyond the mean.

13. Work Flow Over-View of scRNA-Seq

Single-cell RNA-seq has made great strides and become widely available and preferred method for high-throughput single-cell measurements. These measurements are very useful and their usefulness will continue to grow as we invent new ways to think about these data and reduce their noise. Yet, measuring transcript levels alone is insufficient for studying and understating many physiological and pathological processes, not least because the changes of protein levels human across tissues and cell differentiation are poorly predicted by the corresponding changes in mRNA levels.

14. Computational Methods for scRNA-Seq

Multiple technical artefacts such as cell capture, library preparation, and sequencing procedures effect on scRNA-seq. External spike-in RNA molecules (ERCC RNA spike-in mix) are used to quantify technical variability. Use of external RNA spike-ins have several drawbacks such as over- or underrepresentation of spike-in molecules in the sequencing library, inefficient cell lysis, and biasness of coverage. Heterogeneous gene expression pattern will be observed in homogeneous cell population during cell cycle or circadian rhythm. In T cells, housekeeping genes showed homogenous expression pattern than secreted proteins. Cellular transitions can be analysed by hierarchical clustering or PCA-like methods. Such as PCA, self-organizing maps (SOMs) and hierarchical clustering are used to identify different lung progenitors cells in scRNA-seq data.

15. Single Cell DNA Analysis: scDNA-Seq

With single-cell DNA sequencing, the genomic heterogeneity of cell populations can be explored at the level of the individual cell. Genetic changes, such as point mutations and copy number variation occurring during disease and normal development processes, are profiled using the minute amounts of DNA from single cells. Applications include analysis of genetic heterogeneity within unicellular and multicellular organisms, detection of chromosomal anomalies in germ line cells, preimplantation genomic screening of embryos, and defining the

genetic composition of tumors for developing more targeted therapies. DNA single-cell sequencing (SCS) becomes more challenging than RNA SCS because two copies of each DNA molecule are present in single cell whereas thousands of copies of RNA molecules are present in single cell. Technical difficulties such as nonuniformity of coverage, allelic dropout (ADO) events, false-positive (FP) errors, and false-negative (FN) errors are reported due to whole-genome-amplification (WGA) from limited amount of input material. Commonly used WGA methods are degenerative-oligonucleotide-PCR (DOP-PCR) and multiple-displacement-amplification (MDA). DOP-PCR accurately determines copy number whereas it generates low physical coverage. DOP-PCR in combination with flow-sorting of nuclei and NGS is used to determine high-resolution copy number profiles from single mammalian cells. Due to low physical coverage, DOP-PCR showed poor performance to detect mutations. To determine high physical coverage, Phi29 or Bst polymerases is used in MDA. MDA becomes unable to determine DNA copy number accurately due to nonuniform coverage or read depth. In another method such as multiple annealing- and looping-based amplification cycles (MALBAC), circular DNA fragment formation by Bst polymerase and adaptor ligation PCR are used to determine copy number and single nucleotide variants (SNVs) with extremely high FP error rates. Single nucleus exome sequencing (SNES) is used to reduce technical error rates which are generated during single-cell sequencing of exomes and genomes. After WGA, libraries are prepared from the amplified DNA and NGS is performed at Illumina platform. To reduce the cost, barcoded single-cell libraries are pooled together for multiplexed sequencing. Barcoded libraries are also applied to capture exome sequencing with higher coverage depth.

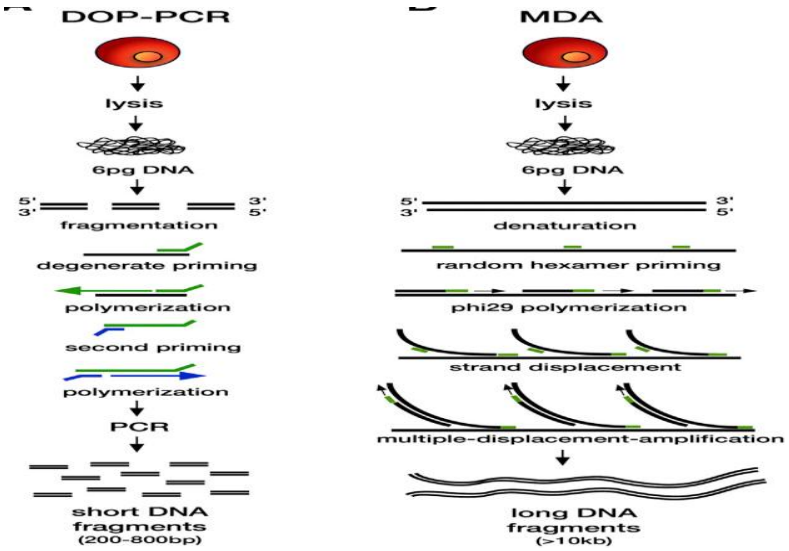

Figure 3: WGA Methods for Single-Cell DNA Sequencing (Ref: Molecular Cell 58, May 21, 2015; http://dx.doi.org/10.1016/j.molcel.2015.05.005)

Single-cell DNA sequencing of single sperm cells are used to determine 22.8 recombination events, 5–15 gene conversion events, and 25–36 de novo mutations in each sperm cell. MALBAC is used to identify 26 recombination events per single sperm cell of an Asian individual and 43 crossover events per oocyte from 8 individual females. SCS determined that rate of aneuploidy and recombination was higher in oocytes compared to sperm. MALBAC was used to determine copy number changes and point mutations in polar bodies prior to implantation to avoid the genetic transmission of diseases in IVF.

16. Single-Cell Epigenomic Sequencing Methods

Technical challenges become more in Epigenomic profiling of single cells. In epigenomic sequencing, bisulfide treatment of DNA is performed prior to sequencing. DNA modifications by bisulfide treatment becomes unable to amplify with DNA polymerases. Reduced single-cell representation bisulfite sequencing (scRRBS) was used to measure cytosine methylation at 1.5 million CpG sites in a single cell. Single cell bisulfite sequencing was used to measure cytosine modifications in DNA of mouse embryonic stem cells which showed that global demethylation occurred in embryonic development. Single-cell bisulfite sequencing (scBS-Seq) along with scRNA-Seq in hepatocellular carcinoma was used to identify differential methylation within two subpopulations. scRRBS showed that after fertilisation demethylation rate is faster in the paternal pronucleus than in the maternal one.

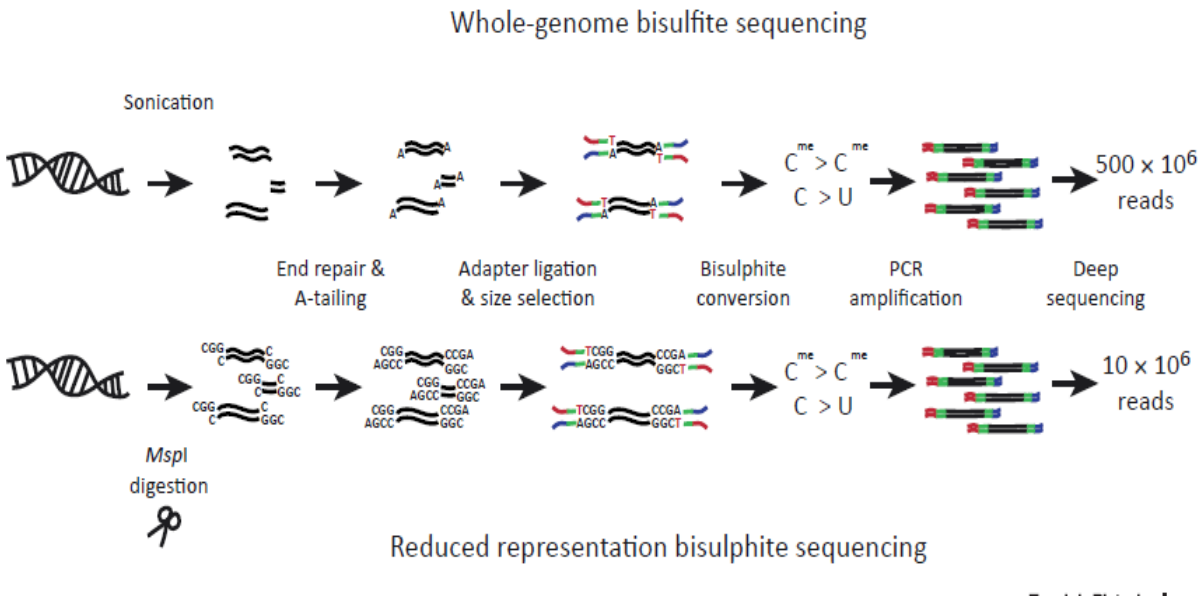

Figure: Schematic Comparison of Whole-Genome Bisulfite Sequencing and Reduced Representation Bisulfite Sequencing (*Ref: Trends in Biotechnology, September 2018, Vol. 36, No. 9 https://doi.org/10.1016/j.tibtech.2018.04.002*)

Combinatorial approach of scBS-Seq and scDNA-seq was used to copy number variants (CNVs). Combined approach of single-cell ATAC-Seq and ChIP-Seq were used to measure the genome-wide landscape of chromatin accessibility at cis-regulatory elements and trans-acting factors which determine functional status of chromatin at a single cell. Single-cell combinatorial indexing-based method (sciCAR) was used to identify cis-regulatory sites to their target genes. Single-cell bisulfite sequencing determines CpGs methylation status. In reduced-representation bisulfite sequencing (RRBS), to reduce DNA loss entire protocol was performed in single tube. Post-bisulfite adaptor tagging (PBAT) showed higher recovery rates and coverage rate for methylation status at CpGs. RRBS is an enrichment-based bisulfite sequencing method in which methylation-insensitive restriction enzyme (MspI) digested template DNA is used for library construction. Methylation-sensitive restriction enzymes and single-cell quantitative PCR (qPCR) are collectively used in single-cell restriction analysis of methylation (SCRAM) approach. Methylation-sensitive restriction enzymes (MSREs) coupled to PCR amplification is used to perform DNA methylation analysis of single cells.

Table. List of Bisulfite based Single-Cell DNA Methylation Profiling Methods

Method	Application
scRRBS	genome-wide method for single-cell

	DNA methylation analysis
scBS-seq	adaptor ligation precedes bisulfite treatment
scWGBS	Coverage based on sequencing depth
snmC-seq	4.7%–5.7% coverage of the entire genome
scMAB-seq	Bisulfite based with M.SssI treatment for detection of 5-formylcytosine (5fC) and 5-Carboxylcytosine (5-caC)

Table. List of Multiomics based Single-Cell DNA Methylation Profiling Methods

Method	Application
scM&T-seq	Combined with RNA-seq
scTrio-seq	Combined with RNA-seq and CNV analysis
scCOOL-seq	Combined with analysis of chromatin state, nucleosome positioning, CNVs, and ploidy
scNMT-seq	Combined with RNA-seq and chromatin accessibility

Single-cell epigenomics can also be used for studying correlations between the epigenetic state of an enhancer and the promoter of it is targeting through long-range looping.

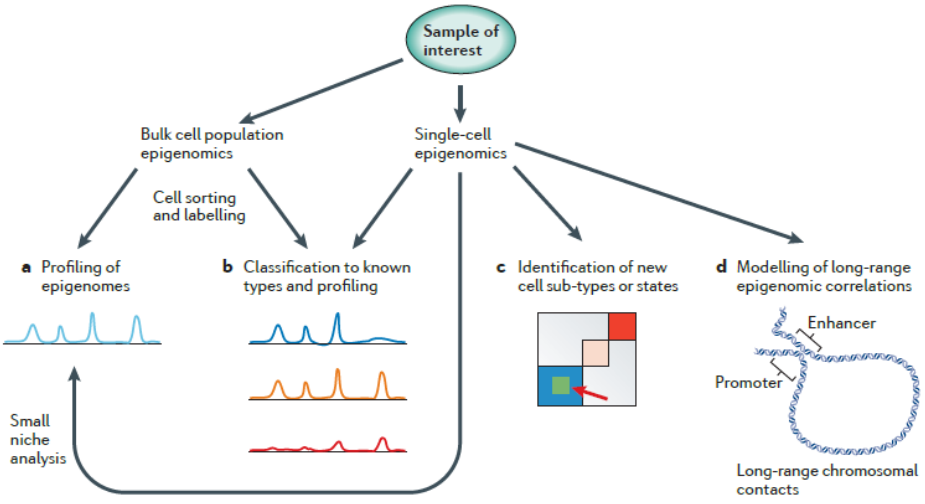

Figure: Schematic representation of applications of single-cell epigenomics (Ref.. *Nature Reviews Genetics | AOP, published online 13 October 2015; doi:10.1038/nrg3980*)

17. Single Cell Proteomic Analysis

As all methods using tandem mass tags, MS measurements are affected by coisolation interferences, which means that about 5 – 10 % of the reporter ion signal for a typical peptide

comes from other peptides. This undesirable contribution can be reduced by using newer instruments with better mass-filters that allow for smaller ion isolation windows. It can also be reduced by simply filtering out peptides with more co-isolation and focusing on those with very limited coisolation or by computationally compensating for it. Single-cell mass-cytometry (also known as CyTOF) with scRNA-Seq was used to characterize tumor cells, adjacent normal tissue, and blood from early lung adenocarcinoma patients to study the functional importance of tumor-infiltrating myeloid cells in anti-tumor immunity. Single cell proteomic tools are used to measure large number of parameters across thousands of single cells at a given time point or handful of parameters in the same cells over time. Single-cell proteomics assays are classified into three categories qualitative methods such as ELISPOT which will qualitatively identify cells that express a given proteins; semi-quantitative methods such as flow/mass cytometry, image cytometry, single cell western blot that measure abundance of proteins and quantitative methods such as SiMoA, Microengraving chip, SCBC that determine protein concentration. The enzyme-linked immunospot (ELISPOT) assay is used for qualitative assessment of cytokine at single cell level. Cytokines are captured by the immobilized antibodies and detected by secondary antibody with enzyme amplification for signal readout. Fluorescence flow cytometry (FFC) is used for analysis of single cell protein. Fluorophore-labeled antibodies are used to analyse primarily, membrane and cytoplasmic proteins in single cells at moderate level. Both FFC and Mass cytometry (CyTOF) are used to measure secreted cytokines. Luminex xMAP (Multi-Analyte Profiling) platform is a bead-based ELISA-like assay which is used to detect multiple proteins in small sample volume. Single molecule array (SiMoA) is used to capture small amount of proteins by using antibody-coated beads.

Table: Characteristics of Single Cell Proteomics Methods

Single Cell Proteomics Method	Characteristics
Fluorescence Activated Cell Sorting (FACS)	cell sorting based on membrane protein cell surface markers by using fluorophore labeled antibodies; colorimetric based method
Mass Cytometry (CyTOF)	More than 30 cytoplasmic proteins assayed per cell, Multiplexing through mass

	Spectrometry, mass tag labeled antibodies are used to stain fixed cells.
Single Cell Barcode Chips (SCBCs)	Spatially-encoded antibody array for fluorescent immunoassays of secreted cytoplasmic protein and integrated protein proteins in lysed cells.
MicroEngraving	Fluorescent based colorimetric immunoassays of secreted proteins.
Single Cell Western Blotting (scWestern)	Identification of cytoplasmic proteins from lysed cells on a microchip by automated Western Blotting

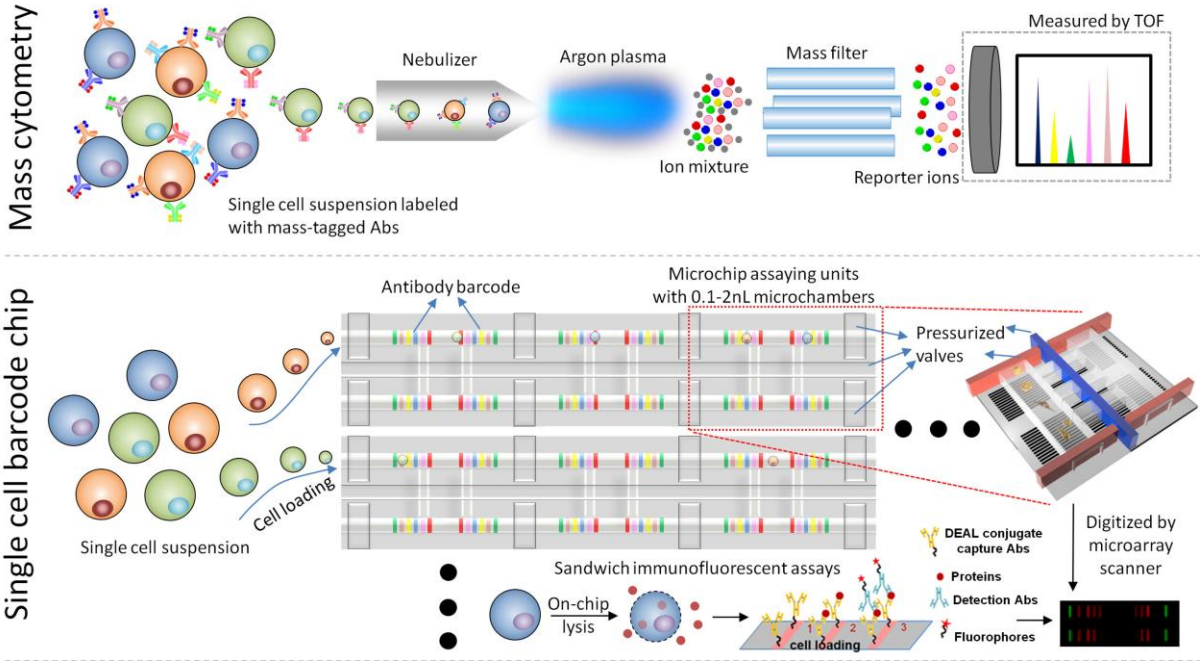

Figure: Schematic representation of multiplex single cell proteomic platforms (*Ref: Proteomics. 2017 February ; 17(3-4): . doi:10.1002/pmic.201600267*).

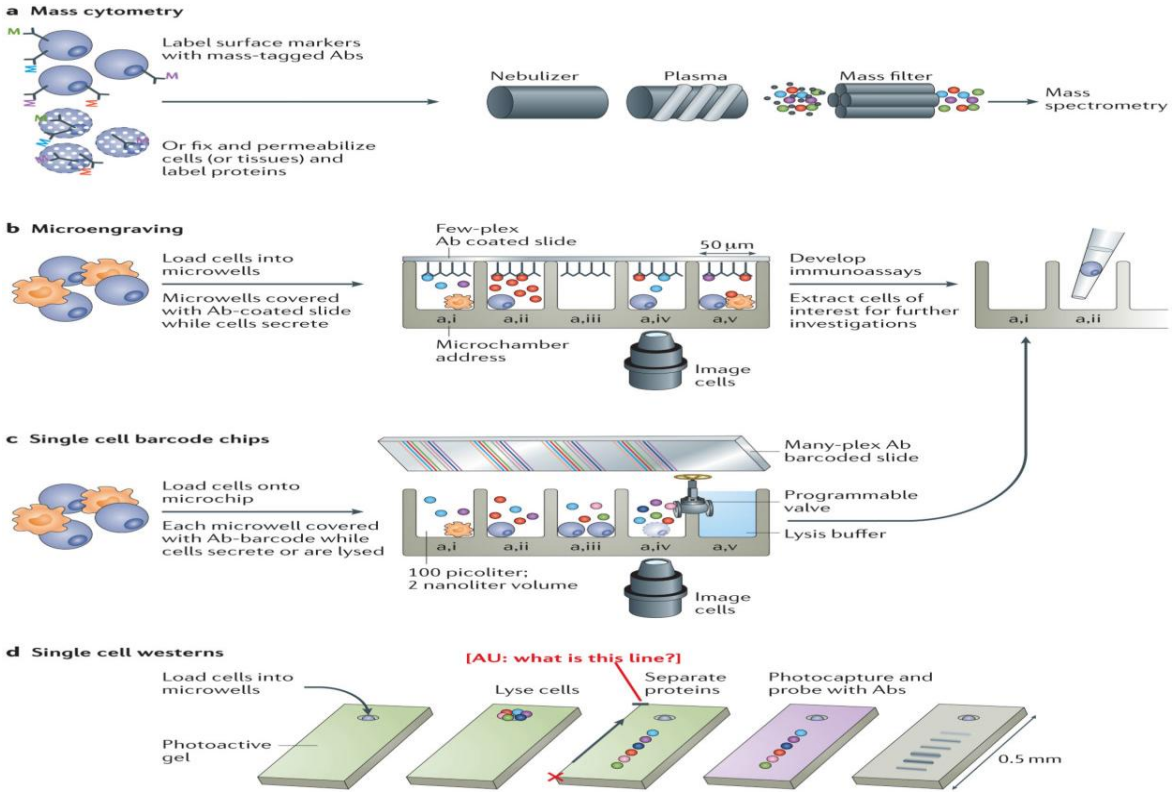

Figure: Schematic representation of single cell proteomics methods *(Ref: Nat Rev Drug Discov. 2016 March ; 15(3): 204–216. doi:10.1038/nrd.2015.16.)*

18. Single Cell Metabolomics:

The dynamics of a cell is always changing. Cells move, divide, communicate, adapt, and are always reacting to their surroundings non-synchronously. Currently, single-cell metabolomics has become the leading field in understanding the phenotypical variations between them, but sample volumes, low analyte concentrations, and validating gentle sample techniques have proven great barriers toward achieving accurate and complete metabolomics profiling. Certainly, advanced technologies such as nanodevices and microfluidic arrays are making great progress, and analytical techniques, such as matrix-assisted laser desorption ionization (MALDI), are gaining popularity with high-throughput methodology. Nevertheless, live single-cell mass spectrometry (LCSMS) values the sample quality and precision, turning once theoretical speculation into present-day applications in a variety of fields, including those of medicine, pharmaceutical, and agricultural industries. While there is still room for much improvement, it is clear that the metabolomics field is progressing toward analysis and discoveries at the single-cell level.

Horseradish peroxidase-templated gold nano-clusters (HRP-AuNCs) are used to detect hydrogen peroxide (H_2O_2) secreted by single cells in droplets. Microfluidic double emulsion-

based FACS (MDE-FACS) approach is used to screen single cells based on their secretomes. In this method, single cells are compartmentalized in emulsions along with specific fluorescence-generating machinery. A droplet-based method is used to detect single circulating tumor cells (CTCs) based on the single-cell metabolism. CTCs in droplets are detected on the basis of principle of Warburg effect which defines the secretion of lactate during glycolysis-based metabolism from CTC by using ratiometric dye. Chip-ESI (electrospray ionization) -MS is used to determine the level of metabolites from breast cancer cells. MALDI-MS has high sensitivity for metabolomic analysis. Microarrays for a mass spectrometry (MAMS) platform is used to analyse thousands of individual cells in a single MS experiment. MAMS chip features arrays are omniphobic material surrounded hydrophilic wells and are used to determine phospholipids in single human epithelial cells (A549).

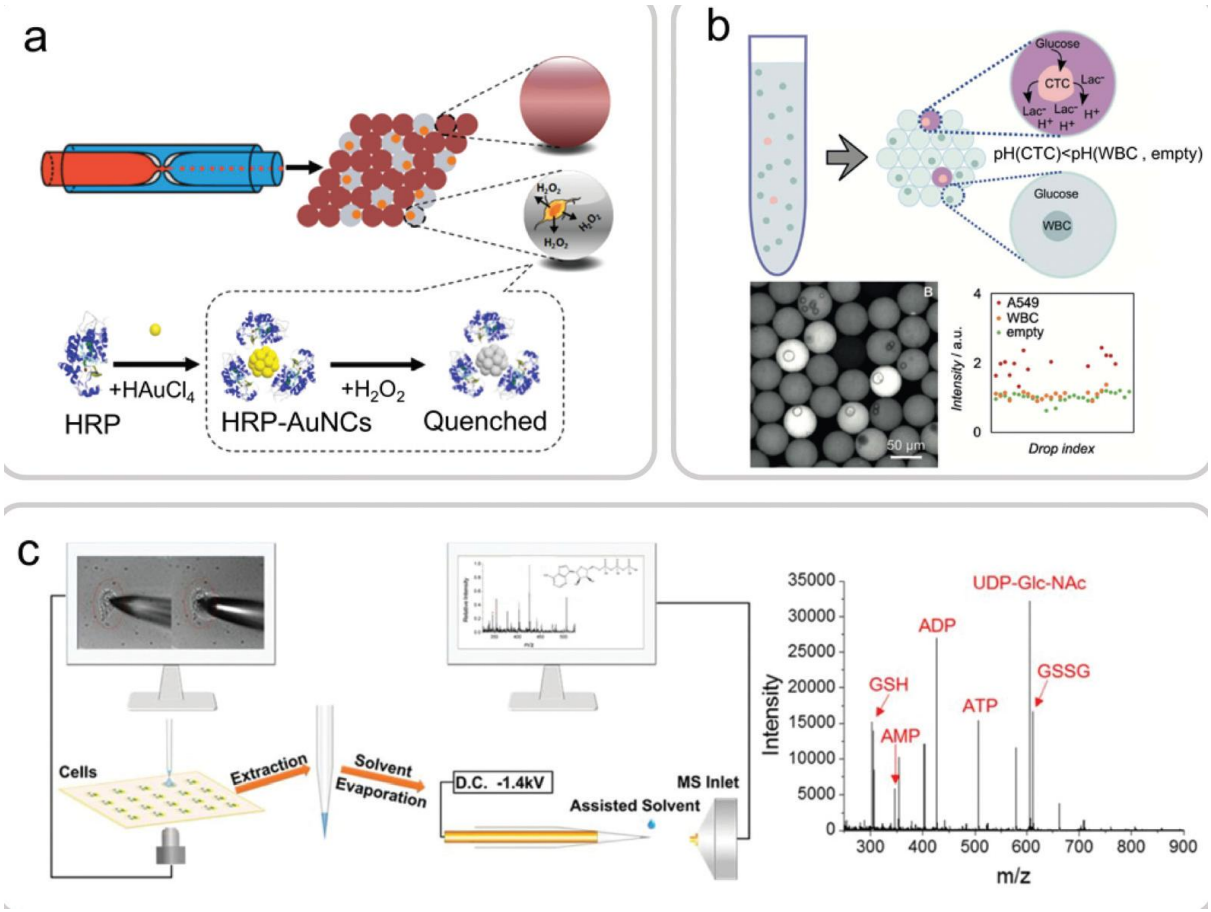

Figure: Schematic representation of Single-cell metabolomics (*Ref: DOI: 10.1039/c8an01503a*) Note: a: Detecting single-cell-secreted hydrogen peroxide in droplets with Au nanoclusters b: single-cell metabolism measurement in droplets c: integrated droplet-microextraction-ESI-MS method

19. Summary :

Interpreting population-average protein levels is fundamentally confounded when samples consist of heterogeneous cells. The most obvious caveat is that the population-average may not be representative for any cell. For example, proteoforms may have bimodally-distributed abundances within the whole heterogeneous cell population. Consider, for example, pancreatic biopsies of two patients containing both α and β cells. Even if the levels of a protein are unimodally-distributed within each cell type, biopsies that have slightly-biased representations of each cell type will yield different population-average measurements. This occurs even if those measurements are perfect. Thus, a population-average measurement might misleadingly suggest differential protein expression when in fact there is no difference whatsoever between the two patients.

The human body is comprised of an estimated 100 trillion cells, each differentiated from a single cell. But, the number of distinct cell types, the molecular basis of their function, and the aggregate functional role in a tissue or an organ are unknown. Mapping the cellular phenotype, like mapping the human genome, is the next frontier in understanding the cellular basis of organism function. A cellular foundation of organism function requires intimate quantitative knowledge of molecular components of single cells and the assembly of the constituent parts in the context of how they interact with one another. Such studies entail developing methods for single cell analysis including, 1) quantitative phenotypic characterization, 2) high throughput individual cell culturing and manipulation, and 3) quantitative model analysis of phenotype. Data generated from the activities of this program will provide insight into functional system development and maintenance while providing means for directed manipulation of cellular systems. At the heart of solving many grand challenges in biomedicine is the true integration of a systems level and synthetically driven manipulation of living cells, which will also be therapeutically relevant. A revolution is underway, unleashed by the study of single cells, in our understanding of cell identity, diversity, development, and function. The emerging technologies for single-cell “omics” for the study of genomes, epigenomes, transcriptomes, and proteomes, are becoming increasingly widespread, with some of the most exciting advances enabling multimodal analyses and systems biology assays that perturb and record at single-cell resolution. Perhaps even more far reaching are single-cell in situ sequencing and imaging-based methods that offer an unparalleled perspective on the biology of individual cells within tissues. The application of

this knowledge for precision medicine is also not far from reality, as one can easily imagine a time when a standard blood test or biopsy will include single-cell profiling to aid in diagnosis or to monitor treatment. More generally, trends within groups – within different cell types, for instance may disappear or even reverse when these groups are combined, as with population-average measurements. This phenomenon is known as Simpson's paradox. Similarly, Simpson's paradox can confound the interpretation of population-average protein levels. If we only consider the average levels of the ith and the jth proteins across cell types, the proteins seem positively correlated. However, paradoxically, within each cell type the abundances of the ith and the jth proteins can have an inverse relationship. Perfect measurements of cell types sorted based on a few markers cannot resolve such phenomena; the true relationship can be observed only by measuring proteins in single cell