# Multiple Platforms for Single Cell Genomics to Enable Biomarker Discovery in Immunotherapy

Ankita Das[1], Kayla Lee[1], Derek Vargas[1], Gavin Washburn[1], Niyati Thosani[1], Vasumathi Kode[1], Nitin Mandloi[2], Jing Wang[1], Amit Chaudhuri[1,2] and Papia Chakraborty[1]

1) MedGenome Inc, Foster City, CA, 2) MedGenome Labs, Bangalore, India; corresponding author : ankita.d@medgenome.com

## Abstract

Single cell genomic approaches can provide valuable insights into the complexity and heterogeneity of the cell types in the context of a tissue or tumor. However, challenges with the preparation of single cell suspensions, good cell viability and efficiently capturing diverse cell types in a mix via appropriate cell capture methods can override the utility of the approaches. We provide validation and application data generated from a range of single cell input types (number of starting cells, viability, and research question) and diverse commercially available platforms (the 10x Genomics Chromium and the FACS based SMART-Seq (Takara Bio). We demonstrate the utility of using the appropriate single cell genomics approach to get relevant information. Recent studies have shown that capturing additional information on cellular phenotypes or features can provide valuable information on cell identities otherwise missed and additional heterogeneity, which in turn facilitates discovery of meaningful biomarkers and understanding of molecular mechanisms of development and disease. To enable for such discovery, we have validated the sC-ATAC seq and CITE-Seq approaches and present data on the utility of those approaches and will present data to highlight the capabilities at MedGenome.

- We present a platform agnostic and flexible approach to utilize single cell profiling of gene expression depending on the research question and the number of cells available.
- We present single cell epigenomic profiling data on immune cells using single cell ATAC seq and show that examining chromatin accessibility can identify cellular identities and mechanisms of response to stimuli.
- We conclude the multi-omic approaches have been validated at MedGenome and can offer it as a service

## Workflow of All Services for MedGenome

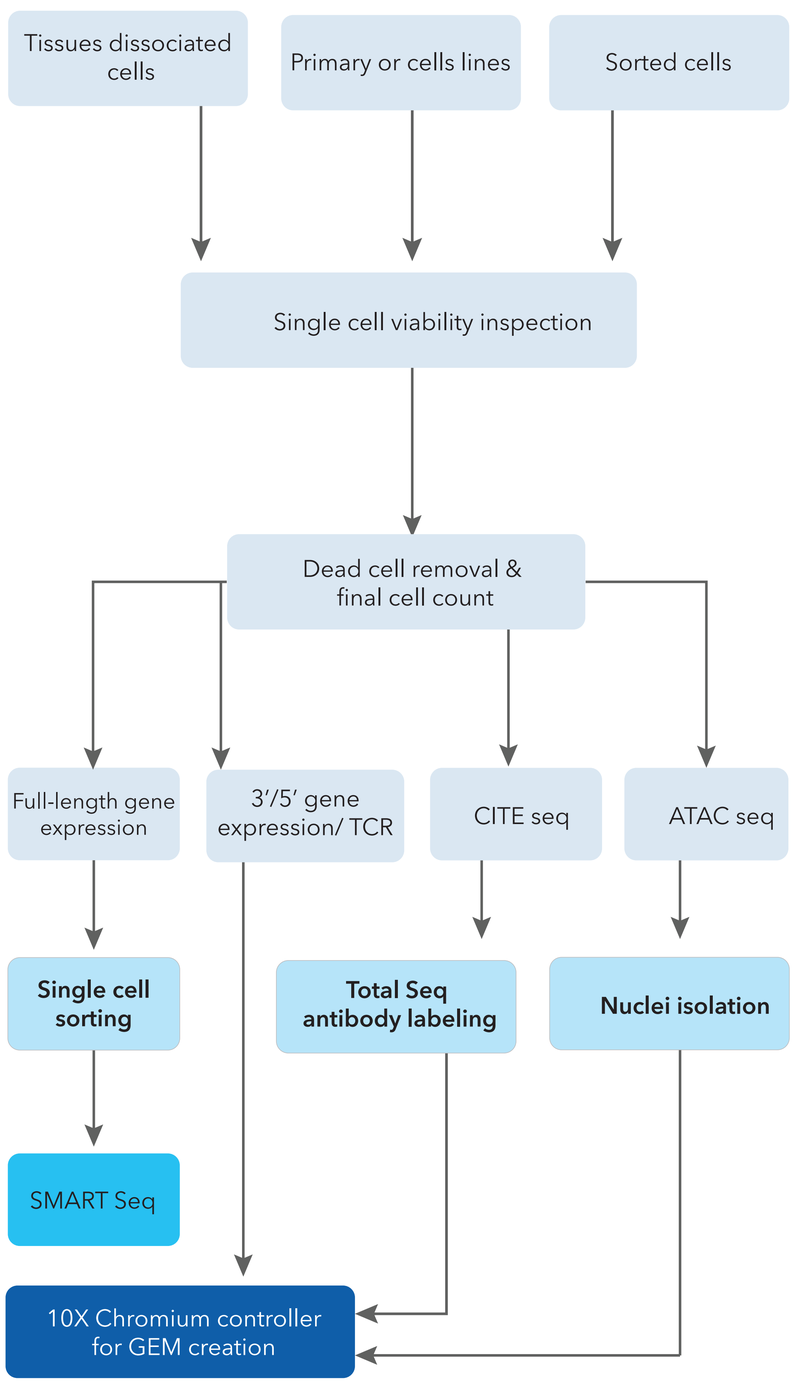



***Figure 1: Schematic representation of the different single cell workflows at MedGenome:*** Starting with single cell suspensions from fresh or frozen dissociated tissues, primary cells, cell lines and FACS sorted cells, cell viability and counting assessments are performed. Depending on the downstream application, cell processing steps are performed. For single cell gene expression, depending on the input cell type or the research question, single cells can be loaded onto the 10x Genomics Chromium controller and library preparation and sequencing can be performed. Single cells can also be sorted into plates with Lysis buffer for the SMARTSeq approach. To perform single cell ATAC seq, nuclei isolation is performed from single cell suspensions and after transposition of the nuclei in bulk, the 10x Chromium controller can be used to generate single nuclei libraries and be sequenced on the Illumina platform. To process cells for characterization of cell surface markers to identify subtypes of cells with CITE-Seq, single cell suspensions can be labeled with antibodies against the cell surface markers. Downstream labeling with feature barcode libraries along with gene expression libraries or single cell TCR repertoire libraries can be generated using the 10x Immune profiling chemistries. A combination multi-omic approach with single cell genomics can enable discovery of additional cellular heterogeneity and characterize biomarkers.

## Single cell gene expression solutions at MedGenome

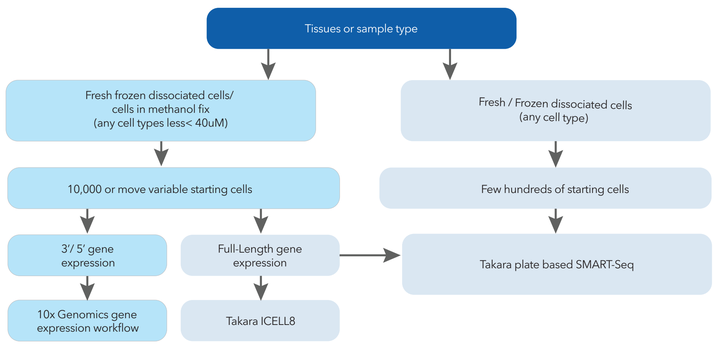
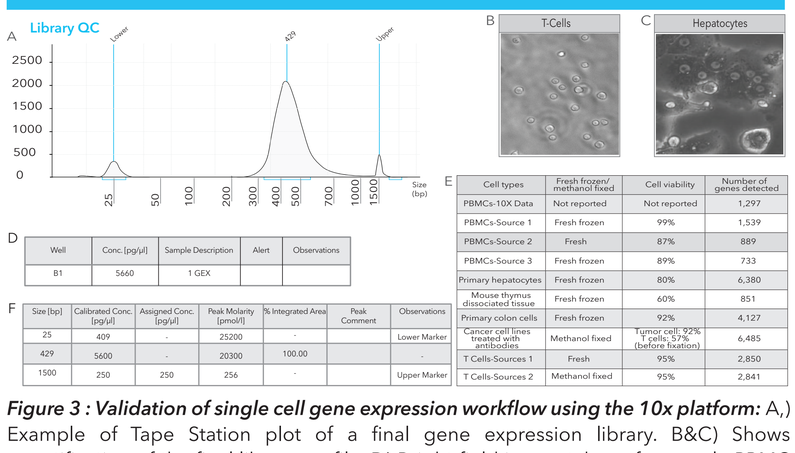



***Figure 2 : Overview of single cell solutions at MedGenome:*** Depending on the type of sample, storage conditions (which affects the viability of the cells and quality of RNA obtained) and the number of cells available, a suitable workflow for library preparation is selected. In general, for fresh frozen tissues, upon optimization of tissue dissociation and freeze thaw conditions by the end user, frozen cell suspensions are further processed through either the 10X Genomics Chromium gene expression solution or the Takara SMART-Seq workflow. In addition to fresh frozen cells, methanol fixed cells can also be processed using the chromium gene expression workflow. For samples with less than a few thousand viable cells, (due to the final recovery of loss of 30% of cells in the processing of cells), we recommend dissociating cells and sorting into plates before the SMART-seq workflow. In addition to the low numbers of cells, if samples include cell types that are not pliable to freeze thaw cycles or have sizes over 40 µM, we recommend proceeding with SMART-seq.

## Validation of 10x gene expression workflow




D

| Well | Conc. [pg/µl] | Sample Description | Alert | Observations |
|---|---|---|---|---|
| B1 | 5660 | 1 GEX | | |

F

| Size [bp] | Calibrated Conc. [pg/µl] | Assigned Conc. [pg/µl] | Peak Molarity [pmol/l] | % Integrated Area | Peak Comment | Observations |
|---|---|---|---|---|---|---|
| 25 | 409 | - | 25200 | - | | Lower Marker |
| 429 | 5600 | - | 20300 | 100.00 | | - |
| 1500 | 250 | 250 | 256 | - | | Upper Marker |

E

| Cell types | Fresh frozen/ methanol fixed | Cell viability | Number of genes detected |
|---|---|---|---|
| PBMCs-10X Data | Not reported | Not reported | 1,297 |
| PBMCs-Source 1 | Fresh frozen | 99% | 1,539 |
| PBMCs-Source 2 | Fresh | 87% | 889 |
| PBMCs-Source 3 | Fresh frozen | 89% | 733 |
| Primary hepatocytes | Fresh frozen | 80% | 6,380 |
| Mouse thymus dissociated tissue | Fresh frozen | 60% | 851 |
| Primary colon cells | Fresh frozen | 92% | 4,127 |
| Cancer cell lines treated with antibodies | Methanol fixed | Tumor cell: 92% T cells: 57% (before fixation) | 6,485 |
| T Cells-Sources 1 | Fresh | 95% | 2,850 |
| T Cells-Sources 2 | Methanol fixed | 95% | 2,841 |

***Figure 3 : Validation of single cell gene expression workflow using the 10x platform:*** A,) Example of Tape Station plot of a final gene expression library. B&C) Shows quantification of the final library profile. D) Bright-field image taken of a sample PBMC where we see an ideal sample with cell size 8-10 µM. F) Bright-field image of hepatocytes (40-50 µM) which is not ideal for processing with 10X Genomics platform and might lead to lower yield of final cells being processed into libraries. As part of our workflow, we evaluate the samples and process, and provide our recommendations on how to analyze the downstream data obtained from the sequencing. F) Example table with cell viability measurements after thawing the fresh frozen samples and post wash and clean up viability. The viability is determined using Trypan blue staining. Before proceeding with processing cells through the 10X Genomics chromium controller, we remove the dead cells using the EasySep™ ( Stem Cell Technologies ) dead cell removal kit.

## Identification of T cell functional response and novel biomarkers of T cell subpopulations using the 10x platform

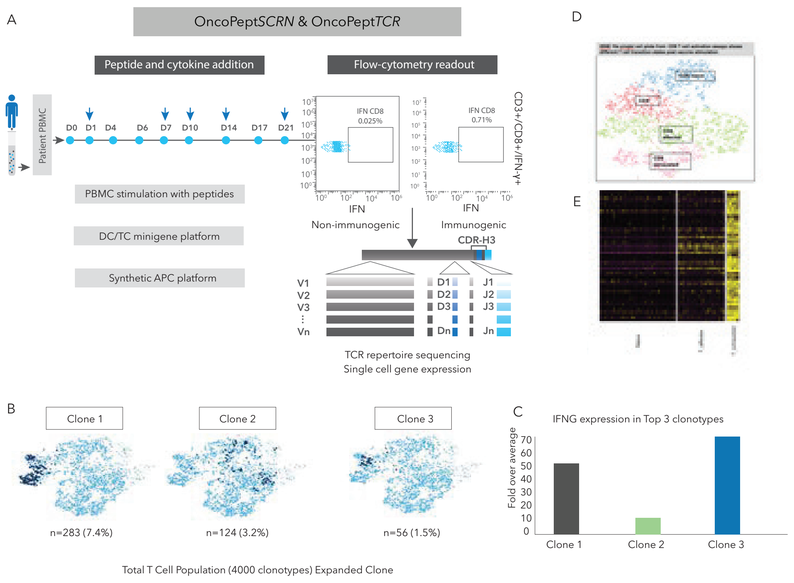
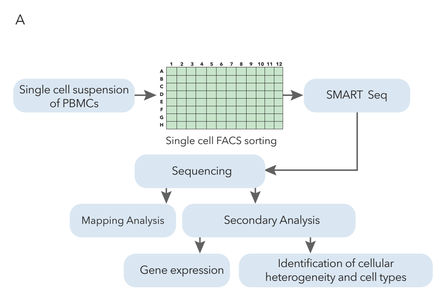



***Figure 4: Application of single cell gene expression coupled to TCR repertoire :*** A) Graphic shows the workflow of an in-vitro neoantigen vaccine screening assay, with a read-out of CD8 T cell activation from PBMCs treated with the peptides. Functional CD8 T cell response is quantitated by IFN-γ production on a flow cytometer. TCR repertoire sequencing and single cell gene expression is then analyzed. B) Single cell TCR repertoire analysis of PBMCs (with positive responses to the peptide, assayed by IFNγ production) identified over 4000 TCR clonotypes, with t-SNE plots showing the top 3 clonotypes and the frequencies are listed below. C) Single cell gene expression of the amplified T cells reveal that there is varying levels of IFNγ in T-cells that are clonally expanded and this information can provide insights on the functional state of the expanded T-cell clones. The single cell TCR and gene expression experiment was performed using the 10x Genomics Immune profiling solution and sequencing was performed using the Illumina Novaseq 6000. 10x Cell Ranger analysis was performed downstream. D) Shows single cell plots from CD8 T cell activation assays. There are different T cell transition states post vaccine stimulation including CD8 native, CD4, CD8 effector and CD8 exhausted. E) shows clustering of the marker genes in the subpopulations identified from the gene expression data.

## Full Length Gene Expression Analysis of Single Cells Using SMART-Seq

A




B

| RNA-Seq | Total Reads | Read Count After Adapter Trimming | Read Count After Contamination Removal | rRNA (%) | MT (%) | Overall % Data Loss | Reads Aligned | Alignment (%) |
|---|---|---|---|---|---|---|---|---|
| P1A10C1 | 184,840 | 184,586 | 173,406 | 2.67 | 3.29 | 6.19 | 142,728 | 82.31 |
| P1A12C1 | 135,528 | 135,308 | 126,440 | 3.45 | 3.02 | 6.71 | 100,628 | 79.59 |
| P1A1C10 | 141,428 | 141,420 | 134,334 | 1.39 | 3.54 | 5.02 | 119,578 | 89.02 |
| P1A2C5 | 134,006 | 133,970 | 124,946 | 4.38 | 2.27 | 6.76 | 105,990 | 84.83 |
| P1A3C1 | 124,624 | 124,346 | 119,760 | 1.64 | 1.92 | 3.90 | 97,278 | 81.23 |
| P1A4C1 | 115,506 | 115,412 | 109,180 | 1.74 | 3.61 | 5.48 | 94,492 | 86.55 |
| P1A5C1 | 131,352 | 131,334 | 117,388 | 7.25 | 3.18 | 10.63 | 96,514 | 82.22 |
| P1A6C1 | 126,692 | 126,450 | 121,432 | 1.59 | 2.24 | 4.15 | 99,568 | 81.99 |
| P1A7C1 | 140,676 | 140,506 | 135,948 | 0.00 | 0.00 | 3.36 | 116,164 | 85.45 |
| P1A8C1 | 94,362 | 94,262 | 89,408 | 0.00 | 0.00 | 5.25 | 76,310 | 85.35 |
| P1A9C1 | 126,368 | 126,184 | 118,948 | 2.05 | 3.57 | 5.87 | 97,208 | 81.72 |
| P1B10C1 | 150,404 | 150,154 | 144,420 | 1.83 | 1.91 | 3.98 | 120,456 | 83.41 |
| P1B11C1 | 143,432 | 143,372 | 135,698 | 2.37 | 2.92 | 5.39 | 119,192 | 87.84 |
| P1B12C1 | 169,114 | 168,880 | 156,592 | 2.80 | 4.40 | 7.40 | 128,644 | 82.15 |
| P1B1C10 | 178,114 | 178,108 | 170,364 | 1.12 | 3.20 | 4.35 | 144,974 | 85.10 |
| P1B2C5 | 177,758 | 177,732 | 172,310 | 1.35 | 1.67 | 3.06 | 118,326 | 68.67 |
| P1B3C1 | 202,732 | 202,622 | 196,308 | 1.45 | 1.56 | 3.17 | 125,106 | 63.73 |
| P1B4C1 | 152,622 | 152,608 | 137,726 | 6.17 | 3.41 | 9.76 | 118,918 | 86.34 |
| P1B5C1 | 178,660 | 178,608 | 172,582 | 1.33 | 1.66 | 3.40 | 144,084 | 83.49 |
| P1B6C1 | 153,508 | 153,206 | 140,578 | 5.55 | 2.43 | 8.42 | 90,968 | 64.71 |
| P1B7C1 | 184,964 | 184,698 | 177,392 | 0.00 | 0.00 | 4.09 | 144,402 | 81.40 |
| P1B8C1 | 122,762 | 122,532 | 117,050 | 2.25 | 2.11 | 4.65 | 96,804 | 82.70 |
| P1B9C1 | 132,408 | 132,314 | 123,160 | 3.17 | 3.67 | 6.98 | 103,596 | 84.11 |

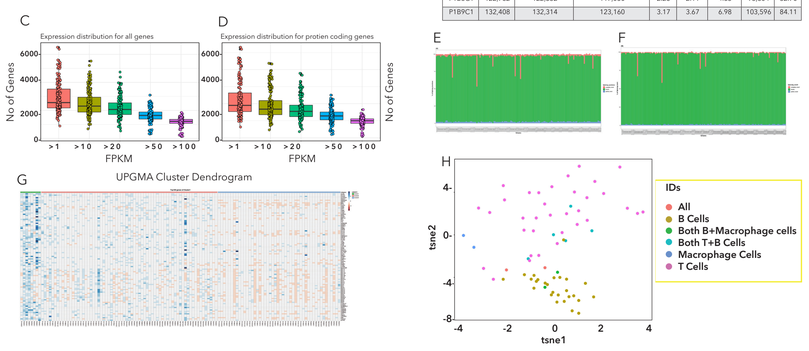



***Figure 5: Schematic workflow of the experimental set up for plate-based SMART-Seq for full length single cell RNA-seq :*** Full length single cell RNA-seq: A) Single cell suspensions were generated using healthy donor PBMCS as per manufacturer's protocols and live cells were gated to be sorted into a 96-well plate with 1X Reaction buffer dispensed into a 96 well plate. The SMART-Seq cDNA synthesis and the Nextera Library preparation were performed at quarter volume. Quality control measurements were performed for individual libraries, and individual libraries were pooled and sequenced using the Illumina NovaSeq 6000 with the PE100 configuration. Each library was sequenced at a depth of 200,000 reads. The data generated was analyzed in-house using standard pipelines. Alignment was performed using STAR (2.4.1) aligner and reads mapping to ribosomal and mitochondrial genome were removed before performing alignment. The raw read counts were estimated using HTSeq-0.6.1. Read counts were then normalized using DESeq2 to get the normalized counts. Additionally, the aligned reads were used for estimating expression of the genes using cufflinks (version:2.2.1). Gene body coverage and splicing analysis was performed on the data (shown in E&F) and then (G&H) clustering analysis was performed to understand the relationship of cells and identify cellular heterogeneity in the sample. Gene expression analysis identify the cell type and cellular heterogeneity in the sample. B) Shows a screenshot of a summary table of the mapping metrics from the sequencing obtained from the libraries. Overall, majority of the samples show percentage of reads aligning to the human genome as 80% and a very low % of ribosomal or mitochondrial reads. C&D) Shows distribution of the number of genes identified with FPKM>1 and above and the distribution of expression for protein coding genes and other transcripts. E & F) Shows the distribution of splice sites identified in each of the cells. G)Shows hierarchical clustering dendrogram of the relationships of the individual cells based on the overall gene expression in individual cells. Based on the gene expression, three major clusters were identified and the top genes in each cluster are listed on the right. F) t-SNE plot shows the cell types within the clusters that were identified from 155 cells. Immune cells such as B cells, CD8 and CD4 T cells and macrophages were identified using this approach.

## Single Cell ATAC Sequencing Identifies Cell Types by Chromatin Accessibility Mapping

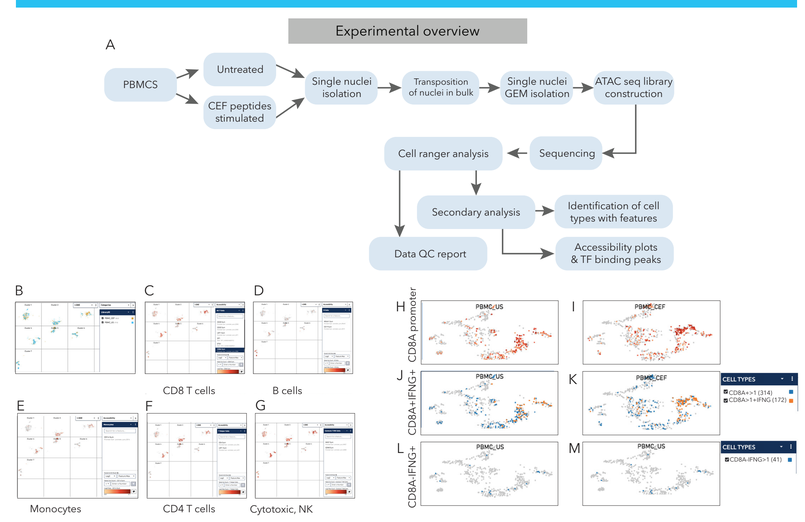



***Figure 6 : Single cell ATAC seq profiling to identify cellular heterogeneity and epigenetic regulation of T-cell activation :*** A) Shows experimental workflow for single cell ATAC sequencing performed using nuclei isolated from PBMCs (unstimulated) and stimulated with a CEF pool of peptides. The nuclei are transposed in bulk and then single nuclei are encapsulated into GEMs using the 10x Genomics Chromium controller. Illumina sequencing libraries are then generated and sequenced on the NovaSeq platform. Analysis of data was performed using Cell Ranger ATAC pipeline from 10x Genomics and secondary analysis was further performed using Loupe Browser. B-G) Transcription factor binding analysis shows the top transcription factor binding motifs that are identified in the open chromatin regions. Taken together, this analysis can be utilized to identify cellular heterogeneity using chromatin accessibility features at a single cell resolution. H-M) Chromatin accessibility profiles of samples identifies cell types with accessible chromatin. Sub-cluster analysis shows chromatin accessibility at promoters of markers of sub-populations of T-cells within the CD8 positive T cells that are effector cells.

## Conclusions and future directions

1) Single cell genomics assays at MedGenome can identify cell types and biomarkers from immune cells in a sensitive manner
2) Recent feature barcoding (CITE-Seq) and epigenomic single cell profiling technologies are currently being validated and will be offered as an end to end solution at MedGenome