**CANCER CLONE FITNESS**

**CLONE : cell or group of cells that is genetically identical to its ancestor cell or group of cells**.

**CANCER CLONE** : A cell that acquires a mutation that increases its fitness will generate more daughter cells than competitor cells that lack that mutation. In this way, **a population of mutant cells**, called a clone, can expand in the neoplasm. Clonal expansion is the signature of natural selection in cancer.

**1 How to define clonal fitness using gene expression**

Progress in defining genomic fitness landscapes in cancer, especially those defined by copy number alterations (CNAs), has been impeded by lack of time-series single-cell sampling of polyclonal populations and temporal statistical models1-7. Here we generated 42,000 genomes from multi-year time-series single-cell whole-genome sequencing of breast epithelium and primary triple-negative breast cancer (TNBC) patient-derived xenografts (PDXs), revealing the nature of CNA-defined clonal fitness dynamics induced by TP53 mutation and cisplatin chemotherapy. Using a new Wright-Fisher population genetics model8,9 to infer clonal fitness, we found that TP53 mutation alters the fitness landscape, reproducibly distributing fitness over a larger number of clones associated with distinct CNAs. Furthermore, in TNBC PDX models with mutated TP53, inferred fitness coefficients from CNA-based genotypes accurately forecast experimentally enforced clonal competition dynamics. Drug treatment in three long-term serially passaged TNBC PDXs resulted in cisplatin-resistant clones emerging from low-fitness phylogenetic lineages in the untreated setting. Conversely, high-fitness clones from treatment-naive controls were eradicated, signalling an inversion of the fitness landscape. Finally, upon release of drug, selection pressure dynamics were reversed, indicating a fitness cost of treatment resistance. Together, our findings define clonal fitness linked to both CNA and therapeutic resistance in polyclonal tumours.

**2 scRNA-seq is the best tool**

Single-cell RNA sequencing finds its main applications in immunology (covered [here](https://omictools.com/blog/single-cell-rna-sequencing-immunology)), cancerology, and the study of development. This technology has already permitted to refine our comprehension of differentiation decisions made by histologically identical cells, to identify new cell types and states within organs, and has promising applications in personalized medicine for cancer and biomarker identification.

[Seurat](https://omictools.com/seurat-2-tool) is an R package that enables quality control (QC), analysis, and exploration of single cell RNA-seq data. The software includes three computational methods: (1) unsupervised clustering and discovery of cell types and states, (2) spatial reconstruction of single cell data, and (3) integrated analysis of single cell RNA-seq across conditions, technologies, and species. It can also localize rare subpopulations, and map both spatially restricted and scattered groups.

Seurat main asset is its ability to use data from different sequencing technology, species or condition, and integrate them to identify shared population across data sets and downstream comparative analysis, by identifying shared sources of variation.

Analysis:

What are the favorable conditions for the biggest cluster to:

1. Divide faster

2. What type of mutation is in the biggest cluster

3. Clonal advantages of the biggest cluster

**DATASET**

**In single cell data analysis genes are features and cells are samples**

Download data from cancerSCEM

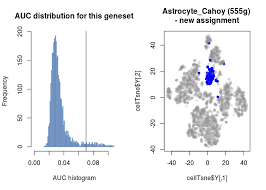
CANCERSCEM : A database of Cancer Single-cell Expression Map. CancerSCEM consists of 208 cancer scRNA-seq datasets across 20 human cancer types, such as lung adenocarcinoma (LUAD), colorectal cancer (CRC) and glioblastoma (GBM).

**Cell markers refer to a type of marker that is specifically expressed at a specific time in a specific cell**, reflecting the growth and differentiation of cells, and can be used to identify specific cells and monitor cell growth and differentiation.

|  |  |  |
| --- | --- | --- |
| No | Marker | Identifier |
| 1 | Astrocyte | AGXT2L1 |
|  |  | GFAP |
|  |  | ALDOC |
|  |  | SLC1A3 |
|  |  | AGT |
|  |  | ALDH1L1 |
| 2 | B Cell | CD19 |
|  |  | MS4A1 |
|  |  | BANK1 |
|  |  | BLK |
|  |  | IRF8 |
|  |  | ABCB4 |
|  |  | ABCB9 |
|  |  | AFF4 |
|  |  | AIDA |
|  |  | AIM2 |
| 3 | Endothelial cell | VWF |
|  |  | PECAM1 |
|  |  | CDH5 |
|  |  | VEGFA |
|  |  | FLT1 |

Assign score with the help of AUCell (R)

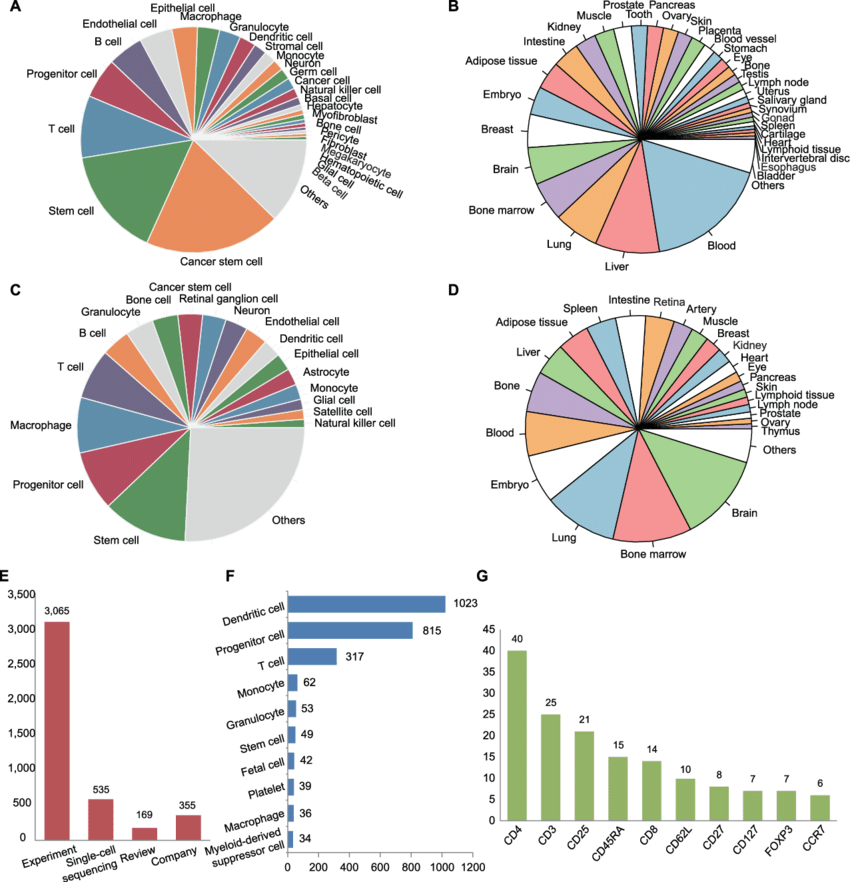
AUCell **allows to identify cells with active gene sets (e.g. signatures, gene modules...) in single-cell RNA-seq data**. AUCell uses the "Area Under the Curve" (AUC) to calculate whether a critical subset of the input gene set is enriched within the expressed genes for each cell.



**Stoufferr's score :** Stouffer's method **combines multiple weighted Z-scores which are calculated in each study**. Although many weight can be introduced but weighting by sample-size is used in meta. oneside.

**GSVA score :** The GSVA enrichment score is **either the maximum deviation from zero (top) or the difference between the two sums (bottom)**. The two plots show two simulations of the resulting scores under the null hypothesis of no gene expression change.

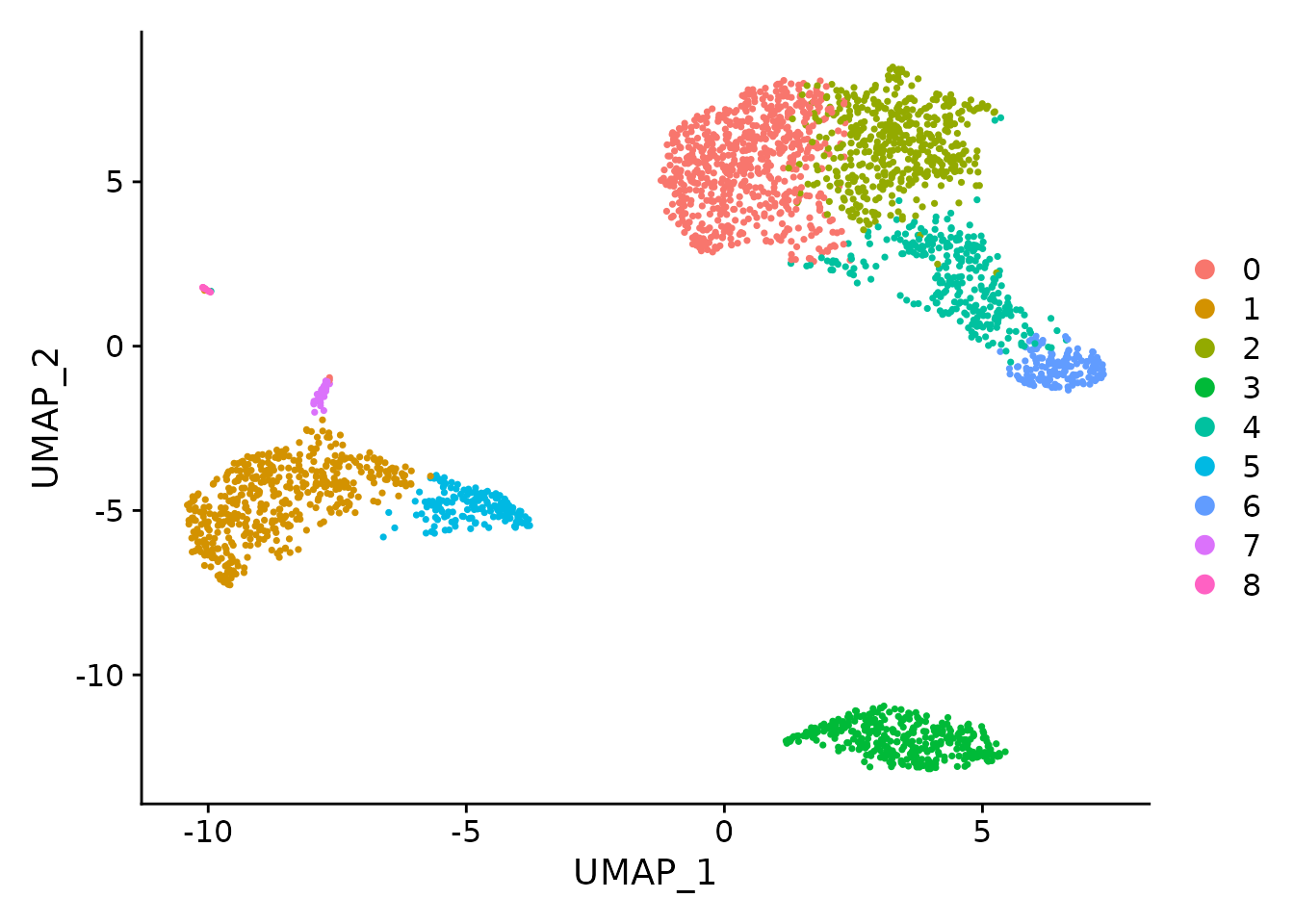
**Cell assignment based on markers given in cancerSCEM with the help of AUCell.**



**Clustering the data**

Seurat v3 applies a graph-based clustering approach, building upon initial strategies in ([Macosko et al](http://www.cell.com/abstract/S0092-8674(15)00549-8)). Importantly, the distance metric which drives the clustering analysis (based on previously identified PCs) remains the same. However, our approach to partitioning the cellular distance matrix into clusters has dramatically improved. Our approach was heavily inspired by recent manuscripts which applied graph-based clustering approaches to scRNA-seq data [[SNN-Cliq, Xu and Su, Bioinformatics, 2015]](http://bioinformatics.oxfordjournals.org/content/early/2015/02/10/bioinformatics.btv088.abstract) and CyTOF data [[PhenoGraph, Levine et al., Cell, 2015]](http://www.ncbi.nlm.nih.gov/pubmed/26095251). Briefly, these methods embed cells in a graph structure - for example a K-nearest neighbor (KNN) graph, with edges drawn between cells with similar feature expression patterns, and then attempt to partition this graph into highly interconnected ‘quasi-cliques’ or ‘communities’.

Seurat can help you find markers that define clusters via differential expression. By default, it identifies positive and negative markers of a single cluster (specified in ident.1), compared to all other cells. [FindAllMarkers()](https://satijalab.org/seurat/reference/FindAllMarkers.html) automates this process for all clusters, but you can also test groups of clusters vs. each other, or against all cells.



**INFERNCV**

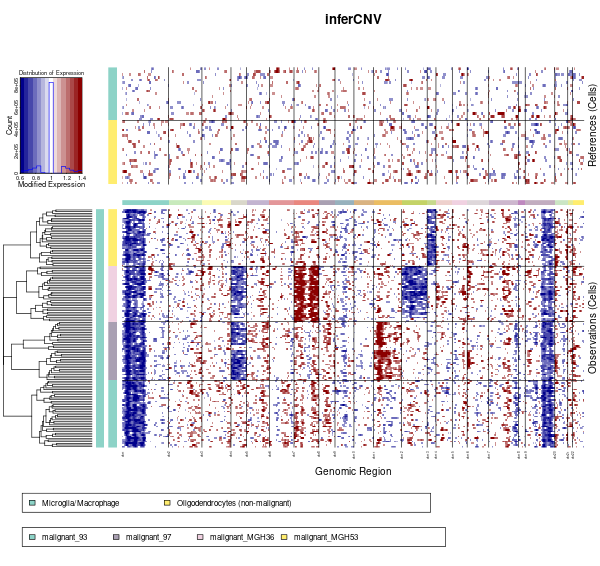
InferCNV is used to explore tumor single cell RNA-Seq data to identify evidence for large-scale chromosomal copy number variations, such as gains or deletions of entire chromosomes or large segments of chromosomes. This is done by exploring expression intensity of genes across positions of the genome in comparison to the average or a set of reference ‘normal’ cells. A heatmap is generated illustrating the relative expression intensities across each chromosome, and it becomes readily apparent as to which regions of the genome are over-abundant or less-abundant as compared to normal cells (or the average, if reference normal cells are not provided).

Installing

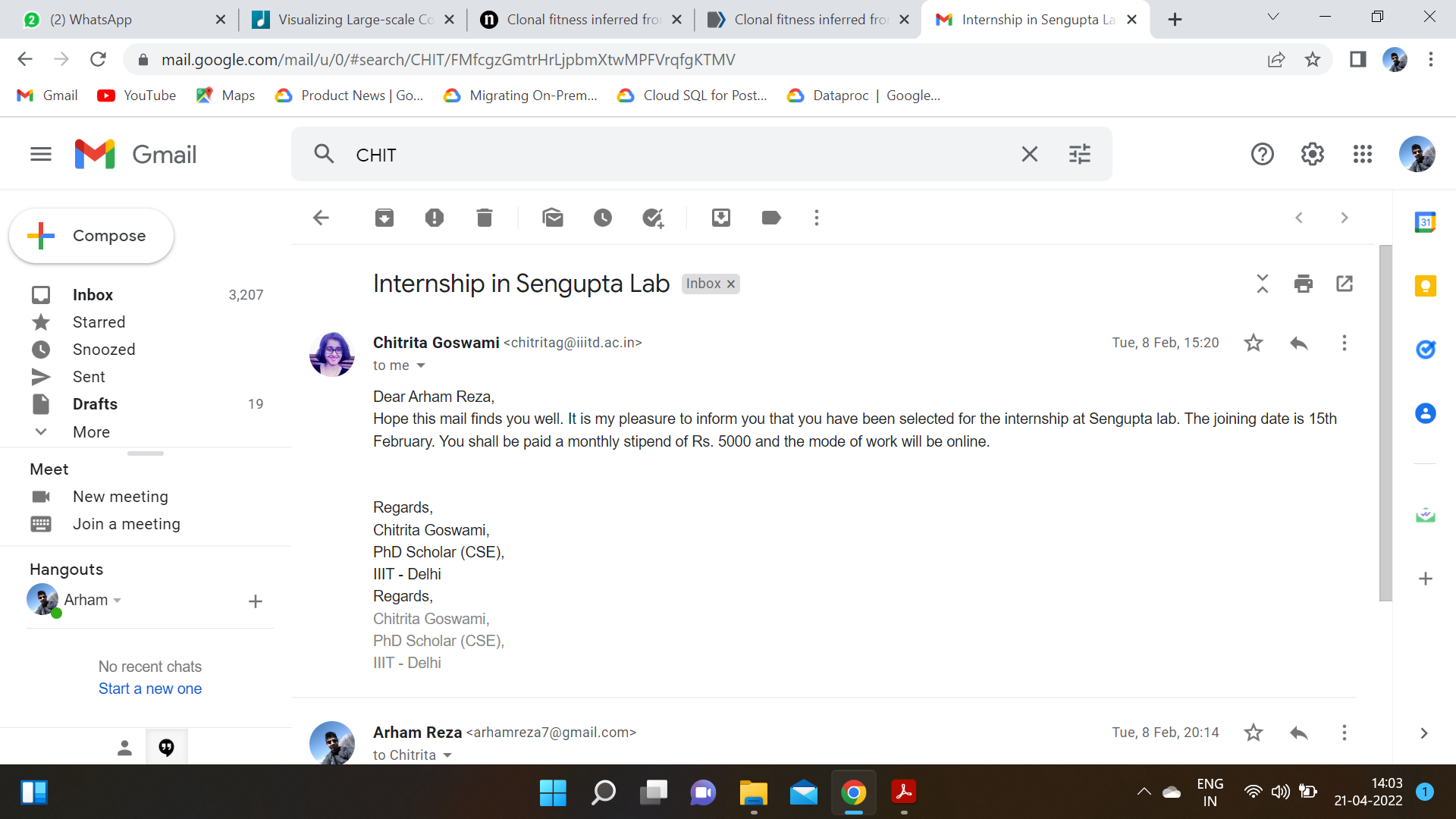
**if** (!requireNamespace("BiocManager", quietly = TRUE))

install.packages("BiocManager")

BiocManager::install("infercnv")



* **We applied this package after clustering the data by seurat to specific celltype of Brain Called Glyoblastoma**
* **This package shows the amplification and the deletion rate of any celltype across its chromosome numbers**
* **The red cluster from the output signifies the amplification rate and the blue cluster signifies the Deletion rate.An Identical celltype shows this deletion and amplification rate in identical postion of the chromosome number.**
* **Thus we can clearly see how the chromosome of a particular clone is amplifying and deleting.Based on this we can identify any celltype**

****