**Scenic Preparation: 102**

To run Scenic an input table with gene expression in each cell is required. From the raw single cell RNA sequencing (scRNA-seq) data only expression information in cells that had a cell type assigned was kept. Genes, that did not meet the following filtering criteria were left out: **sum of counts for a single gene among all cells >2\*0.01\*2\*0.01\*number of cells**. Some differences in nomenclature between our scRNA-seq data and Boulori paper network were identified and subsequently corrected. It was made sure that the maximum amount of Bolouri paper network nodes was included in the Scenic input. Out of 63 nodes, 48 were already added during the gene filtering step and 8 more were added manually afterwards. As a result, a table containing expression data of 10 245 genes among 784 cells was used as Scenic input.

Each Scenic run provides three files containing information about, importance of single interactions between a transcription factor (TF) and a target, composition of a regulon and regulon activity in studied cells. In order to obtain more robust regulons and interactions a total amount of 50 Scenic iterations was performed.

**Scenic clean up: 190**

1. Select the targets from reg file. Save full regulons and only TF-TF interactions.
2. From TF-TF interaction keep only the ones that appear in 40 or more runs (80% of total runs).
3. For these interactions that pass so called frequency filter calculate NIS score.

**Scenic interactions confirmation**

Cistrome

1. Select relevant TFs from Cistrome database.
2. For each TF select top 6000 peaks.
3. Extend each peak by 1 kb.
4. Overlap peaks with human genome to discover target, that will be considered as confirmed by cistrome database.

ATAC seq

1. ATAC-seq data from patients to further confirm the interactions.

Dorothea

1. Use Dorothea as another reference database

Summary

1. Tell how many unique for each source, how many by all three

**Binarize activity for TFs**

1. Working with full regulons (not only TF-TF interactions, but also including other genes)
2. Keep only regulons that pass frequency filter
3. For these regulons conduct new AUCell
4. Heatmap: new\_aucell\_not\_norm.pdf
5. Random Forest to pick 100 TFs that explain the most our cell types
6. Binarize AUCell score for each cell (depending on the cell type) in those 100 TFs. Percentage of cells that are active in each cell type.
7. Heatmap: aucell\_scores\_norm\_top100\_cell\_type\_binarized\_percentage
8. Assign cell type to each TF

**Network of 100 RF TFs**

1. 90% filter on interactions
2. Network: 82\_RF\_TFs.png

**Overlap our network with Bolouri paper Network**

1. 9 overlapping nodes
2. Network: 41\_paper\_overlap\_TFs.png

**Fusion of our network and Bolouri paper network using common nodes**

1. Adapt the nomenclature and format Bolouri paper interactions
2. Fuse networks
3. Network: 66\_bolouri\_RF\_fusion.png

**Reduce the network to relevant nodes**

1. Explain why
2. Network: 66\_bolouri\_RF\_fusion\_curated
3. Network: 66\_bolouri\_RF\_fusion\_curated\_2

**Further reduce the network to only**

1. Explain why
2. Network: 13\_Bonesis\_toy