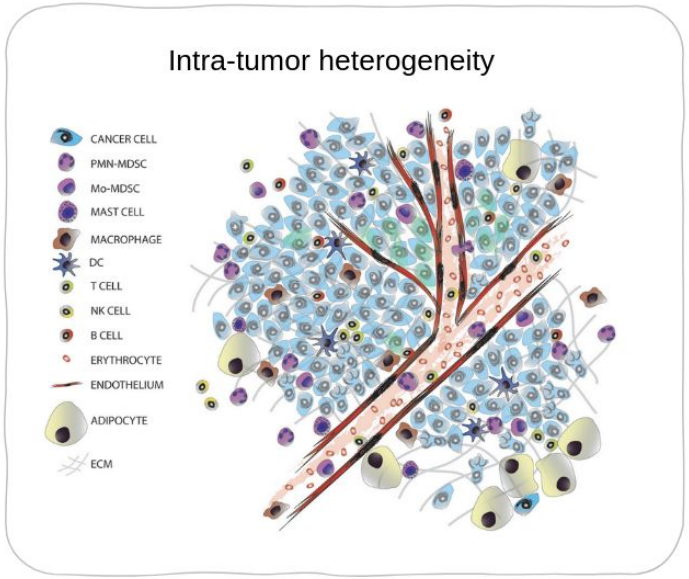
Associating CNAs and Transcriptional States in Cancer based on scRNA-seq Data

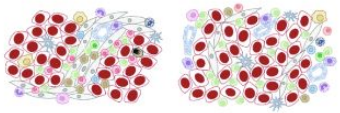
Philip Toma
BSc Student at the Boeva Lab

Contact: tomap@student.ethz.ch

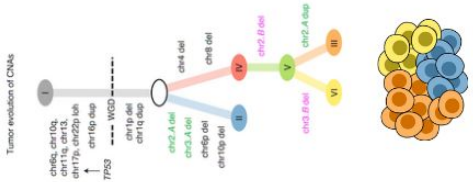
Heterogeneity in tumors



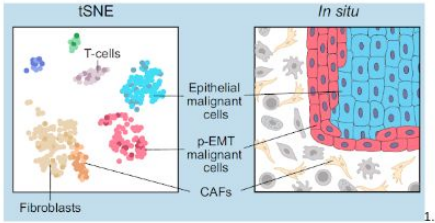
- Different composition of tumor microenvironment (TME)



- Genetic variation between cancer cell subclones



- Transcriptional (and epigenetic) heterogeneity (often not-clonal, unrelated to the mutational profile)



Images adapted from Valdes Mora, et al. 2018, *Frontiers in Immunology*; Geok Wee Tan, et al, *Pathogenes*, 2018; Sidharth Puram et al., *Cell*, 2017

Slide adapted from "**Searching for transcriptional heterogeneity in tumors: a computational quest**" Prof. Valentina Boeva (November 2020)

Quantifying transcriptional heterogeneity (through scRNA-seq) yields non-genetic intra-tumour differences

Quantifying transcriptional heterogeneity (through scRNA-seq) yields non-genetic intra-tumour differences

Measure therapeutic resistance for variations of cancer cells within a tumour

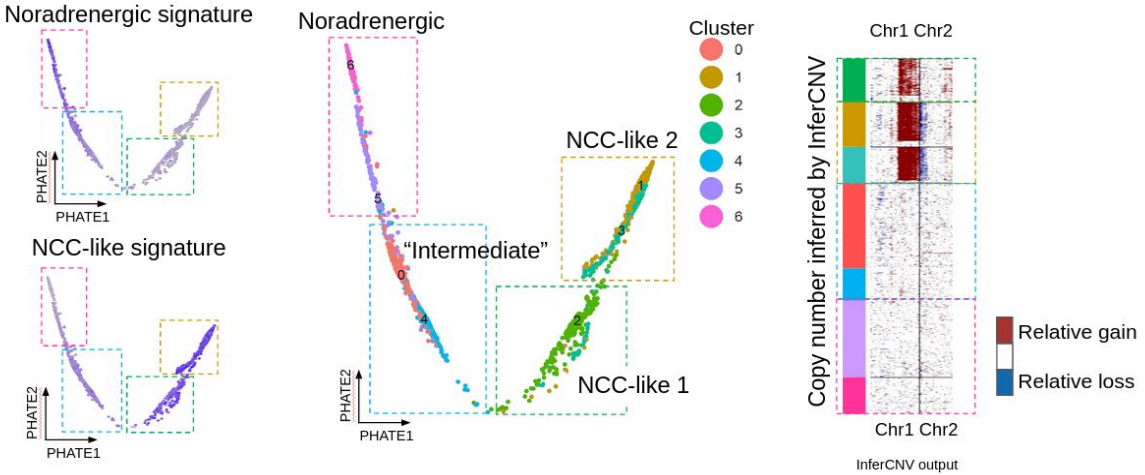
Quantifying transcriptional heterogeneity (through scRNA-seq) yields non-genetic intra-tumour differences

Measure therapeutic resistance for variations of cancer cells within a tumour

Development of effective treatment methods!

Kunihiko Hinohara and Kornelia Polyak, **Intratumoral Heterogeneity: More Than Just Mutations**, *Trends in Cell Biology,* 2019

Can the heterogeneity in neuroblastoma be driven by genomic events?



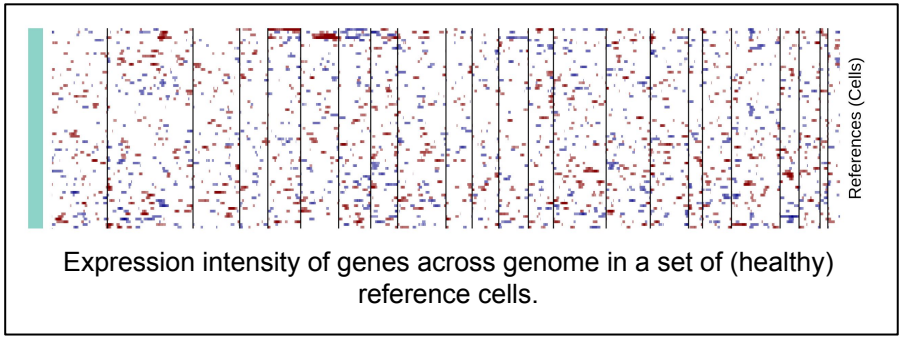
Melissa Saichi

SK-N-SH cell line (enriched in intermediate state)

25.07.2019 32

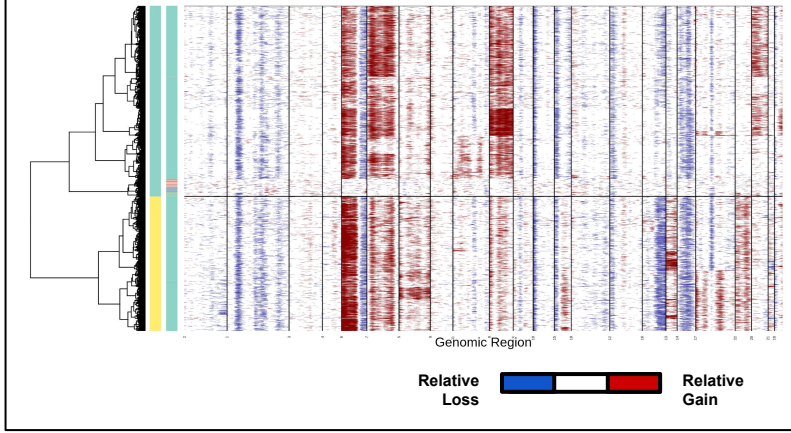
Slide adapted from "**Searching for transcriptional heterogeneity in tumors: a computational quest**"
Prof. Valentina Boeva (November 2020)

Inferring copy number gains and losses relative to a reference genome

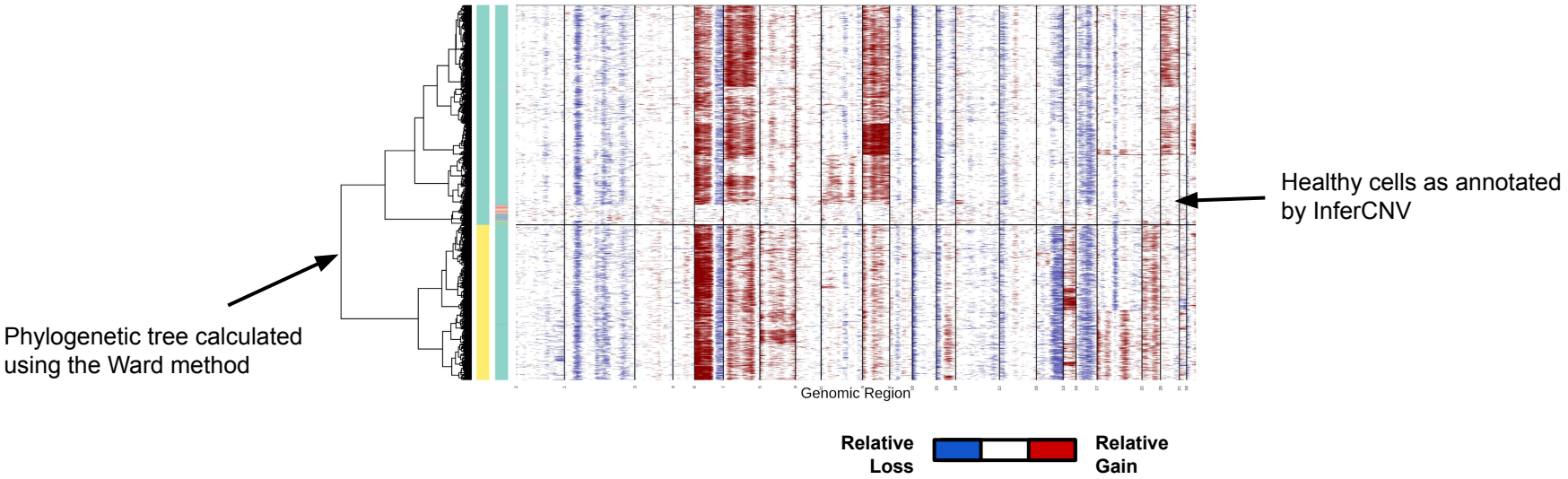


Expression intensity of genes across genome in a set of (healthy) reference cells.

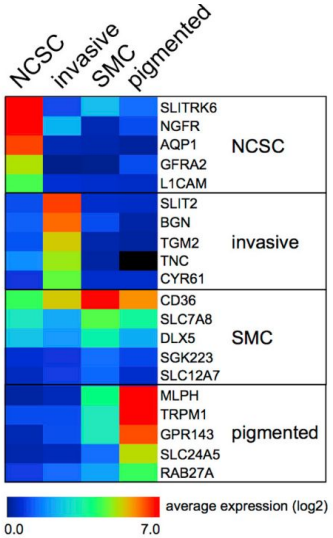
Heatmap of relative expression intensities across each chromosome



InferCNV to infer copy number gains and losses relative to a reference genome



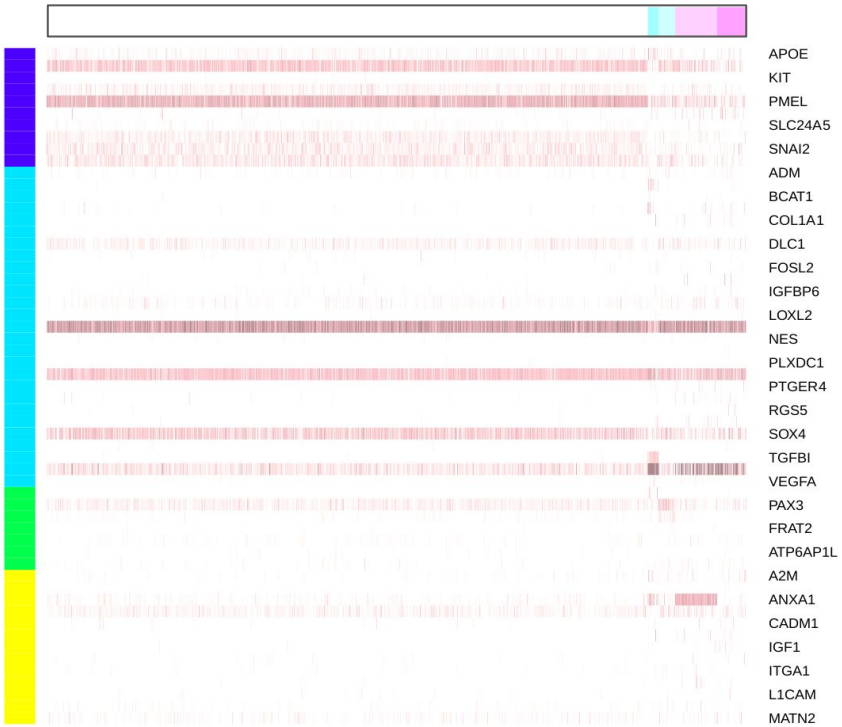
States as proposed by Rambow et al.

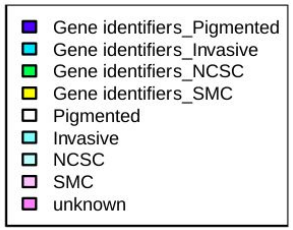


<u># Marker Genes:</u>
49 for Invasive
37 for NCSC
15 for Pigmented
14 for SMC

Florian Rambow et al, **Toward Minimal Residual Disease-Directed Therapy in Melanoma**, *Cell*, 2018

Example Marker Gene Expression Heatmap





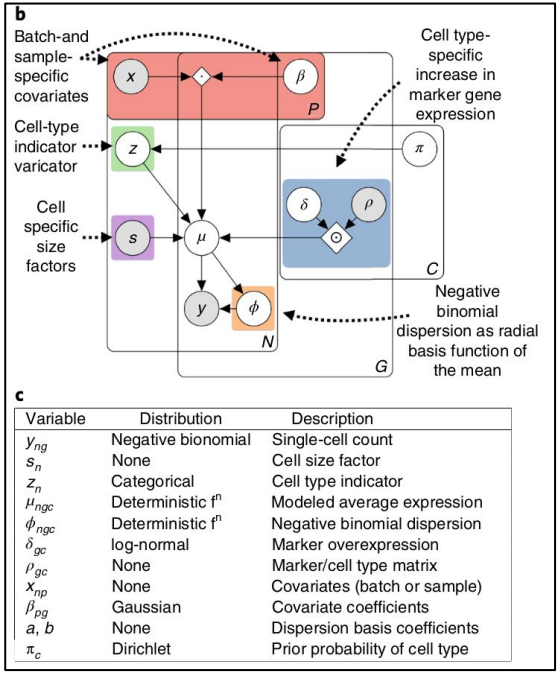
Sample: **MIJUCYK**

Finding links between CNAs and Rambow states requires cell-state annotations!

Finding links between CNAs and Rambow states requires cell-state annotations!

Semi-Supervised Learning
(utilising marker genes proposed by Rambow et al.)

Use marker gene signatures to probabilistically assign cell-types

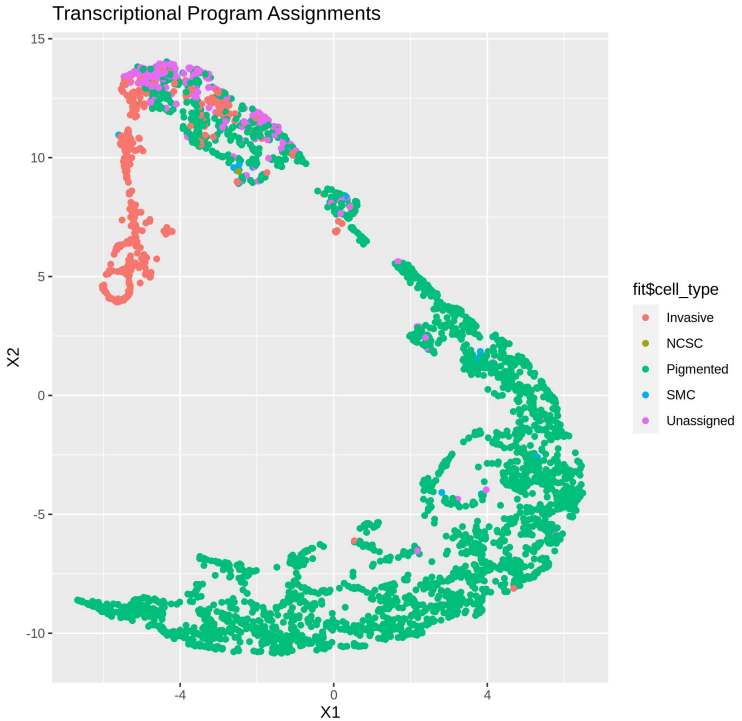


Variable	Distribution	Description
y_{ng}	Negative bionomial	Single-cell count
s_n	None	Cell size factor
z_n	Categorical	Cell type indicator
μ_{ngc}	Deterministic f^n	Modeled average expression
ϕ_{ngc}	Deterministic f^n	Negative binomial dispersion
δ_{gc}	log-normal	Marker overexpression
ρ_{gc}	None	Marker/cell type matrix
x_{np}	None	Covariates (batch or sample)
β_{pg}	Gaussian	Covariate coefficients
a, b	None	Dispersion basis coefficients
π_c	Dirichlet	Prior probability of cell type

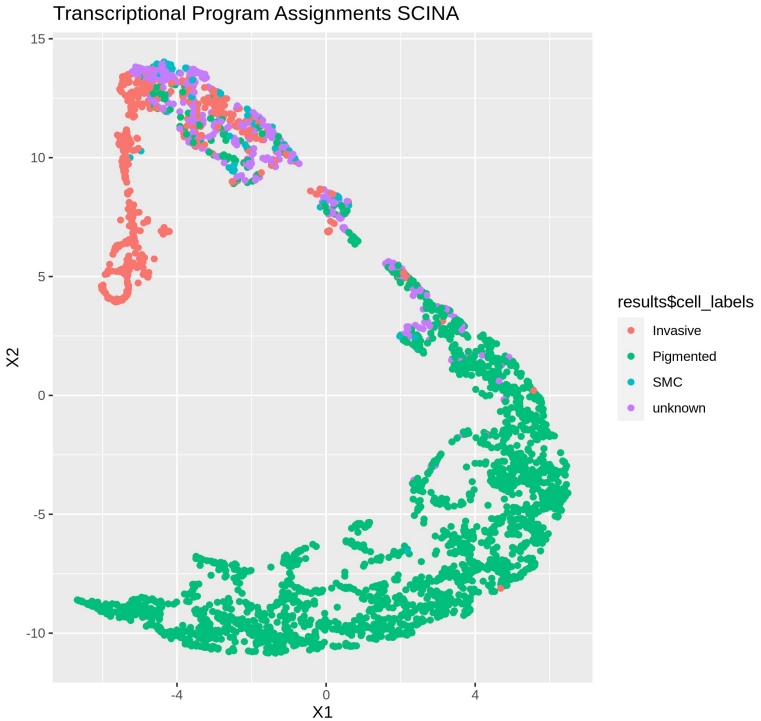
Allen W. Zhang et al, **Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling**, *Nature Methods*, 2019

Assumption: Marker genes are more highly expressed in their identifying type.

→ Using expectation maximization, find the most likely cell type *c*, to which cell *z*□ belongs (given marker gene overexpression estimates)



Learning rate given to Cellassign: 1e-3. With cell cycle removal.



Chosen convergence rate: 0.76 ; no CC-removal.

How can we evaluate the correctness of semi-supervised annotations in the absence of cell type labels?

Use clustering evaluation techniques.

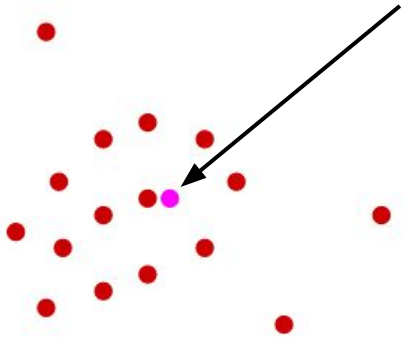
Evaluation Metrics

Intracluster Compactness

1) Calculate the average cell from a cell type.

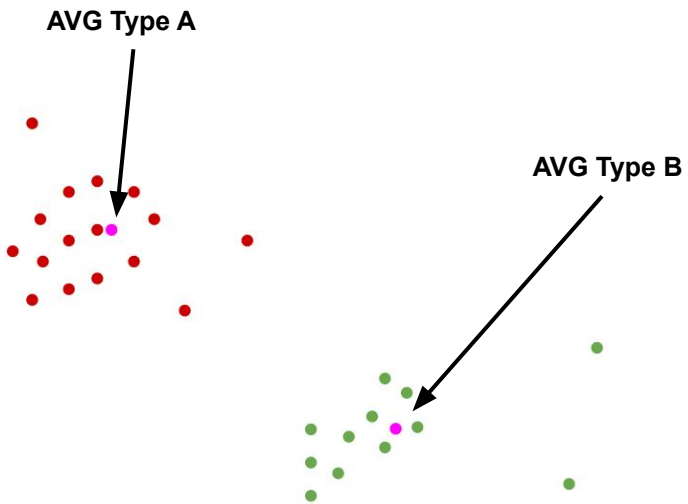
2) For each cell calculate the Pearson correlation to the average of its cell type.

3) Return mean correlation.



Evaluation metrics adjusted from Duan et al. **Learning for single-cell assignment**, *Science Advances*, 2020

Evaluation Metrics



Intercluster Complexity

1) Calculate the average cell from each cell type.

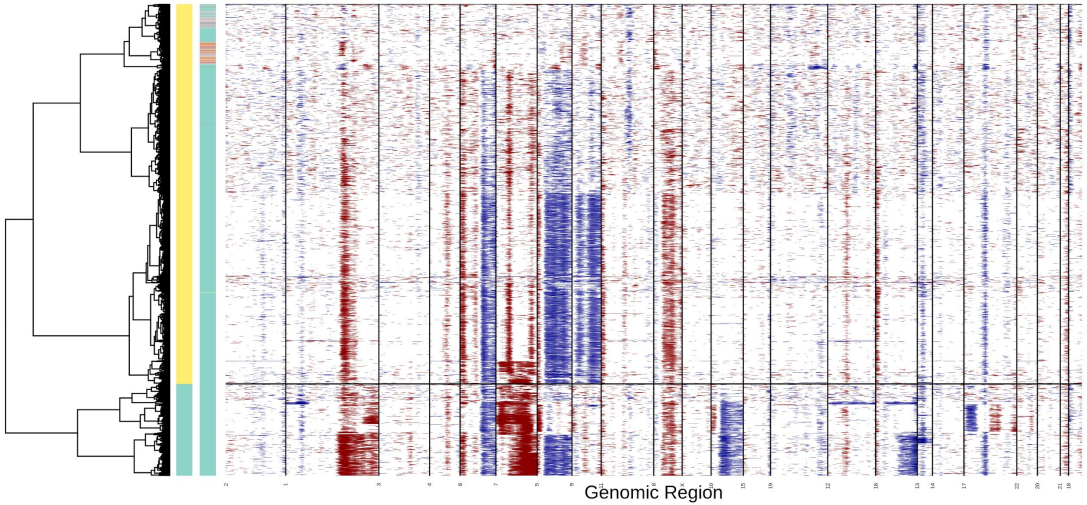
2) For each cell in type A, calculate the Pearson correlation with the avg of type B.

3) Return mean correlation between A and avg of type B.

Evaluation metrics adjusted from Duan et al. **Learning for single-cell assignment**, *Science Advances*, 2020

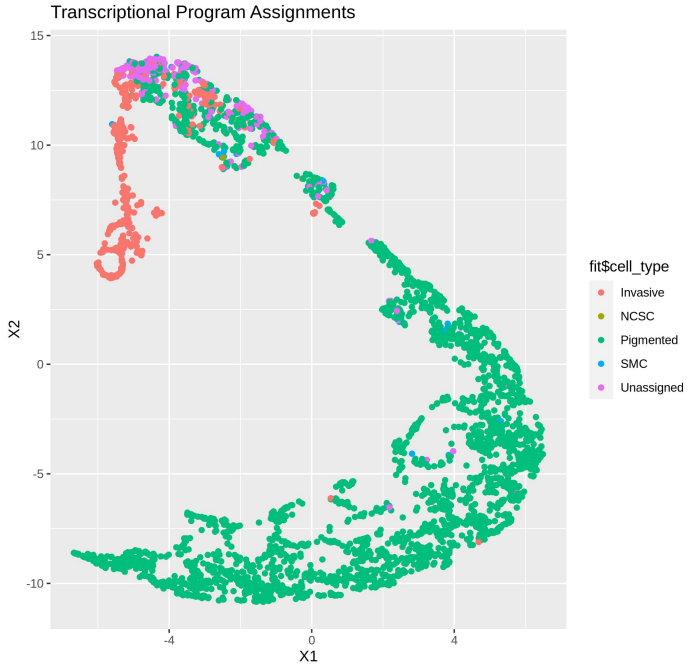
Evaluation Results

Is there a link between CNAs and the transcriptional states described by Rambow?

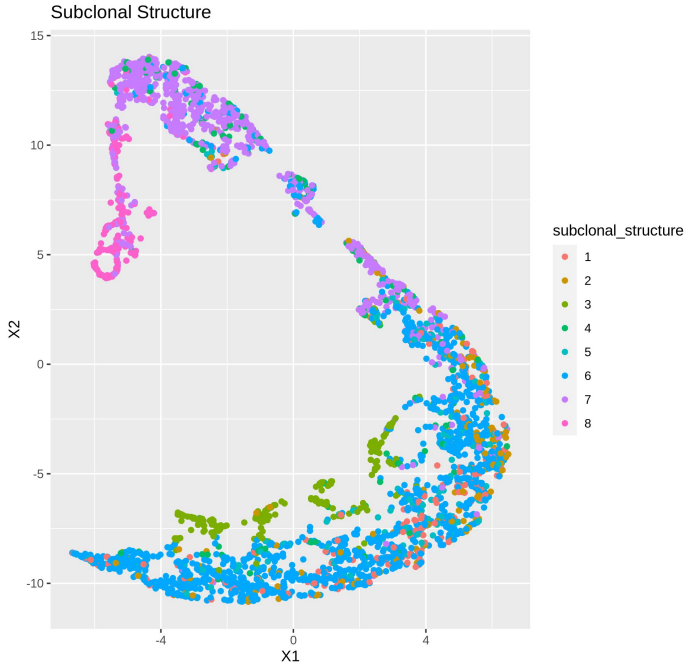


- Use phylogenetic tree to infer subclonal structures
- Which granularity to choose?
- How significant is the overlap between subclones and cell-type annotations?
- **If overlap: which genomic regions drive the differentiation?**

Motivating preliminary results!

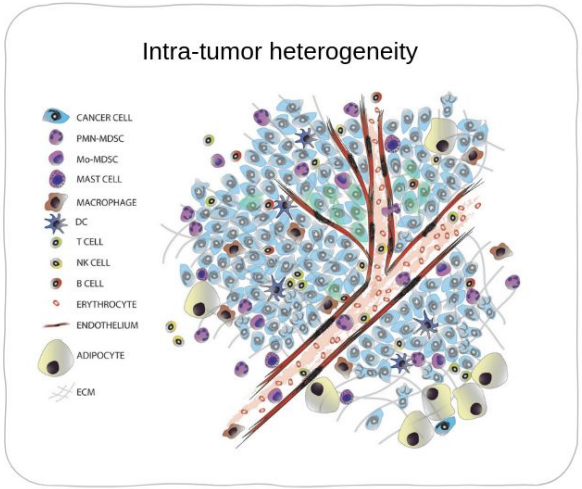


Learning rate given to Cellassign: 1e-3. With cell cycle removal.

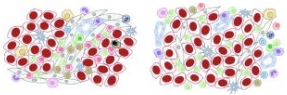


Dendrogram taken from infercnv output. Structures found at level 4 of the hierarchy.

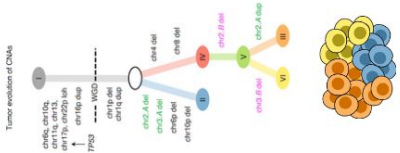
Heterogeneity in tumors



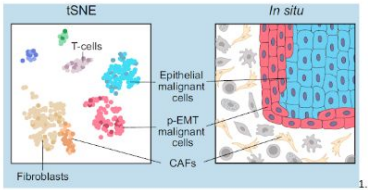
- Different composition of tumor microenvironment (TME)



- Genetic variation between cancer cell subclones



- Transcriptional (and epigenetic) heterogeneity (often not-clonal, unrelated to the mutational profile)



Images adapted from Valdes Mora, et al. 2018, *Frontiers in Immunology*; Geok Wee Tan, et al, *Pathogenes*, 2018; Sidharth Puram et al., *Cell*, 2017

- Infer phylogenetic tree (Ward Method)
- Split sample into sub-clonal structures
- Analyse similarities between
- Analyse genomic regions for CNA

- Use Cellassign and SCINA to generate annotations
- Evaluate using complexity and compactness (intra- and inter-tumour)

Slide adapted from "**Searching for transcriptional heterogeneity in tumors: a computational quest**"
Prof. Valentina Boeva (November 2020)

Associating CNAs and Transcriptional States in Cancer based on scRNA-seq Data

Philip Toma
BSc Student at the Boeva Lab

Contact: tomap@student.ethz.ch