EuroVis 2016
18th EG/VGTC Conference on Visualization
6-10 June 2016, Groningen, the Netherlands

Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets

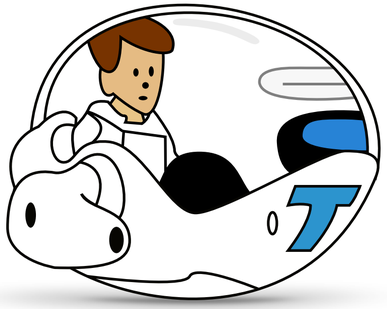
Thomas Höllt[1], Nicola Pezzotti[1], Vincent van Unen[2], Frits Koning[2], Elmar Eisemann[1], Boudewijn Lelieveldt[3], and Anna Vilanova[1]

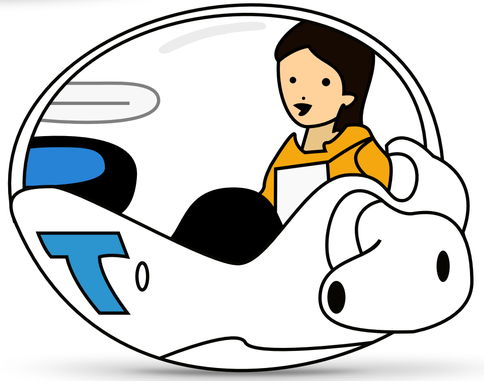
[1]Computer Graphics and Visualization, TU Delft

[2]IHB, Leiden University Medical Center

[3]LKEB, Leiden University Medical Center

Motivation





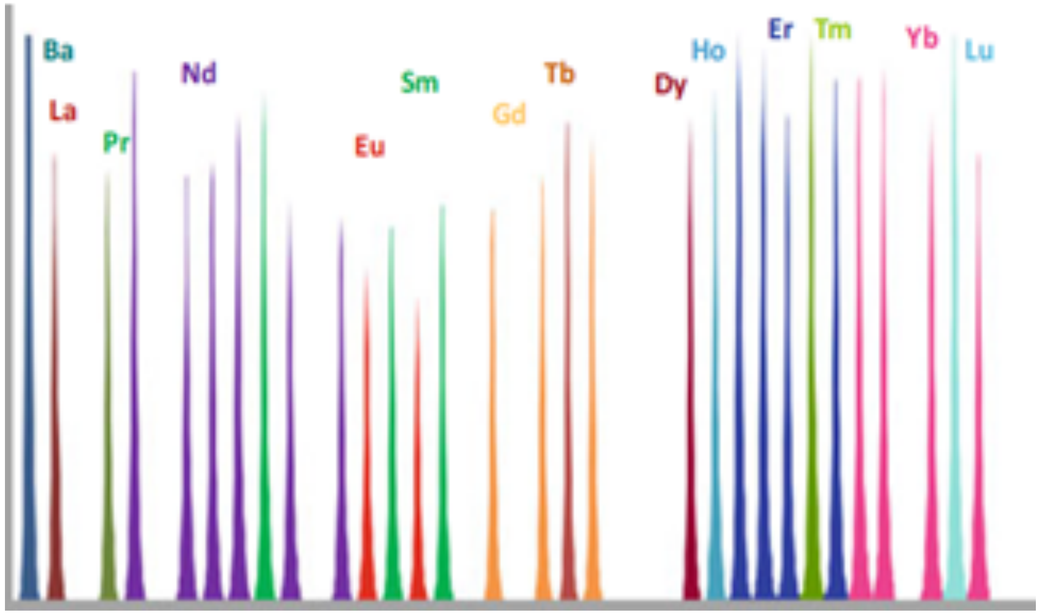
- People react differently to exposure to disease (and treatment)
- Differences in the immune system can be responsible
- Analyze differences on cellular level

Mass Cytometry I

- Recent development in cell analysis (Commercially available since 2014)
- Cells stained with heavy metals
- Analysis with mass spectrometer
- Allows much higher precision than clinical standard (Flow Cytometry) ~40 features vs ~12 features

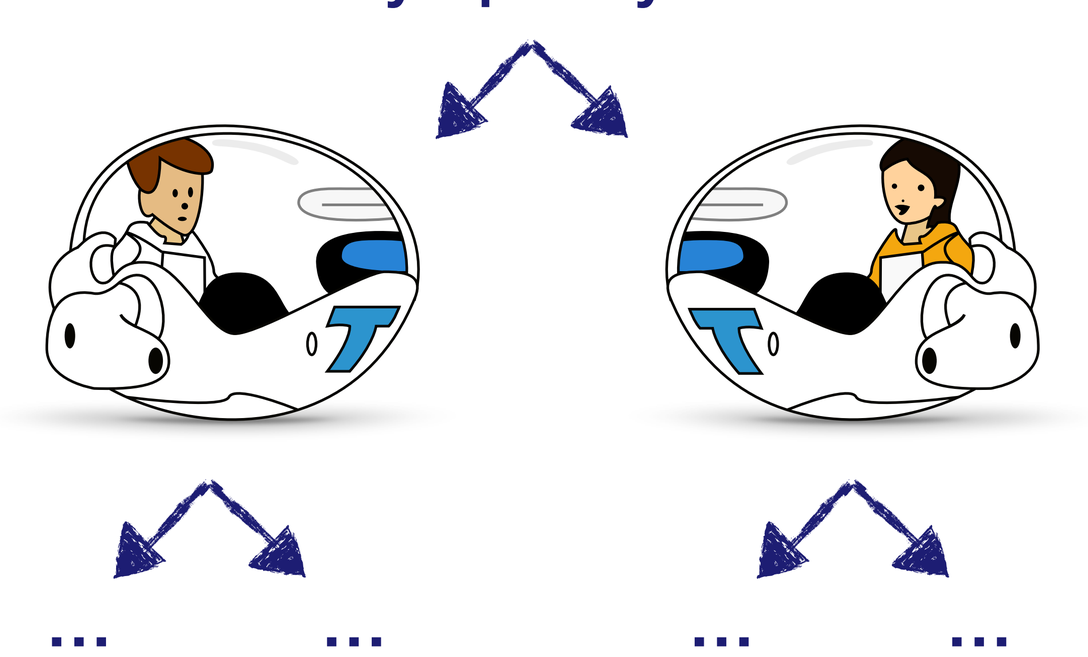
Mass Cytometry II

- ~40 markers // 10.000 immune-system-wide proteins
- Cohort studies ⇨ Millions of Cells
- More markers lead to
  - More depth
    - ⇨ Identification of unknown cell types
  - More breadth
    - ⇨ Exploration of multiple lineages



Mass Cytometry II

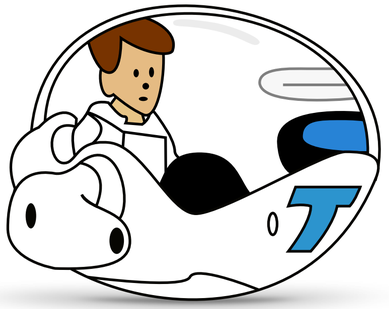
- ~40 markers // 10.000 immune-system-wide proteins
- Cohort studies ⇨ Millions of Cells
- More markers lead to
  - More depth
    ⇨ Identification of unknown cell types
  - More breadth
    ⇨ Exploration of multiple lineages



Mass Cytometry II

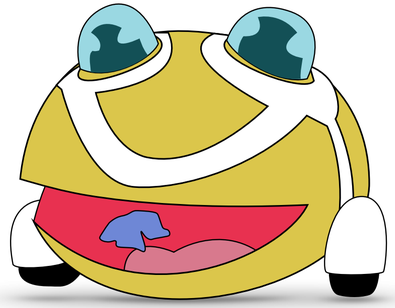
- ~40 markers // 10.000 immune-system-wide proteins
- Cohort studies ⇨ Millions of Cells
- More markers lead to
  - More depth
    ⇨ Identification of unknown cell types
  - More breadth
    ⇨ Exploration of multiple lineages

T-Lymphocytes



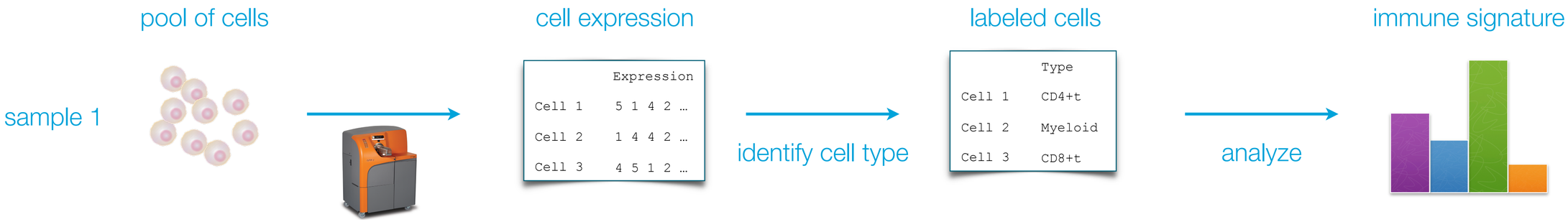
... ...

Macrophage

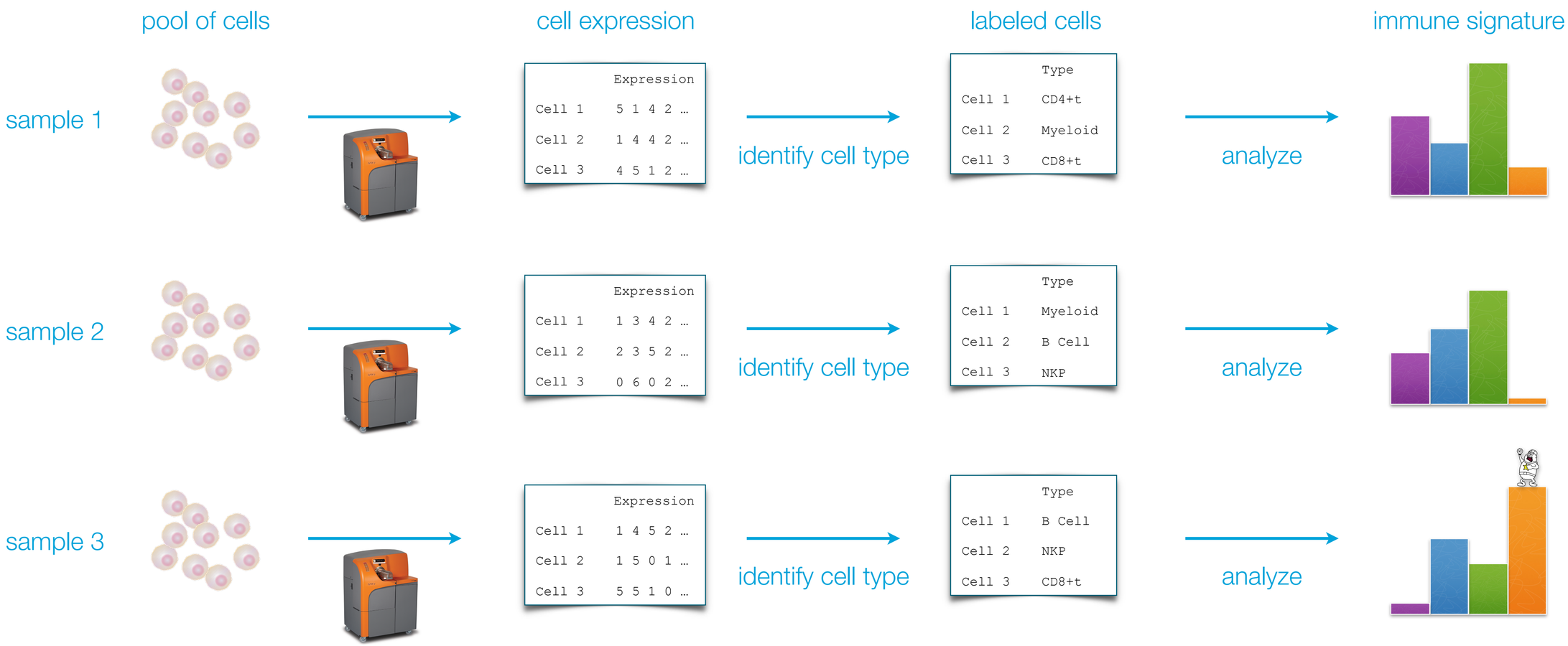


... ...

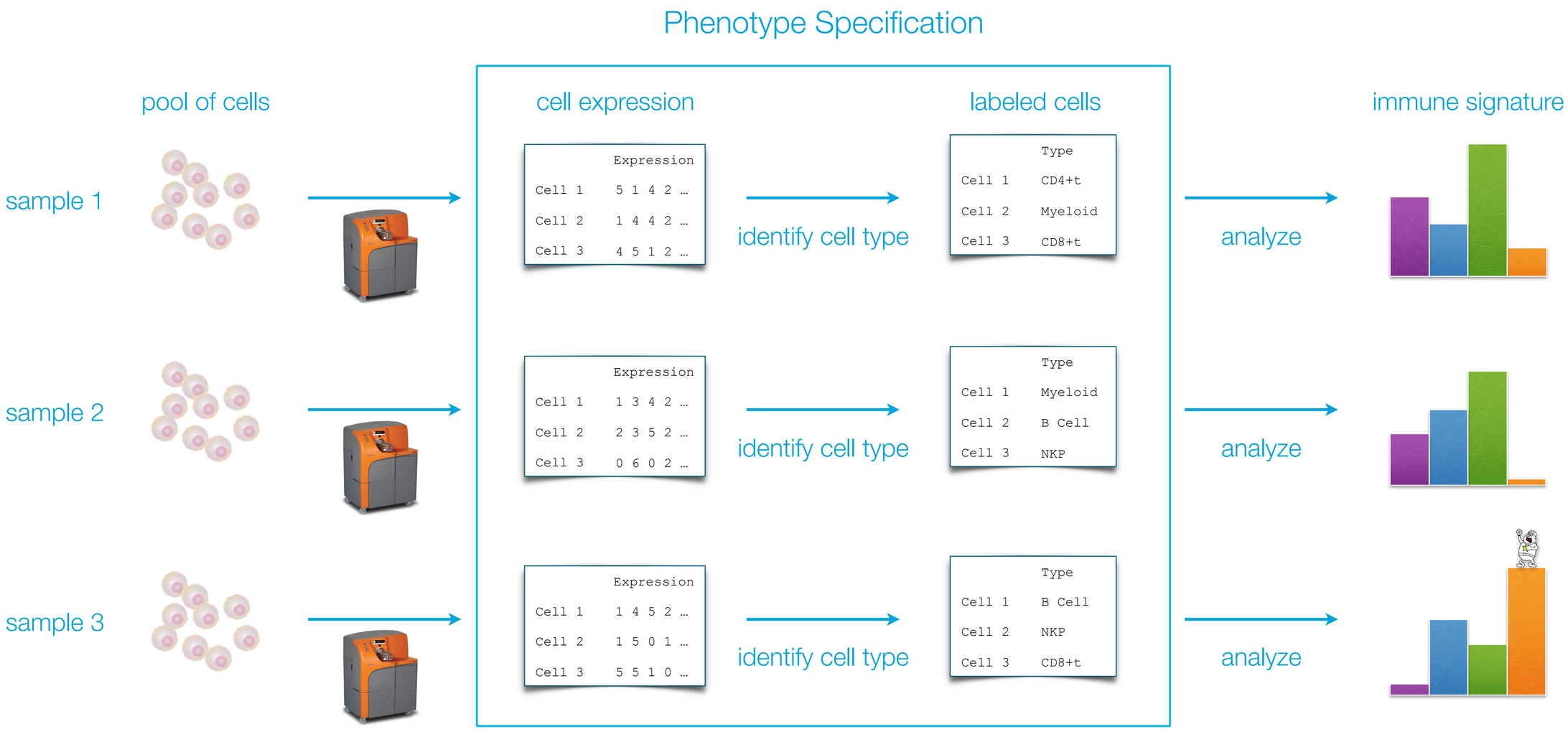
Immune System Analysis



Immune System Analysis



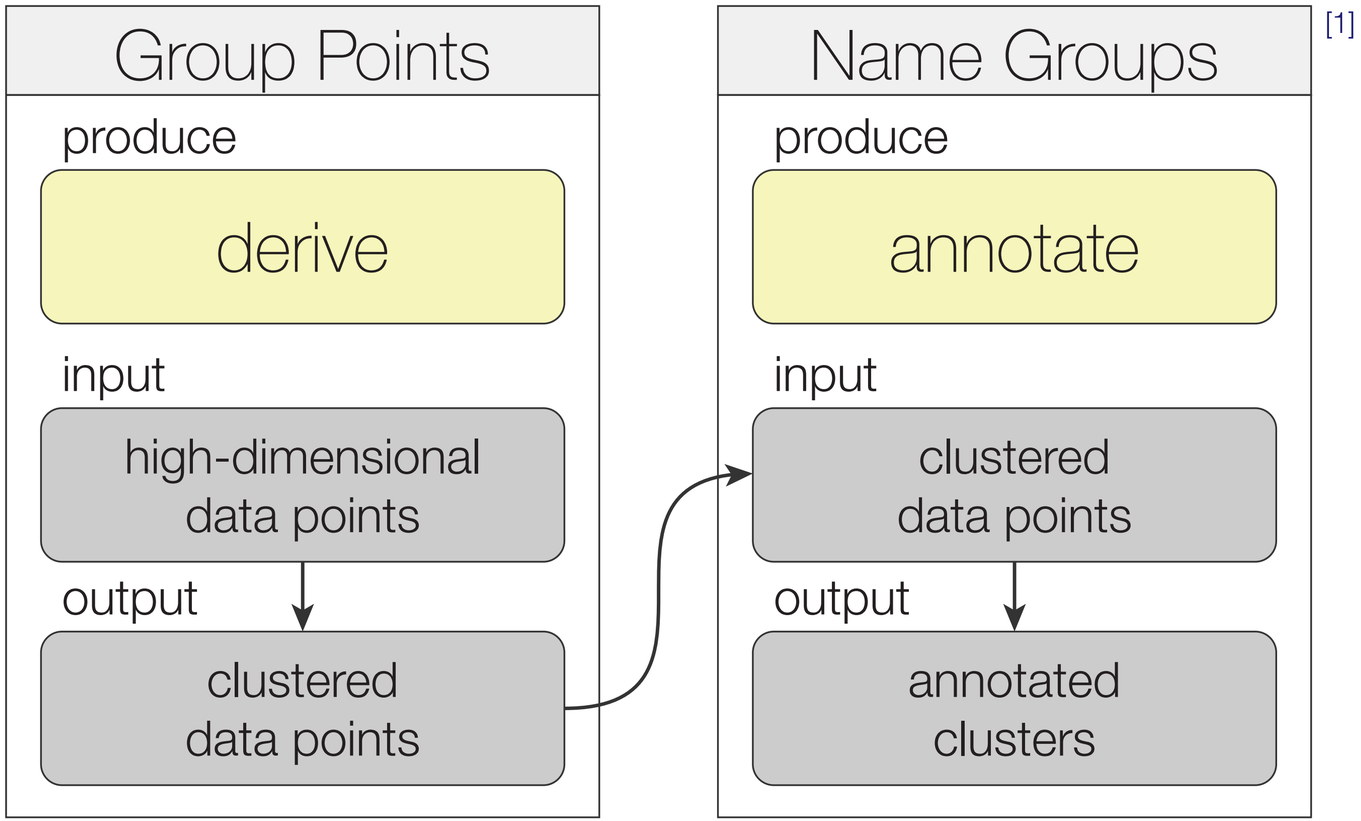
Immune System Analysis



Phenotype Specification

Tasks

- Define groups of similar cells
- Label groups based on expression profile



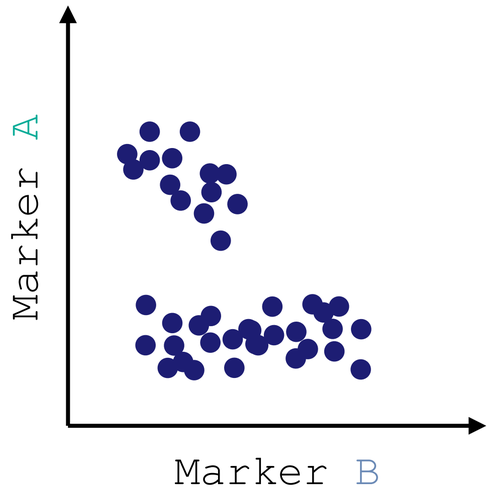
[1]

[1] Brehmer and Munzner, *A multi-level typology of abstract visualization tasks*. TVCG, 2013

2D Scatterplot Gating

- Select 2 markers

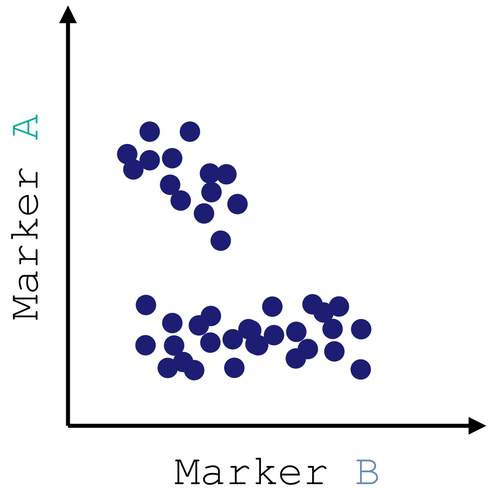
A B



2D Scatterplot Gating

- Select 2 markers

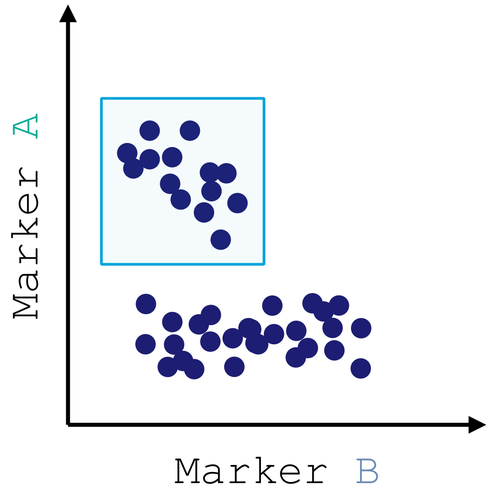
A B



2D Scatterplot Gating

- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low

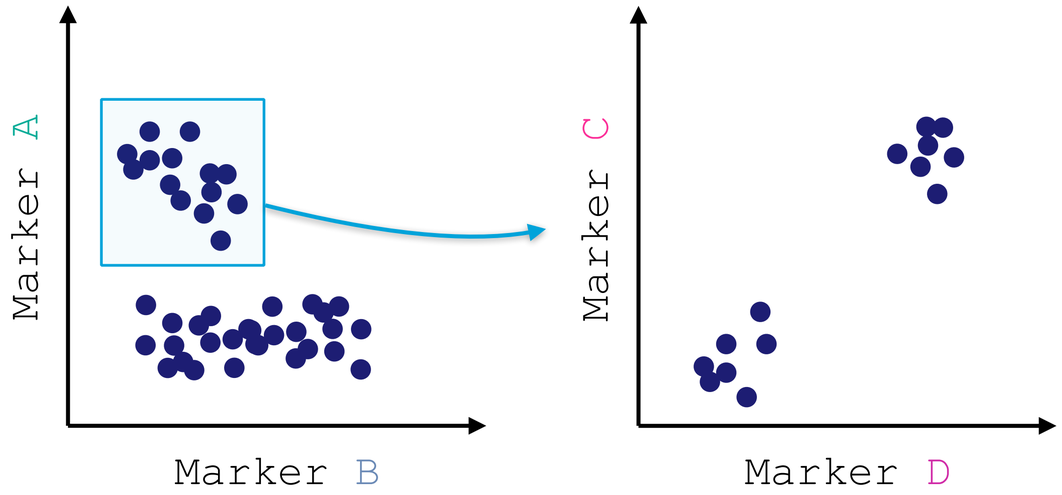
$A^+ B^-$



2D Scatterplot Gating

- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low
- Chose different markers with selection

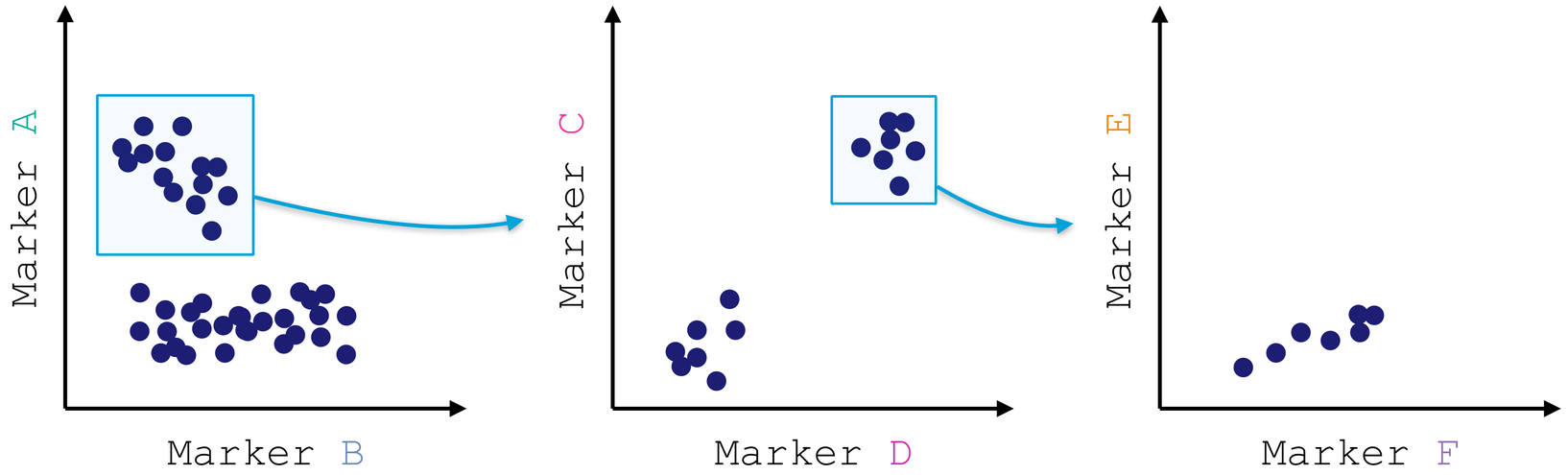
$A^{+}\ B^{-}\ C\ D$



2D Scatterplot Gating

- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low
- Chose different markers with selection

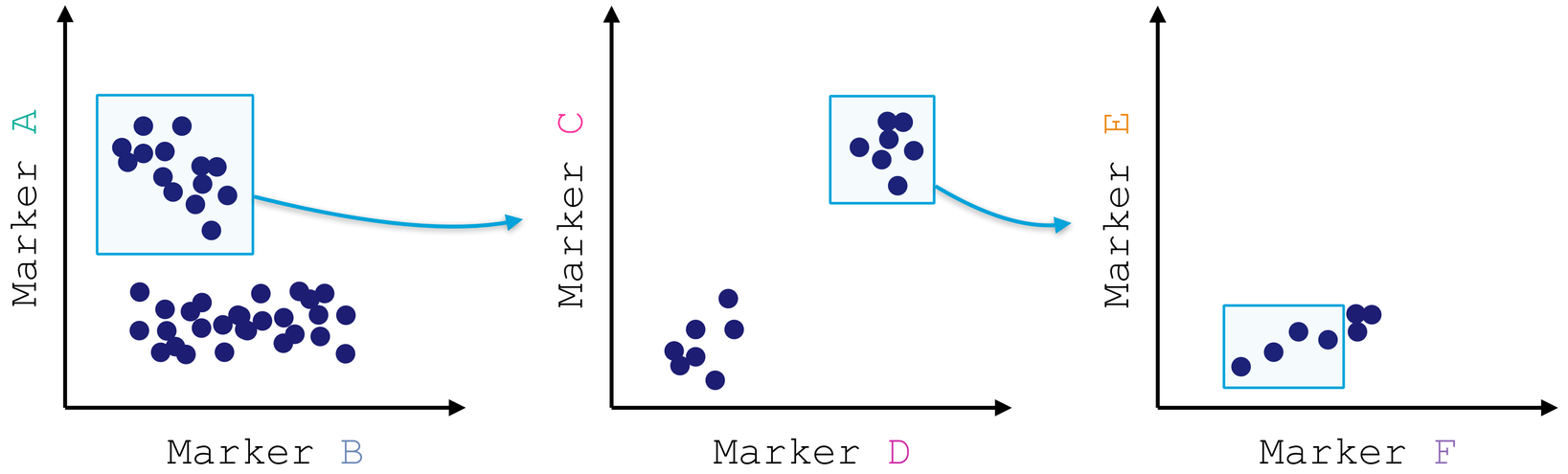
$A^{+}\ B^{-}\ C^{+}\ D^{+}\ E\ F$



2D Scatterplot Gating

- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low
- Chose different markers with selection

$A^+ B^- C^+ D^+ E^- F^-$

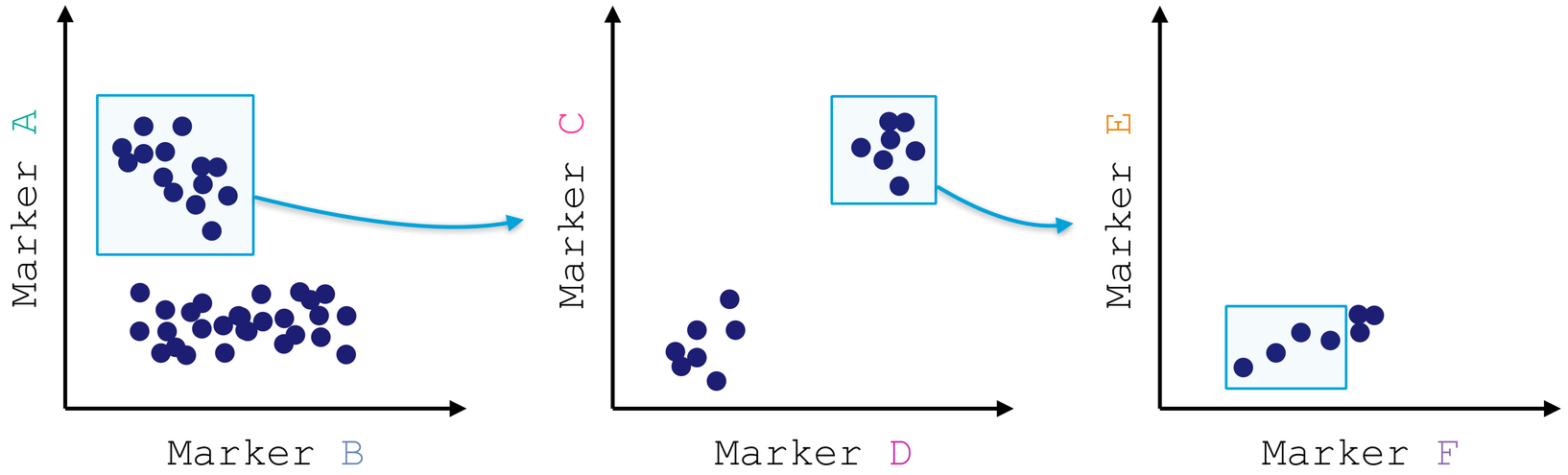


2D Scatterplot Gating

- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low
- Chose different markers with selection

$A^+ B^- C^+ D^+ E^- F^-$

⇨ Cell Type X



2D Scatterplot Gating

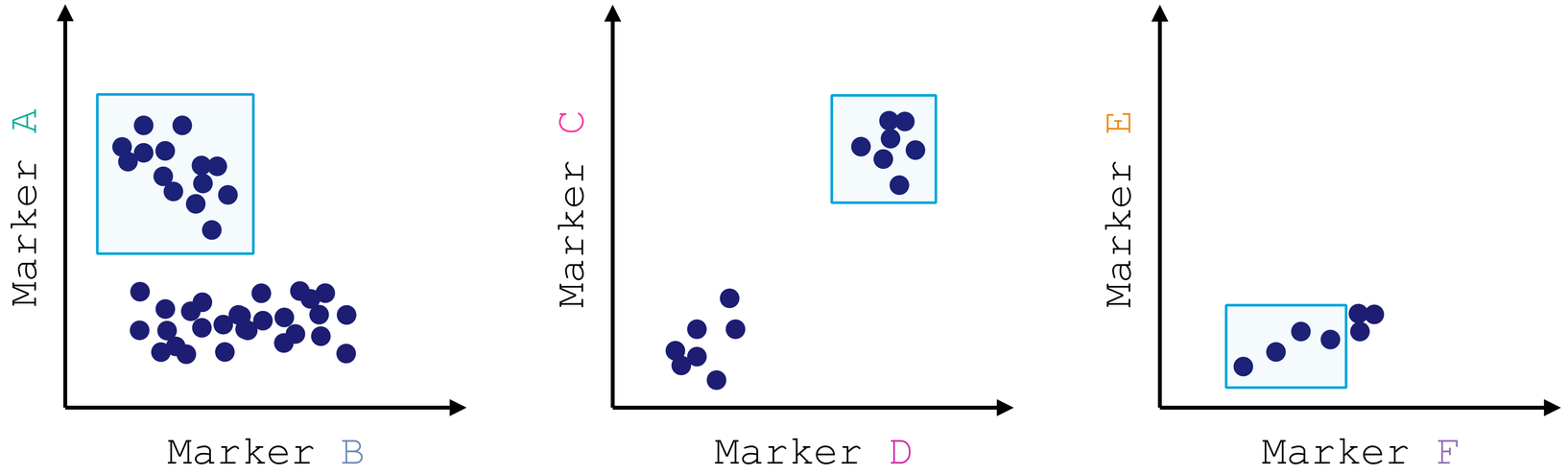
- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low
- Chose different markers with selection

$A^+ B^- C^+ D^+ E^- F^-$

⇨ Cell Type X

- Large user bias
- 40 markers

⇨ 2^{40} combinations



2D Scatterplot Gating

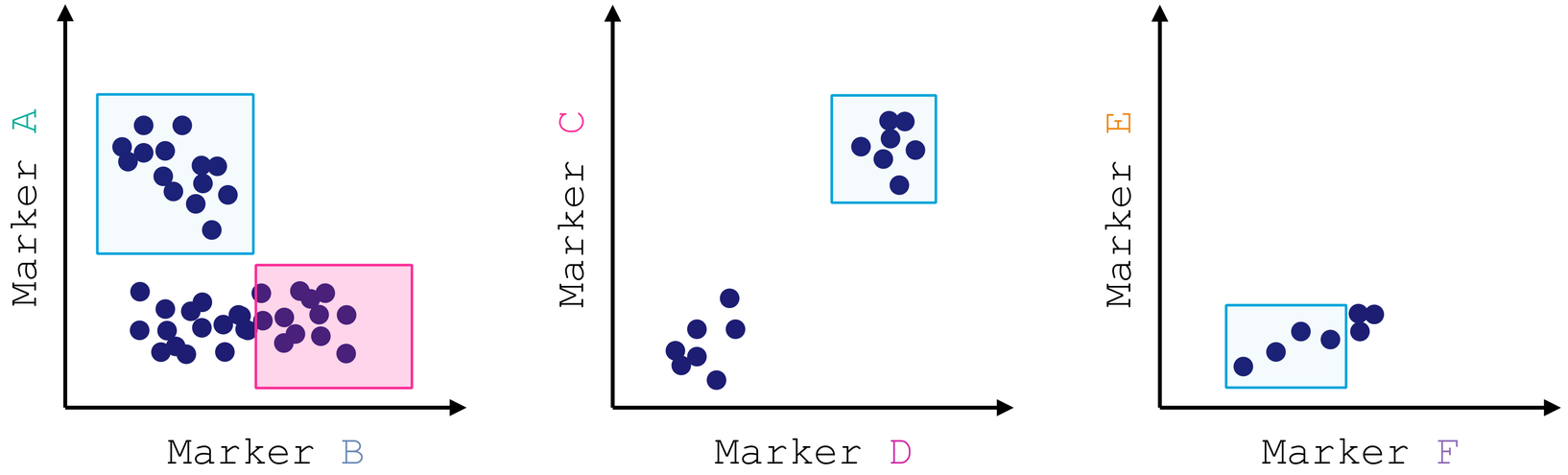
- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low
- Chose different markers with selection

$A^+ B^- C^+ D^+ E^- F^-$

⇨ Cell Type X

- Large user bias
- 40 markers

⇨ 2^{40} combinations



2D Scatterplot Gating

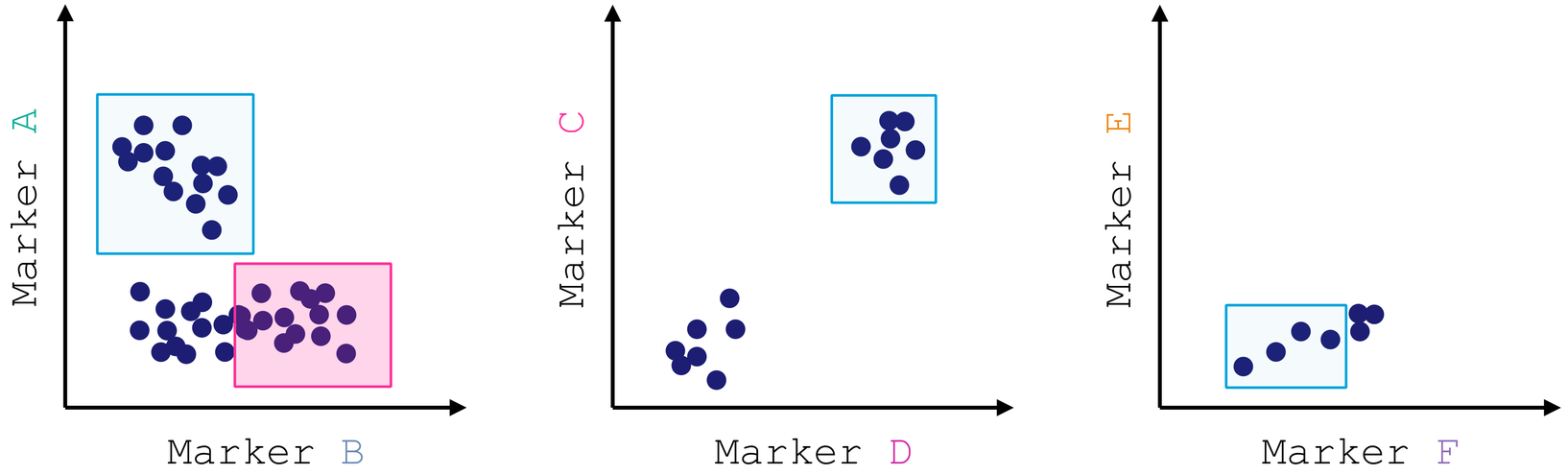
- Select 2 markers
- Select (gate) a range of 2 markers
  - E.g. high/low
- Chose different markers with selection

$A^+ B^- C^+ D^+ E^- F^-$

⇨ Cell Type X

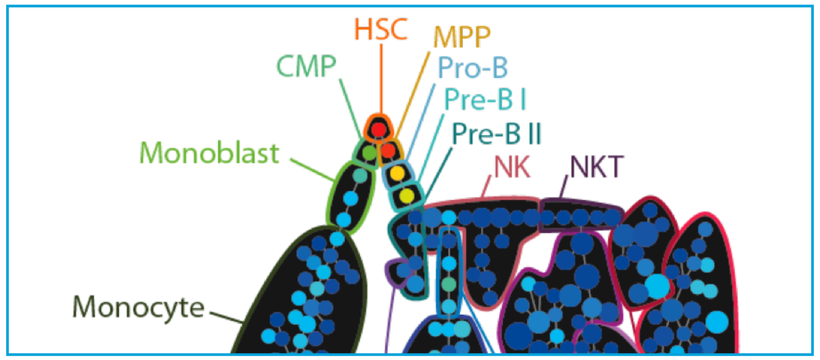
- Large user bias
- 40 markers

⇨ 2^{40} combinations

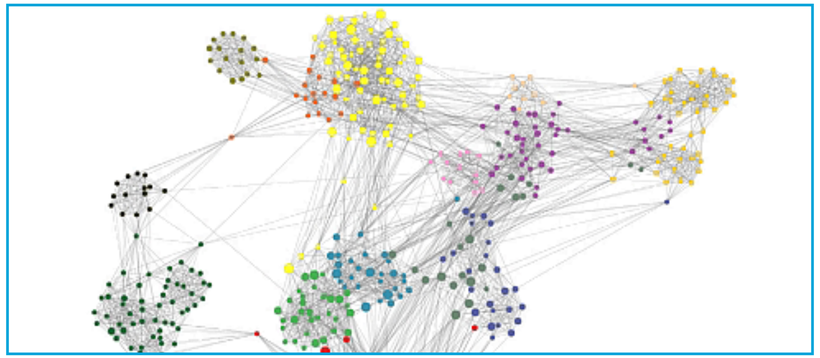


State of the Art I

Clustering-based

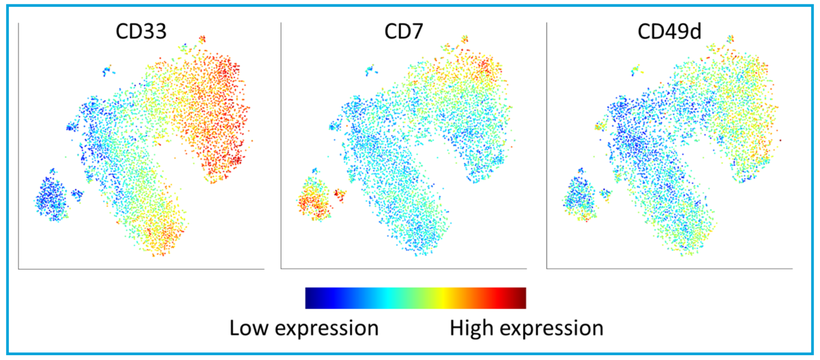


[2] Qiu et al., *Extracting a Cellular Hierarchy from High-dimensional Cytometry Data with* ***SPADE***, Nature Biotechnology, 2011.

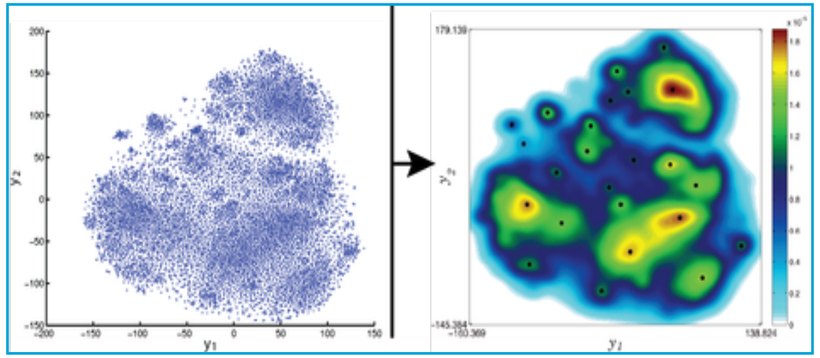


[3] Levine et al., *Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis (**Phenograph**)*, Cell 2015

Dimensionality-Reduction-based



[4] Amir et al., ***viSNE*** *enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia*, Nature Biotech. 2013.



[5] Shekhar et al., Automatic classification of cellular expression by nonlinear stochastic embedding (**ACCENSE**), PNAS, 2014

State of the Art II

Clustering-based

+ Fast
+ Handle large data

- Less precise than tSNE
- Less control

Dimensionality-Reduction-based

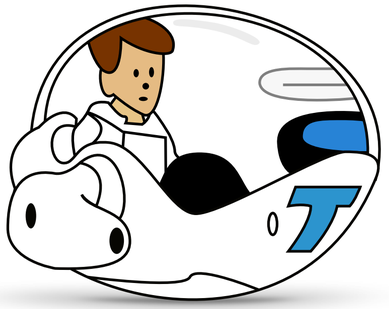
+ State of the art precision
+ Access to single data points

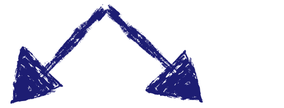
- Slow
- Limited data size

Mass Cytometry

- ~40 markers // 10.000 immune-system-wide proteins
- Cohort studies ⇨ Millions of Cells
- More markers lead to
  - More depth
    ⇨ Identification of unknown cell types
  - More breadth
    ⇨ Exploration of multiple lineages

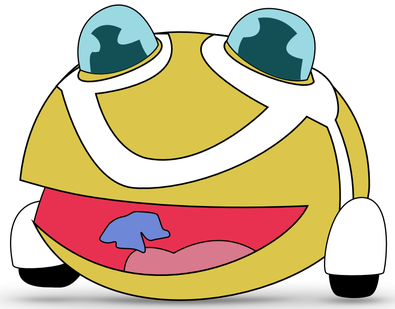
T-Lymphocytes





... ...

Macrophage

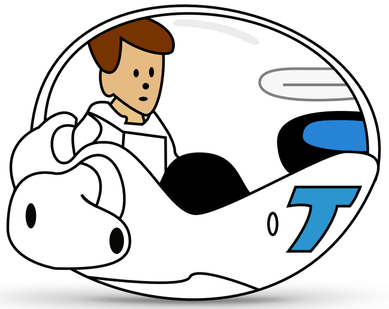


... ...

Mass Cytometry

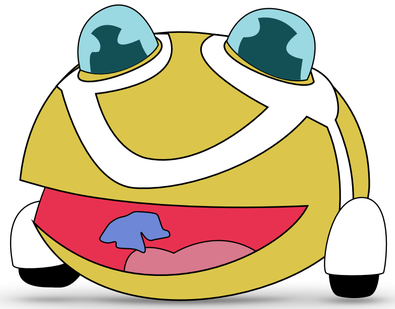
- ~40 markers // 10.000 immune-system-wide proteins
- Cohort studies ⇨ Millions of Cells
- More markers lead to
  - More depth
    ⇨ Identification of unknown cell types
  - **More breadth**
    ⇨ **Exploration of multiple lineages**

T-Lymphocytes



... ...

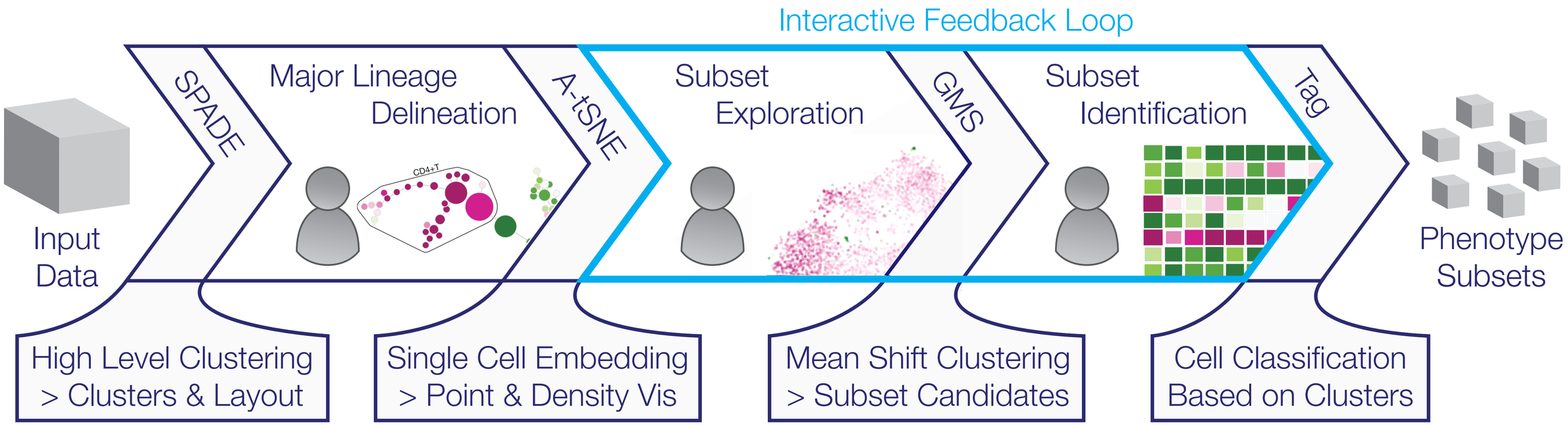
Macrophage



... ...

Multi-Level Workflow

Multi-Level Workflow

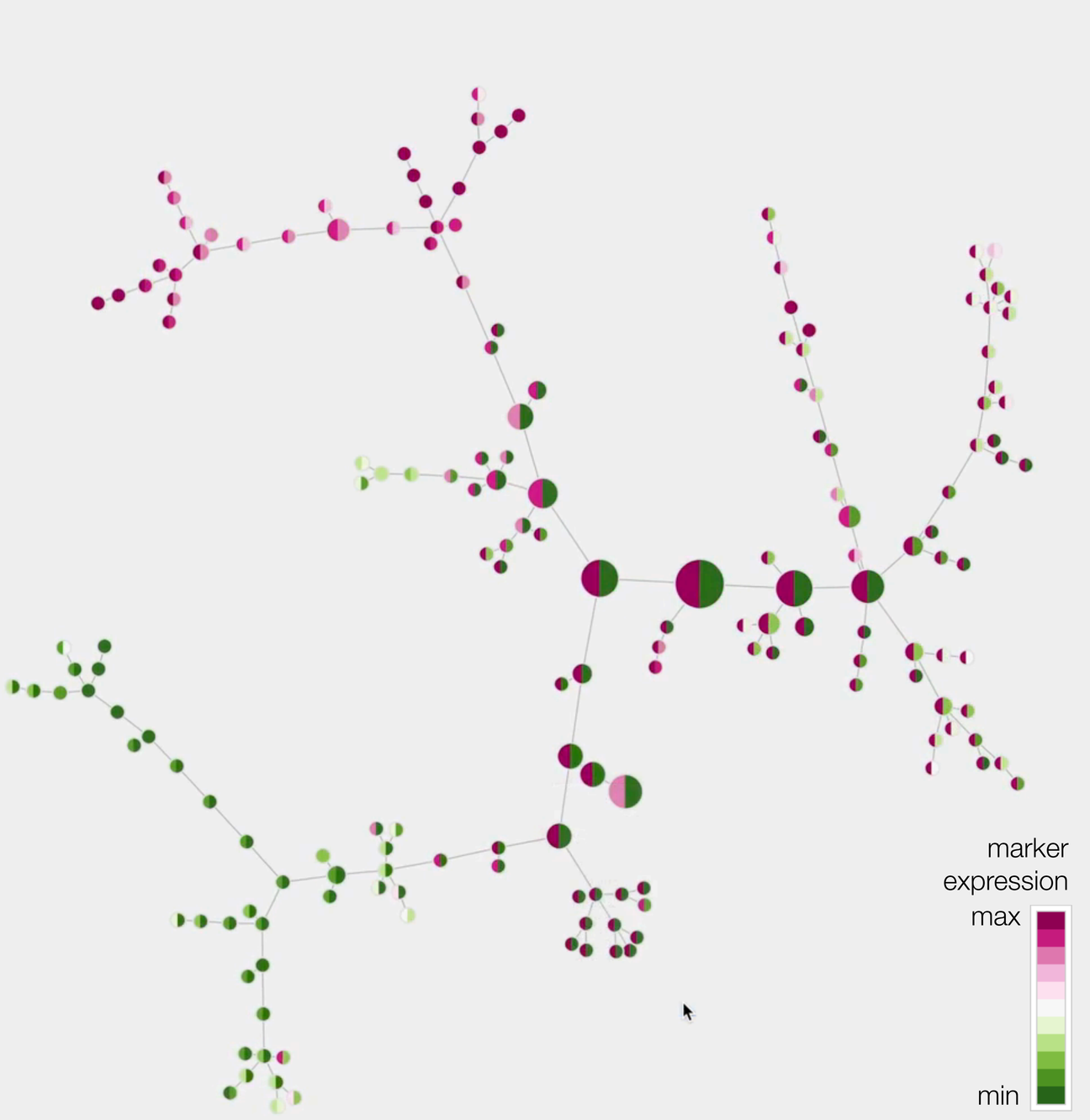


SPADE I

- Nodes represent clusters

  (ideally cells of a single phenotype)

  - Size: number of cells
  - Color: marker expression
- Links based on MST



A-tSNE

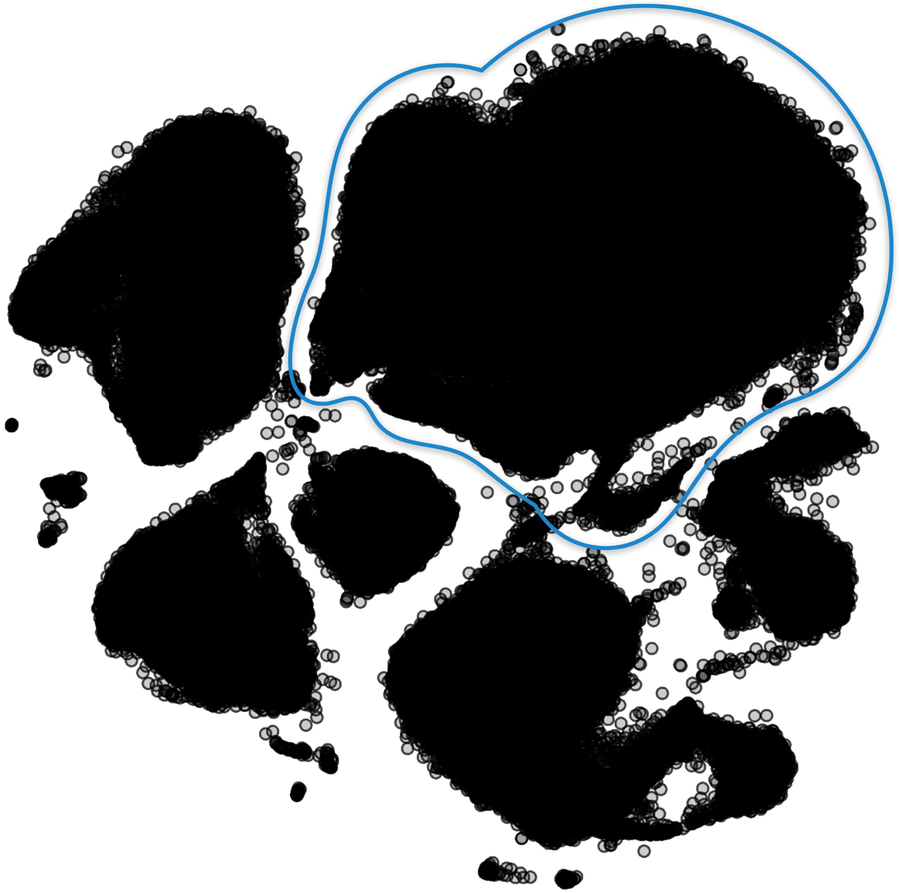
- tSNE[6] optimized to
  - Preserve local structure (clusters)
  - Target 2D or 3D spaces
- A(pproximated)-tSNE[7]
  - Reduces pre-processing time

[6] v.d. Maaten & Hinton, *Visualizing High-Dimensional Data Using t-SNE,* Journal of Machine Learning Research, 2008

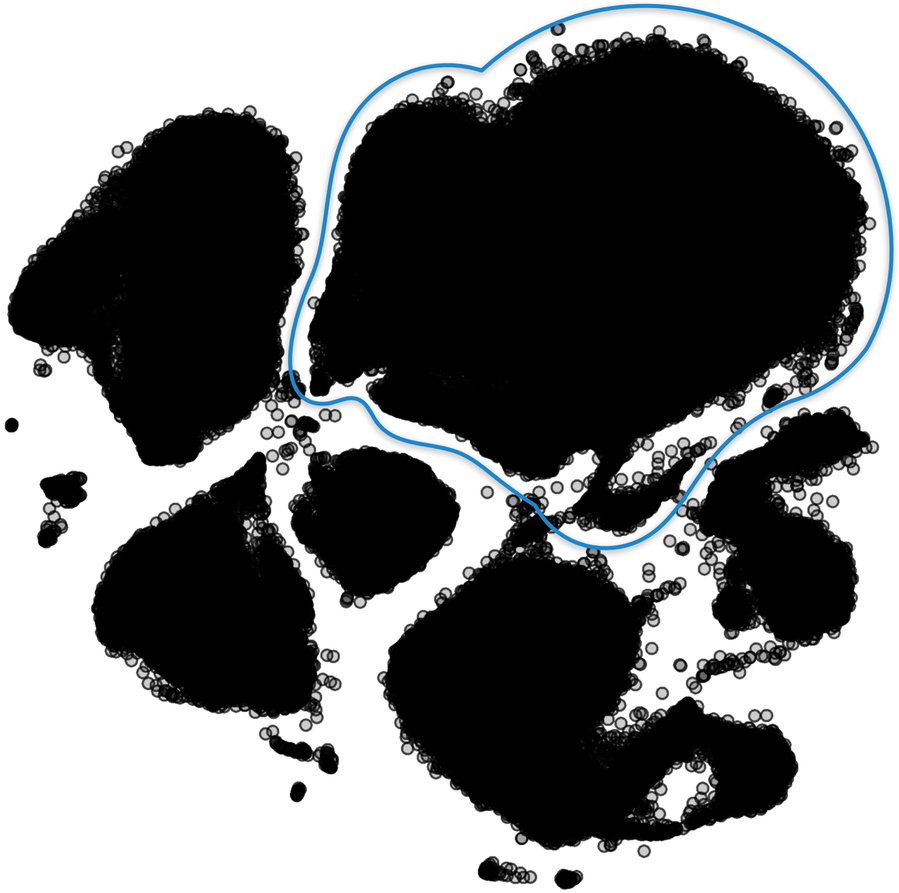
[7] Pezzotti et al., *Approximated and User Steerable tSNE for Progressive Visual Analytics,* TVCG, 2016

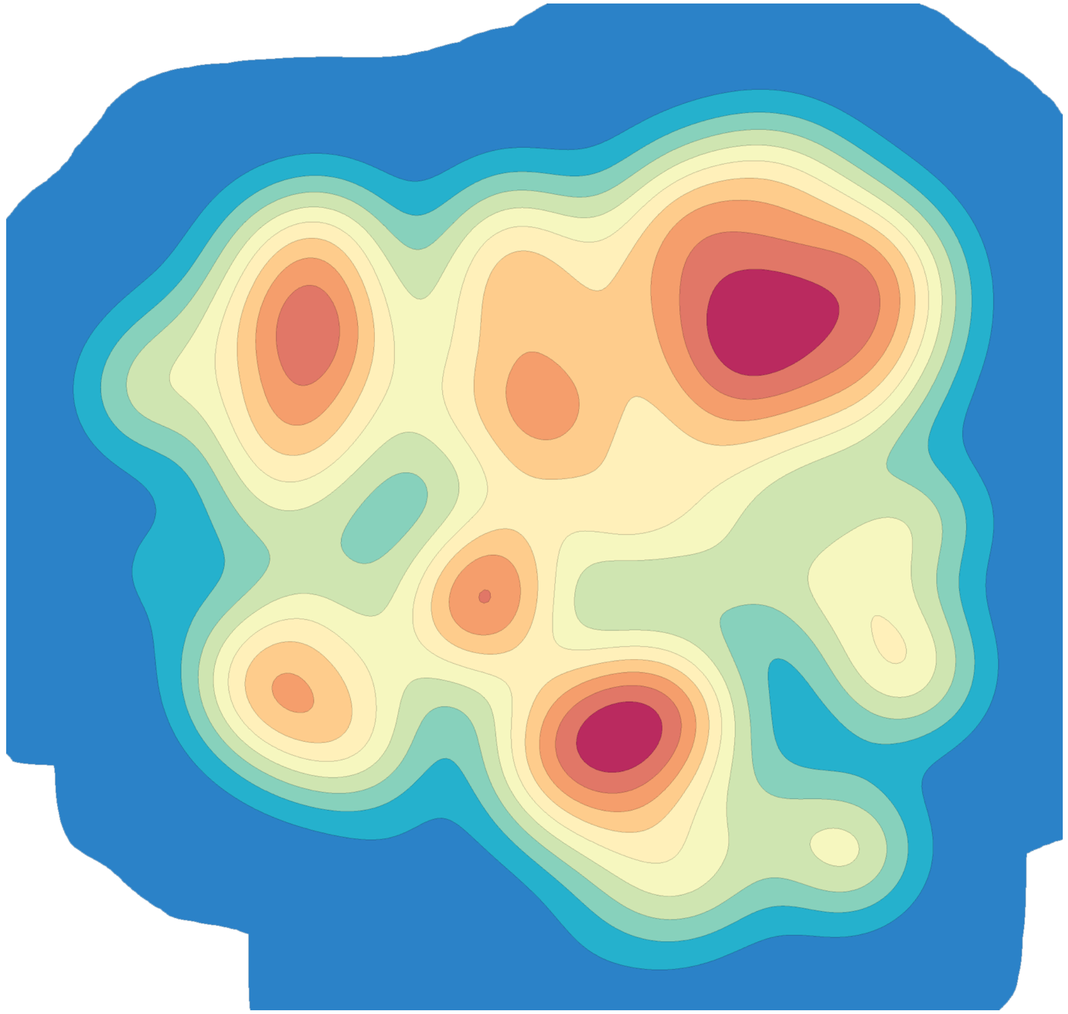
Analyzing tSNE Embedding

Analyzing tSNE Embedding

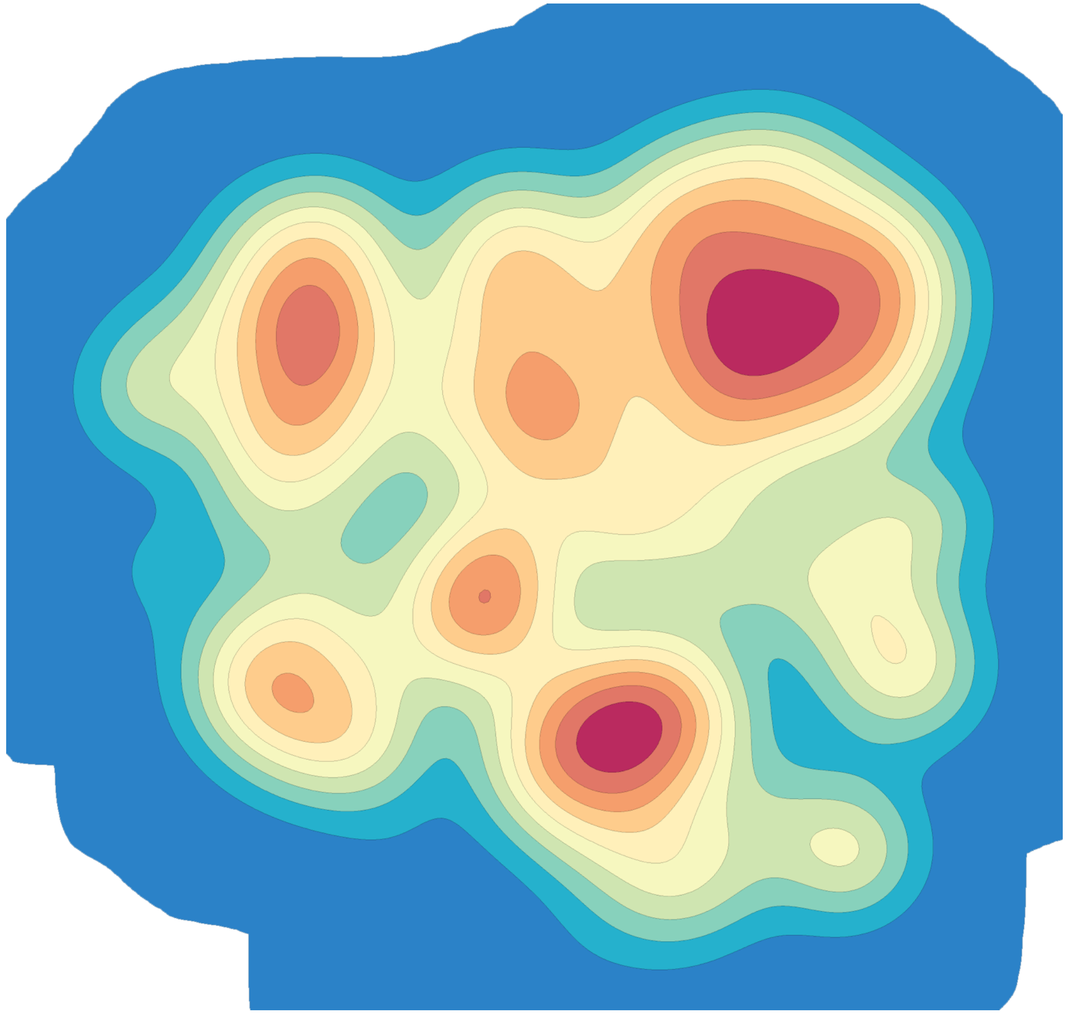


Analyzing tSNE Embedding

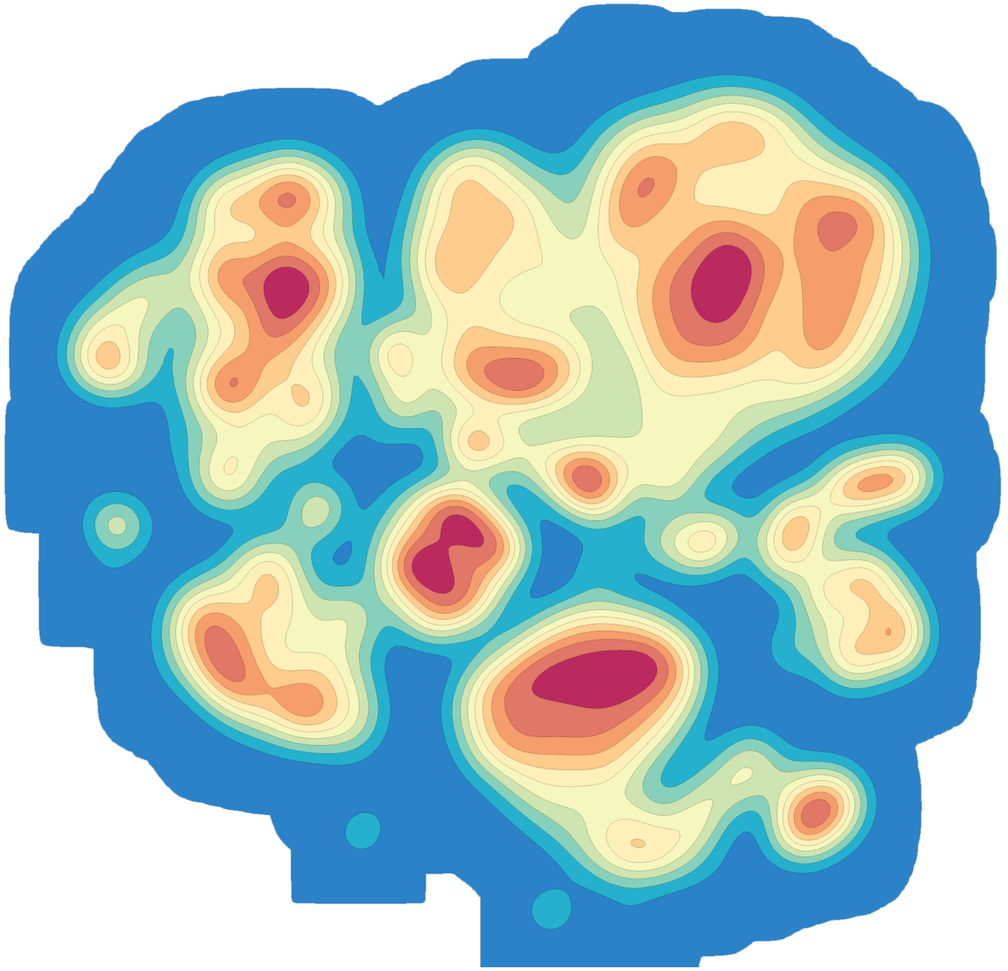




Analyzing tSNE Embedding



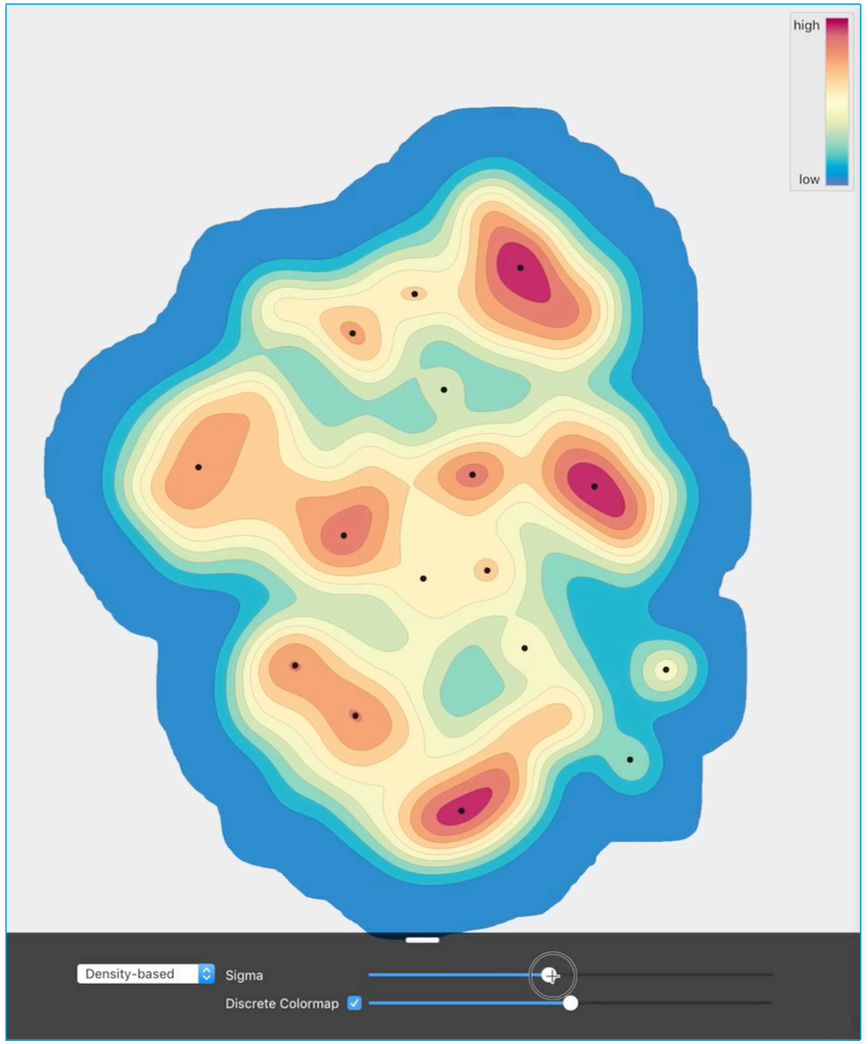
Analyzing tSNE Embedding



Interactive Mean Shift Clustering

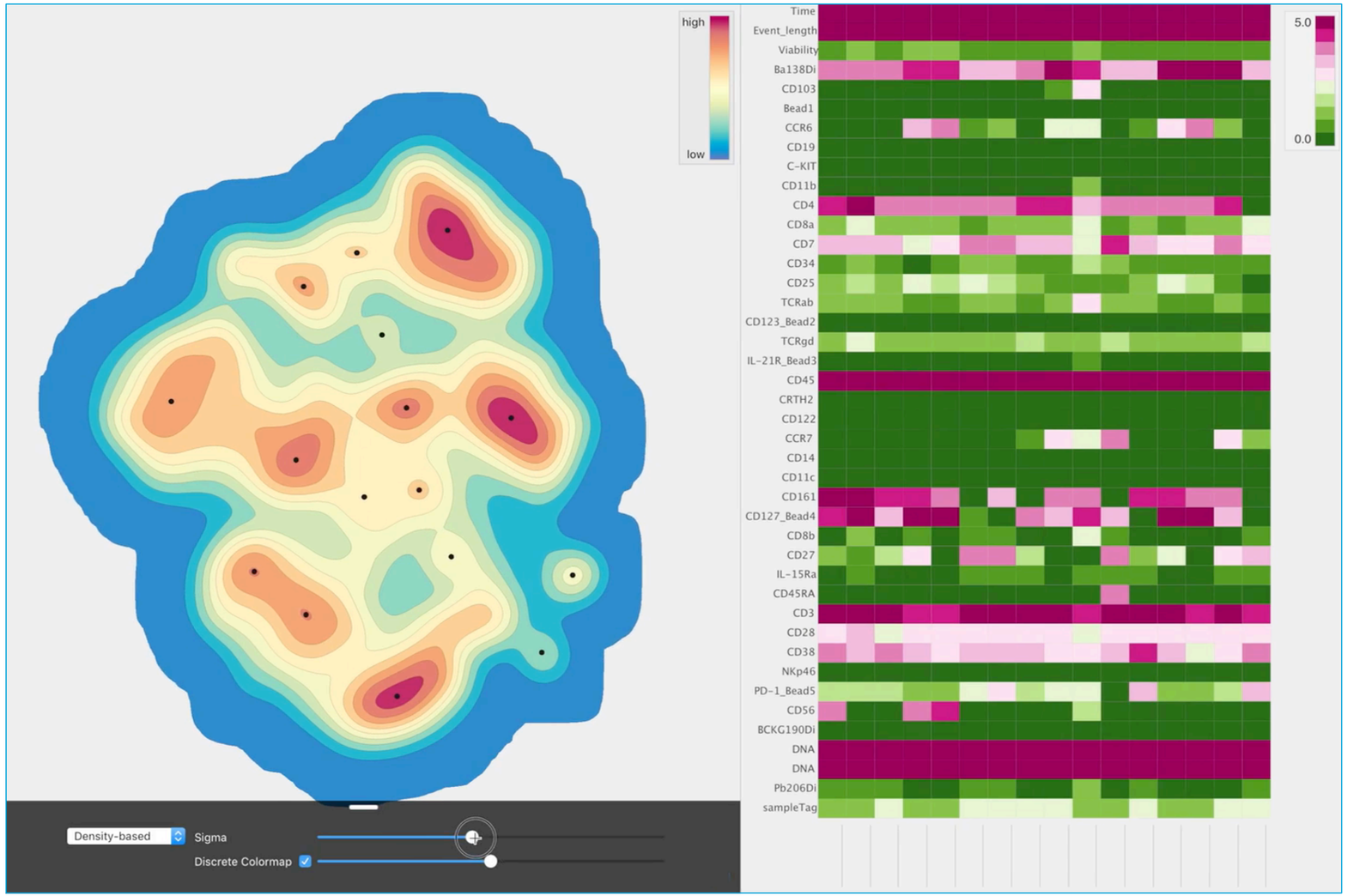
- GPU-based
  - Density estimation[8]
  - Image-based Mean Shift segmentation
  - Interactive rates

    <100ms for hundreds of thousand data points
- On-the-fly visualization in heatmap



[8] Lampe and Hauser, *Interactive visualization of streaming data with kernel density estimation*. IEEE Pacific Vis, 2011

Interactive Mean Shift Clustering



Linked Heatmap

Results

Study

- Prototype study shows effectiveness of workflow[9]
- 3x speedup on small test case

	Total	Lineage Delineation	Subset Computation	Subset Processing
Protoype[9]	108 min	27 min	29 min	52 min
Ours	39 min	13 min	11 min	15 min

[9] van Unen et al., Tissue- and disease-specific signatures of the human mucosal immune system identified by mass cytometry. Immunity, 2016

Study

- Prototype study shows effectiveness of workflow[9]
- 3x speedup on small test case

	Total	Lineage Delineation	Subset Computation	Subset Processing
Protoype[9]	108 min	27 min	29 min	52 min
Ours	39 min	13 min	11 min	15 min

[9] van Unen et al., Tissue- and disease-specific signatures of the human mucosal immune system identified by mass cytometry. Immunity, 2016

Study

- Prototype study shows effectiveness of workflow[9]
- 3x speedup on small test case

	Total	Lineage Delineation	Subset Computation	Subset Processing
Protoype[9]	108 min	27 min	29 min	52 min
Ours	39 min	13 min	11 min	15 min

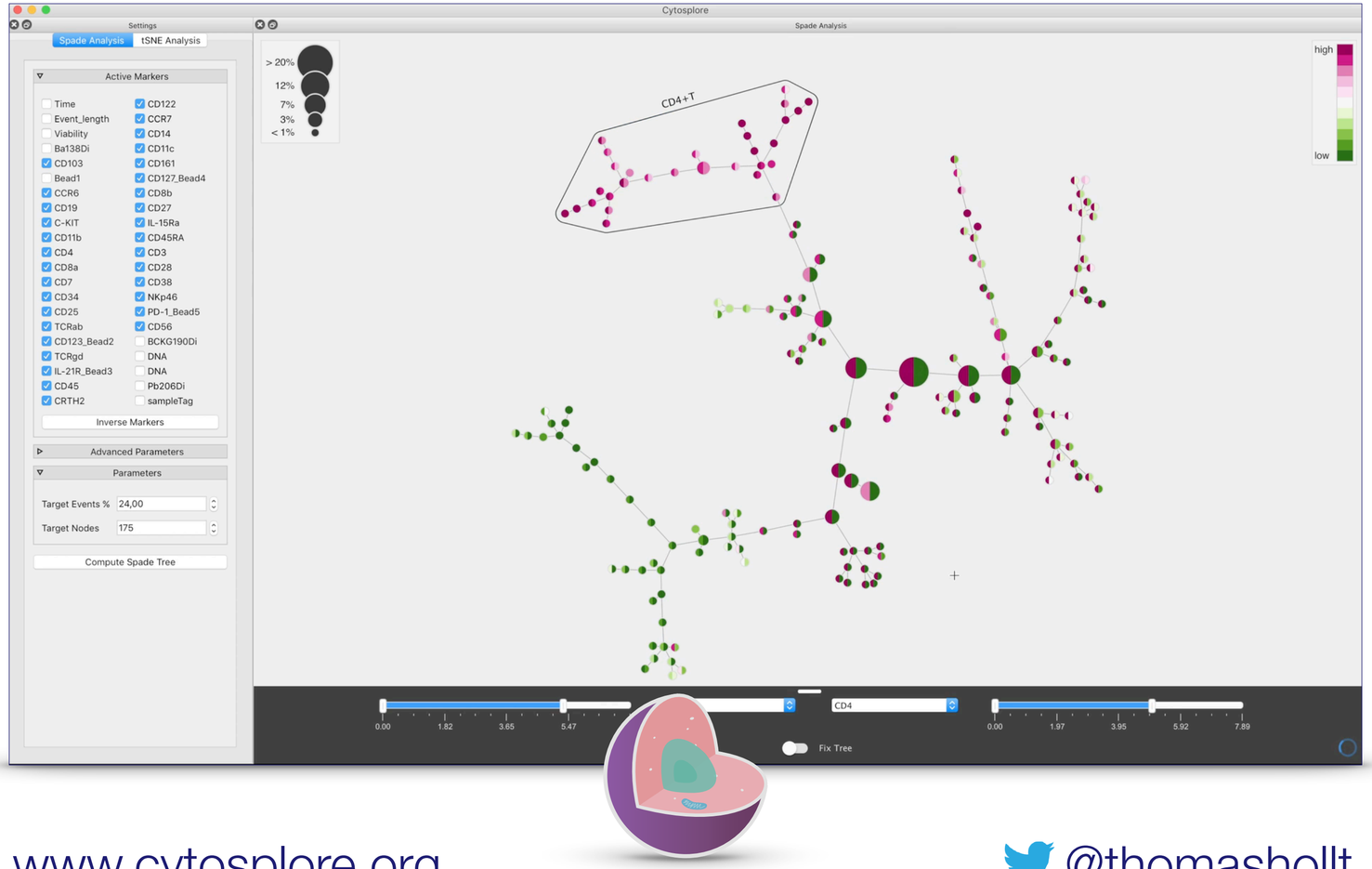
[9] van Unen et al., Tissue- and disease-specific signatures of the human mucosal immune system identified by mass cytometry. Immunity, 2016

Conclusion

- We present a framework for phenotyping large single cell datasets
  - Our multilevel workflow allows the analysis of larger data than before
  - Our user study shows the advantages of our integrated workflow

- Next:
  - Explore speficially designed multi level analysis techniques
  - Extend Cytosplore to support the immune signature analysis

Time for Questions!



www.cytosplore.org

@thomashollt

connecting innovators
Project 12720 VanPlRe

EuroVis 2016