HAU901I

Single-cell transcriptomics

Additional Algorithms

Pr Jacques Colinge

UM/Inserm - IRCM - ICM

Overview

- Single-cell transcriptomics data are potentially very rich and depending on the research project many specialized analyses can be performed
- In these additional slides, we cover two common such analyses:
  - The inference of ligand-receptor interactions between cells (cellular network)
  - The inference of a pseudo-time to follow the differentiation process between related cell populations
- Different tools and approaches exist for the two general questions above, we only one solution for each
- Examples of (uncovered) other analyses: estimation of copy number variation in individual cancer cells and detection of subclones, prediction of the transcription factors activated in each population, etc.

Pseudo-time type-of analyses

- Classical 2D-projections tend to magnify differences between different cells (in different clusters), *e.g.*, t-SNE, or rely on linear relationships between individual transcriptomes, *e.g.*, PCA

- Different authors have proposed methods to reduce data dimensionality such that distances in the projection are close to the distances in the original, high-dimensional space

- For cells that are typically related by a differentiation process such as hematopoeisis, these methods tend to organize the cells along a curve and positions along this curve are related to a pseudo-time representing the stages of differentiation

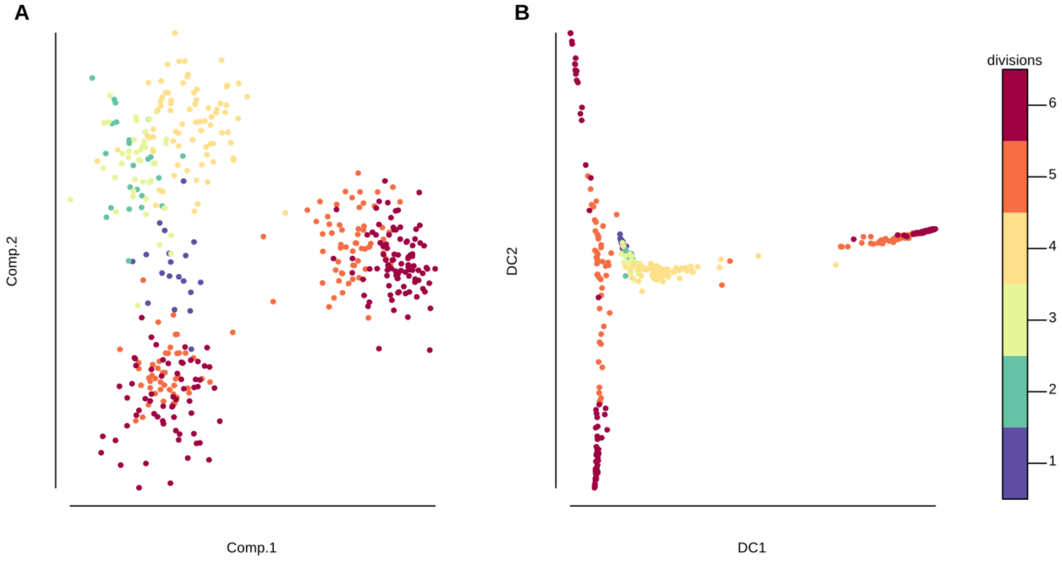
Diffusion maps

- Diffusion maps (DM) (Coifman, PNAS, 2005) originates from the field of manifold dimension reduction like UMAP. It relies on advanced mathematics that are irrelevant here.
- Compared to UMAP, DM better preserve the original distances in the projected space. You can use the Bioconductor package «destiny» that provides a fast implementation adapted to single-cell data (Angerer, Bioinformatics, 2016).
- DM output contains multiple diffusion coordinates sorted in decreasing order of importance, and using the first two provides with a 2D-projection

Six cycles of cell division from zygote to blastocyte (Guo, Dev Cell, 2010).

Left: PCA ; Right: DM

Zygotes lead to different, more differentiated cell types.



Pseudo-time

- DM do not determine any pseudo-time, they are just «more compatible» with this notion than common dimensionalty reduction techniques
- To actually get pseudo-time line(s), we can use the package «Slingshot» (Street, BMC Genomics, 2018)
- Slingshot does not require DM specifically, it can accomodate different upfront dimensionality reductions

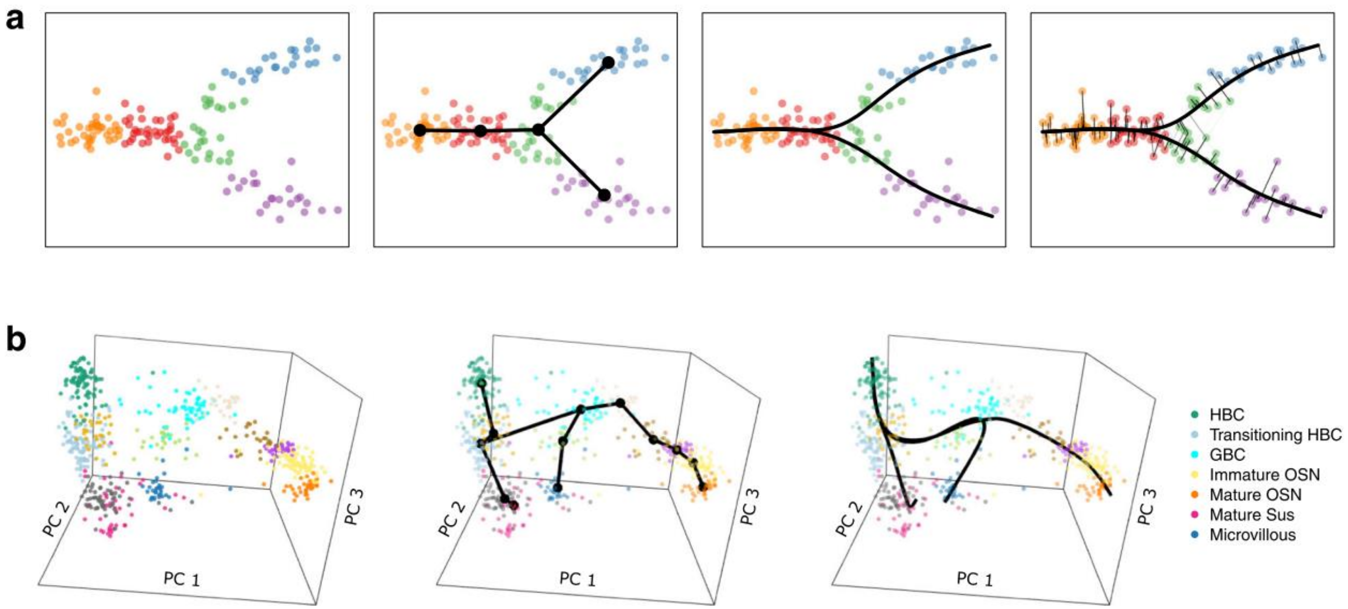
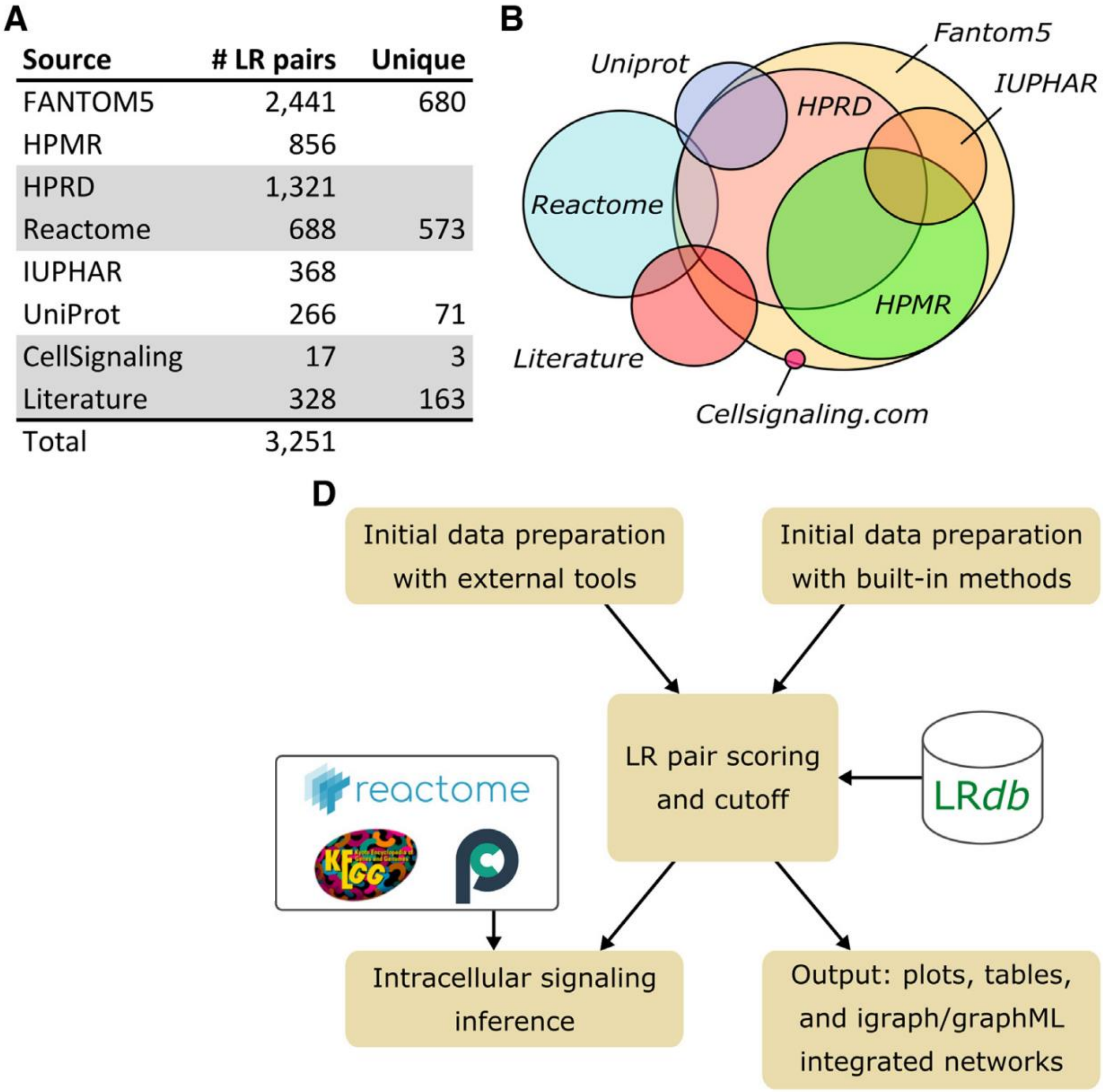


Fig. 1 Schematics of Slingshot's main steps. The main steps for Slingshot are shown for: Panel (**a**) a simple simulated two-lineage two-dimensional dataset and Panel (**b**) the single-cell RNA-Seq olfactory epithelium three-lineage dataset of [26] (see Results and discussion for details on dataset and its analysis). Step 0: Slingshot starts from clustered data in a low-dimensional space (cluster labels indicated by color). For Panel (**b**), the plot shows the top three principal components, but Slingshot was run on the top five. Step 1: A minimum spanning tree is constructed on the clusters to determine the number and rough shape of lineages. For Panel (**b**), we impose some constraints on the MST based on known biology. Step 2: Simultaneous principal curves are used to obtain smooth representations of each lineage. Step 3: Pseudotime values are obtained by orthogonal projection onto the curves (only shown for Panel (**a**))

Cellular networks

- The majority of interactions between cells involve ligand-receptor interactions (LRIs)
- A large number of LRIs are known and can be compiled in a database
- Based on each cell population average expression profiles, it is possible to check whether the ligand and the receptor of a known LRI are expressed
- Based on this idea, we can infer LRIs
- In practice, we need a notion of score to control the false positive rate
- Most tools focus on ligands and receptors that are significantly expressed by one or several populations compared to all the populations of cells
- The real significance of the LRI itself is usually not evaluated!

SingleCellSignalR



A

Source	# LR pairs	Unique
FANTOM5	2,441	680
HPMR	856	
HPRD	1,321	
Reactome	688	573
IUPHAR	368	
UniProt	266	71
CellSignaling	17	3
Literature	328	163
Total	3,251	

D

Initial data preparation with external tools

Initial data preparation with built-in methods

LR pair scoring and cutoff

LRdb

Intracellular signaling inference

Output: plots, tables, and igraph/graphML integrated networks

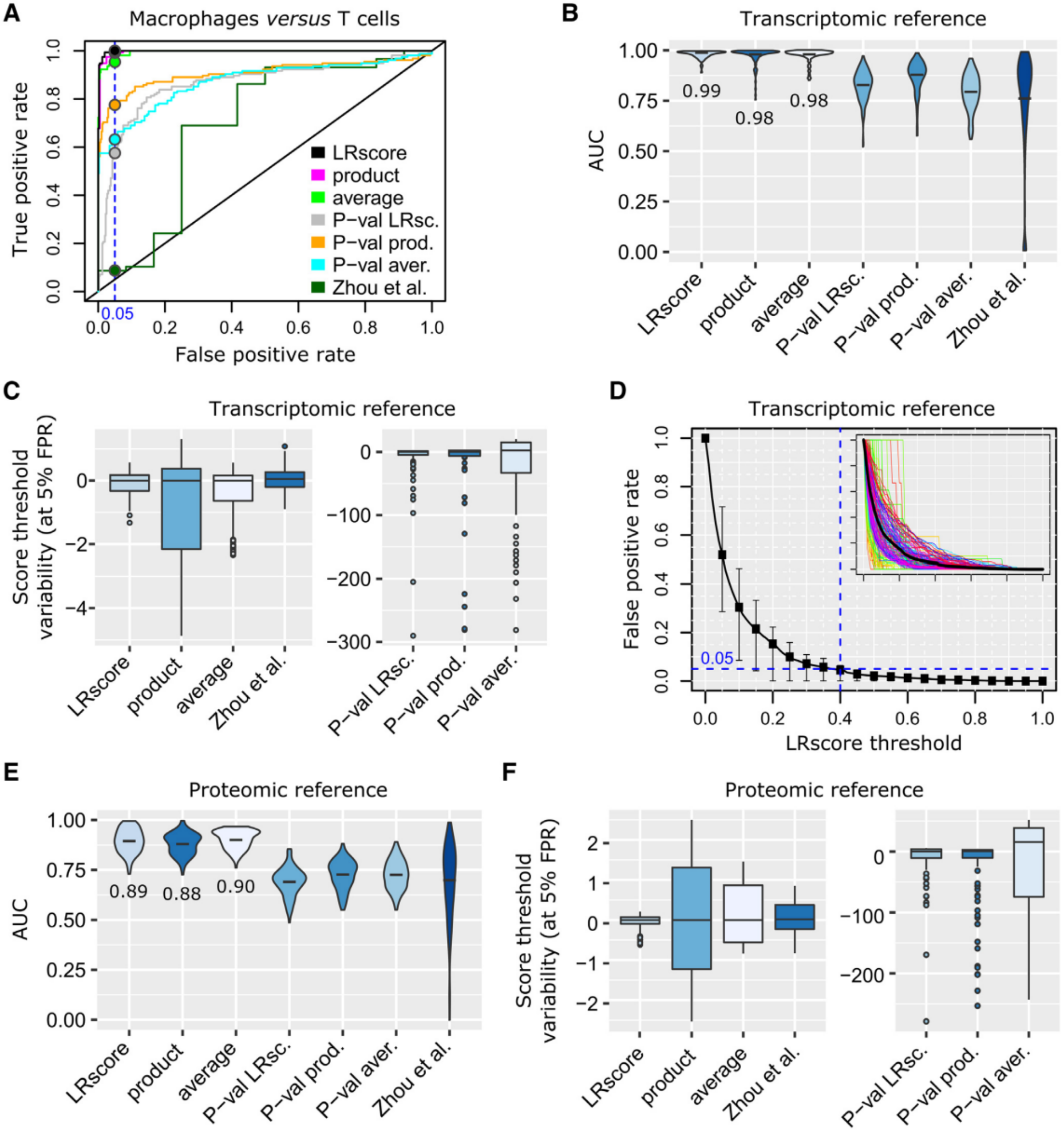
SingleCellSignalR

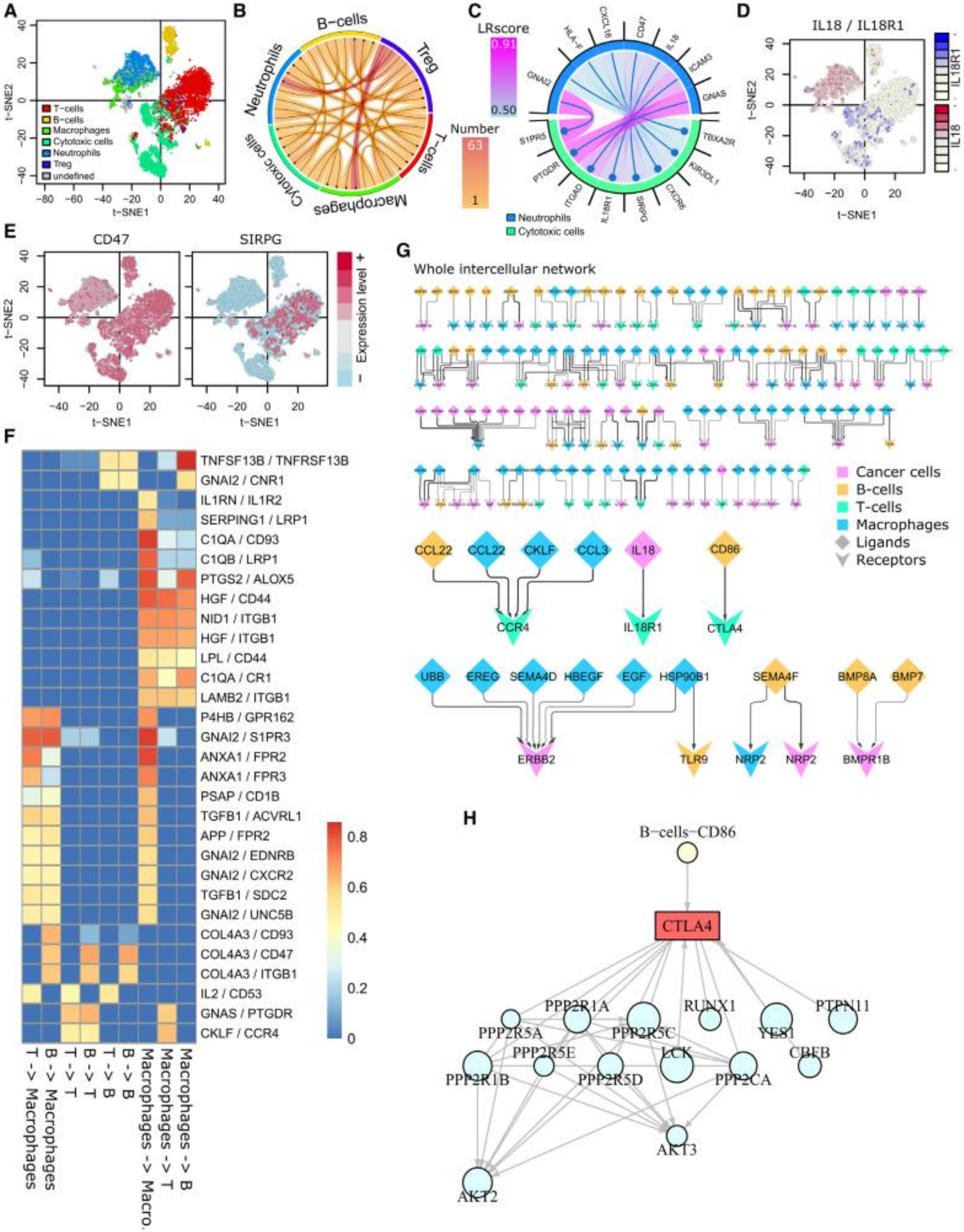
Normalized and log-transformed expression matrix

$$\mathrm{LRscore} = \frac{\sqrt{lr}}{\mu + \sqrt{lr}}$$

μ=matrix average
l=ligand expression
r=receptor expression

ROC curves by comparing with experimental data from expression proteomics and transcriptomics




A
t-SNE2
t-SNE1
T-cells
B-cells
Macrophages
Cytotoxic cells
Neutrophils
Treg
undefined
B
B-cells
Treg
T-cells
Macrophages
Cytotoxic cells
Neutrophils
Number
63
1
C
LRscore
0.91
0.50
HLA-F
CXCL16
CD47
IL18
ICAM3
GNAS
GNAI2
S1PR5
PTGDR
ITGAD
IL18R1
SIRPG
CXCR6
KIR3DL1
TBXA2R
Neutrophils
Cytotoxic cells
D
IL18 / IL18R1
IL18R1
IL18
E
CD47
SIRPG
Expression level
+
-
F
TNFSF13B / TNFRSF13B
GNAI2 / CNR1
IL1RN / IL1R2
SERPING1 / LRP1
C1QA / CD93
C1QB / LRP1
PTGS2 / ALOX5
HGF / CD44
NID1 / ITGB1
HGF / ITGB1
LPL / CD44
C1QA / CR1
LAMB2 / ITGB1
P4HB / GPR162
GNAI2 / S1PR3
ANXA1 / FPR2
ANXA1 / FPR3
PSAP / CD1B
TGFB1 / ACVRL1
APP / FPR2
GNAI2 / EDNRB
GNAI2 / CXCR2
TGFB1 / SDC2
GNAI2 / UNC5B
COL4A3 / CD93
COL4A3 / CD47
COL4A3 / ITGB1
IL2 / CD53
GNAS / PTGDR
CKLF / CCR4
0.8
0.6
0.4
0.2
0
T -> Macrophages
B -> Macrophages
T -> T
B -> T
T -> B
B -> B
Macrophages -> Macro.
Macrophages -> T
Macrophages -> B
G
Whole intercellular network
Cancer cells
B-cells
T-cells
Macrophages
Ligands
Receptors
CCL22
CCL22
CKLF
CCL3
IL18
CD86
CCR4
IL18R1
CTLA4
UBB
EREG
SEMA4D
HBEGF
EGF
HSP90B1
SEMA4F
BMP8A
BMP7
ERBB2
TLR9
NRP2
NRP2
BMPR1B
H
B-cells-CD86
CTLA4
PPP2R1A
RUNX1
PTPN11
PPP2R5A
PPP2R5C
YES1
PPP2R5E
LCK
CBFB
PPP2R1B
PPP2R5D
PPP2CA
AKT3
AKT2