The single cell transcriptome in time and space

Technical Journal Club

Manfredi Carta

26.11.2019

nature

Letter | Published: 08 August 2018

RNA velocity of single cells

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson ✉ & Peter V. Kharchenko ✉

Nature **560**, 494–498(2018) | Cite this article

Background

- RNA abundance indicates the state of individual cells

- Single cell RNA sequencing is now highly accurate, sensitive and high-throughput, but only provides a snapshot at a point of time
  → time-resolved phenomena (e.g. embryogenesis, tissue regeneration) are difficult to analyse

- During development, differentiation occurs on a timescale of hours to days = comparable to the typical half-life of mRNA

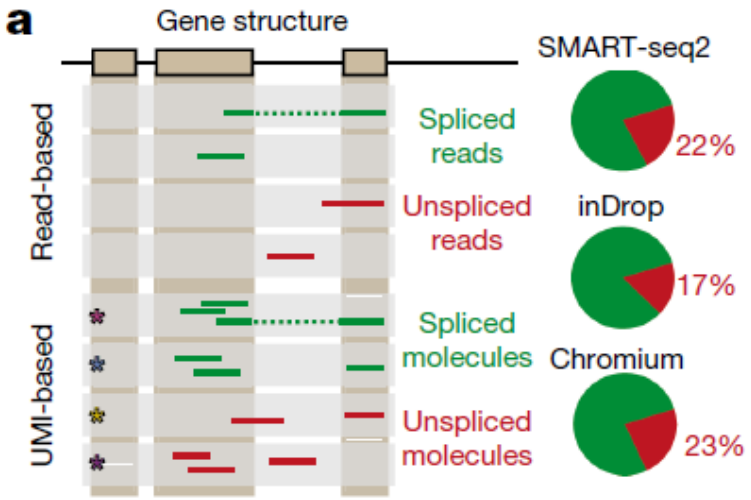
- Here, measuring the relative abundance of unspliced and spliced mRNA allows estimation of the time derivative of the gene expression state: RNA velocity

Background

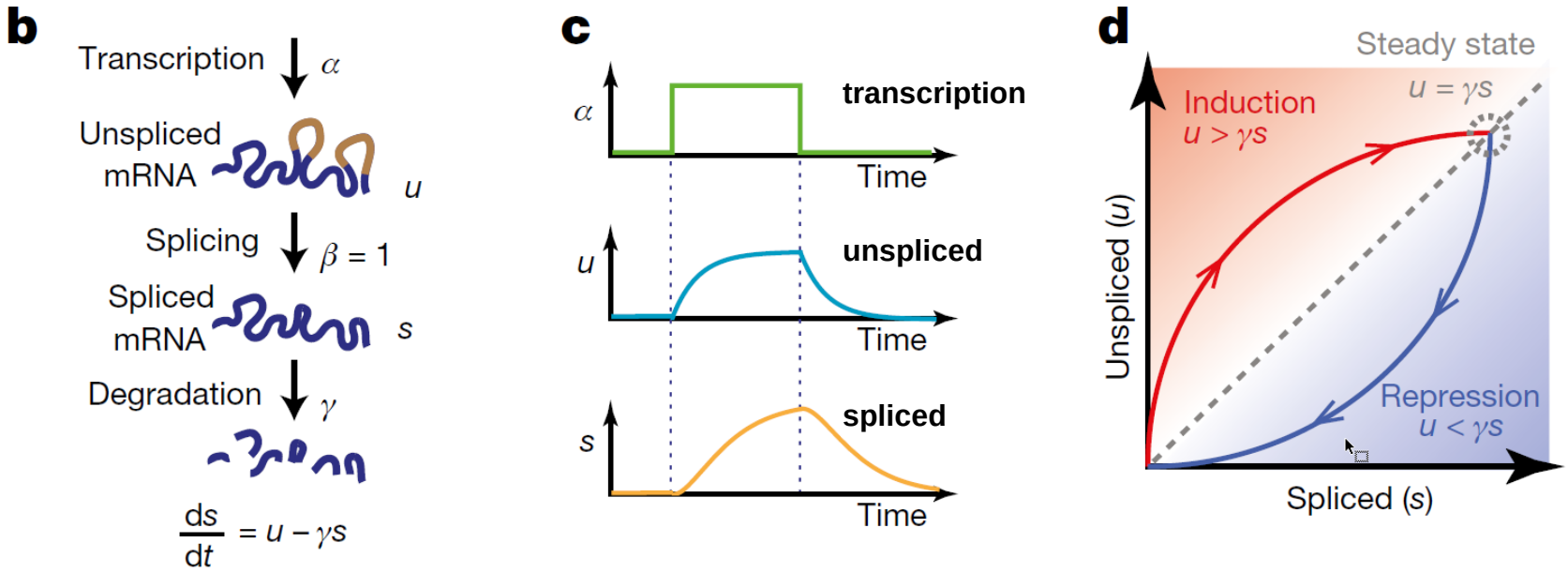
- scRNA-seq protocols rely on oligo-dT primers to enrich poly-A mRNA molecules

- Nevertheless, 15-25% of scRNA-seq reads contain unspliced intronic sequences
  - Originate from secondary priming positions in introns (polyA)
  - Represent unspliced precursor mRNAs

- Time-dependent changes in the abundance of unspliced and spliced mRNA have been observed in HEK cells incubated with 4-thiouridine (4sU)
  - 4sU in incorporated into mRNA, which can then be pulled out
  - Labelled mRNA molecules increase over time



mRNA abundance during a dynamic process

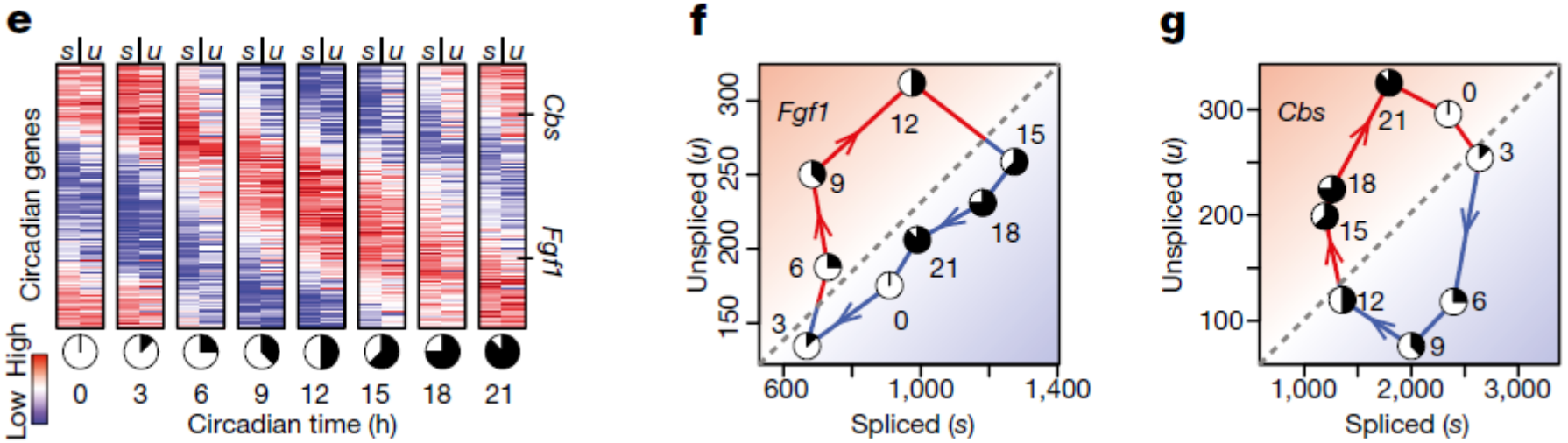


- Red area: increasing expression of a gene → unspliced (u) mRNA are in excess
- γ (diagonal line): constant transcription → equilibrium of unspliced and spliced (s) RNA
- Blue area: decreasing expression → spliced mRNA are in excess

RNA velocity: time derivative of the gene expression state

The balance of unspliced and spliced indicates the future state of the cell

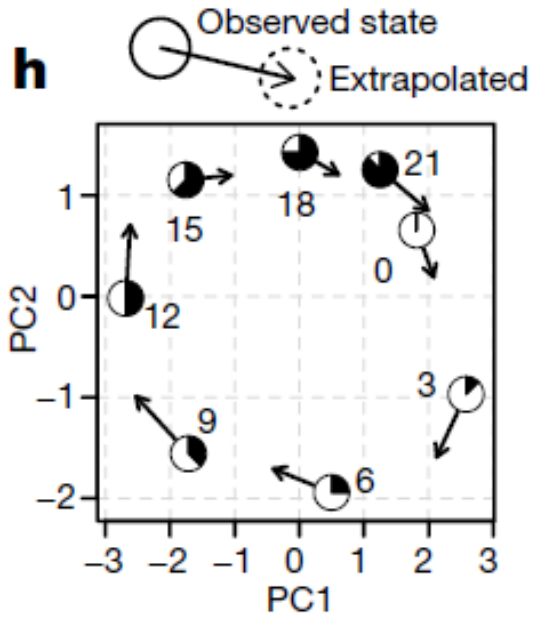
RNA-seq of circadian genes over 24 h (bulk liver mRNA)



Abundance of spliced and unspliced mRNA

Cbs: cystathionine beta-synthase
Fgf1: fibroblast growth factor 1

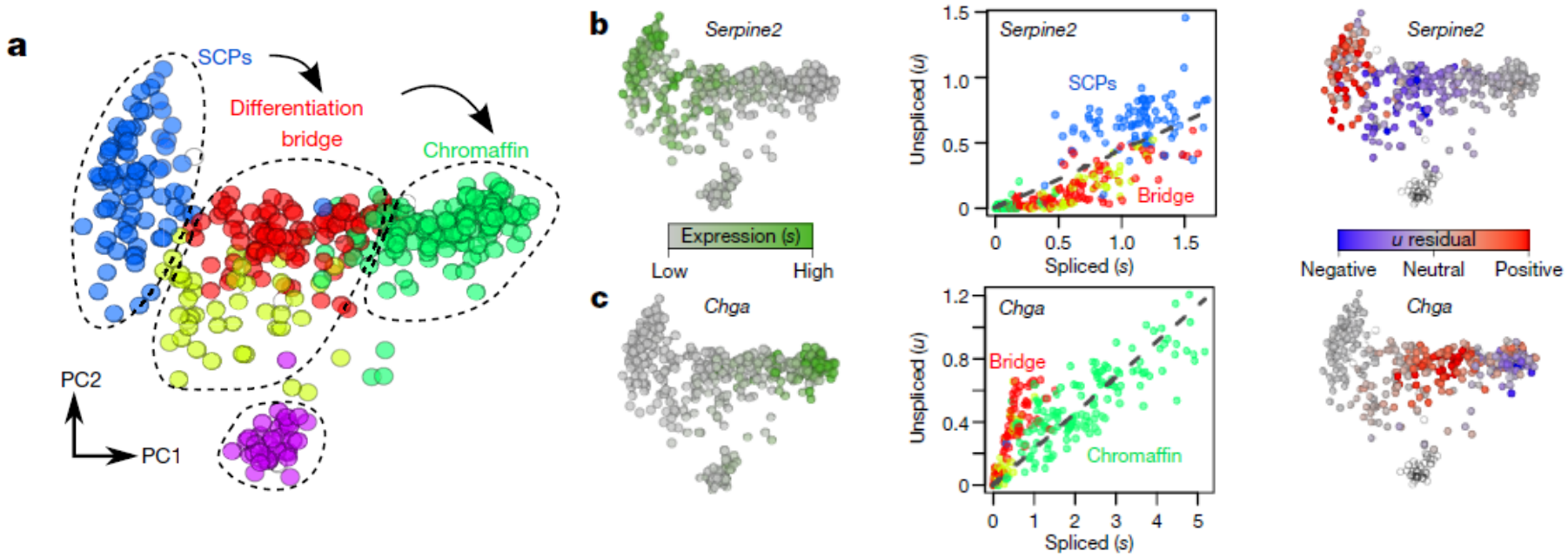
RNA-seq of circadian genes over 24 h (bulk liver mRNA)



- Future state of cell can be predicted based on current state (circle) and velocity estimates (vector)
- Model based on differential equations for every circadian gene

Cbs: cystathionine beta-synthase
Fgf1: fibroblast growth factor 1

RNA velocity during adrenal medulla development

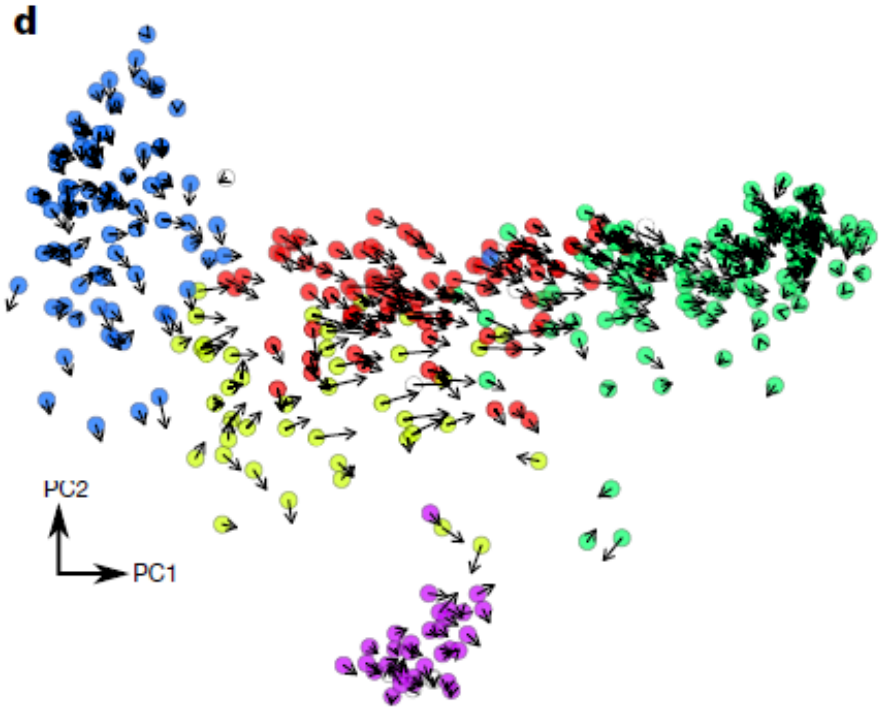


Development of adrenal medulla on embryonic day 12.5

a: Schwann cell precursors (SCP) differentiate into chromaffin cells
b,c: During differentiation, unspliced-spliced phase portraits of many genes deviate from steady-state equilibrium

Furlan et al., Science 2017

RNA velocity during adrenal medulla development

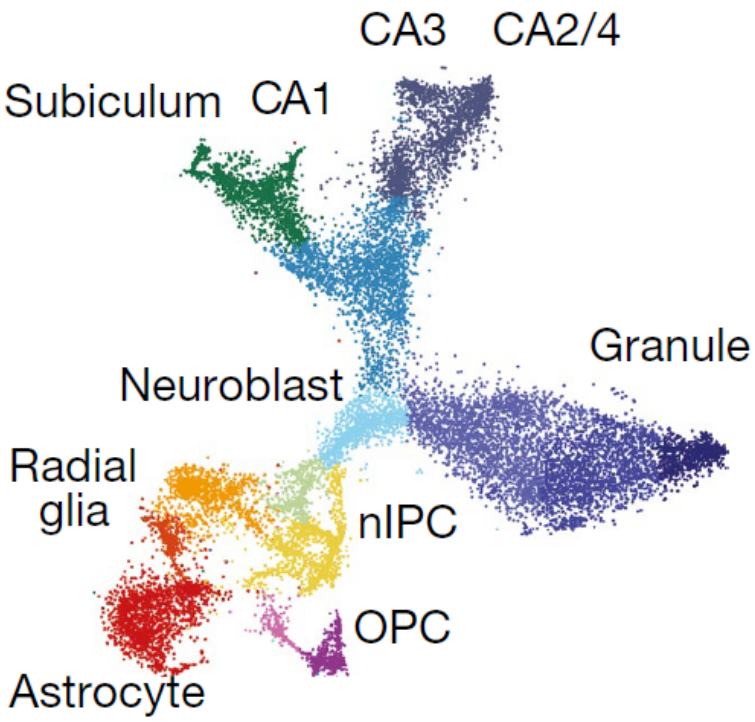


Velocity vectors of individual cells point towards expected fate

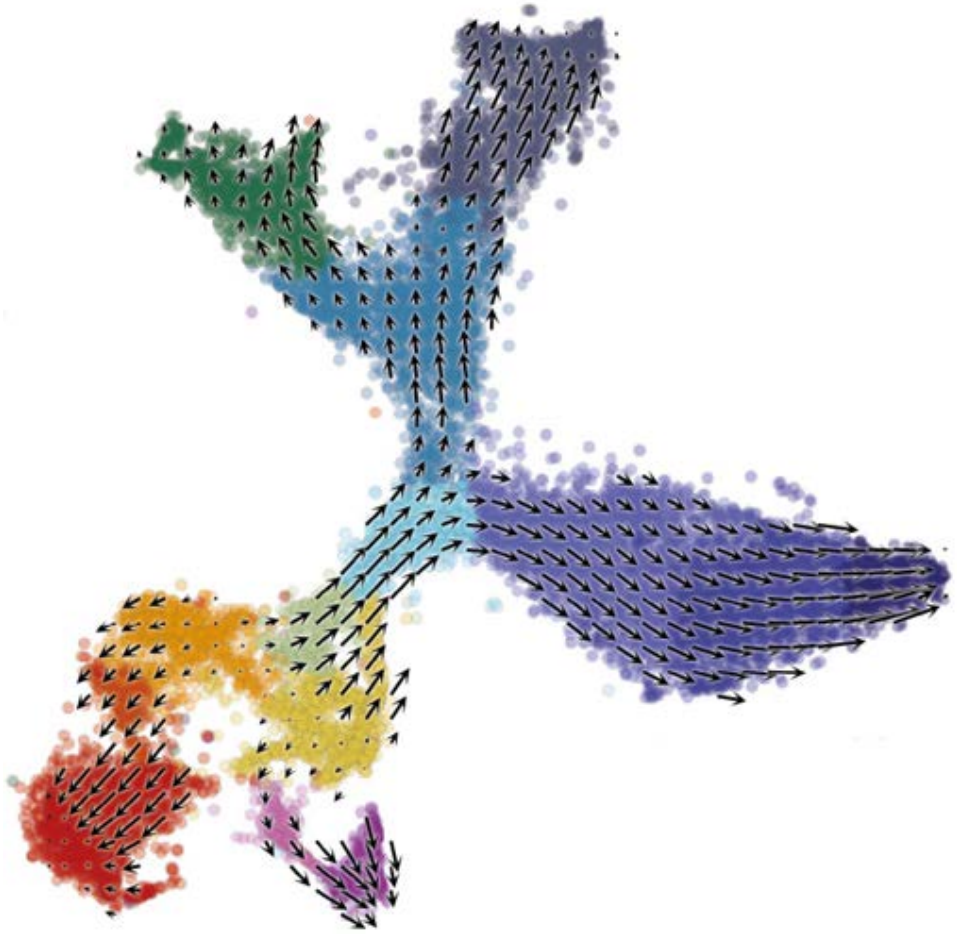
Estimated RNA velocity correlated with changes detected using metabolic labelling

Kriegstein et al., Annu. Rev. Neusci. 2009

RNA velocity: Hippocampal development



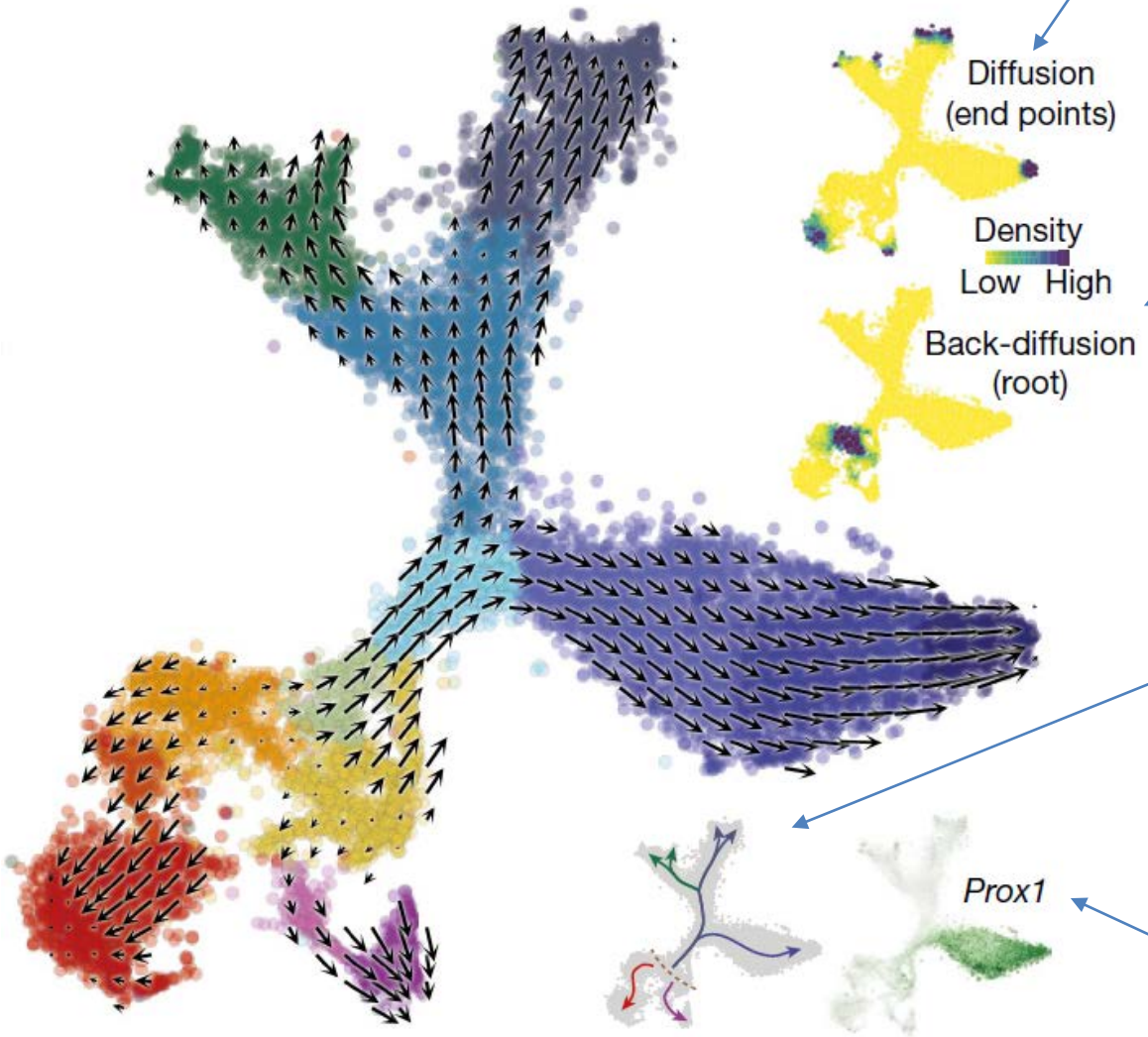
Identification of cell types based on expression of TF (w/o vascular cells)



t-SNE and RNA velocity of hippocampal cells
Arrows: average local velocity

nIPC: neurogenic intermediate progenitor cell
OPC: oligodendrocyte progenitor cell

RNA velocity: Hippocampal development



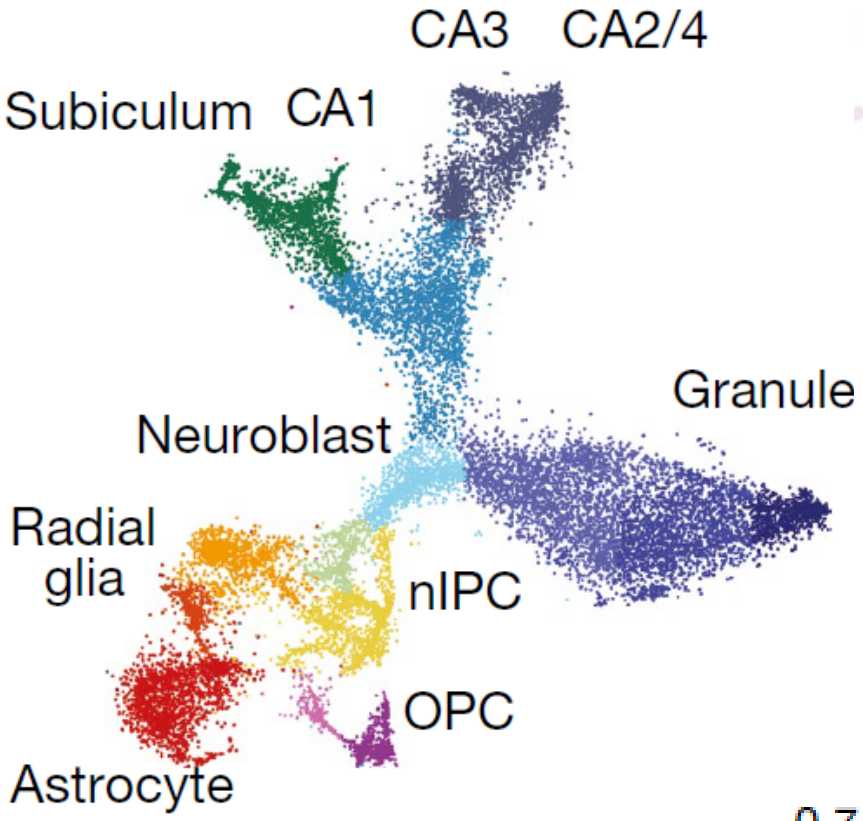
Endpoints = differentiated cells: lowest probability of velocity transition (Markov modelling)

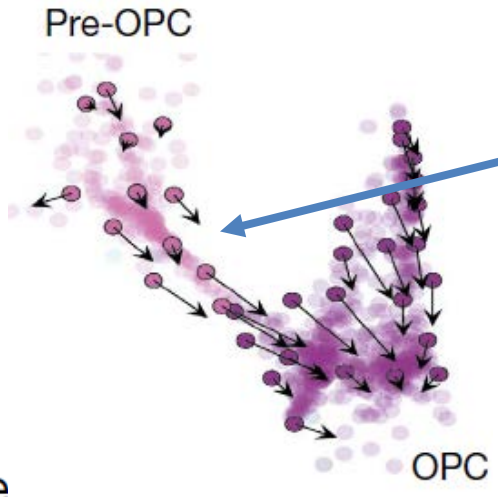
Root = radial glia: highest probability of velocity transition

Summary of velocity field

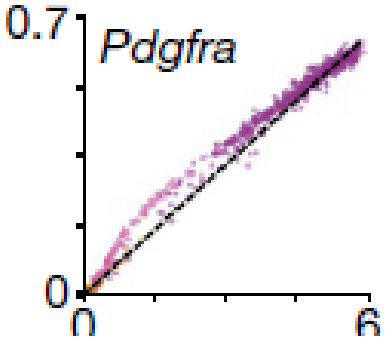
Prox1: transcr. factor required for granule cell development

Fate choice in OPCs





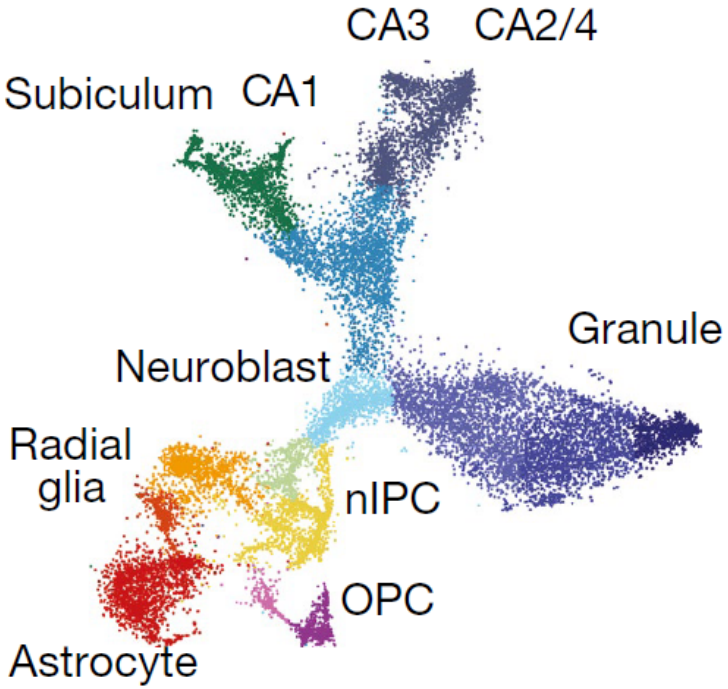
Narrow passage: might represent moment of OPC lineage commitment

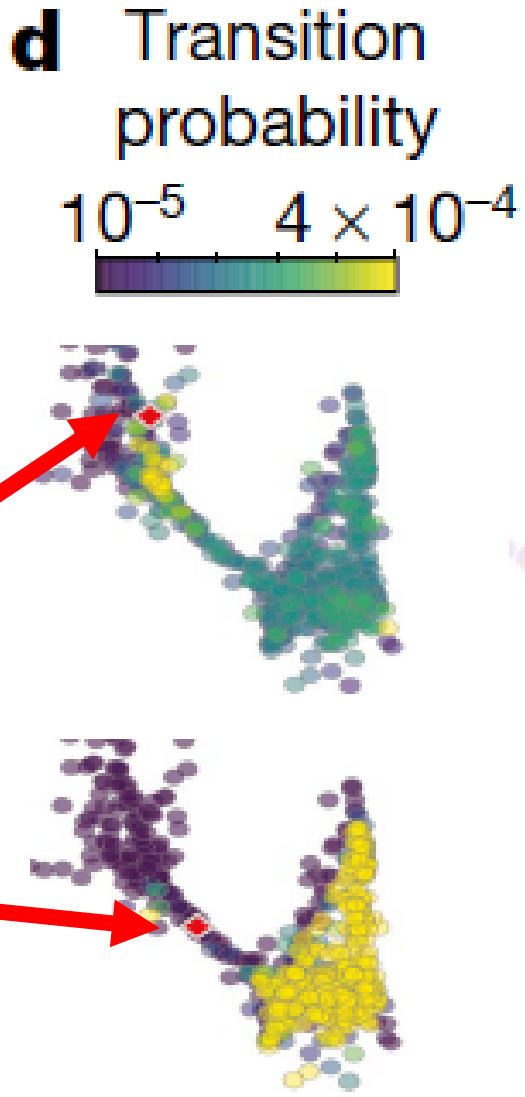


Pdgfra: marker for OPCs

- Induced in pre-OPCs, positive velocity
- Expressed in OPCs, stable

Fate choice in OPCs

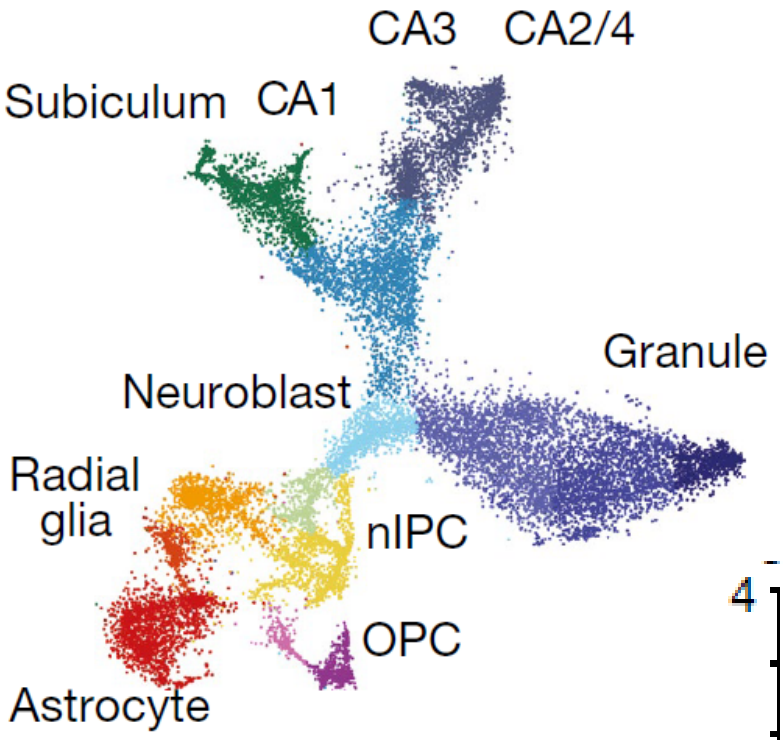


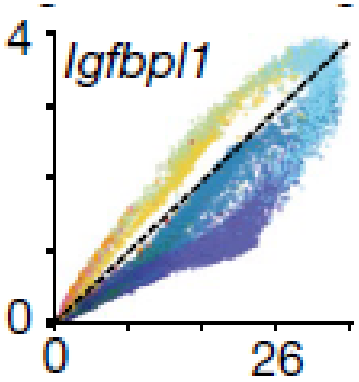


As a cell progresses along the “differentiation bridge”, the probability of transitioning back declines

Transcription factor feedback loops lock the cell into the OPC fate

Fate of neuroblasts





Igfbpl1:

- Expr. in neuroblasts
- positive velocity from radial glia to neurobl.
- Negative velocity from neurobl. to neuronal branches

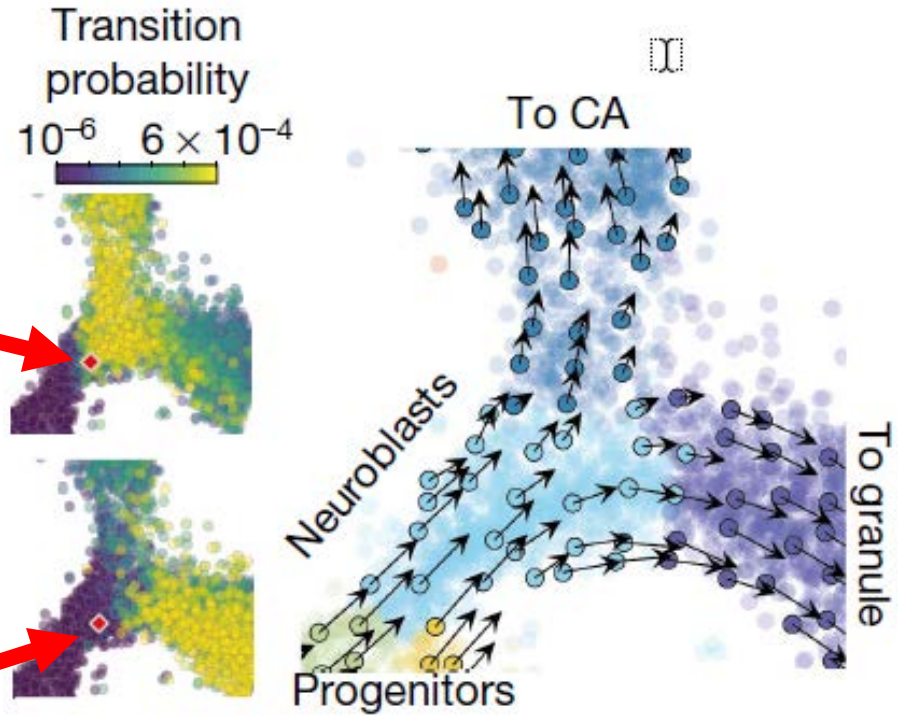
Igfbpl1: insulin-like growth factor binding protein-like 1

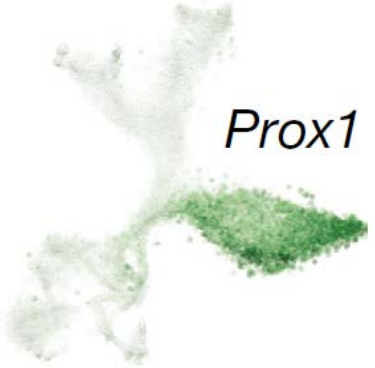
Two transcriptionally similar neuroblasts with different fates

Main difference: *Prox1* expression

Low *Prox1*: CA fate is likely

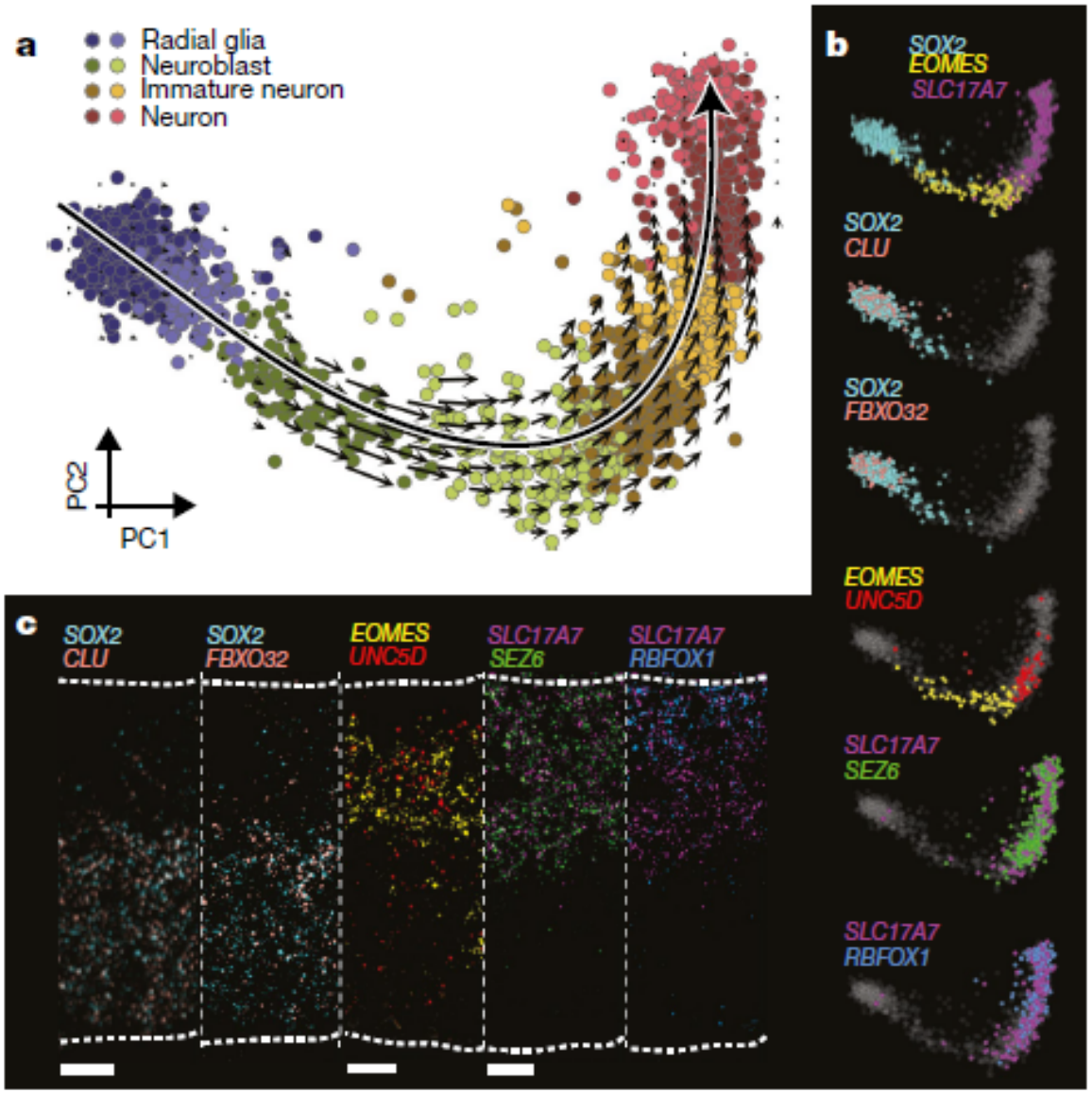
High *Prox1*: granule cell fate is likely





Prox1 is required for granule cell development
Prox1 deletion in neuroblasts → diff. to pyramidal neurons

Human embryonic glutamatergic neurogenesis (10 weeks, forebrain, droplet-based scRNA-seq)



Expression of markers

- *SOX2*: radial glia
- *EOMES*: neuroblasts
- *SLC17A7*: neurons

Multiplexed in-situ hybridisation: Layered expression of markers in tissue corresponds to pseudo-temporal distribution in scRNA-seq data

RNA velocity: Summary

scRNA-seq «snapshot» of unspliced and spliced mRNA abundance can yield information about dynamic temporal processes, such as cell differentiation

- Future state of cell can be predicted based on current state and RNA velocity
- Velocity can be modelled over «pseudotime»
- Stochastic modelling yields probability of transition into other cell states
- RNA velocity can be visualised on PCA / t-SNE plots
  - Note: Cells can have RNA velocities across many independent components simultaneously (e.g. differentiation, maturation, proliferation), which may not be visible in PCA, t-SNE etc.
  - Future algorithms might simultaneously fit a principle component manifold and RNA kinetics

nature

Article | Published: 20 November 2019

Gene expression cartography

Mor Nitzan, Nikos Karaiskos, Nir Friedman ✉ & Nikolaus Rajewsky ✉

Nature (2019) | Cite this article

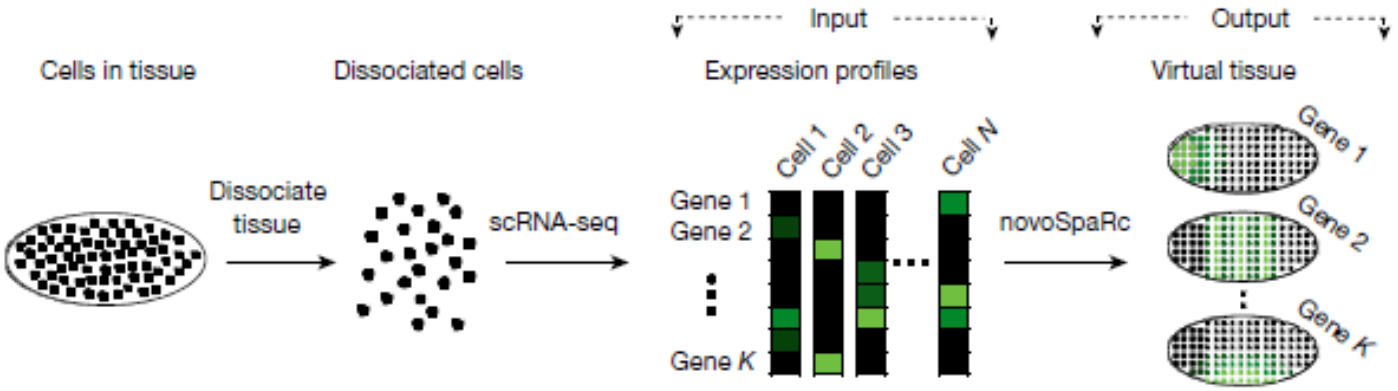
Background

- When performing scRNA-seq, tissues often have to be dissociated
  → Loss of information about spatial relationships and communication between cells

- Existing approaches to reconstruct tissues assign spatial positions to each cell, independently of other cells, by using a marker gene expression reference atlas
  - e.g. a map of in situ RNA patterns (Satija. et al, Nat Biotechnol 2015)
  - No information is currently available on the spatial expression of many genes → precise cell mapping is often impossible

novoSpaRc: Workflow

1. Distances are computed for each pair of cells on graphs for expression space and physical space
2. Distances of pairs of cells are aligned in a way that is consistent with known spatial expression profiles of marker genes (used as anchors)
3. Probabilistic map that assigns each cell a distribution over locations on the physical space is obtained. Mapping of cell pairs is formulated as optimal transport problem.

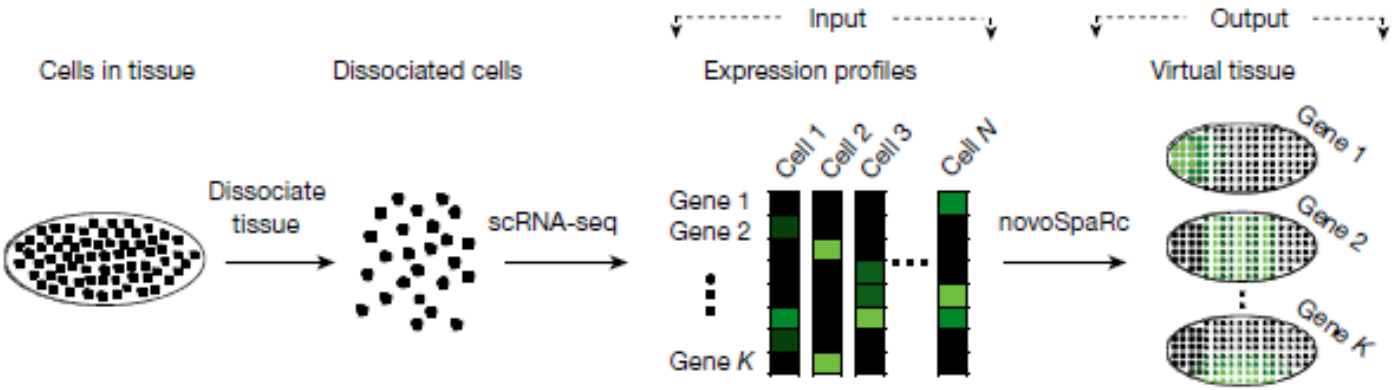


novoSpaRc: Workflow

1. Distances are computed for each pair of cells on graphs for expression space and physical space
2. Distances of pairs of cells are aligned in a way that is consistent with known spatial expression profiles of marker genes (used as anchors)
3. Probabilistic map that assigns each cell a distribution over locations on the physical space is obtained. Mapping of cell pairs is formulated as optimal transport problem.

Optimal transport: Classic example

- *n* mines produce iron ore and *n* factories use the iron ore
- Every mine supplies one factory
- Transport comes at a cost which increases with distance
- What is the optimal transport plan with the lowest cost?

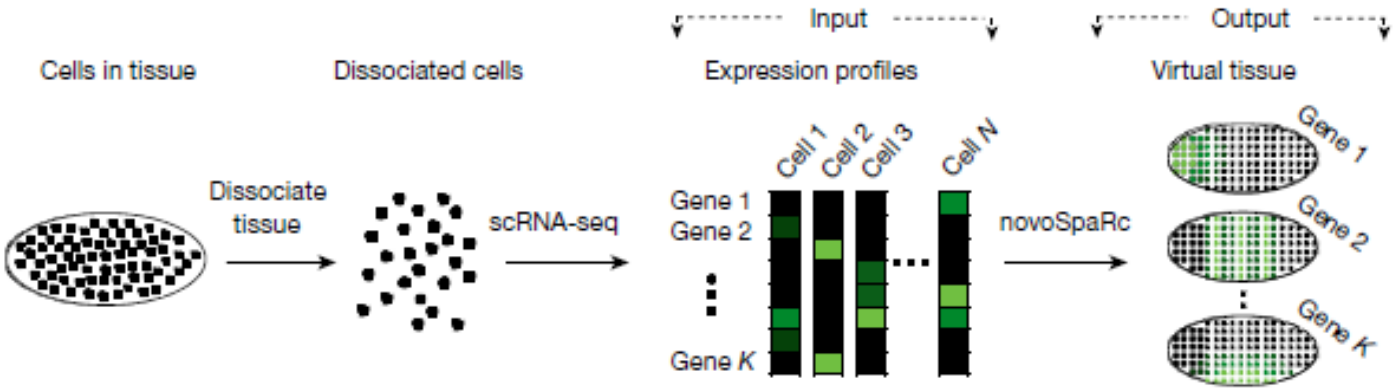


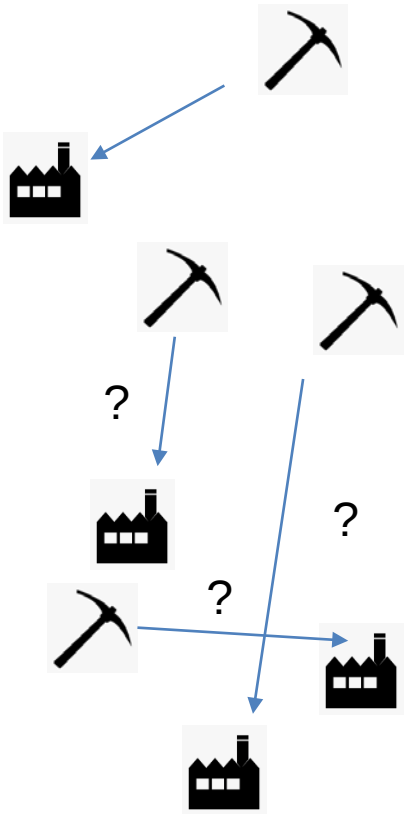
novoSpaRc: Workflow

1. Distances are computed for each pair of cells on graphs for expression space and physical space
2. Distances of pairs of cells are aligned in a way that is consistent with known spatial expression profiles of marker genes (used as anchors)
3. Probabilistic map that assigns each cell a distribution over locations on the physical space is obtained. Mapping of cell pairs is formulated as optimal transport problem.

Optimal transport: Classic example

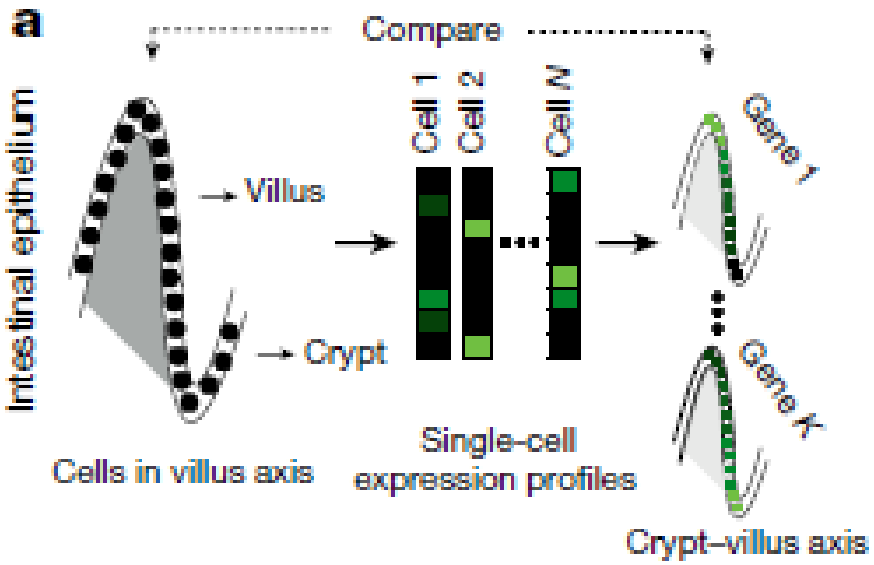
- *n* mines produce iron ore and *n* factories use the iron ore
- Every mine supplies one factory
- Transport comes at a cost which increases with distance
- What is the optimal transport plan with the lowest cost?

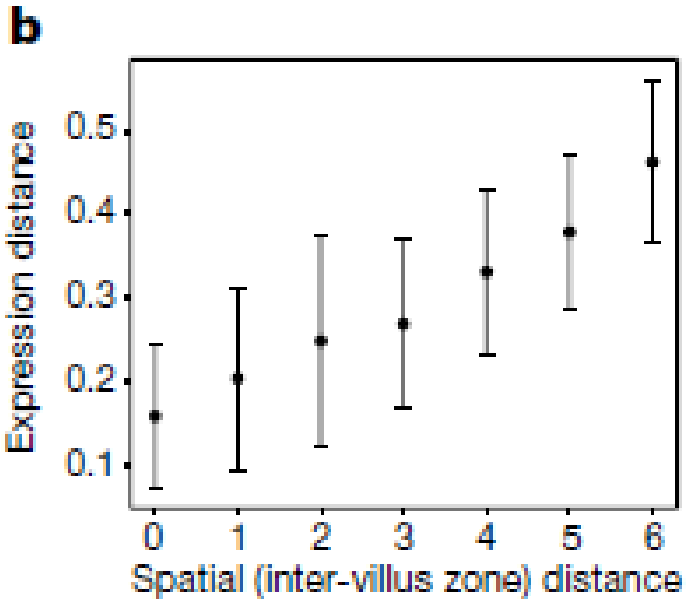




Reconstruction of symmetrical tissues: Intestinal epithelium

Hypothesis: distances in expression state and physical space correspond to each other



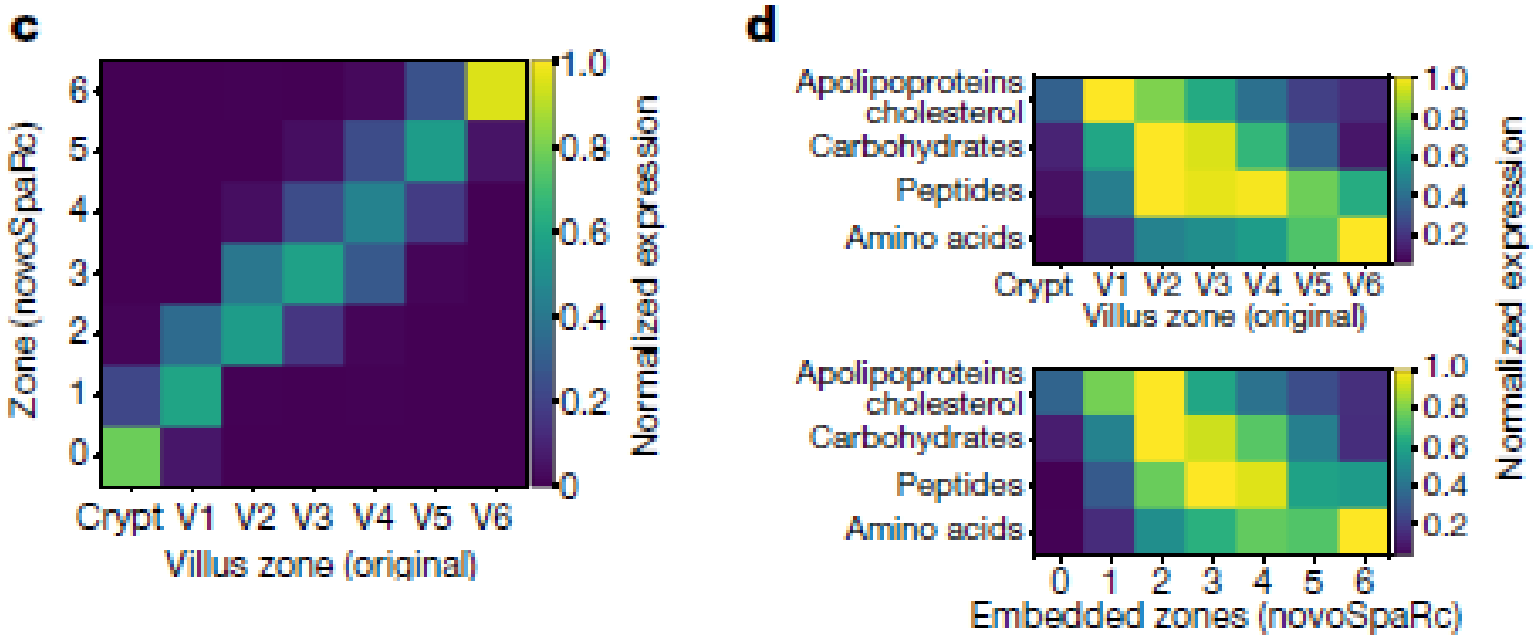


Cells are aligned along the crypt-villus axis

Expression and spatial distance show a monotonic relationship

Reconstruction of symmetrical tissues: Intestinal epithelium

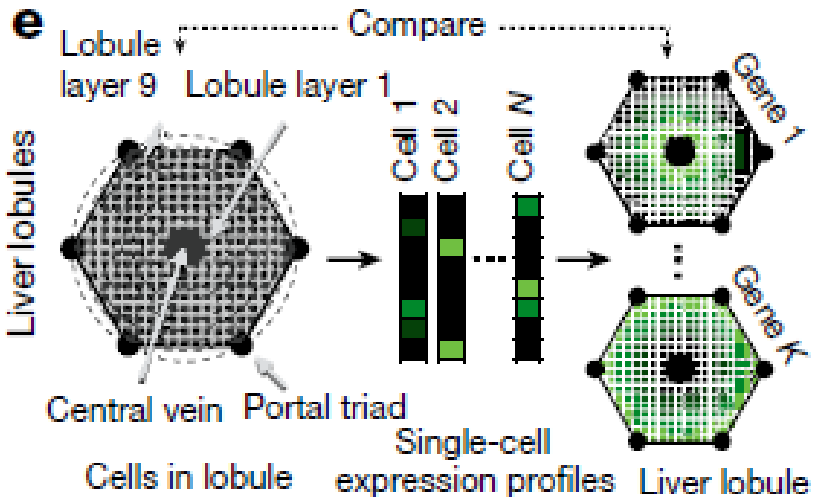
In intestinal epithelium, cells have previously been classified into 7 distinct expression zones (Crypt, V1, V2 … V6)

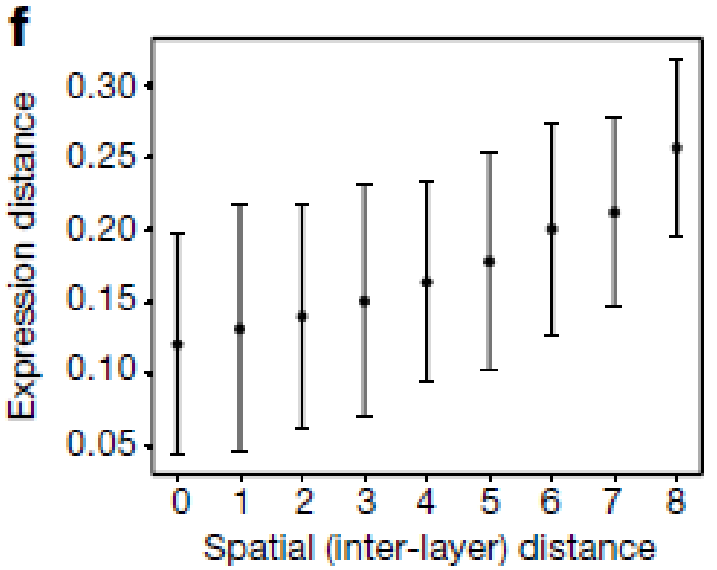


Reconstructed distribution correlates well with actual pattern (r = 0.99)

Reconstruction captured known gene expression “division of labour”

Reconstruction of symmetrical tissues: Liver lobule

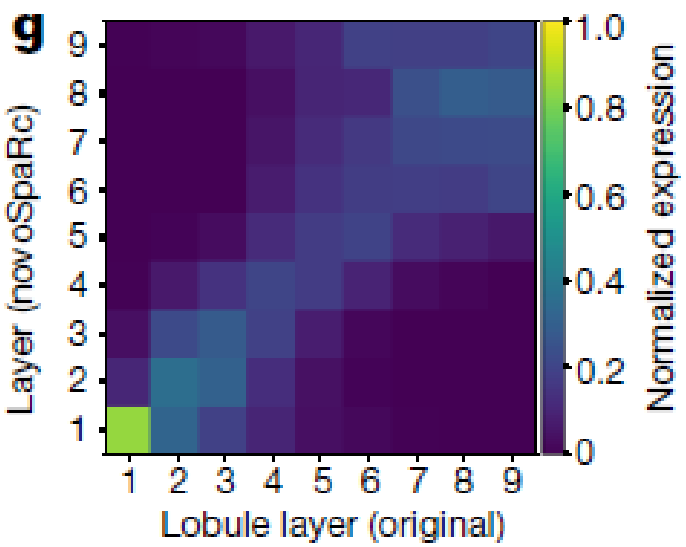


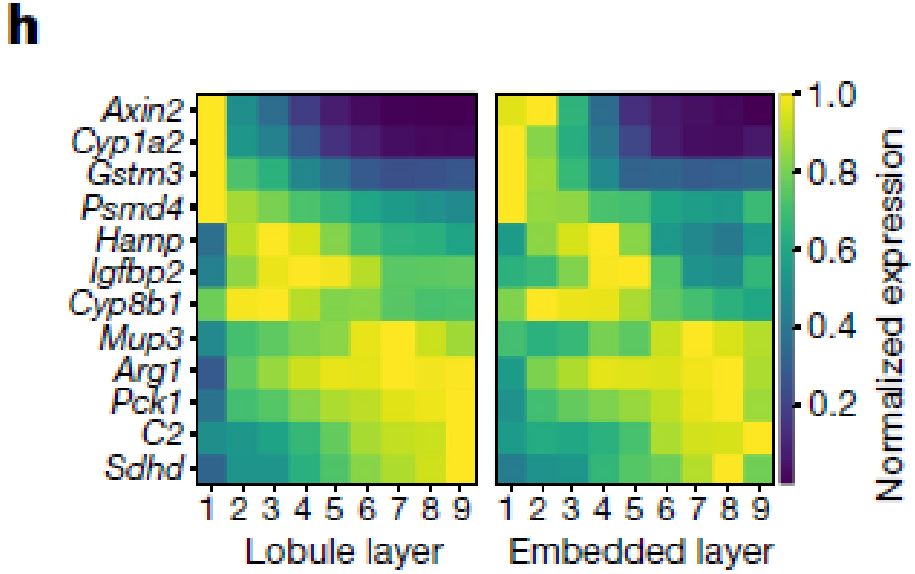


Cells are assigned to 9 layers

Expression and spatial distance show a monotonic relationship

Reconstruction of symmetrical tissues: Liver lobule





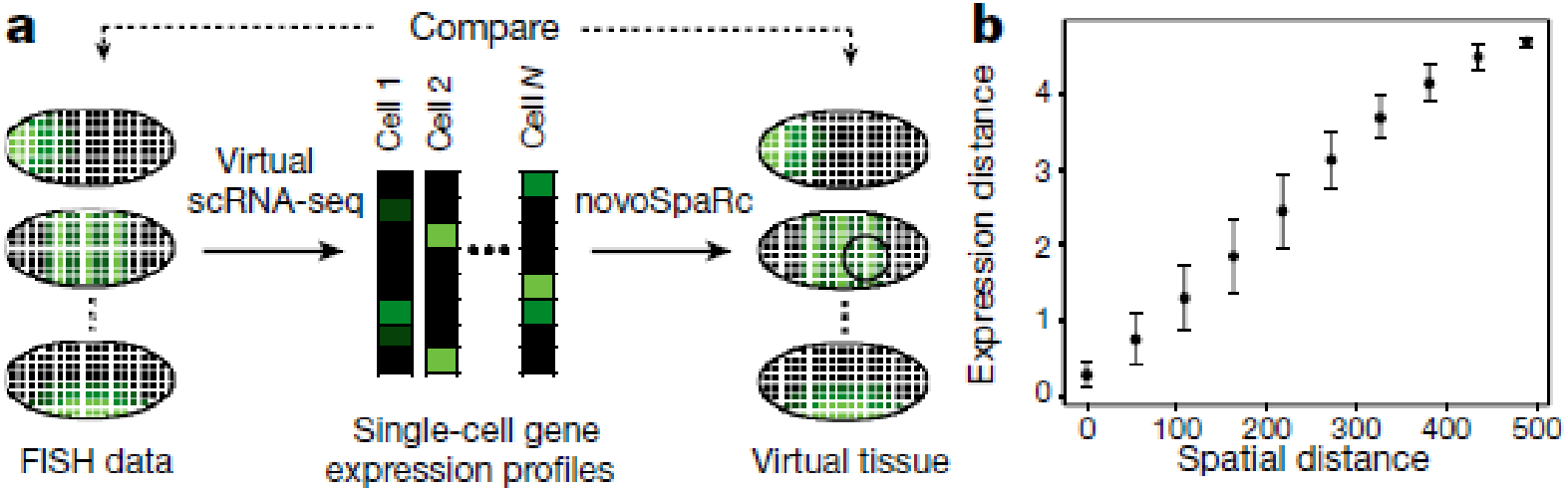
Reconstructed distribution correlates well with actual pattern (*r* = 0.94)

Spatial expression patterns of pericentral and periportal genes can be replicated

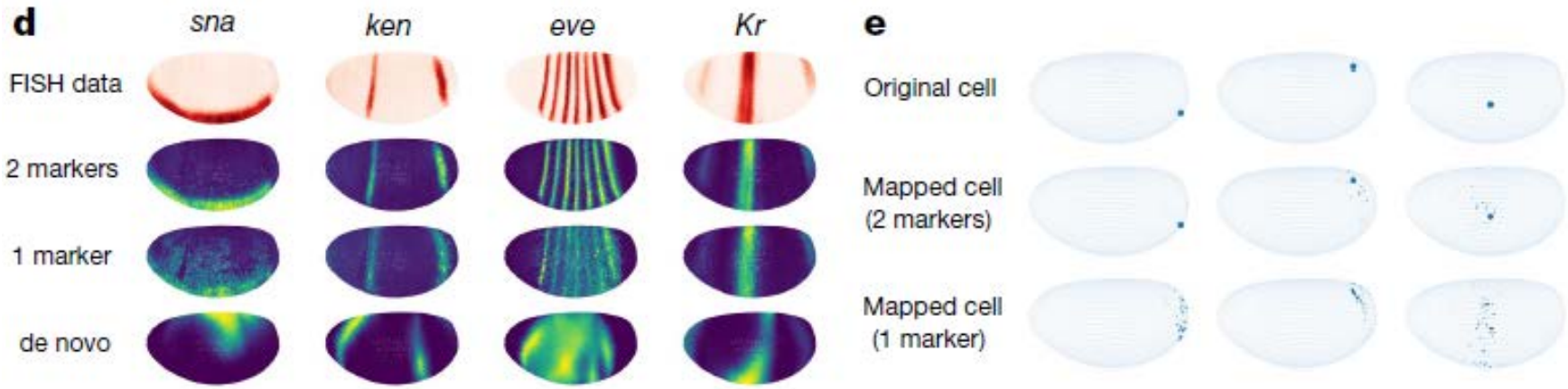
(relevant to next paper)

Reconstruction of a stage 5 *Drosophila* embryo

- In stage 5 of development, *Drosophila* embryos consist of ~6000 cells
- The expression levels of 84 TF have been quantitively registered using FISH (Berkeley *Drosophila* Transcription Network Project, BDTNP)
- scRNA-seq data was obtained from the BDTNP dataset
- Tissue reconstruction was performed 1. de novo and 2. with use of marker genes as reference
- Again, monotonic relationship between expression and spatial distance



Reconstruction results compared to FISH data

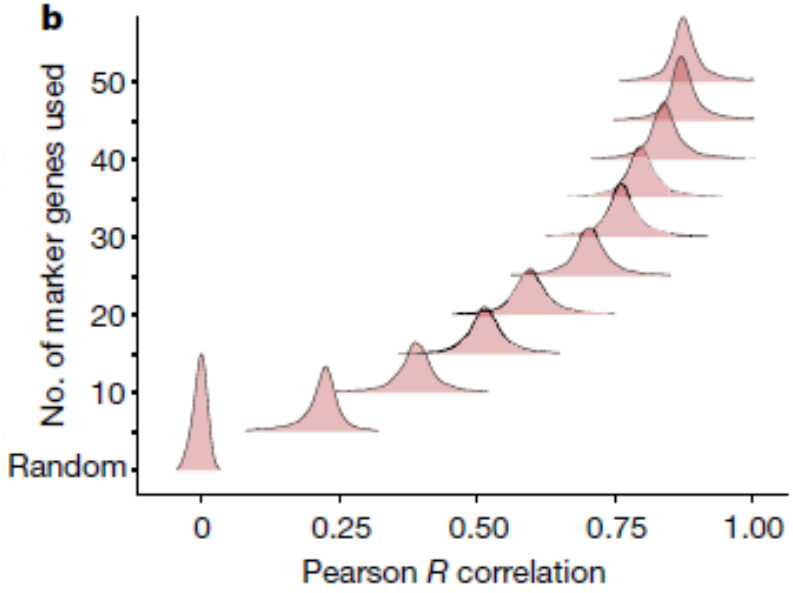


Reconstruction results for 4 TF and individual cells, using 0-2 marker genes as reference, compared to FISH data (top)

2 marker genes are sufficient to create accurate reconstructions

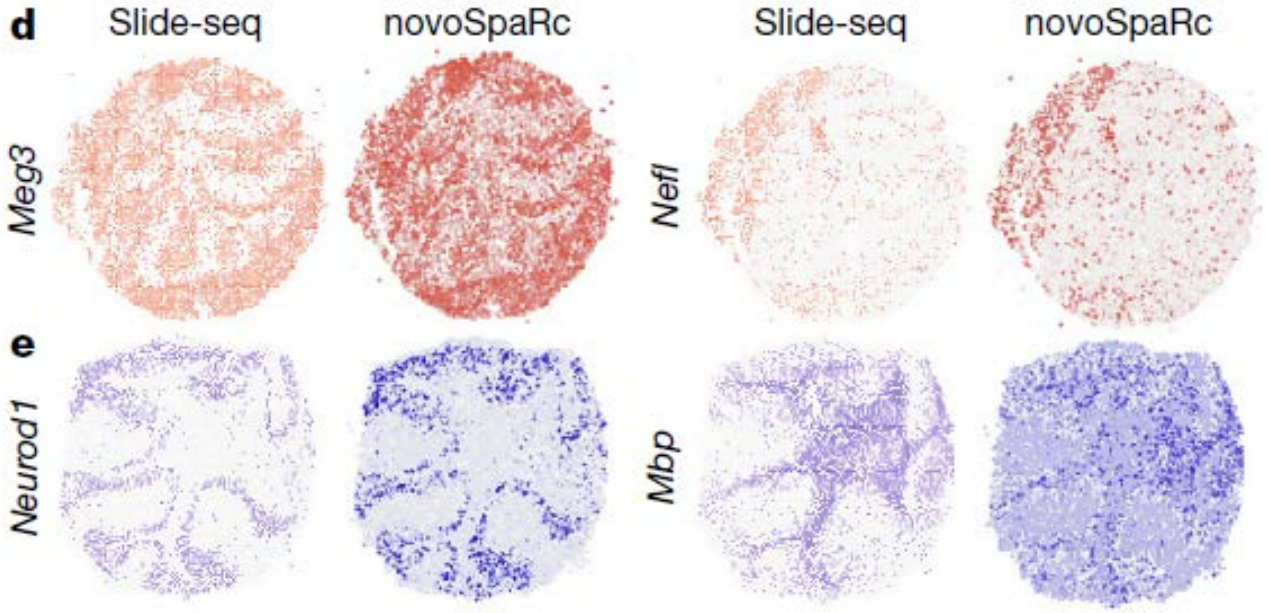
Reconstruction of mouse cerebellar slice

Pcp4: Purkinje cell protein 4



Correlation with known cell distribution increases with no. of employed anchor genes

Reconstruction of mouse cerebellar slice



Meg3: maternally-expressed gene 3

Nefl: neurofil. light chain

Neurod1: neuronal differentiation 1

Mbp: myelin basic protein

Gene expression cartography: Summary

- Diverse biological tissues can be reconstructed from existing scRNA-seq datasets based on a simple hypothesis: There is a structural correspondence between the distances between cells in expression space and in physical space

- Anchoring with known marker genes improves the reconstruction – if not, the reconstr. is subject to global transformations (e.g. mirroring)

- Previously unknown spatially informative genes could be identified (e.g. long non-coding RNAs, transcription factors)

- Can the method handle tissue surfaces in more complex tissues?

nature biotechnology

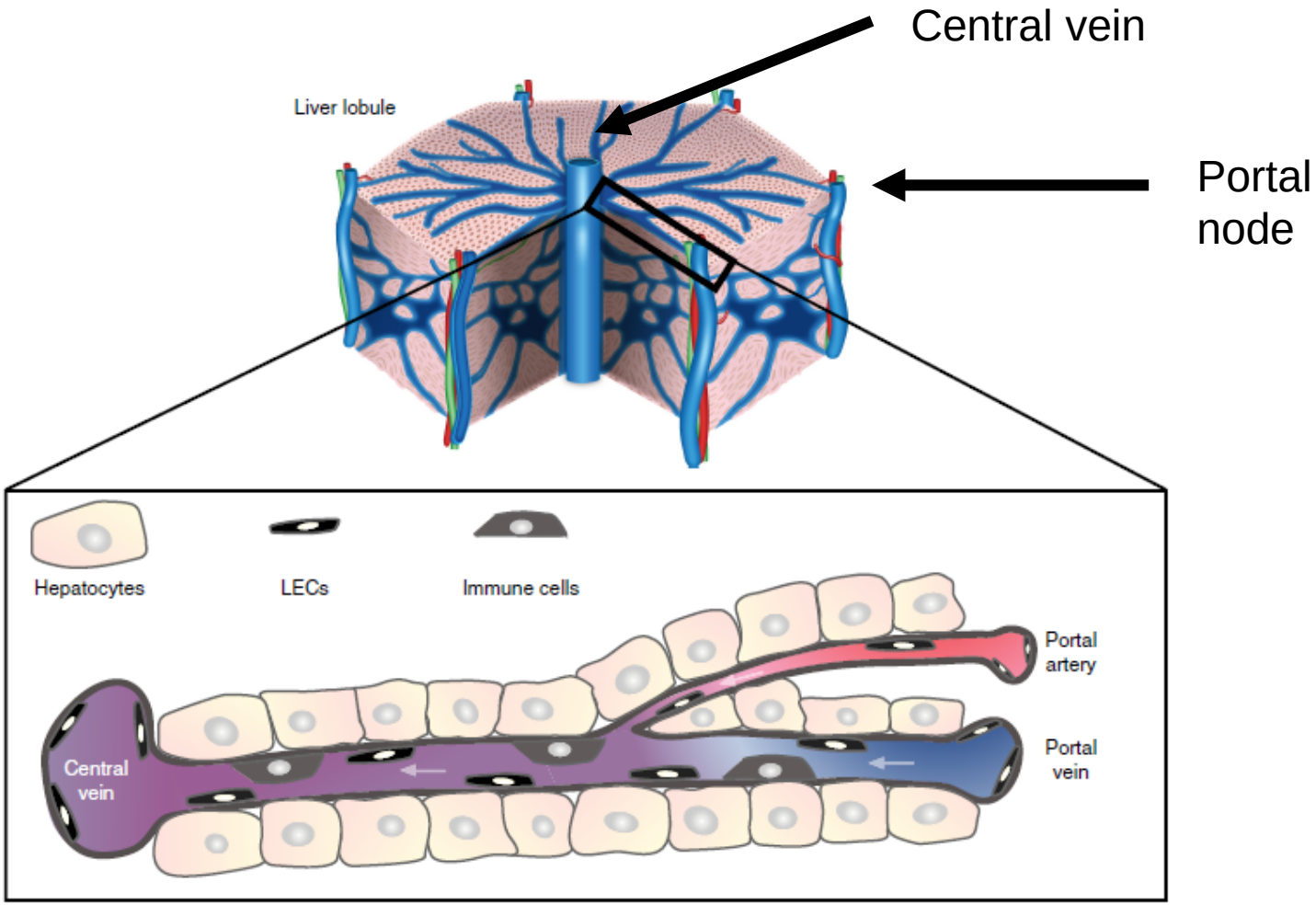
Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells

Keren Bahar Halpern[1,4], Rom Shenhav[1,4], Hassan Massalha[1], Beata Toth[1], Adi Egozi[1], Efi E Massasa[1], Chiara Medgalia[2], Eyal David[2], Amir Giladi[2], Andreas E Moor[1], Ziv Porat[3], Ido Amit[2] & Shalev Itzkovitz[1]

Background

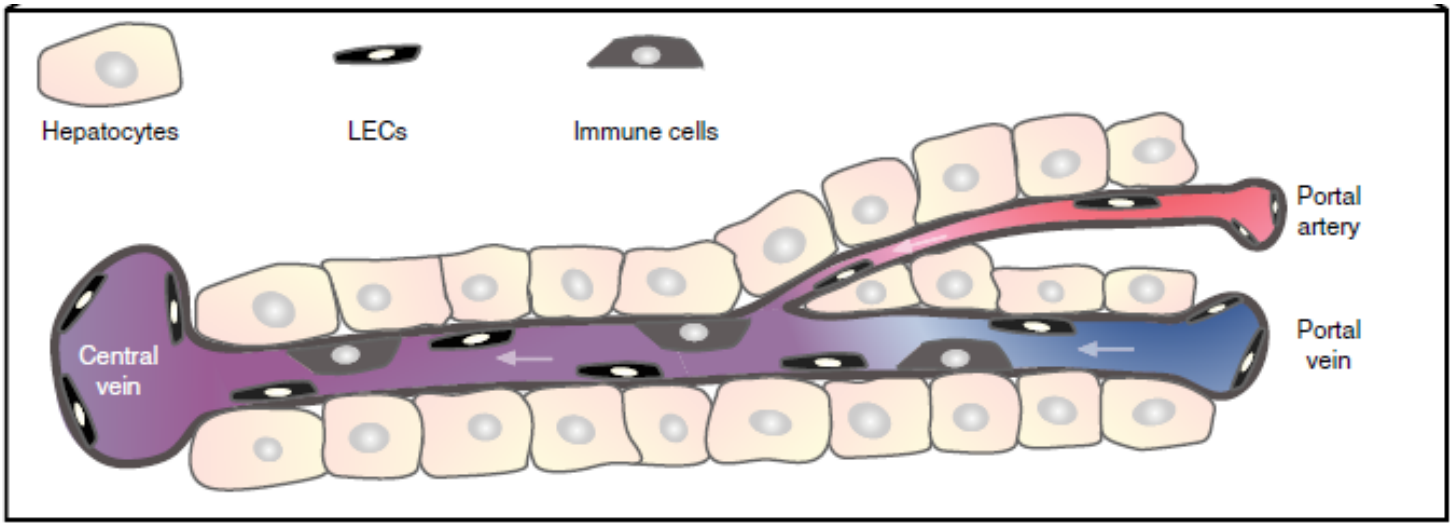
- In spatial transcriptomics, determining zonation of small cells with low mRNA content or without highly expressed landmark genes remains challenging
- In the liver, hepatocytes and diverse non-parenchymal cells (NPCs) are arranged in lobules – repeating, hexagonal units
- Lobules are composed of a central vein, radial sinusoidal networks and portal nodes (arteries, veins and bile ducts)
- Lobule blood vessels are lined with liver endothelial cells (LECs)

Liver lobule



- The lobule microenvironment gives rise to spatial division of labour among hepatocytes, depending on radial coordinates
- It is unknown whether liver non-parenchymal cells exhibit similar spatial division of labour

Liver lobule



- Liver endothelial cells (LECs) make up about 50% of liver NPCs
- Form building blocks of blood vessels, clear endotoxins and bacteria, regulate immune responses, present antigens, secrete morphogens that shape hepatocyte gene expression
- LECs at different lobule radial coordinates are known to posess morphological differences, but their gene expression hasn’t been characterised

Overview

- A previous study used scRNA-seq and FISH to construct a panel of hepatocyte landmark genes, which differed according to radial coordinates (Halpern et al., Nature 2017)

- scRNA-seq of LECs is challenging, given their small size → transcripts of most genes won't be present in individual LECs

- LEC zonation pattern is unknown → no reference available for mapping

Sequence RNA of hepatocyte-LEC pairs → characterise gene expression of hepatocytes (known transcriptional profile) and LECs that are attached to them in the tissue

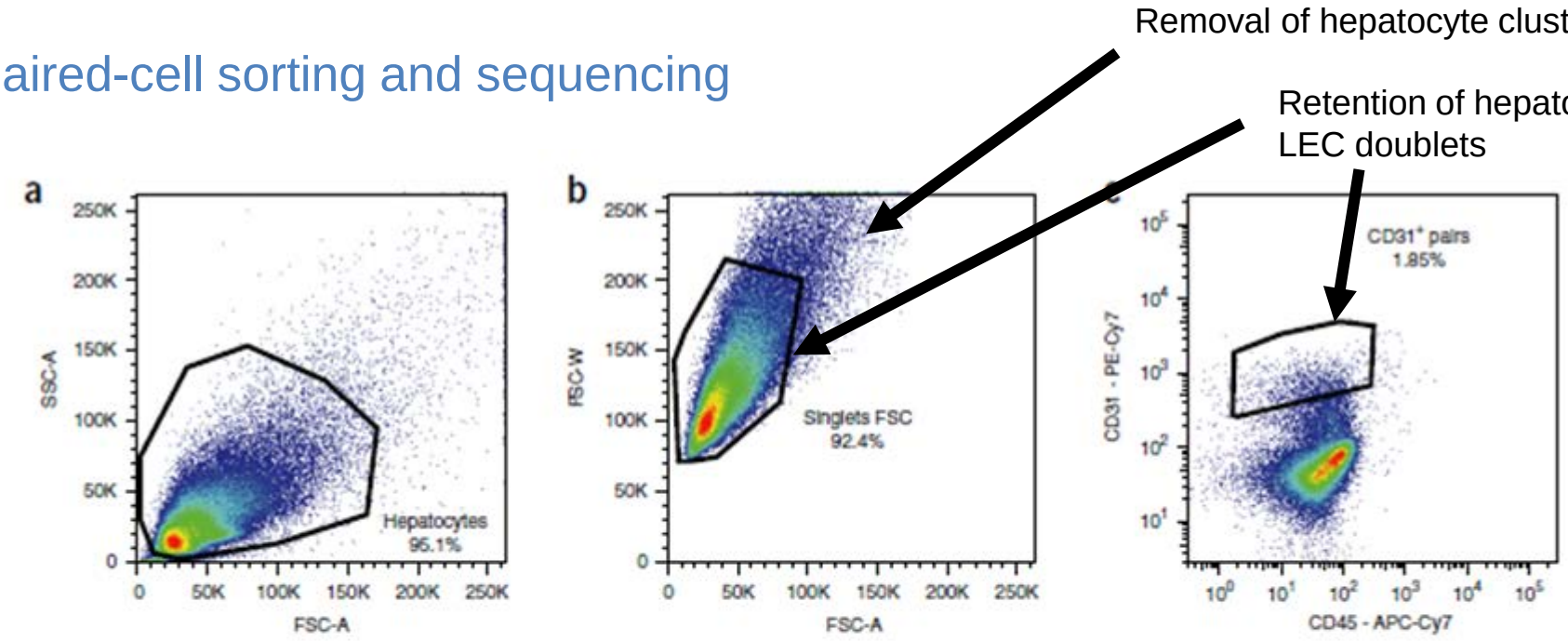
→ Resolution of LEC zonation pattern

Paired-cell sorting and sequencing

- Liver tissue was dissociated with collagenase D, which is **less** efficient than other tissue dissociation enzymes (e.g. Liberase) → **cell pairs could be retained**

- FACS: gating for hepatocytes (based on size) and CD31+ (endothelium) → **hepatocyte-LEC doublets**

Paired-cell sorting and sequencing



Imagestream: sorted pairs

Hepatocyte-LEC doublets are sorted into wells → scRNA-seq

Filtering paired RNA sequencing data

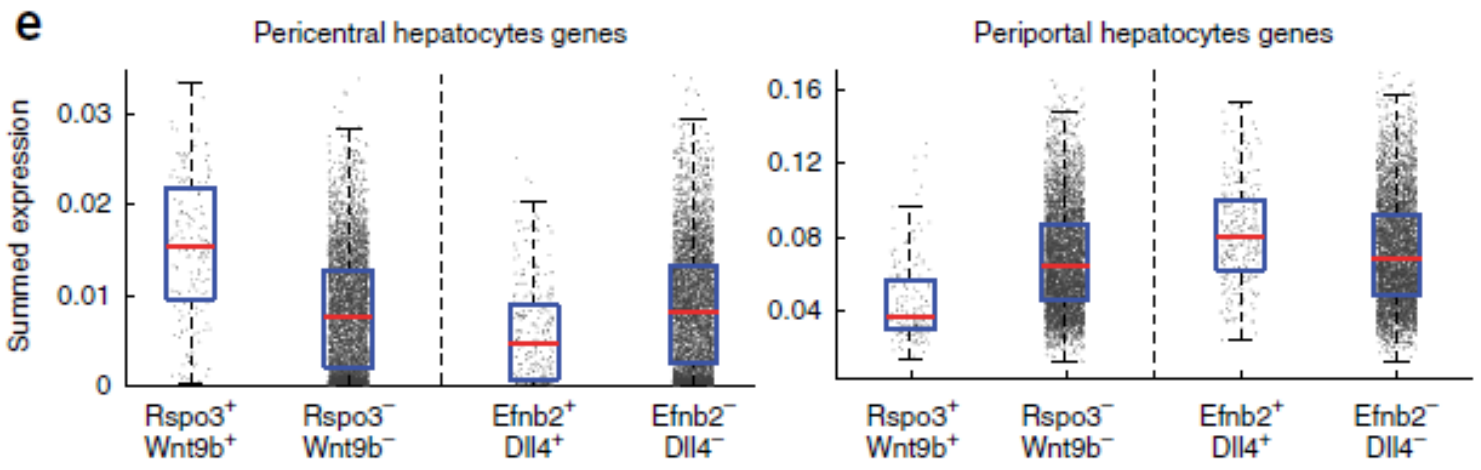
Removal of non-target events

- Filtering out of wells that didn't contain markers of both cell types

Selection of ligand-receptor gene pairs that are known to be zonated

1. *Rspo3*, *Wnt9b*: pericentral
2. *Efnb2*, *Dll4*: periportal

- Pairs with pericentral or periportal genes also showed matching enrichment / depletion of pericentral / periportal hepatocyte transcripts → very few artifactual pairs

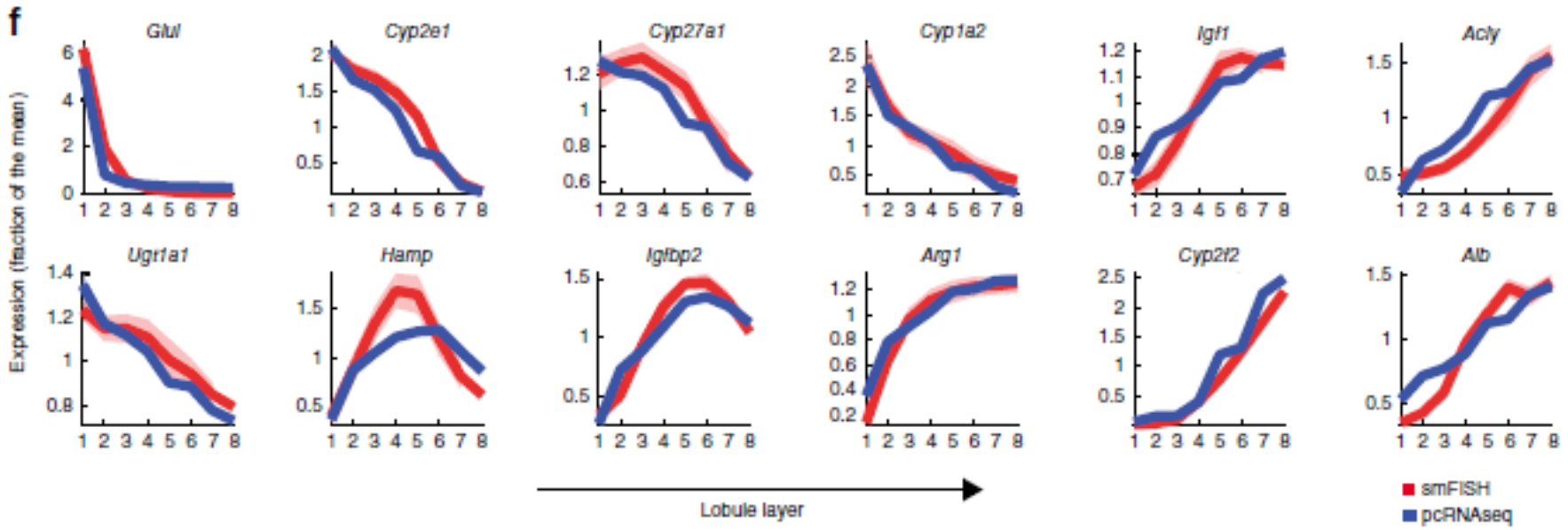


Rspo3: R-spondin 3
Wnt9b: Wnt family 9b
Efnb2: Ephrin B2
Dll4: Delta-like Notch ligand 4

Anatomical reconstruction

Zonation

- Pairs were assigned a scaled radial coordinate (1-8), based on ratio of summed expression of 21 pericentral and 30 periportal hepatocyte landmark genes (previously identified)



Landmark genes with high expression and low inter-mouse variability

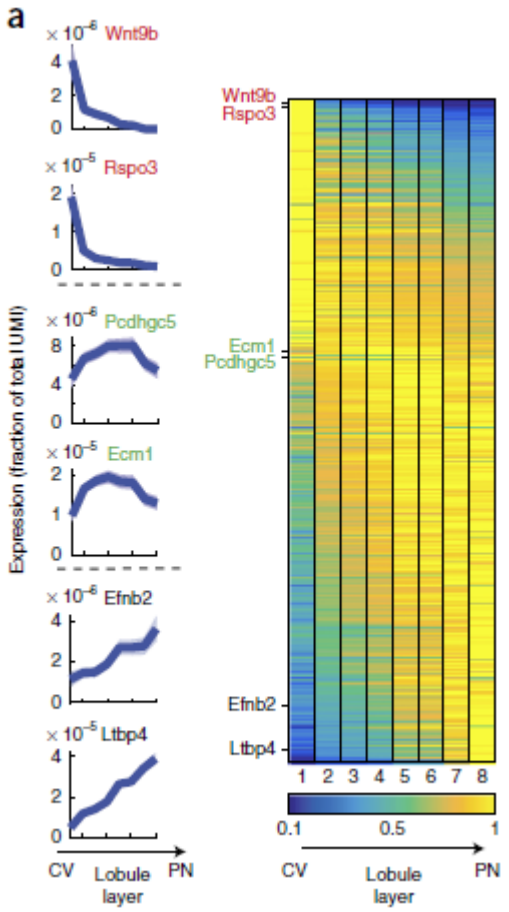
Anatomical reconstruction

Endothelial genes

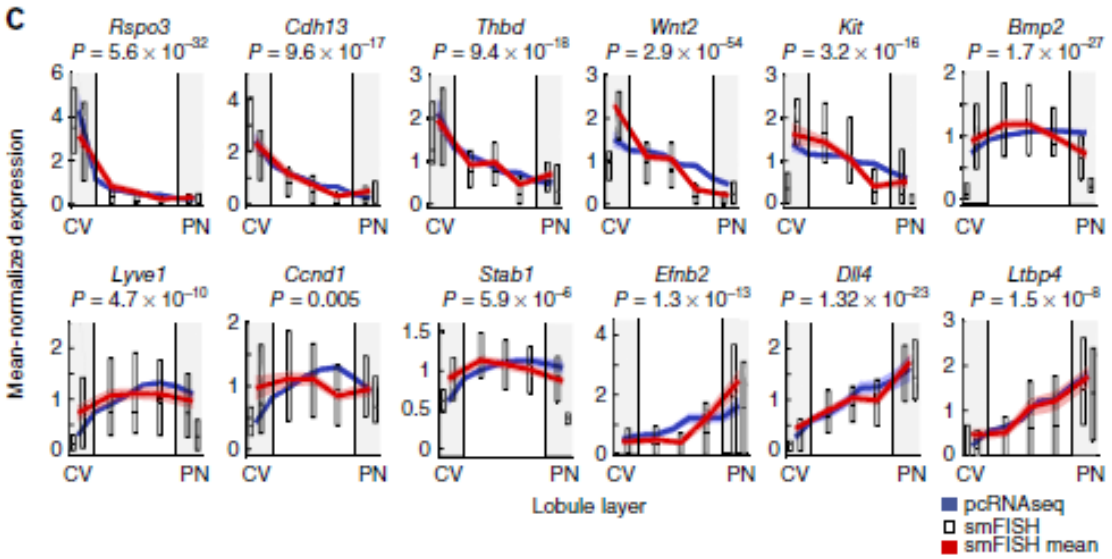
- Focus on genes that are strongly expressed in endothelium, relative to hepatocytes
- Exclusion of genes that are zonated in hepatocytes (otherwise, zonation of endothelium might falsely be attributable to hepatocyte transcripts)
- Removal of immune cell genes
- Selection of endothelial genes that were differentially expressed, according to zonation

→ 1303 LEC-specific genes, 475 of which were zonated (35%)

Zonated LEC-specific genes

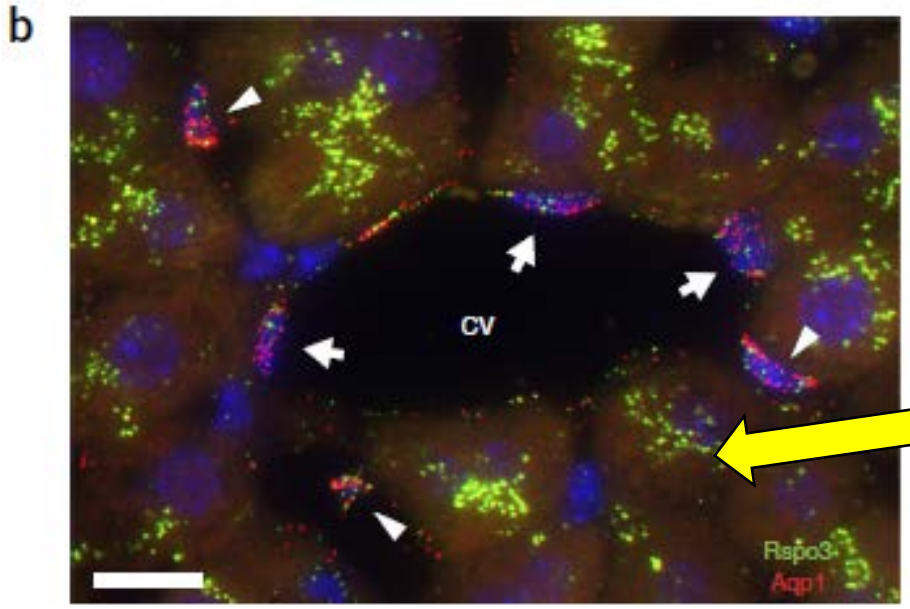


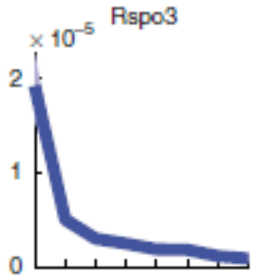
Expression profiles



Zonation patterns can be confirmed by smFISH

FISH of LEC-specific genes

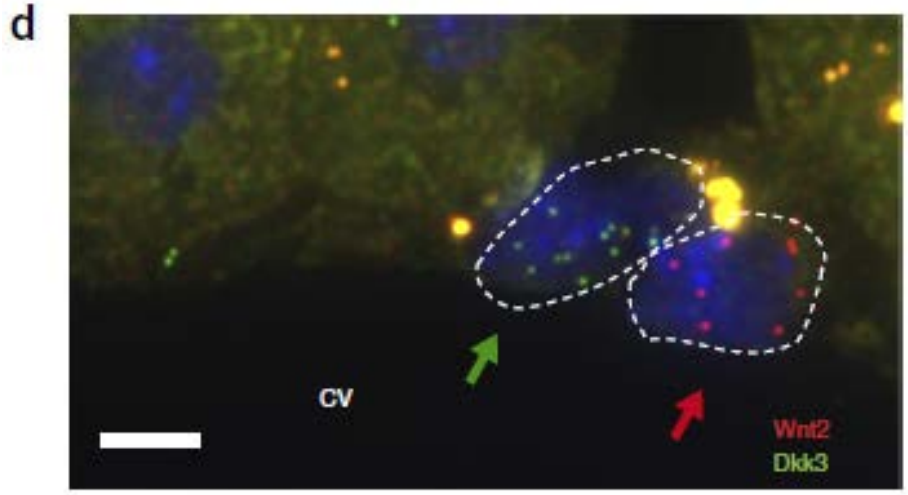




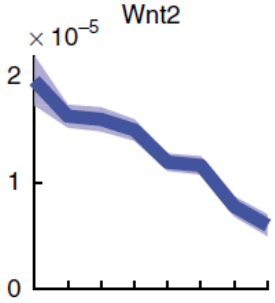
Expression (fraction of total UMI)

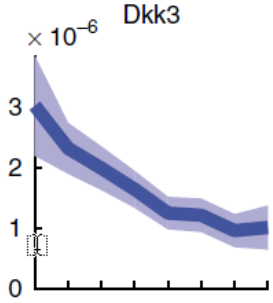
Rspo3 (green) is highly expressed in LEC surrounding the central vein (CV)

Red: Aqp1 (LEC marker)



Scale bar: 5µm





In LEC lining the central vein, Wnt2 (red) and Dkk3 (green) are expressed by distinct cells

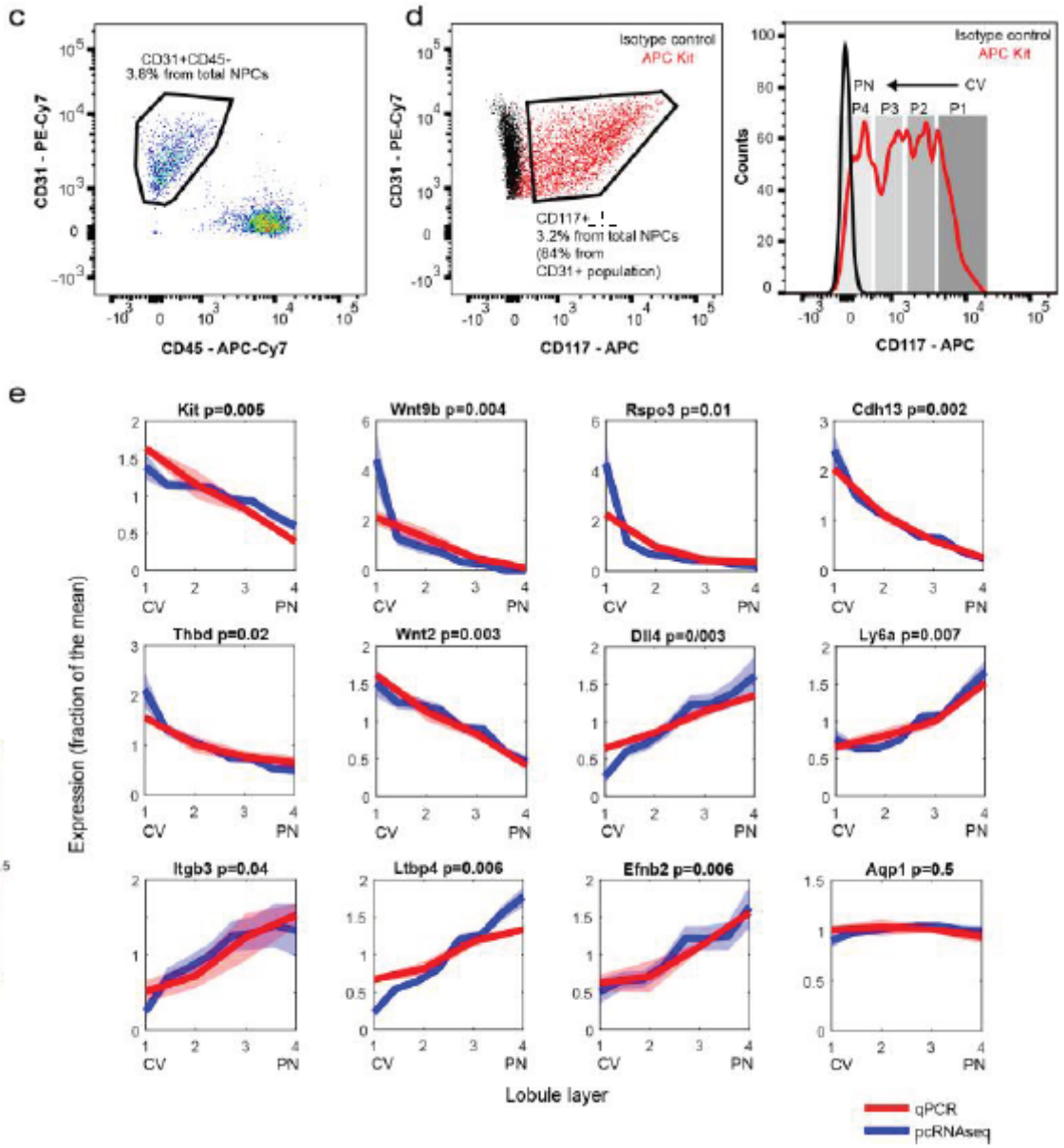
Dkk3: Dickkopf-related protein 3, antagonist of Wnt2

Characteristics of zonated LECs

- Pericentral LECs express *Wnt2, Wnt9b* and *Rspo3*, which are known to be essential for maintenance of hepatocyte zonation
  - Discovery of more pericentral LEC markers: *Thbd, Cdh13, Fabp4,* ***Kit*** (can be used as sorting markers)
  - Differentially expressed in central vein LECs (*Wnt2*↑) and sinusoidal LECs (*Wnt2* ↑ ↑) → finer classification of LECs
  - Genes repressed in pericentral LECs: *Bmp2, Stab1*

Thbd: Thrombomodulin
Cdh13: Cadherin 13
Fabp4: fatty acid binding protein 4
Bmp2: Bone morphogenetic protein 2
Stab1: Stabilin 1

FACS sort based on *Kit* (*Cd117*) expression



- Gating for LECs (CD31+CD45-)
- 4 sorted groups with different levels of *Kit* expression

- The 4 groups show expected levels of zonated LEC gene expression (qPCR, paired cell RNA-seq)

Paired-cell sequencing: Summary

- Development of paired-cell sequencing: spatial information is extracted from endothelial cells, based on spatial information of attached parenchymal cells → reconstruction of location within lobule

- Liver endothelial cells (LECs) showed spatial transcriptomic heterogeneity, based on location along lobule radial axis

- Uncovered molecular signature of pericentral LECs

- Zonated expression of markers (here: *Kit*) allows sorting of specific LEC subsets

- Paired-cell sequencing could be applied to many more tissues, tumours etc.

Limitations

- Hepatocytes are often adjacent to more than one type of LEC (arterial, venous, sinusoidal) → could not be distinguished
- In FISH, different types of endothelium showed different expression of *Rspo3*, *Wnt* etc.

Thank you for your attention