Single cell analysis of midbrain dopamine neurons
- Thomas Perlmann lab (LICR/KI)

Åsa Björklund

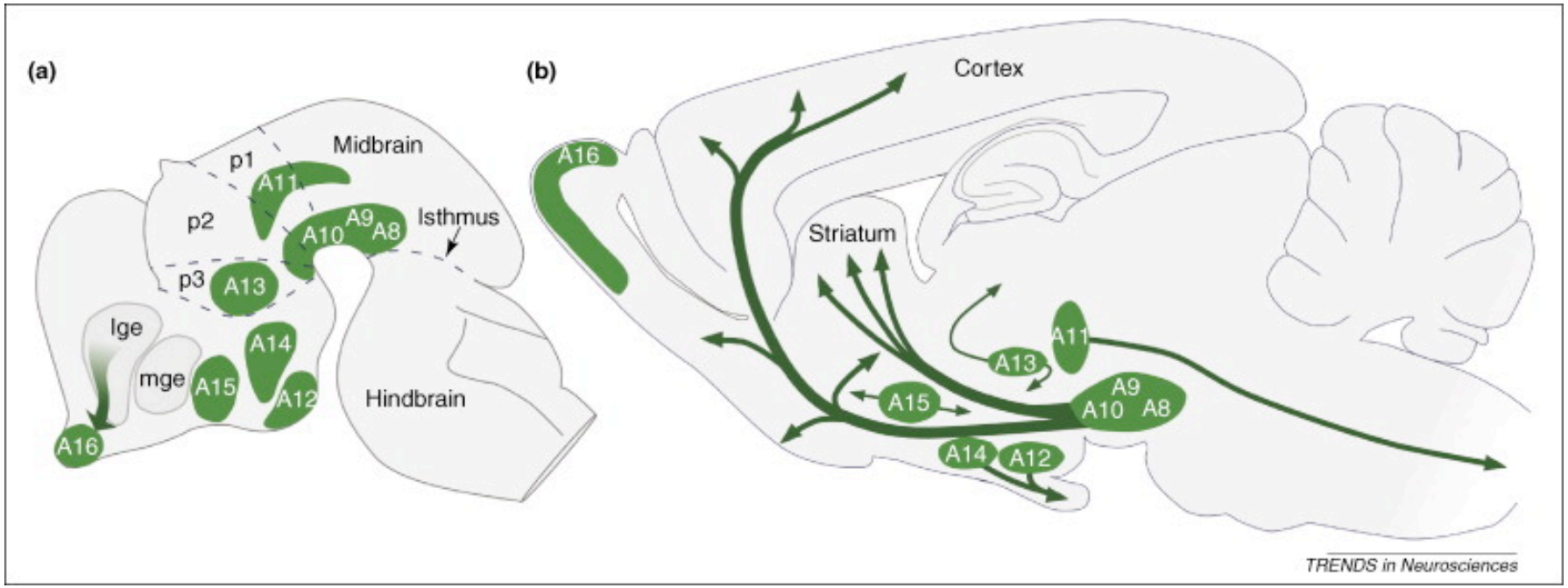
asa.bjorklund@scilifelab.se

Midbrain dopaminergic neurons

Less than 1% of the neurons in a brain
Important for:

- motivation & reward systems
- motor behavior
- working memory

Parkinsons Disease – death of DA neurons in Substantia Nigra
Successful grafting with ES derived DA neurons in 80-90s
Animal trials with iPS derived DA neurons ongoing



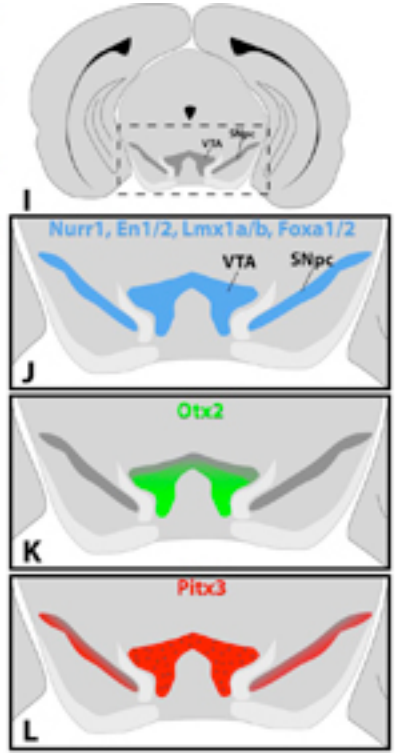
(Björklund et al. 2014)

Pitx3 eGFP knock in mice labels mature DA neurons.

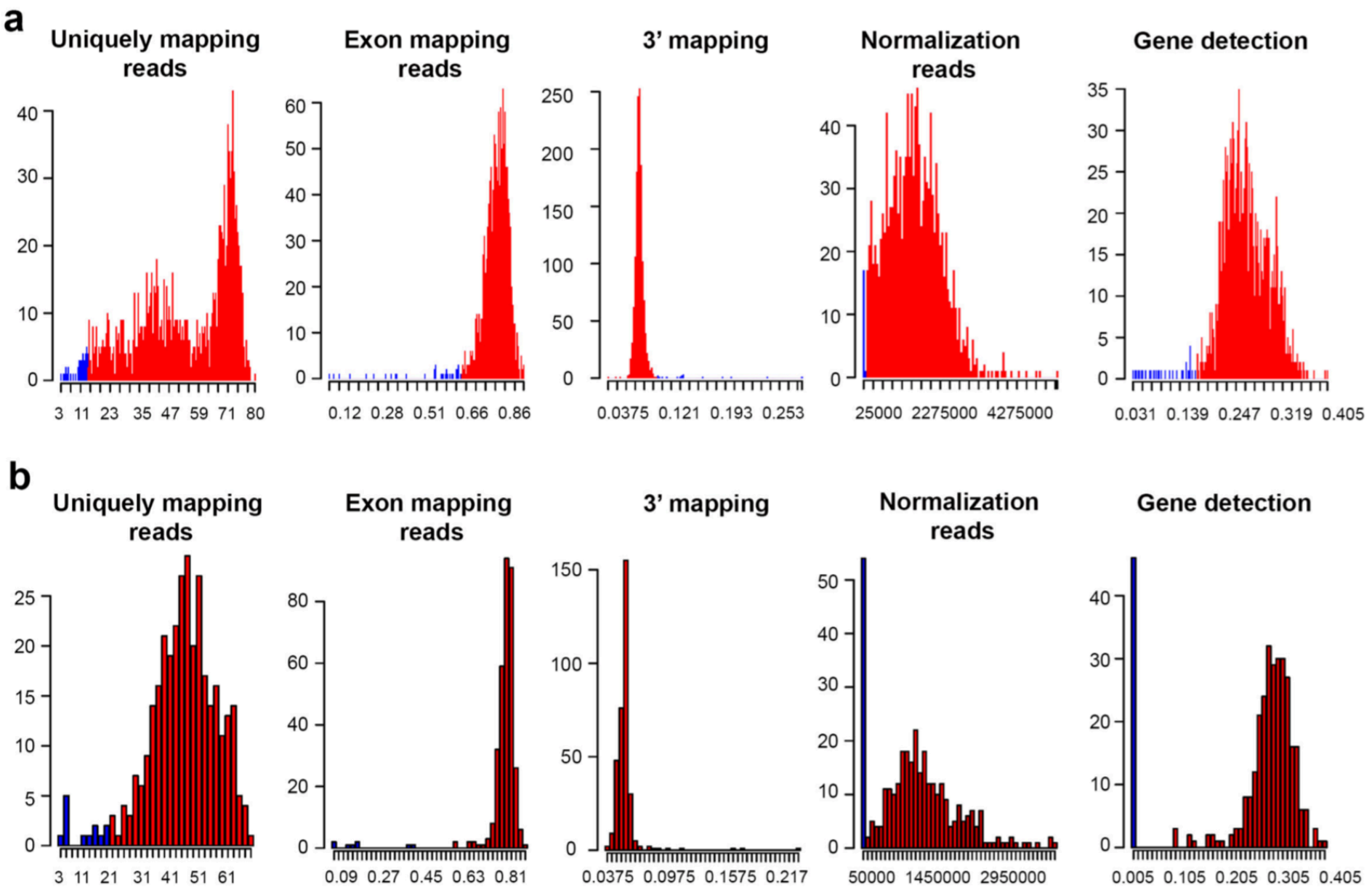
FACS sorted eGFP positive cells from midbrains of two mouse strains.

All libraries prepared in the Perlmann lab

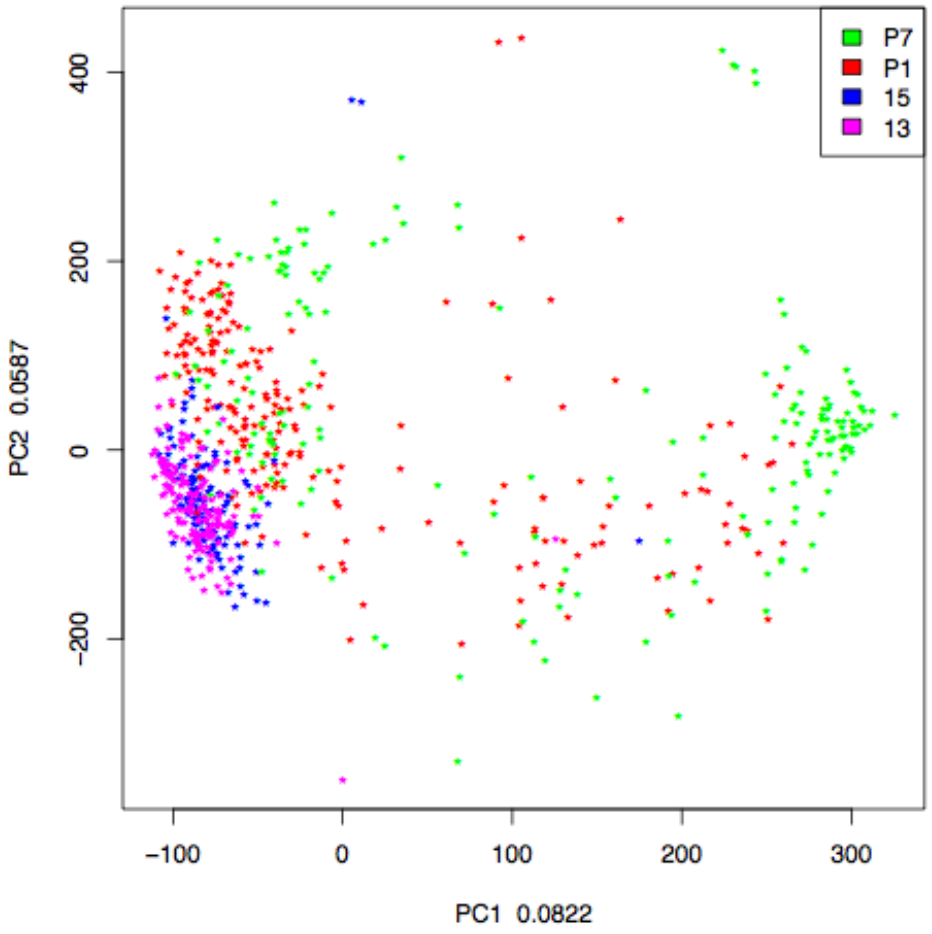
- eGFP Pitx3 heterozygote mice at:
  - Embryo - E13.5, E15.5, E18.5
  - Juvenile - P1, P7
  - Adult - P90
- eGFP Pitx3 homozygote (Pitx3 double KO) mice at:
  - E13.5
  - P1
- Total **1395** SmartSeq2 libraries after quality control

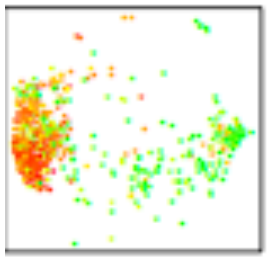


Quality Control

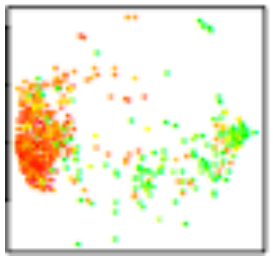


Quality Control – removal of non Pitx3 cells

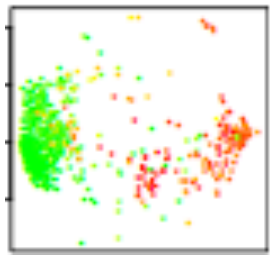




Pitx3 expression

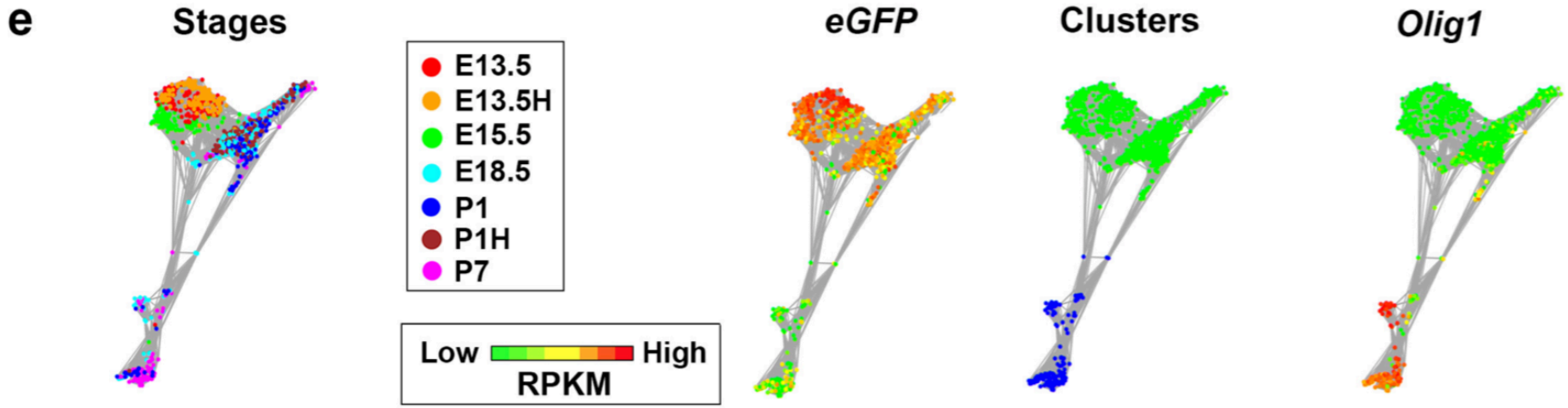


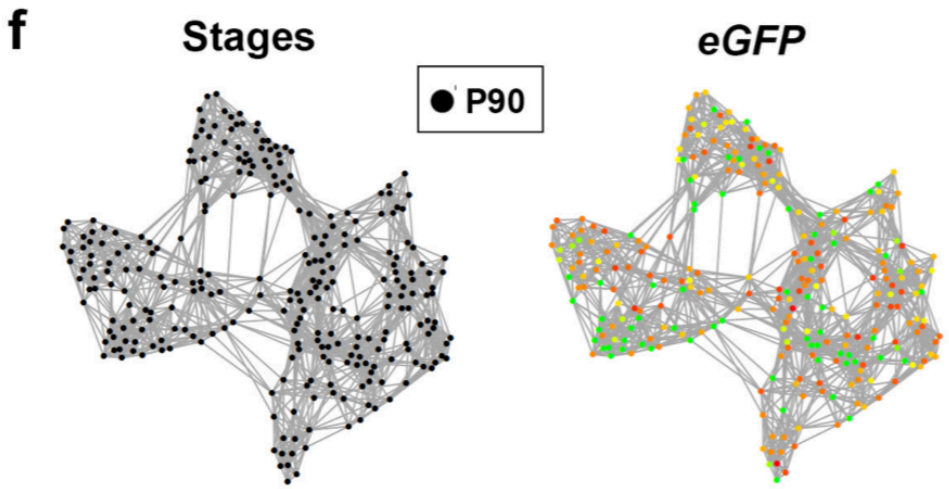
eGFP expression



Olig1 expression

Graph based clustering to remove non-DA neurons





g

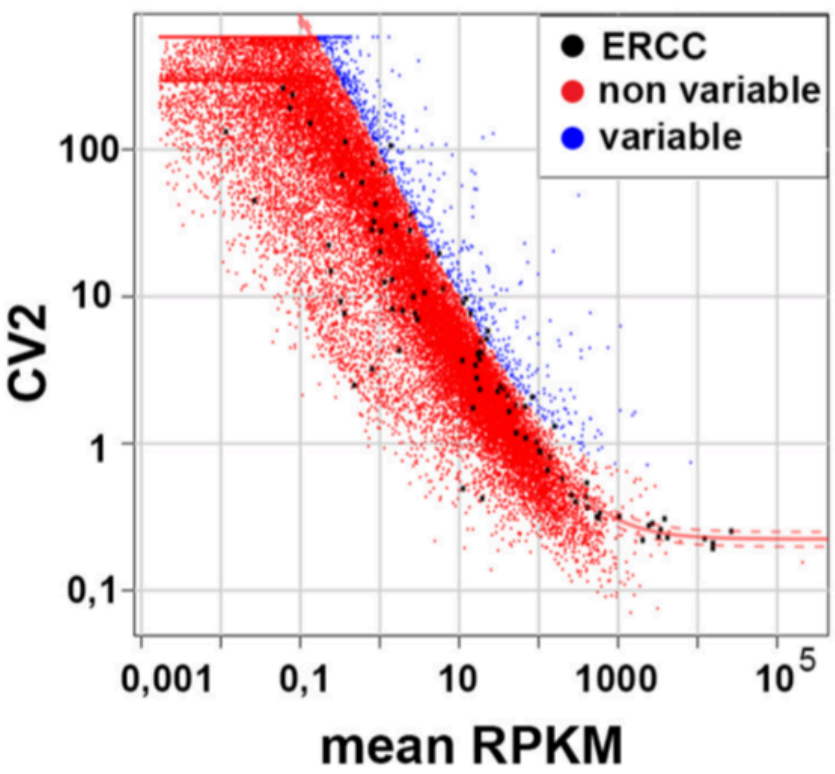
Develop. Stage	Number of Cells
E13.5	135
E15.5	140
E18.5	181
P1	258
P7	80
P90	312
E13.5H	141
P1H	148

Analysis of Pitx3 Heterozygote cells

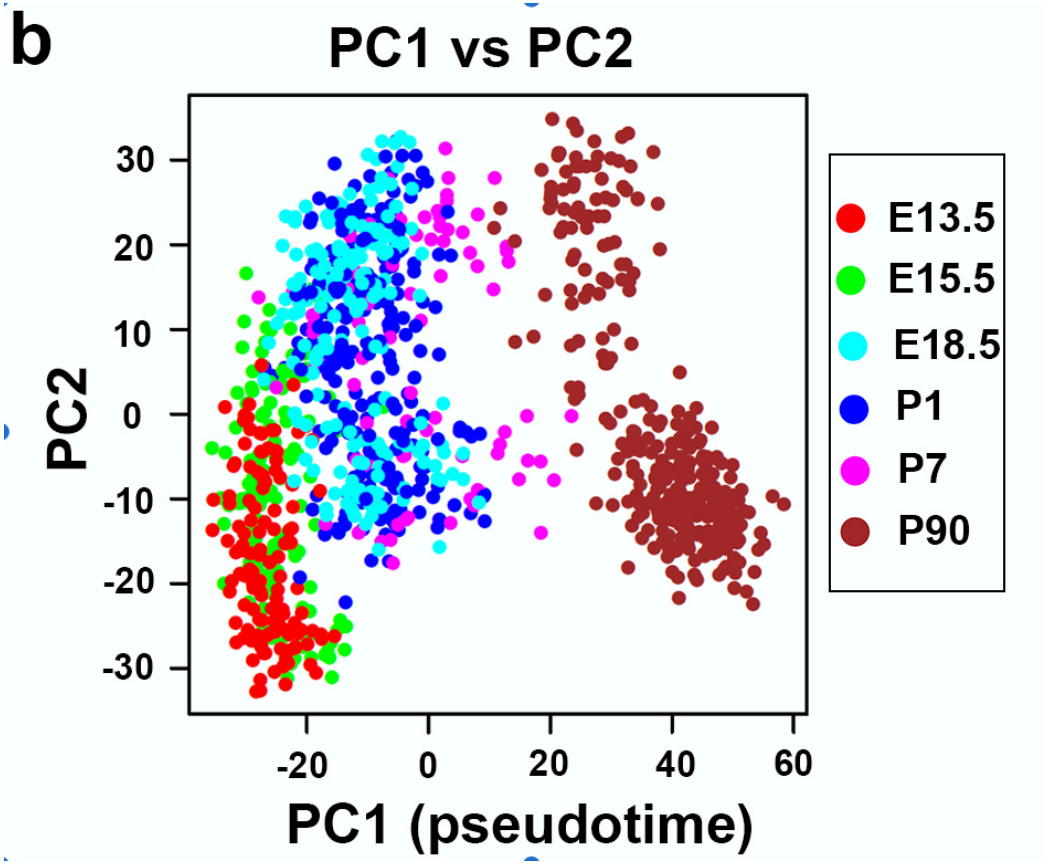
- Main aims
  - Find possible subgroups of cells
  - Find marker genes for the groups
  - Understand development of the lineage

Selection of variable genes

- Using ERCC spike-ins and Brennecke method
  - 453 variable genes using all ages
  - 300 – 800 variable genes using individual ages

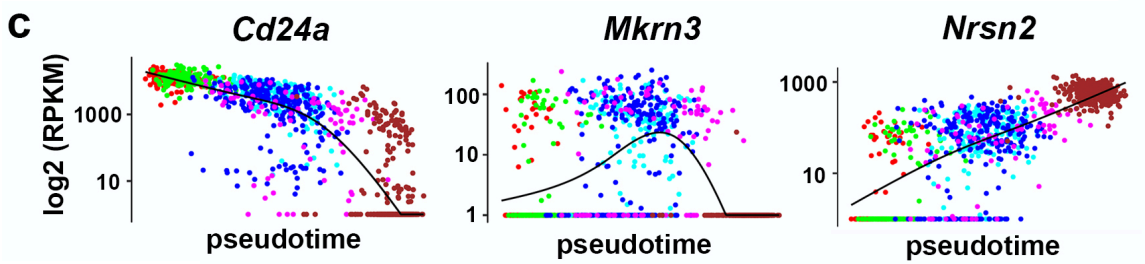


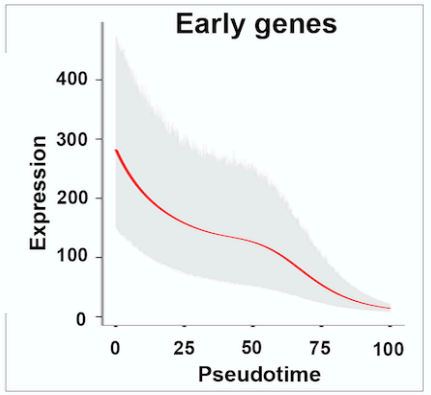
Initial PCA – mainly time separation

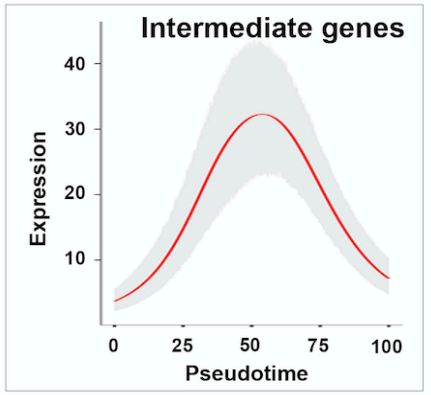


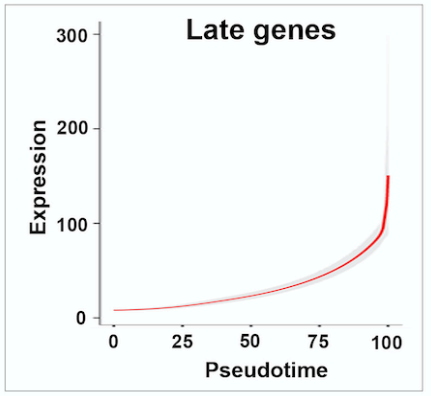
Clear separation by 3 maturation stages: early embryo (E13/E15), perinatal (E18-P7) and adult (P90)

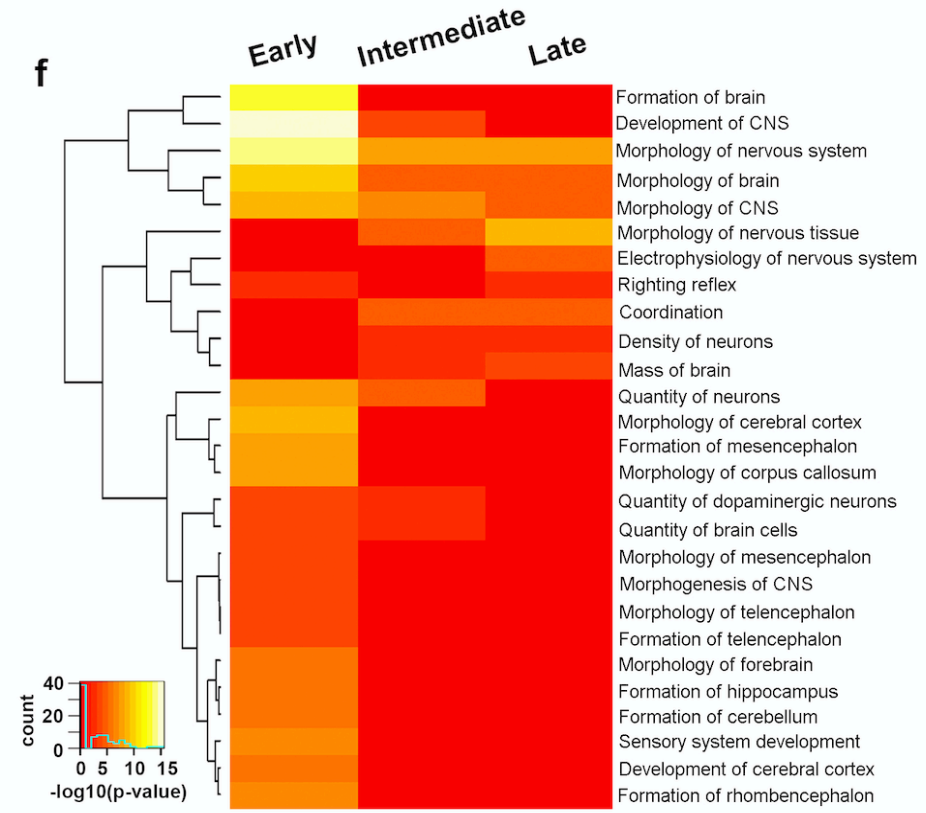
Genes along pseudotime



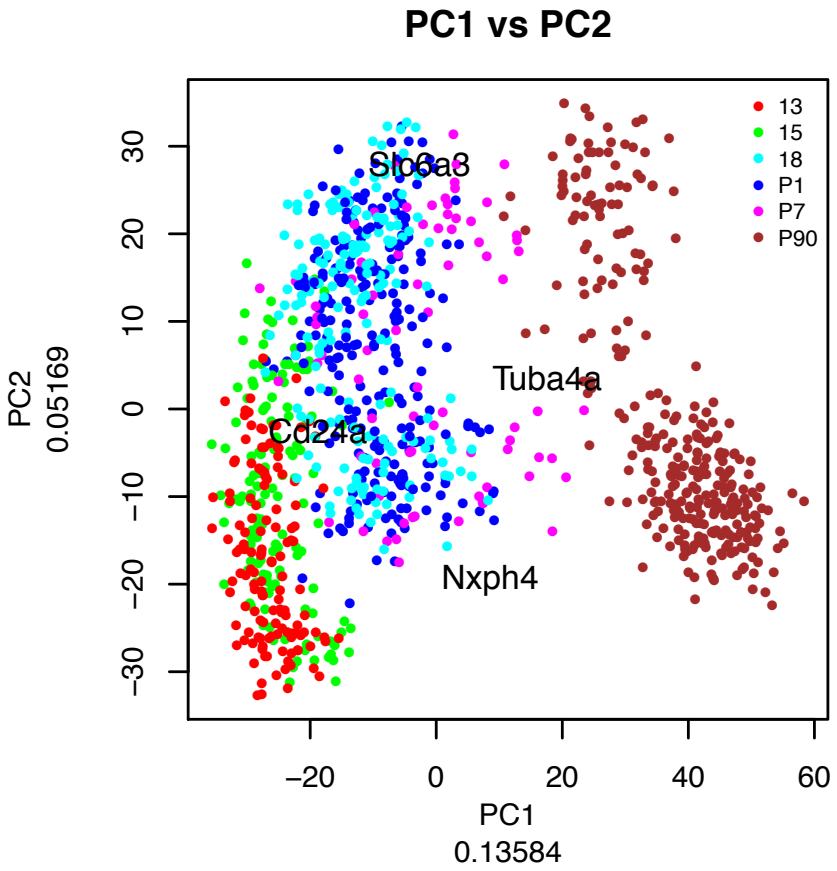


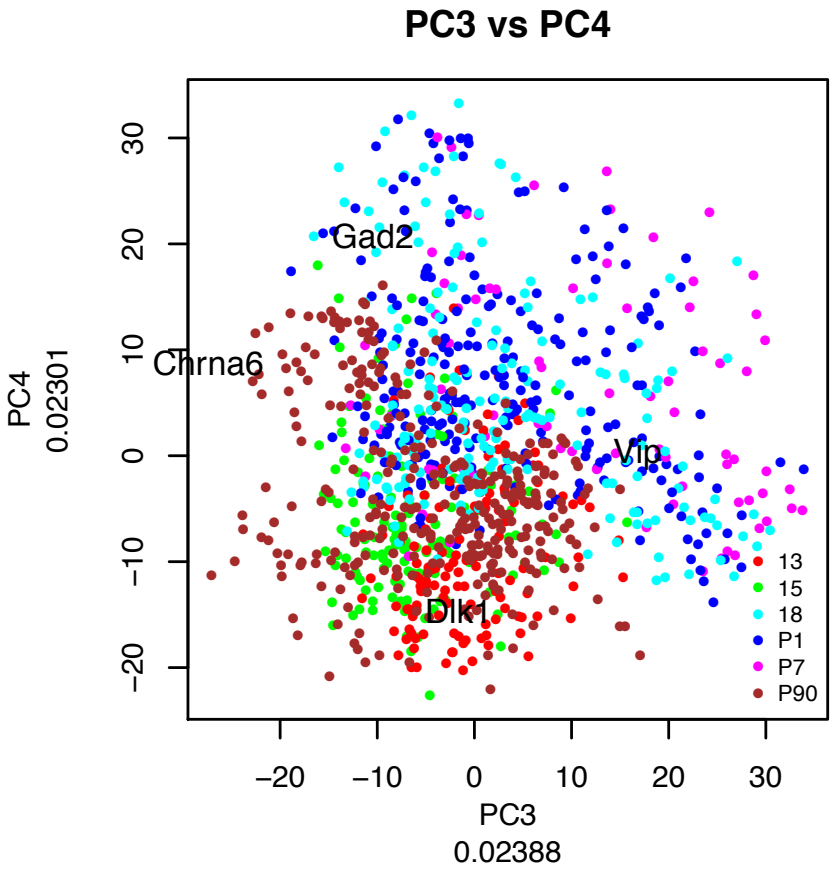






PCA



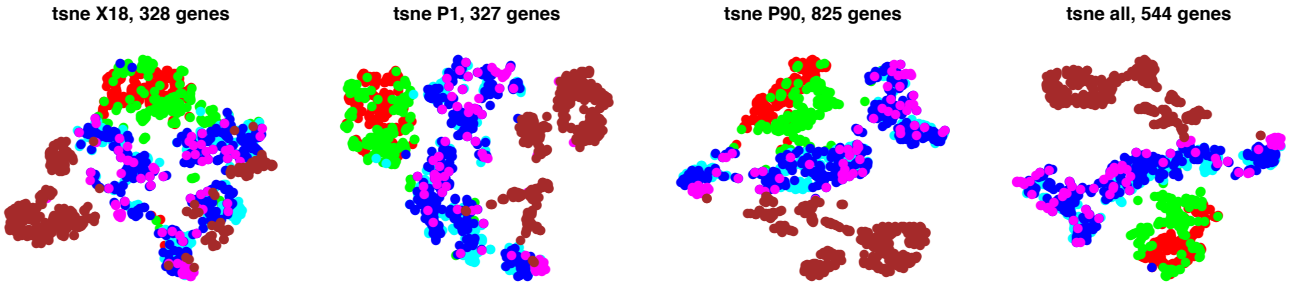


Main problem: All cells are very similar – same celltype. Hard to identify subgroups since the signal is weak relative to developmental time, noise and other sources of variation.

Top loadings of PC2,PC3 seems to define our lineages: Slc6a3 (Dat), Nxph4, Gad2 and Vip

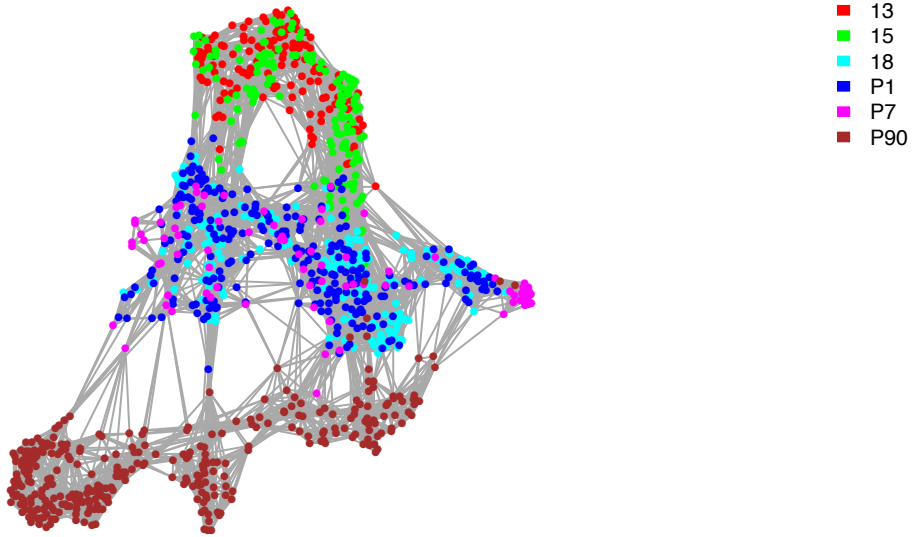
t-distributed stochastic neighbor embedding (tSNE) and igraph

tSNE with different sets of variable genes



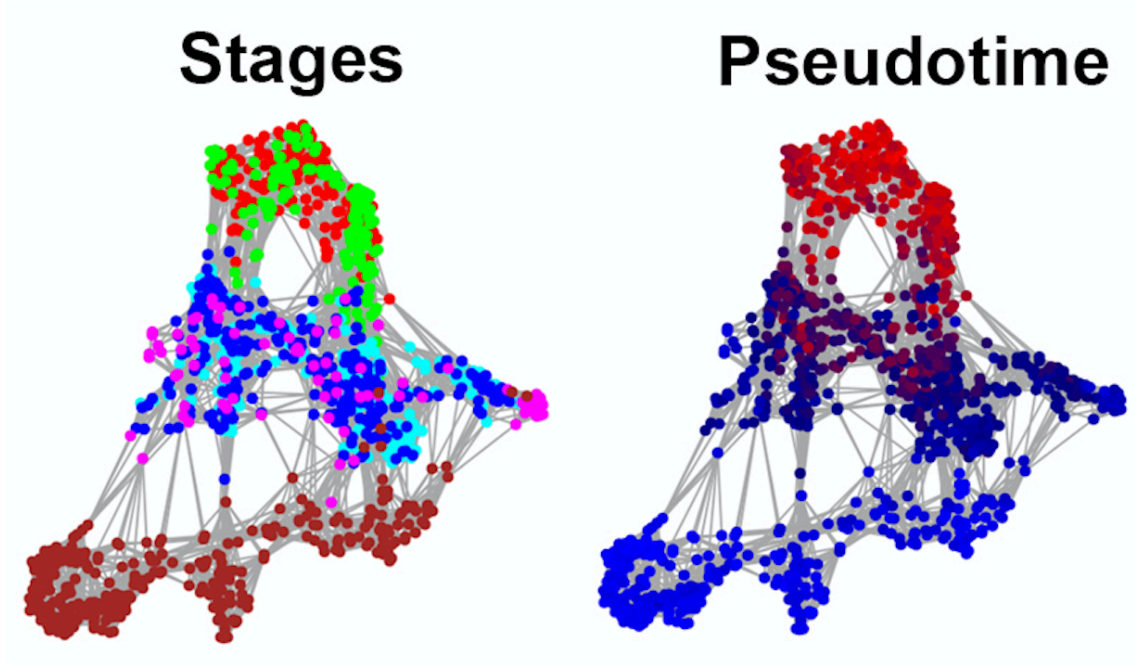
Weighted edges for 5 nearest neighbors – merged for all tSNEs

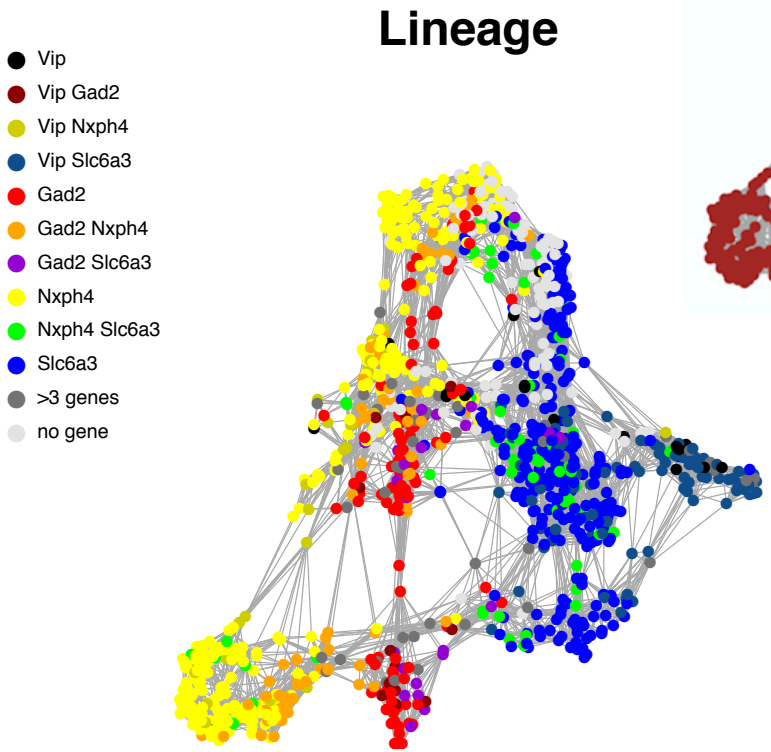
Weighted igraph network



Variable genes at each stage should reflect lineage differences more than developmental differences

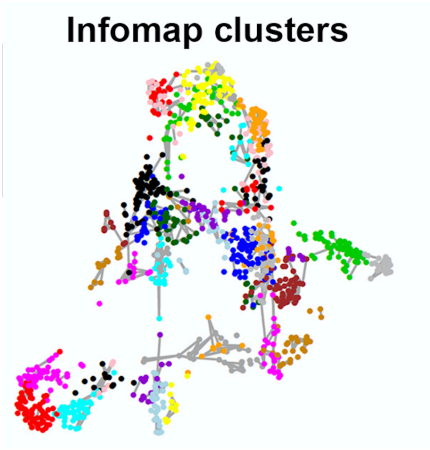
Classifying cells into subgroups





Using genes Vip, Gad2, Nxph4 and Slc6a3 cells are colored by expression of one or more of these genes.

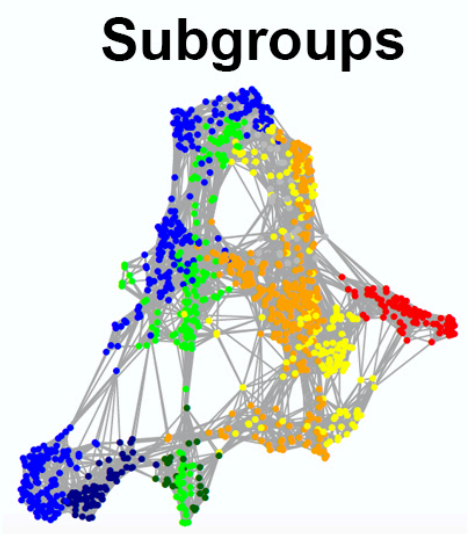
Classifying cells into subgroups



Clusters defined by infomap community detection.

Each cluster classified into a subgroup based on proportion of cells exclusively expressing one of our markers.

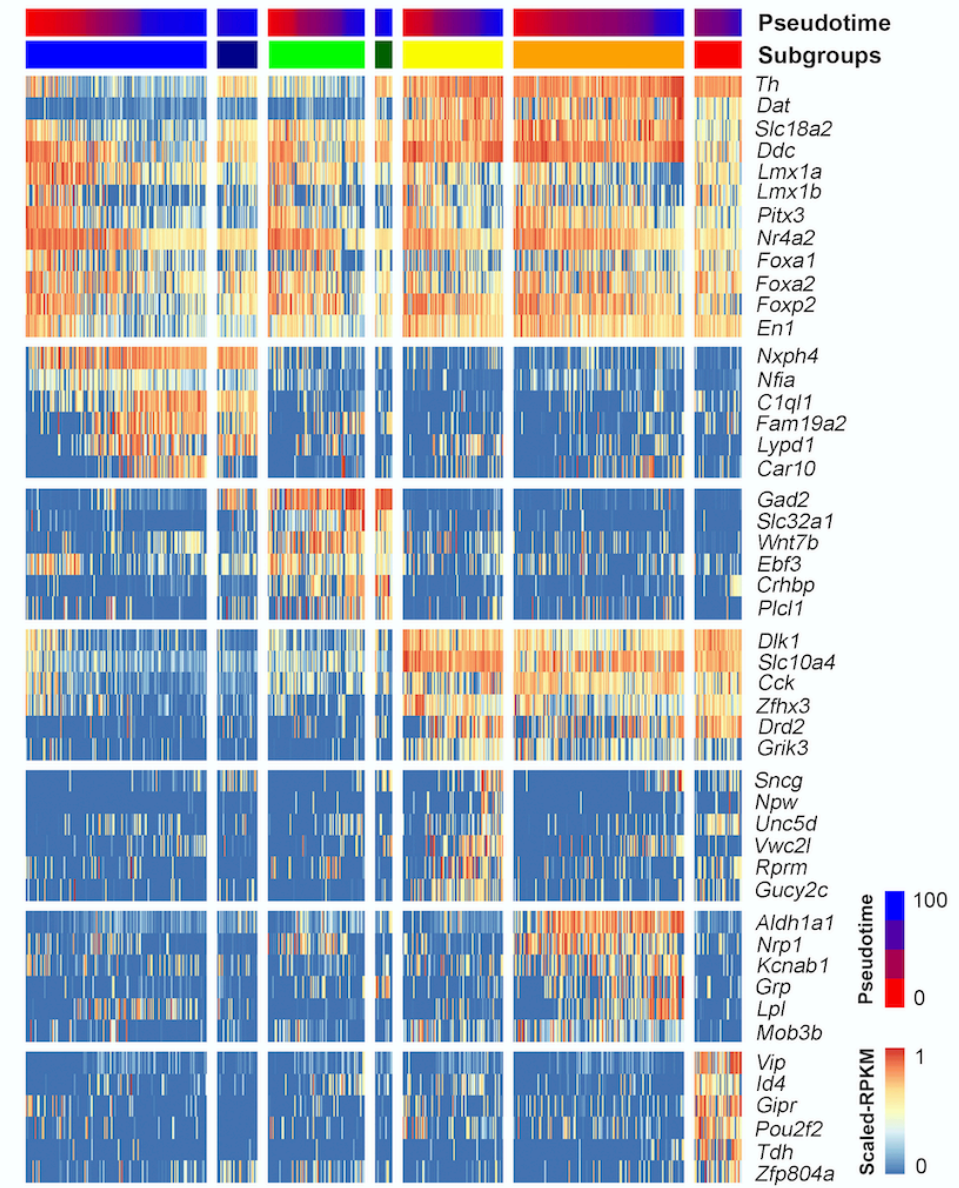
Manual definition of Gad2+ and Nxph4+ lineage into Th high/low.



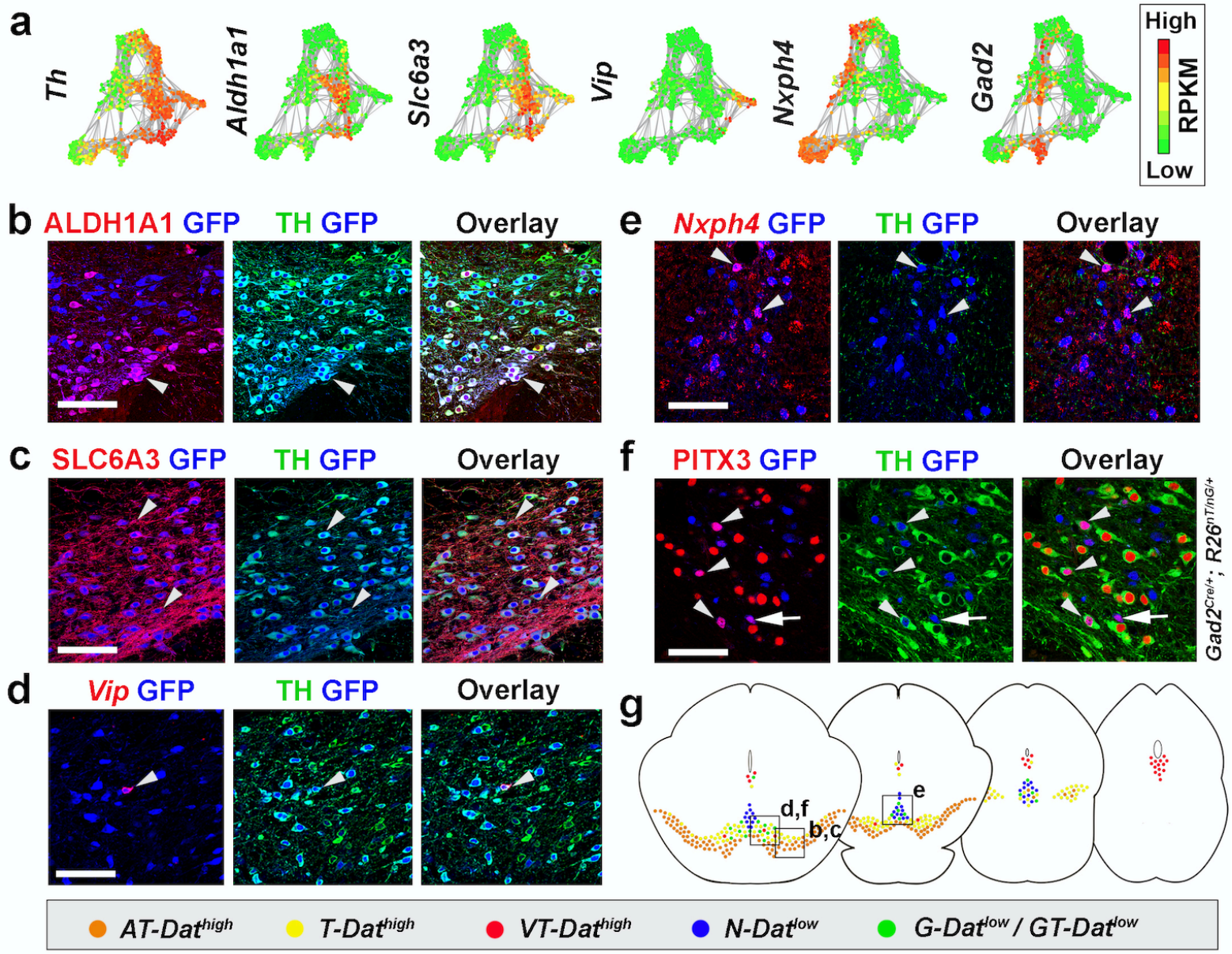
Nxph4 Dat^{low}	*N-Dat^{low}*
Nxph4 Th Dat^{low}	*NT-Dat^{low}*
Gad2 Dat^{low}	*G-Dat^{low}*
Gad2 Th Dat^{low}	*GT-Dat^{low}*
Th Dat^{high}	*T-Dat^{high}*
Aldh1a1 Th Dat^{high}	*AT-Dat^{high}*
Vip Th Dat^{high}	*VT-Dat^{high}*
ND	

Marker gene discovery

Differential expression between the clusters both across all stages and at one developomental stage at a time was done using SAMseq.

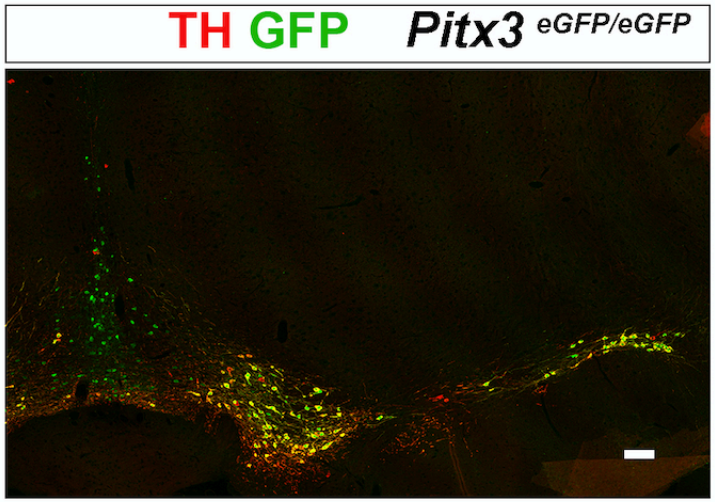


Validations with immunohistochemistry

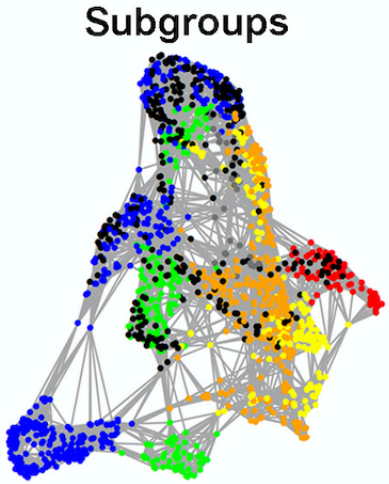


Pitx3 double KO cells

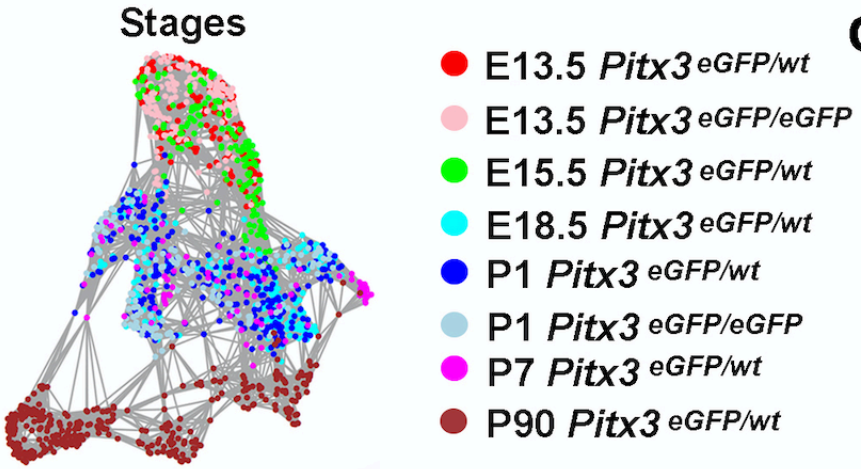
a



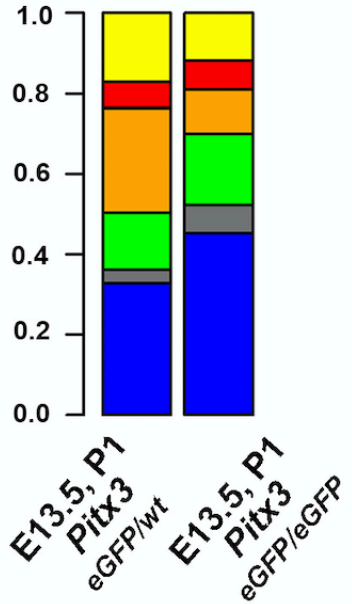
b



c



d

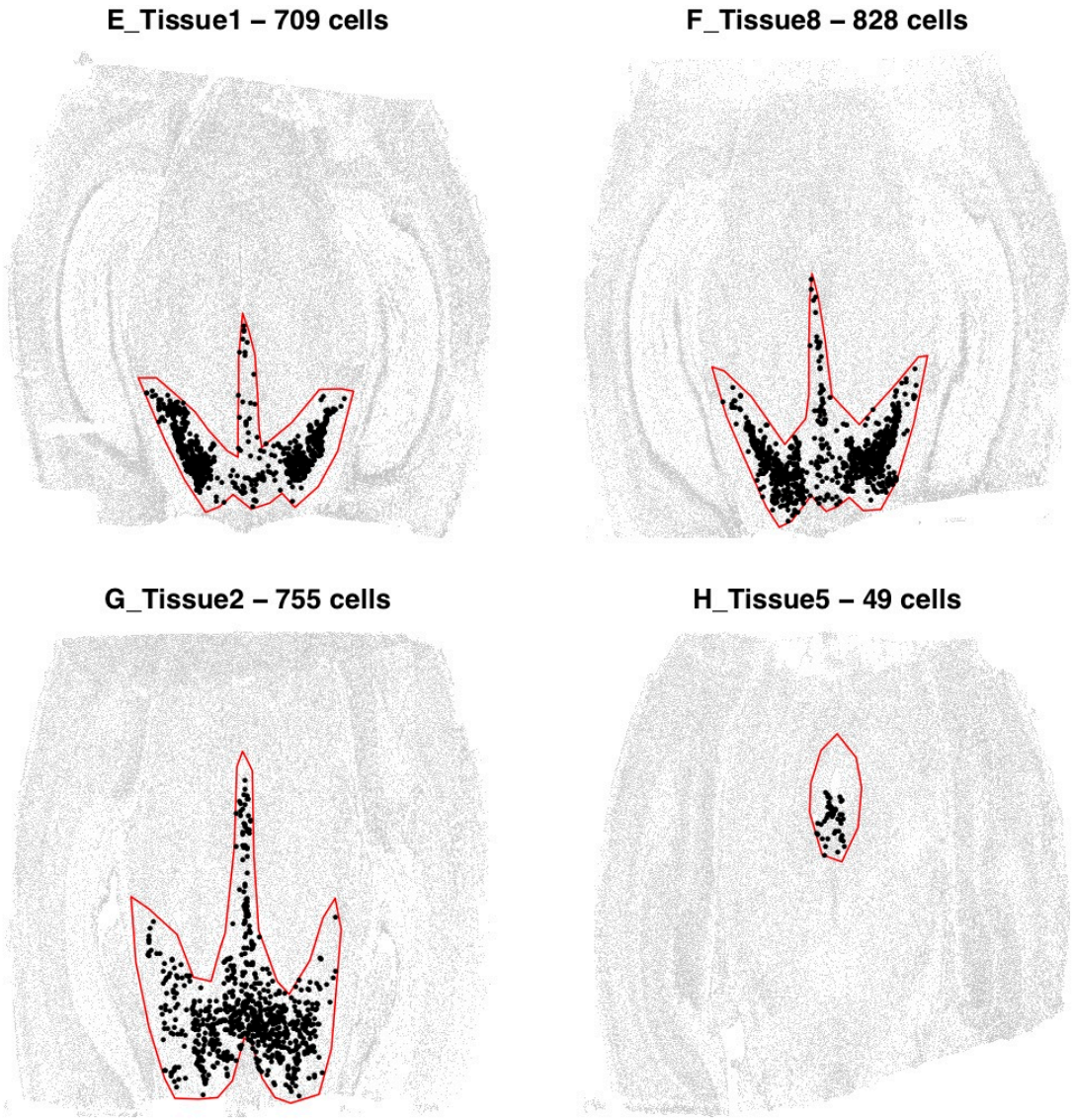


Padlock probe – *in situ* sequencing

- Method developed at Mats Nilsson lab (SU)
- Selected 49 genes among our differential expression data:
  - Fairly high expression
  - Markers for subgroups
  - Some general markers to exclude other celltypes

- Cells defined as expanded area around nuclei (from Dapi staining)

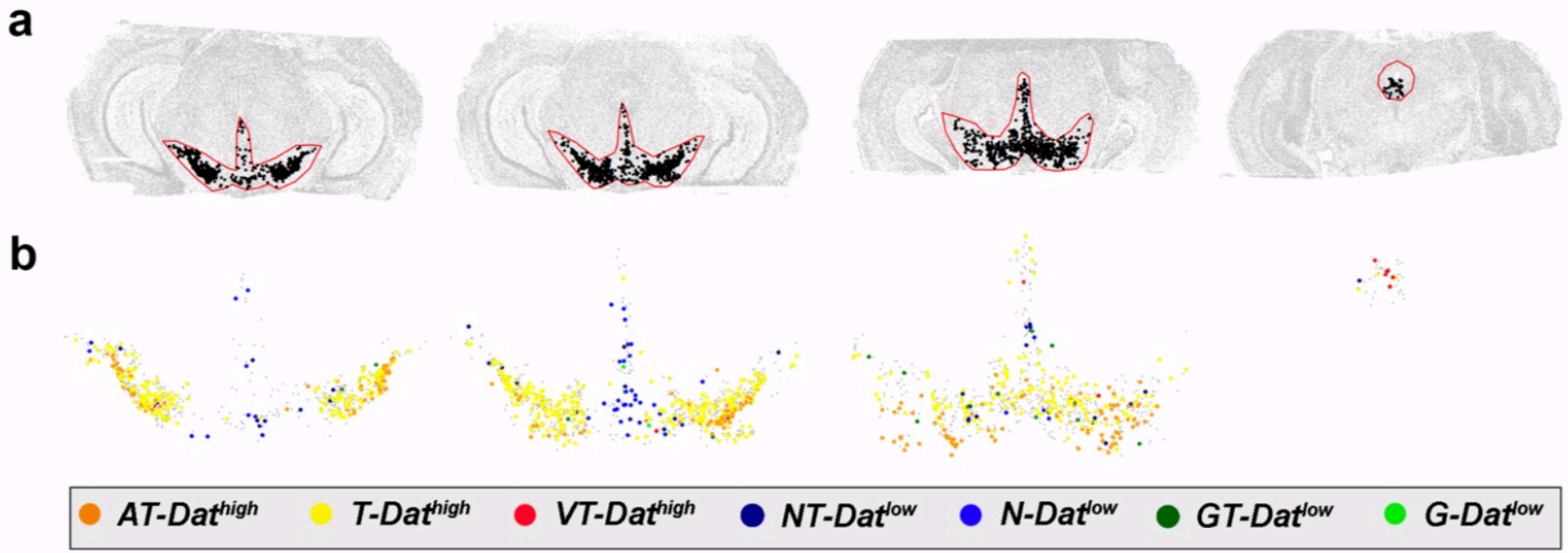
Filter *in situ* data



- Filtering of *in situ* cells for DA neurons:
  - Cells in selected regions
  - Require epression of Pitx3, EGFP or Th.
- Keep only cells with at least 3 subgroup marker genes.
- Out of around 60K cells per section only 2141 cells kept.

Predict subgroup using Random Forest

- Convert both scRNAseq data and in situ data to rank based matrices.
- Train random forest with scRNA-seq data − using cluster membership
- Predict subgroup for *in situ* cells
- Several rounds of training − prediction − only keep consistent predictions.



Conclusions

- Most extensive classification of subgroups of midbrain DA neuron subtypes to date.
- Several verification experiments with antibody staining (also with human tissue), in situ sequencing, retrograde labeling of innervation.
- Main issue in the data was that time separation was much stronger than separation of the subgroups.

Shiny apps

- http://shiny.rstudio.com/
- Interactive R applications used to present the data
- http://rshiny.nbis.se/shiny-server-apps/shiny-apps-scrnaseq/perlmannlab_mouseDA