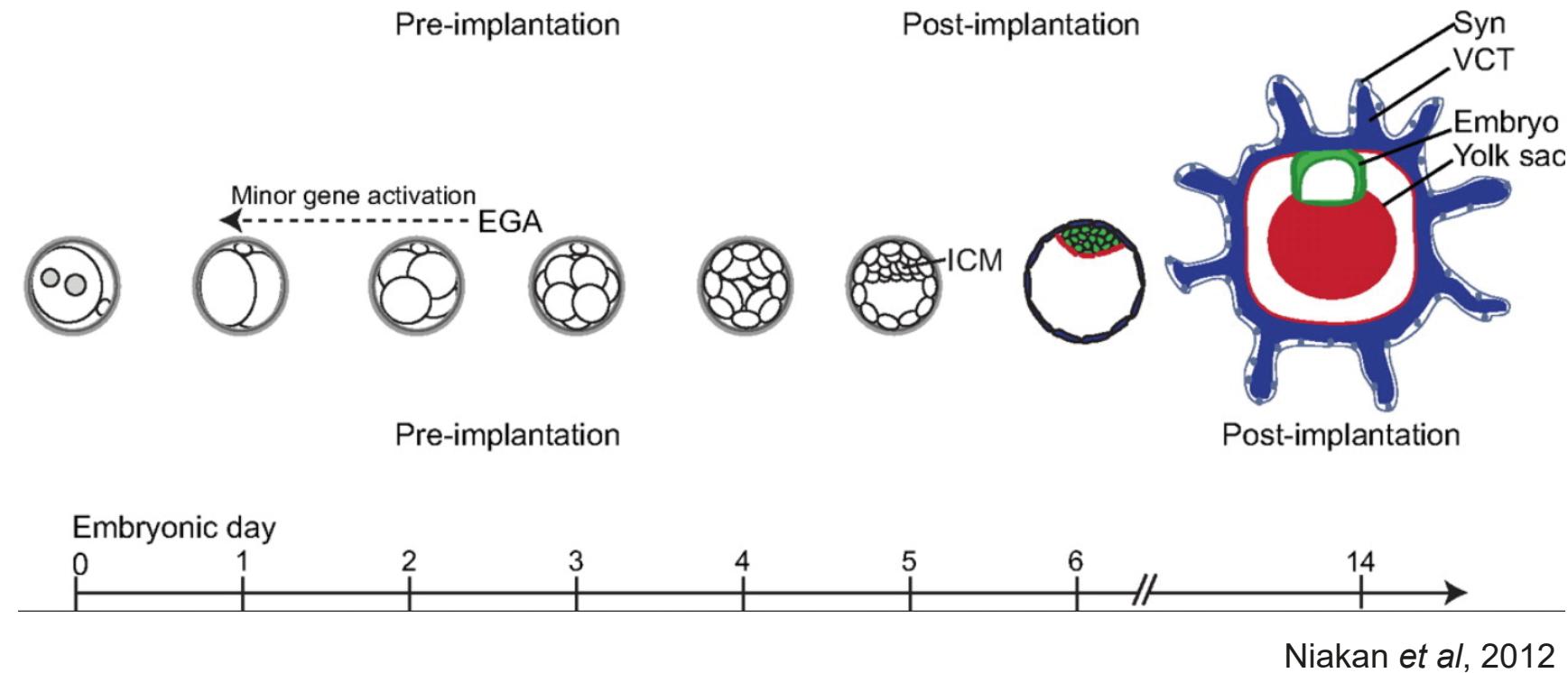


Integrated analysis of the early human embryo from multi-omics single-cell data

Gaël CASTEL, DUBii 2021,
Supervised by Audrey BIHOUEE, Eric CHARPENTIER & Laurent DAVID



BACKGROUND



Timing:

- From fertilization to day 14 of development
- Waves of DNA remethylation

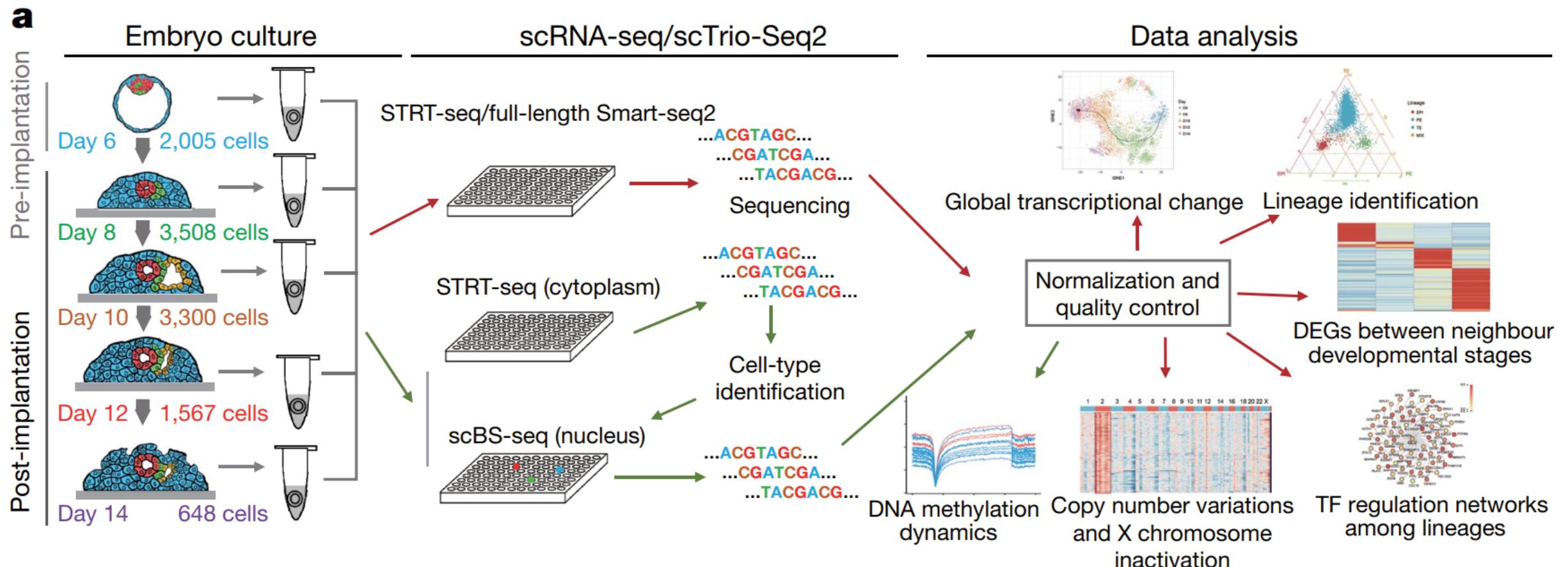
Lineage specification:

- Epiblast > fetus
- Primitive endoderm > yolk sac
- Trophectoderm > placenta

Context:

- Increasing amount of single-cell multi-omics data of the human embryo
 - But no comprehensive integrative analysis
 - Need to identify driver and passenger genes
- > **We will integrate DNA methylation and transcriptomic data at single-cell resolution**

DATASET

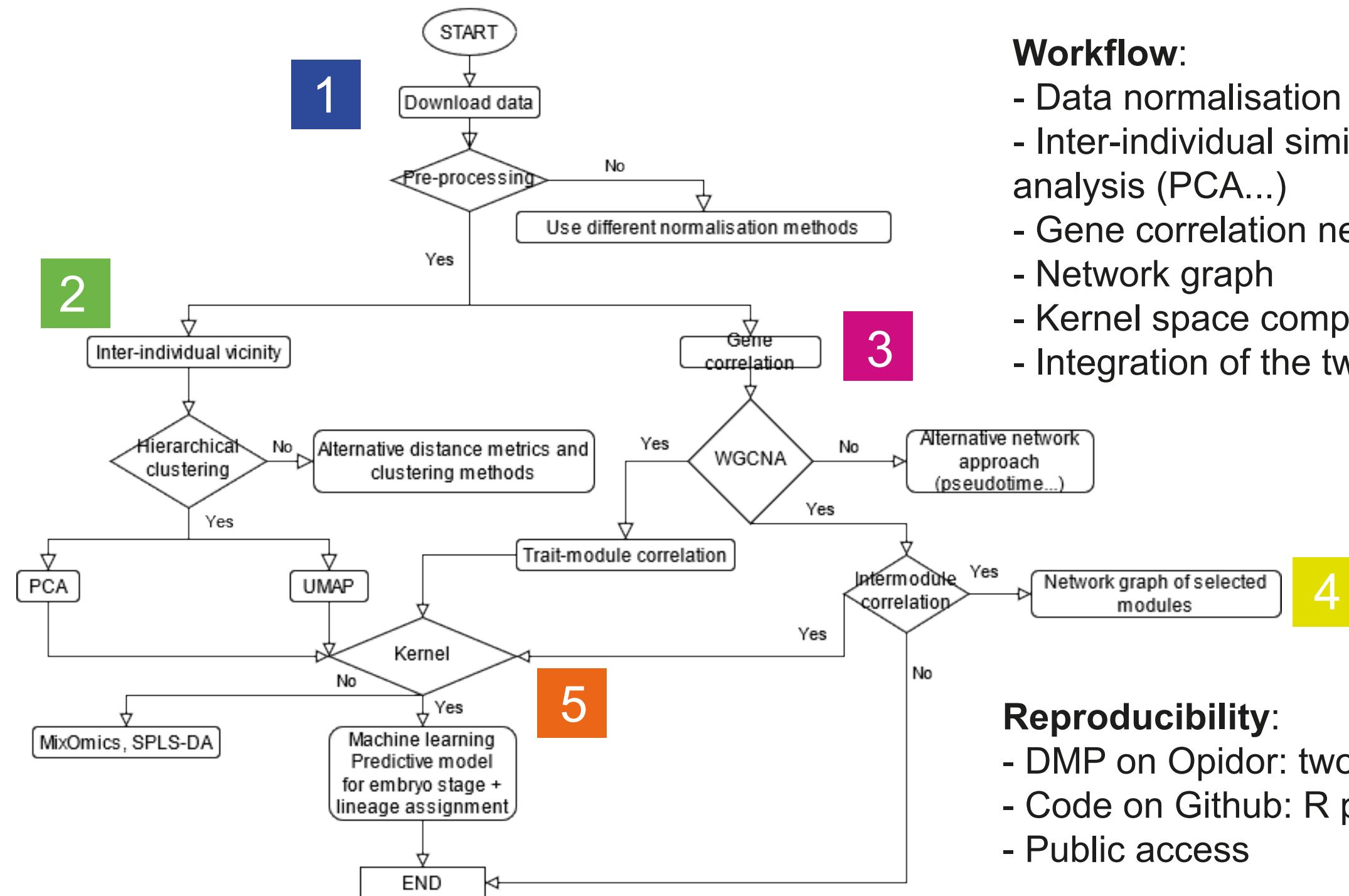


Zhou et al, 2019

- Day 6 to Day 12 human embryos
- 3 lineages (EPI, PE, TE)
- 130 cells

- Gene promoter methylation: single-cell PBAT } from the same cell
- Gene expression: single-cell RNA-Seq }

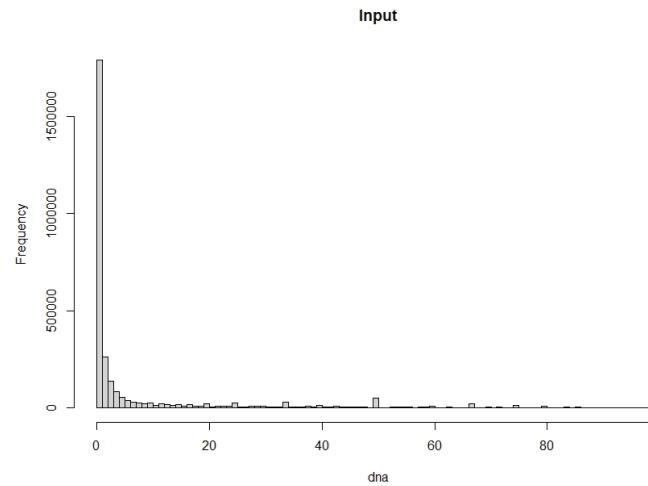
WORKFRAME: FLOWCHART



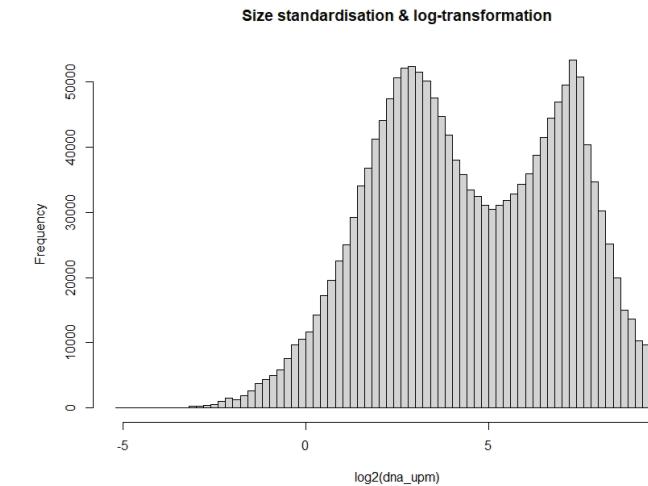
DATA PREPROCESSING

METHYLATION

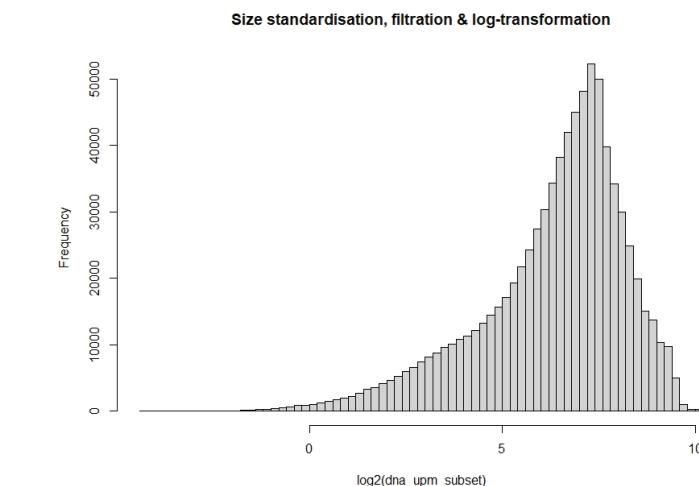
INPUT DATA



SIZE STD + LOG

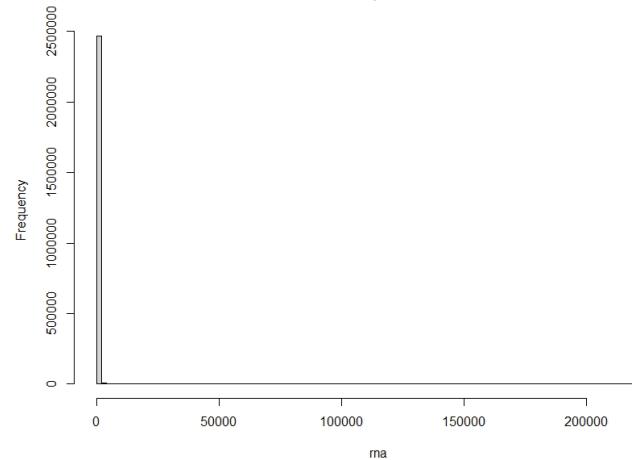


SIZE STD + VAR + LOG

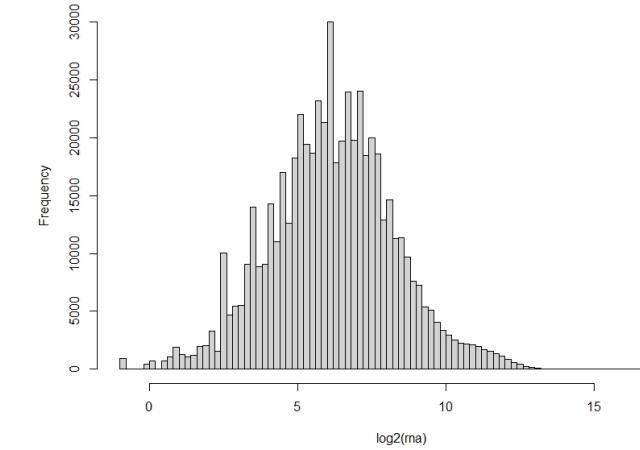


EXPRESSION

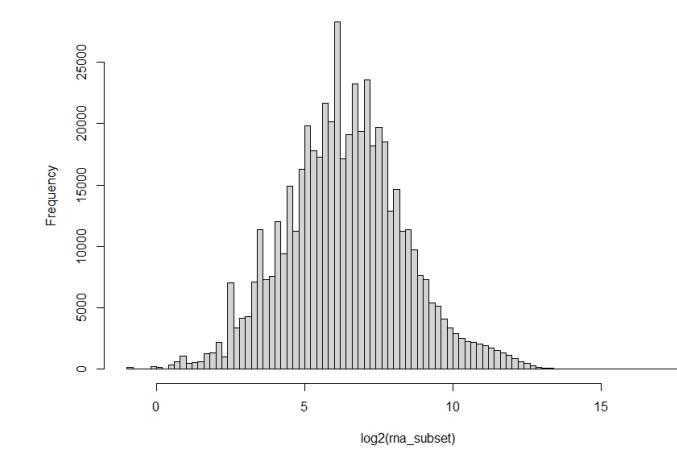
Input



Log-transformation



Size standardisation, filtration & log-transformation

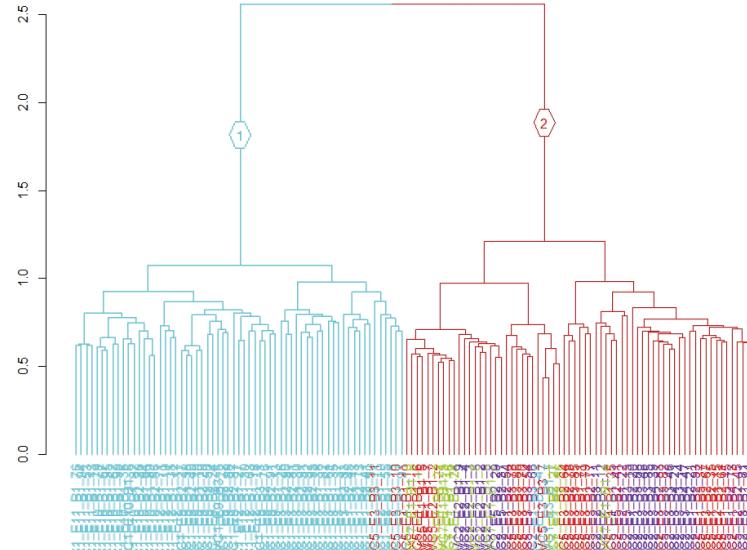


- Notably, DNA methylation levels follow a bimodal law
- After normalisation steps, we approach a Gaussian distribution of the two data types

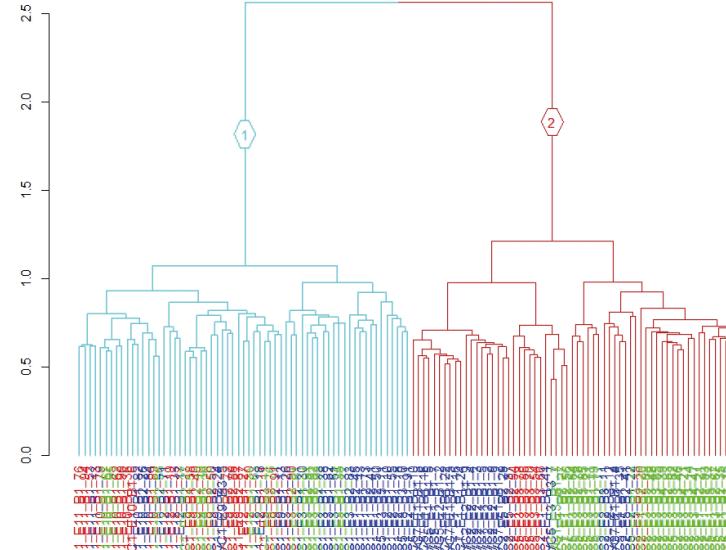
HIERARCHICAL CLUSTERING: CELL VICINITY

coloured by Day

METHYLATION



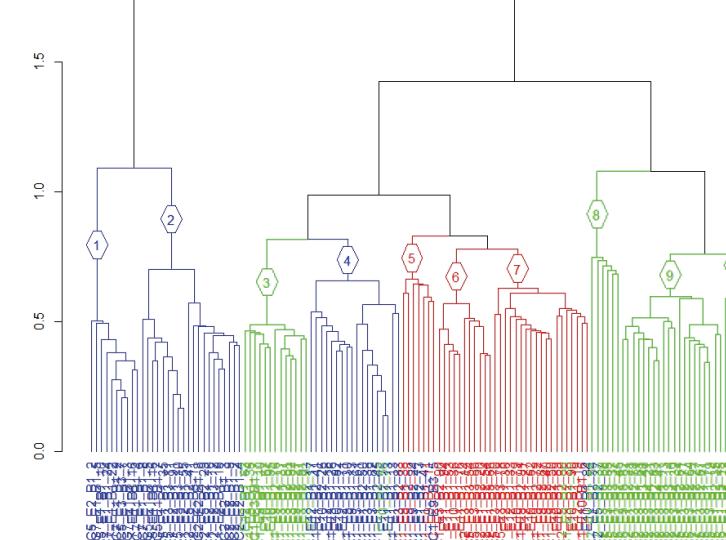
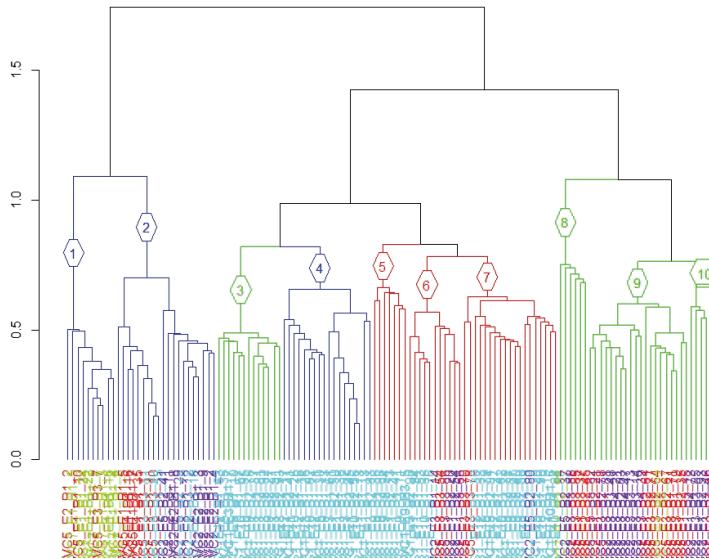
coloured by Lineage



- Gene promoter methylation is strongly linked to developmental stage until day 6

- Lineage specific patterns emerge at later timepoints

EXPRESSION



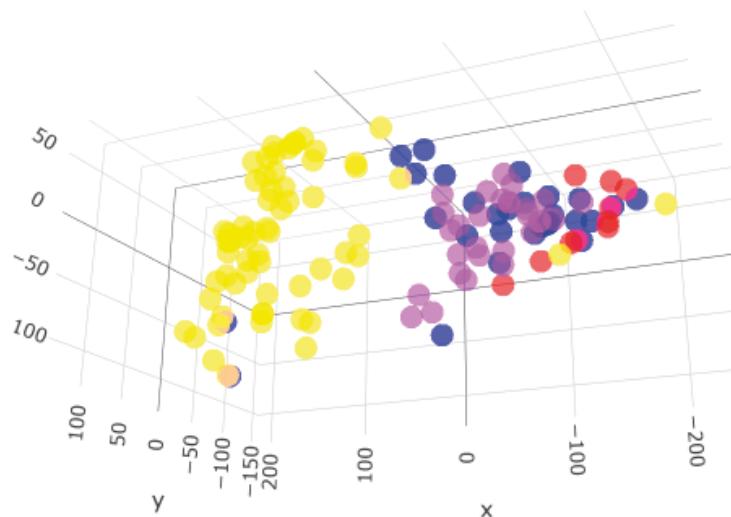
- Gene expression patterns are specific to cell lineages

DIMENSIONALITY REDUCTION: CELL VICINITY

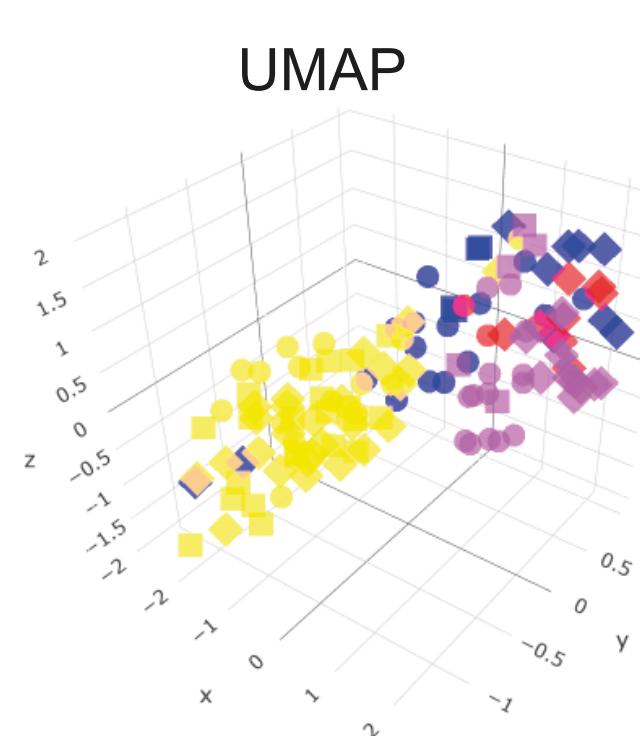
6

METHYLATION

PCA

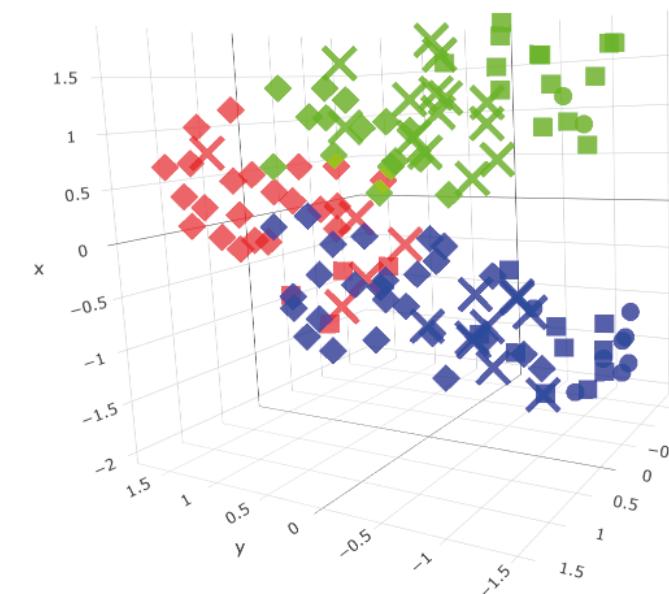
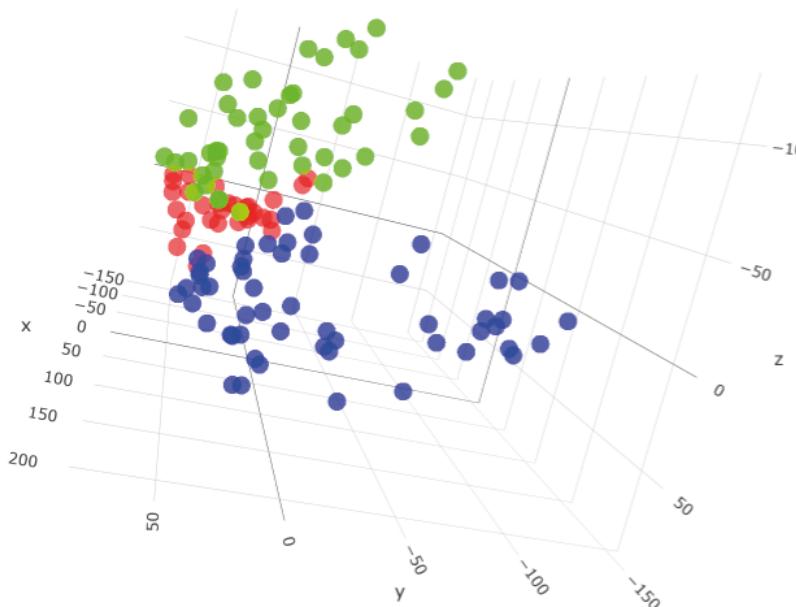


UMAP



- D10
- D12
- D6
- D8
- D10
- Epi
- D10
- ◆ D12
- PE
- ◆ D12
- TE
- D6
- Epi
- D6
- PE
- D6
- TE
- D8
- Epi
- D8
- PE
- D8
- ◆ TE

EXPRESSION

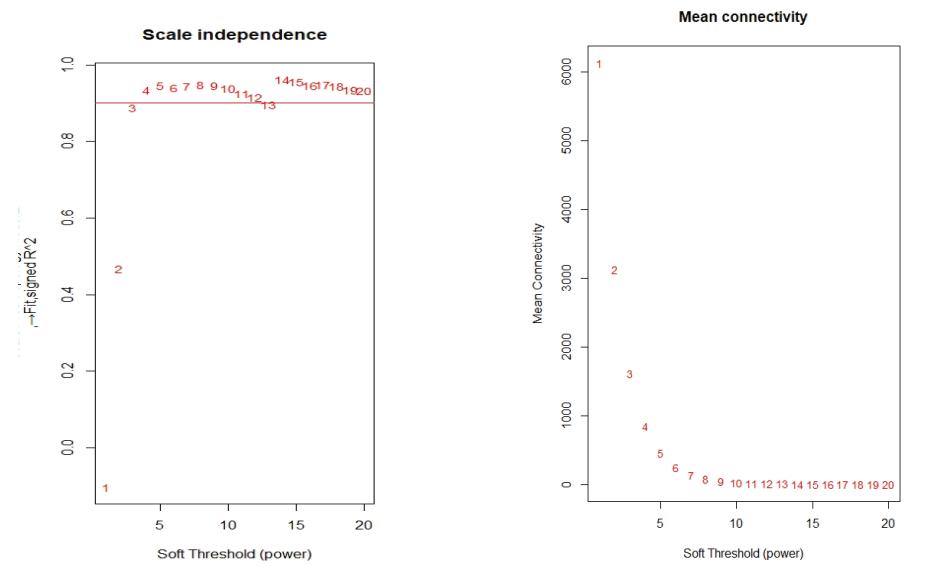


- Epi
- PE
- TE
- Epi
- D10
- ◆ Epi
- ◆ D6
- Epi
- ◆ D8
- PE
- D10
- PE
- D12
- PE
- D6
- PE
- D8
- X
- TE
- D10
- TE
- D12
- ◆ TE
- D6
- X
- TE
- D8

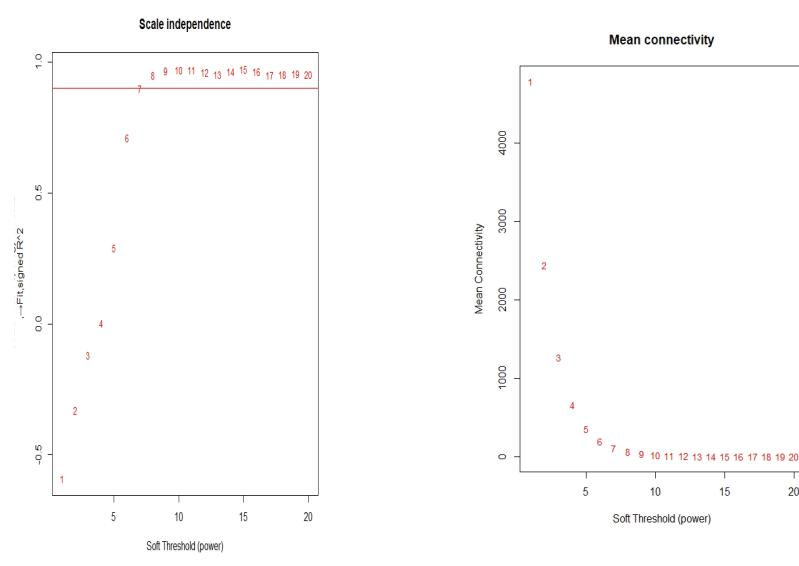
- PCA & UMAP produce close results
- Day 6 is clearly separated from later stages on DNA methylation
- Gene expression segregates the 3 earliest lineages: EPI, PE, TE

WGCNA

METHYLATION



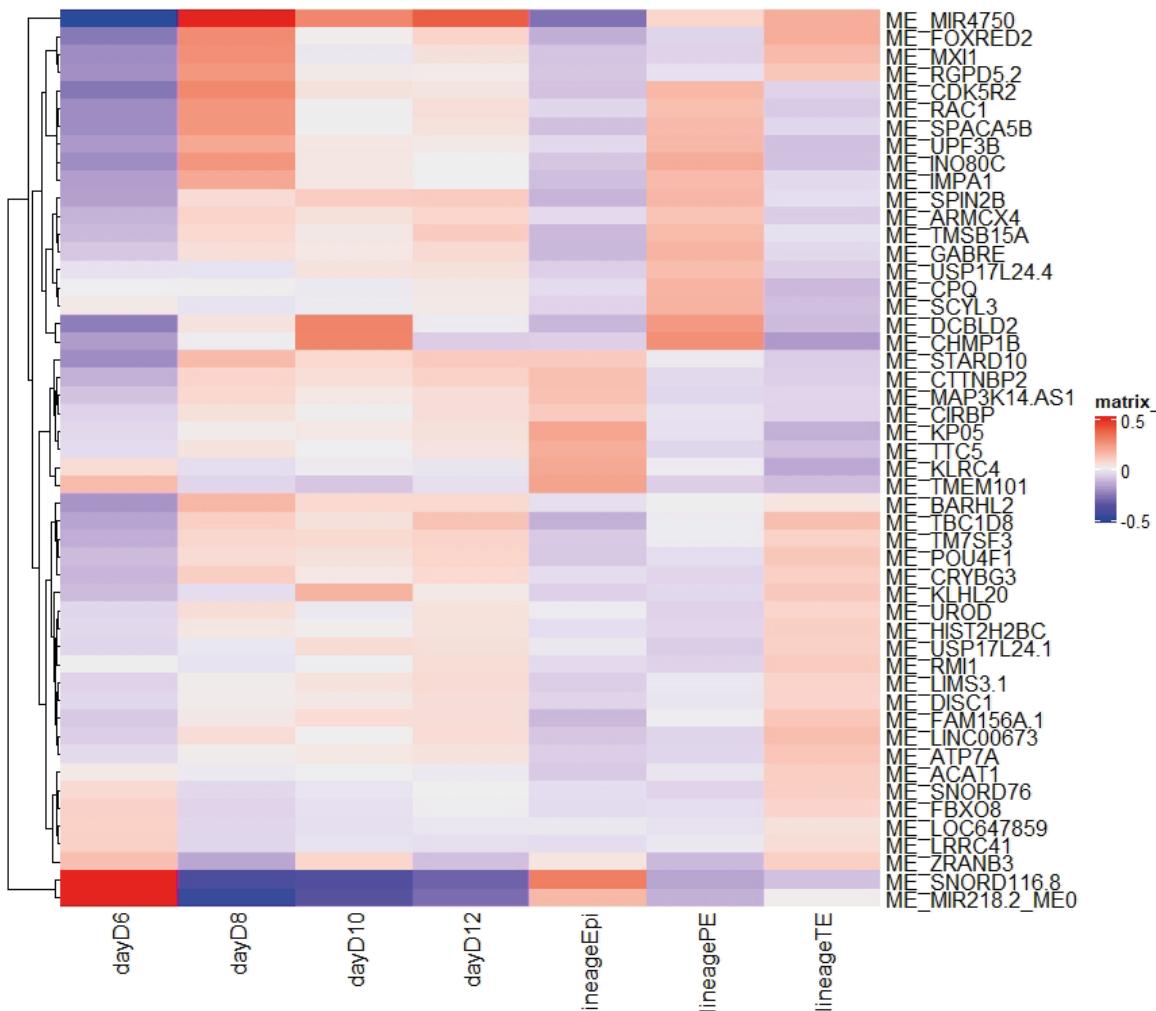
EXPRESSION



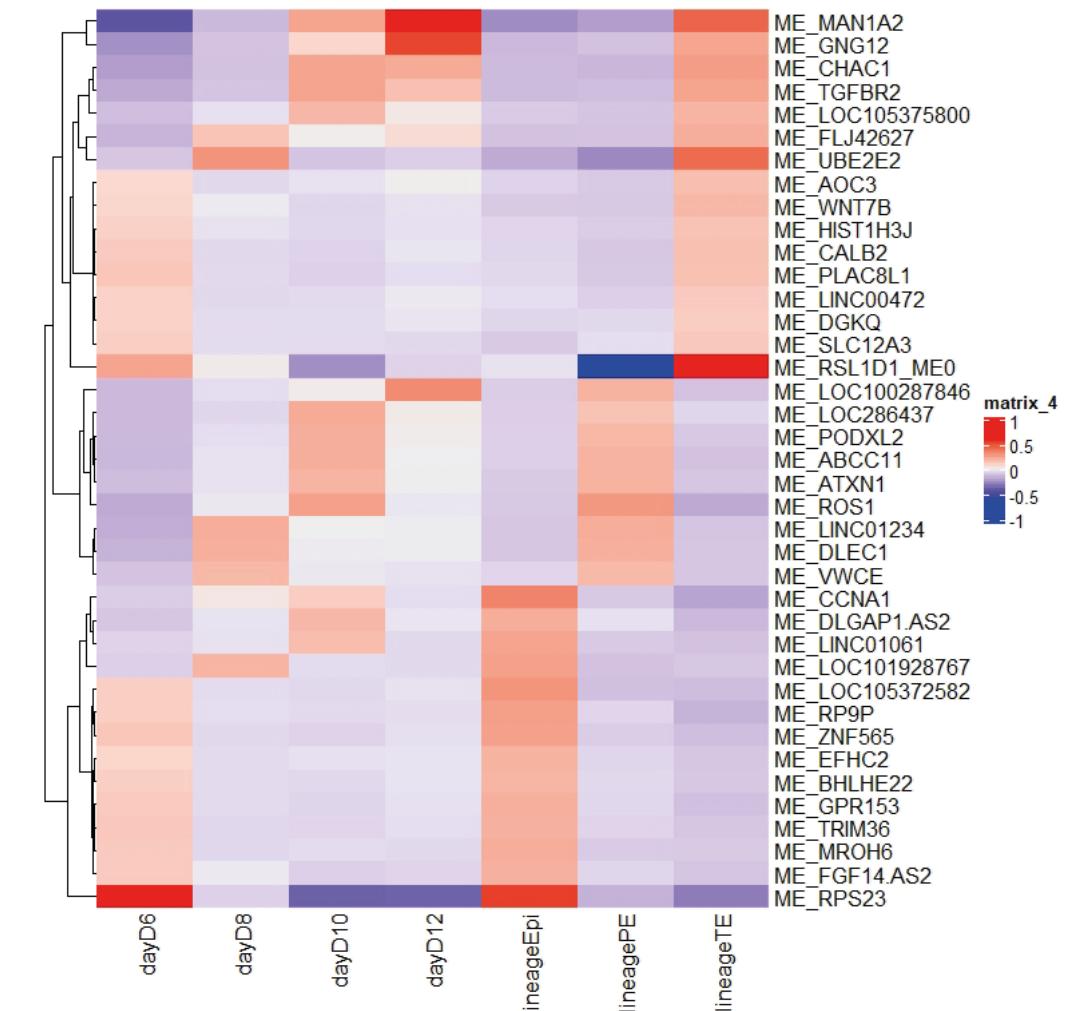
- We set a soft-power of 12 for both datasets
- This choice is made according to Langfelder and Horvath guidelines, as a minimum for signed networks
- Multi-threads calculation on IFB cluster nodes
- WGCNA on DNA methylation produces 49 modules
- WGCNA on gene expression produces 38 modules

MODULE-TRAIT CORRELATION

METHYLATION

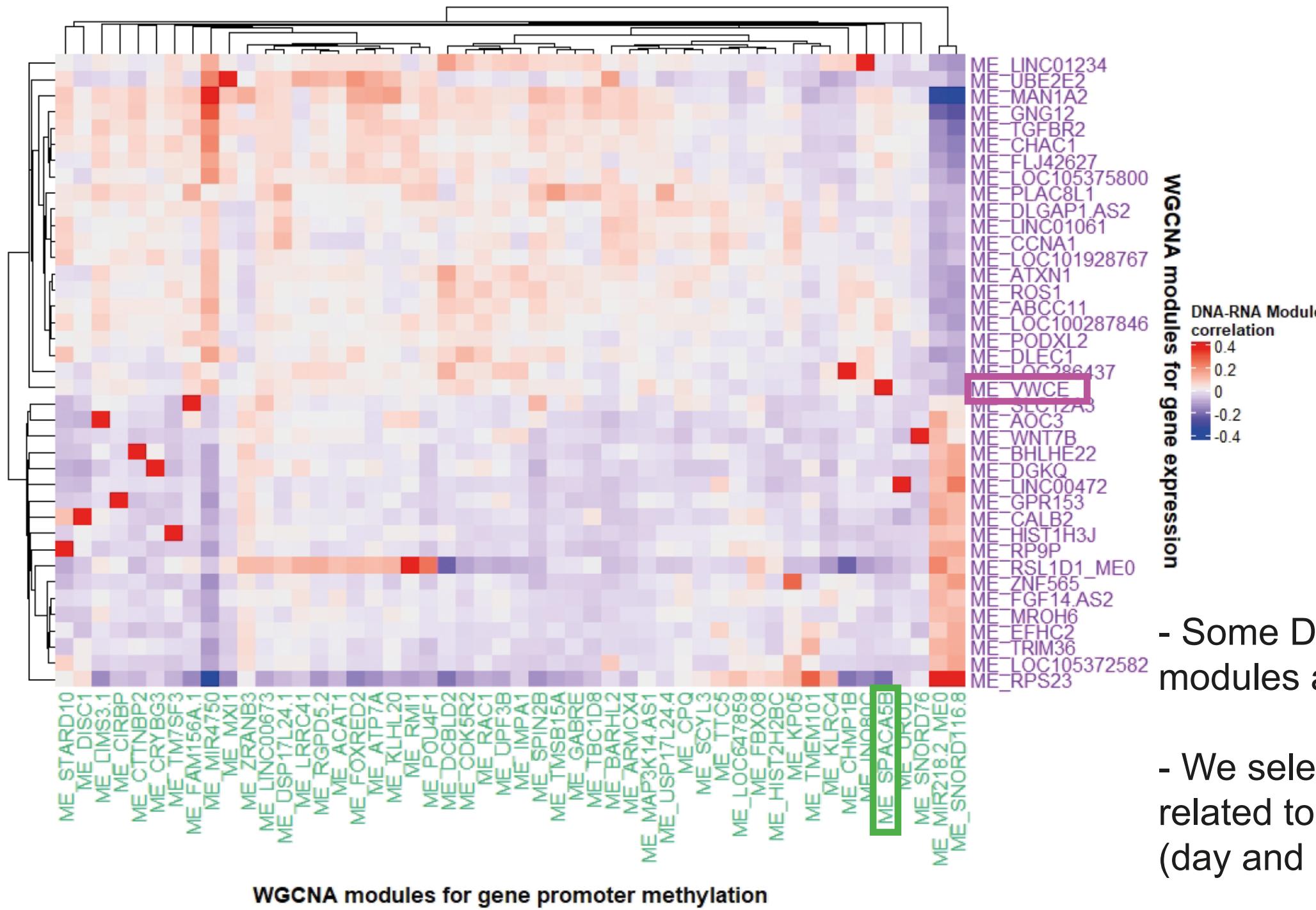


EXPRESSION



- DNA and RNA WGCNA modules recapitulate developmental stage and cell lineages
- Patterns are almost exclusive between groups

INTER-MODULE CORRELATION

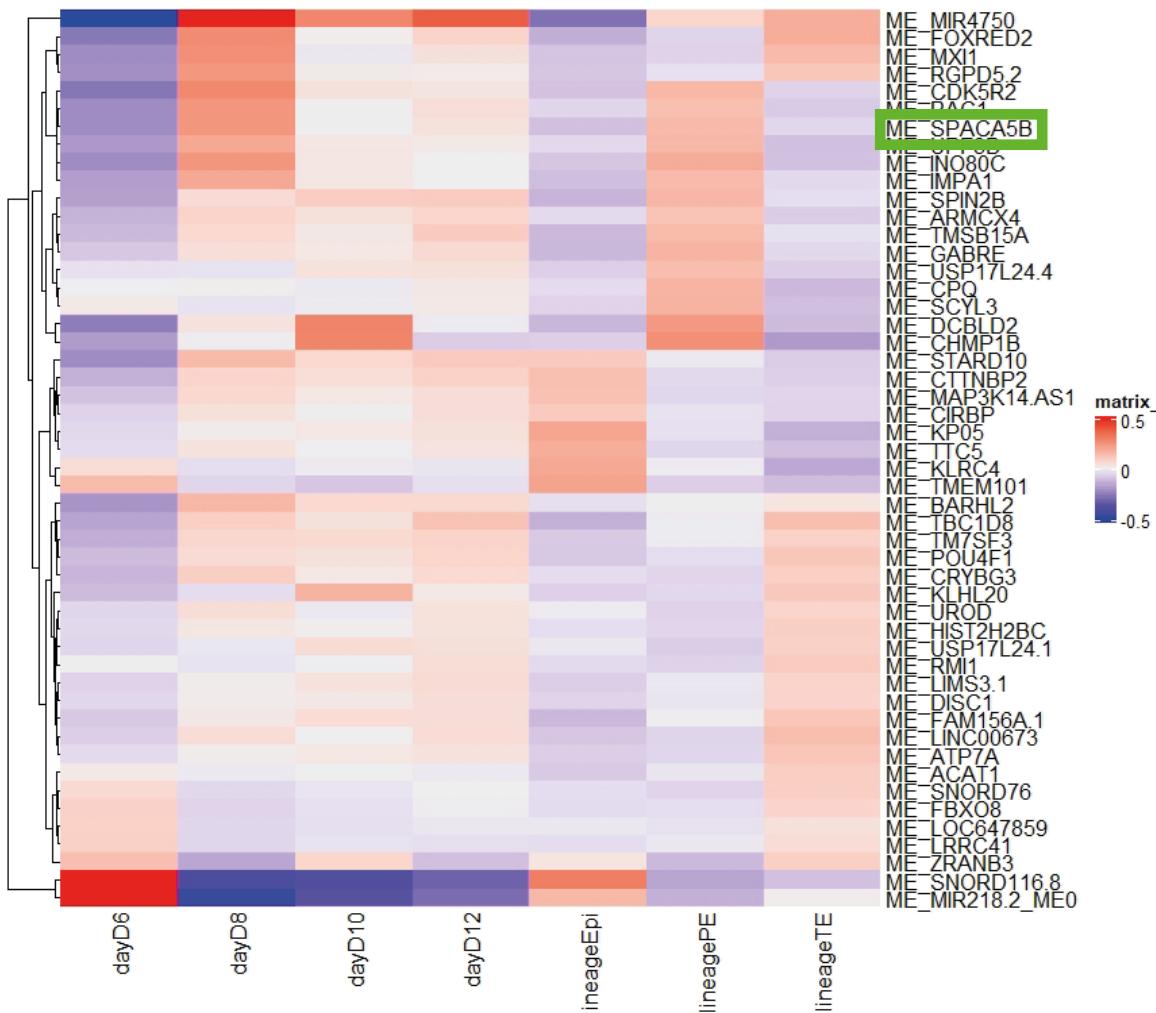


- Some DNA and RNA WGCNA modules are correlated
- We select those pairs related to traits of interest (day and lineage)

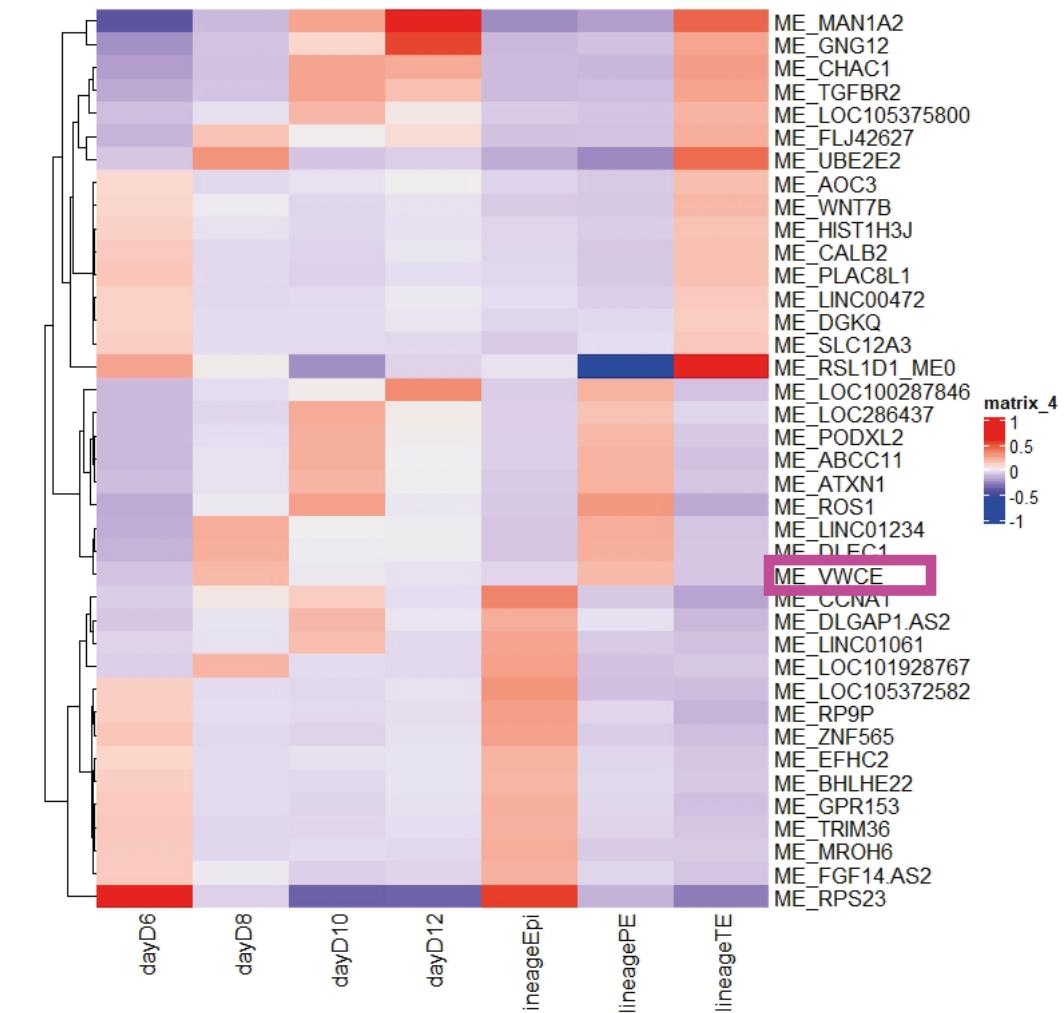
MODULE-TRAIT CORRELATION

10

METHYLATION



EXPRESSION

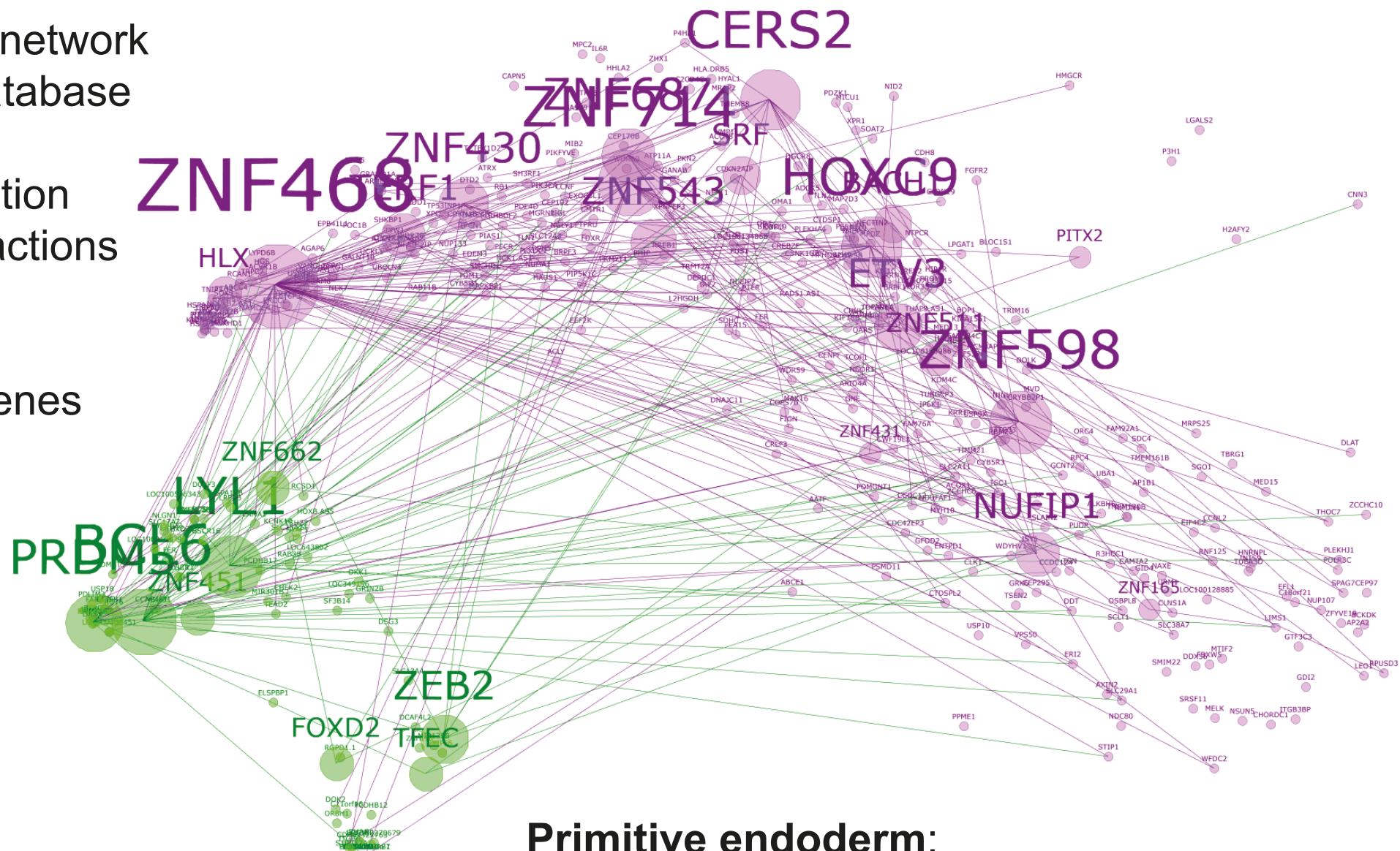


- DNA and RNA WGCNA modules recapitulate developmental stage and cell lineages
- Patterns are almost exclusive between groups

NETWORK: SPACA5B - VWCE (PE)

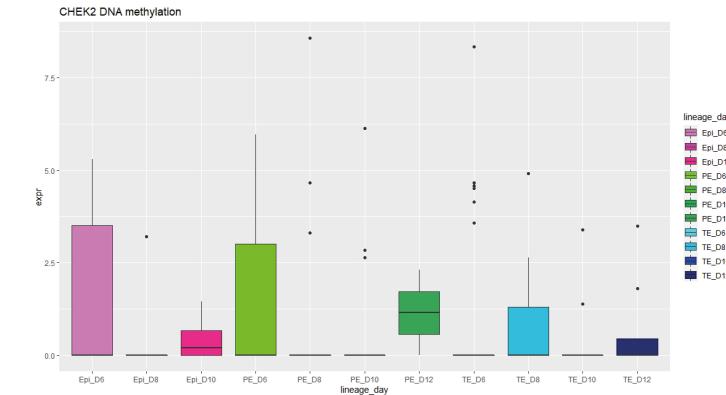
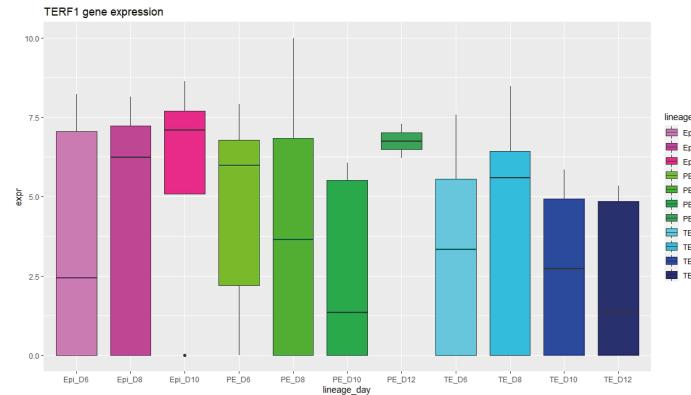
11

- We build the module pair network
 - We query the TF2DNA database
 - We show known transcription factor-gene promoter interactions between the two modules
 - This reveals connected genes

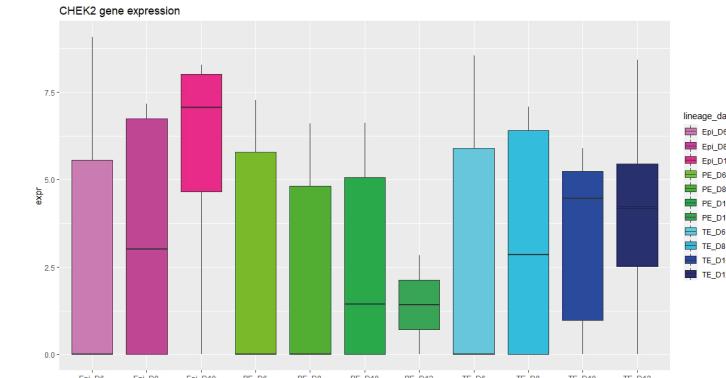


Primitive endoderm:

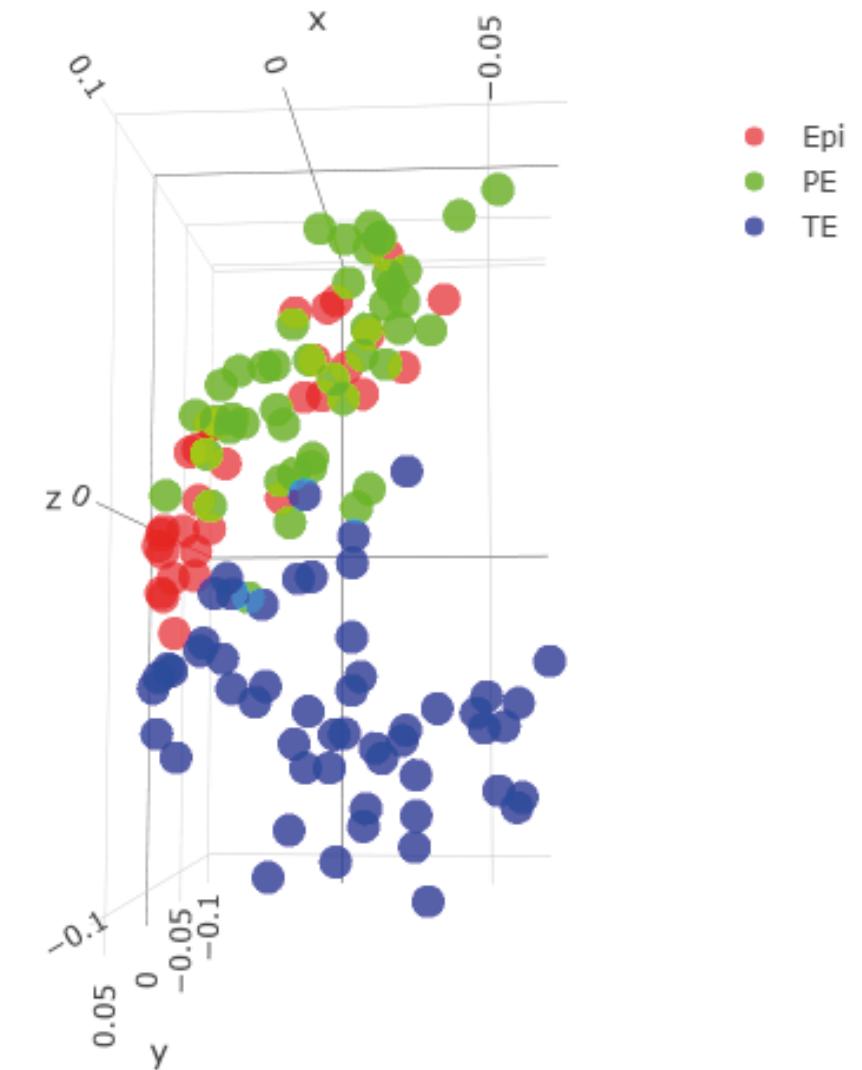
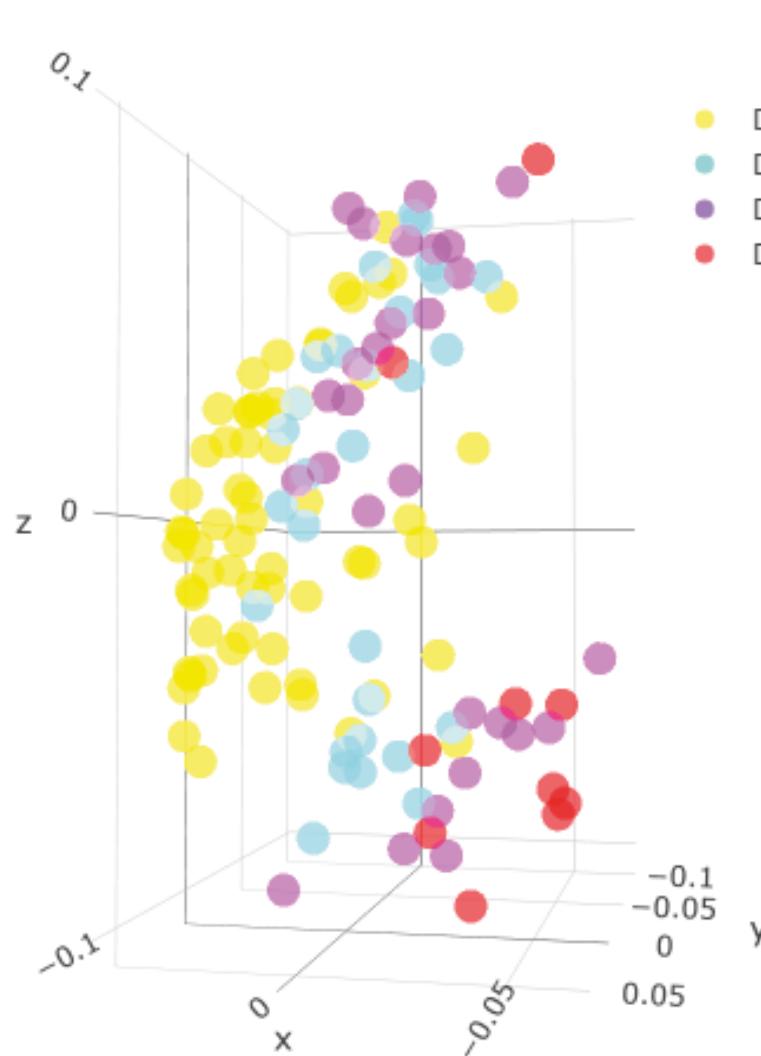
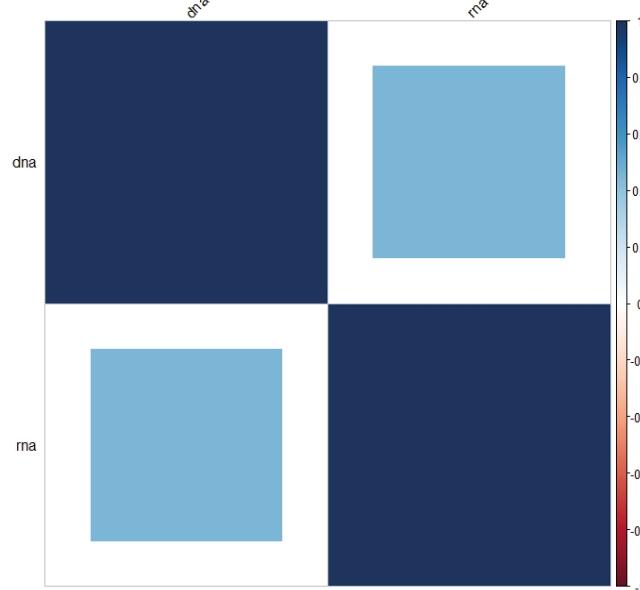
- DNA module: FOXD2, PRDM5, ZEB2...
 - RNA module: HOXC9, HLX, ETV3...



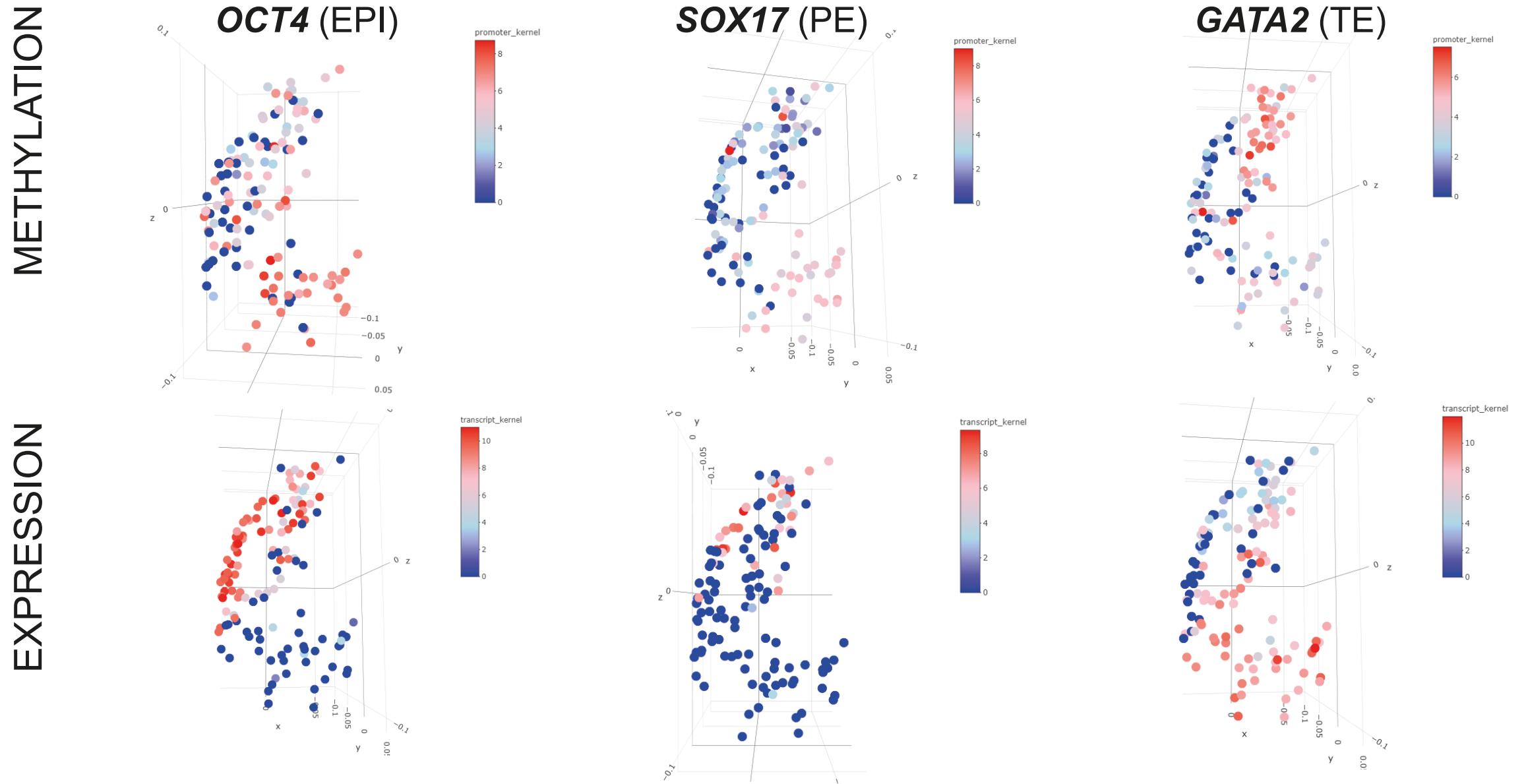
- TERF1, member of the PE RNA module, is highly expressed in PE at day 12
- TERF1 binds the CHEK2 gene on DNA
- CHEK2 gene is highly methylated and lowly expressed in PE at day 12



> **Hypothesis:** this might be an active molecular mechanism in the human embryo



- Kernel calculation combines the two data types (Mariette and Villa-Vialaneix, 2017)
- Embryo cells spread within the kernel space, according to developmental stage & cell lineages



- We validate kernel with marker gene methylation and expression patterns
- New analysis on the human embryo resulting from a combination of different omics data
- Interconnection of gene methylation and expression during early human development

Conclusion:

- We combined DNA methylation and transcriptome data
- Methylation is related to developmental stage
- Transcriptome is related to lineage formation
- Kernel recapitulates both stage and cell lineages (ICM vs TE)

Difficulties:

- No access to FASTQ files
- Hard time normalising the data
- Complicated choice for WGCNA soft-power thresholding

Perspective:

- Investigate gene module interactions
- Experimental validation
- Machine learning
- Predictive model: assign stage and lineage to embryo cells from multi-omics data

ACKNOWLEDGEMENT

- Audrey Bihouée
- Eric Charpentier
- Laurent David
- Thomas Fréour
- Alexandre Godmer
- Dimitri Meistermann
- Simon Chevolleau



REFERENCES

- Langfelder, P., Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008). <https://doi.org/10.1186/1471-2105-9-559>
- Mariette, Jerome & Villa-Vialaneix, Nathalie. (2017). Unsupervised multiple kernel learning for heterogeneous data integration. *Bioinformatics* (Oxford, England). 34. 10.1093/bioinformatics/btx682.
- Niakan, K. K., Han, J., Pedersen, R. A., Simon, C., & Pera, R. A. (2012). Human pre-implantation embryo development. *Development* (Cambridge, England), 139(5), 829–841. <https://doi.org/10.1242/dev.060426>
- Zhou, F., Wang, R., Yuan, P., Ren, Y., Mao, Y., Li, R., Lian, Y., Li, J., Wen, L., Yan, L., Qiao, J., & Tang, F. (2019). Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature*, 572(7771), 660–664. <https://doi.org/10.1038/s41586-019-1500-0>