INRAE

Région Ile de France

ifip institut du porc

Deciphering the molecular network controlling the biology of the pig blastocyst and its cellular interactions

Adrien Dufour[1], Sarah Djebali[2], Stephane Ferchaud[3], Yoann Bailly[3], Patrick Manceau[3], Frederic Martins[4], Bertrand Pain[5], Sylvain Foissac[6], Jérôme Artus[7], Hervé Acloque[1*]

1 Paris-Saclay University, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France ; 2 IRSD, Toulouse University, INSERM, INRAE, ENVT, Toulouse, France ; 3 GenESI, UE 1372 Génétique, Expérimentations et Systèmes Innovants, 86480 Rouillé, France ; 4 Toulouse Biotechnology Institute (TBI), Plateforme Genome et Transcriptome (GeT-Biopuces), Toulouse University, CNRS, INRAE, INSA, Toulouse, France ; 5 University of Lyon 1, INSERM, INRAE, Stem Cell and Brain Research Institute, U1208, USC1361, Bron, France ; 6 GenPhySE, Toulouse University, INRAE, ENVT, Castanet Tolosan, France ; 7 INSERM U1310, Paris Saclay University, Villejuif, France

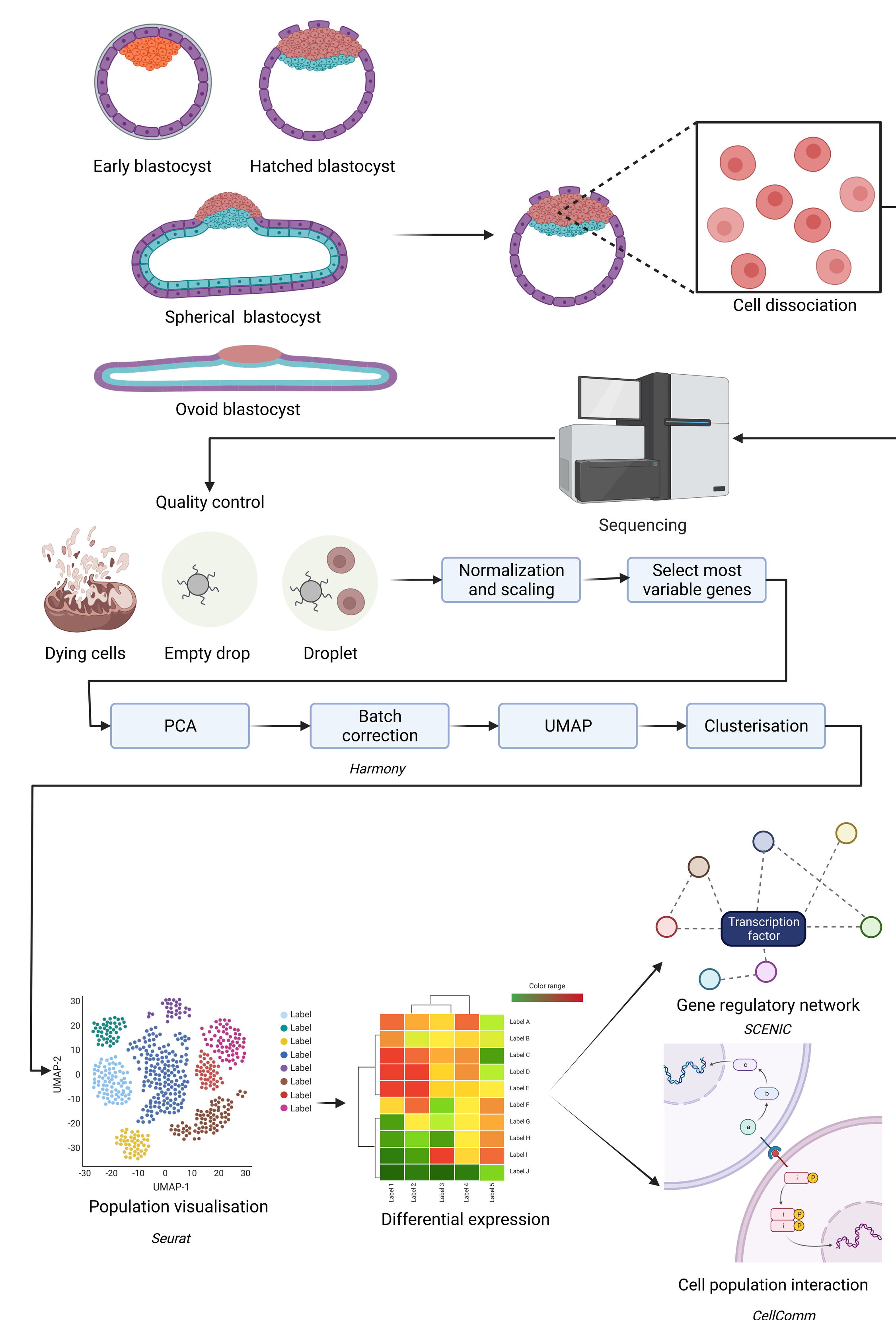
*herve.acloque@inrae.fr

1. INTRODUCTION

The embryonic development of the pig differs from that of humans and mice from the blastocyst stage and is characterized by much later implantation. Theses changes occurring before implantation could thus affect the biology of embryonic pluripotent cells, regulation network controlling pluripotency may not be conserved in mammal species.To better understand the biology of pig embryos before implantation and the molecular networks at play in the pig blastocyst, we produced a large dataset of single-cell RNAseq at different embryonic states (early, hatched, spherical and ovoid blastocysts). These data will help us to identify gene regulatory network and cellular interaction which can give us potential kety players of pluripotent stem cells and cytokine and inhibitor for stem cells cultural medium.

2. METHODS

scRNAseq datasets were cleaned, filtered and represent a total of 40,000 cells. Pre-processing steps including normalization, batch removal and principal components analysis werer performed for each states using Seurat [1] and Harmony [2]. Identification of the differents clusters was done from known markers from the litterature. On those lin eages, differential expression analysis was performed using Seurat ConservedMarker function. Regulon inference was performed using SCENIC [3] with a pig-based database : the transcription factor motif association was performed by orthology between human and pigs genes. Afterwards, a pig genome ranking was calculated by scoring putative regulatory regions for the presence of homotypic clusters of motif instances. Population communication have been inferred using the CellComm package [4], ligand-receptor have been predicted using the LIANA package [5] and the end of the pseudo-pathway is the regulons predicted previously with SCENIC.



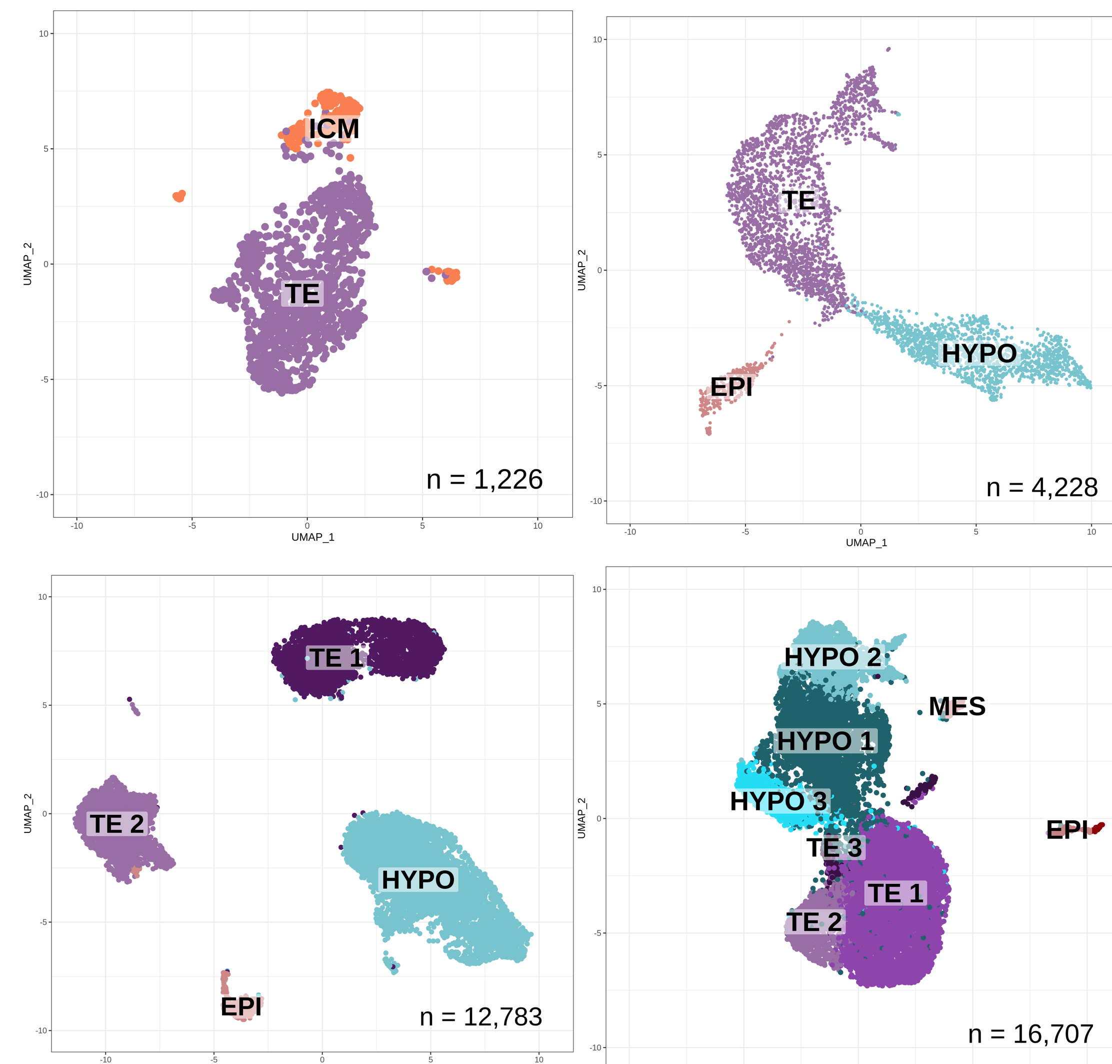
Schema of preprocessing of scRNAseq data

Single-cell multiomics datasets (coupled RNAseq & ATACseq) at different embryonic states (hatched, spherical and ovoid blastocysts) were also produced. Quality controls and pre-processing steps (peak calling, normalization, batch correction) were completed through ArchR package [6], then the cluster assignment was performed with an alignment [7] on our previous scRNAseq datasets.

3. Results

3.1 Identification of cell population

Visualization of gene expression and differential expression allowed to assign clusters to the three main embryonic population.

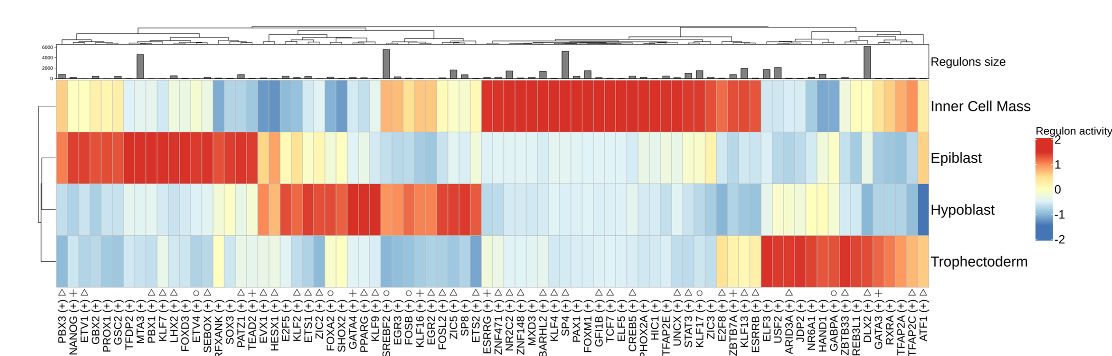


UMAP of cells at different states (E5, E7, E9, E11)

At the early blastocyst stage (E5) we identified a trophectoderm (TE, in violet) and an inner cell mass (ICM in orange) populations. At the hatched blastocyst stage (E7), we identified three main cell lineages : TE, hypoblast (HYPO, in cyan) and the epiblast (EPI, in red). At the spherical blastocyst stage (E9), we identified the 3 main cell lineages and a sub populations of TE. At the ovoid blastocyst stage (E11), we identified multiple subpopulations within TE and HYPO. We also noticed the emergence of a mesendoderm population indicating the onset of gastrulation.

3.2 Gene regulatory network

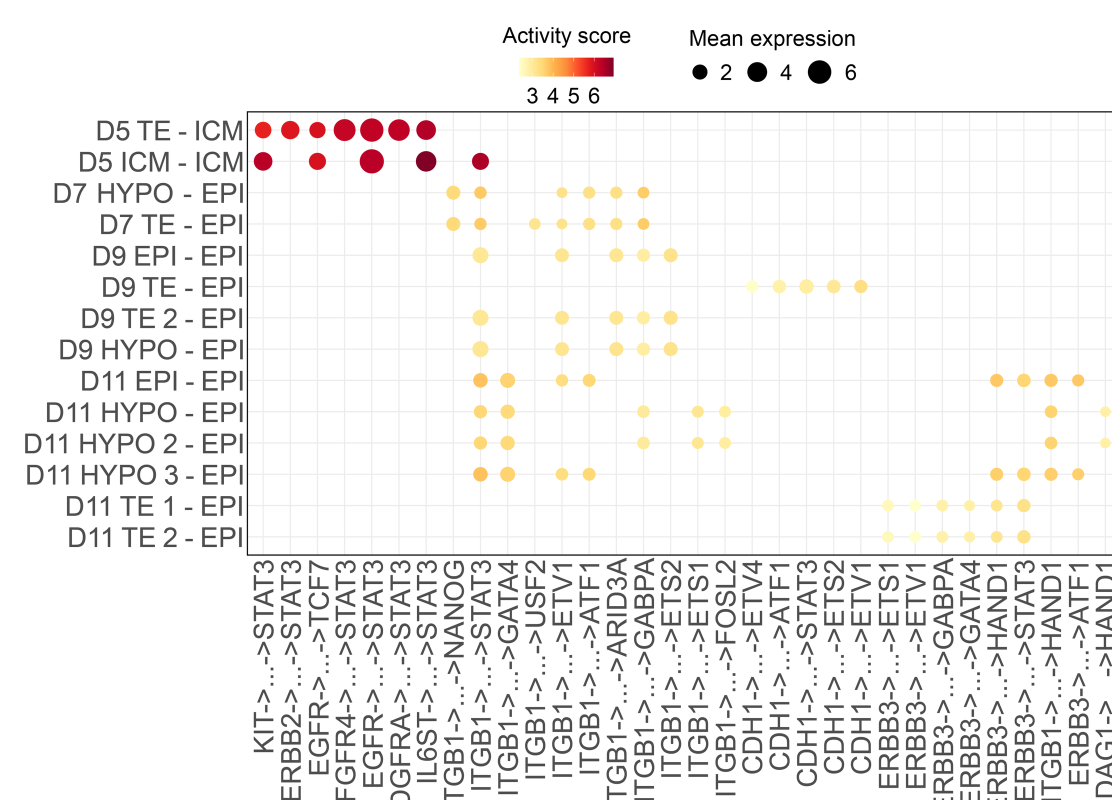
Identification of modules of regulation (regulons).



Heatmap of the regulons activity by cell population

We discovered nearly 300 regulons in the analysis of our cells. We kept the twenty regulons for each population and we compared our result to others studies on pigs [8] and human [9]. We found transcription factors who have been described in litterature (*NANOG*, *GATA4*, *GATA3*). We also identified new candidates like *ETV4* (EPI), *KLF17* (ICM), *FOSB* (HYPO), *GAPBA* (TE).

Identification of interaction between cell populations from a ligand to the final transcription factor



Dotplot of CellComm interaction and pathways

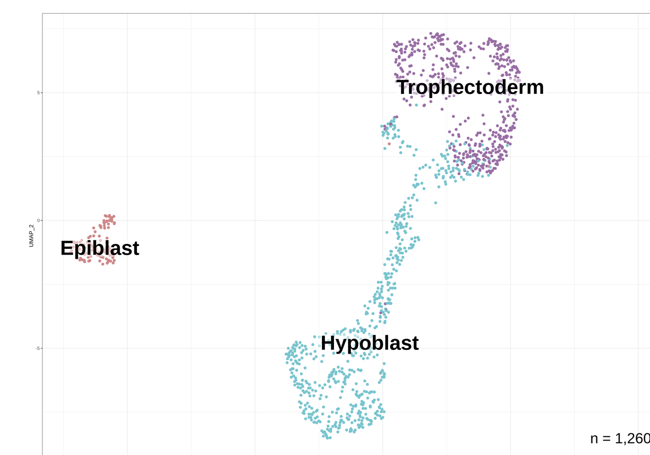
The dot plot shows the active pathways induced either by the EPI (autocrine signal) or adjacent cells (TE or HYPO) to the activation of specific transcription factors in the EPI.

We notice a strong signal in the early blastocyst stage with the activation of *STAT3* regulon in the ICM which may be activated by the Jak-Stat pathway and *NANOG* in the EPI.

We noticed that all tissues do not send signals, for example at E7 EPI does not have autocrine signals.

3.3 scMultiomics RNAseq & ATACseq

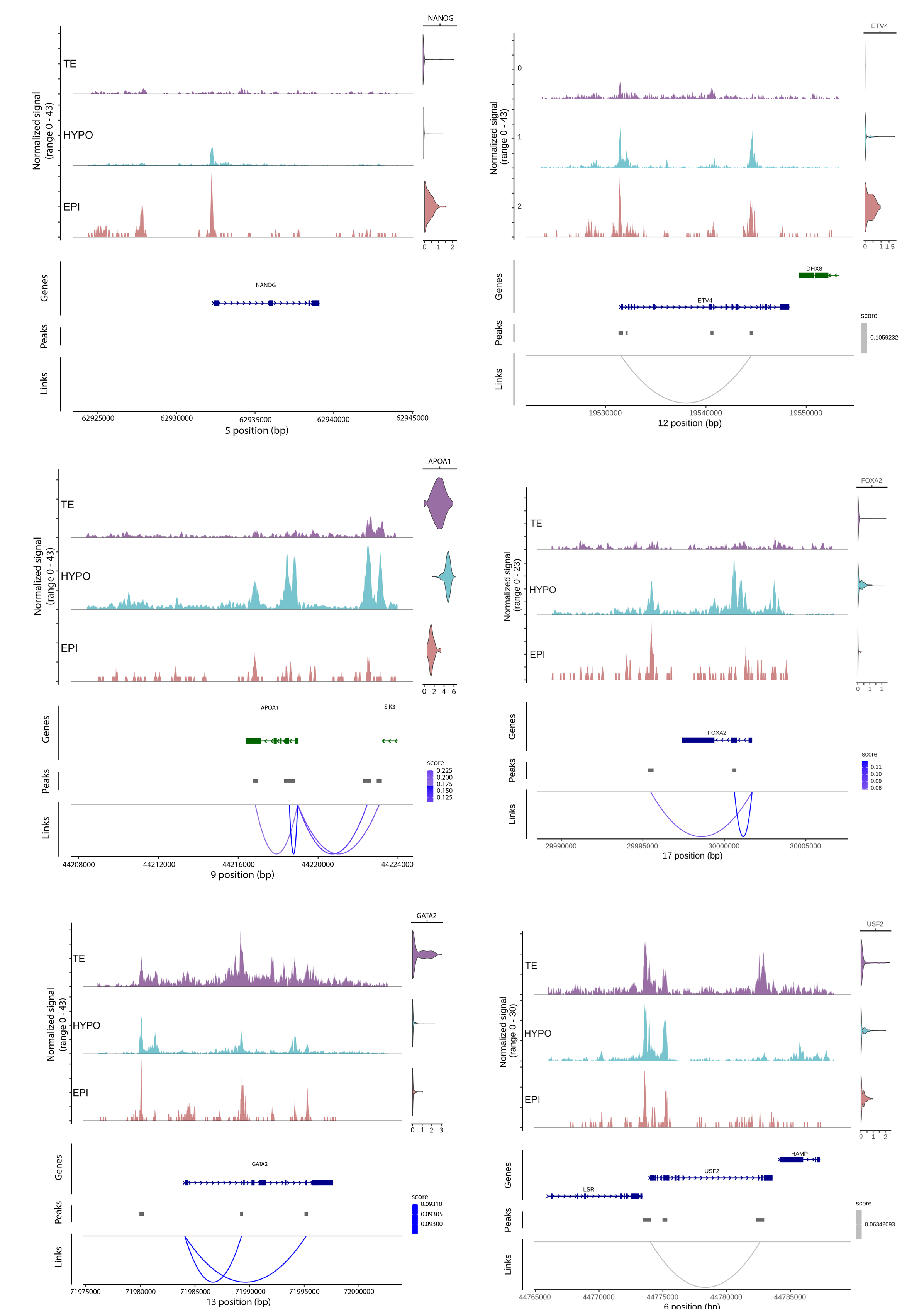
Assignation of cell populations to the multiomics datasets



UMAP of scMultiomics cells at states E7

The mapping of our cells on the scRNAseq datasets allowed us to retrieve the three main lineage.

Visualisation of gene expression and chromatin accessibility of four genes markers for the EPI, TE, HYPO



Profil of known and new population marker genes (NANOG, ETV4, USF2, GATA2, APOA1, FOXA2) at E7

We confirm the population specificity of important marker of the three main lineage, we can see that the RNAseq expression (violinplot) and the ATACseq signal are correlated with the assigned population.

4. CONCLUSION

Our analyses have allowed us identify the three lineage at all states with the expression of known embryo markers like *POU5F1*, *GATA3*, *COL18A1*. Gene regulatory network inference allowed to find known regulon like *NANOG*, *ZFP42*, *KLF7* and new regulons which regulate pluripotent genes like *ETV4*. We then achieved to rely those regulators to potential activators or inhibitors from others populations. We retrieve the significance of the Jak-*STAT3* pathway at the early blastocyst stage and that of *ITGB1* and *CDH1* for the biology of the epiblast. Future work will highlight the link between active regulation networks and chromatin accessibility to drive blastocyst development.

REFERENCES

1. Butler, et al. 2018, Nat Biotechnol 36:411-420
2. Korsunsky, et al. 2019, Nat Methods 16:1289-1296
3. Aibar, et al. 2017, Nat Methods 14:1083-1086
4. Da Rocha, et al. 2022, Nat Cell Biol 24:579-589
5. Dimitrov, et al. 2021, Preprint
6. Granja, et al. 2021, Nat Genet 53:403-411
7. Stuart, et al. 2021, Nat Methods 18:1333-1341
8. Zhi, et al. 2022, Cell Res 32:383-400
9. Meistermann, et al. 2021, Cell Stem Cell 28:1625-1640.e6

Created with BioRender Poster Builder