

Rotation Report

Liu Keqin

2020/11/01

Abstract

The mammalian embryonic implantation is a critical and unique stage in embryonic development. Although it is now better understood how differential expression of genes regulates embryonic cells to undergo cell line separation and progressive loss of pluripotency in pre-implantation. However, because of the difficulty of obtaining human embryos after implantation *in vivo*, it still remains unclear how the gene expression network regulate the loss of pluripotency and progressive determination of cell fate during embryo implantation. In this study, we briefly review the mechanisms of cell lineage segregation events and loss of pluripotency during early embryonic development in humans and mouse, and preliminarily explore the association between gene expression profiles and cell fate regulation during human embryonic implantation.

Research goals and contents

We wanted to analyze the differential gene expression and temporal changes during the human embryonic implantation period using bioinformatics tools to initially investigate the specific mechanisms of loss of pluripotency and cell fate commitment in human embryonic cells. In this study, we acquired publicly available single-cell transcriptome sequencing data of human embryonic implantation (gene expression matrix in TPM format) from the GEO database, and then performed cell clustering, GO enrichment analysis and KEGG pathway analysis on the acquired single-cell data by Seurat(Maaten and Hinton, 2008), Gene Ontology(Ashburner et al., 2000; Consortium, 2019) and KEGG databases(Kanehisa, 2019; Kanehisa and Goto, 2000) in R language respectively.

Research progress and conclusions

1. Single-cell RNA-seq clustering analysis

We processed the original TPM matrix (column names as sample names, row names as gene symbols) which is obtained from the GEO database by adding 1 and log10 and then used Seurat(Maaten and Hinton, 2008) in R to select its top 2000 differentially expressed genes to run t-SNE to obtain the cellular clustering result. It can be observed from Fig1. that all cell samples were divided into a total of 15 clusters.

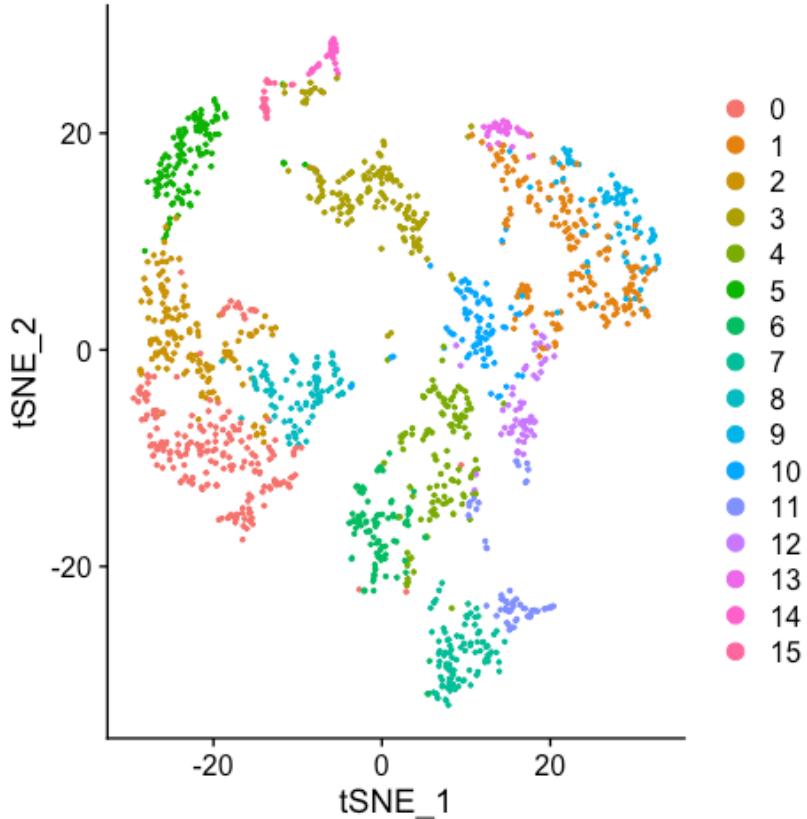
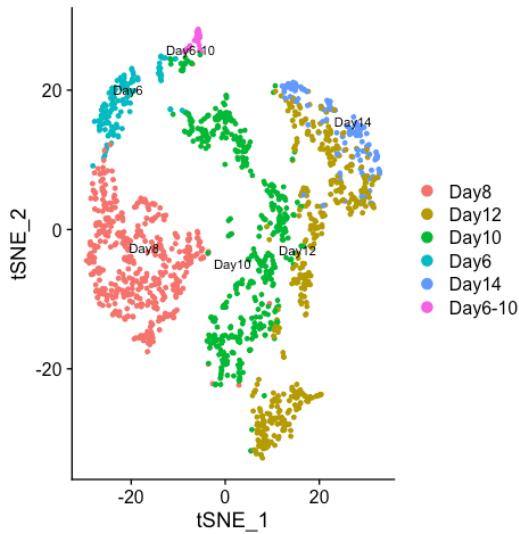
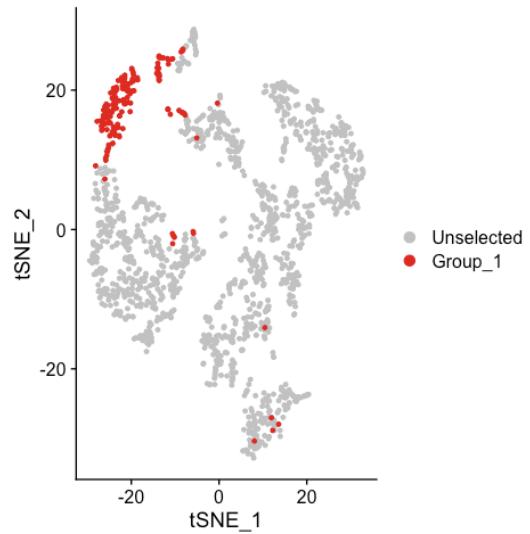


Fig. 1. Cellular clustering results

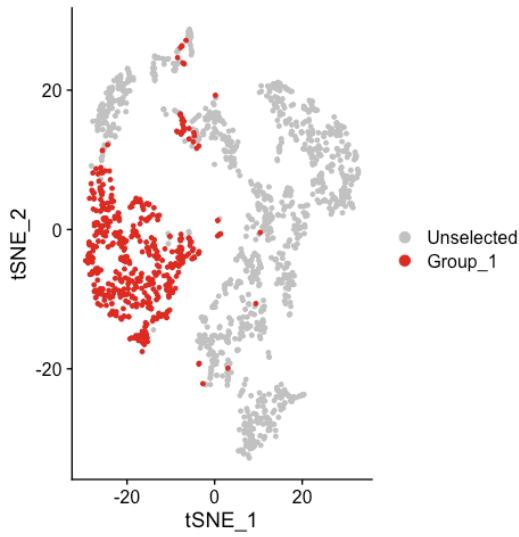
We then annotated the cell clustering result with the number of days of embryonic development according to the stage of embryonic development in which the sample cells were located Fig2. According to the description of the GEO database, all samples were taken from five different embryonic developmental stages: the sixth day, the eighth day, the tenth day, the twelfth day and the fourteenth day(Zhou et al., 2019). Among them, samples taken from the sixth day were pre-implantation embryos and samples taken from the eighth and subsequent days are post-implantation embryos. In the annotated results of cell clustering Fig2.(i), we obtained six clusters of different stages. With the exception of the samples from Day 6 to Day 14, we observed that a few samples from Day 6 to Day 10 overlapped on cluster 14, so cluster 14 is additionally annotated as Day 6-10.



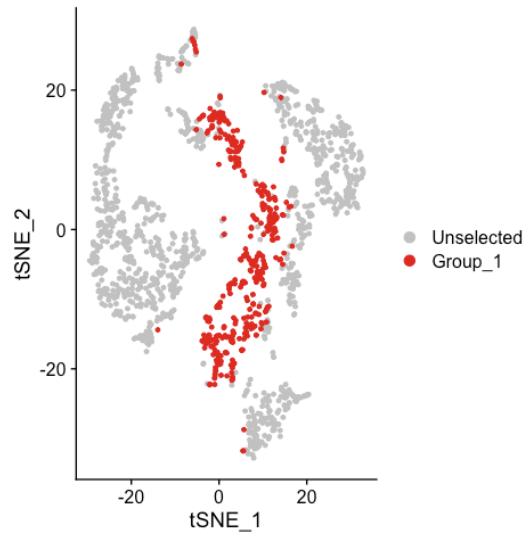
(i)



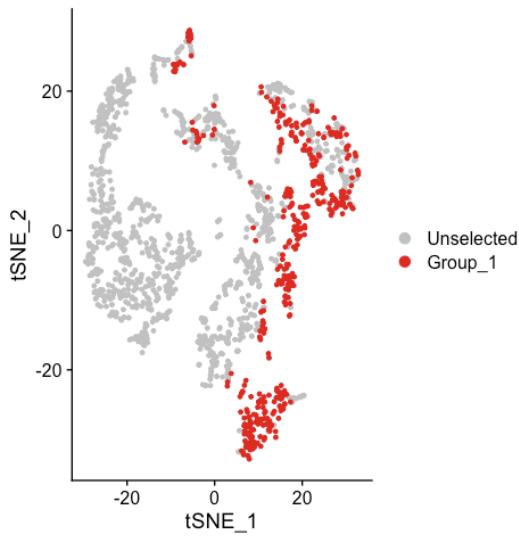
(ii)



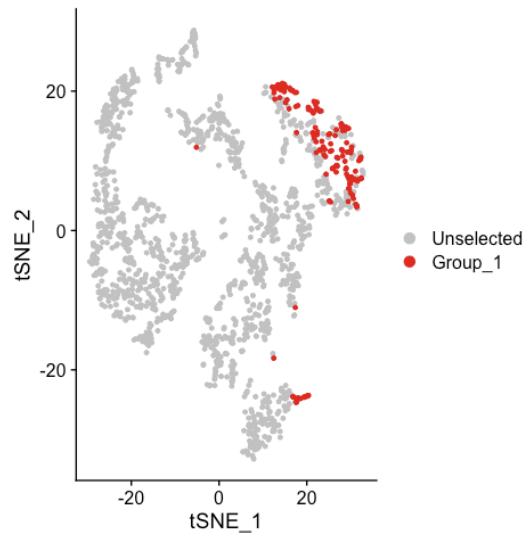
(iii)



(iv)



(v)



(vi)

Fig. 2. Cellular clustering results by embryonic stages

Figure.2 (i)shows all embryonic stages in cellular clustering result. (ii)Day 6: 167 samples; (iii)Day 8: 464 samples; (iv)Day 10: 371 samples; (v)Day 12: 407 samples; (vi)Day 14: 153 samples.

According to the original publication(Zhou et al., 2019) and CellMarker database(Zhang et al., 2019), We obtained cell markers for EPI, PE and TE cells. EPI, PE and TE cells marked by these markers are shown in Fig3.(ii), (iii) and (iv). According to the Fig3.(i), the number of TE cells was the largest, with fewer EPI and PE cells.

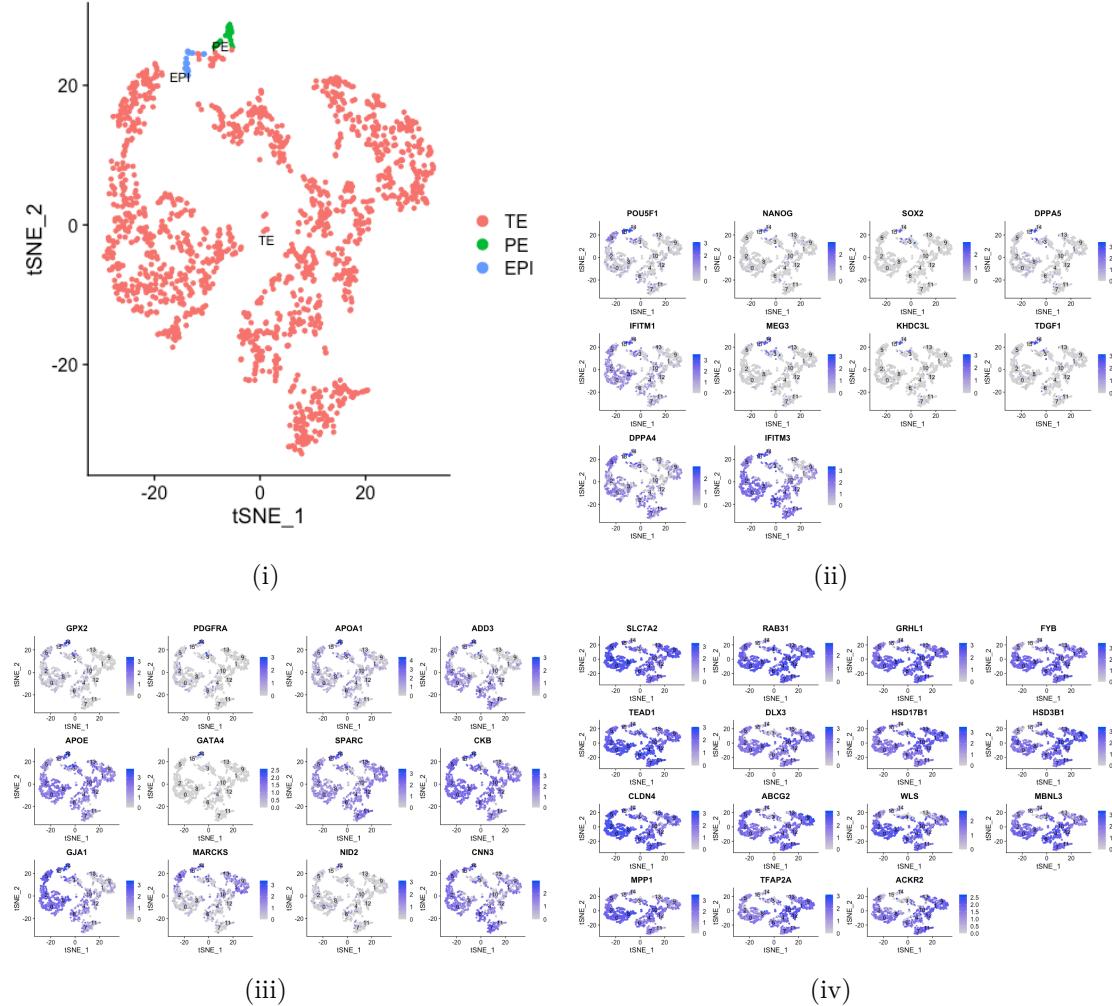


Fig. 3. Cellular clustering results by embryonic cell types

Figure.3 (i)shows embryonic cell type (EPI, PE and TE) in cellular clustering result. (ii)Cells marked by POU5F1, NANOG, SOX2, DPPA5, IFITM1, MEG3, KHDC3L, TDGF1, DPPA4 and IFITM3. (iii)Cells marked by GPX2, PDGFRA, APOA1, ADD3, APOE, GATA4, SPARC, CKB, GJA1, MARCKS, NID2 and CNN3. (iv)Cells marked by SLC7A2, RAB31, GRHL1, FYB, TEAD1, DLX3, HSD17B1, HSD3B1, CLDN4, ABCG2, WLS, MBNL3, MPP1, TFAP2A and ACKR2.

Then we used cell markers mentioned in Fig.3 to label EPI, PE and TE cells in all 5 embryonic stage. The result shows in Fig.4. We noticed that the expression of pluripotency genes like NANOG and Sox2 seems very low accross all samples Fig.4. NANOG is known as a transcription factor that plays an essential role in maintaining the pluripotency and self-renewal capacity of ES cells(Zhang et al., 2010). During mouse embryo development, expression of NANOG is gradually down-regulated as the embryo implants. NANOG mRNA is first detected in the interior cells of the compacted morulae and disappears in

the trophectoderm in the blastocyst stage. After implantation, expression of NANOG is down-regulated, but still can be detected in germ cells of the genital ridges of E11.5 mouse embryos(Chambers et al., 2003). In addition, downregulation of NANOG can induce both mouse and human ES cell differentiation to extra-embryonic lineages(Hough et al., 2006). Besides, Sox2 is also an important transcriptional regulator in pluripotent stem cells (PSCs). Together with NANOG, Sox2 controls gene expression in PSCs and maintain their pluripotency. Sox2 expression is initially detected in cells at the morula stage, becoming more specifically located in the ICM of blastocyst and epiblast(Avilion et al., 2003) during the latter stages, which implies that Sox2 may have important roles in the formation of early pluripotent embryonic cells(Zhang and Cui, 2014). Thus, the low expression of both NANOG and Sox2 across all samples may shows that the pluripotency of the embryo has been reduced at the implantation stage.

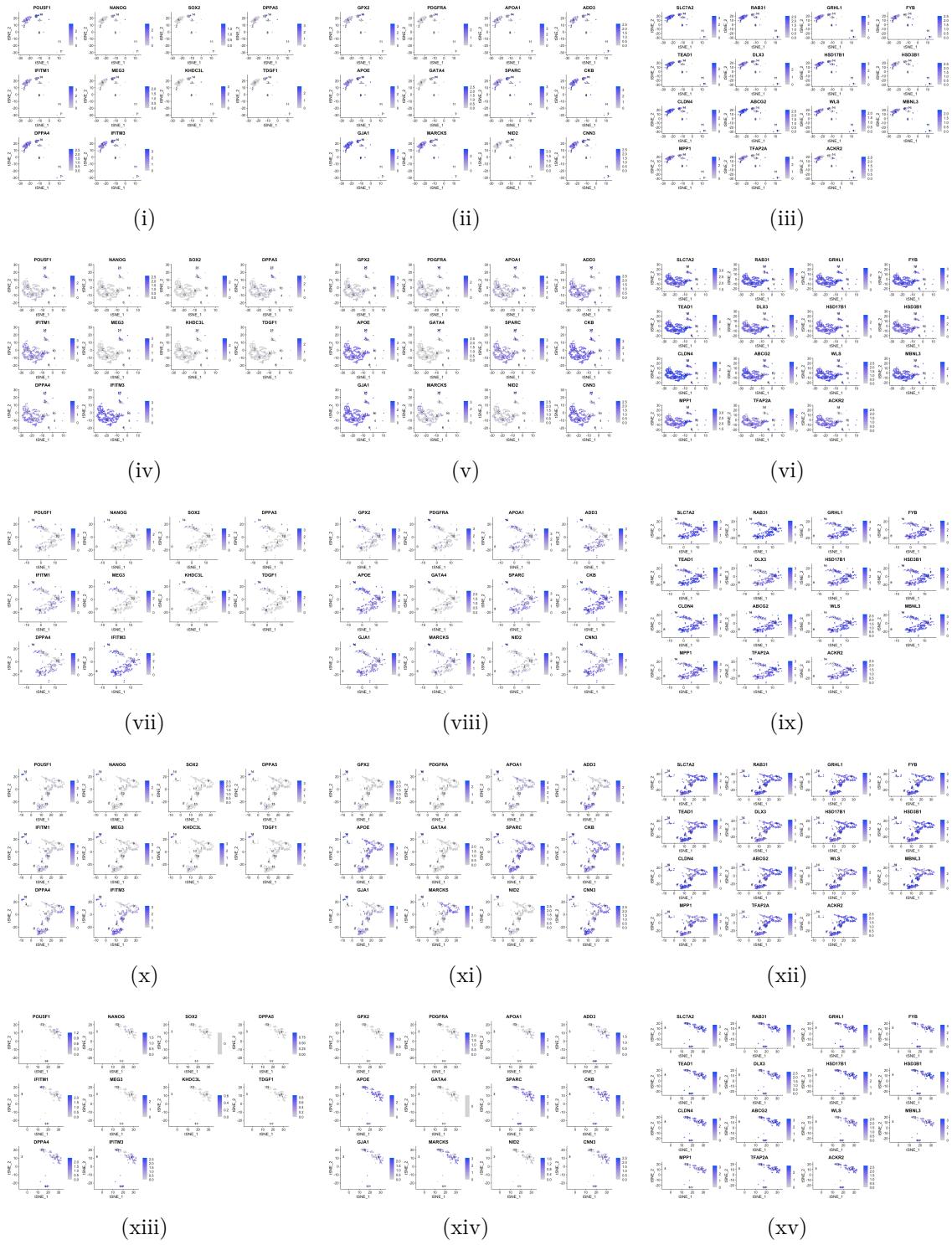


Fig. 4. Cells marked by EPI, PE and TE cell markers in all embryonic stages

Figure.4 (i)(ii)(iii)Cells on the 6th day marked by EPI, PE and TE cell markers. (iv)(v)(vi)Cells on the 8th day marked by EPI, PE and TE cell markers. (vii)(viii)(ix)Cells on the 10th day marked by EPI, PE and TE cell markers. (x)(xi)(xii)Cells on the 12th day marked by EPI, PE and TE cell markers. (xiii)(xiv)(xv)Cells on the 14th day marked by EPI, PE and TE cell markers.

And we noticed the especial expression of GATA4 across samples in Fig4.(ii)(v)(viii)(xi)(xiv), which seems that GATA4 is expressed at very low levels in imbreros at Day6 to Day12 stage and barely expressed at Day14 stage. Therefore we decided to check the expression of the family of GATA transcription factor in all samples. The result shows in Fig.5. We

found that GATA2 and GATA3 are highly expressed in almost all stages of embryonic cells, but low in cluster 14, which happens to be PE cells according to Fig3.(i). Meanwhile, GATA4 is only highly expressed in cluster 14, GATA6 is generally expressed at low levels, and GATA5 is only expressed in few embryonic cells. It's been proven by previous studies that GATA proteins are involved in the regulation of key genes expressed by the trophectoderm that underpin the transition from the morula to trophoblast, and trophectoderm maintenance(Bai et al., 2013). In mice, it's known that the mRNA and proteins of all six GATA factors were detected during the embryonic development process. GATA2 and GATA3 regulate development and differentiation of hematopoietic lineages(Fujiwara et al., 1996; Pandolfi et al., 1995; Tsai et al., 1994), while GATA4, GATA5 and GATA6 are involved in cardiac development and endodermal derivatives(Molkentin et al., 1997, 2000; Morrissey et al., 1998). Based on Fig.5, we might be able to infer that at the stage of embryonic implantation, the process of development and differentiation of hematopoietic lineages is active, while the building process of cardiac development and endodermal derivatives is less active.

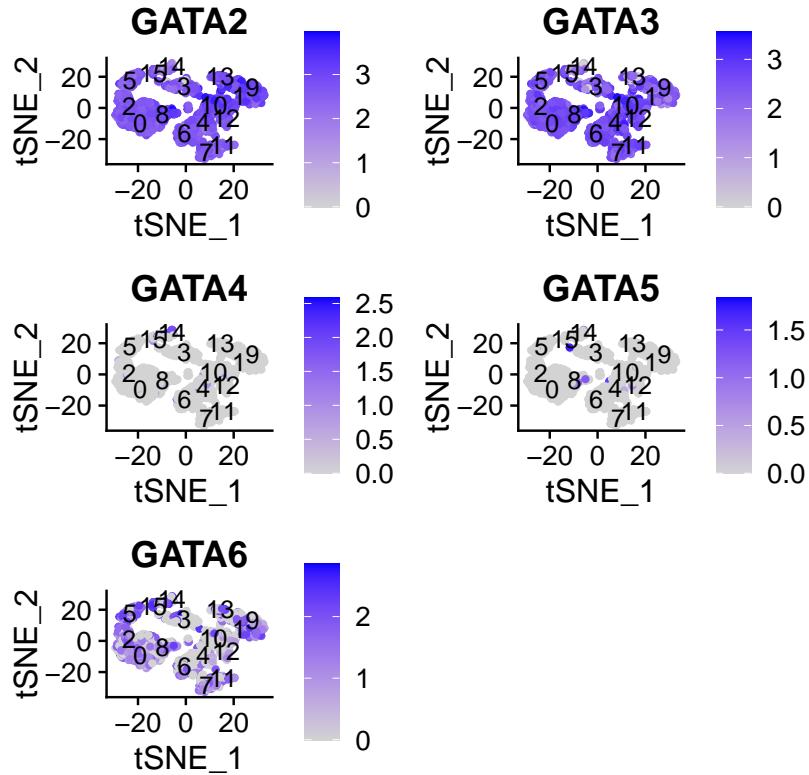


Fig. 5. Expression of GATA transcription factors across all samples

2. Analysis of cell marker and DE genes in EPI, PE and TE

With Fig.6(i)(iii)(v), we can clearly see that the EPI cells labeled by POU5F1, MEG3 and TDGF1 are mainly concentrated in cluster15, while the PE cells labeled by GPX2, PDGFRA, APOA1, ADD3, APOE, GATA4, SPARC, CKB, GJA1, MARCKS, NID2 and CNN3 are mainly concentrated in cluster14, which corroborates the cell clustering results that we observed in Fig.3. Notably, the expression level of POU5F1 has a peak on cluster15 which shows in Fig.6(ii). In previous studies, several key regulators like Oct4(also

known as Oct3 and encoded by POU5F1), Sox2, and Nanog have been identified that are essential both for the formation of the ICM during mouse preimplantation development and for self-renewal of pluripotent ESCs(Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003). During mouse preimplantation development, zygotic Pou5f1 expression is activated at the four-cell stage and is later restricted to the pluripotent cells of the ICM and epiblast. In the mouse postimplantation embryo, Pou5f1 expression is down-regulated upon EPI differentiation and its expression is maintained only in the primordial germ cells. In addition, Oct4 is highly expressed in human and mouse ESCs, and its expression diminishes when these cells differentiate and lose pluripotency(Palmieri et al., 1994). Thus, we may be able to infer that the cluster15 or EPI with high expression of Oct4 appear to retain some pluripotency throughout the deposition process. However, the expression levels of NANOG, Sox2, and KHDC3L in EPI and other cells approached almost zero, which, considering that NANOG and Sox2 are also representative cellular pluripotency factors, may suggests that the pluripotency of EPI cells during the implantation is already low.

Additionally, it is noteworthy that two higher peaks of MEG3 expression levels occurred in cluster14 and cluster15, i.e. PE and EPI cells, respectively which also shows in Fig.6(ii). MEG3 is imprinted in paternally allele and thus expressed predominantly from the maternal allele. It generates alternatively-spliced long non protein-coding RNAs that overlap with the microRNAs Mir770 and Mir1906-1 and transcripts from this gene function is negative regulators of growth(Bult et al., 2019). Hypomethylation of MEG3 allows its production from the maternal allele. MEG3 increases EMT and migration of trophoblast cells and decreases apoptosis via down regulation of NFkB, caspase 3 and Bax(Basak and Ain, 2019). Besides, many previous studies have reported that aberrant imprinting is related to the efficiency of animal cloning(Chung et al., 2003; Dean et al., 2001) and the plasticity of pluripotent stem cells(Stadtfeld et al., 2010). In the mouse, MEG3 in the imprinted Dlk1–Dio3 gene cluster was aberrantly silenced in the majority of iPSCs derived from different cell types at various stages of differentiation. These aberrant expressions hamper the developmental potential of iPSCs. Aberrant silencing of maternally imprinted MEG3 has been shown in the early stage of reprogramming; however, this status was not recovered in the majority of piPSCs(Cheng et al., 2012; Liu et al., 2014). Therefore, the expression status of MEG3 may be closely related to the plasticity potential. And in this case, we may be able to deduce that EPI and PE cells remain partially pluripotent during embryonic implantation.

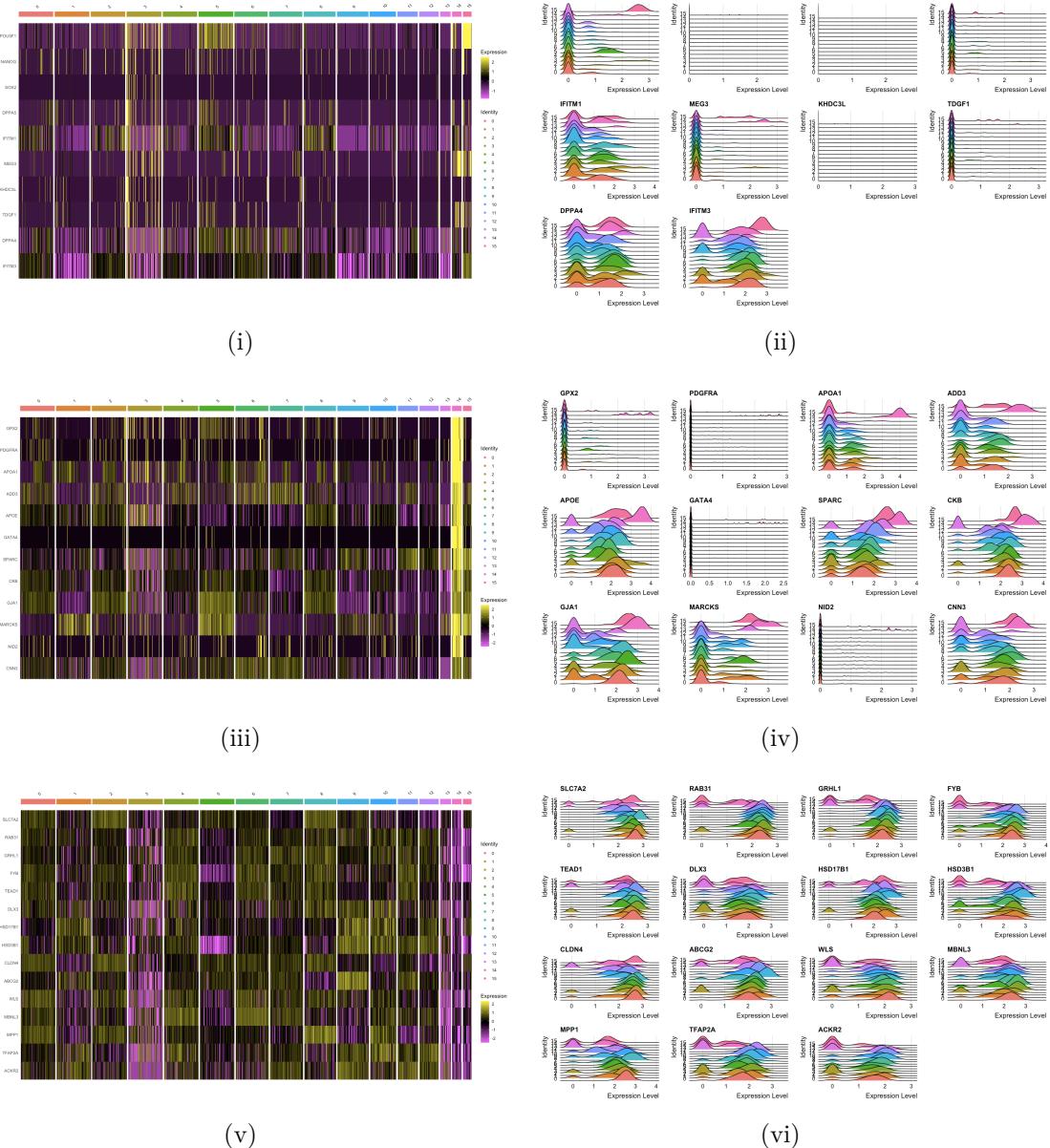


Fig. 6. Heatmaps and Ridgeplots of EPI, PE and TE cells

Figure 6 (i) A heatmap and (ii) a ridge-plot of EPI of cell markers. (iii) A heatmap and (iv) a ridge-plot of PE cell markers. (v) A heatmap and (vi) a ridge-plot of TE cell markers.

We then looked at the top 100 genes that were differentially expressed across all samples and the top 96 of these are shown on the Fig. 7. We decided to perform GO enrichment analysis on these visually examined differentially expressed genes (Ashburner et al., 2000; Consortium, 2019).

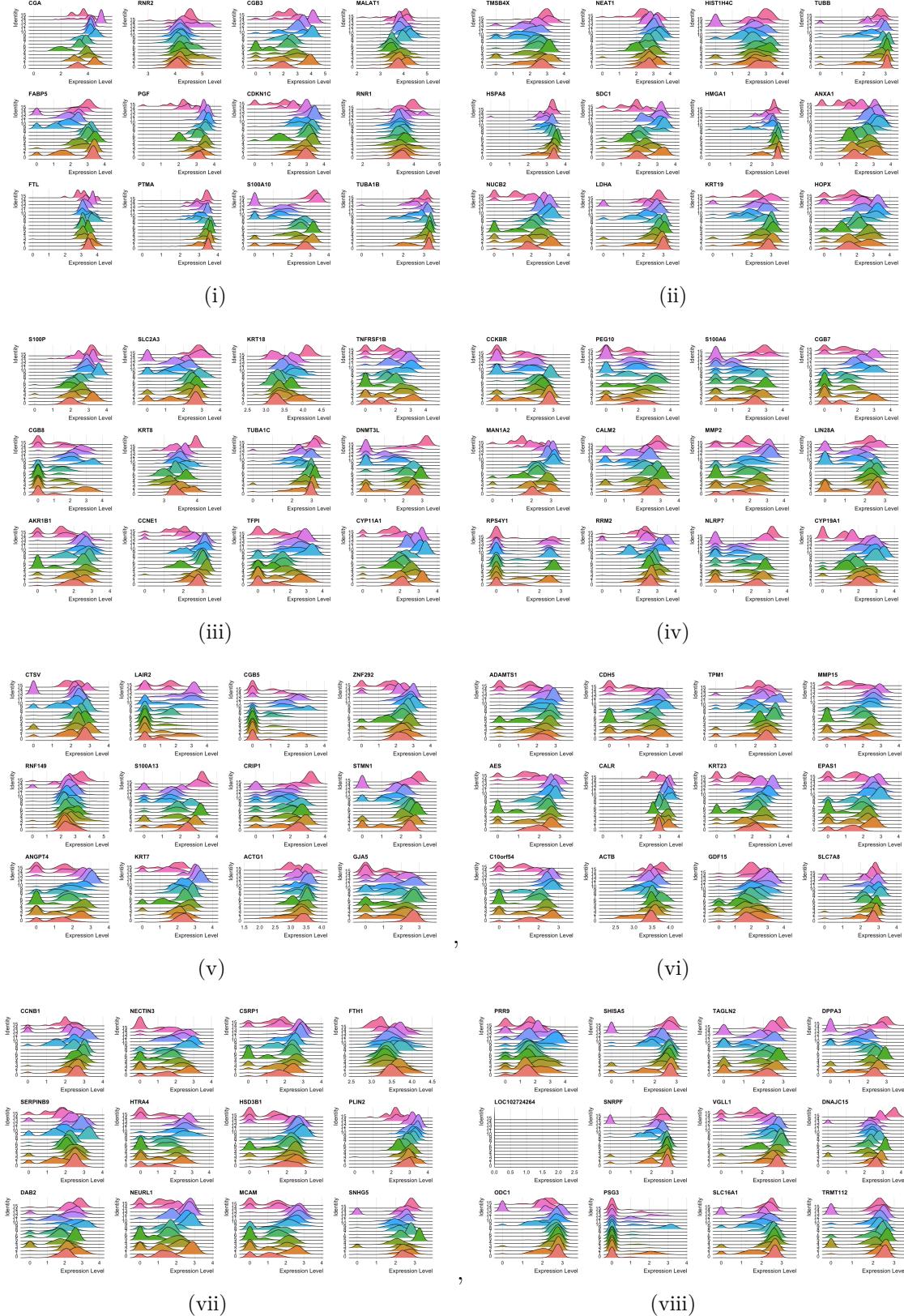


Fig. 7. Top 96 genes for differential expression

Figure 7 (i) Top 12 genes for DE. (ii) Genes in the top 13 to 24 of DE rankings. (iii) Genes in the top 25 to 36 of DE rankings. (iv) Genes in the top 37 to 48 of DE rankings. (v) Genes in the top 49 to 60 of DE rankings. (vi) Genes in the top 61 to 72 of DE rankings. (vii) Genes in the top 73 to 84 of DE rankings. (viii) Genes in the top 85 to 96 of DE rankings.

3. Downstream enrichment analysis

We observed from Fig.8(i) that there are high generatio of cell adhesion molecule binding, strcuctural constituent of cytoskeleton, ubiquitin-like protein ligase binding, cadherin binding, iron ion binding and kinase regulator activity. Considering the p-value, strcuctural constituent of cytoskeleton and iron ion binding have a high degree of credibility, which is also confirmed in the (ii).

And in (iv), it shows the main components of KEGG pathway. For further analysis, we visualized the KEGG pathway in the next step by KEGG Database(Kanehisa, 2019; Kanehisa and Goto, 2000).

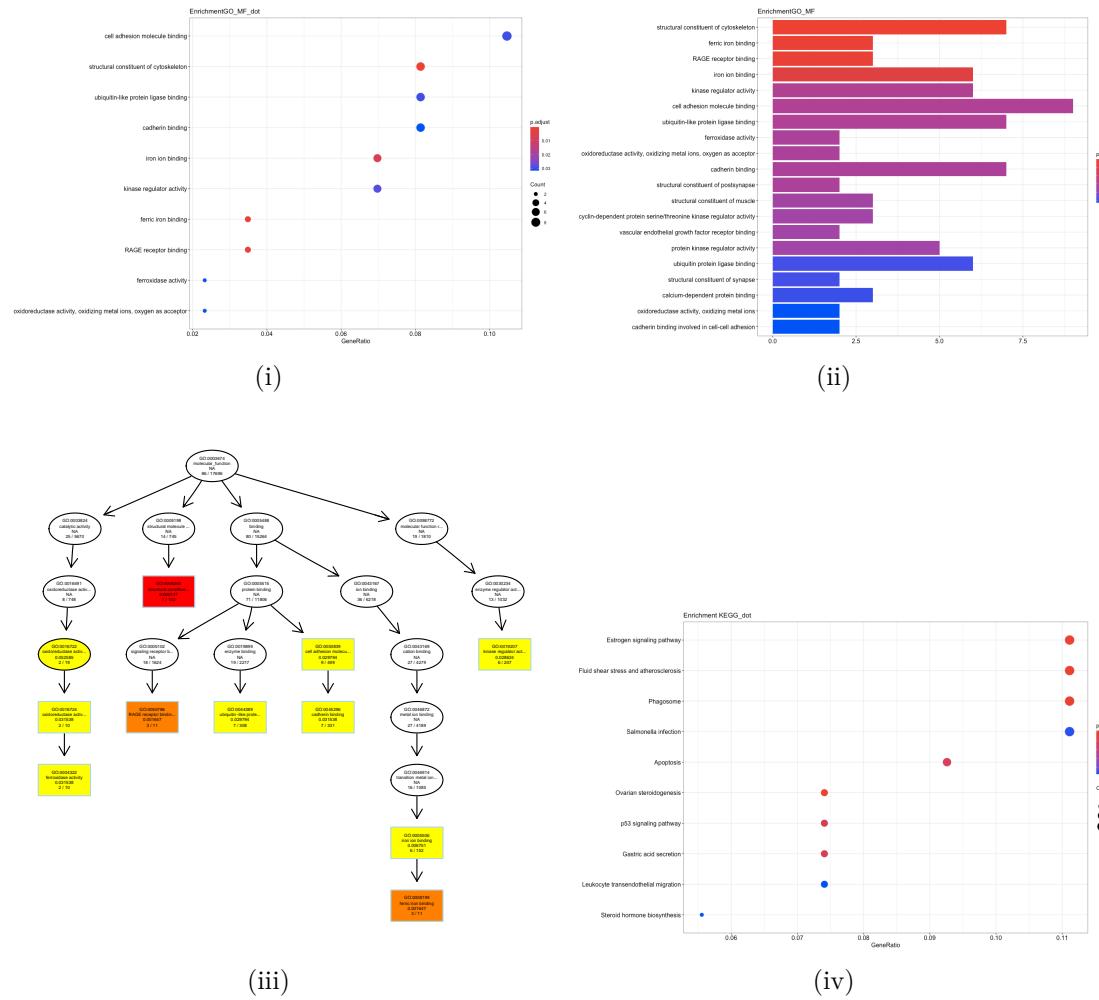


Fig. 8. GO Enrichment analysis

Figure.8 (i)Enrichment GO dotplot. (ii)Enrichment GO barplot. (iii)GOgraph plot. (iv)Enrichment KEGG dotplot.

In Fig.9(ii), (iii) and (iv) we can observe the relationship between the functional modules of differentially expressed genes. There are several main functional modules among them:

iron ion binding which interacts selectively and non-covalently with iron (Fe) ions, ferric iron binding which interacts selectively and non-covalently with ferric iron (Fe(III)), RAGE receptor binding which interacts selectively and non-covalently with the RAGE receptor (the receptor for advanced glycation end-products) and RAGE receptor binding which is the action of a molecule that contributes to the structural integrity of a cytoskeletal structure.

In Fig.9(i), we ranked the KEGG pathways in descending order according to the size of generation, and selected six KEGG pathways with higher generation, which includes RAPI SIGNALING PATHWAY, P53 SIGNALING PATHWAY, PHAGOSOME, APOPTOSIS, OVARIAN STEROIDOGENESIS and ESTROGEN SIGNALING PATHWAY. Their visualization results are shown in Fig.10.

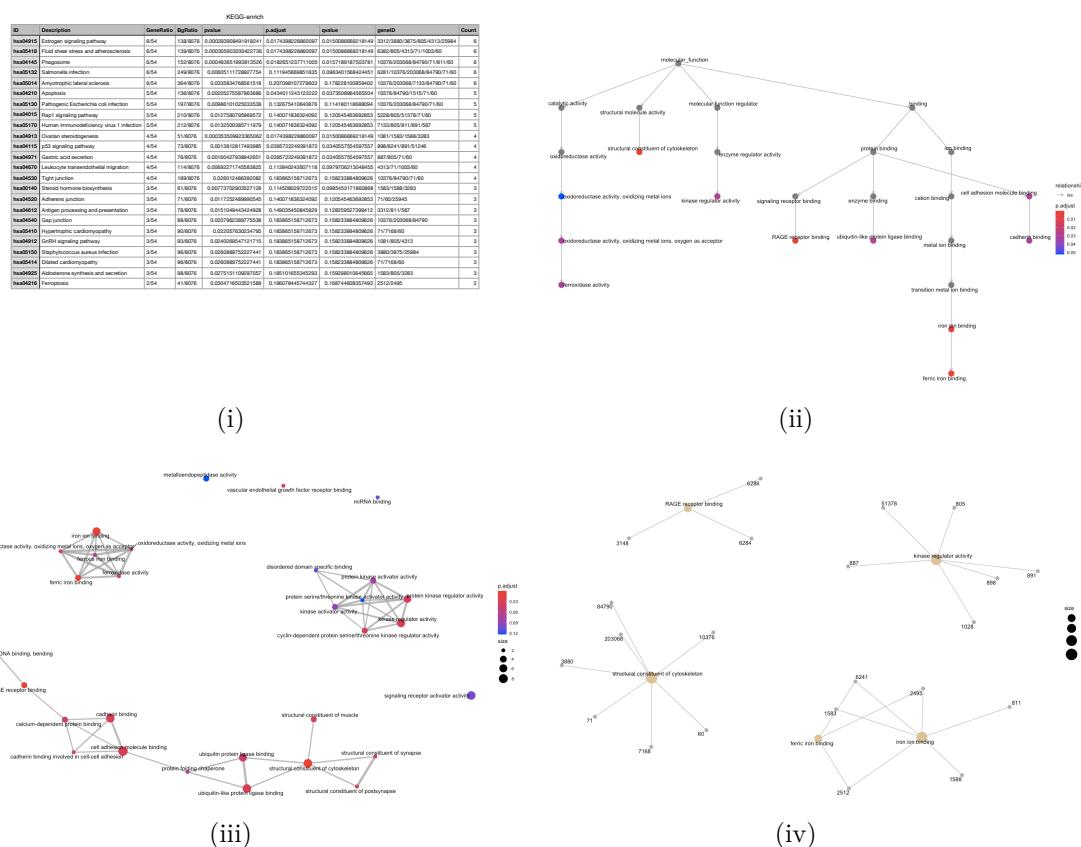


Fig. 9. KEGG analysis

Figure.9 (i)The table of KEGG pathways involved. (ii)GO plot. (iii)EMA plot. (iv)CNE plot.

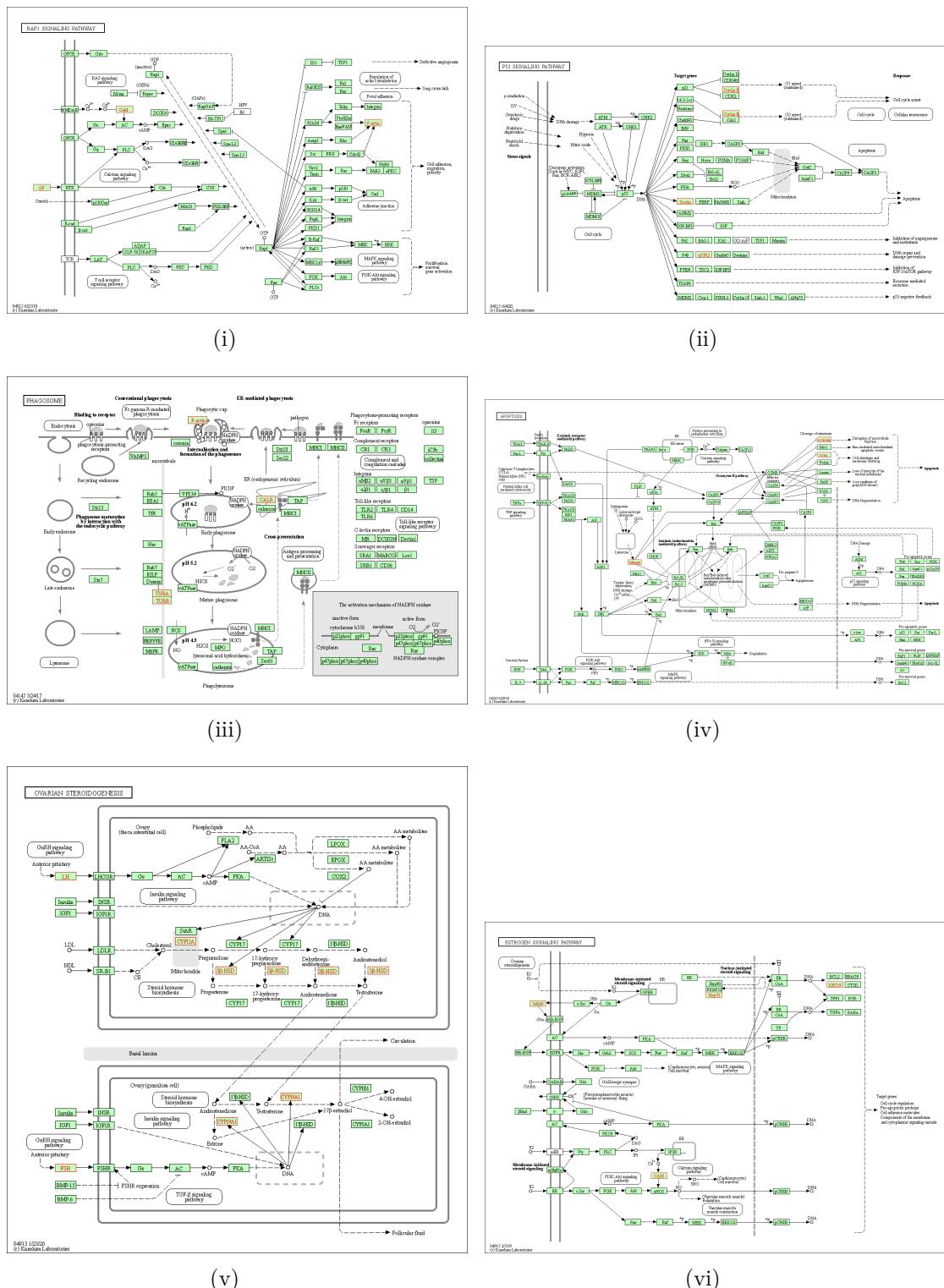


Fig. 10. KEGG Pathway

Figure.10 (i)Rap1 signaling pathway. (ii)p53 signaling pathway. (iii)Phagosome. (iv)Apoptosis. (v)Ovarian steroidogenesis. (vi)Estrogen signaling pathway.

Future research plans

Data availability

Code availability

References

- Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.
- Ariel A Avilion, Silvia K Nicolis, Larysa H Pevny, Lidia Perez, Nigel Vivian, and Robin Lovell-Badge. Multipotent cell lineages in early mouse development depend on sox2 function. *Genes & development*, 17(1):126–140, 2003.
- Hanako Bai, Toshihiro Sakurai, James D Godkin, and Kazuhiko Imakawa. Expression and potential role of gata factors in trophoblast development. *Journal of Reproduction and Development*, 59(1):1–6, 2013.
- Trishita Basak and Rupasri Ain. Long non-coding rnas in placental development and disease. *Non-Coding RNA Investig*, 3:14, 2019.
- Carol J Bult, Judith A Blake, Cynthia L Smith, James A Kadin, and Joel E Richardson. Mouse genome database (mgd) 2019. *Nucleic acids research*, 47(D1):D801–D806, 2019.
- Ian Chambers, Douglas Colby, Morag Robertson, Jennifer Nichols, Sonia Lee, Susan Tweedie, and Austin Smith. Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5):643–655, 2003.
- De Cheng, Zhenzhen Li, Yajun Liu, Yi Gao, and Huayan Wang. Kinetic analysis of porcine fibroblast reprogramming toward pluripotency by defined factors. *Cellular Reprogramming (Formerly “Cloning and Stem Cells”)*, 14(4):312–323, 2012.
- Young Gie Chung, Sarayu Ratnam, J Richard Chaillet, and Keith E Latham. Abnormal regulation of dna methyltransferase expression in cloned mouse embryos. *Biology of reproduction*, 69(1):146–153, 2003.
- Gene Ontology Consortium. The gene ontology resource: 20 years and still going strong. *Nucleic acids research*, 47(D1):D330–D338, 2019.
- Wendy Dean, Fátima Santos, Miodrag Stojkovic, Valeri Zakhardtchenko, Jörn Walter, Eckhard Wolf, and Wolf Reik. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proceedings of the National Academy of Sciences*, 98(24):13734–13738, 2001.
- Yuko Fujiwara, Carol P Browne, Kerrianne Cunniff, Sabra C Goff, and Stuart H Orkin. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor gata-1. *Proceedings of the National Academy of Sciences*, 93(22):12355–12358, 1996.
- Shelley R Hough, Ian Clements, Peter J Welch, and Kristin A Wiederholt. Differentiation of mouse embryonic stem cells after rna interference-mediated silencing of oct4 and nanog. *Stem cells*, 24(6):1467–1475, 2006.

Minoru Kanehisa. Toward understanding the origin and evolution of cellular organisms. *Protein Science*, 28(11):1947–1951, 2019.

Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

Yajun Liu, Yangyang Ma, Jeong-Yeh Yang, De Cheng, Xiaopeng Liu, Xiaoling Ma, Franklin D West, and Huayan Wang. Comparative gene expression signature of pig, human and mouse induced pluripotent stem cell lines reveals insight into pig pluripotency gene networks. *Stem Cell Reviews and Reports*, 10(2):162–176, 2014.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Kaoru Mitsui, Yoshimi Tokuzawa, Hiroaki Itoh, Kohichi Segawa, Mirei Murakami, Kazutoshi Takahashi, Masayoshi Maruyama, Mitsuyo Maeda, and Shinya Yamanaka. The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells. *Cell*, 113(5):631–642, 2003.

Jeffery D Molkentin, Qing Lin, Stephen A Duncan, and Eric N Olson. Requirement of the transcription factor gata4 for heart tube formation and ventral morphogenesis. *Genes & development*, 11(8):1061–1072, 1997.

Jeffery D Molkentin, Kevin M Tymitz, James A Richardson, and Eric N Olson. Abnormalities of the genitourinary tract in female mice lacking gata5. *Molecular and Cellular Biology*, 20(14):5256–5260, 2000.

Edward E Morrisey, Zhihua Tang, Kirsten Sigrist, Min Min Lu, Fang Jiang, Hon S Ip, and Michael S Parmacek. Gata6 regulates hnf4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes & development*, 12(22):3579–3590, 1998.

Susan L Palmieri, Werner Peter, Heike Hess, and Hans R Schöler. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Developmental biology*, 166(1):259–267, 1994.

Pier Paolo Pandolfi, Matthew E Roth, Alar Karis, Mark W Leonard, E Dzierzak, Frank G Grosveld, James Douglas Engel, and Michael H Lindenbaum. Targeted disruption of the gata3 gene causes severe abnormalities in the nervous system and in fetal liver hematopoiesis. *Nature genetics*, 11(1):40–44, 1995.

Matthias Stadtfeld, Effie Apostolou, Hidenori Akutsu, Atsushi Fukuda, Patricia Follett, Sridaran Natesan, Tomohiro Kono, Toshi Shioda, and Konrad Hochedlinger. Aberrant silencing of imprinted genes on chromosome 12qf1 in mouse induced pluripotent stem cells. *Nature*, 465(7295):175–181, 2010.

Fong-Ying Tsai, Gordon Keller, Frank C Kuo, Mitchell Weiss, Jianzhou Chen, Margery Rosenblatt, Frederick W Alt, and Stuart H Orkin. An early haematopoietic defect in mice lacking the transcription factor gata-2. *Nature*, 371(6494):221–226, 1994.

Jingyu Zhang, Xia Wang, Bing Chen, Zhifeng Xiao, Wenmei Li, Youyong Lu, and Jianwu Dai. The human pluripotency gene nanog/nanogp8 is expressed in gastric cancer and associated with tumor development. *Oncology letters*, 1(3):457–463, 2010.

Shuchen Zhang and Wei Cui. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World journal of stem cells*, 6(3):305, 2014.

Xinxin Zhang, Yujia Lan, Jinyuan Xu, Fei Quan, Erjie Zhao, Chunyu Deng, Tao Luo, Liwen Xu, Gaoming Liao, Min Yan, et al. Cellmarker: a manually curated resource of cell markers in human and mouse. *Nucleic acids research*, 47(D1):D721–D728, 2019.

Fan Zhou, Rui Wang, Peng Yuan, Yixin Ren, Yunuo Mao, Rong Li, Ying Lian, Junsheng Li, Lu Wen, Liying Yan, et al. Reconstituting the transcriptome and dna methylome landscapes of human implantation. *Nature*, 572(7771):660–664, 2019.