

The role of tissue-restricted antigens in the early embryonic development

18.07.2022

Clara Certa, Yixin Chen, Linda Kaupp, Alewtina Towara

Supervisor: Dr. Maria Dinkelacker, Dr. Carl Herrmann

Tutor: Ian Dirk Fichtner

Data analysis for students of Molecular Biotechnology

Heidelberg University

Contents

1. Abstract	1
2. Introduction	1
2.1 Early embryogenesis	1
2.2 Chemokines	1
2.3 Tissue-restricted antigens	2
2.4 Aim of this project	2
3. Materials and methods	2
3.1 Datasets	2
3.1.1 Microarray dataset	2
3.1.2 TRA dataset	2
3.2 Packages used in R	2
3.3 Quality control and normalization of the data	3
3.3.1 Single chip control	3
3.3.2 RNA degradation plot	3
3.3.3 Normalization	3
3.4 Principal component analysis	4
3.5 K-means clustering	4
3.6 Limma analysis	4
3.7 Gene set enrichment analysis	5
4. Results	5
4.1 Principal component analysis	5
4.2 Limma analysis	5
4.2.1 Venn diagram	6
4.2.2 Volcano plot	7
4.2.3 Gene set enrichment analysis	8
4.3 Chemokines	9
5. Discussion	9
5.1 General embryonic development	9
5.2 Maternal to zygotic transition	9
5.3 Switch of energy metabolism	10
5.4 Implantation	10

6. Supplementary material	11
6.1 Tissue distribution of original TRA data (Dinkelacker, 2019)	11
6.2 Top 10 up-regualted TRA genes in the compared stages	11
6.2.1 Expression plots of certain top 10 genes	12
6.3 Gene set enrichment analysis	14
7. References	16

1. Abstract

Research, especially in the field of early embryogenesis is important in many ways. It is not only required for a better understanding of genetic defects of the embryos in an early stage but also important for the improvement of assisted human conception. In this report we investigated gene expression data of the early stages of embryogenesis up to the formation of the blastocyst. As the formation of tissues only occurs after the stages displayed by our dataset, we expected to not observe any tissue specific development and therefore no differentially expressed tissue-restricted antigens (TRAs) during this period. However, we found numerous differentially expressed TRAs that display the developmental processes in early embryogenesis accurately. For instance, the inquired genes reflect the activation of transcription of the embryonic genome as well as the switch of energy metabolism in the early embryo. Furthermore, we found three differentially expressed chemokines that play a role in implantation and embryonic development. These insights may contribute to a deeper understanding of embryo implantation and therefore successful conception.

2. Introduction

2.1 Early embryogenesis

After the fertilization, the human embryo begins to cleave every 24 hours. The individual cells can be detected until the 16-cell stage. Till then every single cell is able to form a whole fetus as well as the fetal placenta. After the 16-cell stage, an event called compaction occurs. The individual cells, or blastomeres, become flat and thereby maximize the cell-cell contacts to one another. The result of compaction is the morula stage. The next step is the formation of the blastocyst which is the first differentiative step in early embryo development. The outer cells form an epithelial layer around the embryo called the trophectoderm. The trophectoderm is later needed for the implantation in the uterine endometrium. As soon as the epithelial layer is built those cells start the ingestion of fluid which leads to the formation of the blastocoel cavity. The blastocoel is additionally containing a small number of undefined cells which will later form the actual embryo. Those cells are referred to as the inner cell mass (Leese *et al.*, 1993). Interestingly, the activation of the embryo genome is thought to only start around the 4-cell stage. Prior, the development is driven by maternal stores of RNA which were passed during oogenesis. This process is called the maternal to zygotic transition (MZT)(Ko, 2016). The metabolic activity itself only starts rising at the late morula stage. The two reasons for that are the increased energy demand caused by protein biosynthesis as the number of dividing cells goes up and the ion pumping into the blastocoel cavity. During this phase the energy metabolism of the embryo changes from using aerobic substrates such as pyruvate, lactate, and glutamine to using glucose (Leese *et al.*, 1993). This switch results in a 2.7-fold rise of oxygen consumption in the trophectoderm where the oxidative phosphorylation takes place. As a consequence, the production of harmful reactive oxygen species (ROS) increases. The embryo therefore has to invent a system to prevent high amounts of ROS and at the same time prepare repair mechanisms for fixing the damage (Kaneko *et al.*, 2013).

2.2 Chemokines

The secretion of signal proteins known as chemokines plays an important role in several cellular processes by causing, for example, the movement of epithelial cells or leukocytes (Raman *et al.*, 2011). Those effects also occur during pregnancy. It has been shown that chemokines are essential for the implantation of the embryo and for embryonic development (Hannan *et al.*, 2007)

2.3 Tissue-restricted antigens

Our immune system is pivotal for protecting our body from exogenous pathogens. It consists of two parts, the innate immune system and the adaptive immune system. The innate immune system reacts very quickly to foreign substances but is restricted to recognizing conserved structures like bacterial cell wall components, whereas the adaptive immune system develops after birth and is able to target a great variety of antigens. An important part of the adaptive immune response are T cells which recognize antigens with specific T cell receptors (TCRs). In this process, it is crucial that the immune system solely attacks foreign pathogens and not the body's own tissues. To gain self-tolerance, a selection process takes place in the thymus where T cells are screened for T cell receptors that can recognize antigens but not self-antigens. First, the T cells undergo positive selection. In this step, only those T cells that can recognize self-major histocompatibility complex (MHC) proteins are selected to survive. Thereafter, a negative selection process takes place which prevents the emergence of autoreactive T cells. This is based on triggering of apoptosis in T cells with TCRs that are capable of recognizing self-antigens in the context of self-MHC. These self-antigens are referred to as tissue-restricted antigens (TRAs) and are presented by medullary thymic epithelial cells (mTECs) (Murphy *et al.*, 2012). TRAs are genes which are expressed considerably higher in certain tissues compared to other tissues. Thus, TRAs are characteristic for the tissues they belong to. However, in mTECs, TRAs are expressed promiscuously to ensure that all tissues are represented in the thymus. Hence, every autoreactive T cell is sorted out (Dinkelacker, 2007).

2.4 Aim of this project

This project aims to investigate the expression of TRAs in early embryonic development. Tissue-specificity plays an important role in embryonic development as embryonic stem cells differentiate step by step to eventually form organs. This development could be monitored by analyzing gene expression data of different stages in embryogenesis. In this project, we are investigating the preimplantation stages of embryonic development up to the formation of the blastocyst. In the course of our work, we will determine the differentially expressed genes (DEG) between the first few stages of embryonic development and extract the differentially expressed TRAs and chemokines for further analysis of their function.

3. Materials and methods

3.1 Datasets

In order to study the cellular and tissue development during human early embryogenesis, we analyzed a microarray dataset published by Xie *et al.* in 2010. We examined the amount of differentially expressed TRAs in our dataset by comparing it with the TRA dataset provided by Dr. Dinkelacker in 2019.

3.1.1 Microarray dataset

The microarray dataset by Xie *et al.* contained three datasets for three species: bovine, mouse and human. We decided to focus our project on the human dataset. The data was generated by RNA extraction, amplification, and hybridisation onto Affymetrix microarrays. For early human embryogenesis, they prepared three replicates at 1-cell stage, 2-cell stage, 4-cell stage, 8-cell stage, morula, and blastocyst each. Each replicate includes 95,659 transcripts.

3.1.2 TRA dataset

As a reference TRA dataset, we chose the GTEx RNAseq human dataset provided by Dr. Dinkelacker in 2019. It consists of 60,131 transcripts from 53 different tissues. We used the TRA dataset and Ensembl gene information to annotate our raw data.

3.2 Packages used in R

During the analysis, a number of packages were used which are listed in Table 1. The code was written using the 4.2.0 version of R.

Table 1: List of all R and Bioconductor packages used in our project.

affy 1.74.0	vsn 3.64.0	ggplot2 3.3.6	AnnotationDbi 1.58.0	VennDiagram 1.7.3
enrichplot 1.16.1	limma 3.52.2	dplyr 1.0.9	tidyverse 1.3.1	hexbin 1.28.2
hgu133plus2hsenstcdf 25.0.0	cluster 2.1.3	factoextra 1.0.7	org.Hs.eg.d 3.15.0	go.db 3.15.0
hgu133plus2hsenstprobe 25.0.0	pheatmap 1.0.12	clusterProfiler 4.4.4	tinytex 0.40	RColorBrewer 1.1-3

3.3 Quality control and normalization of the data

Before analyzing the microarrays, it is important to determine the quality of the used raw data. This is done by the procedure called Quality Control (QC) where potential mistakes in the experiment can be identified.

3.3.1 Single chip control

The first step of QC is the single-chip analysis where major quality problems like fingerprints, local irregularities, imprints of pipette tips and extreme light intensities can be detected by a visual inspection of the microarray images. After looking at our chips, we assumed that the GSM456645 chip could be of bad quality due to a smudge on the bottom edge.

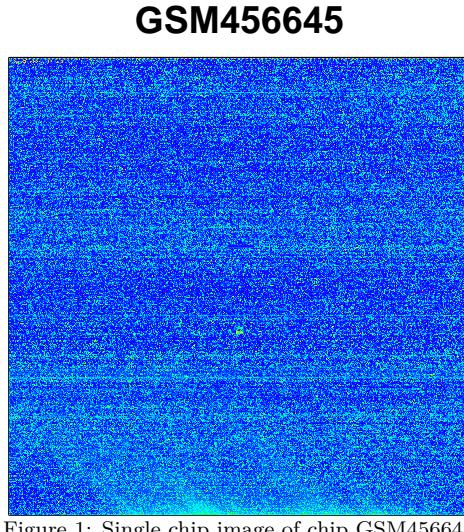


Figure 1: Single chip image of chip GSM456645

3.3.2 RNA degradation plot

In the next step, we analyzed the RNA degradation that is an indicator for bad quality RNA. Due to the start of the RNA degradation at the 5' end, the intensities of probes at the 5' end must be lower than at the 3' end. In our RNA degradation plot, where the mean intensity for each probe set is plotted versus the probe number, no high slope and disagreements between the arrays were detected.

3.3.3 Normalization

By performing the variance stabilization normalization (vsnrma) (Huber *et al.*, 2002), the data is corrected for biases from non-biological sources, like labeling efficiency and scanner setup. We also measured the quality of the variance stabilization by plotting the row standard deviation versus row means.

The red line in the meanSdplot is not an approximately horizontal as it should be. This could be caused by the standard deviation at high expression levels being higher than at low levels even if the relative deviation from mean is identical.

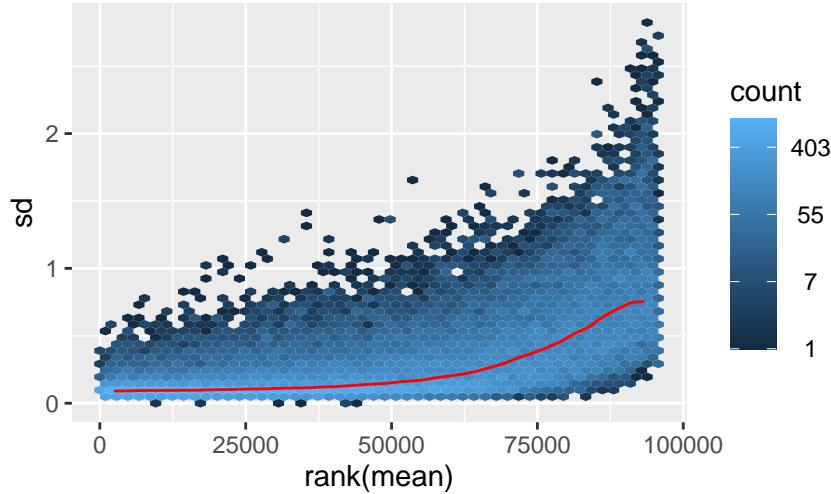


Figure 2: MeanSd plot

Different plots were reviewed in order to determine the degree of the systematic variation in the raw data. For both the pre-normalized data and the normalized data, density plots and boxplots were generated.

The boxplots visualize the variation of the gene expression for each chip. After the normalization, no boxes were significantly elevated or more spread out than other arrays.

In the density plots for each chip the density function is plotted versus log expression. After the vsnrma normalization, the distribution of all curves in the plot showed a vicinity.

For the control of the normalization, scatter plots were produced which compare the log fold intensity change between an array and a reference array.

As a consequence of our QC, we decided to keep all chips for further analysis because no serious outliers were found after the normalization.

3.4 Principal component analysis

Principal component analysis (PCA) is a statistical method to reduce the dimensionality of a data set. It is an orthogonal linear transformation of the data to a new coordinate system, so that the greatest variance of the data is demonstrated by the first coordinate. The new coordinates are called principal components, which are linear combination of variables from the data (Jolliffe, 2002).

3.5 K-means clustering

K-means is a clustering method for which one first must indicate the number of clusters (k). Later the number can be determined by an elbow-plot or silhouette-plot. The algorithm starts by choosing random centroids of k clusters. Each object is now assigned to the cluster for which the Euclidean distance between centroid and object is the smallest. In the next step, the centroids get updated by calculating the mean of each cluster containing the new objects. Those steps are repeated till the cluster assignments of the objects stop changing (Bortz and Schuster, 2010).

3.6 Limma analysis

For analysis of differential gene expression we conducted limma analysis. Therefore, the limma package was employed, which facilitates statistical testing of expression level changes of thousands of genes at once. For that, limma uses linear models to calculate the test statistic between sample group expression levels (Ritchie *et al.*, 2015).

3.7 Gene set enrichment analysis

Gene set enrichment analysis (GSEA) is a method to identify groups of genes in a dataset which are overrepresented more than expected in any known set of genes. These gene sets can be groups of genes that share a common biological function, chromosomal location, or regulation. GSEA uses statistical approaches to find groups of genes that are significantly enriched or depleted. The goal of this method is to better understand the biological function of groups of genes in a data set (Subramanian *et al.*, 2005). We used Gene Ontology as input data base.

4. Results

4.1 Principal component analysis

In order to verify that the variance in our dataset results from the biological background more than from the technical process, we performed a principal component analysis by plotting PC1 versus PC2. Moreover, we thereby achieved a greater insight into possible interesting stages for examination.

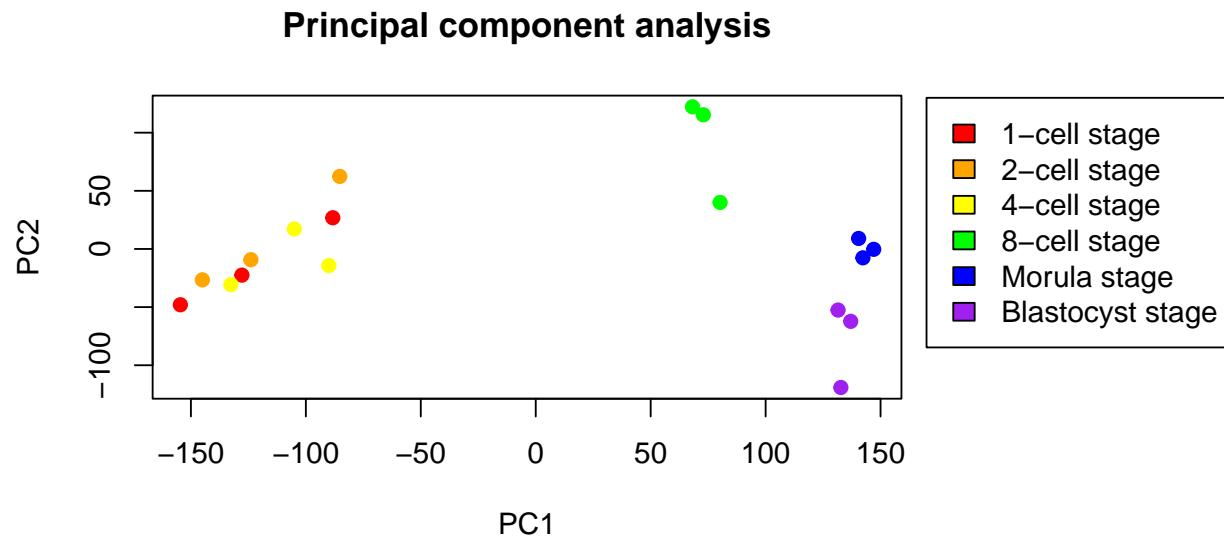


Figure 3: Principal component analysis showing PC1 versus PC2

We grouped our replicates using k-means clustering on the first four principal components which explain 90% of the variance. By performing a silhouette plot we indicated four as the number of clusters. The clusters were composed as follows: cluster 1 (1-cell, 2-cell and 4-cell stage), cluster 2 (8-cell stage), cluster 3 (morula stage) and cluster 4 (blastocyst stage). These clusters can also be visually observed in our PCA (see figure 3). Therefore, we did not observe any technical effects in our microarray data. Due to this result, we focused our project on investigating the difference in gene expression between 1-cell stage and 8-cell stage, 8-cell stage and morula and morula and blastocyst.

4.2 Limma analysis

We performed limma analysis using the normalized expression data to discover the DEGs between every combination of stages. We filtered the resulting DEGs for a p-value of 0.01 and annotated the genes with Ensembl and the TRA data. Based on our relevant clusters found by k-means clustering, we solely focused on the DEGs between stages 1-cell to 8-cell, 8-cell to morula and morula to blastocyst for further analysis. For the comparison of the 1-cell to 8-cell stage, we found 23,444 DEGs. For 8-cell stage to morula stage there were 6,392 DEGs and for morula stage to blastocyst we got 5,139 DEGs.

We wanted to investigate our hypothesis and determine whether there are differentially expressed TRAs in our dataset. Thus, we displayed the distribution of TRAs differentially expressed in our chips according to their max tissue (see figure 4).

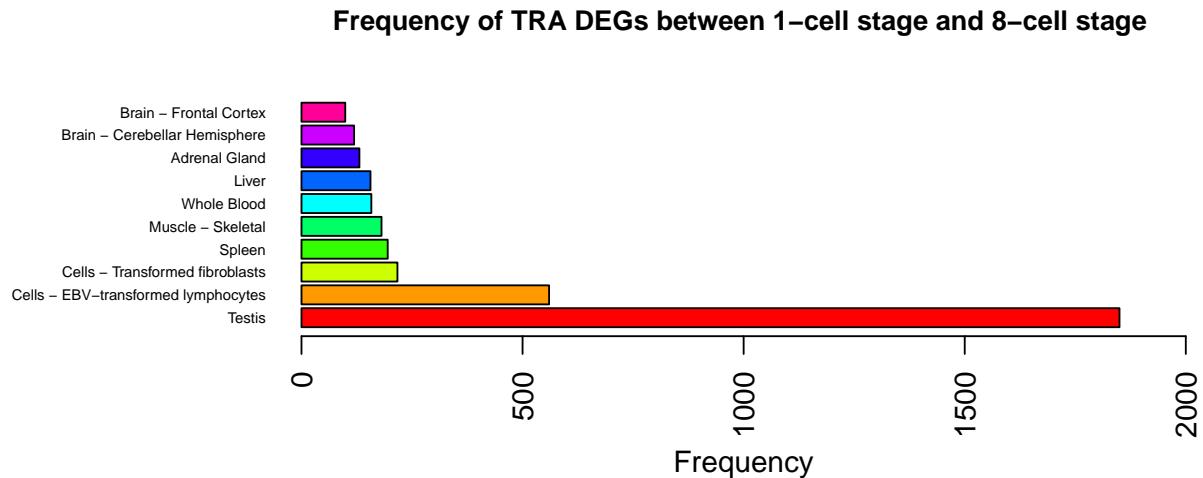


Figure 4: Frequency of TRA DEGs between 1-cell stage and 8-cell stage sorted by tissues

It seems like even though there is no tissue development in the observed stages, certain TRAs are still differentially expressed. We also analyzed the original TRA dataset provided by Dr. Dinkelacker (2019) to substantiate the validity of this distribution. We noted a high level of testis genes in the original data (see figure 9).

To visualize our results based on limma and to find out interesting TRA genes that play a significant role in embryogenesis, we decided to use a Venn diagram and volcano plots to process our results.

4.2.1 Venn diagram

To gain an overview of TRA genes that are important in all stages, we used a Venn diagram to plot the intersection of differentially expressed TRA genes comparing the stages 1- to 8-cell, 8-cell to morula as well as morula to blastocyst (see figure 5). We extracted genes that are continuously up- and down-regulated along the three stage transitions, since they might be important for embryogenesis.

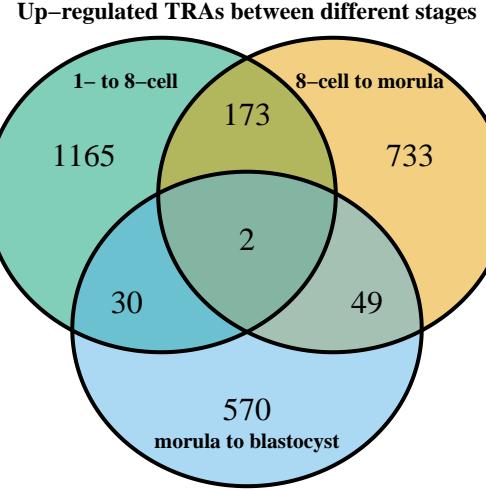


Figure 5: Venn diagram of up-regulated TRA genes

We found five transcripts which encoded three TRA genes (DAB2, PERP, CHI3L2) that are continuously up-regulated in all three stage transitions and six TRA genes (OTX2, TENT5C, TPRXL, TUBB7P, THBS4, TCL1A) that are continuously down-regulated in all three stage transitions.

4.2.2 Volcano plot

After looking at TRA genes that are important for all stages, we focused on each individual stage transition. In order to display the differential gene expression between each two stages, we created three volcano plots for three stage transitions and colored the significantly DEGs with a p-value cutoff of 0.01 and a log fold change ($\log_{10}FC$) cutoff of 2. To extract the most biologically significant genes, we identified the top 10 up- and down-regulated TRA transcripts that are most significantly expressed and have the highest logFCs. These genes are labeled in the volcano plots (see figure 6).

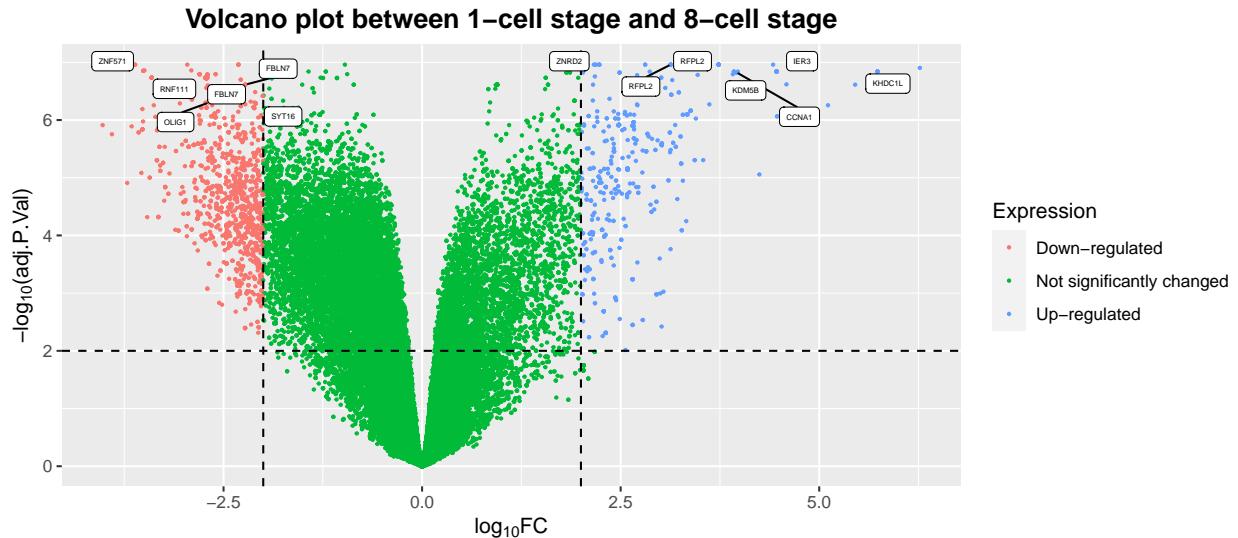


Figure 6: Volcano plot comparing the 1-cell and 8-cell stage

The top 10 up-regulated TRA genes were displayed in table 2-4 in the supplementary material. The function of the interesting TRA genes are investigated in discussion.

After extracting the top 20 TRA DEGs of the three different stage transitions from the limma tables, we compared them with 20 genes which explain the most variance in PCA by conducting PCA separately with the six replicates from each stage transition. Only genes with the highest loading from PC1 are chosen. We found that 46.67% of the TRA DEGs between the 1- and 8-cell stage can be found in PCA top variance genes. 33.33% of TRA DEGs can be found in PCA top variance genes between 8-cell and morula stage and 45.45% between morula and blastocyst stage. Therefore, the top TRA DEGs can be considered explaining the variance between different stages and are relevant for the biological differences between them.

4.2.3 Gene set enrichment analysis

After extracting the top up- and down-regulated TRA DEGs, we decided to perform GSEA to find out if our DEGs were generally enriched in certain gene sets. Using the clusterProfiler package, we conducted GSEA for all three stage transitions and received the following results.

In the dot plot of GSEA (see figure 14) which visualizes the activated or suppressed top gene sets enriched in our DEGs from stages 1- to 8-cell, a lot of ribosome-related gene sets are highly and significantly activated, like “cytosolic ribosome”, “ribosome” and “ribosome subunit” as cellular component. Also in the results table (saved in Github repository), several transcription-related gene sets are highly activated, like “basal transcription machinery binding” and “general transcription initiation factor binding” as molecular function. No interesting gene sets can be found in the dot plot of GSEA from stages 8-cell to morula (see figure 15). However, by looking at the GSEA result table (saved in Github repository), 27 activated mitochondrial gene sets are enriched in our set of DEGs, like “mitochondrial DNA replication” and “positive regulation of mitochondrion organization” as biological process. In the dot plot for stages morula to blastocyst (see figure 7), several ion-related gene sets are enriched in our set of DEGs with a gene ratio of nearly one, like “anion channel activity” and “chloride channel activity” as molecular function and “chloride transport” as biological process.

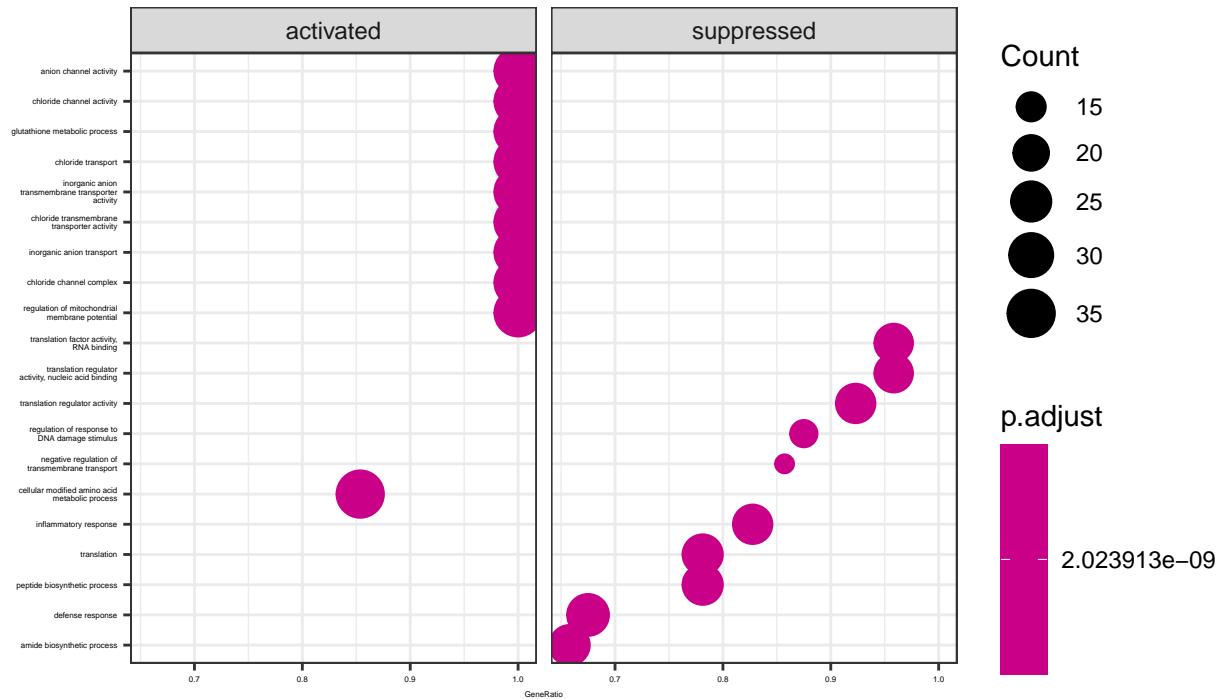


Figure 7: Gene set enrichment analysis comparing morula and blastocyst stage

4.3 Chemokines

We looked for proteins with a chemokine motif in all of our DEGs including non-TRA genes. We visualized the following results in a heatmap (see figure 8):

Between the stages 1-cell to 8-cell CXCL14 was up-regulated with a logFC of approximately 1 and the chemokine receptor CCR3 was down-regulated (logFC = -1.12). Between the stages 8-cell to morula the expression of CCL15 was up-regulated with a logFC of approximately 2.45 whereas in morula to blastocyst the expression was down-regulated with a logFC of -2.66.

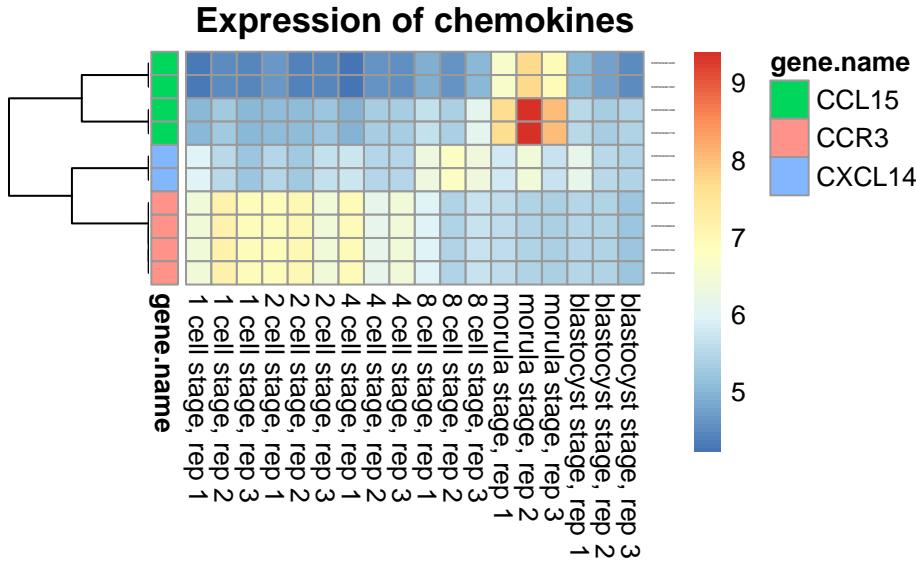


Figure 8: Differential expression of different chemokines

5. Discussion

We originally assumed that no TRAs can be found in our dataset, since our data only consists of stages from early embryogenesis, which means that the development of tissues has not taken place yet. However, we did find numerous differentially expressed TRAs, indicating that TRAs are not only expressed by their specific tissue, but also by undifferentiated embryonic cells.

5.1 General embryonic development

As might be expected, the TRAs that are significantly up-regulated over all three investigated stages have important functions in embryonic development. DAB2 for example is involved in the formation of the visceral endoderm (Morris *et al.*, 2002) while PERP plays a critical role in stratified epithelial development (Ihrie *et al.*, 2005).

5.2 Maternal to zygotic transition

Most importantly, our results indicate the onset of MZT: The zygote switches from a transcriptionally silent stage, in which maternal mRNA is dominant in the cell, to genome activation and maternal mRNA starts to get degraded (Schier *et al.*, 2007).

The results show that the MZT is initiated around 4-cell stage. This can be seen in the differential expression of the TRA genes like KDM5B. KDM5B, a testis TRA gene which codes for the enzyme histone lysine demethylases 5B, is found highly up-regulated between 1-cell stage and 8-cell stage with a logFC of 5.23. Together with KDM5C, KDM5B is involved in mammalian embryo transcription activation and DNA damage response (Glanzner *et al.*, 2020). The onset of MZT can also be indicated by the GSEA results: several

transcription-related gene sets are found enriched in this stage transition according to GSEA, suggesting that the embryonic genome is activated. With the activation of embryonic genome, an activation of ribosome-related gene sets can also be observed in GSEA, indicating that the embryo has started to synthesize its own ribosomes for protein translation to support rapid cell division and cell growth.

5.3 Switch of energy metabolism

During our limma analysis of the DEGs we also found genes indicating a change of energy source in the mammalian embryo development. In the late morula stage, glucose becomes the predominant exogenous energy substrate, while prior the energy was mostly derived from aerobic substrates like pyruvate, lactate, and glutamine (Leese *et al.*, 1993). At the same time, the oxygen consumption increases. This is a result of ATP synthesis in the trophectoderm which is driven by oxidative phosphorylation. Unfortunately, the production of reactive oxygen species (ROS) thereby also rises (Kaneko *et al.*, 2013). One indicator for this event is that in our data the SLC2A3 gene was significantly up-regulated in the morula stage (see figure 11). SLC2A3 encodes for a glucose transporter (GLUT3) which is responsible for the diffusion of glucose across plasma membranes (Ziegler *et al.*, 2020). At the same time, our GSEA results between 8-cell and morula stage also show a great enrichment of mitochondrial-related gene sets, indicating that the cell organelles are activated, in which oxidative phosphorylation takes place. Furthermore, we found genes encoding for proteins which prevent excess accumulation of ROS during oxidative phosphorylation (TEAD4, see figure 12)(Kaneko *et al.*, 2013) and proteins which defend the cell against oxidative stress (NQO1, see figure 13)(Ross *et al.*, 2021). This energy homeostasis is thought to be caused by increased energy demands due to two processes: the increased protein synthesis because the embryo starts its net growth and the pumping of sodium ions by the Na⁺/K⁺ ATPase in the trophectoderm to form the blastocoel (Leese *et al.*, 2008).

5.4 Implantation

Even though our data only includes preimplantation stages, we still can observe the first differentiative event in embryogenesis: the formation of the blastocyst. The blastocoel, a cavity with fluid in blastocyst, is formed by the Na⁺/K⁺ ATPase in the trophectoderm that draws water. This can be confirmed by the GSEA dot plot between morula stage and blastocyst stage, in which ion transport-related pathways are highly enriched. At the same time, the outer cells start to form tight junctions and build an epithelial layer which is called the trophectoderm. This layer is later involved in the implantation process meaning the blastocyst will settle in the uterine endometrium (Leese *et al.*, 1993). To ensure a successful implantation several arrangements must be made in advance. One of them is the expression of STMN1 (see figure 10) which promotes the migration and invasion of the trophoblast by regulating the expression of N-cadherin and E-cadherin. N-cadherin promotes invasion whereas E-cadherin suppresses the invasion (Tian *et al.*, 2015). In our data one can observe a significant increase in STMN1 expression during the morula stage which indicates that the preparation for implantation starts long before the actual implantation event.

As described in the introduction, chemokine expression is important during the embryonic development. For successful implantation, new blood vessels need to be formed in the endometrium (Torry *et al.*, 2007). CCL15 showed to induce the process of angiogenesis (Hwang *et al.*, 2004). Our analysis showed that this chemokine is first up-regulated between the stages 8-cell and morula. This also demonstrates the process of implantation preparation. However, we could not explain the down-regulation between the stages morula and blastocyst. Our findings on the up-regulation of CXCL14 hint that it could also play a significant role in embryogenesis. This chemokine is related to CXCL12 whose expression showed to encourage angiogenesis and a positive pregnancy course (Koo *et al.*, 2021). In line with previous studies, where CXCL14 is involved in the same embryonic pathways as CXCL12 (Nassari *et al.*, 2017), the expressions seem logical. All in all, it can be said, that our analysis points up that chemokines play a role in early embryonic development. But ensuing steps for further analysis and research should be taken to study the mechanism by which those signal proteins are involved in embryogenesis.

6. Supplementary material

6.1 Tissue distribution of original TRA data (Dinkelacker, 2019)

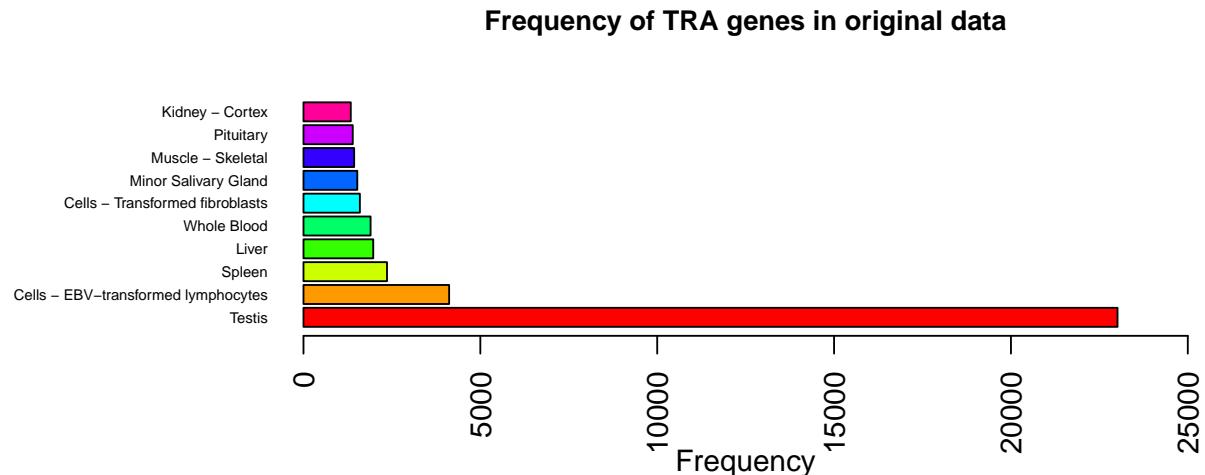


Figure 9: Frequency of TRAs in the original data (Dinkelacker, 2019) sorted by tissues

6.2 Top 10 up-regulated TRA genes in the compared stages

Table 2: Up-regulated genes comparing 1-cell and 8-cell stage

ensembl.transcript	logFC	gene.name
ENST00000491153	5.202884	KDM5B
ENST00000338791	4.979271	IMPDH1
ENST00000348127	4.979271	IMPDH1
ENST00000469328	4.979271	IMPDH1
ENST00000527413	3.125206	ZNRD2
ENST00000370388	7.343357	KHDC1L
ENST00000460428	3.550244	SFPQ
ENST00000376377	4.868293	IER3
ENST00000302097	5.501666	ZNF280A
ENST00000531462	3.530502	TIMM10B

Table 3: Up-regulated genes comparing 8-cell and morula stage

ensembl.transcript	logFC	gene.name
ENST00000447510	2.561530	PROM1
ENST00000505450	2.561530	PROM1
ENST00000513448	2.561530	PROM1
ENST00000513946	2.561530	PROM1
ENST00000539194	2.561530	PROM1
ENST00000399728	3.010510	STMN1
ENST00000455785	2.878334	STMN1
ENST00000358409	2.083279	TEAD4
ENST00000359864	2.083279	TEAD4
ENST00000543435	3.731967	SLC2A3

Table 4: Up-regulated genes comparing morula and blastocyst stage

ensembl.transcript	logFC	gene.name
ENST00000237612	5.066903	ABCG2
ENST00000379047	3.934499	NQO1
ENST00000439109	3.914990	NQO1
ENST00000374208	2.486778	CRYBG2
ENST00000374211	2.400154	CRYBG2
ENST00000561500	3.198884	NQO1
ENST00000564043	3.227720	NQO1
ENST00000353492	2.380233	ATP6V0A4
ENST00000265162	2.802331	ENPEP
ENST00000297325	2.640385	SUN3

6.2.1 Expression plots of certain top 10 genes

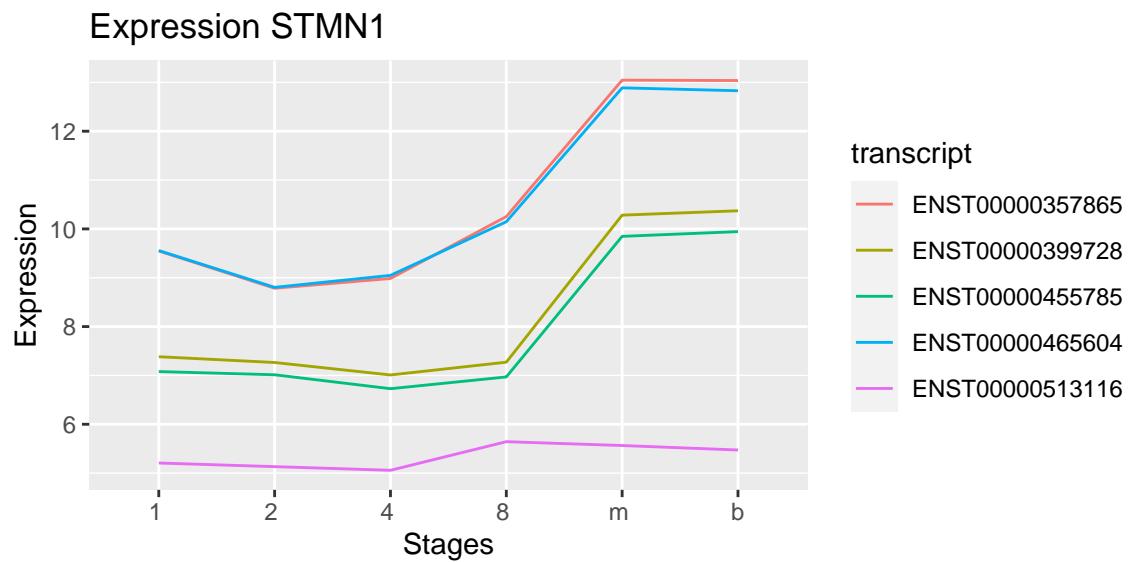


Figure 10: Expression plot of Stathmin 1

Expression SLC2A3

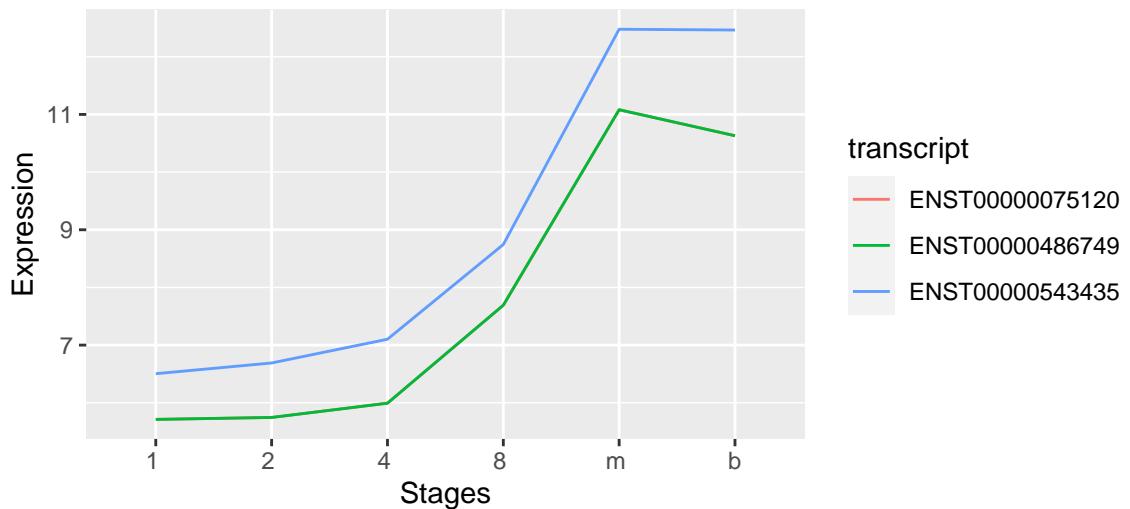


Figure 11: Expression plot of Solute Carrier Family 2 Member 3

Expression TEAD4

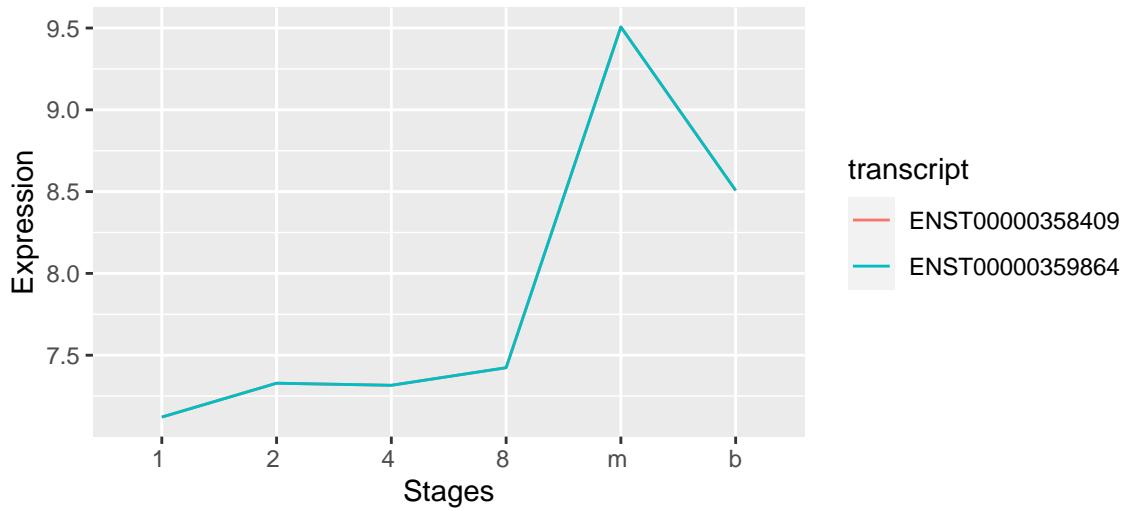


Figure 12: Expression plot of TEA Domain Transcription Factor 4

Expression NQO1

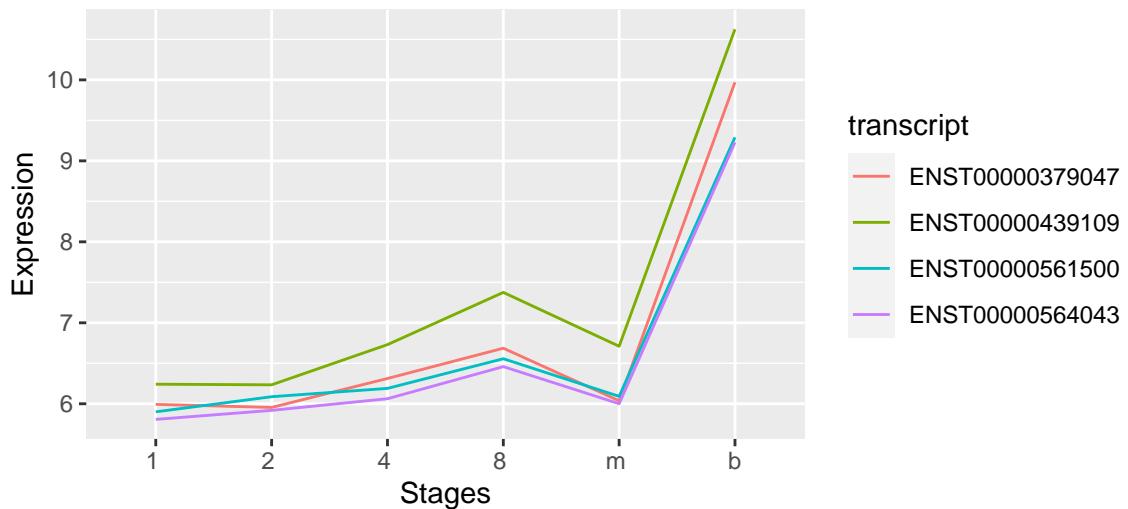


Figure 13: Expression plot of NAD(P)H Quinone Dehydrogenase 1

6.3 Gene set enrichment analysis

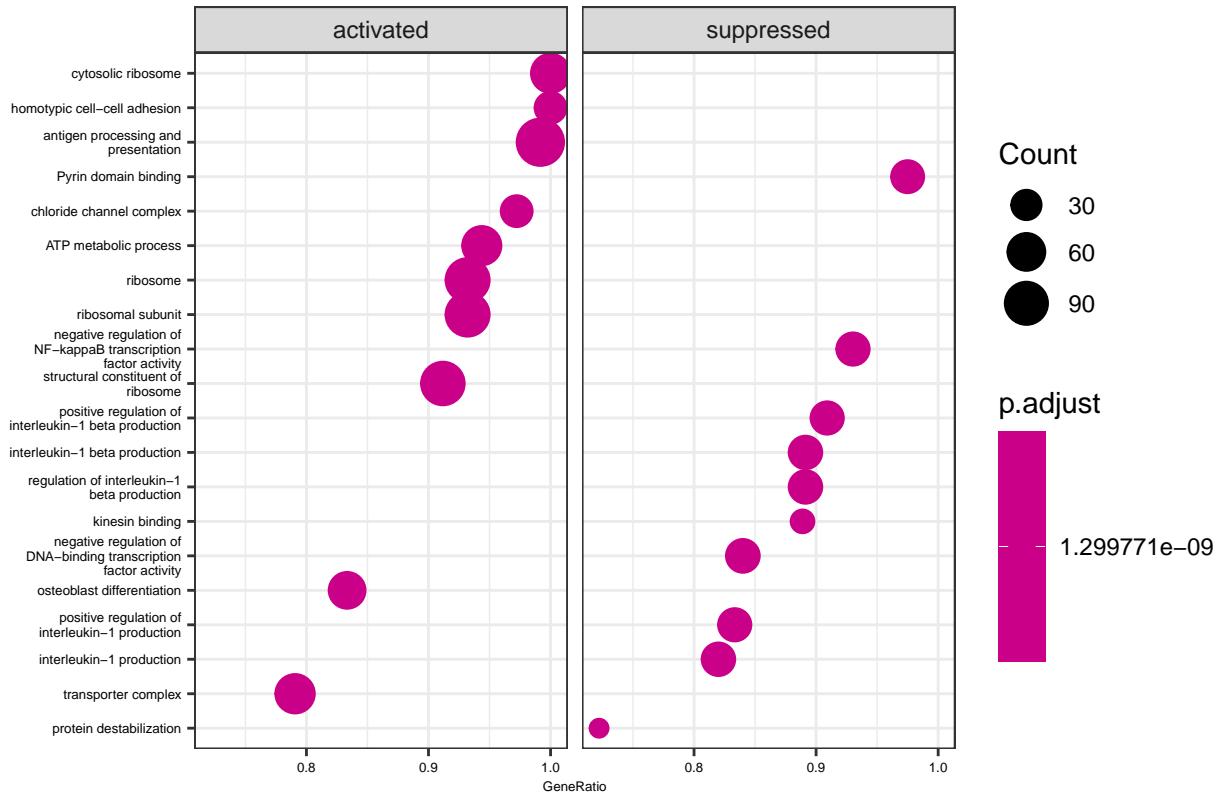


Figure 14: Gene set enrichment analysis comparing 1-cell and 8-cell stage

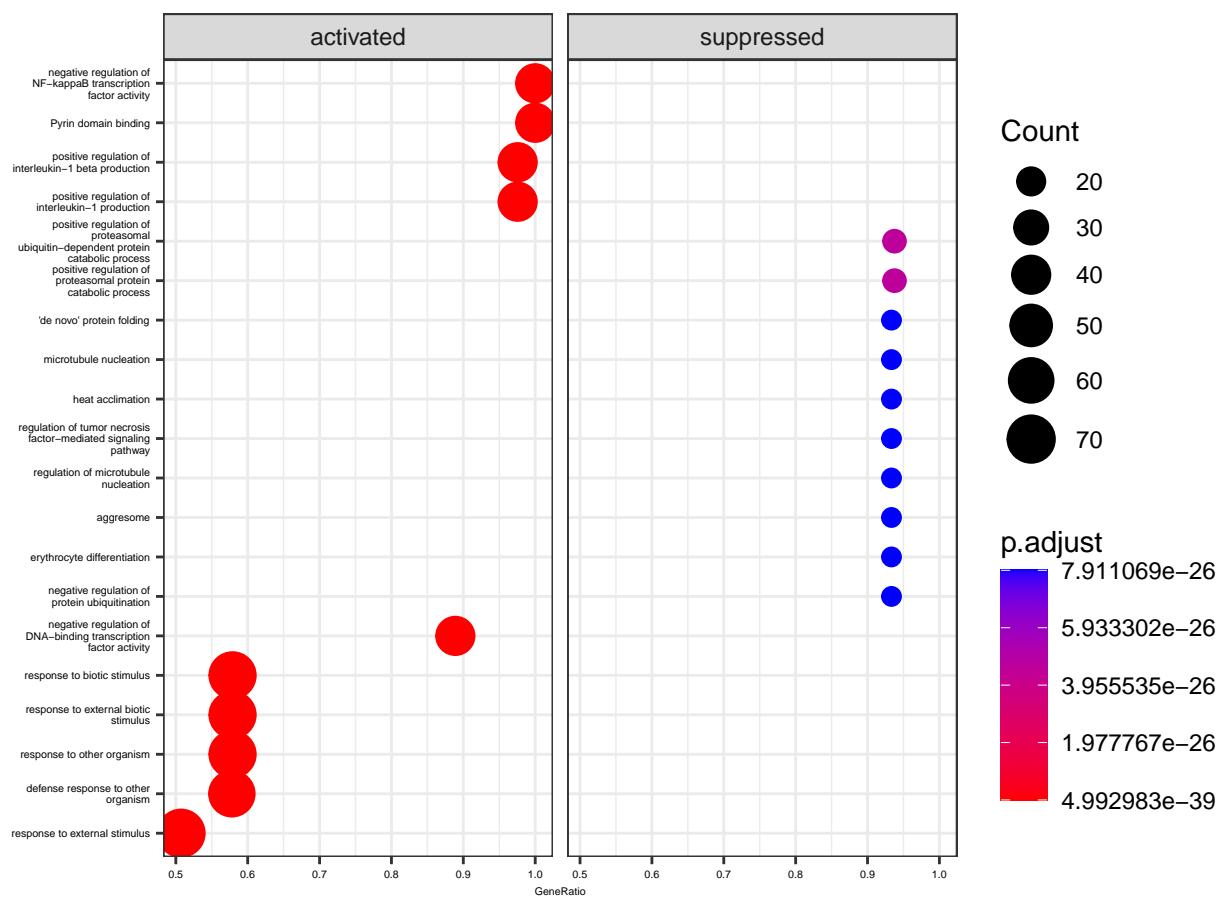


Figure 15: Gene set enrichment analysis comparing 8-cell and morula stage

7. References

- Bortz, J., Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler, 7 Edition (Springer Berlin, Heidelberg). 10.1007/978-3-642-12770-0.
- Corkery, D.P., Holly, A.C., Lahsaee, S., and Dellaire, G. (2015). Connecting the speckles: Splicing kinases and their role in tumorigenesis and treatment response. *Nucleus* 6, 279-288. 10.1080/19491034.2015.1062194.
- Dinkelacker, M.A. (2007). A database of genes that are expressed in a tissue-restricted manner to analyse promiscuous gene expression in medullary thymic epithelial cells. Diplomarbeit (Albert-Ludwigs-Universität).
- Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A., Dudoit, S. . (2005). Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Springer New York, NY). 10.1007/0-387-29362-0.
- Glanzner, W.G., Gutierrez, K., Rissi, V.B., de Macedo, M.P., Lopez, R., Currin, L., Dicks, N., Baldassarre, H., Agellon, L.B., and Bordignon, V. (2020). Histone Lysine Demethylases KDM5B and KDM5C Modulate Genome Activation and Stability in Porcine Embryos. *Front Cell Dev Biol* 8, 151. 10.3389/fcell.2020.00151.
- Hannan, N.J., and Salamonsen, L.A. (2007). Role of chemokines in the endometrium and in embryo implantation. *Current Opinion in Obstetrics and Gynecology* 19, 266-272. 10.1097/GCO.0b013e328133885f.
- Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18 Suppl 1, S96-104. 10.1093/bioinformatics/18.suppl_1.s96.
- Hwang, J., Kim, C.W., Son, K.N., Han, K.Y., Lee, K.H., Kleinman, H.K., Ko, J., Na, D.S., Kwon, B.S., Gho, Y.S., and Kim, J. (2004). Angiogenic activity of human CC chemokine CCL15 in vitro and in vivo. *FEBS Lett* 570, 47-51. 10.1016/j.febslet.2004.06.023.
- Ihrie, R.A., Marques, M.R., Nguyen, B.T., Horner, J.S., Papazoglu, C., Bronson, R.T., Mills, A.A., Attardi, L.D. (2005). Perp is a p63-regulated gene essential for epithelial integrity. *Cell* 120, 843-856. 10.1016/j.cell.2005.01.008.
- Jolliffe, I.T. (2002). Principal Component Analysis, 2nd Edition (Springer New York, NY). 10.1007/b98835.
- Kaneko, K.J., and DePamphilis, M.L. (2013). TEAD4 establishes the energy homeostasis essential for blastocoel formation. *Development* 140, 3680-3690. 10.1242/dev.093799.
- Ko, M.S.H. (2016). Chapter Three - Zygotic Genome Activation Revisited: Looking Through the Expression and Function of Zscan4. In *Current Topics in Developmental Biology*, M.L. DePamphilis, ed. (Academic Press), pp. 103-124. 10.1016/bs.ctdb.2016.04.004.
- Koo, H.S., Yoon, M.-J., Hong, S.-H., Ahn, J., Cha, H., Lee, D., Ko, J.-E., Kwon, H., Choi, D.H., Lee, K.-A., et al. (2021). CXCL12 enhances pregnancy outcome via improvement of endometrial receptivity in mice. *Scientific Reports* 11, 7397. 10.1038/s41598-021-86956-y.
- Leese, H.J., Baumann, C.G., Brison, D.R., McEvoy, T.G., and Sturmy, R.G. (2008). Metabolism of the viable mammalian embryo: quietness revisited. *Mol Hum Reprod* 14, 667-672. 10.1093/molehr/gan065.
- Leese, H.J., Conaghan, J., Martin, K.L., and Hardy, K. (1993). Early human embryo metabolism. *Bioessays* 15, 259-264. 10.1002/bies.950150406.
- Morris, S.M., Tallquist, M.D, Rock, C.O., Cooper, J.A. (2002). Dual roles for the Dab2 adaptor protein in embryonic development and kidney transport. *EMBO J* 21, 1555-1564. 10.1093/emboj/21.7.1555.
- Murphy, K., and Weaver, C. (2018). Janeway Immunologie, 9th Edition (Springer Spektrum, Berlin, Heidelberg). 10.1007/978-3-662-56004-4.
- Nassari, S., Blavet, C., Bonnin, M.A., Stricker, S., Duprez, D., and Fournier-Thibault, C. (2017). The chemokines CXCL12 and CXCL14 differentially regulate connective tissue markers during limb development. *Sci Rep* 7, 17279. 10.1038/s41598-017-17490-z.

- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43, e47-e47. 10.1093/nar/gkv007.
- Raman, D., Sobolik-Delmaire, T., and Richmond, A. (2011). Chemokines in health and disease. *Exp Cell Res* 317, 575-589. 10.1016/j.yexcr.2011.01.005.
- Ross, D., and Siegel, D. (2021). The diverse functionality of NQO1 and its roles in redox control. *Redox Biol* 41, 101950. 10.1016/j.redox.2021.101950.
- Schier, A.F. (2007). The maternal-zygotic transition: death and birth of RNAs. *Science* 316, 406-407. 10.1126/science.1140693.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102, 15545-15550. doi:10.1073/pnas.0506580102.
- Tian, F.J., Qin, C.M., Li, X.C., Wu, F., Liu, X.R., Xu, W.M., and Lin, Y. (2015). Decreased stathmin-1 expression inhibits trophoblast proliferation and invasion and is associated with recurrent miscarriage. *Am J Pathol* 185, 2709-2721. 10.1016/j.ajpath.2015.06.010.
- Torry, D.S., Leavenworth, J., Chang, M., Maheshwari, V., Groesch, K., Ball, E.R., and Torry, R.J. (2007). Angiogenesis in implantation. *J Assist Reprod Genet* 24, 303-315. 10.1007/s10815-007-9152-7.
- Wang, W., Dong, J., Chen, B., Du, J., Kuang, Y., Sun, X., Fu, J., Li, B., Mu, J., Zhang, Z., et al. (2020). Homozygous mutations in RECQL cause female infertility characterised by multiple pronuclei formation and early embryonic arrest. *J Med Genet* 57, 187-194. 10.1136/jmedgenet-2019-106379.
- Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res* 20, 804-815. 10.1101/gr.100594.109.
- Ziegler, G.C., Almos, P., McNeill, R.V., Jansch, C., and Lesch, K.P. (2020). Cellular effects and clinical implications of SLC2A3 copy number variation. *J Cell Physiol* 235, 9021-9036. 10.1002/jcp.29753.