

The role of tissue restricted antigens (TRAs) in heart development during embryogenesis

Team 4

7/18/2022

Contents

1 Abstract	1
2 Introduction	1
3 Materials	2
4 Methods	2
5 Results	4
6 Discussion	10
7 Appendix	12

1 Abstract

The embryogenesis is one of the most fascinating processes before life onsets. With our provided data set (*Mus musculus*) from the one cell stadium to a fully formed blastocyst before preimplantation we were able to evaluate the changes in genetic expression patterns necessary for specific tissue and thus organ development in depth. With the provided TRA database by Dr. Dinkelacker we were able to examine the connection of Tissue Restricted Antigens (TRAs), the establishment of self-tolerance due to the upregulation of TRAs in specific cell stages and tissues. Through the means of differential gene expression (DGE) and gene set enrichment analysis (GSEA) were able to achieve further results and thus contribute to the overall understanding of embryogenesis and tissue formation. A particularly interesting outcome of upregulated gene transcripts in the fourth to eighth cell stages, connected with the TRA data set implies an upregulated gene activation of hypoxia relevant genes. Furthermore, the most upregulated TRA transcripts could be found in the heart tissue, consequently we conclude that the upregulation of hypoxia genes and thus the sensitive regulation of oxygen levels in the early embryo are one of the most important events in the embryogenesis. Moreover, we could confirm the timing of zygote genome activation (ZGA) by clustering and the sum of the total gene expression changes, as well as the downregulation of histone activity right after discard of maternal transcripts. Working with a bovine data set (*Bos taurus*) for comparison, we concluded that the different development and regulation of metabolic pathways is due to a divergent time dependent cell differentiation and hence change of genetic expression. In summary, we could show that the expression of TRAs in the early stages of tissue development are essential and connected with the tissue formation and thus a vital cardiovascular system.

2 Introduction

Mouse development and organogenesis occur as early as compaction and the formation of a blastocyst before preimplantation of the mouse embryo. After the fertilization the most important stages of development are the two cell; four cell and eight cell stages as well as morula and blastocyst. A blastocyst already contains three different cell types: trophectoderm; epiblast and the endoderm (Kojima *et al.*, 2014). The blastula is roughly reached after 72 hours (Ciemerych and Sicinski, 2005). The very first two cycles after fertilization have a lengthened duration compared to the fourth and eighth cell stages. This is due to the chromatin remodeling and the decondensation of maternal and paternal chromatin in order to gain a functional nucleus (Ciemerych and Sicinski, 2005). The dynamic cell changes are controlled by so called D-cyclins, many transcription factors, and mainly performed by DNA- and Histone methylases and demethylases (Mihajlović and Bruce, 2017), (Sha *et al.*, 2019). From the one cell stage till the beginning of the fourth cell stage, the minor and the major zygote genome activation (ZGA) take place in *Mus musculus* (Mihajlović and Bruce, 2017). This implies that from now

on the development will be directed by the zygote's genome transcripts, while the maternal mRNA transcripts will be degraded, and thus the expression pattern will drastically change (AOKI, 2022). We will concentrate on the change of the gene expression between the fourth and eighth cell stages.

Since mammalian embryos develop under low oxygen conditions, managing these conditions and providing enough oxygen for morphogenesis, cell proliferation and tissue formation is essential. In order to prevent cell stress during embryogenesis, there are hypoxia sensitive genes which will be activated (Dunwoodie, 2009a). One of the most important factors for this matter is the Hypoxia Inducing Factor (HIF). HIF binds to the HIF-Responsive element, which is encoded by three genes. Whenever HIF is absent or epigenetically silenced, the morphogenesis of the heart is impaired. Especially affected is the formation of the endothelium in the cardiovascular muscles and the chamber formation of the heart. In order to develop a healthy cardiovascular system, HIF is essential (Krishnan *et al.*, 2008).

Tissue restricted antigens (TRAs) play a rather hidden and enigmatic role in embryonic development. With the aim of establishing functioning T cells, which recognize intruders such as pathogens via T cell receptors (TCRs), the T cells need to be trained (Alberts *et al.*, 2017). The positive and negative selection in the thymus allows T cells to recognize self-antigens which are displayed by MHC molecules on the cell surface. The expression and regulation are controlled by AIRE autoimmune regulator and Fezf2 (Monteleone-Cassiano *et al.*, 2022). The role of TRAs in the crucial stages of embryonic development is yet unknown, just as the immune suppressive impact of Fezf2 regulator in those cells (Takaba and Takayanagi, 2017). We worked with (Xie *et al.*, 2010) bovine and mouse data sets.

3 Materials

3.1 R and RStudio

This project was entirely done in R (R Core Team, 2022) version 4.2.0 (2022-04-22) and RStudio (RStudio Team, 2021) version 2021.09.0.

3.2 Affy Packages

The microarray chips used in the research of Xie *et al.* are Affymetrix GeneChips (mouse: MOE430A, bovine: bovine). In order to analyse these chips we installed the affy package (Gautier *et al.*, 2004), using Bioconductor. Affy is an R package that is used to analyse gene chips of the affymetrix type. Some of its many functions are to read in data and do quality control checks. The data are read in as .CEL files.

3.3 Brainarray and loading the Chip Description Files of mouse and bovine

The chip description files (CDF) of our two data sets (mouse and bovine) were downloaded using BrainArray (Dai *et al.*, 2005). BrainArray is an online data bank that gathers re-analyzed existing Affymetrix Genechip data “with updated probe set definitions”,(Dai *et al.*, 2005) to offer custom CDF files with better gene annotations and calculations.

3.4 Bioconductor

Bioconductor (Morgan, 2022) gathers different packages that are used in R, in order to widen the analysis of gene expression data sets. Most of the packages that we used in our project are installed through Bioconductor, this includes: limma (Phipson *et al.*, 2016), affy (Gautier *et al.*, 2004), VSN (Huber *et al.*, 2002), GSEA (Thomas *et al.*, 2011) and AnnotationDbi (Pagès *et al.*, 2022).

3.5 Tidyverse

Tidyverse is a collection of packages used for “data import, tidying, manipulation, visualization, and programming” (Wickham *et al.*, 2019). It is analogue to Bioconductor.

4 Methods

4.1 Quality Control (QC)

4.1.1 Mouse chips

After reading in the data, we examined the chips of the mouse data set to see if any of the chips have quality issues. The chips contain data for different cell stages: one-, two-, four-, eight-, morula-, blastocyst-cell stage. Each stage has three replicates (GSM456661 - GSM456678). First, we read the chips as images in order to see if they differ from the overall expression trend. We noticed three chips that seemed to differ. Chip GSM456666 is distinctly over-expressed and two

other, GSM456674 and GSM45675, were under-expressed. As an example, one chip is shown.

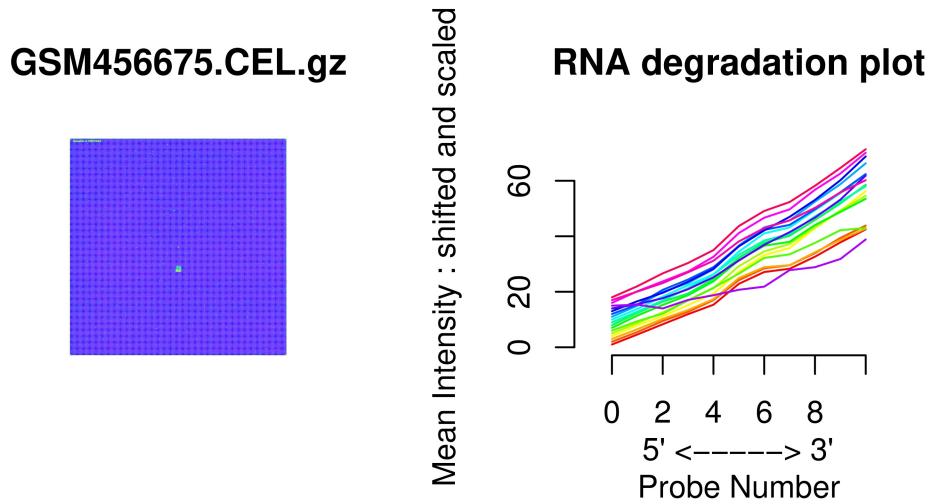


Figure 1: Single chip control and RNA degradation plot mouse

The second step in the quality control was done through an RNA degradation plot on the data set, that is shifted and scaled. The RNA degradation plot follows the degradation of the RNA by targeting the probe set in different regions of the selected transcript, the central section, the 3' and the 5'. This allows assessing the degradation rate of individual transcripts by examining the 3'/5' probe-set signal ratios. A good RNA degradation plot would show a steady upward trend with minimal crossing. In our case we can see that the orange line follows a different trend and crosses the others. This could be due to the three chips that have low quality. We are left with only one replicate of the morula cell stage, which could lead to inaccuracy. The quality issues can be caused by dye irregularities or binding of the targets to the probe. How well the bound probes to the targets are saturated and their distance to the 3' end of the transcript can also have an effect on the quality (Fasold and Binder, 2012).

4.1.2 Bovine

The same procedure was done for the bovine data set. Through the quality control of the bovine chips, we saw that the last chip had quality issues, as the dye showed a difference from the rest. This can be seen by plotting the RNA degradation plot of the 16 chips, as the line of GSM456642 (blastocyst, second replicate) crosses the rest of the lines. The plots can be seen in the appendix

4.2 Variance Stabilization Normalization

Variance stabilization normalization (VSN) is a statistical method, that is used for micro-arrays to reduce background noise, optical illusions and dye irregularities (Huber *et al.*, 2002). It is done through a log transformation in order to get a better concept of perception. It includes three main steps, the normalization, which is done through data calibration, the mean-variance-dependence of the model, and a variance stabilizing transformation (Huber *et al.*, 2002). For both data sets, the VSN was visualized using different plotting techniques.

4.2.1 Mean versus Standard deviation plot

The quality of the VSN can be visualized using the mean versus standard deviation plot (meanSDplot). The standard deviation should not have a strong correlation to the mean/variance and thus the red line of the median estimator should be horizontal (Dinkelacker, 2021).

4.2.2 Density Plot

The density plot is used to plot the density function against the log intensity of each chip. If the curves are well adjusted after the VSN, this would mean that the normalization was successful. We excluded the low quality chips mentioned above for the rest of our analysis.

4.3 Hierarchical Clustering

We proceeded to cluster the 15 mouse chips and created a distance matrix using the euclidean distance. After that the hierarchical clustering was done using the average linkage method. This was plotted and a dendrogram was formed. The bigger the height difference the more different the groups are.

4.4 Finding TRAs in our data set

Through the available data set provided by Dr. Dinkelacker (Dinkelacker, 2021), we were able to match the TRAs with our mouse data set using R.

4.5 Differential Gene Expression Analysis

The differential gene expression (DGE) analysis “refers to the analysis and interpretation of differences in abundance of gene transcripts within a transcriptome” (Conesa *et al.*, 2016). It is done in R Studio using the limma package in Bioconductor (Phipson *et al.*, 2016). Limma uses the linear model as an approach for the DGE, by simply forming a design matrix “which indicates in effect which RNA samples have been applied to each array” (Ritchie *et al.*, 2015) and a contrast matrix, where we define which objectives will be compared to each other. In our case the contrast matrix compares cell stages to each other and the design matrix designs a matrix that groups the chips by the cell stage they belong to. After that a linear model will be fit to our design matrix, and in the end the contrast matrix will be fit with the linear model. Limma uses the Bayes method in order to use probability to represent all uncertainty within the model. Here it moderates the standard errors of the estimated log-fold changes. It is calculated using the Bayesian Theorem, which is then used for hypothesis testing, in our case a t-test. The differential gene expression was performed for the mouse and bovine data sets, respectively.

4.6 Gene Set Enrichment Analysis

A gene set enrichment analysis (GSEA) is used to identify if a set of genes is enriched in expression. The analysis uses previous knowledge in order to see if a set of genes is related by shared criteria. This criteria can be a certain pathway or a functional classification (Thomas *et al.*, 2011). The GSEA is based on the results from the DGE that includes the results of the t-test and the p-values. Additionally we use The Molecular Signatures Database (MSigDB), a resource that contains annotated gene sets for our species and pathway analysis (Dolgalev, 2022). Using the annotation packages for *Mus musculus* and *Bos taurus* gives us information from different identifiers (Carlson, 2022). For us, the GSEA will help with enriching pathways that might play a role in tissue formation.

5 Results

5.1 VSN

5.1.1 Mouse

We normalized our data using VSN. To see if the VSN was successful we used the meanSD plot and the density probability against the log intensity. The mean versus standard deviation, shows a slight upward trend of the red line which represents the median estimator. If all chips were of good quality, the median estimator would show a horizontal line.

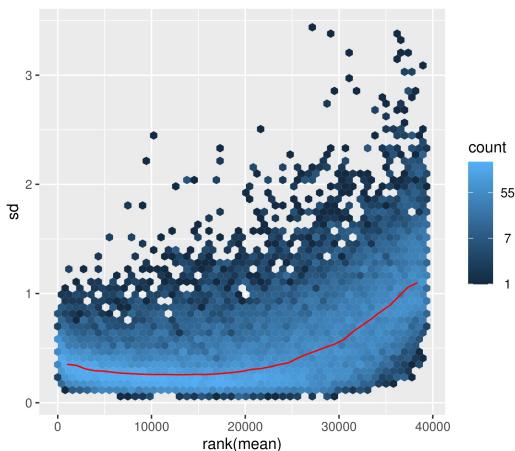


Figure 2: meanSD Plot of mouse

The density plot shows that all chips align with slight differences.

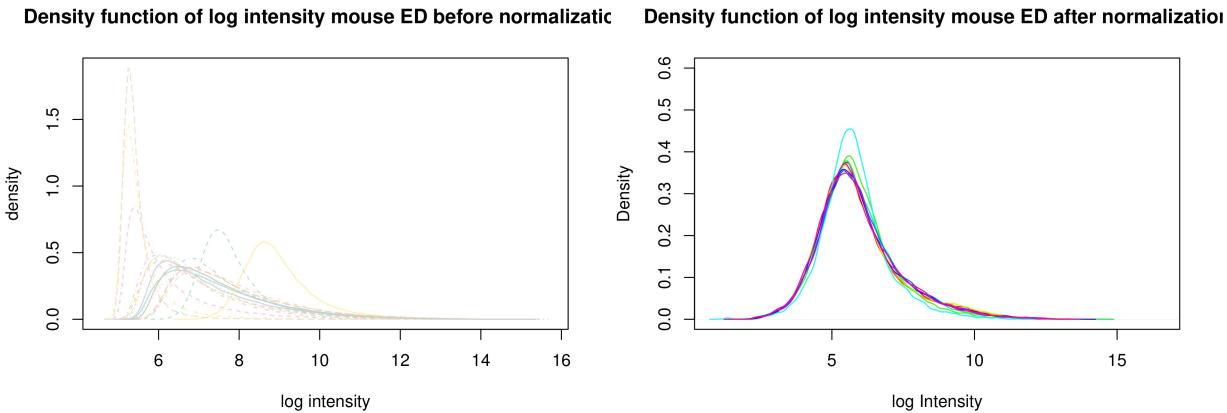


Figure 3: Density function of log intensity before and after normalization

Additionally, we plotted the VSN results using boxplots. The boxplot after VSN shows that the median of all chips align but has more outliers than the one before VSN.

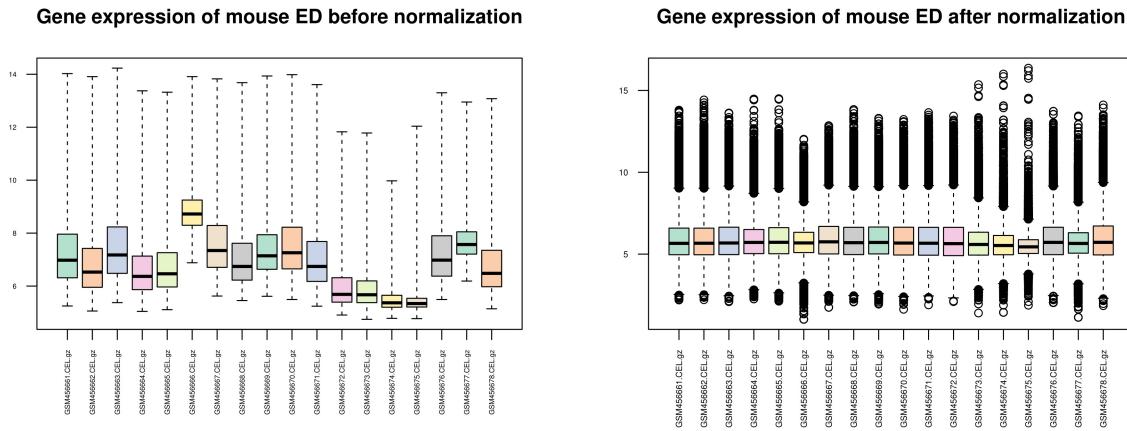


Figure 4: Gene expression of mouse embryonic development before and after normalization

After all those steps were done, we decided to discard three chips in total and took out the control transcripts.

5.1.2 Bovine

VSN of bovine data proved to be successful as seen in the density and boxplots. Analogue to the mouse data boxplots, we can see that due to the VSN the outliers are increased in the boxplot, but the median is equivalent in all cell stages. The density plot shows a successful VSN as the lines of the different cell stages align together. The meanSD plot as well as the box plots and the density function can be seen in the appendix.

5.2 Principal Component Analysis

We performed a principal component analysis, to see if the different replicates of the chips would show high correlation if they belonged to the same stage and to find out if we can reduce the dimension of our data set. Here, the variables are the different transcripts expressed (39281) in the 15 mouse samples. A scree plot was done in order to see how many principal components are needed, which are two as those explain around 50% of all the data variance. The PCA was done using the PCA function in R. Through the ggplot2 we can see that GSM456661 and GSM456662 (one cell stage) have excellent correlation. GSM456663 is close to them as well. This is also the case in GSM456667 and GSM456668 (four cell stage) but here, the third replicate is farther away Blastocyst stage (GSM456676 and GSM456678) are closer to each other which hints a good correlation. What one can notice is that the second replicate of the blastocyst (GSM456677) is a lot further away but closer to morula (GSM456673).

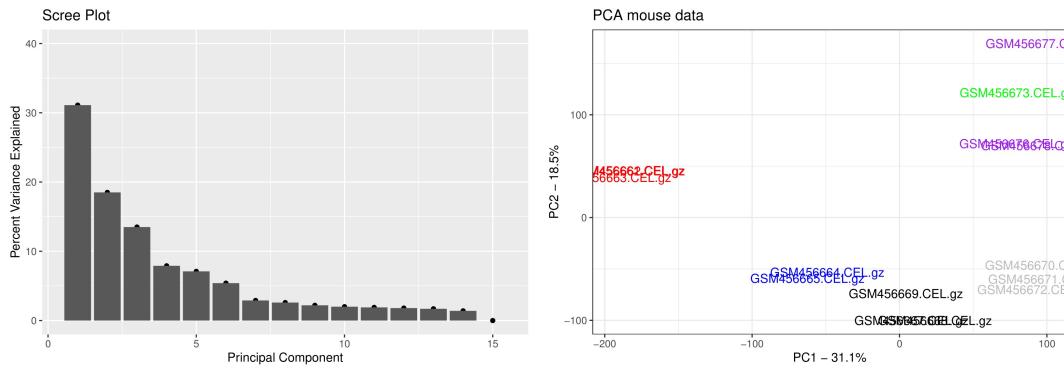


Figure 5: Scree plot and PCA of mouse data

5.3 Hierarchical Clustering

Hierarchical clustering analysis is based on an algorithm that calculates distances between the objects and forms clusters. Before we clustered we created a distance matrix using the euclidean distance. Based on the distance matrix we plotted a dendrogram in order to see which clusters differ the most from each other. This is based on the height of the branches. Based on the plot we can see that GSM456661 to GSM456663 differ significantly from the rest of the chips. The clusters with the biggest height difference is between GSM456661- GSM456663 and GSM456676-GSM456678.

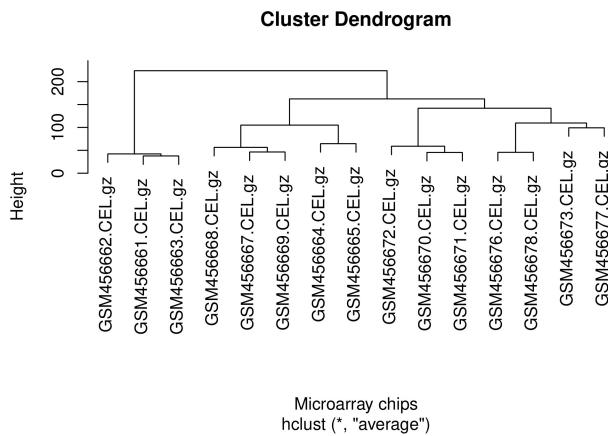


Figure 6: Clustering of mouse

5.4 Tissue Restricted Antigens in the mouse data set

Through R, we were able to match the mouse data set to the TRA dataset provided by Dr. Dinkelacker (Dinkelacker, 2021).

We plotted the frequency of the TRAs in the different tissues and found that the highest amount of TRAs occur in testis.

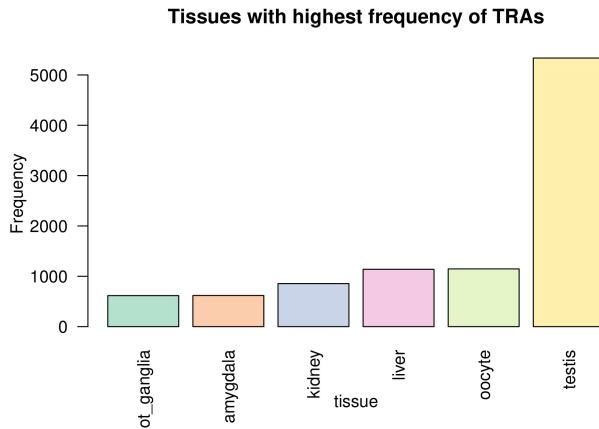


Figure 7: Tissues with highest frequency of TRAs

Upon matching, we found that our data set contains 6188 TRA transcripts which translate to 3255 genes. The TRAs were matched with their respective tissue to create a dataframe that contains the ensemble transcripts of the TRAs, their expression value in each microarray chip and their tissue.

5.5 Differential Gene Expression (DGE)

5.5.1 Mouse

Using limma, we performed a DGE in order to see if the deregulation of the TRAs vary in the different cell stages. The analysis performs a t-test on the expression set and gives us the result of the test with the p-value. The DGE simplifies this and gives us a matrix with three different values that correspond to the state of the TRA. 1 is assigned if the TRA is upregulated in the contrast between the two stages. 0 is assigned if the TRA has not significantly changed and -1 if it is underexpressed.

Using a volcano plot of the top ten most over- and underexpressed gene transcripts, we were able to show the statistical significance versus magnitude of fold change. The genes with a deviating fold change can be seen on either side of the expected fold change. The genes highlighted in red have a significant change of expression values. Here we could show that H1foo, a histone linker gene, is overall the most down regulated throughout all stages.

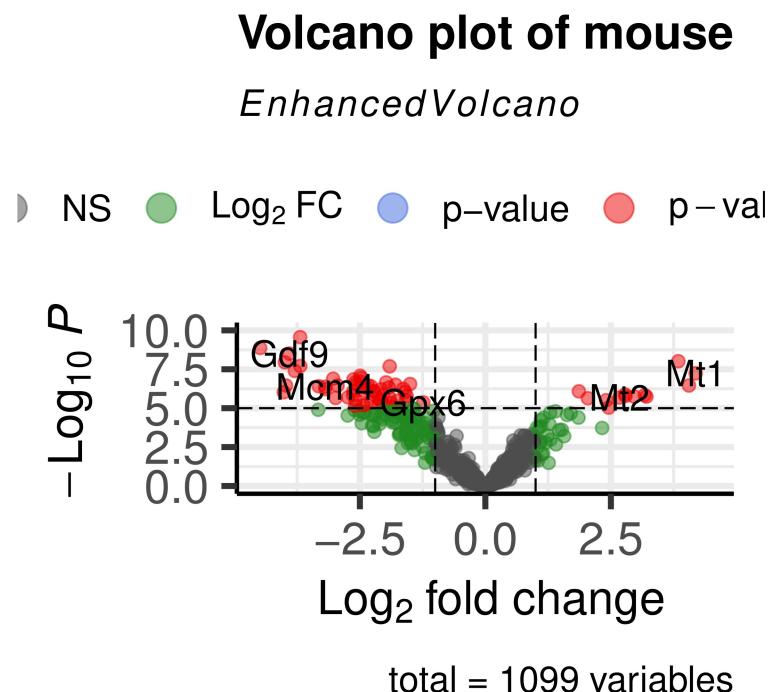


Figure 8: Volcano plot of mouse

Upon plotting the deregulated TRAs of the mouse using a bar plot, we saw that the most overexpressed TRAs are between the fourth and the eighth cell stages. The most amount of underexpressed TRAs are between the first and second cell stages.

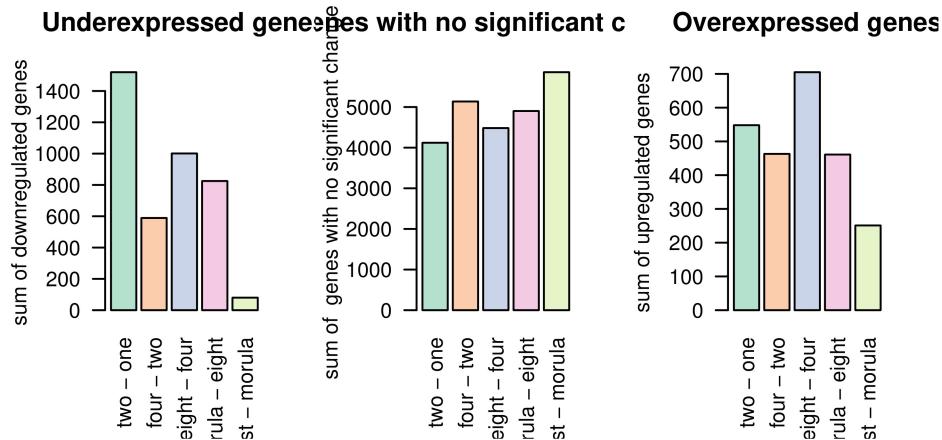


Figure 9: Gene expression in different cell stages mouse

In order to visualize the tissues with most expressed TRAs we used a treemap to visualize the total share of overexpressed TRAs in the tissues between the fourth and eight cells stages. We saw the the tissue with the most amount of overexpressed TRAs is the blastocyst followed by the heart.

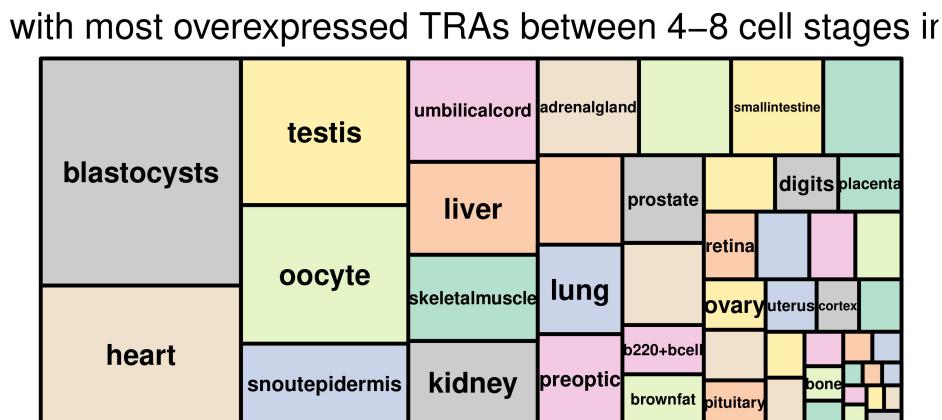


Figure 10: Tissues with most overexpressed TRAs in four to eight cell stage

As our main objective is to focus on the earliest biological function, we decided to focus on the heart, which is the earliest organ that to develop. In total we have 149 TRAs in the heart, 62 of those are overexpressed. We visualized the heart TRAs in the different cell stages using a box-plot .

Using a VENN diagram, we saw that there are three TRAs differentially expressed that can be found in cell stage two to four, four to eight, and eight to morula.

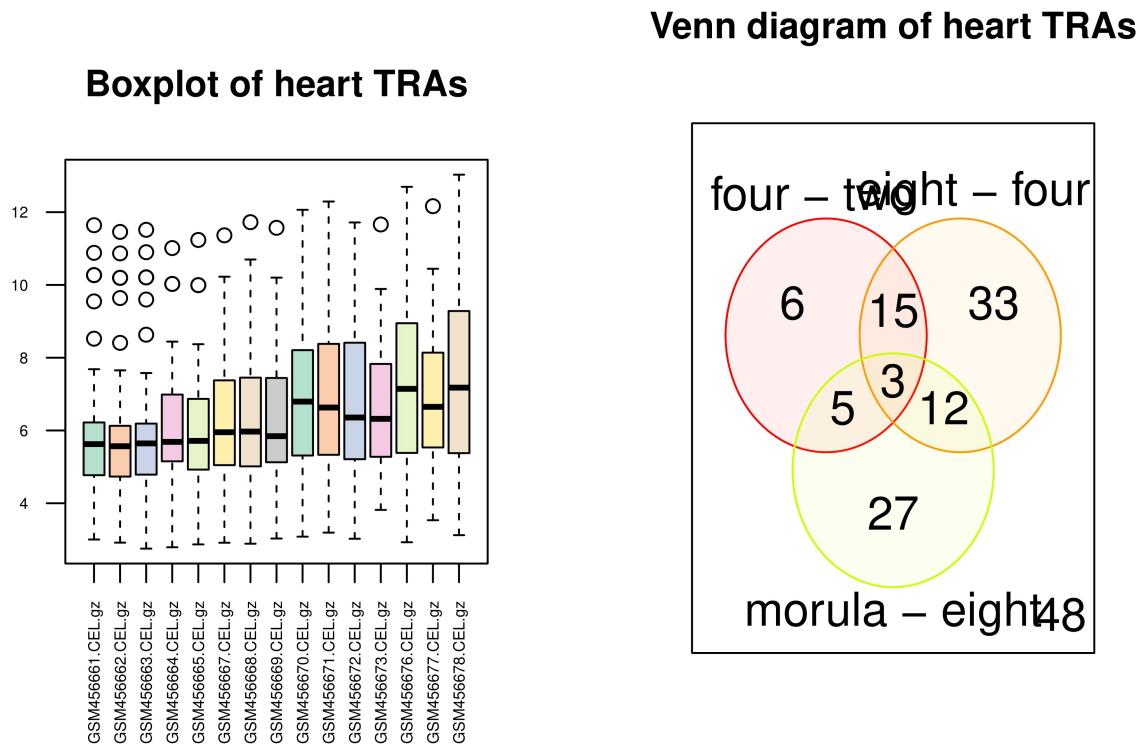


Figure 11: Boxplot and VENN diagram of heart TRAs in mouse

5.5.2 Bovine

After performing the DGE with our data set, we choose the Volcano Plot to present our results. The red dots are marking the significantly over or under expressed genes. We could only find 14 significantly differentially over or under expressed genes. Here we could specify a significant upregulation of the SERPINA14 gene and a downregulation of two genes which are FUS and FXR1.

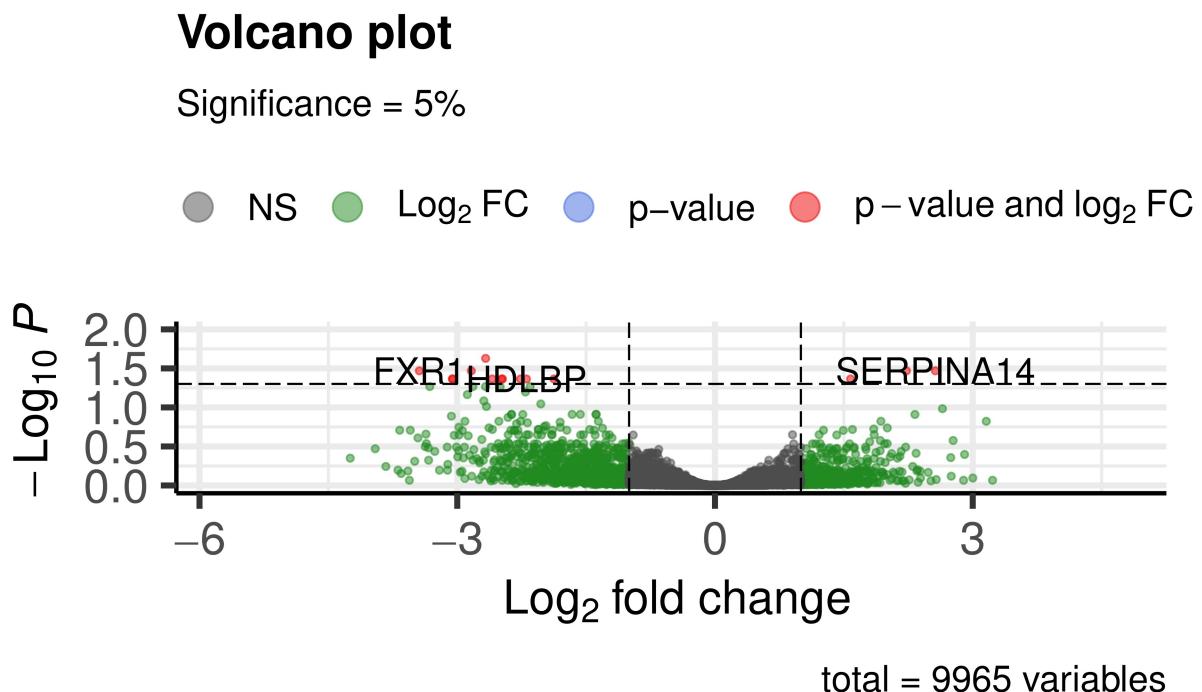


Figure 12: Volcano plot of bovine data

5.6 GSEA

5.6.1 Mouse

The gene set enrichment analysis was performed for the mouse data set which was prior matched with the TRA data. The GSEA was executed with the molecular data base MSigDB, we used the “Hallmark gene sets” collection available from the website. The most overexpressed pathway, with a 5% p-Value, which was found with GSEA is “Hallmark Hypoxia”. With these results we achieved an enrichment score (ES) of 0.38 and an associated p-Value of 4%. This implies a scientifically enriched and overexpressed pathway in our data set.

5.6.2 Bovine

The GSEA was also performed on our bovine data set without matched TRA data. The analysis revealed two pathways. “Epithelial Mesenchymal Transition” which was highly upregulated with an ES of 0.39 and an associated p-Value of 0.1%. “Oxidative Phosphorylation” is the second found pathway which is downregulated in our data with an ES of 0.46 and a similarly associated p-Value of 0.1%.

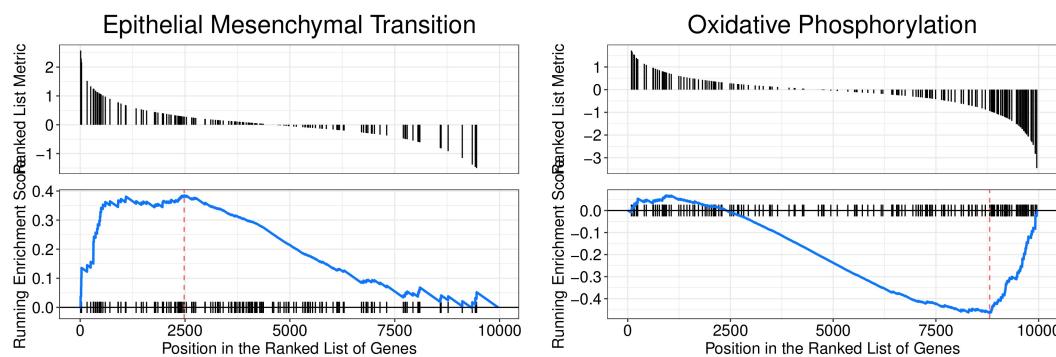


Figure 13: GSEA of mouse and bovine

6 Discussion

The performed PCA showed a high correlation between the tested chips, hence we conclude that the replicates of the chips belong to the same cell stage. With an overall variance explanation of 50% the PCA results are overall satisfying. A varied gene expression of GSM456663 would explain why this chip is not closely displayed in the ggplot with the other replicates of the same cell stage. This might also be a possible explanation for the second replicate of the blastocyst cell stage, which shows the same trend.

We performed a clustering analysis in order to group the chips that are most similar to each other, and to see how much the chips of different cell stages vary. Interestingly, we saw that the chips of the first cell stage differ completely from the rest, as they are clustered solitary. Biologically this can be explained, as the major zygote genome activation did not happen yet. In the first cell stage the cell has two gametes. The gametes need to form a new organism with totipotent cells in order to initiate embryogenesis. This happens through a process called maternal to zygotic transition, this allows the production of zygotic gene products instead of the maternal products (Schulz and Harrison, 2019). Another aspect that can be deduced from the clustering analysis is that GSM456677 (blastocyst, second replicate) is clustered with GSM456673 (morula first replicate). A reason for this can be that through the quality control, we excluded the second and third replicate of the morula cells due to their low quality. This could lead to the clustering of the morula replicate with the blastocyst replicate as the genetic material of the morula is characteristically closer to the blastocyst genetic material than to the eighth cell stage.

With the performed DGE across all stages we could see that the biggest changes of genetic expression patterns are between the fourth and eighth cell stage. For this matter we plotted the sum of the upregulated gene transcripts throughout the cell stages in figure. Specifically, the changes apply to overexpressed gene transcripts in this cell stage. This result aligned with the results by Xie *et al.* which describe that the biggest changes in genetic expression are to be found after ZGA in mouse after the second cell stage, since the maternal transcripts are removed, and the zygote's own genetic expression is activated. We did not find any significant results in the downregulation of the genes in the fourth and eighth cell stage. The biggest downregulation of gene transcripts ensues in the one to two cell stage, that is why we decided to exclude these results and to work henceforth with the upregulated gene transcripts in the cell stage we focused on. We plotted the sum of gene transcripts with no significant change but decided that this has no noteworthy meaning for our further analysis.

We carried out a volcano plot and could show that four out of the top ten underexpressed gene transcripts which we found belong to the gene H1foo. H1foo belongs to H1 Histone family. H1foo is involved in global epigenetic regulation and when active, will impair cell pluripotency and oocyte-specific linker histones. It is involved in formation of condensed chromatin, which cannot be read by DNA Polymerases (Schulz and Harrison, 2019). Since a downregulated histone activity advances the ZGA, this result aligns with the very early ZGA of the mouse in the first and second cell stage (Paul and Knott, 2014). Through the VENN diagram we saw that there are three transcripts that are expressed in the cell stages two to eight. The transcripts code for the gene Idh2, isocitrate dehydrogenase 2. Idh2 is found in the mitochondria, that catalyzes the oxidative decarboxylation of isocitrate to 2-oxoglutarate in the citrate cycle. It controls the mitochondrial redox balance, which is a first line of defense against the oxidative damage. If this balance is not held, oxidative stress can be enhanced which results in the increase of reactive oxygen species (ROS). This can lead to cardiac hypertrophy, which is a risk factor of heart failure (Ku *et al.*, 2015). The importance of the Idh2 gene explains why it is expressed in the cell stages two to eight.

The upregulated SERPINA14 in bovine encodes for uterine serpins, which are produced by the two pregnancy hormones progesterone and estrogen (Moffatt *et al.*, 1987). The role of serpines in the embryogenesis is yet not fully understood, however it many results point toward a control of a fully functioned immune response in the early pregnancy (Malathy *et al.*, 1990). FUS is an RNA and protein binding protein just as FXR1. These two genes were significantly downregulated. A downregulation of RNA binding proteins (RBPs) during the development of a embryo might seem contradictory because of the ongoing cell and tissue differentiation where huge amounts of gene transcripts are needed, however a downregulation of RBPs leads to a more strict control of inflammatory response during the pregnancy which prevents early abort of the embryo (Enerothe *et al.*, 1990; Khalaj *et al.*, 2016).

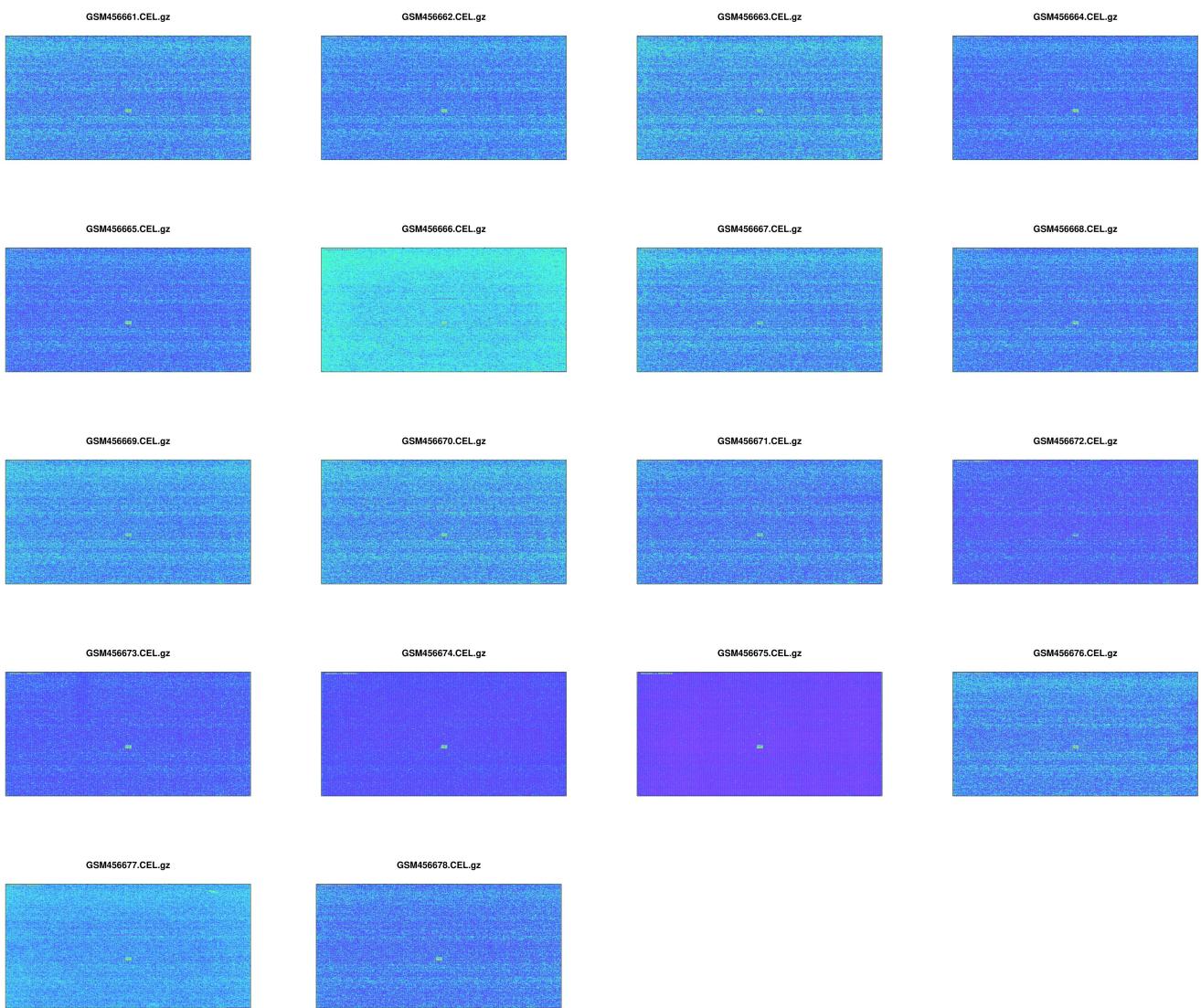
GSEA was carried out for our mouse data. The upregulated pathway which could be found is “Hallmark Hypoxia”. Hypoxia is the state of low oxygen in an organism. The embryo develops in a low oxygen environment. Those conditions are managed and regulated by the so called hypoxia inducible factors (HIF), which also act as a transcription factor together with HIF Response Element. The hypoxic state will upregulate these genes, which we can see in our data, since oxygen is a key provider to the placenta and cardiovascular development (Dunwoodie, 2009b). Since heart development is the first one to take place in an embryo, the supervision of HIF genes or hypoxia sensitive genes is essential. Especially because the process of heart tissue formation is inevitably connected with HIF and vascular endothelial growth factor (VEGF) (Patterson and Zhang, 2010). Aligning this with our previous results, that the transcripts for the heart tissue are the second most expressed in our data, it is consequential and logical that the most upregulated pathway we found is of importance for cardiovascular development. The metabolic pathway of hypoxia is upregulated in our data so we conclude it is not only essential but necessary for a healthy morphogenesis of the heart.

The GSEA which was performed for our bovine data, had two results. One of them being an upregulation of “Epithelial Mesenchymal Transition” and a downregulation of the “Oxidative Phosphorylation” pathway. During embryogenesis the cells need to be pluripotent in order to achieve an active proliferation state in which a normal morphogenesis can take place (Kim *et al.*, 2014). EMT induces polarized mesenchymal cells which enables cell to cell adhesion, pluripotency, and many other functions. Embryonic development requires these functions especially for stem cell differentiation (Kim *et al.*, 2017). The second result, which displayed a downregulation of oxidative phosphorylation occurs in mitochondrial development and in cancer cells, changing their metabolism. A downregulation of this pathway is especially useful for cancer cells (Santacatterina *et al.*, 2016).

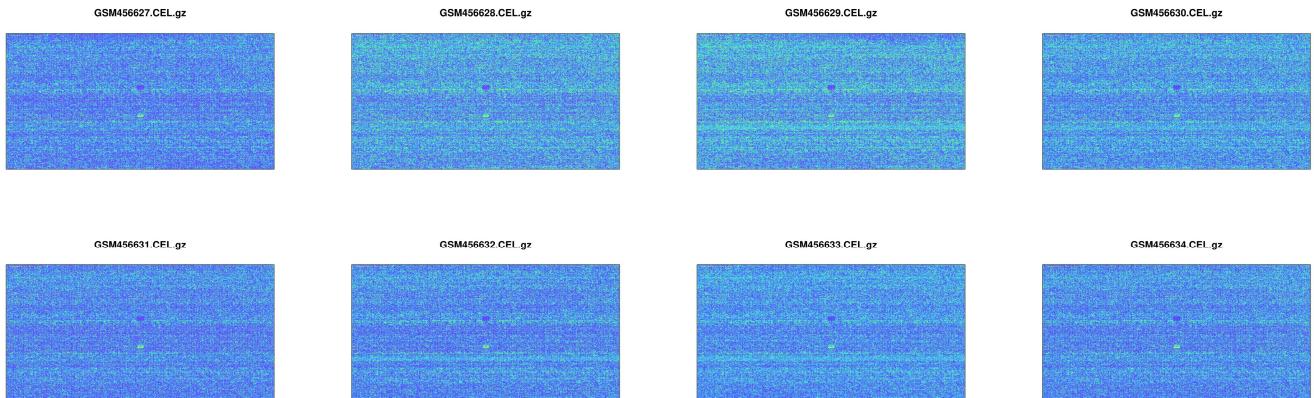
An interesting link between hypoxia and oxidative phosphorylation is that both are antagonists. As the oxidative phosphorylation is the primary metabolic pathway for ATP production and utilizes oxygen for the process, a hypoxic state can result in insufficient ATP production leading to cell death. Due to this phenomenon, a downregulation of the oxidative phosphorylation pathway is obligatory to limit the concentration of reactive oxygen species and oxygen depletion (Wheaton and Chandel, 2011), while in a hypoxic state. Inborn cardiomyopathy can be caused by high concentration of ROS during embryonic heart development (Zhao *et al.*, 2019).

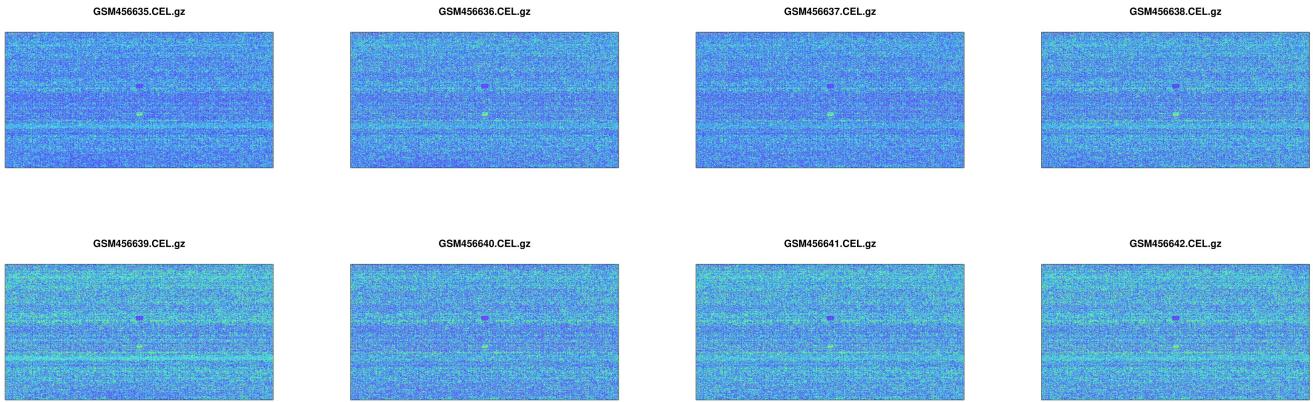
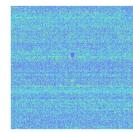
7 Appendix

7.1 Mouse Images

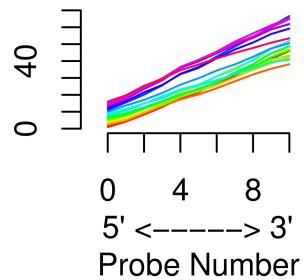


7.2 Bovine Images



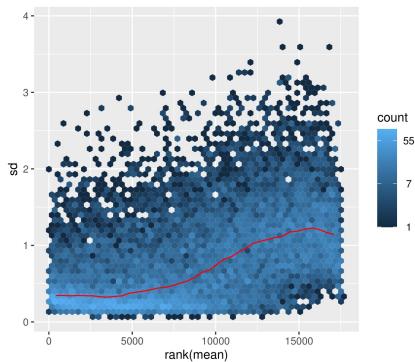
**GSM456642.CEL.gz****RNA degradation plot**

Mean Intensity : shifted and scale



```
## Quality control
bovine
```

```
## Results
```

**Figure 14:** meanSD plot of bovine

Gene expression bovine embryonic development before normalization: Gene expression bovine embryonic development after normalization

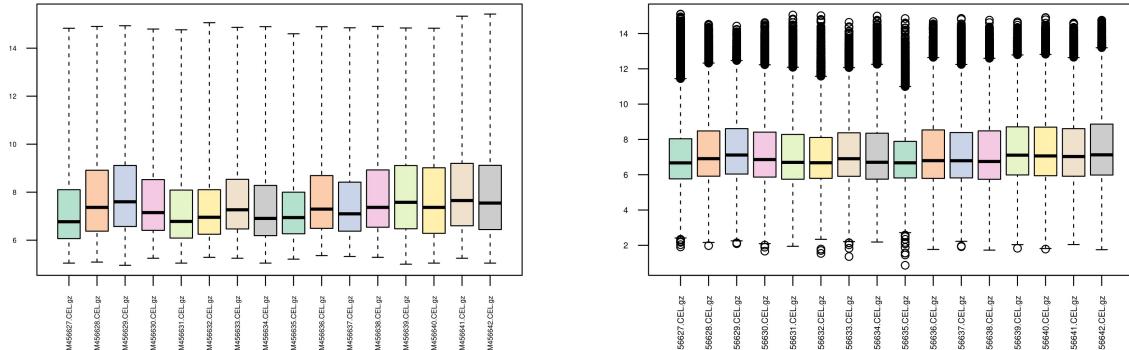


Figure 15: Gene expression bovine embryonic development before and after normalization

Density function of log intensity bovine ED before normalization: Density function of log intensity bovine ED after normalization

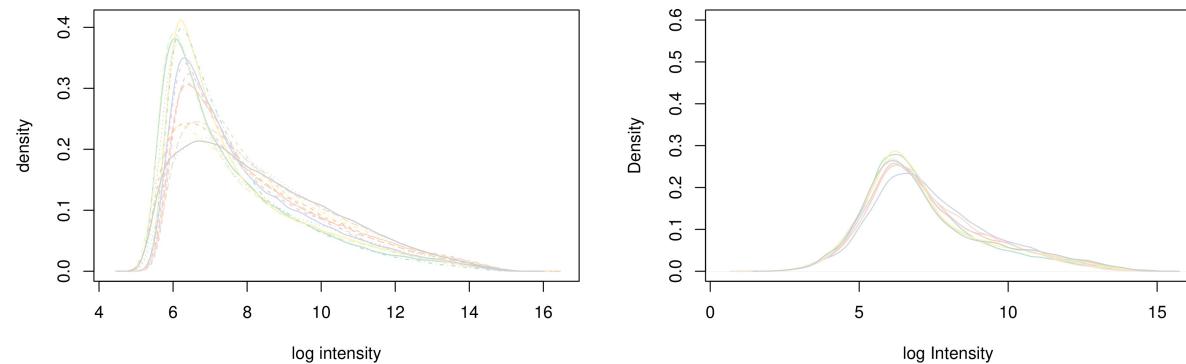


Figure 16: Density function of bovine before and after normalization

Session Info

```
## R version 4.2.0 (2022-04-22 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 22000)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Germany.utf8  LC_CTYPE=German_Germany.utf8
## [3] LC_MONETARY=German_Germany.utf8 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.utf8
##
## attached base packages:
## [1] stats4      stats       graphics    grDevices   utils       datasets   methods
## [8] base
##
## other attached packages:
## [1] tinytex_0.40          formatR_1.12
## [3] treemap_2.4-3         RColorBrewer_1.1-3
## [5] magrittr_2.0.3         org.Bt eg.db_3.15.0
## [7] org.Mm eg.db_3.15.0   msigdbr_7.5.1
## [9] clusterProfiler_4.4.4   EnhancedVolcano_1.14.0
```

```

## [11] ggrepel_0.9.1          icd.data_1.0
## [13] Rcpp_1.0.8.3           limma_3.52.0
## [15] knitr_1.39             forcats_0.5.1
## [17] stringr_1.4.0           purrr_0.3.4
## [19] readr_2.1.2              tidyverse_1.3.6
## [21] tibble_3.1.7             ggplot2_3.3.6
## [23] tidyverse_1.3.1           rstudioapi_0.13
## [25] pheatmap_1.0.12          hexbin_1.28.2
## [27] dplyr_1.0.9               cluster_2.1.3
## [29] bovinebtensprobe_25.0.0  bovinebtencdf_25.0.0
## [31] moe430ammensprobe_25.0.0 moe430ammencdf_25.0.0
## [33] AnnotationDbi_1.58.0     IRanges_2.30.0
## [35] S4Vectors_0.34.0          vsn_3.64.0
## [37] affy_1.74.0              Biobase_2.56.0
## [39] BiocGenerics_0.42.0

##
## loaded via a namespace (and not attached):
## [1] shadowtext_0.1.2          readxl_1.4.0          backports_1.4.1
## [4] fastmatch_1.1-3            plyr_1.8.7             igraph_1.3.2
## [7] lazyeval_0.2.2             splines_4.2.0          BiocParallel_1.30.3
## [10] gridBase_0.4-7             GenomeInfoDb_1.32.2   digest_0.6.29
## [13] yulab.utils_0.0.5          htmltools_0.5.2        GOSeqSim_2.22.0
## [16] viridis_0.6.2              G0.db_3.15.0           fansi_1.0.3
## [19] memoise_2.0.1              tzdb_0.3.0             Biostrings_2.64.0
## [22] graphlayouts_0.8.0         modelr_0.1.8           enrichplot_1.16.1
## [25] colorspace_2.0-3           blob_1.2.3             rvest_1.0.2
## [28] haven_2.5.0                xfun_0.30              crayon_1.5.1
## [31] RCurl_1.98-1.6             jsonlite_1.8.0         scatterpie_0.1.7
## [34] ape_5.6-2                  glue_1.6.2              polyclip_1.10-0
## [37] gtable_0.3.0                zlibbioc_1.42.0        XVector_0.36.0
## [40] scales_1.2.0                DOSE_3.22.0            DBI_1.1.3
## [43] xtable_1.8-4               viridisLite_0.4.0      gridGraphics_0.5-1
## [46] tidytree_0.3.9              bit_4.0.4              preprocessCore_1.58.0
## [49] httr_1.4.3                 fgsea_1.22.0           ellipsis_0.3.2
## [52] pkgconfig_2.0.3             farver_2.1.0           dbplyr_2.2.1
## [55] utf8_1.2.2                 labeling_0.4.2          later_1.3.0
## [58] ggplotify_0.1.0             tidyselect_1.1.2       rlang_1.0.2
## [61] reshape2_1.4.4              munsell_0.5.0          cellranger_1.1.0
## [64] tools_4.2.0                 cachem_1.0.6           downloader_0.4
## [67] cli_3.3.0                  generics_0.1.2          RSQLite_2.2.14
## [70] broom_1.0.0                evaluate_0.15          fastmap_1.1.0
## [73] yaml_2.3.5                 ggtree_3.4.0            babelgene_22.3
## [76] bit64_4.0.5                fs_1.5.2               tidygraph_1.2.1
## [79] KEGGREST_1.36.2             ggraph_2.0.5            nlme_3.1-157
## [82] mime_0.12                  aplot_0.1.6             DO.db_2.9
## [85] xml2_1.3.3                 compiler_4.2.0          png_0.1-7
## [88] affyio_1.66.0               reprex_2.0.1            treeio_1.20.0
## [91] tweenr_1.0.2                stringi_1.7.6           highr_0.9
## [94] lattice_0.20-45             Matrix_1.4-1            vctrs_0.4.1
## [97] pillar_1.7.0                lifecycle_1.0.1         BiocManager_1.30.18
## [100] data.table_1.14.2          bitops_1.0-7            httpuv_1.6.5
## [103] patchwork_1.1.1             qvalue_2.28.0           R6_2.5.1
## [106] promises_1.2.0.1            bookdown_0.27          gridExtra_2.3
## [109] codetools_0.2-18             MASS_7.3-56             assertthat_0.2.1
## [112] withr_2.5.0                GenomeInfoDbData_1.2.8 parallel_4.2.0
## [115] hms_1.1.1                  grid_4.2.0              ggfun_0.0.6
## [118] rmarkdown_2.14                ggforce_0.3.3           shiny_1.7.1
## [121] lubridate_1.8.0

```

- Alberts, B, Johnson, A, Lewis, J, Morgan, D, Raff, M, Roberts, K, and Walter, P (2017). Molecular biology of the cell, W.W. Norton & Company.
- AOKI, F (2022). Zygotic gene activation in mice: Profile and regulation. *Journal of Reproduction and Development* 68, 79–84.
- Carlson, M (2022). Org.mm.eg.db: Genome wide annotation for mouse.
- Ciemerych, MA, and Sicinski, P (2005). Cell cycle in mouse development. *Oncogene* 24, 2877–2898.
- Conesa, A et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol* 17.
- Dai, M et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research* 33(20), e175.
- Dinkelacker, M (2021). Chromosomal clustering of tissue restricted antigens, Heidelberg University Library.
- Dolgalev, I (2022). Msigdb: MSigDB gene sets for multiple organisms in a tidy data format.
- Dunwoodie, SL (2009a). The role of hypoxia in development of the mammalian embryo. *Developmental Cell* 17, 755–773.
- Dunwoodie, SL (2009b). The role of hypoxia in development of the mammalian embryo. *Dev Cell* 17, 755–773.
- Eneroth, M, Mandahl, N, Heim, S, Willén, H, Rydholm, A, Alberts, KA, and Mitelman, F (1990). Localization of the chromosomal breakpoints of the t(12;16) in liposarcoma to subbands 12q13.3 and 16p11.2. *Cancer Genet Cytogenet* 48, 101–107.
- Fasold, M, and Binder, H (2012). Estimating RNA-quality using GeneChip microarrays. *BMC Genomics* 13, 186.
- Gautier, L, Cope, L, Bolstad, BM, and Irizarry, RA (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 307–315.
- Huber, W, von Heydebreck, A, Sueltmann, 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–S104.
- Khalaj, K, Luna, RL, França, MER de, Oliveira, WH de, Peixoto, CA, and Tayade, C (2016). RNA binding protein, tristetraprolin in a murine model of recurrent pregnancy loss. *Oncotarget* 7, 72486–72502.
- Kim, D, Xing, T, Yang, Z, Dudek, R, Lu, Q, and Chen, Y-H (2017). Epithelial mesenchymal transition in embryonic development, tissue repair and cancer: A comprehensive overview. *J Clin Med* 7, 1.
- Kim, Y-S, Yi, B-R, Kim, N-H, and Choi, K-C (2014). Role of the epithelial-mesenchymal transition and its effects on embryonic stem cells. *Exp Mol Med* 46, e108.
- Kojima, Y, Tam, OH, and Tam, PPL (2014). Timing of developmental events in the early mouse embryo. *Seminars in Cell Developmental Biology* 34, 65–75.
- Krishnan, J, Ahuja, P, Bodenmann, S, Knapik, D, Perriard, E, Krek, W, and Perriard, J-C (2008). Essential role of developmentally activated hypoxia-inducible factor a for cardiac morphogenesis and function. *Circulation Research* 103, 1139–1146.
- Ku, HJ, Ahn, Y, Lee, JH, Park, KM, and Park, J-W (2015). IDH2 deficiency promotes mitochondrial dysfunction and cardiac hypertrophy in mice. *Free Radic Biol Med* 80, 84–92.
- Malathy, PV, Imakawa, K, Simmen, RC, and Roberts, RM (1990). Molecular cloning of the uteroferrin-associated protein, a major progesterone-induced serpin secreted by the porcine uterus, and the expression of its mRNA during pregnancy. *Mol Endocrinol* 4, 428–440.
- Mihajlović, AI, and Bruce, AW (2017). The first cell-fate decision of mouse preimplantation embryo development: Integrating cell position and polarity. *Open Biology* 7, 170210.
- Moffatt, J, Bazer, FW, Hansen, PJ, Chun, PW, and Michael Roberts, R (1987). Purification, secretion and immunocytochemical localization of the uterine milk proteins, major progesterone-induced proteins in uterine secretions of the Sheep1. *Biol Reprod* 36, 419–430.
- Monteleone-Cassiano, AC, Dernowsek, JA, Mascarenhas, RS, Assis, AF, Pitol, D, Moreira, NCS, Sakamoto-Hojo, ET, Issa, JPM, Donadi, EA, and Passos, GA (2022). The absence of the autoimmune regulator gene (AIRE) impairs the three-dimensional structure of medullary thymic epithelial cell spheroids. *BMC Molecular and Cell Biology* 23.
- Morgan, M (2022). BiocManager: Access the bioconductor project package repository.
- Pagès, H, Carlson, M, Falcon, S, and Li, N (2022). AnnotationDbi: Manipulation of SQLite-based annotations in bioconductor.
- Patterson, AJ, and Zhang, L (2010). Hypoxia and fetal heart development. *Curr Mol Med* 10, 653–666.
- Paul, S, and Knott, JG (2014). Epigenetic control of cell fate in mouse blastocysts: The role of covalent histone modifications and chromatin remodeling. *Mol Reprod Dev* 81, 171–182.
- Phipson, B, Lee, S, Majewski, IJ, Alexander, WS, and Smyth, GK (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann Appl Stat* 10, 946–963.
- R Core Team (2022). R: A language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing.
- Ritchie, ME, Phipson, B, Wu, D, Hu, Y, Law, CW, Shi, W, and Smyth, GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43, e47.
- RStudio Team (2021). RStudio: Integrated development environment for r, Boston, MA: RStudio, PBC.
- Santacatterina, F et al. (2016). Down-regulation of oxidative phosphorylation in the liver by expression of the ATPase

- inhibitory factor 1 induces a tumor-promoter metabolic state. *Oncotarget* 7, 490–508.
- Schulz, KN, and Harrison, MM (2019). Mechanisms regulating zygotic genome activation. *Nat Rev Genet* 20, 221–234.
- Sha, Q-Q, Zhu, Y-Z, Li, S, Jiang, Y, Chen, L, Sun, X-H, Shen, L, Ou, X-H, and Fan, H-Y (2019). Characterization of zygotic genome activation-dependent maternal mRNA clearance in mouse. *Nucleic Acids Research* 48, 879–894.
- Takaba, H, and Takayanagi, H (2017). The mechanisms of t cell selection in the thymus. *Trends in Immunology* 38, 805–816.
- Thomas, MA, Yang, L, Carter, BJ, and Klaper, RD (2011). Gene set enrichment analysis of microarray data from pimephales promelas (rafinesque), a non-mammalian model organism. *BMC Genomics* 12, 66.
- Wheaton, WW, and Chandel, NS (2011). Hypoxia. 2. Hypoxia regulates cellular metabolism. *Am J Physiol Cell Physiol* 300, C385–93.
- Wickham, H et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software* 4, 1686.
- Xie, D, Chen, C-C, Ptaszek, LM, Xiao, S, Cao, X, Fang, F, Ng, HH, Lewin, HA, Cowan, C, and Zhong, S (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Research* 20, 804–815.
- Zhao, Q, Sun, Q, Zhou, L, Liu, K, and Jiao, K (2019). Complex regulation of mitochondrial function during cardiac development. *J Am Heart Assoc* 8, e012731.