Contents

Final Report - Datascience MoBi 2022

Gaining insight on human early organogenesis through TRA expression analysis

Topic 04 - Team 01

Paul Christmann, Joshua Eigenmann,
David Jewanski, Verena Merke

Date: June 17, 2022

Supervisor: Dr. Maria Dinkelacker

Tutor: Ian Fichtner

Institute of Pharmacy and Molecular Biotechnology

Heidelberg University

---

1 Abstract

A single cell turns into a complex multicellular organism during embryogenesis. While the morphological steps are mostly understood, the role of molecular pathways as well as transcriptional regulation in embryonic development is still a topic of current research. Tissue restricted antigens (TRAs) may be a key to understand the connection gene expression and morphological consequence. Since the expression of TRA is specific to their relative tissue, we could draw conclusions about organ development from TRA expression patterns between week 4 and 9 of embryogenesis. Using microarray data from Yi H *et al.* (2010), we created a dataset of differentially expressed transcripts, including genes not found by Yi H *et al.*, by using limma analysis. Based on this data, we could successfully link developmental steps to TRA expression patterns for some of the analyzed organs. Furthermore, we found individual genes associated with processes in neurogenesis. Finally, we attempted without success to define specific TRAs as biomarkers for organogenesis.

2 Introduction

2.1 TRAs as a tool to gain insight into embryonic development

Tissues restricted antigens (TRAs) are originally a concept from immunology. There, they are sets of genes for auto-antigens that enable a negative selection in the thymus, which prevents autoimmune reactions. TRAs are ordered in clusters and controlled by autoimmune regulator in medullary epithelial cells and represent the diversity of antigens in the different tissues of an organism (Dinkelacker, 2019; Murphy & Weaver, 2018). For our research, we use a broader definition, that TRAs are genes, which are expressed more than 5 times the median gene expression in less than 5 different tissues (Dinkelacker, 2019). In contrast to housekeeping genes, this makes TRAs highly specific for individual tissues. Therefore, the existence and functionally of an organ should be correlated with the expression of its associated TRAs. We will use this connection to gain insight into organ development based on gene expression data between week 4 and 9 of embryogenesis.

2.2 Embryonic development during the observed timeframe

In order to compare the results based on expression patterns, it is essential to give an overview about organogenesis during this time. We will summarize the developments in following organs: the liver, testes, spleen, heart, stomach, skin, skeletal muscles and brain.
First, the liver sprout begins to form in week 3 after gestation. From week 4, it develops hepatocyte precursors and is innervated by veins. Between week 5 to 9 the production of gallic acid starts. Glycogen granules develop by week 8 and glycogen synthesis starts in the following week (Deutsch, 2013).
The testes initially develop as non sex-specific gonads. At week 5, the first germ cells appear in the gonades. The gender-specific development into ovaries and testes starts at week 7 (Benninghoff, 1993).
The spleen appears at week 6 of embryonic development. Blood vessels in the organ develop from week 8 to 9 (James & Jones 1983). The spleen plays a major role in the human immune system. To that end, B-lymphocytes are present within this tissue from week 12, while T-cells can be found not earlier than week 14. Previously, these cells develop in the liver starting week 9 and in the thymus from week 7 respectively (Hayward, 1983).
At week 3, the heart consists only of a preliminary tube. From then on, it undergoes extensive growth and development with cambers and ventricles forming. The fundamental layout is already present by week 5. Then, further remodeling takes place until around week 7, at which point the major steps are already completed (Ulfig, 2009; Hikspoors *et al.*, 2022).
The stomach develops from the foregut. The primitive gut divides into foregut, midgut and hindgut by week 4. The stomach is first visible at the end of week 4 (Kluth *et al.*, 2013). Gastric pits form by week 8, while essential cell types like enteroendocrine cells and mucous cells appear between the 10th and 15th week. Stomach acid is only secreted from week 32 onward (Esrefoglu *et al.*, 2017).
Skin development starts immediately after gastrulation at week 3. The ectoderm further develops to the nervous system and skin epithelium. There, the epidermal differentiation is illustrated through the expression of keratin genes. Adhered cells (periderm) create a protective layer for the ectoderm during weeks 4 to 8 (Hu *et al.*, 2018).
The skeletal muscles from mesoderm first in form of myoblasts that later, between week 10 and 13, fuse to form myotubes and then differentiated muscle fibers. The proteins necessary for muscle formation appear the earliest at 7 weeks, with more being expressed from week 9 and 10. Muscle fibers only form from week 15 onwards (Romero *et al.*, 2013).
With neurulation happens around week 4 and the major parts of the brain are already visible by week 9. Between theses stages, characteristic steps of neuronal development such as neuronal proliferation and differentiation starting at week 4 as well as neuronal migration and synapse formation starting at week 9 take place (National Research Council and Institute of Medicine, 2009; Müller & Hassel, 2018). These processes are influenced through signals provided for instance by chemokines.

2.3 Chemokines and brain development

Chemokines are a group of small proteins, acting as chemoattractors on effector cells. They are classified in 4 groups labelled as alpha to delta, depending on the position of their first cysteines (Cys). In the alpha group (CXC), they are separated by a single aminoacid. In the beta group (CC), Cys are next to each other. In the gamma group (C), only one Cys is present. In the delta group (CX3C), they are separated by three aminoacids. (Yusuf *et al.*, 2005). Chemokines induce cell migration by binding to their respective receptors. It is known, that the CXCL12/CXCR4 signalling pathway plays an important role in the neuronal cell migration during embryonic development (Tiveron & Cremer 2008).

3 Methods

3.1 Programming language and Libraries

We used the programming language R version 4.2.0 and its IDE RStudio to draw statistical conclusions. We installed the library packages (Table 1) from CRAN and bioconductor, which an open software library statistical genomics. Annotation packages for microarrays were provided by brainarray.

Table 1: *All libraries used for the code of this report, libraries installed from CRAN, Bioconductor and brainarray*

Library	Version	Library	Version	Library	Version	Library	Version
affy	1.74	AnnotationDbi	1.58	biomaRt	2.52	cluster	2.1.3
clusterProfiler	4.4.4	cowplot	1.1.1	dplyr	1.0.9	enrichplot	1.16.1
factoextra	1.0.7	ggbiplot	0.55	ggforce	0.3.3	GGally	2.1.2
ggplot2	3.3.6	ggplotify	0.1.0	ggpubr	0.4.0	ggrepel	0.9.1
ggsci	2.9	ggupset	0.3.0	grid	4.2.0	gridExtra	2.3
gt	0.6.0	gtExtras	0.4.1	hexbin	1.28.2	hgu133plus2hsenstcdf	25.0
hgu133plus2hsenstprobe	25.0	igraph	1.3.2	kableExtra	1.3.4	limma	3.52
magick	2.7.3	magrittr	2.0.3	org.Hs.eg.db	3.15	pheatmap	1.0.12
png	0.7	Rcpp	1.0.9	RCurl	1.98	readxl	1.4
rentrez	1.2.3	Rfssa	2.0.1	stringr	1.4	svglite	2.1
tidyverse	1.3.1	treemapify	2.5.5	VennDiagram	1.7.3	viridis	0.6.2
vsn	3.64	webshot	0.5.3	XML	3.99		

3.2 Dataset

We obtained the data set from Yi H *et al.* (2010). It contains human embryonic data, which covers every week between the 4th and 9th week with three replica at each point in time, hence the data from 18 embryos were acquired. The timezone covers the Carnegie stages 10-23, finishing the process of embryogenesis and organogenesis. This peroid of embryogenesis is highly regulated with considerable differential gene expression. Overall, the data set suits the requirements for our purpose.

3.2.1 Affymetrix U133 plus 2.0 human GeneChip array and importing its data

The data was generated using Affymetrix U133 plus 2.0 human GeneChip arrays, coted slides with matrices for screening purposes. The HG-U133 Plus 2.0 allows the detection of about 50,000 transcripts and include 62 control transcripts.

We downloaded the raw data from the Gene Expression Omnibus with the Accession Number of GSE15744. We imported it with the library package *affy*, which allows more manageable data analysis and manipulation of microarray intensity values.
To access the data remotely, we uploaded it to the cloud-based repository hosting service github.

3.2.2 Quality control of the suface images and of RNA Degradation

As shown in Fig. 1A the surface of the microarrays show no spatial artefacts, fingerprints, irregular dye or stripes. Some differences in overall brightness are visible but marginal.

Furthermore, we tested for low RNA quality chips. Coated matrices degrade under unfavorable conditions, which negatively affects raw intensities (Fasold & Binder 2013). By plotting the RNA degradation for 3'–5' strand, we compare the different chips (Fig. 1B) and verify the overall chip quality for data analysis.

3.3 Normalising the data set

Intensity values of different chips are affected by statistical variance and random fluctuation. To access the biological relevant variation, the raw data is normalised. We chose the vsn rma normalization with its library *vsn* according to Huber *et al.* (2002). This library is designed to process microarray intensity values. It calibrates data and applies *generalized log*-transformation, which is an adjusted natural logarithm and preserves statistical significance.

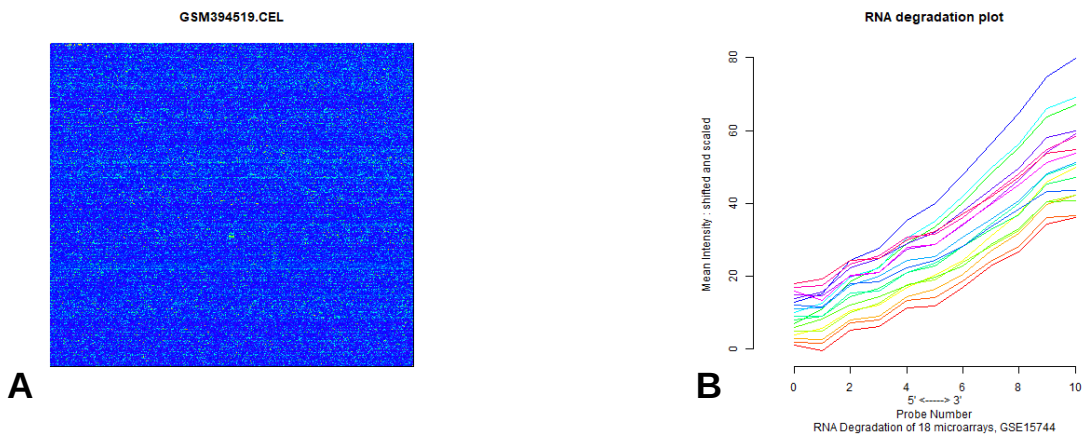


Figure 1: Quality control: Selected surface image of a microarray shows no damage or artefacts and RNA degradation plot shows slight irregularities and verifies the data. ***A****: The microarray inspection shows no irregularities and every chip is accepted for further data analysis.* ***B****: Some crossing lines can be seen, especally the microarray GSM394519. We decided that the inconsistencies are minor though, and kept all microarrays to avoid the loss of potentionally relevant data.*

3.3.1 Quality control of the vsn normalization

To verify the transformed data intensity values, some test can be performed (supplementary Fig. 7). After the normalization, the rank of the mean of the intensity values and their standard deviation should not correlate. To control this, we plotted the rank of the mean against the standard deviation, which resulted in a reasonable horizontal line indicated in red (supplementary Fig. 7A). To further control the normalization we visualized the intensity values. We used boxplots to compare each of the 18 microarray separately by its mean, median and variance (supplementary Fig. 7B). We deduced that all arrays fit.. The second option gives us the ability to zoom in even further. The intensity levels of three replica should be the same, since they were taken at the same time. We can use scatterplots to compare single intensity levels. With one of the replica applied on the x and y axis respectively, we should see a scatterplot following the linear function y = x since the same transcript should show the same intensity in both replica (supplementary Fig. 7C).

3.4 Analytical methods

1. **Annotation**: To make sense out of the intensity values they need to be associated to common data with known properties. We applied the data frame *ensembl_103.txt* provided by Dr. Dinkelacker, to annotate our data and yield the appropriate transcript ID for the Probe ID of the microarray. To annotate for TRAs, we applied another data frame by Dr Dinkelacker called *tra.2017.human.gtex.5x.table.tsv*.

2. **Limma** The *limma* package determines among many other things the changes of gene expression over time in intensity values of microarrays. It facilitates advanced statistical algorithms to calculate the necessary coefficients of a linear model for every intensity value in the data set. It uses information borrowing, quantitative weighting, variance modelling and data preprocessing, while not subset the data (Ritchie. *et al.* 2015). Because the linear model was casted on every intensity value, statistical tests called Empirical Bayes can determine differential expressed transcripts via t-statistics and their associated p-values.

3. **Over representation analysis** The statistical method over-representation analysis determines among other thing the over represented function of genes with associated transcripts in a subset of a mother data set with annotated transcripts and known functions. Categories for functions can be accessed via gene ontology.

4 Results

4.1 Limma analysis

To filter our data for biological interesting data, we performed *limma* analysis to extract differentially expressed transcripts. Our threshold for significance is an Benjamini-Hochberg adjusted p-value of 0.01 or below. We found changes of gene

expression in 1,814 transcripts. For all of those, we found at least 40 differentially expressed transcripts within our dataset. (Fig. 3). These numbers are sufficient for further analysis within individual tissues.
Volcano plot for differentially expressed transkripts between week 4 and 9 was made to evaluate limma analysis. Genes are categorized as up-regulated, down-regulated or not differentially expressed (Supplementary: Fig. 8)). We further used genes between all weeks with an adjusted p-value smaller than 0.01. All results will be based on this dataset with differentially expressed transcripts.

4.2 Comparison of differential expressed transcripts between the original paper and our method

We acquired the data from the original paper from the supplemental materials. We imported it into a data frame annotated it with the use of our data set because the both have the same probe set IDs from the same microarrays.

4.2.1 Venn diagram

To determine the number of transcripts which overlap between the data of the paper and our data, we plotted a Venn diagram (Fig. 14). The intersection showed a number of transcripts worthy for further analysis.

4.2.2 Verifing the trends postulated in the paper with our data

In contrast to our method using limma, the authors of the original paper used *One-way analysis of variance* with a p-value of 0.05 to determine the differential expressed transcripts (Yi H *et. al*, 2010). In the paper they provided an annotation of transcripts that were regulated *up*, *down* or showed an *arch*. We used k-means clustering to determine, if we see these trends in our data as well (Fig. 2).
Fig.2A and B shows great accordance with the annotated pattern. We repeated both plots with our differentially expressed transcripts and the results were highly similar. The figure can be found on our github (Report/Comparison with paper differentially expressed.png).
For Fig. 2C cluster 1, 6, 7, 8 clearly show an *arch* regulated pattern, but for clusters 4 and 5 the data rather appears to be *down* regulated, or *up* regulated in the case of cluster 10. And the final clusters 2, 3 and 9 with a total of 156 transcripts are not differentially expressed at all. When we cluster our differentially expressed transcripts annotated with the *arch* pattern (Fig. 2D), we receive only 10% of the amount of transcripts which are clearly clusters.
Additionally, Fig. 2E reveals, that there clearly *up* and *down* regulated genes of 646 transcripts or about 10%, that are missed by the method used by the authors of the paper.

4.3 TRAs can infer a basic timeline of organ development

4.3.1 Differentially expressed transcripts can be linked to all analyzed tissues

The TRA dataset covers 53 distinct tissues (Fig. 3). The minimum of differentially expressed transcripts in tissue were 46 stomach-linked transcripts, the highest reached 837 TRAs for the testes.
Nonetheless, there is a significant overlap between the TRAs associated with different tissues, especially as each transcript is on average linked to 7.4 different sub-tissues or tissues. This overlap is illustrated by supplementary Fig. 11. Furthermore, a detailed heatmap for the shared TRAs between tissues is shown in the supplementary Fig. 9.

4.3.2 The expression of all TRAs associated with a tissue cannot be used to infer organ development

In this research, we attempt to draw conclusions about the developmental state of a tissue based on the expression of genes associated with it alone. For all differencially expressed transcripts associated with one tissue, we analyzed the share of transcripts above a certain expression level over time, as shown in Fig. 4A. Furthermore, we observed trends within the median expression of these trascnrips (Fig. 4B). Since both metrics only depicted miniscule changes, we hypothesized that distinct, counteracting trends in expression exist within one tissue. Thus, k-means clustering was used to determine groups of TRAs with similar expression patterns. For each of these clusters, the median expression was plotted as shown in Fig. 4C. The clustering graphs for all tissues can be found on our github repository.

For many tissues, as shown here exemplary with the spleen, the clustering revealed two or more clusters that could be characterized as either an upregulation or a downregulation. In order to determine the indications for organ development, we analyzed the functions of the transcripts belonging to the two clusters.

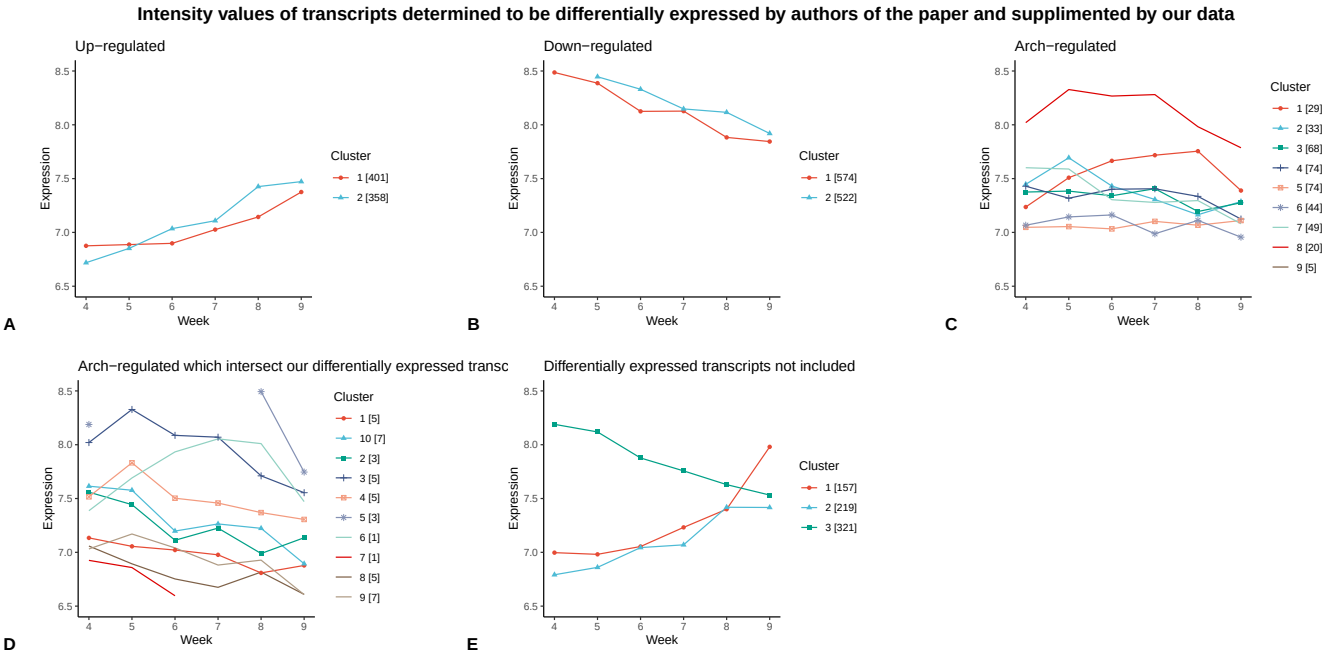


Figure 2: Clustering of the common transcripts reveals similar trends and but our additional data shows missing differentially expressed transcripts. *We chose data with transcript IDs both in the data set of the original paper and in our TRA data. Additionally they had to be annotated as* up*, down,* or *arch* by the authors of the paper. The* up *and* down *regulated transcript-data (**A** and **B**) follow the same pattern as postulated in the paper. For the* arch *regulated transcript-data (**C**) we see clusters, with same matching the* arch *pattern and some who are rather undetermined. Of 396 transcripts only 39 of our differentially expressed transcripts by* limma *analysis share this* arch *property (**D**). Here we clearly see differential gene expression. In **E** we plotted data of our differentially expressed transcripts. There are* up *and* down *regulated genes determined by our* limma *analysis, that are not included in the data set of the authors of the original paper*

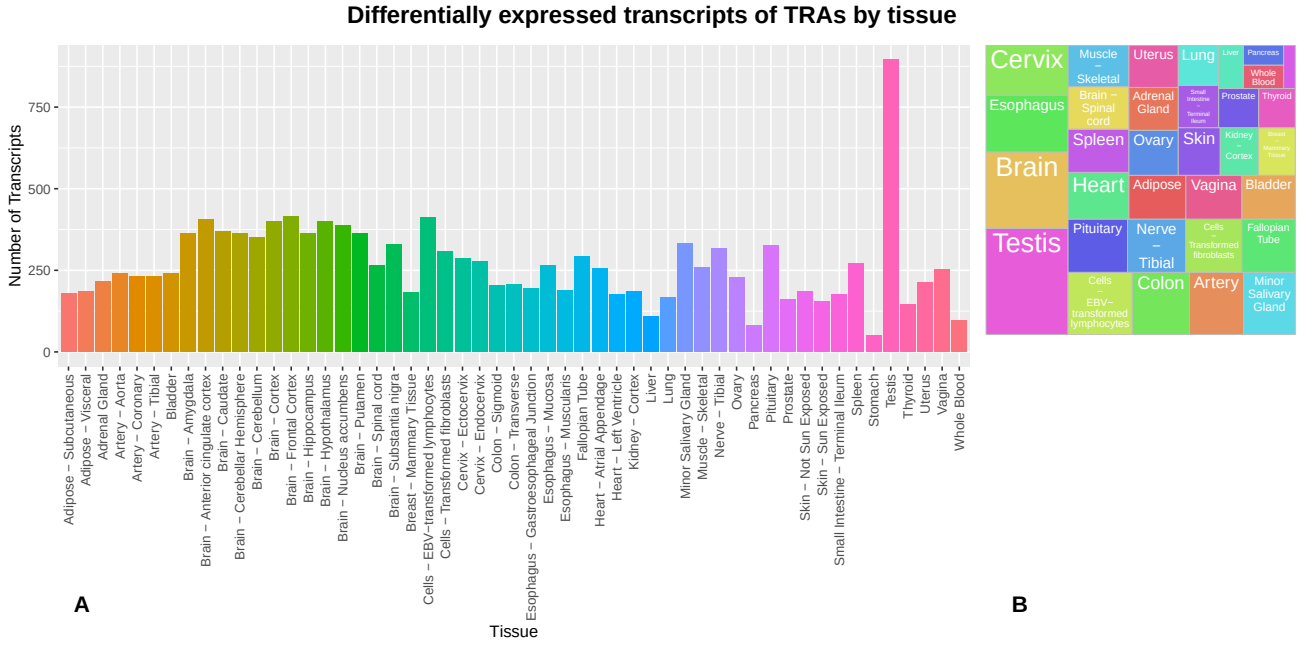


Figure 3: *A. The number of transcripts associated with each tissue, including subtissues, is displayed. B. The plot shows the share of TRAs associated with each tissue. Subtissues are subsumed under their main tissue.*

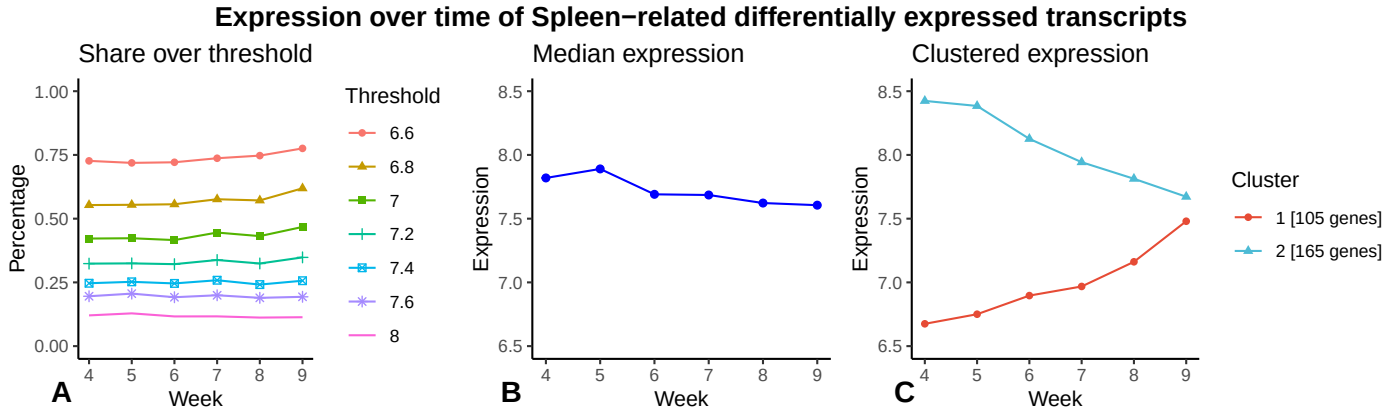


Figure 4: *A. For different expression thresholds the share of differentially expressed transcripts with higher expressions than the threshold is depicted. B. The median expression for all spleen-associated TRAs is shown for each point in time. C. For k-means clustring, the silhouette score determined an optimum of two clusters. The median expression analog to B is plottet for each of these clusters.*

4.3.3 The clusters of up- and downregulated transcripts can be linked to distinct gene functions.

As shown in supplementary Fig. 10, the spleen is a clear example of two dinstict clusters with one consisting of upregulated previously inactive genes and one with downregulated highly active genes. For all these differentially expressed transcripts, we used the NCBI gene database to get a functional annotation. Of the 98 upregulated genes, 48 had a functional annotation. 17 of those were clearly associated with immune system or blood functions and thus relevant for the functional thymus. We further found 157 downregulated genes. There, 70 were annotated and 45 of those displayed a relation to the cell cycle or cell division. The tables of the transcripts with a relevant function are visible in the suppplementary material (Fig. 12 and Fig. 13).

4.3.4 Overrepresentation Analysis can create plots that signify organ development

For this analysis, the eight tissues with the most meaningful results were chosen.In Fig. 5, the most important functions for these tissue were determined through overrepresentation analysis. In addition, the Expression of the associated transcripts was plotted.

For the spleen (Fig. 5A), we determined a largely constant expression of immune-related genes throughout the time frame, with a slight increase in some functions from week 7-9. The brain (Fig. 5B) showed an increase in neuron projection morphogenesis from a previously inactive state (expression < 6.8) in week 4 to a significant expression (>7.5) by week 8. Synaptic signaling stayed at relatively constant expression levels. Heart-associated functions (Fig. 5C) can be grouped into two categories. Cardiac functions (cardiac muscle tissue development, heart contraction) are highly expressed in week 4 and fall continously until week 8. In contrast, general muscle gene sets are rising from originally lower expression levels during the observed time. The liver (Fig. 5D) shows no clear expression patterns, with some metabolic functions increasing through time (organic hydroxy compound metabolic process) while others stay mostly constant (cellular amino acid metabolic process) or fall (organic acid catabolic process). In contrast, skeletal muscle gene sets (Fig. 5E) show a very clear trend. After a mostly slight increase between week 4 and 8, a sharp rise in expression levels is visible from week 8 to 9. The testis-associated sets (Fig. 5F) continuously decrease in expression from week 5 onward. For the stomach (Fig. 5G), we found a initially high expression in week 4 that then falls until week 6 and then increases again towards week 9. Finally, the skin-associated functions (Fig. 5H) all displayed a constant rise in expression levels from week 5 to 9.

4.4 Hypothesis: Neural TRA expression patterns reflect morphological brain development

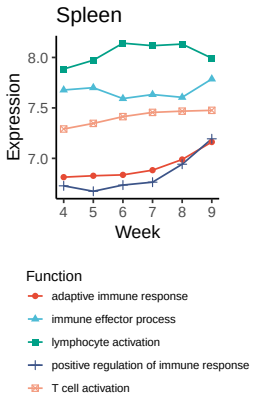
Our data covered 13 different brain subtissues, whose TRAs showed a signficant overlap. Hence, we filtered the differentially expressed transcripts for those that are only related to one of all 13 subtissues. The discovered genes were examined using the NCBI database. The resulting genes of interest can be categorized as follows:
1.Genes of Ion channels 2.Genes for neuronal development 3.Genes of cytokines

Main functions from overrepresentation analysis plotted by tissue

Spleen

FUNCTION	ADJ_P_VALUE
lymphocyte activation	1×10^{-65}
T cell activation	3×10^{-56}
adaptive immune response	4×10^{-53}
immune effector process	5×10^{-50}
positive regulation of immune response	8×10^{-50}
regulation of cell activation	3×10^{-49}
regulation of leukocyte activation	4×10^{-49}
regulation of lymphocyte activation	8×10^{-48}
leukocyte cell-cell adhesion	2×10^{-42}
immune response-regulating signaling pathway	1×10^{-41}

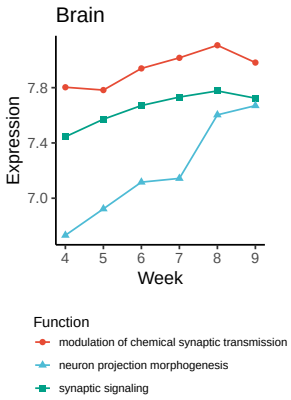
A



Brain

FUNCTION	ADJ_P_VALUE
synaptic signaling	9×10^{-43}
trans-synaptic signaling	9×10^{-43}
chemical synaptic transmission	9×10^{-43}
anterograde trans-synaptic signaling	9×10^{-43}
modulation of chemical synaptic transmission	9×10^{-25}
regulation of trans-synaptic signaling	9×10^{-25}
neuron projection morphogenesis	2×10^{-18}
cell projection morphogenesis	8×10^{-18}
plasma membrane bounded cell projection morphogenesis	2×10^{-17}
cell part morphogenesis	2×10^{-17}

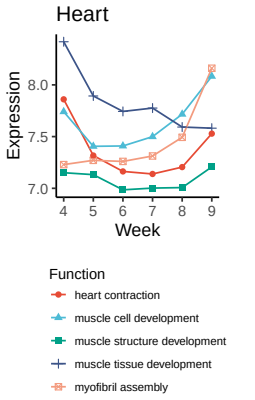
B



Heart

FUNCTION	ADJ_P_VALUE
muscle cell development	3×10^{-27}
muscle structure development	1×10^{-26}
myofibril assembly	2×10^{-24}
striated muscle cell development	2×10^{-24}
muscle tissue development	3×10^{-24}
cardiac muscle tissue development	3×10^{-24}
muscle system process	3×10^{-24}
striated muscle tissue development	2×10^{-23}
muscle cell differentiation	5×10^{-23}
heart contraction	1×10^{-22}

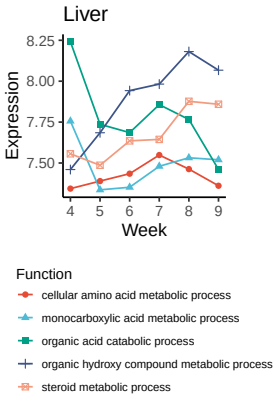
C



Liver

FUNCTION	ADJ_P_VALUE
monocarboxylic acid metabolic process	7×10^{-51}
small molecule catabolic process	7×10^{-39}
organic acid catabolic process	4×10^{-36}
carboxylic acid catabolic process	2×10^{-35}
steroid metabolic process	7×10^{-33}
small molecule biosynthetic process	4×10^{-31}
cellular amino acid metabolic process	1×10^{-28}
fatty acid metabolic process	6×10^{-28}
organic hydroxy compound metabolic process	9×10^{-28}
alpha-amino acid metabolic process	3×10^{-27}

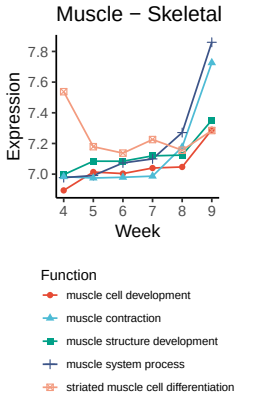
D



Muscle - Skeletal

FUNCTION	ADJ_P_VALUE
muscle structure development	7×10^{-29}
muscle system process	8×10^{-27}
muscle organ development	5×10^{-25}
muscle cell development	1×10^{-21}
muscle contraction	1×10^{-21}
muscle cell differentiation	4×10^{-20}
striated muscle cell differentiation	4×10^{-19}
myofibril assembly	3×10^{-18}
striated muscle cell development	3×10^{-18}
striated muscle contraction	1×10^{-16}

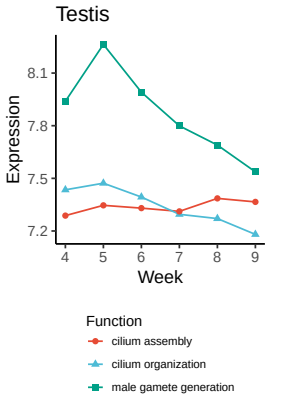
E



Testis

FUNCTION	ADJ_P_VALUE
male gamete generation	1×10^{-35}
spermatogenesis	3×10^{-34}
microtubule-based movement	2×10^{-26}
cilium organization	2×10^{-26}
cilium assembly	2×10^{-23}
microtubule cytoskeleton organization	2×10^{-23}
meiotic cell cycle	2×10^{-23}
cilium movement	3×10^{-20}
nuclear division	1×10^{-18}
meiotic cell cycle process	2×10^{-18}

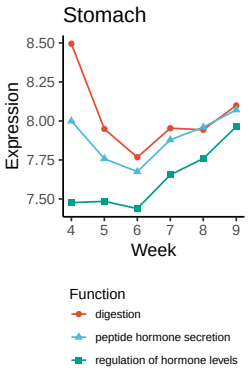
F



Stomach

FUNCTION	ADJ_P_VALUE
digestion	1×10^{-4}
regulation of hormone levels	6×10^{-4}
gastric acid secretion	2×10^{-3}
peptide hormone secretion	2×10^{-3}
peptide transport	2×10^{-3}
peptide secretion	3×10^{-3}
digestive system process	6×10^{-3}
amide transport	6×10^{-3}
hormone transport	6×10^{-3}
hormone secretion	6×10^{-3}

G



Skin

FUNCTION	ADJ_P_VALUE
epidermis development	2×10^{-21}
skin development	9×10^{-21}
keratinization	6×10^{-17}
keratinocyte differentiation	2×10^{-14}
epidermal cell differentiation	1×10^{-10}
intermediate filament cytoskeleton organization	9×10^{-10}
intermediate filament-based process	1×10^{-9}
intermediate filament organization	6×10^{-9}
regulation of water loss via skin	2×10^{-8}
establishment of skin barrier	1×10^{-7}

H

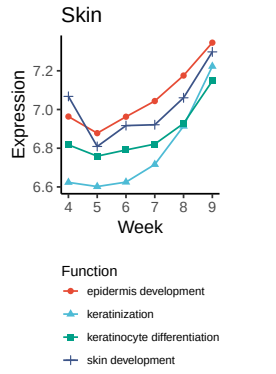


Figure 5: *For eight tissues A to H, the tables show the 10 most significant results from overrepresentation analysis. There, the functional GO annotation of all transcripts in our dataset (not only the differentially expressed ones) was compared to the functions associated with all TRAs of one tissue. The significance was determined by an adjusted p-value. Of those 10 gene sets, up to 5 were chosen to represent different groups of functions and avoid the overlap of highly related processes. For each of those, the median expression for each week was calculated and plotted, as seen in the expression graphs.*

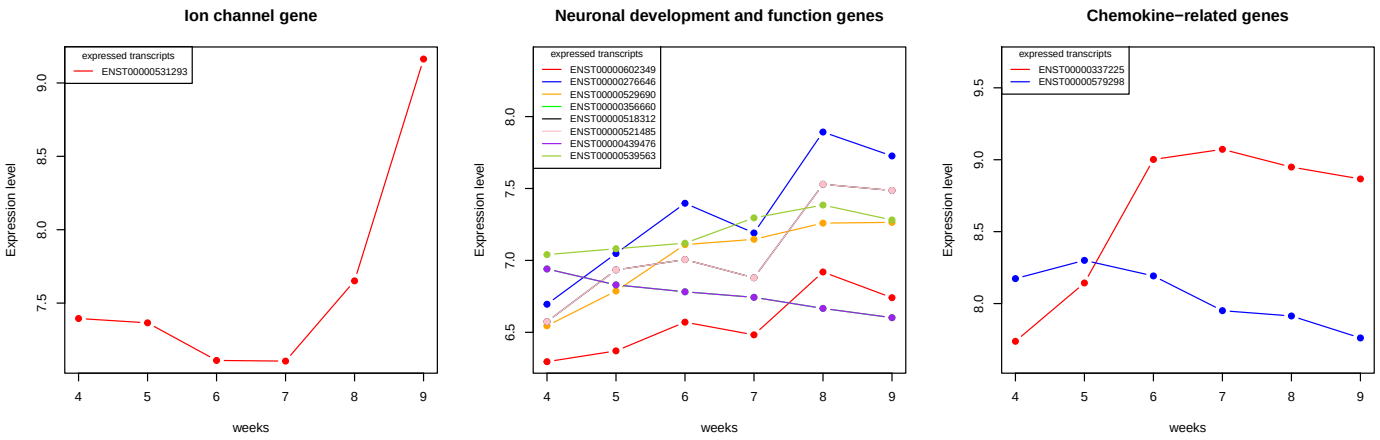


Figure 6: *The gene expression of different transcripts coding for an ion channel (left), neuronal development and function proteins (middle) and chemokine-related genes (right) was plotted for week 4 to 9.*

4.4.1 Ion channel

Ion channels play an important role in the function of neurons. We discovered that ENST00000531293 is highly expressed in the nucleus accumbens. It shows a significant increase between weeks 7 to 9 and codes for SLN sarcolipin which is a sarcoplasmic reticulum Ca-ATPase (Fig. 6 left).

4.4.2 Genes for neuronal development and function

The second group are genes with a specific role in neuronal development and function. There, we discovered that ENST00000276646 and ENST00000529690 expression increases significantly over time. Both transcripts are associated with the cerebellar hemisphere and code for SYBU (syntabulin), a protein that contributes to activity-dependent presynaptic assembly in neuronal development.
Furthermore, we found four transcripts for axon goudance: ENST00000602349 codes for NXPH1 (neurexophilin 1) which forms a tight complex with neurexins. These proteins promote adhesion between axons and dendrites. The transcript shows a strong rise in expression, especially from weeks 7 to 8, and is connected to the anterior cingulate cortex (Fig. 6 middle).
ENST00000518312 and ENST00000521485 encode for SNAP91 (synaptosome associated protein 91), which plays a role in regulation of clathrin-dependent endocytosis. Therefore, SNAP91 is important for essential axonal functions of neurons like postsynaptic density (Overhoff et al. 2020). The gene is associated with the cerebellar hemisphere and also shows a significant increase between weeks 7 to 8 (Fig. 6 middle).
In addition, ENST00000539563 encodes for LSAMP (limbic system associated membrane protein) which plays a role in axon guidance. The encoded preprotein is processed into a neuronal surface glycoprotein which functions as an adhesion molecule during axon guidance and neuronal growth in the developing limbic system. ENST00000539563 is associated with the Putamen, a part of the basal ganglia that are associated with the limbic system- ENST00000356660 and ENST00000439476 code for BDNF (brain derived neurotrophic factor). A binding of BDNF to its receptor promotes neuronal survival. Both transcripts show an identical decline in expression over the weeks. Nonetheless, ENST00000356660 is connected to cerebellar hemisphere while ENST00000439476 to related to the hippocampus.

4.4.3 Chemokine related genes

The last group are chemokines, proteins with an important role in the signaling process during neuronal development (Tiveron 2008). We discovered ENST00000337225, which encodes for CXCL14 (alpha class chemokin ligand). It shows a strong increase in expression between week 4 to 6 and is associated with the anterior cingulate cortex. ENST00000579298 encodes for NUP85 (nucleoporin 85), a protein component of the Nup107-160 subunit of the nuclear pore complex. NUP85 can bind to CCR2 (a receptor for beta class chemokines). ENST00000579298 is related to the frontal cortex and shows a decline between week 5 to 9 (Fig. 6 right).

4.5 No TRAs could be identified as key biomarkers for developmental progression

Principal component analysis (PCA) was performed on a matrix containing all differentially expressed genes throughout the 6 weeks. This was performed to reduce dimension while keeping most of the data's variance. To get a better grasp of how much variance is explained by each principal component (PC), the percentage variance of each PC together as well as the cumulative variance were plotted. This visualization helps to determine how many PCs are needed to explain a significant amount of the data's variation. The chosen PCs can then be used for further analysis. The first three PCs already explain over 80% of the data's variance, with the first PC alone accounting for over 60% (supp. Fig. 16 A).

For further analysis, each transcript was awarded a rank depending on how much it contributes to each PC. Then, a score was calculated for maximum contribution to PC1 and minimal loadings for all other relevant PCs (PC2-5). We took a closer look at the five best socring transcripts for PC1, as this PC alone already explains the brunt of the variance of our data. These transcripts encode for three different proteins: Low-Density Lipoprotein Receptor Class A Domain-Containing Protein 4 (LDLRAD4) encoded by ENST00000399848, Calcium-activated potassium channel subunit beta-2 (KCNMB2) encoded by ENST00000432997 and ENST00000452583 and SLIT And NTRK-Like Protein 3 (SLITRK3) encoded by ENST00000241274 and ENST00000475390.

The transcripts of ENST00000432997 and ENST00000452583 encode for the same Protein (KCNMB2). Fitting to that, both transcripts consistently display the same expression level over all weeks. The same applies to the transcripts of ENST00000241274 and ENST00000475390, as they also encode for the same protein (SLITRK3) and therefore also display the same expression levels. The transcripts for LDLRAD4 show the highest overall expression levels. Their expression pattern also does not display many similarities to those of KCNMB2 and SLITRK3. However, KCNMB2 and SLITRK3 show a highl similar curve over all weeks (supp. Fig. 16 B).

5 Discussion

5.1 Comparison with original paper

In contrast to our *limma* analysis to generate data with biological relevance, the authors of the original paper (Yi H *et al.*, 2010) of our data set used *One-way analysis of variance* to acquire differential expressed transcripts. We compared both data subsets by k-means clustering and plotting the overlapping Intensity values (Fig. 2). By looking at the trends of the cluster for *up* and *down* regulated gene expression, we can validate the method used by the authors. In contrast the *arch* regulated genes-clusters do not present a clear result and show a weakness of the method used. It falsely detects some genes as differentially expressed, assuming our *vsn* normalization is correct, which is indicated by the quality control (section 3.3.1).
Furthermore, we showed in our Venn diagram (Fig. 14), that our method detected differentially expressed transcripts not included in their data. This questions the significance of either the papers results or again our *vsn* normalization and subsequent *limma* analysis, which is rather unlikely given our quality control and plots.
The discrepancy can be explained by the publication date, which is year 2010. Now there are more advanced algorithms that determine differentially expressed transcripts more precisely. The *vsn* and the *limma* package are up to date with frequent advances (Ritchie. et al. 2015).
Overall, this ensured the quality of our data set and differentially expressed transcripts while also pointing out some flaws of the data set of the paper.

5.2 TRAs can infer a basic timeline of organ development

In our analysis, we have shown that a number of TRAs are differentially expressed (section 4.3.1) between week 4 and 9 of human embryonic development in each of the analyzed tissues. Nonetheless, the expression levels of TRAs associated with one tissue do not constitute a useful metric for the organ's development (section **??**). This can be explained by the fact that within one tissue's TRAs, there are multiple groups of genes both distinct in expression patterns (clustering in section **??**) and function (analysis of spleen gene functions in section 4.3.3). Thus, we determined that the expression over time of functional gene sets linked to specific tissues through overrepresentation analysis is a more meaningful metric for organ development.

This approach was used in section 4.3.4 for eight different tissues. For the spleen, the results of our analysis (Fig. 5A) largely do not reflect the embryonic development (section 2.2). While some of the immune-related gene sets are already expressed in week 4, the spleen only develops in week 6 and contains immune cells by week 12. This shows that while the spleen plays a role in the immune system and such gene sets are therefore rightly linked to the spleen, the expression of these transcripts alone does not necessarily relate to the development of the organ. It is still noteworthy that functions

related to the adaptive immune system increase in expression form week 7 onward, which correlates with the beginning of T-cell development in the thymus.
The observed timeframe is an important part of brain development (section 2.2). This is also visible in the expression data (Fig. 5B), with a already high but still continously increasing expression of synaptic gene sets. Furthermore, as the brain starts to form, the expression of neuron projection morphogenesis transcripts increases continuously from week 5 to 8.
At week 4, the clearly heart-associated gene sets (Fig. 5C) are at their highest expression level and decrease until week 8. The cardiac muscle tissue development transcripts still remain highly expressed (>7.5). This corresponds to the early development of the heart as noted in the introduction (section 2.2). It is noteworthy that the heart contraction gene set rises in expression again from week 8 to 9, but here an explanation is not possible without further analyzing the individual genes.
The liver-associated TRAs showed no clear expression pattern (Fig. 5D). Thus, even though the liver forms mostly during the analyzed timeframe (section 2.2), we cannot link the gene expression to the organ's development. The detected functions are mostly metabolic pathways whose activity could also be related to processes outside the liver. As a result, it is plausible that their expression is independent of liver development.
The skeletal muscle functions are expressed only late within the observed time, as shown by the large increase in expression from week 8 to 9 (Fig. 5E). As muscle fibers begin to develop later than week 9 and the first related proteins appear from week 7 on (section 2.2), these expression data correspond well to the embryonic development.
The testis gene sets decrease in expression from week 5 onward (Fig. 5E). This is in contrast to the embryonic development, where the gonads start to form at around the same time (section 2.2).
For the stomach, the expression pattern indicates a decrease until week 6 followed by rising expression levels until week 9 (Fig. 5G). However, the literature indicates that these results are unrelated to the stomach development. Functions like digestion or peptide hormone secretion are impossible to occure at this time, since the specific cells needed for this only appear later in embryogenesis (section 2.2). Therfore, the cause of the changing expression would have to be determined through a more in-depth analysis of the involved genes.
Finally, the skin shows an increased expression of related genes sets from week 5 to 9 (Fig. 5H). This broadly reflects the embryonic development, with the epidermis starting to form in week 4 (section 2.2). We also found this expression pattern in the keratinization gene set that is suggested by literature as a good indicator for skin formation.

5.2.1 Brain subtissue-specific gene expression reflects neuronal development processes

First, we found a significant increase in the expression of SLN sarcolipin (a Ca-ATPase) (Fig. 6 left). Since Ca(2+) is an essential cofactor for actin dependent cell migration, this might be related to neuronal migration starting week 9 (section 2.2).
Additionally, the strong correlation between the expression of SNAP91 genes and one SYBU gene (ENST00000276646) (Fig. 6 middle) was noteworthy. Both proteins are associated with the cerebellar hemisphere and contribute to endocytosis (Overhoff et al. 2020).
In addition we identified a significant increase in expression of NXPH1 (Fig. 6 middle). An upregulation of this factor can prepare the process of synapse formation which starts at week 11 (section 2.2).
We further identified a strong LSAMP expression associated to the putamen as well as NXPH1 linked with the anterior cingulate cortex. (Fig. 6 middle). As adhesion molecules for axon guidance, their increases in expression (from week 6/7 on) could be in preparation for synapse formation at week 9 (section 2.2).
The neuronal cell migration is strongly dependent on chemoattractors like chemokines (Tiveron 2008). We identified a significant increase in CXCL14 between weeks 5 to 6 that maintains a high expression level afterwards. Fig. 6 right). This could be an accumulation for neuronal migration, starting by week 9 (section 2.2). A decline in NUP85 expression is also notable (Fig. 6 right). NUP85 can bind to CCR2, hence a decline in reduce the chances for this binding. A consequence might be that more CCR2 receptors are free for beta type chemokin mediated signals. Overall, there were two cases were two closely related transcripts (linked to the same gene) showed identical expression values. This phenomenon could be caused by failures in annotation, e.g. associating two transcripts with one probe.

5.3 No TRAs could be identified as key biomarkers for developmental progression

Principal component analysis helped to identify transcripts that have a high contribution to our data's variance. The high variance explained by PC1 alone made it possible to focus our further analysis only on this PC. Ranking the transcripts according to their contribution led us to identify the three proteins LDLRAD4, KCNMB2 and SLITRK3 as main contributors to only principal component 1. Nonetheless, the loadings for these genes are not significantly larger than those of other genes. Therefore, the information gained trough PCA was not sufficient to reliably classify these or any TRAs as biomarkers for embryonic development. More in-depth analysis is necessary to determine whether PCA is

an insufficient method to identify TRAs these biomarkers, or if using TRAs as markers is an unsuccessful approach in general.

5.4 Conclusion

The overarching goal of this research was to gain insight on human embryonic development between week 4 and 9 through the expression patterns of tissue-restricted antigens (TRAs). Under this main objective, we divided our research into distinct parts. First, we managed to find differentially expressed TRAs associated with all tissues (section 4.1) and determined that our method of analysis even has advantages compared to the original paper (section 4.2). Then, we used these results to try to determine the developmental steps occurring during the observed time frame (section 4.3). While not all TRA-specific expression patterns could be linked to morphological events, we were able to show that the expression patterns for some organ-specific gene sets are highly correlated with milestones in development (section 5.2). An in-depth analysis of brain subtissues proved diffcult due to the high degree of TRA overlap between them. Still, we found genes related to neuronal function that were TRAs for only one subtissue (section 4.4). Based on their expression, we could draw some conclusions on neuronal development (section 5.2.1). Finally, we tried to define individual genes as biomarkers for embryonic development through PCA (section 5.3). There, we could not define such markers effectively based on the selection of genes through principal component loadings. All in all, we were still able to show that the morphological events during embryogenesis are reflected in gene expression and prove that differential expression analysis can be a valid method for embryonic research.

6 References

Benninghoff, A. (1993), Makroskopische Anatomie, Embryologie und Histologie des Menschen, 15. Auflage, München, Wien, Baltimore, Urban und Schwarzenberg.

Deutsch J. (2013), Embryologie und Physiologie der Leber, Pädiatrische Gastroenterologie, Hepatologie und Ernährung, 375–87, Berlin Deutschland.

Dinkelacker M., (2019), Chromosomal clustering of tissue restricted antigens.

Esrefoglu M., Taslindere E. & Cetin A. (2017), Development of the Esophagus and Stomach, Bezmialem sci. *5*, 175-82.

Fasold M. & Binder H., (2013), AffyRNADegradation: control and correction of RNA quality effects in GeneChip expression data, Bioinformatics, *29*, 129-31.

Hayward AR., (1983), The human fetus and newborn: development of the immune response, Birth Defects Orig. Artic. Ser. *19*, 289-94.

Hikspoors J.P.J.M., Kruepunga N., Mommen G.M.C. et al., (2022), A pictorial account of the human embryonic heart between 3.5 and 8 weeks of development, Commun. Biol. *5*, 226.

Hu MS., Borrelli MR., Hong WX., Malhotra S., Cheung ATM., Ransom RC., Rennert RC., Morrison SD., Lorenz HP. & Longaker MT., (2018) Embryonic skin development and repair, Organogenesis *14*, 46-63.

Huber W., von Heydebreck A., Sültmann H., Poustka A., Vingron M., (2002), Variance stabilization applied to microarray data calibration and to the quantification of differential expression, Bioinformatics. *18* ,Suppl 1, 96-104.

James F. & Jones M.D., (1983), Development of the Spleen, Lymphology *16*, 83-89, Georg Thieme Verlag Stuttgart, New York.

Kluth D., Jaeschke-Melli S. & Fiegel H., (2003),The embryology of gut rotation, Semin. Pediatr. Surg. *12*, 275-279.

Müller W. A. & Hassel M., (2018), Entwicklungsbiologie und Reproduktionsbiologie des Menschen und bedeutender Modellorganismen, 6. Auflage, Springer-Verlag GmbH Deutschland, Berlin, Deutschland.

Murphy K. & Weaver C., (2018), Janeway Immunologie, 9. Auflage, Springer-Verlag GmbH Deutschland, Berlin, Deutschland.

National Research Council and Institute of Medicine, (2009), Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities, Washington, DC: The National Academies Press.

Overhoff M., De Bruyckere E. & Kononenko N- L., (2020), Mechanisms of neuronal survival safeguarded by endocytosis and autophagy, J. Neurochem. *157*, 263-296.

Ritchie ME., Phipson B., Wu D., Hu Y., Law CW., Shi W. & Smyth GK., (2015), limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. *43*, e47.

Romero N. B., Mezmezian M .& Fidziańska A. (2013), Pediatric Neurology Part III, Chapter 137 - Main steps of skeletal muscle development in the human: Morphological analysis and ultrastructural characteristics of developing human muscle, Handb. Clin. Neurol. *113*, 1299-1310.

Tiveron M.C. & Cremer H., (2008), CXCL12/CXCR4 signalling in neuronal cell migration, Curr. Opin. Neurobiol. *18*, 237-244.

Ulfig N., (2009), Kurzlehrbuch Embryologie, Georg Thieme Verlag Stuttgart,New York.

Yi H., Xue L., Guo MX., Ma J., Zeng Y., Wang W., Cai JY., Hu HM., Shu HB., Shi YB. et al., (2010), Gene expression atlas for human embryogenesis, FASEB. J. *24*, 3341-50..

Yusuf F., Rehimi R., Dai F. & Brand-Saberi B., (2005), Expression of chemokine receptor CXCR4 during chick embryo development, Anat. Embryol. *210*, 35-41.

7 Supplementary

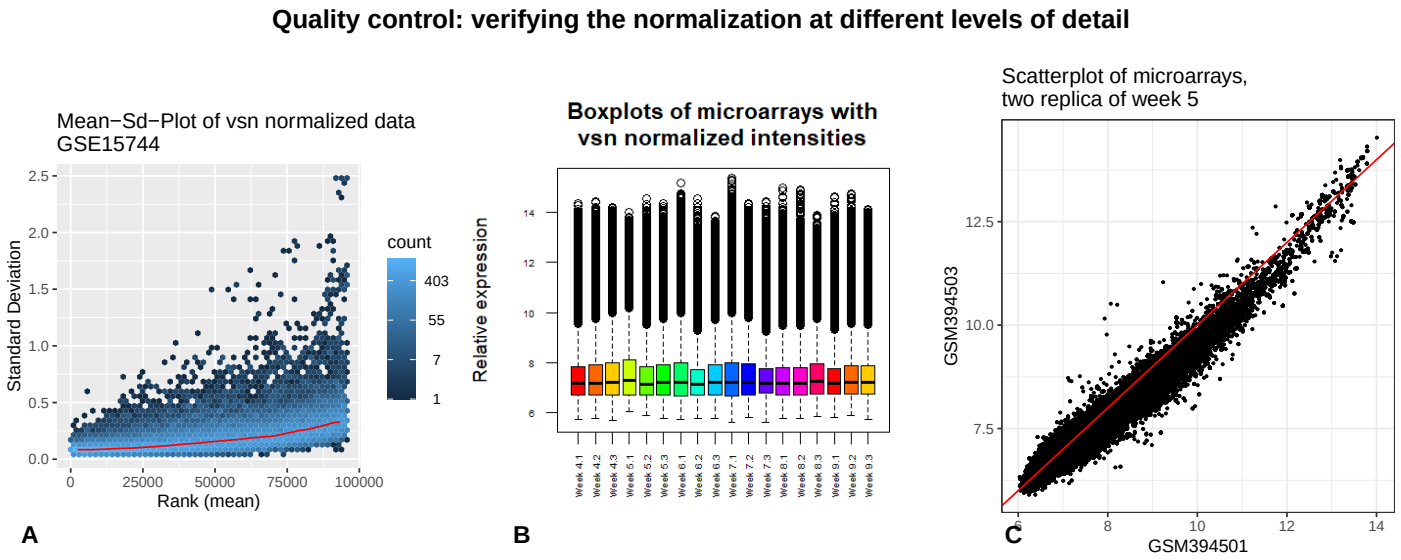


Figure 7: *The plots support the use of the vsn normalization on our data set. A: The red line is close to horizontal, although it shows some correlation at high intensity leves. B: The boxplots show nice allignment of the mean intensity values. Some outliers are given but can neglected given the 0.25 and 0.75 quantile. C: A selected Scatterplot is shown. Very slight banana shaped structure can be seen, but only marginal. Overall the quality control confirmes successful vsn normalization*

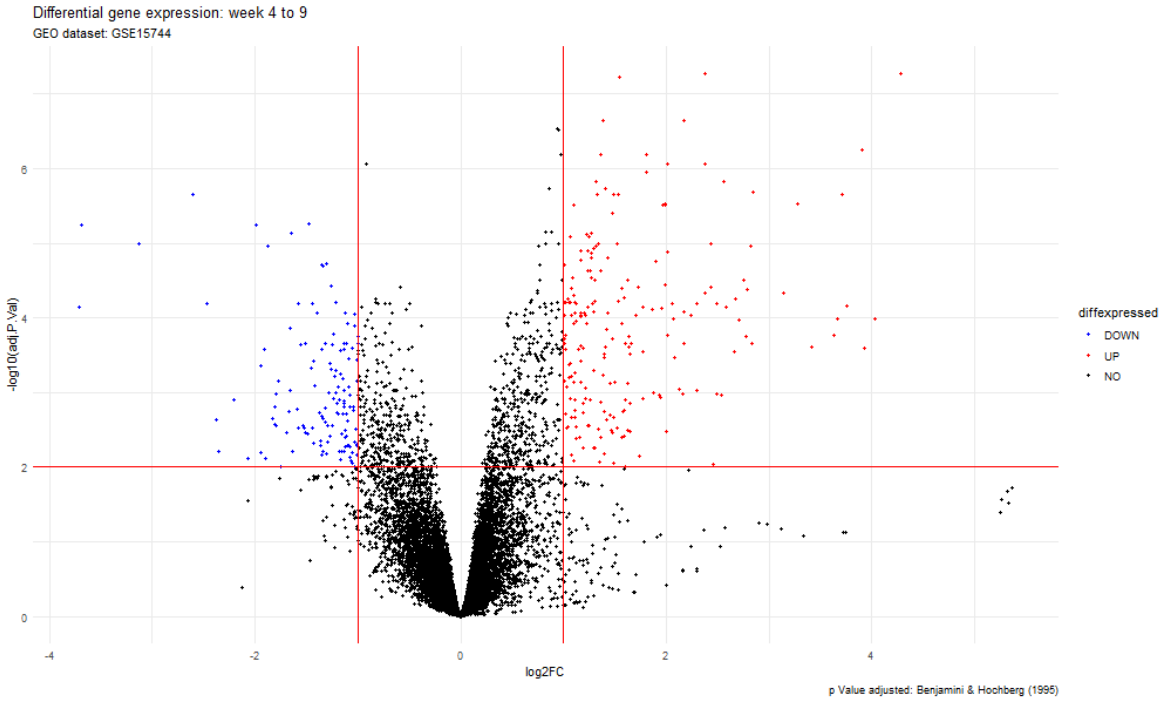


Figure 8: *Volcano plot for differential gene expression between week 4 and 9. The adjusted P-value boundary was set at 0.01. The logFC boundarys were -1 and 1,which reflects a doubling in the expression. Genes are up-regulated if the logFC value is larger than 1 and the adjusted P-value is smaller than 0.01 and down-regulated if the logFC value is smaller than 1 and the adjusted P-value is smaller than 0.01. Genes which didn't belong in those groups were declared as not differentially expressed.*

NULL

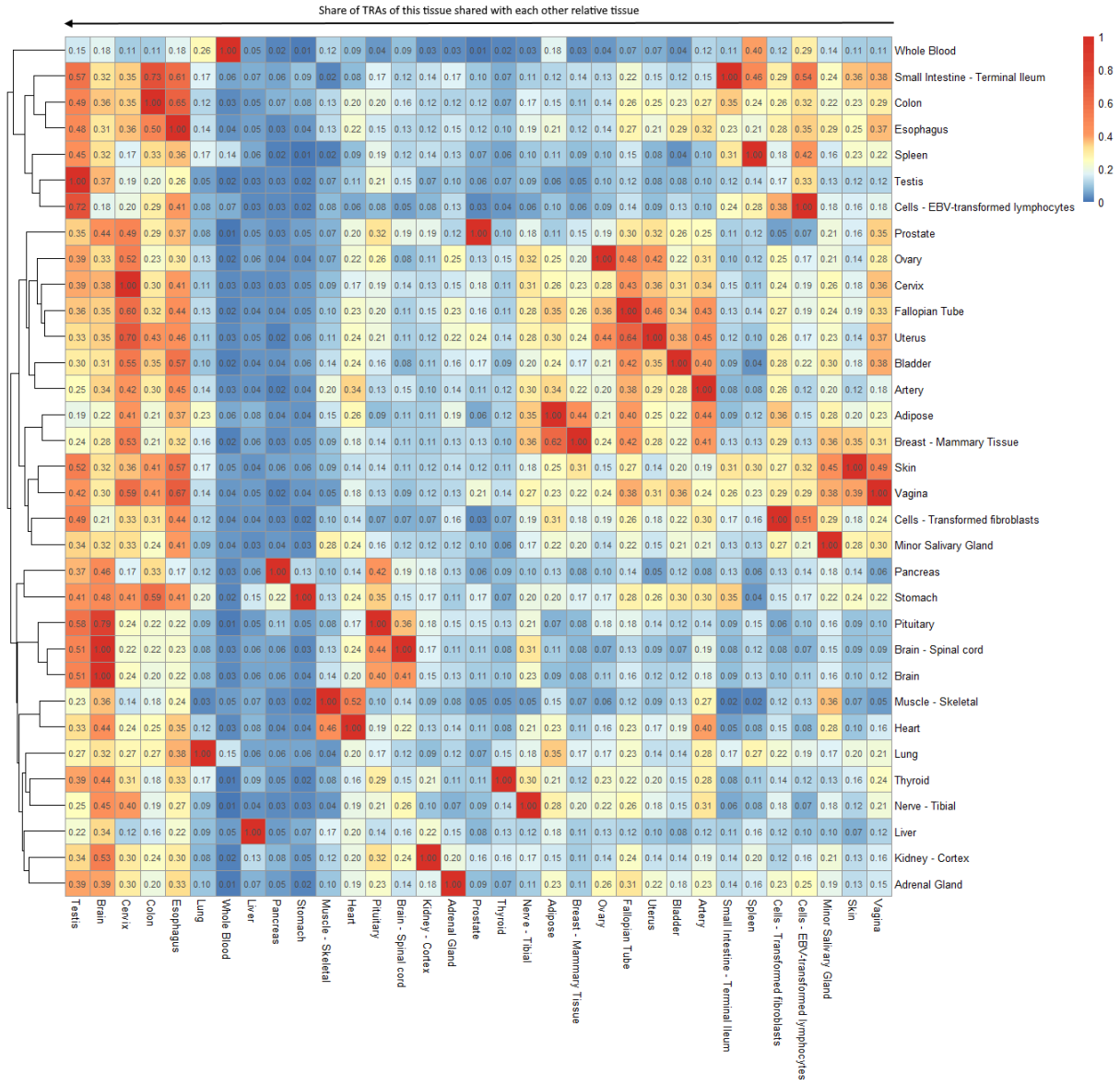


Figure 9: *For each of the tissues on the right, the share of its TRAs that are also associated with the tissue in the respective column is displayed.*

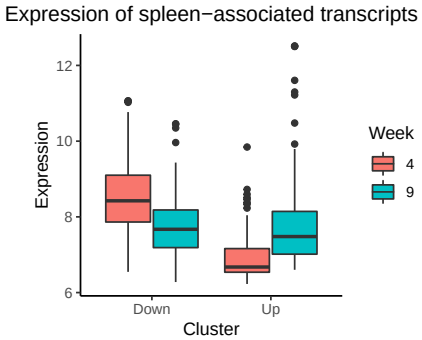


Figure 10: *A further look on the expression of transcripts in the up- and downregulated clusters shows that the upregulated transcripts are close to the minimum expression level between 6 and 7 in week 4 and showing expressions between 7 and 8.5 by week 9. In contrast, the downregulated genes have very high expression levels (8-9) by week 4 and decrease to a more moderate expression between 7 and 8.5 analogous to the upregulated transcripts.*

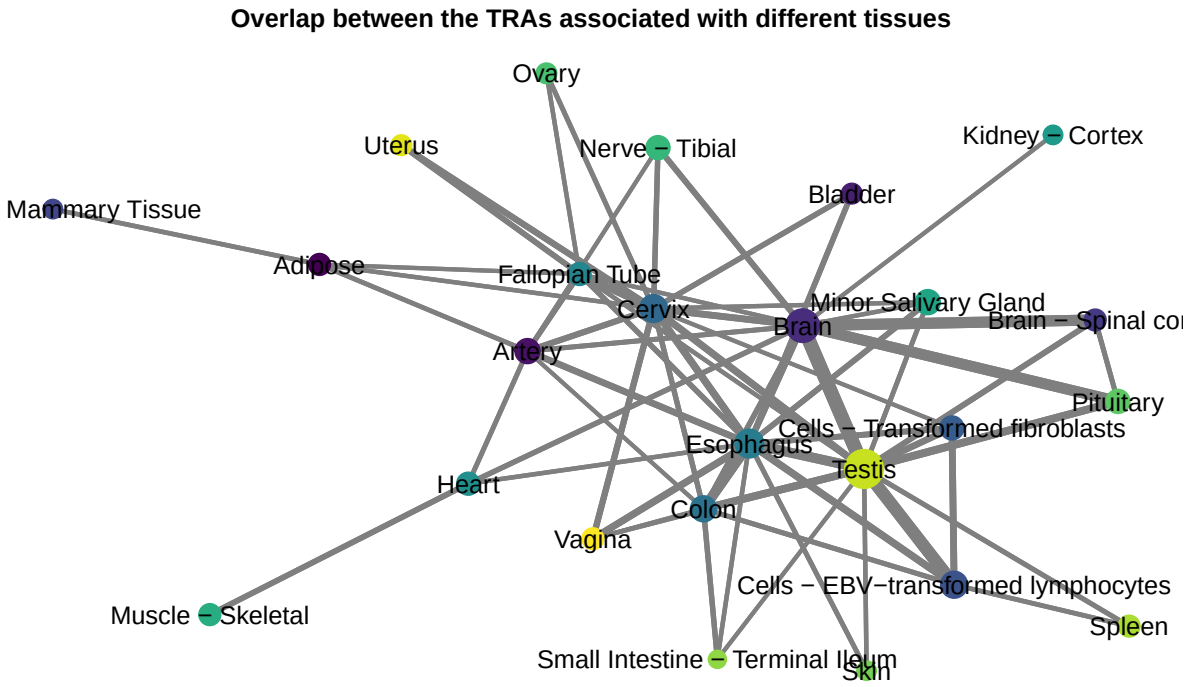


Figure 11: *Each tissue is displayed as a node with its size representing the number of transcripts associated with it. The edges show the shared TRAs between the linked tissues, with only links corresponding to more than 100 common TRAs visible.*

Spleen - Upregulated

List of differentially epxressed genes

TRANSCRIPT	GENE	PROTEIN	SUMMARY	EXPRESSION_OVER_TIME
ENST00000251296	IGSF21	immunoglobin superfamily member 21	Member of the immunoglobulin superfamily	
ENST00000259698	RIPOR2	RHO family interacting cell polarization regulator 2	Mediates polarization of T cells and neutrophils	
ENST00000259989	FGFBP2	fibroblast growth factor binding protein 2	Secreted by cytptpxic lymphocytes	
ENST00000265162	ENPEP	glutamyl aminopeptidase	Can upregulate blood pressure	
ENST00000286758	CXCL13	C-X-C motif chemokine ligand 13	B lymphocyte chemoattractant	
ENST00000318041	IL11RA	interleukin 11 receptor subunit alpha	IL-11 receptor	
ENST00000335295	HBB	hemoglobin subunit beta		
ENST00000358511	COL6A6	collagen type VI alpha 6 chain	Contains von Willebrand factor domains	
ENST00000368732	S100A8	S100 calcium binding protein A8	May function as a cytokine	
ENST00000374429	CXCL12	C-X-C motif chemokine ligand 12	Plays a role in immune surveillance	
ENST00000394635	CFI	complement factor I	Regulating the complement cascade	
ENST00000395388	HLA-DRA	major histocompatibility complex, class II, DR alpha	HLA class II	
ENST00000400131	CHODL	chondrolectin	Endocytosis of pathogen	
ENST00000404220	IFNAR2	interferon alpha and beta receptor subunit 2	Receptors for interferons alpha and beta	
ENST00000445105	FGF12	fibroblast growth factor 12	Fibroblast growth factor	
ENST00000512148	CFI	complement factor I	Regulating the complement cascade	
ENST00000598319	FCGRT	Fc gamma receptor and transporter	Protect the antibody from degradation	

Figure 12: *Table of all spleen-associated genes from the upregulated cluster with a function related to the spleens overall purpose (immune and blood-related genes)*

Spleen - Downregulated

List of differentially epxressed genes

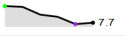
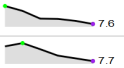
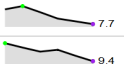
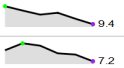
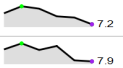
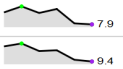
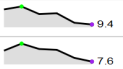
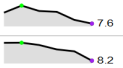
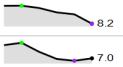
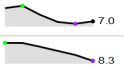
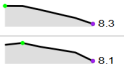
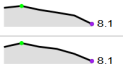
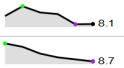
TRANSCRIPT	GENE	PROTEIN	SUMMARY	EXPRESSION_OVER_TIME
ENST00000233505	CENPA	centromere protein A	Component of the centromere	
ENST00000240423	NCAPH	non-SMC condensin I complex subunit H	Interphase chromosome condensation	
ENST00000251496	NCAPG	non-SMC condensin I complex subunit G	Chromosome condensation and stabilization	
ENST00000260359	NUSAP1	nucleolar and spindle associated protein 1	Spindle microtubule organization	
ENST00000260731	KIF11	kinesin family member 11	Spindle dynamics	
ENST00000287598	BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	Spindle checkpoint function	
ENST00000300093	PLK1	polo like kinase 1	Highly expressed during mitosis	
ENST00000302759	BUB1	BUB1 mitotic checkpoint serine/threonine kinase	Central role in mitosis	
ENST00000306324	HOXD4	homeobox D4		
ENST00000310955	CDC20	cell division cycle 20	Nuclear movement in the cell cycle	
ENST00000313288	TACC3	transforming acidic coiled-coil containing protein 3	Stabilization of the mitotic spindle	
ENST00000327331	CDCA8	cell division cycle associated 8	Regulator of mitosis	
ENST00000335658	DPPA4	developmental pluripotency associated 4	Stem cell pluripotency	
ENST00000348459	HELLS	helicase, lymphoid specific	Involved with cellular proliferation	
ENST00000348581	EXO1	exonuclease 1	Mismatch repair	
ENST00000349718	CMTM7	CKLF like MARVEL transmembrane domain containing 7	Regulates G1/S transition	
ENST00000352331	KIF23	kinesin family member 23	Moves chromosomes	
ENST00000352433	PTTG1	PTTG1 regulator of sister chromatid separation, securin	Anaphase-promoting complex substrate	
ENST00000355112	BLM	BLM RecQ like helicase	Unwinds Holliday junction	
ENST00000357481	ACIN1	apoptotic chromatin condensation inducer 1	Induces apoptotic chromatin condensation	
ENST00000359803	MDK	midkine	Promotes cell growth	
ENST00000361953	FOXM1	forkhead box M1	Transcriptional activator in proliferation	
ENST00000366955	CENPF	centromere protein F	Associates with centromer-kinetochore complex	
ENST00000371566	ORC1	origin recognition complex subunit 1	Initiation of DNA replication	
ENST00000372462	CDC20	cell division cycle 20	Regulatory in the cell cycle	
ENST00000374403	KIF4A	kinesin family member 4A	Maintaining chromosome integrity in mitosis	
ENST00000380026	CENPE	centromere protein E	Stable spindle microtubule capture	
ENST00000380756	DNMT3A	DNA methyltransferase 3 alpha	CpG methylation	
ENST00000382011	MCM5	minichromosome maintenance complex component 5	Initiation of DNA replication	
ENST00000395284	CDK1	cyclin dependent kinase 1	Subunit of M-phase promoting factor	
ENST00000395405	ZWINT	ZW10 interacting kinetochore protein	Involved in kinetochore function	
ENST00000396863	MYBL2	MYB proto-oncogene like 2	Cell cycle progression	
ENST00000404276	CHEK2	checkpoint kinase 2	Cell cycle checkpoint regulator	
ENST00000412359	BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	Spindle checkpoint function	
ENST00000414849	NUSAP1	nucleolar and spindle associated protein 1	Spindle microtubule organisation	
ENST00000420246	TP53	tumor protein p53	Inducing cell cycle arrest	
ENST00000427946	NCAPH	non-SMC condensin I complex subunit H	Subunit of condensin complex	
ENST00000428239	CDC7	cell division cycle 7	Cell division cycle protein	
ENST00000433496	RFC4	replication factor C subunit 4	DNA polymerase accessory	
ENST00000515001	CENPH	centromere protein H	Centromere-microtubule complex	
ENST00000518483	EXO1	exonuclease 1	Mismatch repair	
ENST00000520452	PTTG1	PTTG1 regulator of sister chromatid separation, securin	Anaphase-promoting complex substrate	
ENST00000534871	AURKB	aurora kinase B	Segregation of chromosomes	
ENST00000535254	BUB1	BUB1 mitotic checkpoint serine/threonine kinase	Central role in mitosis	
ENST00000561208	MEIS2	Meis homeobox 2	Contributor to developmental programs	

Figure 13: *Table of all spleen-associated genes from the downregulated cluster with a function related to the cell cycle of cellular division*

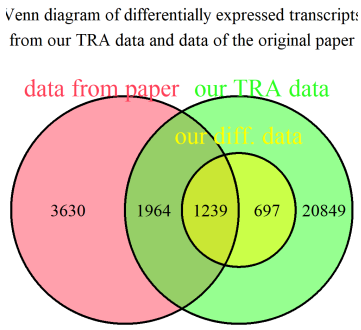


Figure 14: ***The Venn diagram shows overlapping sections between our data and the data from the original paper.*** *According to the paper, of the its 6,833 transcripts are differentially expressed. This overlapps with our general data with all TRAs (24,749 transcripts) and with our differentially expressed (diff.) data according to our limma analysis with a p-value of 0.01 (1,837 transcripts)*

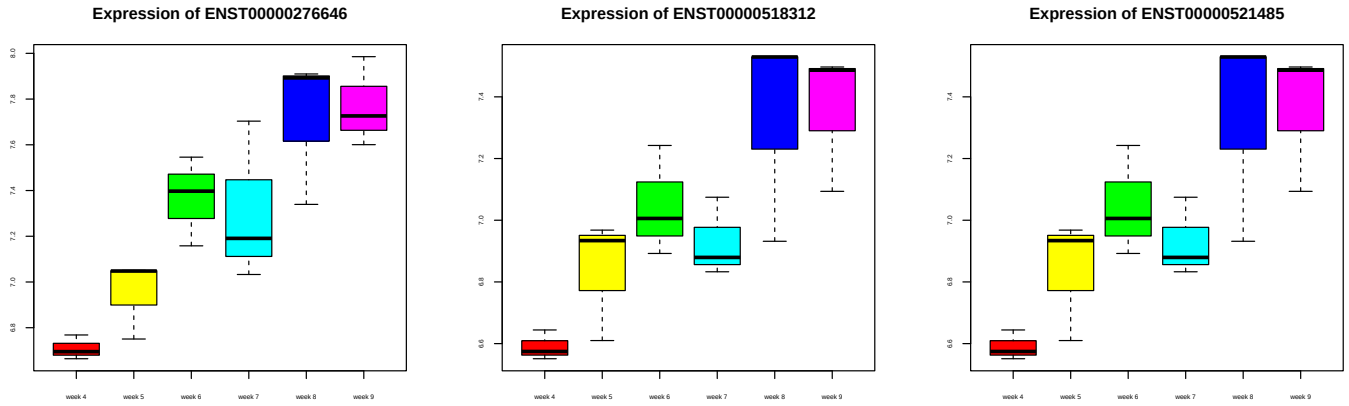


Figure 15: *A boxplot of ENST00000276646, ENST00000518312 and ENST00000521485 was made to evaluate the scattering of the three samples per week for the 6 weeks from our dataset.*

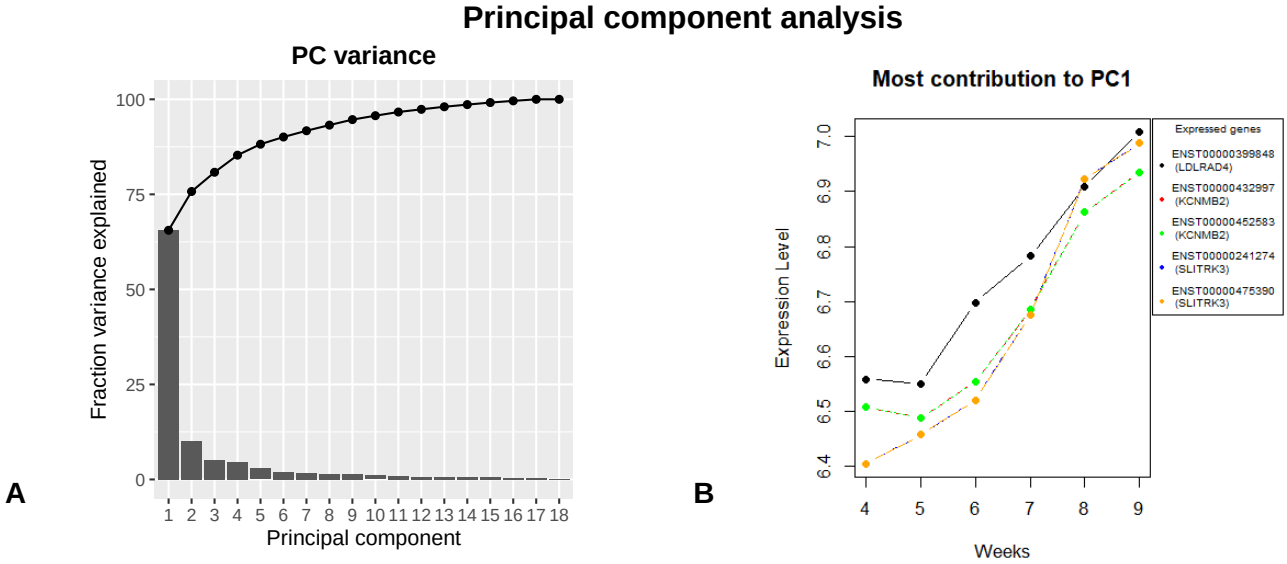


Figure 16: *A. Variance explained by each individual PC as well as cumulative variance explained by the PCs together. B. Expression levels of the five transcripts that contribute the most to PC1.*