# Test4

Paul Christmann

2022-07-12

# TRAs can infer a basic timeline of organ development

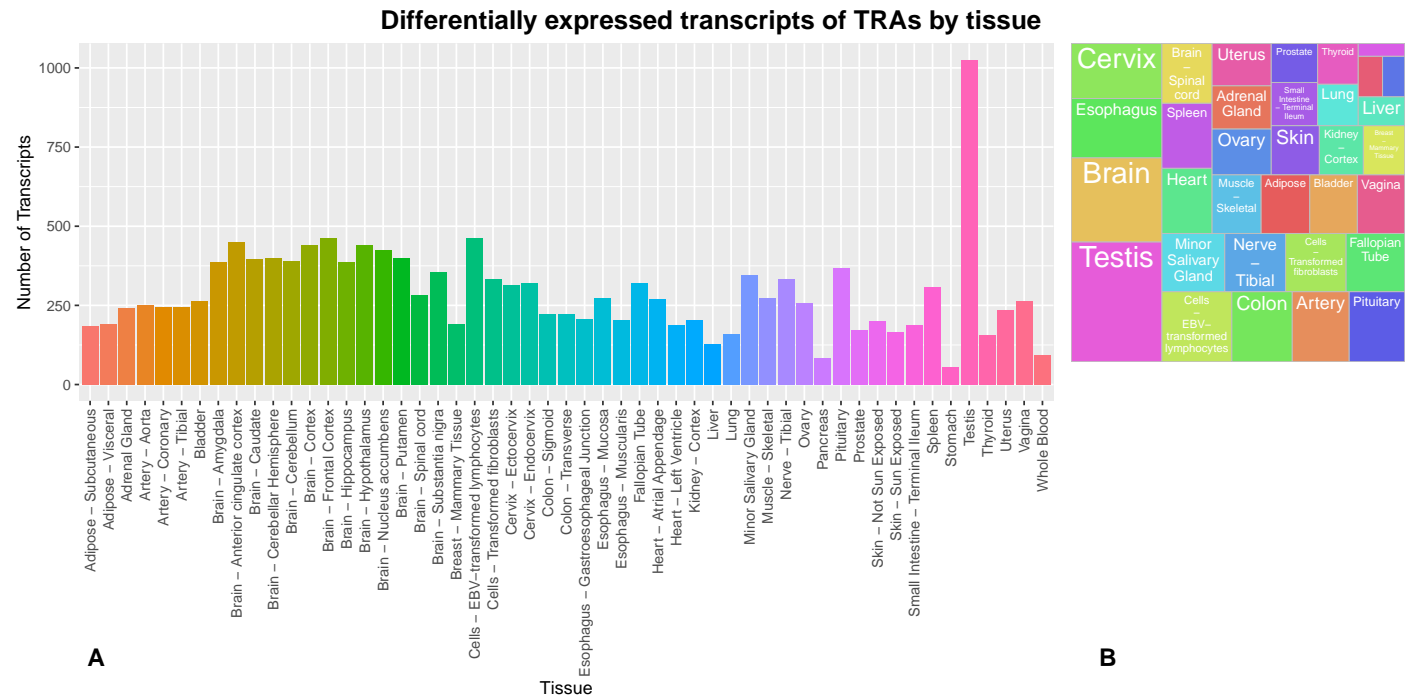## Differentially expressed transcripts can be linked to all analyzed tissues



*Figure 1:* 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.*

The TRA dataset covers 53 distinct tissues. For all of those, we found at least 40 differentially expressed transcripts within our dataset. The minimum were 46 stomach-linked transcripts, the highest reached 837 TRAs for the testes.These numbers are sufficient for further analysis of the gene expression within individual tissues. Therefore, all results will be based on our dataset with differentially expressed genes from limma analysis.

Nonetheless, it should be noted that 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. @ref(fig:tissue-links). Furthermore, a detailed heatmap for the shared TRAs between tissues is shown in the supplementary Fig. @ref(fig:heatmap-tissues).

## 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. @ref(fig:expression-plot)A. Furthermore, we observed trends within the median expression of these trascnrips (Fig. @ref(fig:expression-plot)B). 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. @ref(fig:expression-plot)C. 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.

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

As shown in supplementary Fig. @ref(fig:spleen-boxplot), 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

**Expression over time of Spleen–related differentially expressed transcripts**
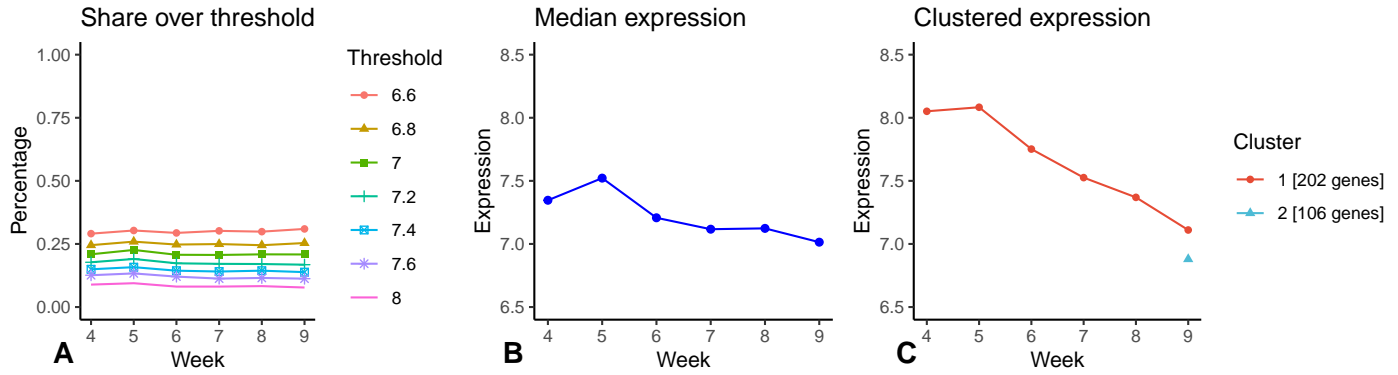
*Figure 2:* A. For different expression thresholds the share of differentially expressed transcripts with higher expressions than the threshold is depicted. B. The median expression for all spleen-associated TRAs is shown for each point in time. C. For k-means clustring, the silhouette score determined an optimum of two clusters. The median expression analog to B is plottet for each of these clusters.

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. @ref(fig:spleen-up-table) and Fig. @ref(fig:spleen-down-table)).

**Overrepresentation Analysis can create plots that signify organ development**

For this analysis, the eight tissues with the most meaningful results were chosen.In Fig. @ref(fig:ORA-plot), 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. @ref(fig:ORA-plot)A), 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. @ref(fig:ORA-plot)B) 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. @ref(fig:ORA-plot)C) 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. @ref(fig:ORA-plot)D) 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. @ref(fig:ORA-plot)E) 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. @ref(fig:ORA-plot)F) continuously decrease in expression from week 5 onward. For the stomach (Fig. @ref(fig:ORA-plot)G), 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. @ref(fig:ORA-plot)H) all displayed a constant rise in expression levels from week 5 to 9.

# PCA

The first three PCAs were now plotted against each other. This helps to analyze the genes contributions to the PCAs. The more parallel a gene's vector is to the PCA axis, the more the gene contributes to that PCA. The lenght of the vectors shows how much variability of the gene is explained by the PCAs. The longer the vector, the better it is represented in this dimension. The angels between the vectors give insight about the correlation between the genes. Small angels demonstrate a high correlation whereas opposite angles demonstrate a high negative correlation.

par

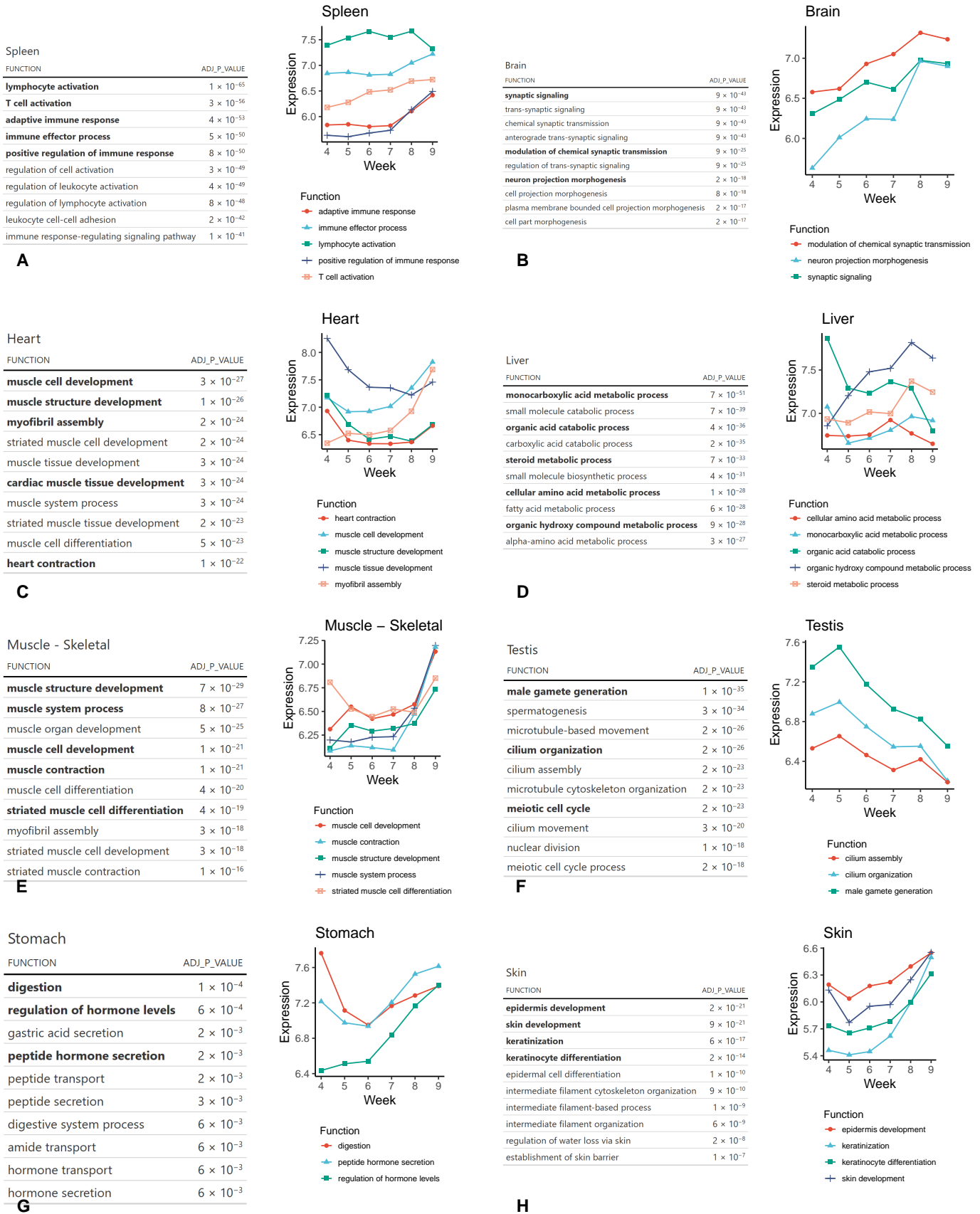# Main functions from overrepresentation analysis plotted by tissue

### Spleen

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

**A**

Spleen plot — Expression vs Week (4–9)

Function:
- adaptive immune response
- immune effector process
- lymphocyte activation
- positive regulation of immune response
- T cell activation

### Brain

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

**B**

Brain plot — Expression vs Week (4–9)

Function:
- modulation of chemical synaptic transmission
- neuron projection morphogenesis
- synaptic signaling

### Heart

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

**C**

Heart plot — Expression vs Week (4–9)

Function:
- heart contraction
- muscle cell development
- muscle structure development
- muscle tissue development
- myofibril assembly

### Liver

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

**D**

Liver plot — Expression vs Week (4–9)

Function:
- cellular amino acid metabolic process
- monocarboxylic acid metabolic process
- organic acid catabolic process
- organic hydroxy compound metabolic process
- steroid metabolic process

### Muscle - Skeletal

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

**E**

Muscle – Skeletal plot — Expression vs Week (4–9)

Function:
- muscle cell development
- muscle contraction
- muscle structure development
- muscle system process
- striated muscle cell differentiation

### Testis

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

**F**

Testis plot — Expression vs Week (4–9)

Function:
- cilium assembly
- cilium organization
- male gamete generation

### Stomach

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

**G**

Stomach plot — Expression vs Week (4–9)

Function:
- digestion
- peptide hormone secretion
- regulation of hormone levels

### Skin

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

**H**

Skin plot — Expression vs Week (4–9)

Function:
- epidermis development
- keratinization
- keratinocyte differentiation
- skin development

***Figure 3:*** *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.*

# Discussion

## 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 genes. We compared both data subsets by k-means clustering and plotting the overlapping Intensity values (Figure @ref(fig:paper-trends-cluster)). 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 @ref(QC-normalization)).
Furthermore, we showed in our Venn diagram (Figure @ref(fig:Venn-Diagram)), that our method detected differentially expressed genes 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 genes 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 genes while also pointing out some flaws of the data set of the paper.

## TRAs can infer a basic timeline of organ development {dis-organ}

In our analysis, we have shown that a number of TRAs are differentially expressed (section @ref(organ-overview)) 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 @ref(organ-clustering)). 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 @ref(organ-clustering)) and function (analysis of spleen gene functions in section @ref(organ-tables)). 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 @ref(organ-ora) for eight different tissues. For the spleen, the results of our analysis (Fig. @ref(fig:ORA-plot)A) largely do not reflect the embryonic development (section @ref(intro-tissues)). 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 @ref(intro-tissues)). This is also visible in the expression data (Fig. @ref(fig:ORA-plot)B), 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. @ref(fig:ORA-plot)C) 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 @ref(intro-tissues)). 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. @ref(fig:ORA-plot)D). Thus, even though the liver forms mostly during the analyzed timeframe (section @ref(intro-tissues)), 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. @ref(fig:ORA-plot)E). As muscle fibers begin to develop later than week 9 and the first related proteins appear from week 7 on (section @ref(intro-tissues)), these expression data correspond well to the embryonic development.
The testis gene sets decrease in expression from week 5 onward (Fig. @ref(fig:ORA-plot)E). This is in contrast to the embryonic development, where the gonads start to form at around the same time (section @ref(intro-tissues)).
For the stomach, the expression pattern indicates a decrease until week 6 followed by rising expression levels until week 9 (Fig. @ref(fig:ORA-plot)G). 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 @ref(intro-tissues)). 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. @ref(fig:ORA-plot)H). This broadly reflects the embryonic development, with the epidermis starting to form in week 4 (section @ref(intro-tissues)). We also found this expression pattern in the keratinization gene set that is suggested by literature as a good indicator for skin formation.

## Brain subtissue-specific gene expression reflects neuronal development processes

First point to discuss is the significant increase in the gene expression of Ca(2+)ATPase (SLN sarcolipin) by factor 2 in logarithmic scale (Fig. @ref(fig:gene-expression-brain) left). A cause of this might be the process of neuronal migration which starts at week 9 (section @ref(intro-tissues)). Ca(2+) is an essential cofactor for actin dependent cell migration.

Another point is the strong correlation between the expression of SNAP91 genes and one SYBU gene (ENST00000276646) (Fig. @ref(fig:gene-expression-brain) middle). Both proteins were associated with the Cerebellar Hemisphere and contribute to endocytosis, which is essential for functional neurons and contributes to neuronal survival (Overhoff et al. 2020). One point to mention here is, that ENST00000521485 and ENST00000518312 were both associated with SNAP91, nevertheless to different isoforms. But both show identical correlation in expression, hence a failure in annotation might be possible.

In addition we identified a significant increase in expression of NXPH1 (Fig. @ref(fig:gene-expression-brain) middle), this refers to the NXPH1-promoted adhesion between axons and dendrites. An upregulation of this supplementary factor can prepare the process of synapse formation which starts at week 11 (section @ref(intro-tissues)).

We further identified a strong LSAMP expression associated to the putamen , and a NXPH1 associaated to Anterior cingulate cortex. (Fig. @ref(fig:gene-expression-brain) middle). The increased gene expression of LSAMP starts at week 6 and by NXPH1 between week 7 and 8. Tmight be a preparation for synapse formation at week 9, hence LSMP and NXPH1 are adhesion molecules in axon guidance (section @ref(intro-tissues)).

In addition, we identified that two transcripts associated to two different tissues show identical correlation, this can be caused by false annotations. Nevertheless, they show a strong downregulation of BDNF (Fig. @ref(fig:gene-expression-brain) middle). This factor normally promotes neuronal survival. A downregulation of BDNF can be a preparation for the phase of programmed cell death, which starts at week 20 in neuronal development (section @ref(intro-tissues)).

The neuronal cell migration is strongly dependent on chemoatractors like chemokines (Tiveron 2008). We identified a significant increase in CXCL14 between weeks 5 to 6 and a maintaining high expression level for the following weeks (Fig. @ref(fig:gene-expression-brain) right). This could be an accumulation for neuronal migration, starting at week 9 (section @ref(intro-tissues)).

A decline in NUP85 expression is noteable between week 5 to 9 (Fig. @ref(fig:gene-expression-brain) right). NUP85 can bind to CCR2, hence a decline in NUP85 genexpression reduce the chances for this binding. A consequence might be that more CCR2 receptors are free for beta type chemokin mediated signals.

## Hypothesis: Specific TRAs can be used as key biomarkers for the development of individual organs

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.

##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 @ref(limma)) and determined that our method of analysis even has advantages compared to the original paper (section @ref(paper-comp)). Then, we used these results to try to determine the developmental steps occurring during the observed time frame (section @ref(organ)). 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 @ref(dis-organ)). 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 @ref(brain)). Based on their expression, we could draw some conclusions on neuronal development (section @ref(dis-brain)). Finally, we tried to define individual genes as biomarkers for embryonic development through PCA. 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.