

# Test3

Paul Christmann

2022-07-12

## 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. Therefore, we analyzed the share of differentially expressed transcripts above a certain expression level over time, as shown in Fig. ???A. Furthermore, we observed trends within the median expression of all differentially expressed transcripts associated with a tissue (Fig. ???B). Since both metrics only showed in miniscule changes, we hypothesized that distinct, counteracting trends in expression existed 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. ???C.

## Normalising the data set

Intensity values of different chips are affected by sample preparation and array manufacturing and processing resulting in statistical variance and random fluctuation. To access the biological relevant variation the raw data needs to be transformed by normalisation. We chose the vsn rma normalization with its library *vsn* according to Huber *et al.* (2002).

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 continuously 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.

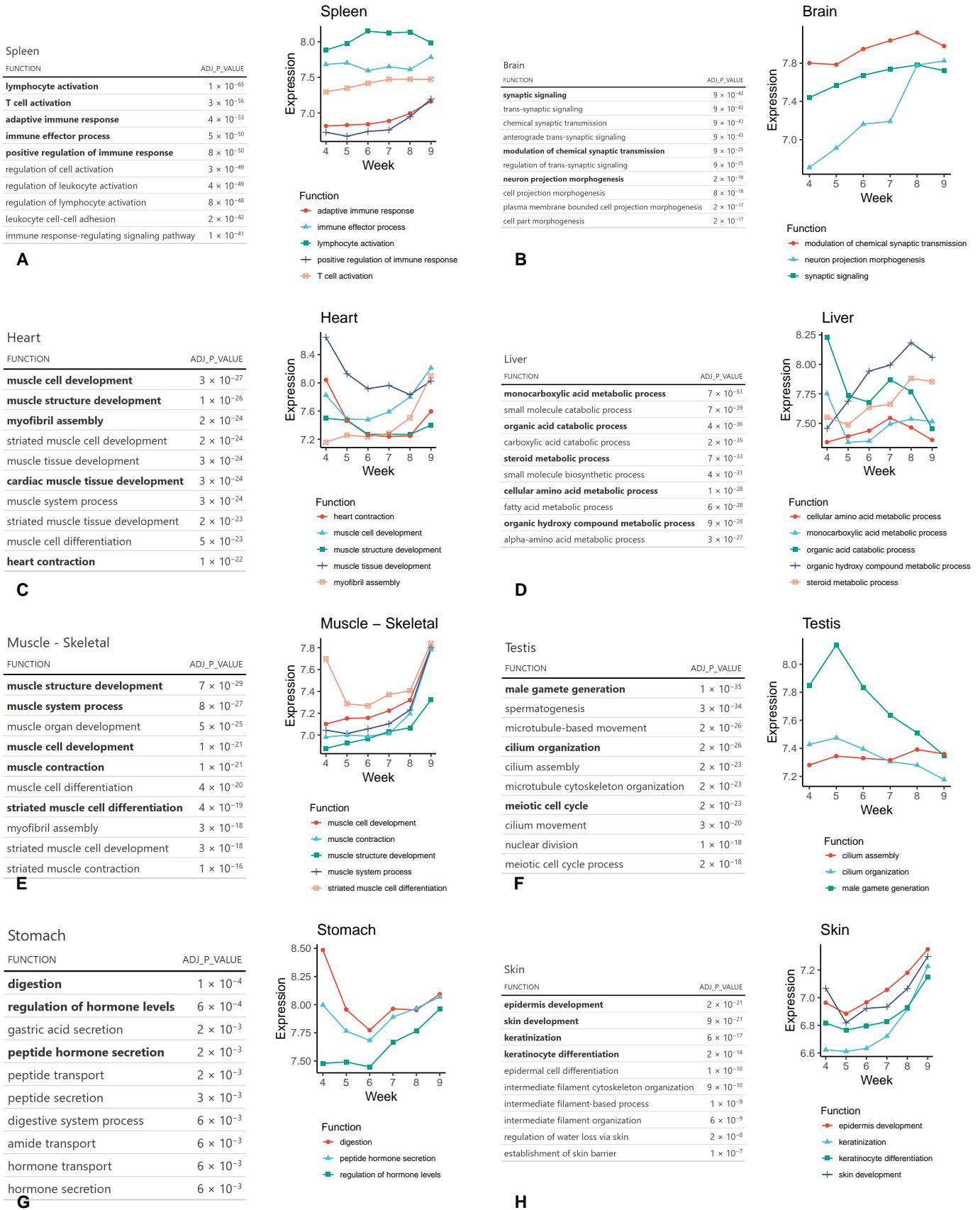
Finally, the skin-associated functions (Fig. @ref(fig:ORA-plot)H) all displayed a constant rise in expression levels from week 5 to 9.

## Discussion

### Hypothesis: TRAs can infer a timeline of organ development similar to the results by Yi et al. 2010

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)).

## Main functions from overrepresentation analysis plotted by tissue



**Figure 1:** 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 functions, 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.

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 by 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 from 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 continuously 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 occur at this time, since the specific cells needed for this only appear later in embryogenesis (section @ref(intro-tissues)). Therefore, 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 through 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.

## Dataset

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

### Affymetrix U133 plus 2.0 human GeneChip array

The data was generated from embryos by using Affymetrix U133 plus 2.0 human GeneChip arrays. RNA microarrays are slides coated with oligonucleotides as matrices which screen for thousands of transcripts. The HG-U133 Plus 2.0 allows the detection of about 50,000 transcripts and uses quality control matrices. The Affymetrix chip includes 62 control transcripts, whose intensities are imported together with the acquired data.

## Importing the data set

We downloaded the raw data to a local harddrive from the Gene Expression Omnibus with the Accession Number of GSE15744. We imported it with the help of the library *affy* and is connected to the correct Annotation by the *brainarray* package. The *affy* package allows more manageable data analysis and manipulation of microarray intensity values.