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

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.

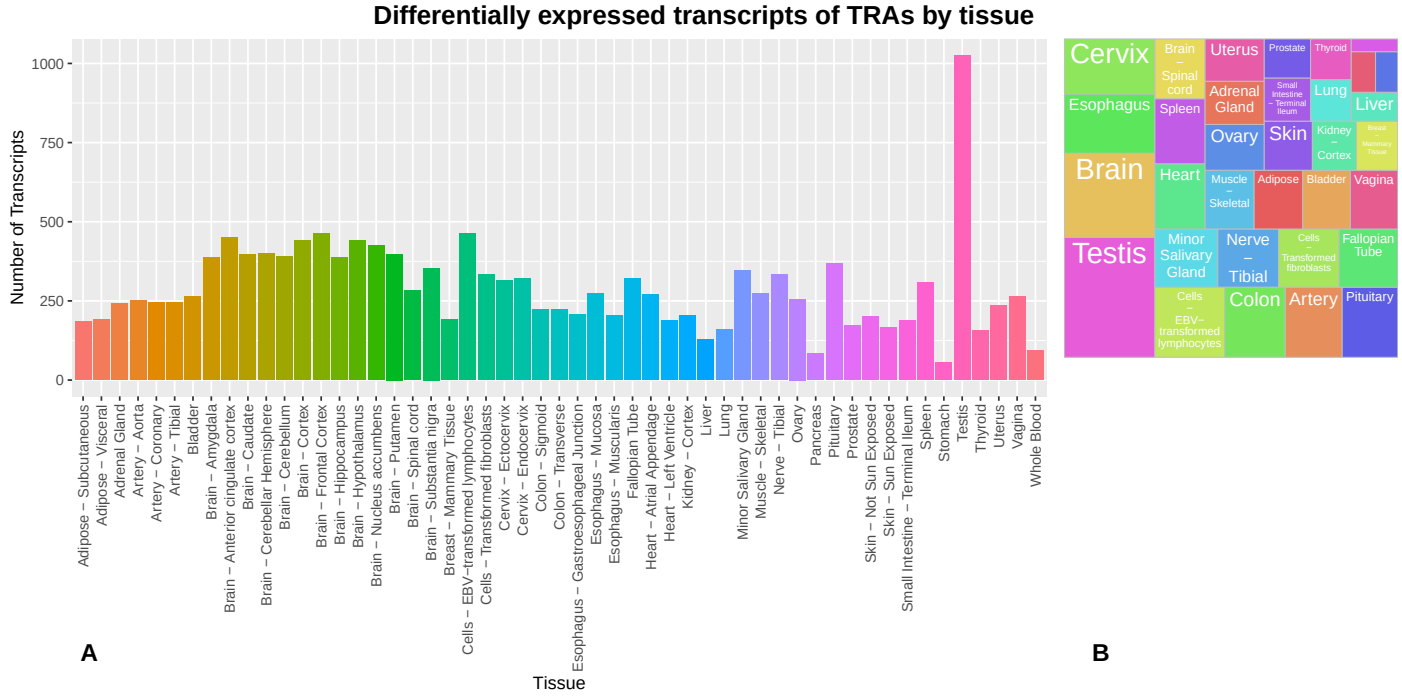


Figure 1: *A. The number of transcripts associated with each tissue, including subtissues, is displayed. B. The plot shows the share of TRAs associated with each tissue. Subtissues are subsumed under their main tissue.*

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 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).

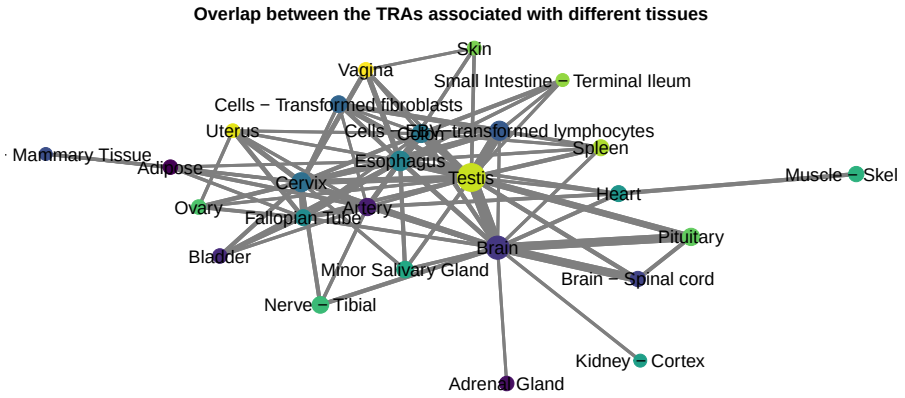


Figure 2: *Each tissue is displayed as a node with its size representing the number of transcripts associated with it. The edges show the shared TRAs between the linked tissues, with only links corresponding to more than 100 common TRAs visible.*

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.

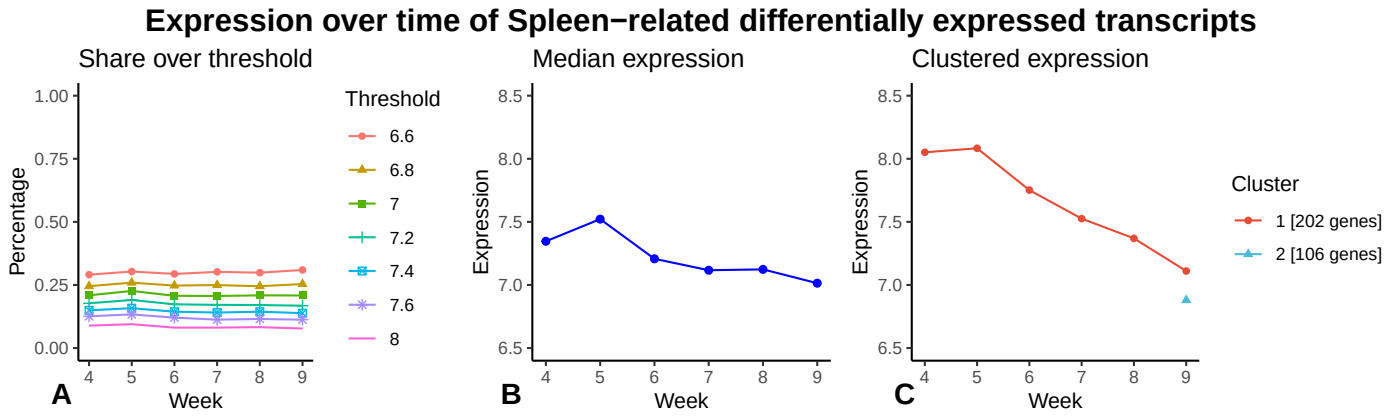


Figure 3: *A. For different expression thresholds the share of differentially expressed transcripts with higher expressions than the threshold is depicted. B. The median expression for all spleen-associated TRAs is shown for each point in time. C. For k-means clustring, the silhouette score determined an optimum of two clusters. The median expression analog to B is plottet for each of these clusters.*

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.

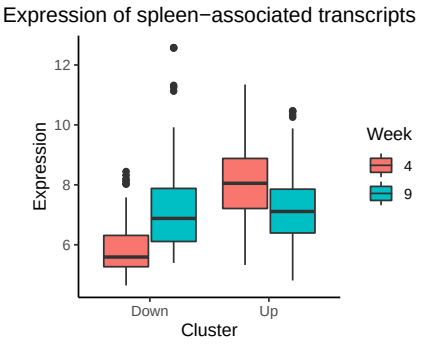


Figure 4: *A further look on the expression of transcripts in the up- and downregulated clusters shows that the upregulated transcripts are close to the minimum expression level between 6 and 7 in week 4 and showing expressions between 7 and 8.5 by week 9. In contrast, the downregulated genes have very 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.*

As shown in 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 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.

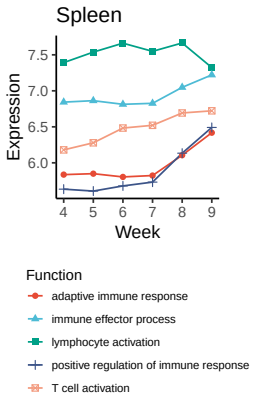
For the spleen (Fig. @ref(fig:ORA-plot)A), we determined a largely constant expression of immune-related genes throughout the time frame, with a slight increase in some functions from week 7-9. The brain (Fig. @ref(fig:ORA-plot)B) showed an increase in neuron projection morphogenesis from a previously 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.

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}

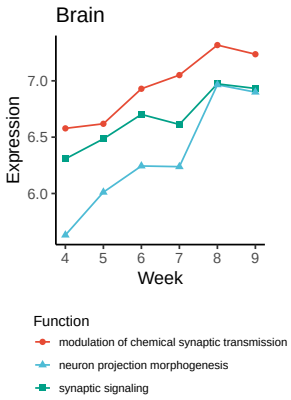
A



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}

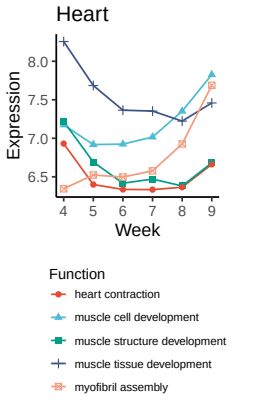
B



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}

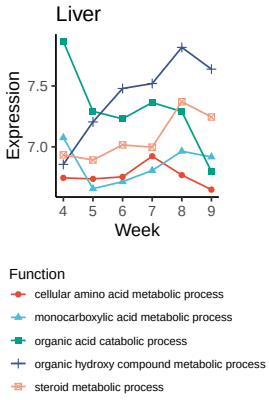
C



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

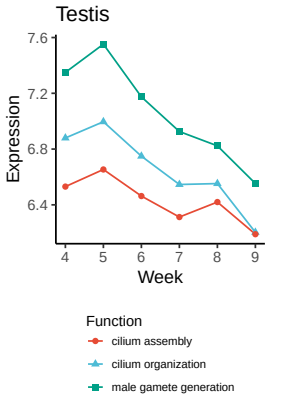
FUNCTION	ADJ_P_VALUE
muscle structure development	7×10^{-29}
muscle system process	8×10^{-27}
muscle organ development	5×10^{-25}
muscle cell development	1×10^{-21}
muscle contraction	1×10^{-21}
muscle cell differentiation	4×10^{-20}
striated muscle cell differentiation	4×10^{-19}
myofibril assembly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

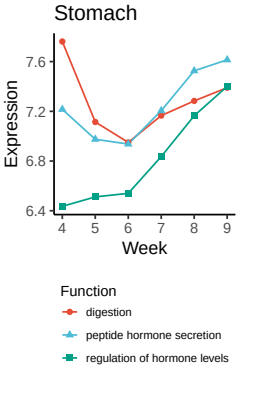
F



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

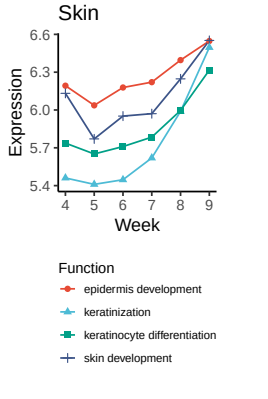


Figure 5: *For eight tissues A to H, the tables show the 10 most significant results from overrepresentation analysis. There, the functional GO annotation of all transcripts in our dataset (not only the differentially expressed ones) was compared to the functions associated with all TRAs of one tissue. 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.*