Final Report

Gruppe 4

2022-07-03

Contents

0.1 TRAs can infer a basic timeline of organ development

0.1.1 Differentially expressed transcripts can be linked to all analyzed tissues

The TRA Data covers 53 distinct tissues. For all of those, we found at least 40 differentially expressed transcripts within our dataset. The minimum was found with 46 stomach-linked transcripts, the maximum were 837 TRAs for the testes.

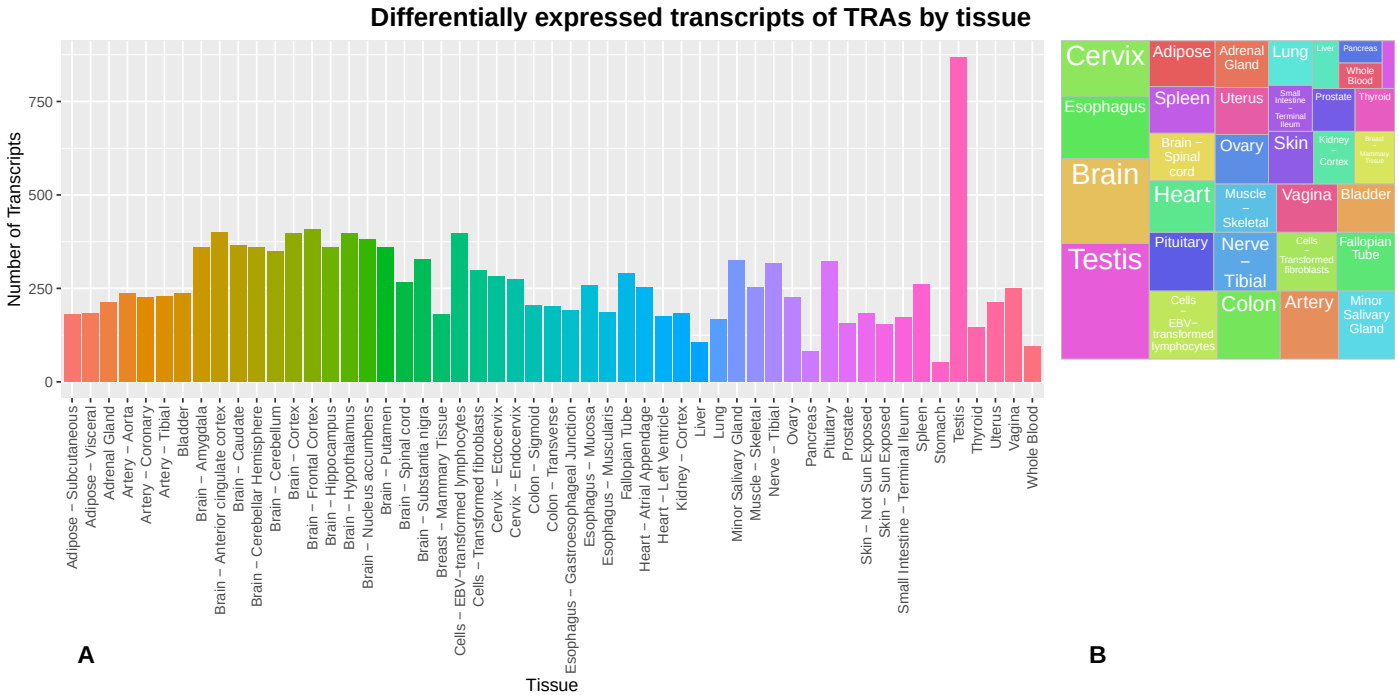


Figure 1: *A. The number of transcripts associated with each tissue, including subtissues, is displayed. B. The share of TRAs associated with each tissue. Subtissues are subsumed under their main tissue.*

These numbers are sufficient for further analysis of the gene expression within individual tissues. Therefore, all further analysis 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 further illustrated by Fig. 2. Furthermore, a detailed heatmap for the shared TRAs between tissues is shown in the supplementary Fig. 6.

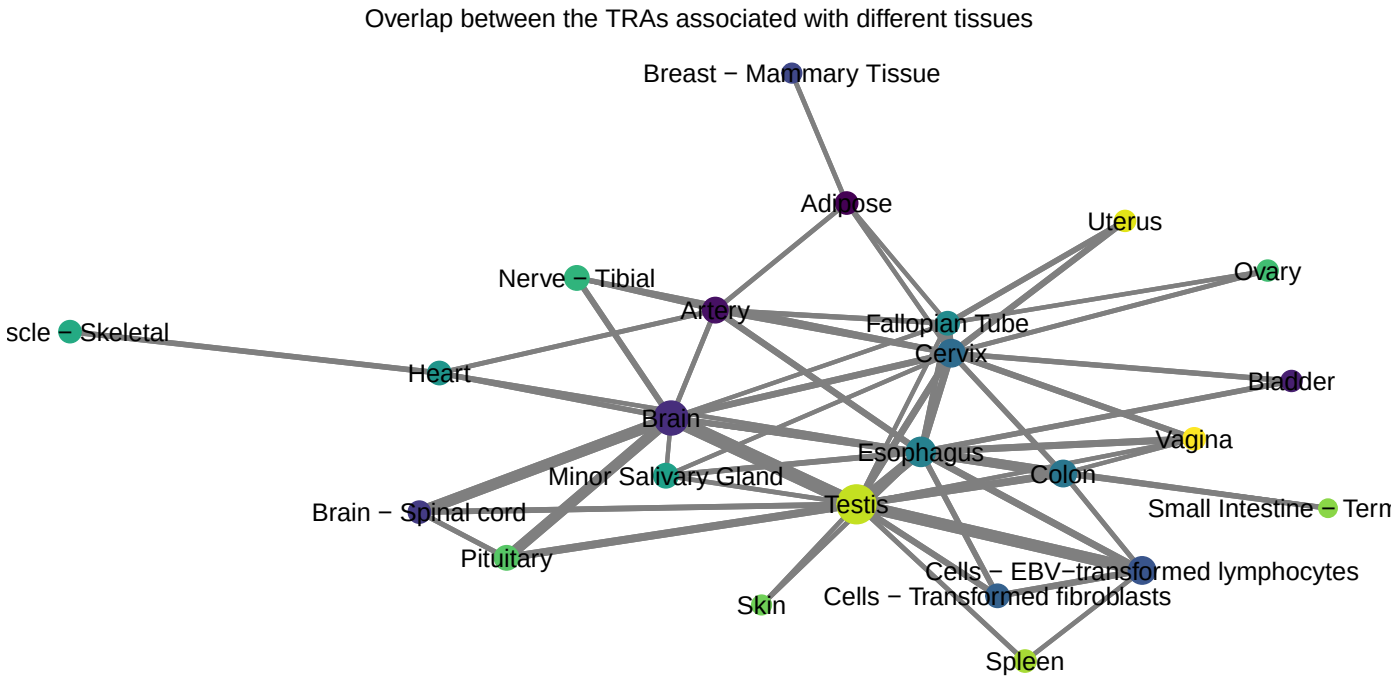


Figure 2: *Each tissue is displayed as a node with its size representing the number of transcripts associated with it. The edges show the shared TRAs between the linked tissues, with only links corresponding to more than 100 common TRAs visible.*

0.1.2 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. Therefore, we analyzed the share of differentially expressed transcripts above a certain expression level over time, as shown in Fig. 3A. Furthermore, we observed trends within the median expression of all differentially expressed transcripts associated with a tissue (Fig. 3B). Since both metrics only showed in miniscule changes, we hypothesized that distinct, counteracting trends in expression exsisted 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. 3C.

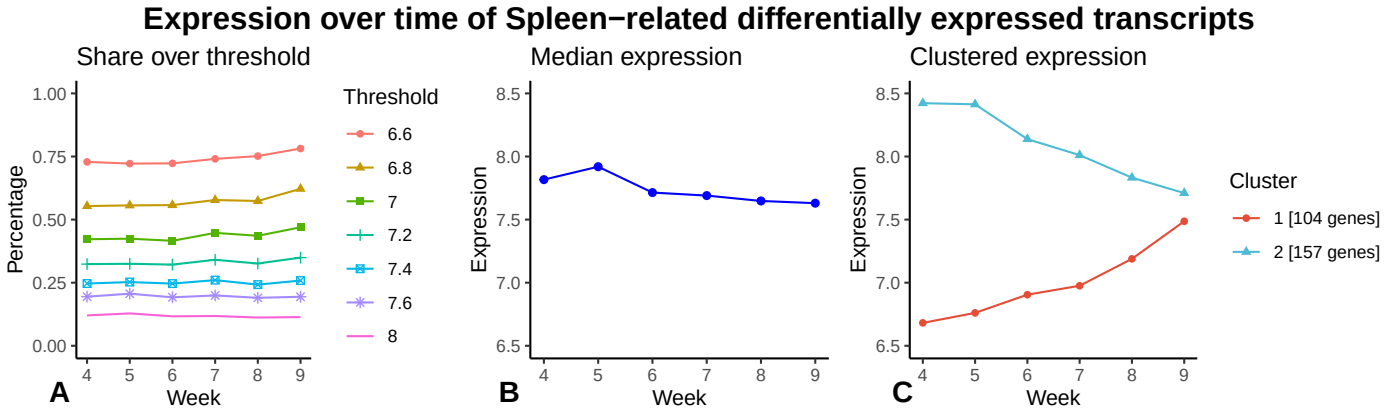


Figure 3: *A. For different expression thresholds the share of differentially expressed transcripts with higher expressions than the threshold is depicted. B. The median expression for all spleen-associated TRAs is shown for each point in time. C. For k-means clustring, the silhouette score determined 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 each be characterized as either an upregulation or a downregulation. In order to analyze the indications for organ development, we analyzed the functions of the transcripts belonging to the two clusters.

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

For all differentially expressed spleen-associated transcripts, we used the NCBI gene database to get a functional annotation.

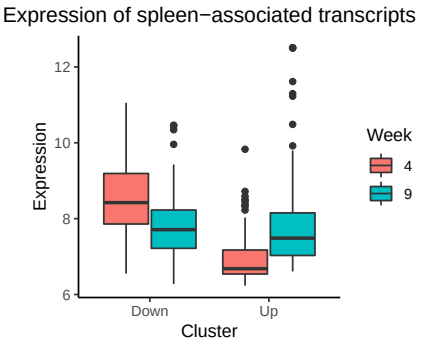


Figure 4: *A further look on the expression of transcripts in the up- and downregulated clusters shows that the upregulated transcripts are close to the minimum expression level between 6 and 7 in week 4 and showing expressions between 7 and 8.5 by week 9. In contrast, the downregulated genes have very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

0.1.4 Overrepresentation Analysis can create plots that signify organ development

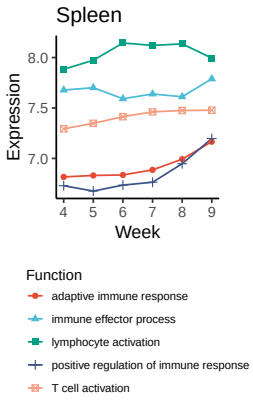
For this analysis, the eight tissues with the most meaningful results were chosen.In Fig. @ref(fig: ORA-plot), the most important functions for these tissue were determined through overrepresentation analysis. In addition, the Expression of the associated transcripts was plotted.

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}

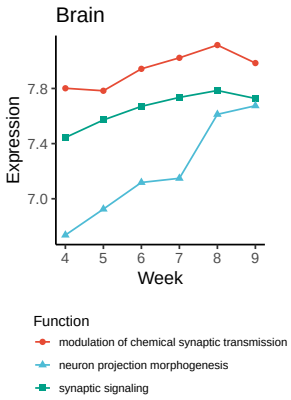
A



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}

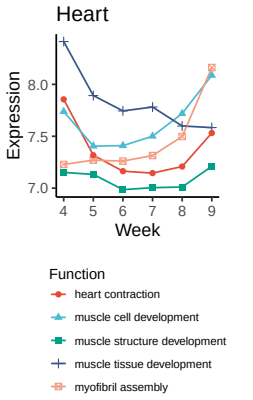
B



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}

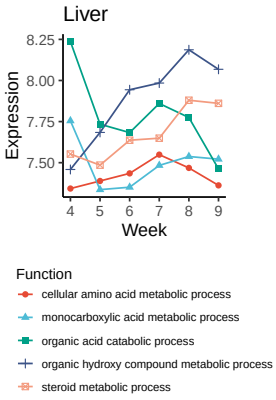
C



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}

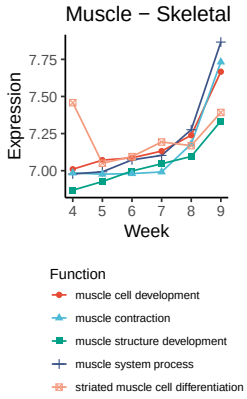
D



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}

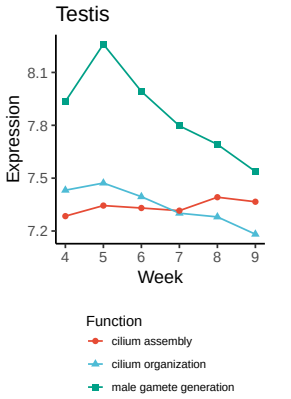
E



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}

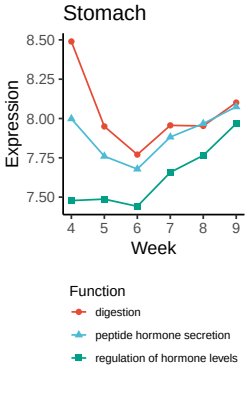
F



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

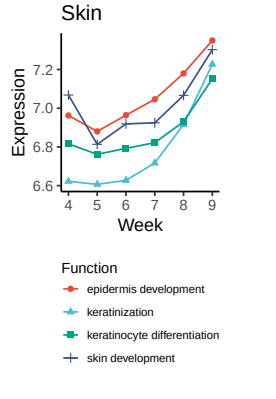


Figure 5: *For eight tissues A to H, the tables show the 10 most significant results from overrepresentation analysis. There, the functional GO annotation of all transcripts in our dataset (not only the differentially expressed ones) was compared to the functions associated with all TRAs of one tissue.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

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.

1 Supplementary

1.1 Organ development

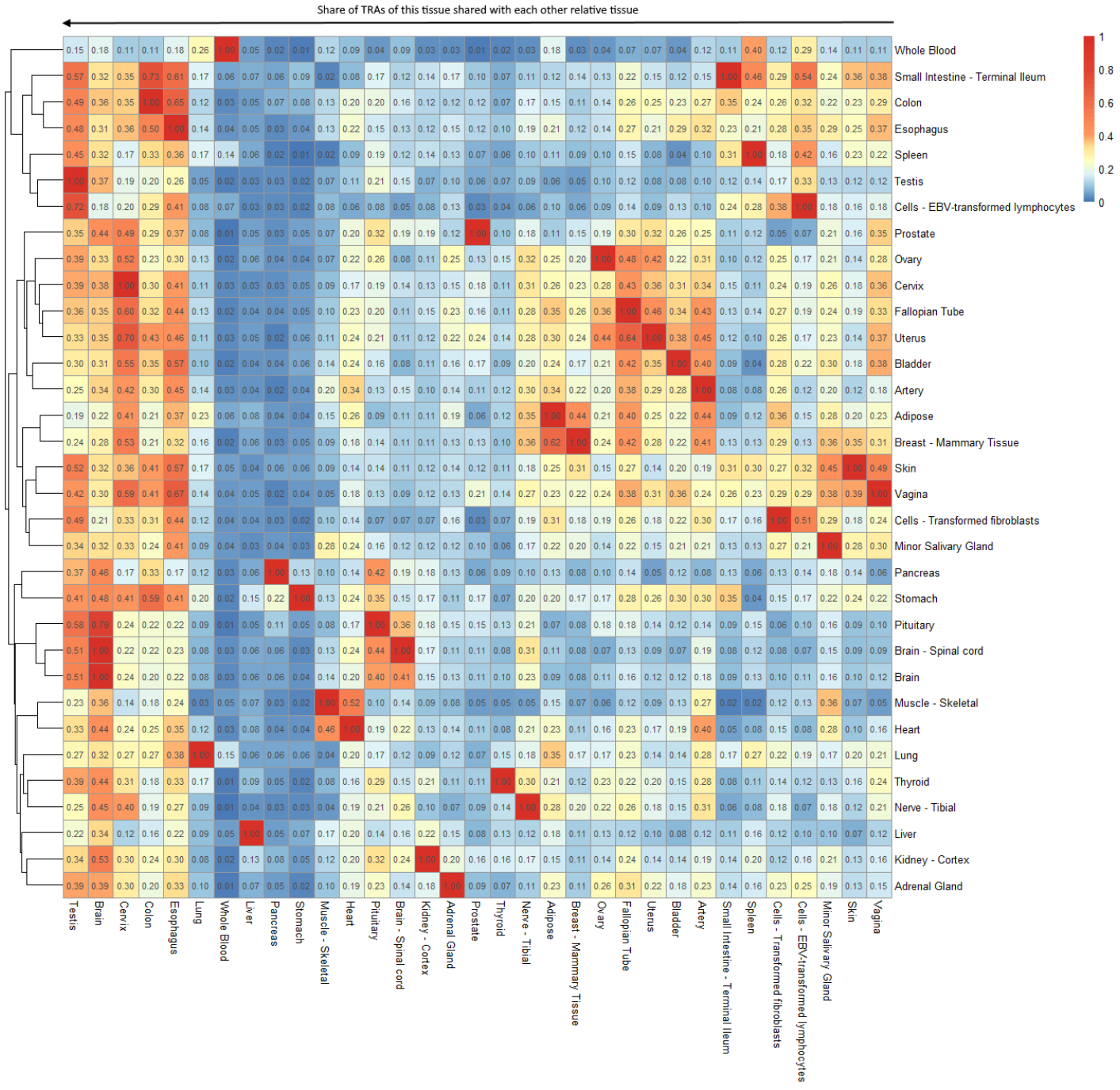


Figure 6: *For each of the tissues on the right, the share of its TRAs that are also associated with the tissue in the respective column is displayed.*

Spleen - Upregulated

List of differentially epxressed genes

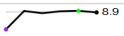
TRANSCRIPT	GENE	PROTEIN	SUMMARY	EXPRESSION_OVER_TIME
ENST00000251296	IGSF21	immunoglobin superfamily member 21	Member of the immunoglobulin superfamily	
ENST00000259698	RIPOR2	RHO family interacting cell polarization regulator 2	Mediates polarization of T cells and neutrophils	
ENST00000259989	FGFBP2	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 7: *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	7.6
ENST00000240423	NCAPH	non-SMC condensin I complex subunit H	Interphase chromosome condensation	7.1
ENST00000251496	NCAPG	non-SMC condensin I complex subunit G	Chromosome condensation and stabilization	7.9
ENST00000260359	NUSAP1	nucleolar and spindle associated protein 1	Spindle microtubule organization	8.6
ENST00000260731	KIF11	kinesin family member 11	Spindle dynamics	8.3
ENST00000287598	BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	Spindle checkpoint function	8.1
ENST00000300093	PLK1	polo like kinase 1	Highly expressed during mitosis	7.7
ENST00000302759	BUB1	BUB1 mitotic checkpoint serine/threonine kinase	Central role in mitosis	7.3
ENST00000306324	HOXD4	homeobox D4		6.8
ENST00000310955	CDC20	cell division cycle 20	Nuclear movement in the cell cycle	8.3
ENST00000313288	TACC3	transforming acidic coiled-coil containing protein 3	Stabilization of the mitotic spindle	7.7
ENST00000327331	CDCA8	cell division cycle associated 8	Regulator of mitosis	7.8
ENST00000335658	DPPA4	developmental pluripotency associated 4	Stem cell pluripotency	6.7
ENST00000348459	HELLS	helicase, lymphoid specific	Involved with cellular proliferation	7.2
ENST00000348581	EXO1	exonuclease 1	Mismatch repair	7.5
ENST00000349718	CMTM7	CKLF like MARVEL transmembrane domain containing 7	Regulates G1/S transition	7.6
ENST00000352331	KIF23	kinesin family member 23	Moves chromosomes	7.7
ENST00000352433	PTTG1	PTTG1 regulator of sister chromatid separation, securin	Anaphase-promoting complex substrate	9.4
ENST00000355112	BLM	BLM RecQ like helicase	Unwinds Holliday junction	7.2
ENST00000357481	ACIN1	apoptotic chromatin condensation inducer 1	Induces apoptotic chromatin condensation	7.9
ENST00000359803	MDK	midkine	Promotes cell growth	9.4
ENST00000361953	FOXM1	forkhead box M1	Transcriptional activator in proliferation	7.6
ENST00000366955	CENPF	centromere protein F	Associates with centromer-kinetochore complex	8.2
ENST00000371566	ORC1	origin recognition complex subunit 1	Initiation of DNA replication	7.0
ENST00000372462	CDC20	cell division cycle 20	Regulatory in the cell cycle	8.3
ENST00000374403	KIF4A	kinesin family member 4A	Maintaining chromosome integrity in mitosis	8.1
ENST00000380026	CENPE	centromere protein E	Stable spindle microtubule capture	8.1
ENST00000380756	DNMT3A	DNA methyltransferase 3 alpha	CpG methylation	7.7
ENST00000382011	MCM5	minichromosome maintenance complex component 5	Initiation of DNA replication	8.1
ENST00000395284	CDK1	cyclin dependent kinase 1	Subunit of M-phase promoting factor	8.7
ENST00000395405	ZWINT	ZW10 interacting kinetochore protein	Involved in kinetochore function	8.9
ENST00000396863	MYBL2	MYB proto-oncogene like 2	Cell cycle progression	7.6
ENST00000404276	CHEK2	checkpoint kinase 2	Cell cycle checkpoint regulator	7.4
ENST00000412359	BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	Spindle checkpoint function	8.1
ENST00000414849	NUSAP1	nucleolar and spindle associated protein 1	Spindle microtubule organisation	8.7
ENST00000420246	TP53	tumor protein p53	Inducing cell cycle arrest	7.3
ENST00000427946	NCAPH	non-SMC condensin I complex subunit H	Subunit of condensin complex	7.1
ENST00000428239	CDC7	cell division cycle 7	Cell division cycle protein	8.0
ENST00000433496	RFC4	replication factor C subunit 4	DNA polymerase accessory	8.3
ENST00000515001	CENPH	centromere protein H	Centromere-microtubule complex	7.6
ENST00000518483	EXO1	exonuclease 1	Mismatch repair	7.5
ENST00000520452	PTTG1	PTTG1 regulator of sister chromatid separation, securin	Anaphase-promoting complex substrate	9.4
ENST00000534871	AURKB	aurora kinase B	Segregation of chromosomes	7.7
ENST00000535254	BUB1	BUB1 mitotic checkpoint serine/threonine kinase	Central role in mitosis	7.3
ENST00000561208	MEIS2	Meis homeobox 2	Contributor to developmental programs	9.3

Figure 8: *Table of all spleen-associated genes from the downregulated cluster with a function related to the cell cycle of cellular division*