Test3

Paul Christmann

2022-07-12

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 doi: 10.1093/nar/gkv007. 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 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 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 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.

Hypothesis: Neural TRA expression patterns reflect morphological brain development

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 , which is associated to the putamen (Fig. @ref(fig:gene-expression-brain) middle). The recognized increase in gene expression of LSAMP starting at week 6 might be a preparation for synapse formation at week 9, hence LSMP is an adhesion molecule in axon guidance (section @ref(intro-tissues)).
We further 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.

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.

##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.