PROJECT 2 – SARS-MICE

Discussion part

HUYNH Minh-Anh

Sorbonne Université

UE 5BM578 “Workshop in immunological data analysis”

Discussion

Overall, the analyses can be split into two themes, an analysis from the perspective of all strains and an analysis from the perspective of time points. Analysis of all strains reveals that only the SARS-WT strains have a similar gene expression phenotype, and that the other four conditions (SARS-icSARS, SARS-WT, SARS-BatSRBD, mock) possess a similar expression phenotype, through the analysis obtained from the heatmap and the MDS representation.

These results seem surprising since one could possibly expect the icSARS-CoV and Bat-SRBD strains to have a similar phenotype to SARS-WT. The current Bat-SRBD strain could be explained by the fact that the researchers wanted to modify the spike protein of interest to perhaps make the strain non-pathogenic (approaching a mock phenotype). The icSARS-CoV strain is, by definition, a modified clone of the SARS-WT strain, which would also have lost its pathogenicity, even though we do not know why due to a lack of context, since it approaches the mock phenotype. The analyses should therefore not only identify the differences between the major groups but could also potentially identify the reasons why both the SARS-BatSRBD and SARS-icSARS strains were able to achieve a mock phenotype.

A limitation of the large-scale transcriptomic analyses we performed is the lack of possibility to see in detail the differences associated with each of these strains. Indeed, subtle gene expression relating to specific pathways would be lost when compared to huge pathways having more data and thus being more easily identifiable as more genes are co-expressed. Moreover, gene co-expression does not necessarily correlate with significant functional associations, as many genes in the same pathway do not have similar transcript profiles. Novel functional associations are thus harder to discover (Uygun et al., 2016). Therefore, it is important to explore rather exhaustively multiple dataset combinations, similarity measures, clustering algorithms and parameters, and thus validate our own results with additional analysis.

Notably, the over-expressed genes highlighted in the volcano plot **(Figure 2)** could signal over-expressed metabolic pathways that could be identified on Reactome in further precise analysis, and thus be able to have another observation perspective in which metabolic pathways were left untouched for the non SARS-WT condition and see in which way they differ. Indeed, we can see that even though both SARS-BatSRBD and SARS-icSARS are similar to mock in gene expression, hierarchical clustering by the dendrogram of the heatmap **(Figure 4)** could still group SARS-icSARS together, suggesting that this group slightly differs from SARS-BatSRBD even in a minor way. The present analysis has identified that only CXCL9 and CXCL10 were differentially expressed SARS-icSARS mice *versus* mock condition. Those cytokines form the CXCL9, -10, -11/CXCR3 axis (Tokunaga et al., 2018) which is known to regulate immune cell migration, differentiation, and activation. Particularly, they induce Th1 polarization and activates the immune cells in response to IFN-γ.

Interestingly, SARS-BatSRBD and the mock group are not perfectly orderly clusterized, and thus we could hypothesize that SARS-BatSRBD is a strain most similar to the mock condition, and not having a significatively different impact from mock phenotype, which is evidenced by the fact that no differential gene expression were found significant, compared to mock.

It would therefore be relevant to propose an additional analysis between BatSRBD, icSARS, and mock in order to confirm or refute the similarity of expression phenotypes, starting with an

MDS for example. Similarly, since the changes between strains may not be identical, the interactions between proteins may differ, and an analysis using a StringDB approach could clarify the differences.

Additionally, the outliers identified are strangely identical to the cluster they relate to, suggesting that they are not a biological outlier but an experimental one, which possibly had a mislabeled group.

Concerning the comparison of the response to SARS-WT strain at various time points, data analysis suggests that overall, the response to SARS-WT infection is of a classic immune response consisting of NK cell and T cell lymphocyte infiltration with chemotaxis markers, along with various pro-inflammatory cytokines, especially in early (D1 and D2) condition involving Th1 cell and Th17 polarization, along with some Th2 polarization-associated cytokines. Those genes start to become less activated the longer the infection gone which seems to make sense.

Possible limitations of our current analysis involve the fact that we did not explore the obtained gene lists furthermore into specific pathway analysis which would have given additional perspective, and that we did not use StringDB, thus some supplementary research still remain possible, given additional perspective.

References

Tokunaga, R., Zhang, W., Naseem, M., Puccini, A., Berger, M.D., Soni, S., McSkane, M., Baba, H., and Lenz, H.-J. (2018). CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation – A target for novel cancer therapy. Cancer Treat Rev *63*, 40–47.

Uygun, S., Peng, C., Lehti-Shiu, M.D., Last, R.L., and Shiu, S.-H. (2016). Utility and Limitations of Using Gene Expression Data to Identify Functional Associations. Plos Comput Biol *12*, e1005244.