

# Test4

Paul Christmann

2022-07-12

# Using Principal component analysis to find biomarkes for developmental progression

Principal component analysis (PCA) was performed on a matrix containing all differentially expressed genes throughout the 6 weeks. This was performed to reduce dimension while keeping most of the data’s variance. To get a better grasp of how much variance is explained by each principal component (PC), the percentage variance of each PC together as well as the cumulative variance were plotted. This visualization helps to determine how many PCs are needed to explain a significant amount of the data’s variation. The chosen PCs can than be used for further analysis. The first three PCs already explain over 80% of the data’s variance, with the first PC alone accounting for over 60% (Fig. @ref(fig:PCA\_analysis) A).

For further analysis, each transcript was awarded a rank depending on how much it contributes to each PC. Then, a score was calculated for maximum contribution to PC1 and minimal loadings for all other relevant PCs (PC2-5). We took a closer look at the five best socring transcripts for PC1, as this PC alone already explains the brunt of the variance of our data. These transcripts encode for three different proteins: Low-Density Lipoprotein Receptor Class A Domain-Containing Protein 4 (LDLRAD4) encoded by ENST00000399848, Calcium-activated potassium channel subunit beta-2 (KCNMB2) encoded by ENST00000432997 and ENST00000452583 and SLIT And NTRK-Like Protein 3 (SLITRK3) encoded by ENST00000241274 and ENST00000475390.

The transcripts of ENST00000432997 and ENST00000452583 encode for the same Protein (KCNMB2). Fitting to that, both transcripts consistently display the same expression level over all weeks. The same applies to the transcripts of ENST00000241274 and ENST00000475390, as they also encode for the same protein (SLITRK3) and therefore also display the same expression levels. The transcripts for LDLRAD4 show the highest overall expression levels. Their expression pattern also does not display many similarities to those of KCNMB2 and SLITRK3. However, KCNMB2 and SLITRK3 show a highl similar curve over all weeks (Fig. @ref(fig:PCA\_analysis) B).

Principal component analysis

