Mice Exercise Experiment – Insights using multidimensional multiomic analysis

Bivariate analysis

In transcriptomics, differential expression analysis provides unidimensional insights, examining the separation and distribution of gene expression changes one gene at a time. While useful, this approach overlooks the complex interactions between genes. To push differential expression one step further, we plan to explore multidimensional genetic relationships by analyzing gene pairs. First, we use a preliminary feature selection technique to reduce greatly the number of genes of interest to run a more in-depth analysis, two genes at a time. This reduces the number of genes of interest from over 30,000 down to a few hundred. In this experiment, we apply Relief to select 100 genes of interest and combine those 100 genes in pairs, generating 4950 gene pairs that can be used as features to train 2D Support Vector Machines (SVMs). We repeated the same pipeline to the proteomics data extracted from the muscle tissue and the proteomics data extracted from the mice serum, as well as assessing four different mice groups comparison, being I. NE PBS vs E PBS, II. NE AmAc vs E AmAc, III. NE PBS vs E AmAc, and IV. E PBS vs E AmAc. We stored the scores in the file "Bivariate_Results_TTU.xlsx" file and summarized the findings below.

I. NE PBS vs E PBS

The genes that differentiate between the NE PBS group and E PBS mice groups are many. The top 3 ranked in the transcriptomic data are Hoxc5, 1110065P20Rik and Vkorc1. Some gene pairs have shown differentiability between classes without great standalone performance, unable to differentiate the groups individually, but reaching perfect separation when paired. One of these pairs is Cd209a and Ccdc85b (Figure 1), in "Figures/NEPBSvsEPBS/RNA/7_Cd209a_Ccdc85b.png". Other figures with such behavior are:

- /NEPBSvsEPBS/RNA/4_Gtf2h3_Mta2
- /NEPBSvsEPBS/RNA/10_Megf6_Pigx
- /NEPBSvsEPBS/RNA/14_Gm2a_Sp100

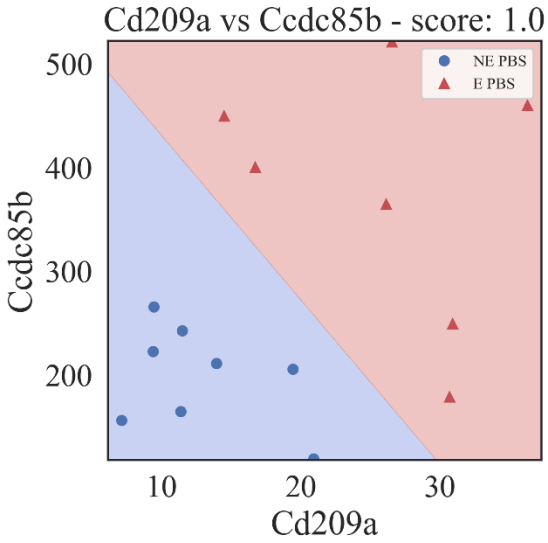

Figure 1. In transcriptomics, several gene pairs are able to separate NE PBS from E PBS mice when paired, one such example is Ccdc85b and Cd209a.

In the tissue proteomics data, the top 3 protein-coding genes found to differentiate well between NE PBS and E PBS are Ilf2, Smyd2 and Hgs. Other promising pairs are:

- NEPBSvsEPBS/PROTTISSUE/3_Ilf2_Myo1c
- NEPBSvsEPBS/PROTTISSUE/7_Obscn_Adss1
- NEPBSvsEPBS/PROTTISSUE/11_Col5a2_Kng1

In serum proteomics, the top 3 protein-coding genes were Hrg, Itgam and F9. Out of the potentially interesting gene pairs, we have:

- NEPBSvsEPBS/PROTSERUM/1_Itgam_Serpinf2
- NEPBSvsEPBS/PROTSERUM/3_Amy1_Sod3

II. NE AmAc vs E AmAc

The genes and protein highlighted in this section represent changes from VWR exercise under hyperammonemia.

We'd hoped the same genes responsible for separating NE PBS vs E PBS would also be responsible for separating NE AmAc vs E AmAc, which would give us some insight into the effects of exercise under different treatments, however, such commonalities were not observed.

In transcriptomics, we have Ccnk, Slc2a3 and Dmxl2 as our top ranked genes, with pairs such as Tmed5 and Hbb-bs (Figure 2) providing a perfect separation between groups when paired, far better than the classification score obtained individually. Other gene pairs with this behavior are:

- /NEAmAcvsEAmAc/RNA/5_Rnf157_C1qtnf9
- /NEAmAcvsEAmAc/RNA/7_Rnf157_Gdi2
- /NEAmAcvsEAmAc/RNA/10_Fam122a_Gdi2

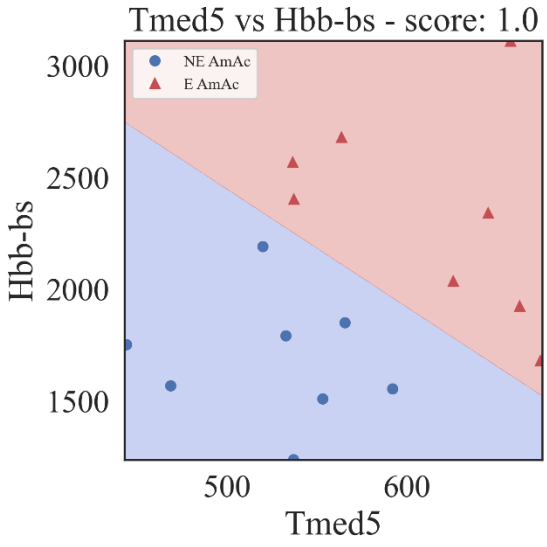

Figure 2. In transcriptomics, the genes Tmed5 and Hbb-bs are able to differentiate the NE AmAc and E AmAc mice groups with 100% accuracy, better than the score obtained by Tmed5 alone (81%) and Hbb-bs (87%).

In tissue proteomics, the top 3 ranked protein-coding genes are Clpp, Stxbp3 and Cbr1. Several gene pairs are able to differentiate between NE AmAc and E AmAc. These gene pairs offer better separability between groups when paired:

- NEAmAcvsEAmAc/PROTTISSUE/1_Otub1_Manf
- NEAmAcvsEAmAc/PROTTISSUE/2_Atp5f1c_Grpel1
- NEAmAcvsEAmAc/PROTTISSUE/6_Coq6_Flnc

In serum proteomics, the top 3 ranked protein-coding genes are Ltf, Chga and Selenop. Interestingly, these genes have very subtle fold changes within the NE AmAc and E AmAc groups, with Ltf at -0.07 log2FC, Chga at -0.13 log2FC and Selenop also at -0.13 log2FC. Figure 3 shows the pairing of Ltf with Psmb3, which are very subtly differentially expressed, but can separate the NE AmAc vs E AmAc. Other gene pairs worth considering are:

- NEAmAcvsEAmAc/PROTSERUM/1_Ltf_Psmb3
- NEAmAcvsEAmAc/PROTSERUM/2_Kng1_Pla2g7
- NEAmAcvsEAmAc/PROTSERUM/11_Ctsd_Serpinf1

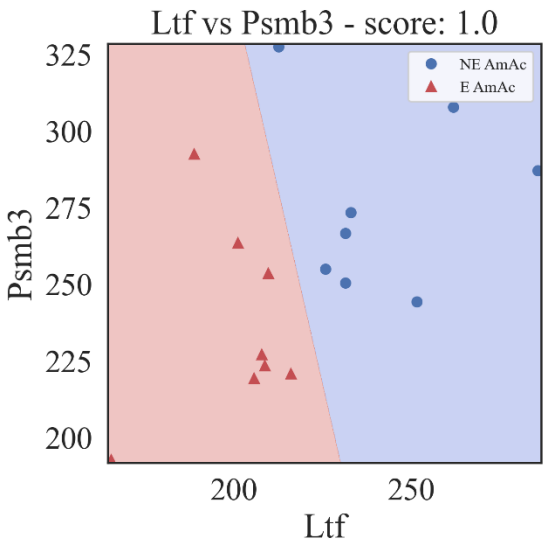

Figure 3. In serum proteomics, the genes Ltf and Psmb3 can separate the NE AmAc and E AmAc with a perfect 1.0 score, even though these genes aren't significantly differentially expressed.

III. NE PBS vs NE AmAc

Comparing the mice from the NE PBS treatment to the NE AmAc mice group shows the genetic changes caused from hyperammonemia. We'd hoped some of the genes and protein responsible for the changes in NE PBS vs NE AmAc would be somewhat the same for E PBS vs E AmAc, but these expectation were not met.

In transcriptomics, the top ranked genes were found to be Pdss1, Gm7049 and Hivep3. A lot of gene pairs obtain perfect accuracy, but most of the gene pairs we looked at provided little advantage over individual gene performance.

In tissue proteomics, the top ranked protein-coding genes were Acot9, Dazap1 and Cyc1. The gene pairs here obtain significantly higher scores when paired, such as the ones in figures:

- NEPBSvsNEAmAc/PROTTISSUE/1_Adhfe1_Ssb
- NEPBSvsNEAmAc/PROTTISSUE/2_Adhfe1_Srrm1
- NEPBSvsNEAmAc/PROTTISSUE/3_Dazap1_Acot9

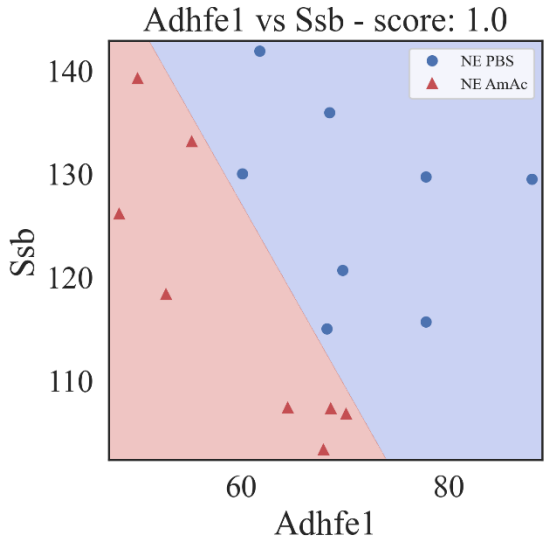

Figure 4. In tissue proteomics, the protein-coding genes Adhfe1 and Ssb are able to differentiate the NE PBS group from the NE AmAc with perfect 1.0 score, which is far better than the score obtained using each gene individually.

In serum proteomics, the top ranked protein-coding genes were Aldoa, Antxr2 and Adipoq. Similarly, pairing genes increases the cross-validation accuracy of our linear models significantly over univariate approaches. Some examples can be seen in figures:

- NEPBSvsNEAmAc/PROTSERUM/2_Antxr2_Psmb5
- NEPBSvsNEAmAc/PROTSERUM/17_Ext2_Fetub
- NEPBSvsNEAmAc/PROTSERUM/25_Plxdc2_Heg1

IV. E PBS vs E AmAc

In transcriptomics, the top genes implicated with hyperammonemia under VWR conditions were Gm48088, Zfp580 and Tmem88.

In tissue proteomics, the top protein-coding genes were Alad, Bcl2l13 and Mtx2.

In serum proteomics, the top protein-coding genes were Aldob, Hsp90ab1 and F7.

Multivariate Analysis

We employed Random Forests to explore higher-order genetic interactions, with the purpose of discovering multidimensional relationships within the omics data. Despite our efforts, the multi-modal methodology using random forests did not yield better performance than single-modality approaches. The code is in "Notebooks/Exercise_Experiments.ipynb".

Closing Remarks

Given the limited sample size of 8 mice per treatment group and the large amount of features that can be selected to separate the groups, it's expected that some of those features will separate the groups out of chance rather than actual biological changes in the mice organisms. Metrics such as False Discovery Rate can help track and identify these instances, however, it's advisable to gather more data before drawing conclusions.