

## Analysis Pipeline

This analysis pipeline script was written by Julián Candia and Ceereena Ubaida-Mohien to ensure the reproducibility of results reported in " Transcriptomics and proteomics of gastrocnemius muscle biopsies in peripheral artery disease" by L. Ferrucci, J. Candia, C. Ubaida-Mohien et al. For questions and/or comments, please contact Julián Candia ([julian.candia@nih.gov](mailto:julian.candia@nih.gov)), Ceereena Ubaida-Mohien ([ceereena.ubaida-mohien@nih.gov](mailto:ceereena.ubaida-mohien@nih.gov)) or Luigi Ferrucci ([Ferruccilu@mail.nih.gov](mailto:Ferruccilu@mail.nih.gov)).

### **FIGURE 1:**

#### **FIG. 1A)**

DEG\_limma\_voom.R: using raw gene expression counts as input, this script was used to filter genes, perform voom quantile normalization, and compute empirical Bayes statistics for differential gene expression.

DEG\_volcano\_coding.R: generates Volcano plot of coding genes.

#### **FIG. 1B)**

DEG\_volcano\_noncoding.R: generates Volcano plot of non-coding genes.

#### **FIG. 1C)**

TPM\_boxplots\_genes.R: generates TPM expression boxplot for a target gene of interest.

#### **FIG. 1D)**

We implemented our GSEA framework (publicly available under GPL-3.0 license at [https://github.com/juliancandia/GSEA\\_Pathway\\_Analysis](https://github.com/juliancandia/GSEA_Pathway_Analysis)) using parameters `n_run = 1000`, `adj_pval_thres = 0.05`, `pwysel_thres = 80`, `gene_sel_thres = 80`, `seed = 123`, `eps = 0`. We performed separate runs for "Hallmarks" and "Canonical\_Pathways". Output files were manually merged and annotated (provided as **Supplementary Data 2**).

DEG\_GSEA\_barplot.R: annotated pathways are visualized using a barplot chart.

### **FIGURE 2:**

#### **FIG. 2A)**

DEP\_volcano.R: generates Volcano plot of proteins.

#### **FIG. 2B)**

ETC\_ProteinHeatmap.R: generates heatmap of ETC complex proteins.

#### **FIG. 2C)**

RibosomeBarplot.R: generates ribosomal protein barplot.

**FIG. 2D)**

We implemented our GSEA framework (publicly available under GPL-3.0 license at [https://github.com/juliancandia/GSEA\\_Pathway\\_Analysis](https://github.com/juliancandia/GSEA_Pathway_Analysis)) using parameters `n_run = 1000`, `adj_pval_thres = 0.05`, `pwy_sel_thres = 80`, `gene_sel_thres = 80`, `seed = 123`, `eps = 0`. We performed separate runs for “Hallmarks” and “Canonical\_Pathways”. Output files were manually merged and annotated (provided as **Supplementary Data 4**).

DEP\_GSEA\_barplot.R: annotated pathways are visualized using a barplot chart.

**FIGURE 3:**

**FIG. 3A-B)**

ETC\_prot\_net.R: calculates full list of protein pairwise correlations and filtered list for FDR-adjusted  $p < 0.05$ , separately for PAD and non-PAD subcohorts. It also generates an annotated list of ETC proteins. These files were uploaded into Cytoscape v. 3.9.1 to generate protein network visualizations (ETC\_networks.cys).

**FIG. 3C -D)**

Co-regulation\_ProteinsPAD\_Non-PAD.R calculates the correlation of ETC complex proteins with PAD and non-PAD.

**FIG. 3E)**

ETC\_prot\_net\_closeness.R: implements algorithm to determine the average closeness centrality of the protein PAD and non-PAD networks, as well as random network ensembles to assess statistical significance. For further details, see “Network analysis” under the “Online methods” section.

**FIG. 3F)**

ETC\_prot\_correl\_comparison.R: performs a statistical comparison between pairwise correlation distributions for PAD and non-PAD, separately determined for each ETC protein complex.

**FIGURE 4:**

PAD\_muscle\_paper.sas: SAS code to generate csv output files, which, in turn, are uploaded into ORIGIN PRO for plotting.

**FIGURE 5:**

**FIG. 5A)**

Prot\_mRNA\_cor.R: calculates protein-mRNA correlations with 95% confidence-level ranges for PAD and non-PAD subcohorts.

Prot\_mRNA\_cor\_groups.R: classifies protein-mRNA pairs based on correlation significance in PAD and non-PAD subcohorts.

Prot\_mRNA\_cor\_plot.R: generates plot of protein-mRNA correlations in PAD (y-axis) vs non-PAD (x-axis) colored by significance group, highlighting genes contained in several target gene sets.

**FIG. 5B-D)**

Prot\_mRNA\_cor\_diff.R: generates detailed plots of PAD vs non-PAD differences for genes contained in several target gene sets.