Homogenate   Synaptosome

raw data

proteome discoverer TMT search settings
mods, fdr, database

ASD_Dev_Hom_Allplexes_Run2_PSM
ASD_Dev_Clinic_allvariables
ASD_Syn_complete_noP5F6_Run1_PSM
ASD_Dev_Clinic_allvariables
ASD_Dev_limma 3 pairs with low abundance dropped

PSMs

data-setup (1)   *Create metadata*
Sums PSMs (unique modifications) into peptides

Peptides   *79,952 x 80*
61,060 x 74

Peptide-protein-gene-maps (1.1)(peptide)   *Map peptides to proteins & genes w/ ensembl database*

35,103 x 80   quality-control (2)(peptide)   *Remove peptides that don't map to any protein*
29,311 x 74   *Remove peptides not quantified in at least 50% samples and pooled controls*

Synaptic protein enrichment: Synaptosome to Homogenate comparison

VAE imputation   VAEIT (f)   vae-python.R   *These need to be in directory*

Hom only   imputation-test (3.1)(peptide-multi-rep)   *Testing imputation w/ & w/o pooled controls included*

imputation-test (3.1)(peptide)   *Simulates missingness of the data & calculates imputation error*

imputation (3)(peptide)   *Calculates softimpute starting values w/ pooled controls*
Imputes missing values for samples only

imputation (3)(peptide)(Hom/Syn)   *Calculates softimpute starting values w/ pooled controls*
Imputes missing values for samples and pooled controls

Peptide level analysis

unimputed data

10,772 x 62
10,761 x 56

roll-up-normalization (4)(protein)   *Removes peptides that map to multiple proteins*
Sums peptide abundance per protein
Performs sample loading normalization

normalization (4)(peptide)   *Performs sample loading normalization*

Retain only pools &\NT subjects in both Hom and Syn datasets
Retain only unique peptides identified in Hom and Syn datasets
synaptosome-enrichment-irs-norm   *Sum peptide abundance per protein*
Performs IRS normalization w/ pool controls

U.norm.prt *2,578*
U.norm.prt *2,455*

Proteins   norm.prt *4,727*
norm.prt *4,287*

Peptides   norm.pep *35,103 x 62*
norm.pep *29,311 x 56*

Proteins   norm.prt *3,852*
norm.prt *3,852*

~ DX + PLEX + AGE + (1|Pair)
~ DX + PLEX + AGE + DX:AGE + (1|Pair)

model-selection (6)(protein)   *Forward variable selection using covariates*
Tested for the interaction of DX:AGE

~ DX + PLEX + AGE + (1|Pair)
~ DX + PLEX + AGE + DX:AGE + (1|Pair)

model-selection (6)(peptide)   *Forward variable selection using covariates*
Tested for the interaction of DX:AGE

~ DX + PLEX + AGE + (1|Pair)
~ DX + PLEX + AGE + (1|Pair)
~ DX + PLEX + AGE + DX:AGE + (1|Pair)

Model selection attempted. Plex is important; however, samples within each plex are identical in Hom and Syn.

Model becomes rank deficient if Plex is used.

~ Tissue + AGE + (1|Subject ID)   These covariates were selected.

lmer-regression (7)(protein)   *Performs linear regression*
Calculates the significance of differential abundance
for the effects of age, dx, or the interaction of dx & age

lmer-regression (7)(peptide)   *Performs linear regression*
Determines DE peptides for the effect of DX & AGE

LMER-regression   *Performs linear regression*
Determines DE proteins for the effect of Tissue
Returns residuals w/ effect of tissue added back

U.MLM.DX   U.MLM.AGE
U.MLM.DX   U.MLM.AGE   U.MLM.DX.AGE

MLM.DX   MLM.AGE
MLM.DX   MLM.AGE
MLM.DX   MLM.AGE   MLM.DX.AGE

MLM.DX   MLM.AGE
MLM.DX   MLM.AGE
MLM.DX   MLM.AGE   MLM.DX.AGE

MLM.Tissue   RE.Tissue

*to be comparable to Hom, DE for MLM.AGE was calculated w/o DX:AGE interaction in the model

lmer-residuals (7.1)(protein)   *Returns residuals that account for all variables or*
specific effects can be added back iteratively

syn-enrichment-plot   *Use SynGO ontolgies to classify proteins as synaptic or nonsynaptic*
Volcano and density plots for fold change estimate
Boxplot select proteins by individual subjects enrichments

RE-DX   RE-AGE
RE-DX+DX:AGE   RE-AGE+DX:AGE   RE-DX+AGE+DX:AGE

age-flexmix-clustering (9)(protein)   *Calculate residuals w/ the effect of AGE (& DX:AGE) added back*
Separate subjects by DX & perform flexmix clustering (k=3)
Perform paired t test between DX group for each protein's AGE.b

prt.clusters
prt.clusters

t test results were calculated
but DE proteins by AGE:DX were taken from the MLM results w/ a q<0.1

Basic plots

MLM.DX **MLM.AGE**
MLM.DX **MLM.AGE** **MLM.DX.AGE**

summary-results-barplot-pvaldist
Creates bargraph of total No. proteins and DE proteins
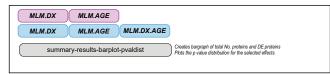*Plots the p value distribution for the selected effects*

MLM.DX **MLM.AGE**
MLM.DX **MLM.AGE**

homogenate-synaptosome-comparison-scatterplot
Plots scatterplot of Hom and Syn estimates of overlapping proteins
Calculates pearson correlation of the overlapping proteins

MLM.-(B+B+X) **U.MLM.DX**
MLM.DX **MLM.DX.AGE** **U.MLM.DX**

volcano-plots
Createss volcano plot using estimates for the effects of DX

MLM.DX **Gandal 2022 DE V1 transcripts**
MLM.DX

Overlapping_proteinandtranscript
Createss scatter plot of overlapping DE proteins and transcripts

RE-AGE **prt.clusters**
RE-AGE+DX:AGE **prt.clusters**

example-developmental-proteins
Calculate residuals w/ the effect of AGE & if needed, DX:AGE, added back
Select representative protein from each cluster (n=3) (smallest q value)
Plot regression line of residual abundance within each DX group

RE-DX+AGE+DX:AGE **prt.clusters**

intDXAGE-trajectory-plot
Calculate residuals w/ the effect of DX & AGE & DX:AGE added back
Select representative protein from each developmental protein cluster (n=9) (smallest q value)
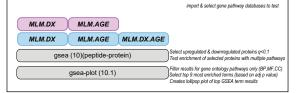*Plot regression line of residual abundance within each DX group*

prt.clusters **MLM.DX.AGE**

ORA-development-barplots
Using webgestalt, perform overrepresentation analysis within developmental subclusters (n=9)
Make barplots of only terms w/ FDR < 0.05 (BPnR or Reactome)
if no term passes FDR, just list proteins (or input string diagrams?)

MLM.DX
MLM.DX **MLM.DX.AGE**

mitoxplorer-gsea-plot
Use mitoxplorer mitochondrial gsea enrichment function on website
Plot bargraph of enriched terms based on combined score (p value and fold change)
Color the percentage of the bar that corresponds to DE by DX and fold change directionality

GSEA

import & select gene pathway databases to test

MLM.DX **MLM.AGE**
MLM.DX **MLM.AGE** **MLM.DX.AGE**

gsea (10)(peptide-protein)
Select upregulated & downregulated proteins q<0.1
Test enrichment of selected proteins with multiple pathways

gsea-plot (10.1)
Filter results for gene ontology pathways only (BP,MF,CC)
Select top 9 most enriched terms (based on adj p value)
Creates lollipop plot of top GSEA term results

Synaptosome total protein enrichment

need Synaptosome_prep excel sheet

synaptosome-total-protein-comparison
Calculates ratio of synaptosome total protein (ug) to mg tissue
Calculates adjusted residuals with the effects of DX & AGE added back

ASD comparisons

MLM.DX
MLM.DX

protein-DE-lists
Select upregulated & downregulated or generally DE proteins q<0.1
Proteins are mapped to their ensembl gene ids or removed if no gene id

transcript-brainregion-list
Filter for upregulated & downregulated DE transcripts FDR<0.1 from Gandal 2022
Retain only the union of transcripts identified at the protein level

transcript-celltype-Wamsley-list2
Filter for upregulated & downregulated DE transcripts FDR<0.1 from Wamsley 2024
Retain only the union of transcripts identified at the protein level

gene-list
Filter for risk genes from Fu 2022 & SFARI
Retain only the union of genes identified at the protein level

Enrichment analysis genelistoverlap.R
This function needs to be loaded
altered original code to 'D=backgroundTotal-(A+B+C)'

transcript-brainregion-enrichment
Performs 2x2 contingency enrichment analysis for each brain region & protein group

transcript-brainregion-enrichment2
Calculates FDR of all comparisons
Creates matrix table of enrichment values

transcript-celltype-wamsley-enrichment2
Performs 2x2 contingency enrichment analysis for each brain region & protein group

transcript-celltype-wamsley-enrichment3
Calculates FDR of all comparisons
Creates matrix table of enrichment values

gene-enrichment
Performs 2x2 contingency enrichment analysis for each brain region & protein group
Calculates FDR of all comparisons
Creates matrix table of enrichment values