

Appendix for Shorthouse et al. Heterogeneity of the Cancer Cell Line Metabolic Landscape

Page 2: Appendix Figure S1

Page 4: Appendix Figure S2

Page 5: Appendix Figure S3

Page 7: Appendix Figure S4

Page 8: Appendix Figure S5

Page 9: Appendix Figure S6

Page 11: Appendix Figure S7

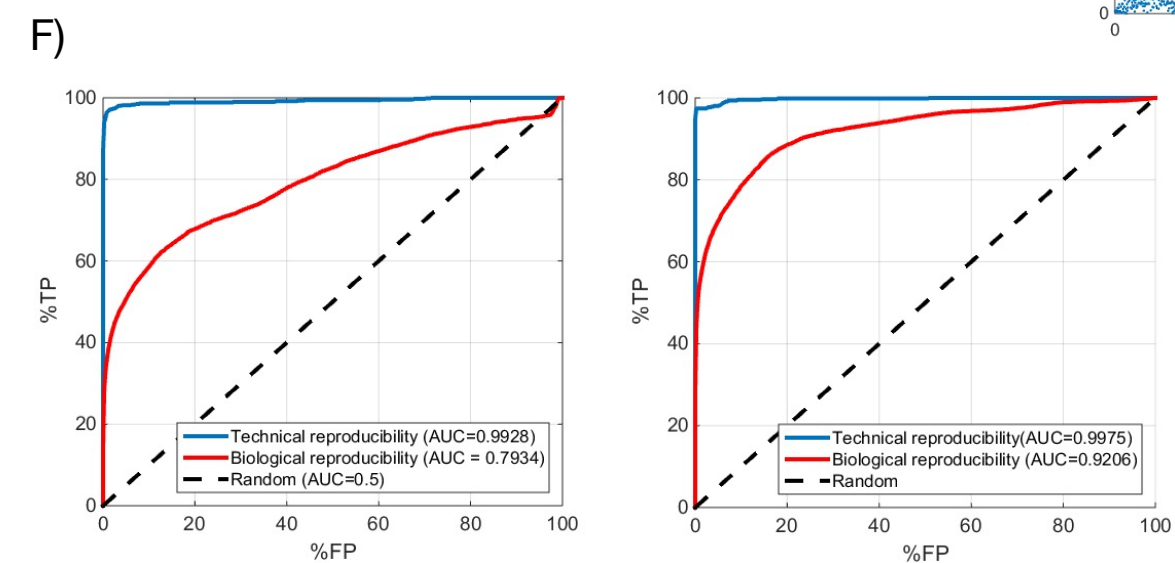
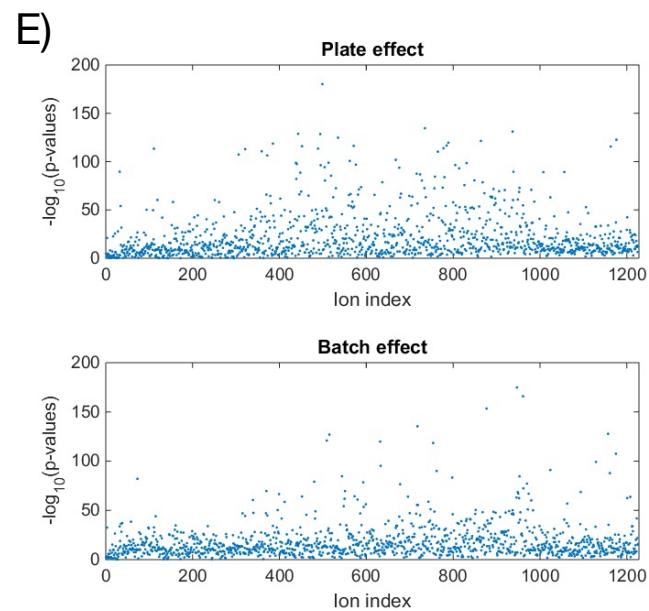
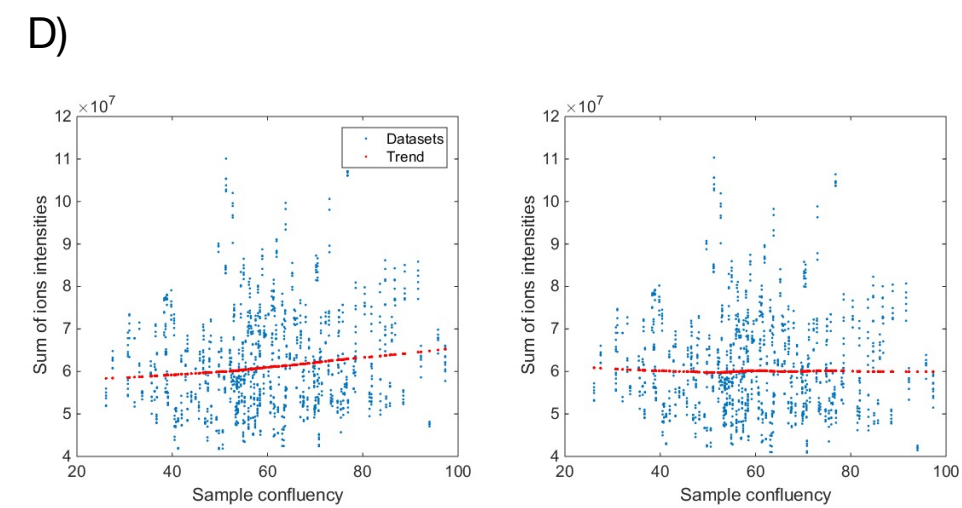
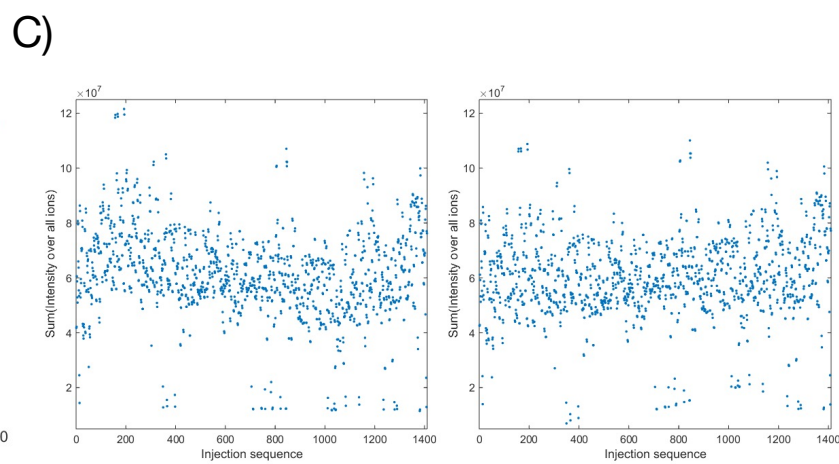
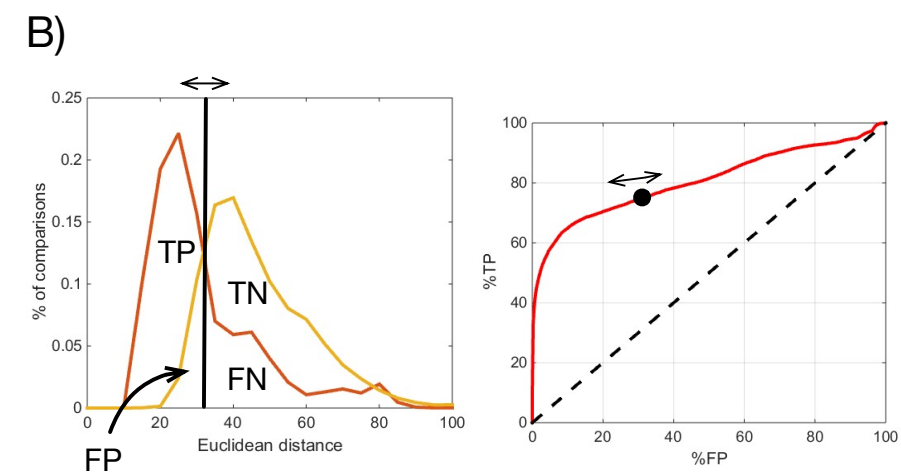
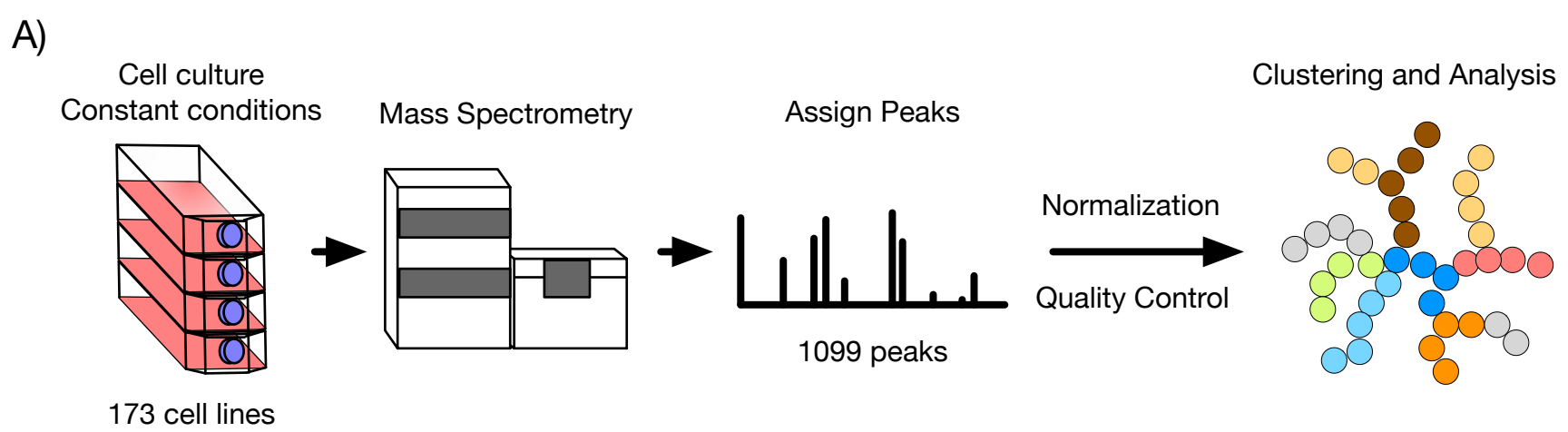
Page 12: Appendix Figure S8

Page 13: Appendix Figure S9

Page 14: Appendix Figure S10

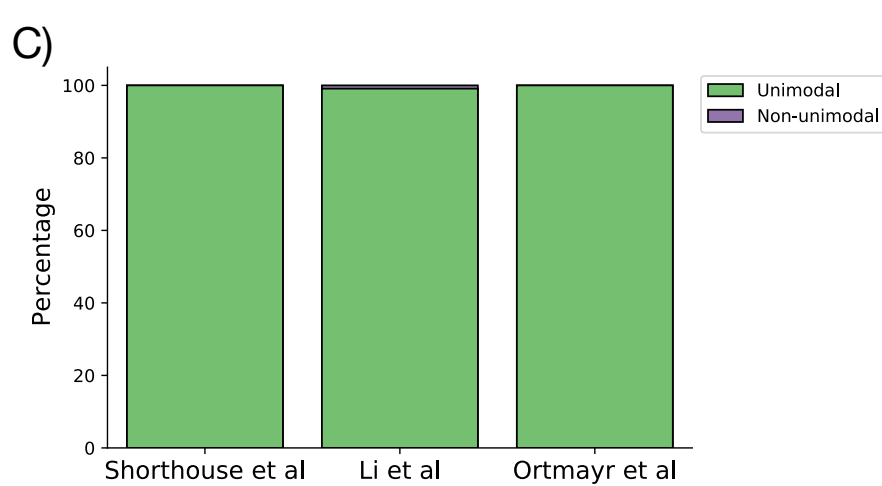
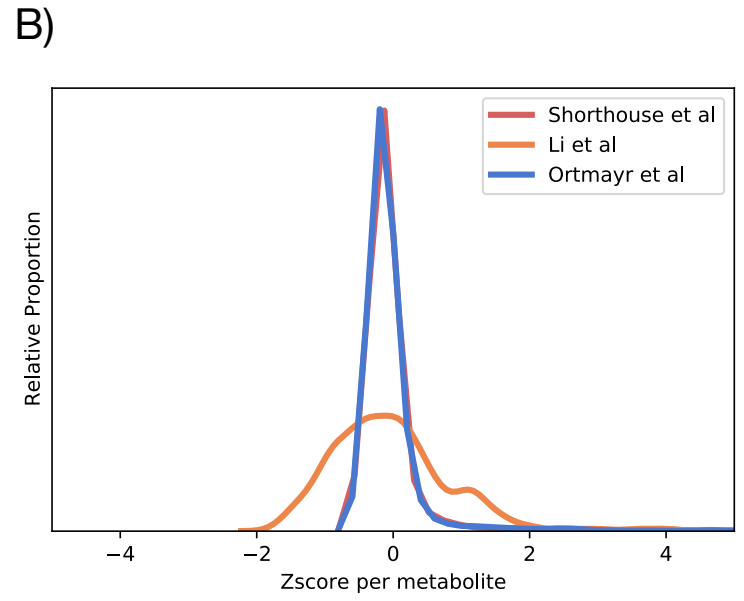
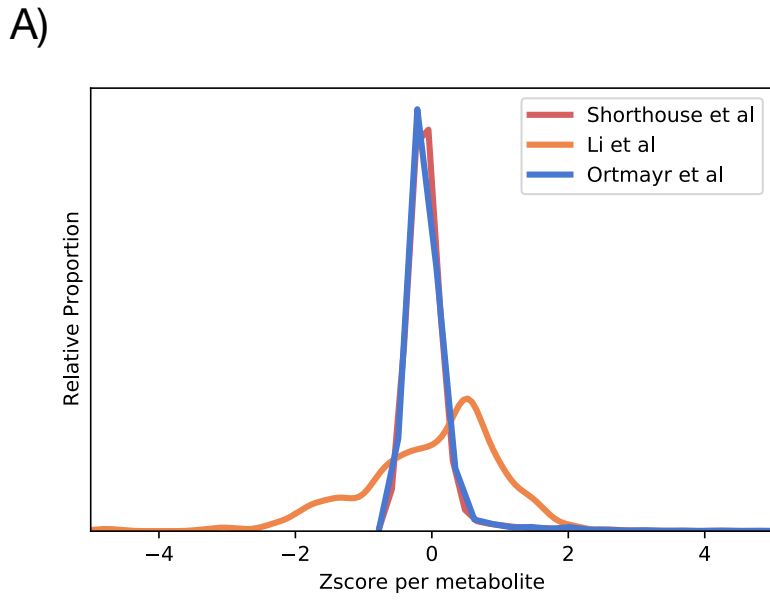
Page 16: Appendix Figure S11

Page 17: Appendix Figure S12



Appendix Figure S1

- A) Schematic of workflow for generating Mass Spectrometry data for 173 cell lines from 11 tissues.
- B) Euclidean distance based ROC analysis performed for biological vs non-biological replicates.
- C) Total mass per injection before (left) and after (right) Lowess smoothing to control for total injection mass.
- D) Total mass per injection before (left) and after (right) Lowess smoothing to control for plate confluence.
- E) ANOVA p-values for correlation of each peak with plate (top) and batch (bottom) effects.
- F) ROC curves for complete metabolism data before (left) and after (right) normalisation and correction. Red lines represent distances between biological replicates, blue lines represent distances between technical replicates.



D)

Metabolite	Correlation	Pvalue vs null distribution
2-hydroxylutarate	0.505	0
Glutamine/Glutamic Acid	0.155	0.028
Citrate	0.277	0.0002
Succinate	0.288	0.0002
Glycine	0.16	0.025

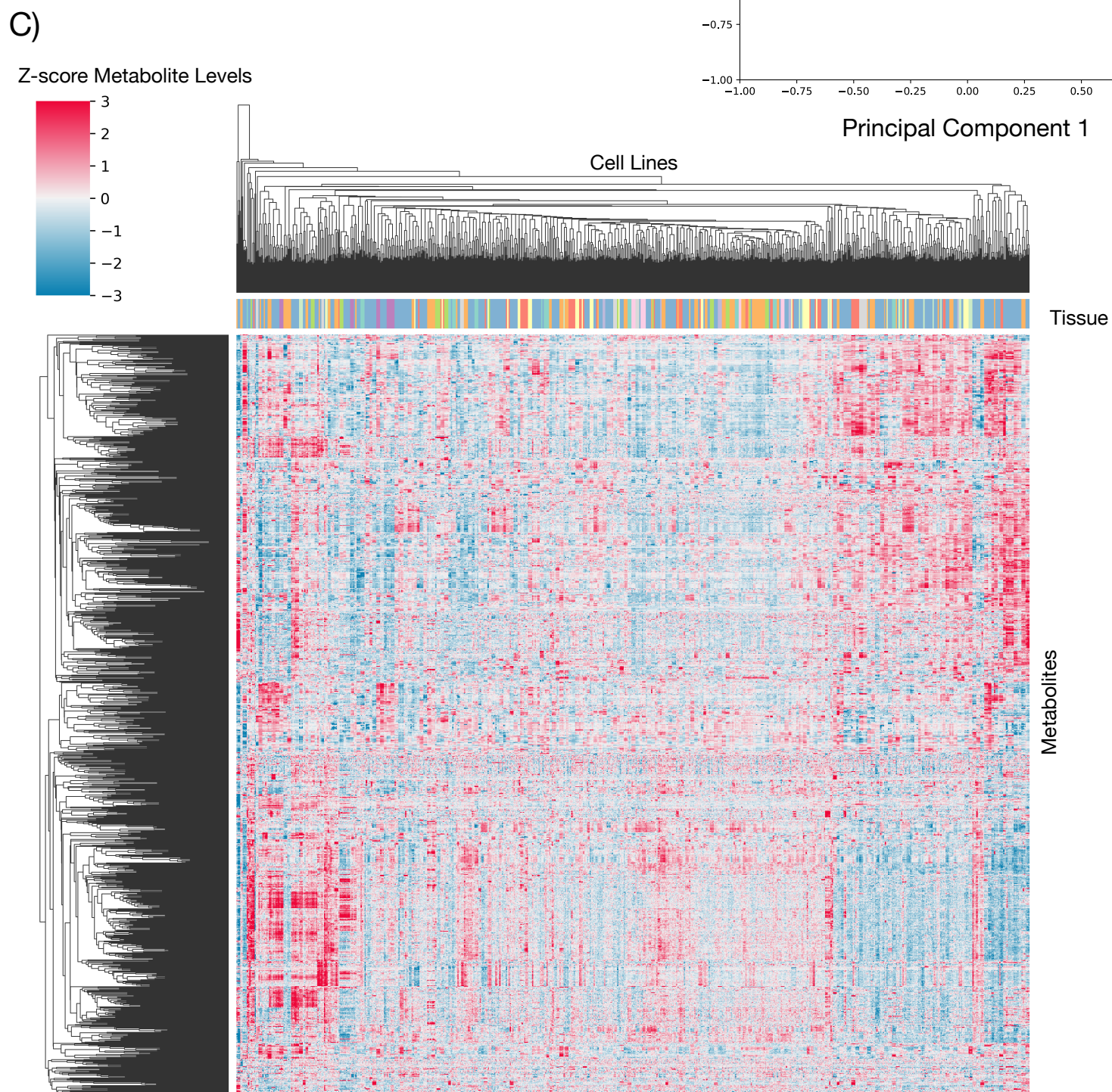
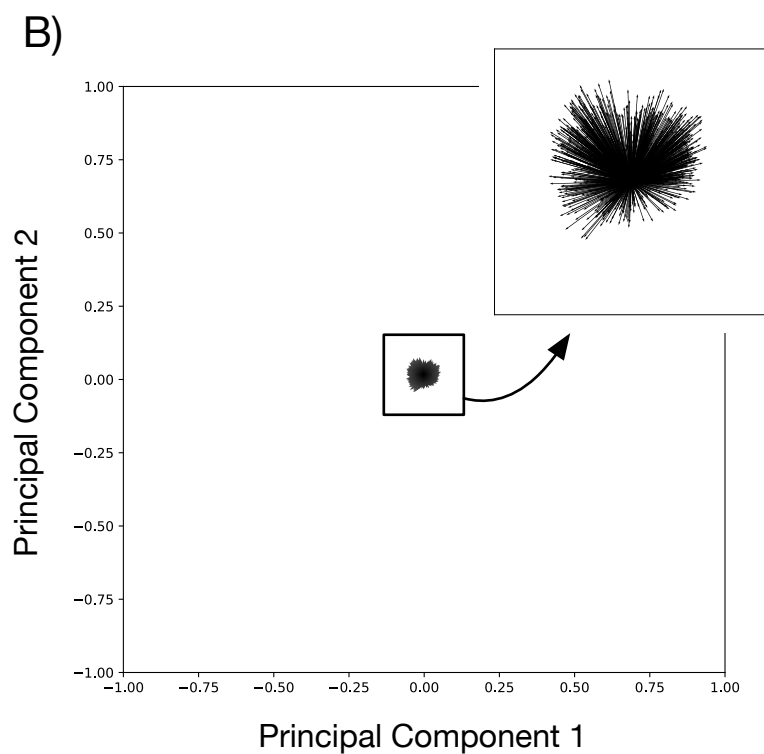
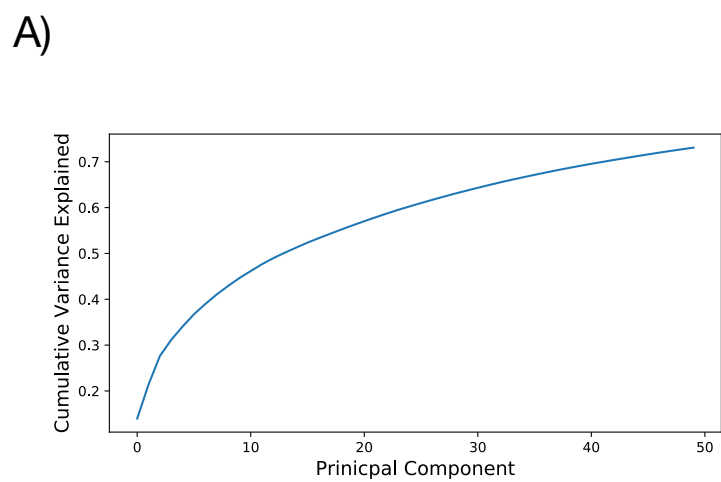
Appendix Figure S2

A) Distribution of z-score means for each metabolite in Shorthouse et al, Li et al, and Ortmayr et al

B) Distribution of standard deviations for each metabolite in Shorthouse et al, Li et al, and Ortmayr et al

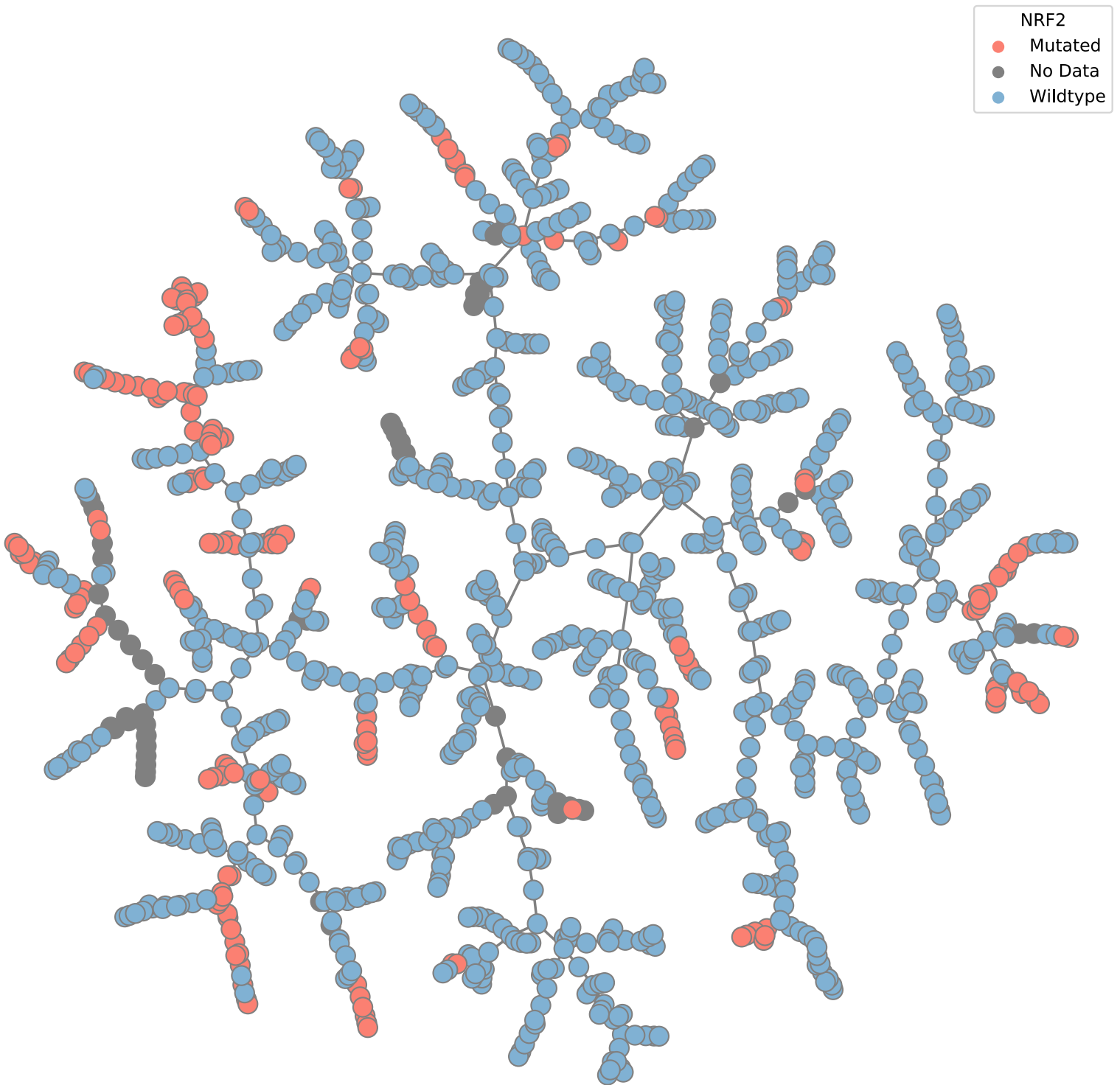
C) Proportion of unimodal metabolite distributions in Shorthouse et al, Li et al, and Ortmayr et al.

D) Statistical comparison of Shorthouse et al with Li et al. Metabolites from both studies and in overlapping cell lines are correlated using Pearson correlation, and a null distribution generated by permuting one dataset 1,000,000 times. The null distribution is used to generate a pvalue for that metabolite.



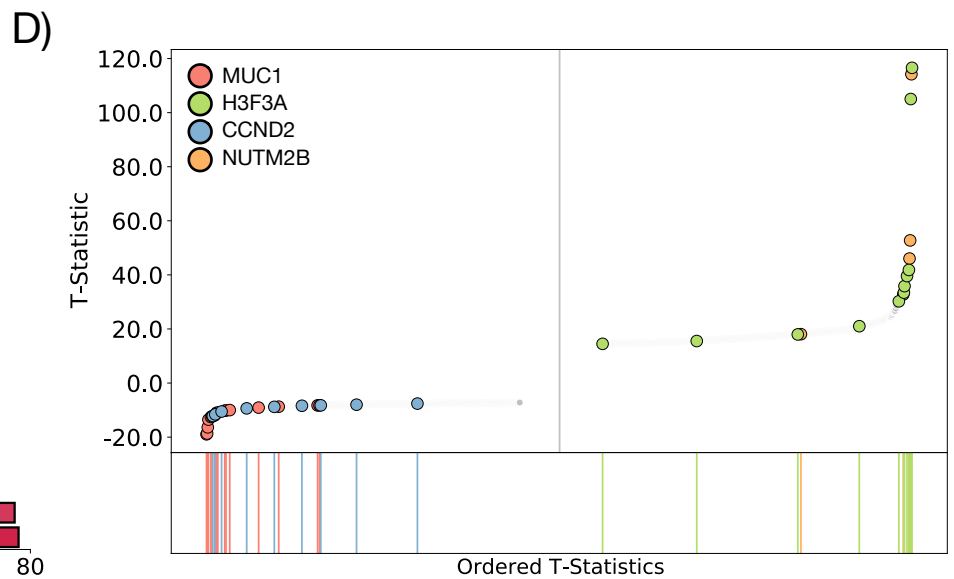
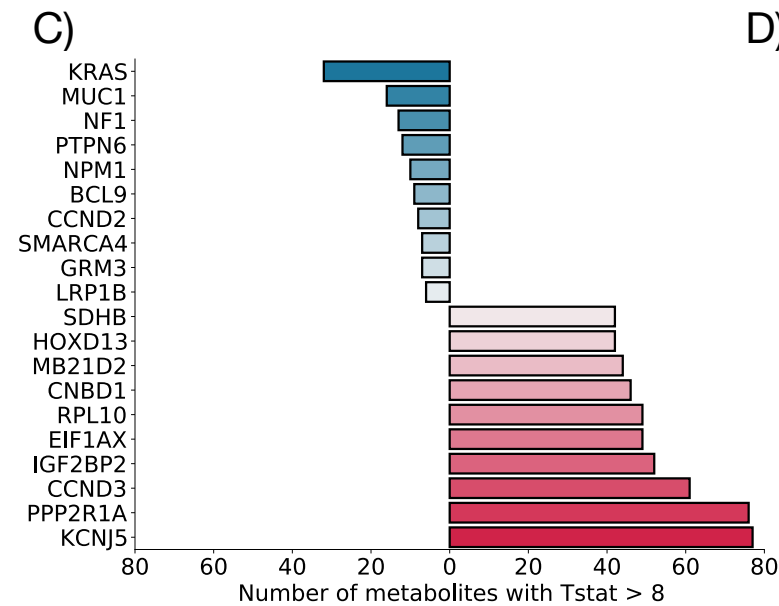
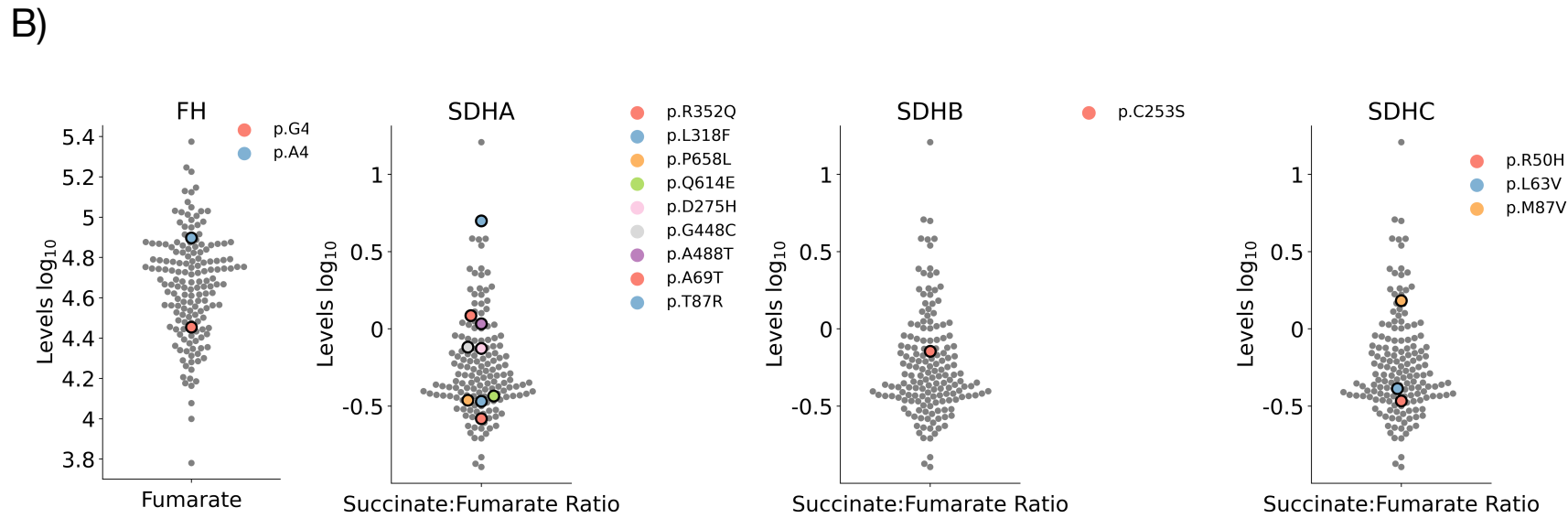
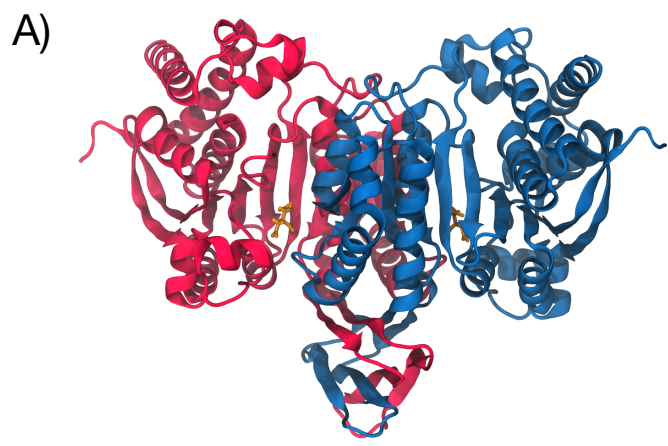
Appendix Figure S3

- A) Cumulative variance explained by principal components of the LC/MS data.
- B) Top two principal components of the LC/MS data, each arrow represents the contribution of one peak to the principal component.
- C) Heatmap of all metabolite levels in each sample, hierarchically clustered.



Appendix Figure S4

Tmap for all samples analysed, coloured by presence of a non-silent mutation to any gene in the NRF2 pathway (defined as: NFKB1, NFKB2, RELA, RELB, IKBKB, IKBKE, IKBKG.)



Supplementary Figure 4

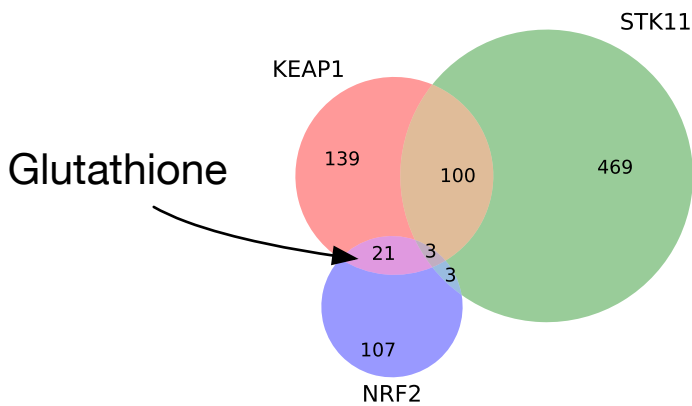
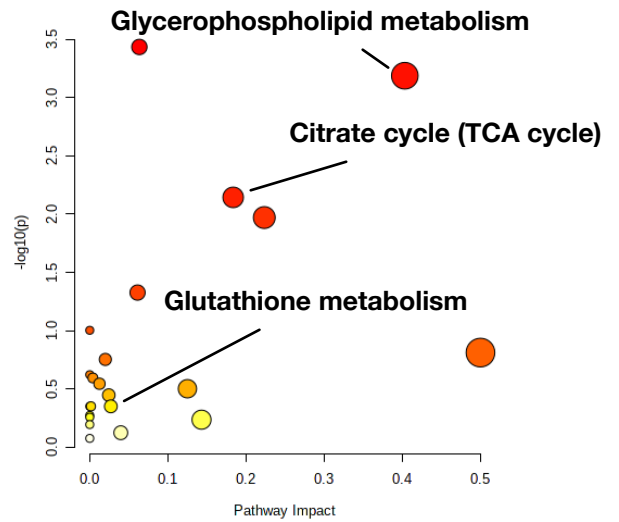
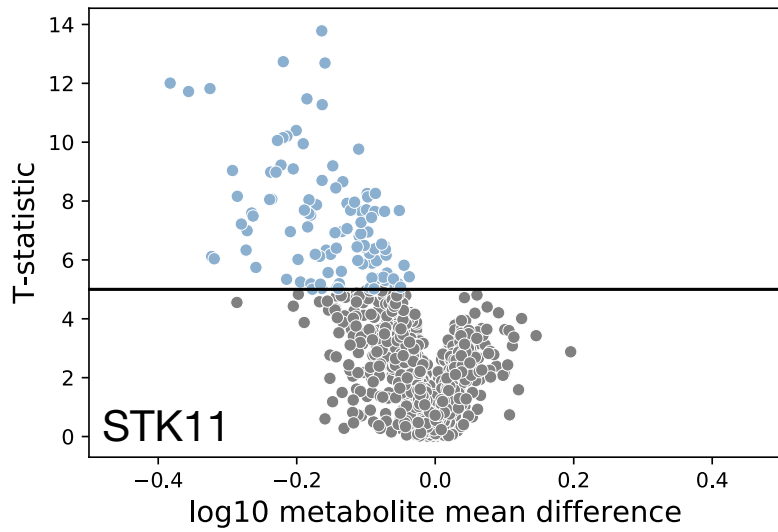
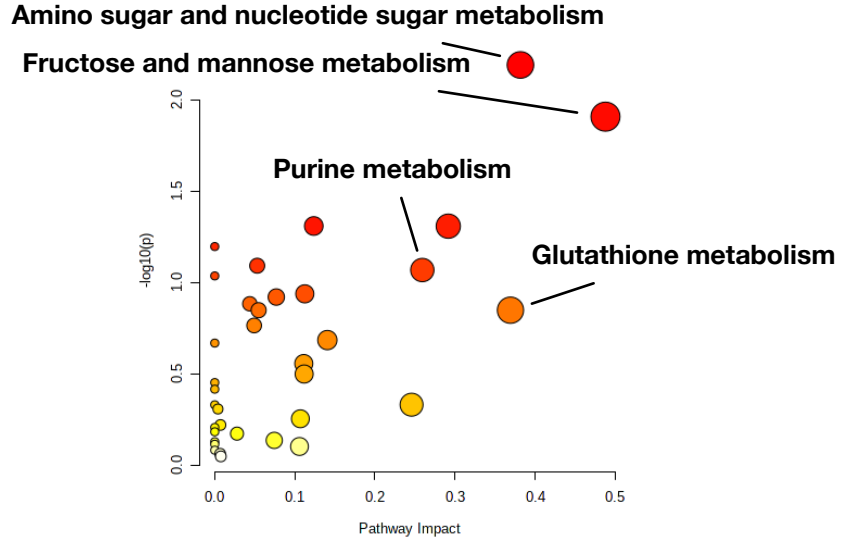
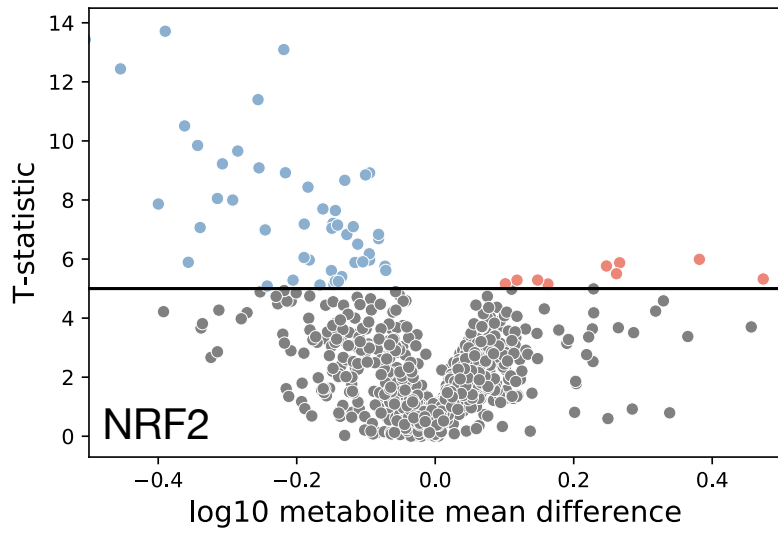
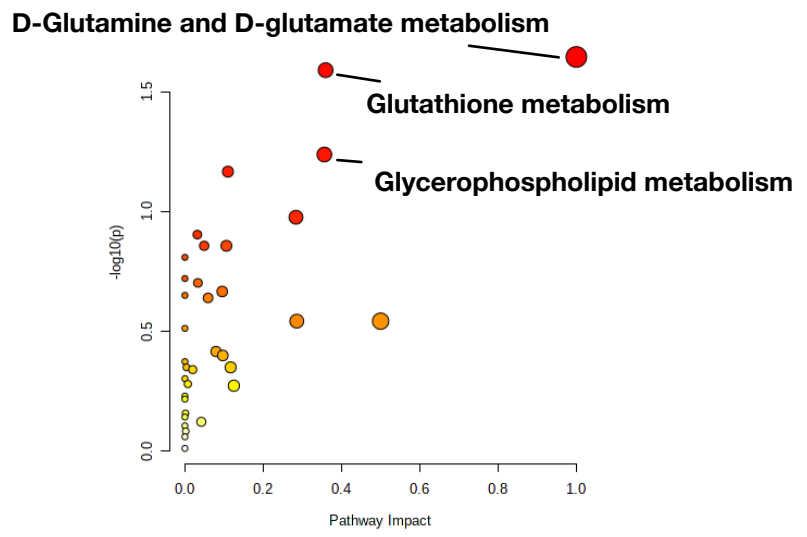
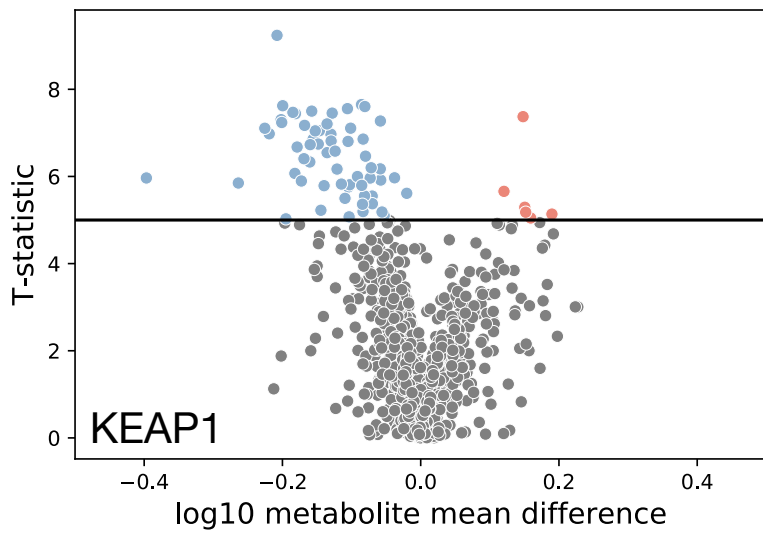
Appendix Figure S5

A) Structure of IDH1, colours show two different subunits.

B) Log₁₀ levels for Fumarate (left) or Fumarate:Succinate ratio (others) for FH, SDHA, SDHB, and SDHC. Coloured dots represent samples with a non-silent mutation in the gene.

C) Top and bottom 10 genes ranked according to the number of metabolites with an absolute T-statistic above 8.

D) Rank plot of the T statistics for mutations in MUC1, H3F3A, CCND2, and NUT2MB. 8.



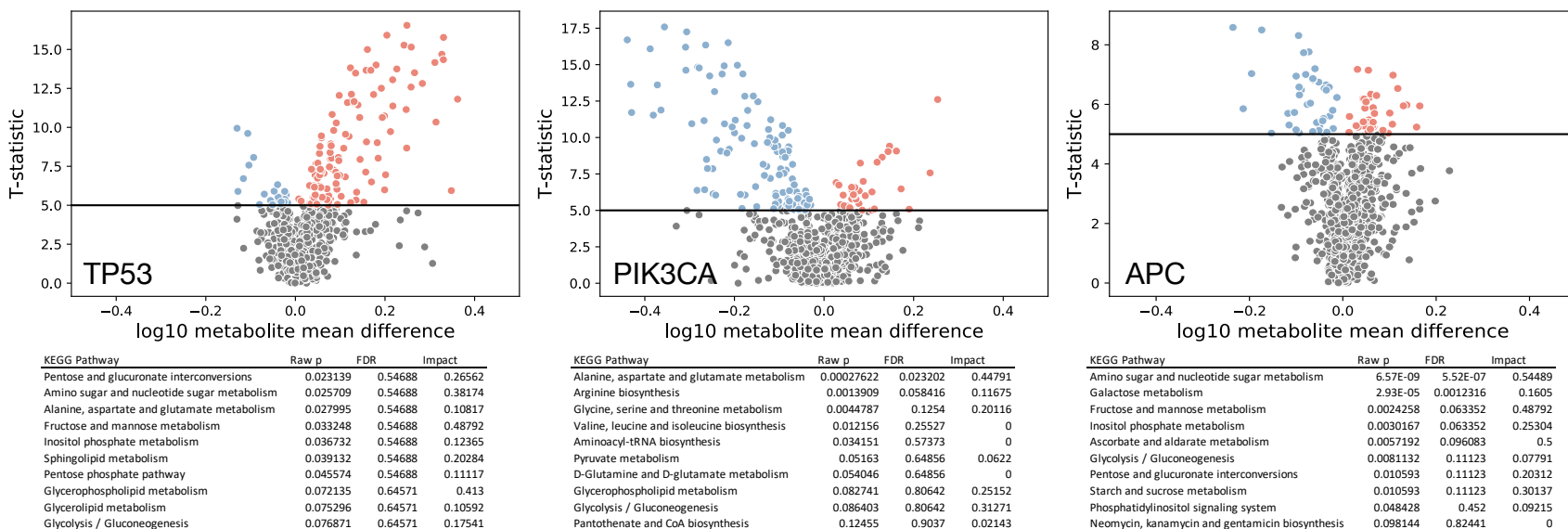
Appendix Figure S6

Differently expressed metabolites for lung cell lines mutated for KEAP1 (top), NRF2 (middle), and STK11 (bottom). Shown are volcano plots of log₁₀ metabolite mean difference between mutated and wildtype cell lines against regression T statistic (left), and pathway enrichment of all metabolites with a T-statistic > 5 (right). Bottom shows a Venn diagram of the overlap between significant metabolites (T-statistic >5) in all 3 mutant cases.

A)

KEGG Pathway	Raw pvalue	FDR	Impact
Aminoacyl-tRNA biosynthesis	5.74E-05	0.004825	0.16667
Pantothenate and CoA biosynthesis	0.0010016	0.042067	0.46786
Phenylalanine, tyrosine and tryptophan biosynthesis	0.0029319	0.078127	1
Valine, leucine and isoleucine biosynthesis	0.0037203	0.078127	0
Glycine, serine and threonine metabolism	0.0084966	0.11641	0.63031
Phenylalanine metabolism	0.0096216	0.11641	0.61904
Sphingolipid metabolism	0.0097009	0.11641	0.44828
D-Glutamine and D-glutamate metabolism	0.012736	0.13373	0
beta-Alanine metabolism	0.038713	0.36132	0.61567
Histidine metabolism	0.053737	0.45139	0.54917

B)



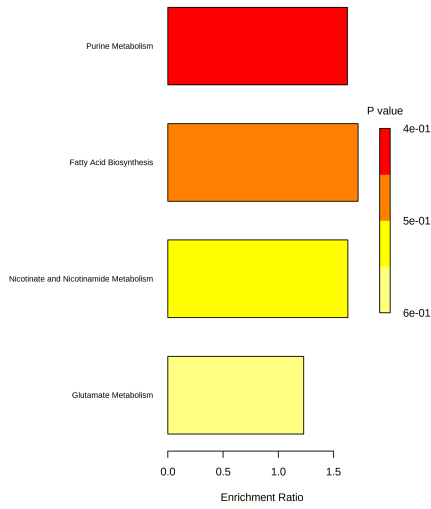
Appendix Figure S7

A) Pathway enrichment table for metabolites significantly associated with KRAS mutations (T-statistic > 5)

B) Volcano plots and pathway enrichment tables for metabolites significantly associated (T-statistic >5) with TP53 (left), APC (middle), and PIK3CA (right).

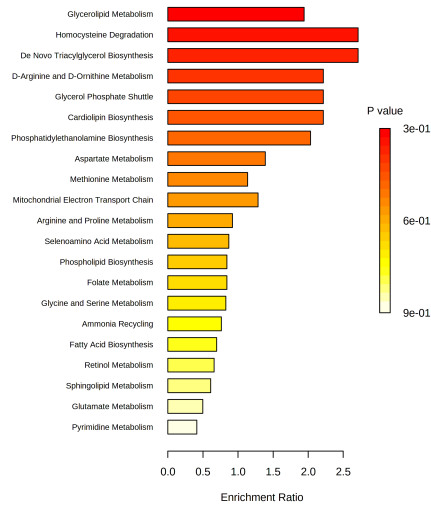
KRAS: G12A

Metabolite Sets Enrichment Overview



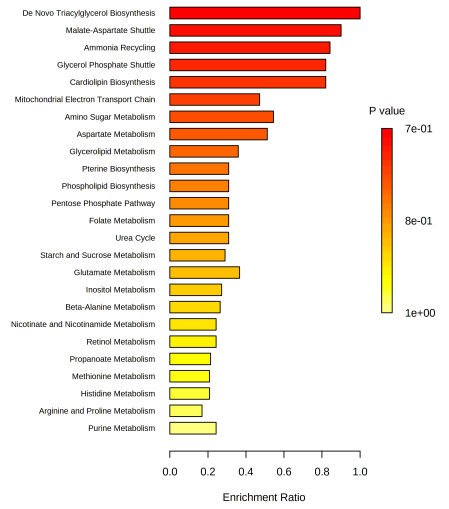
KRAS: G12C

Metabolite Sets Enrichment Overview



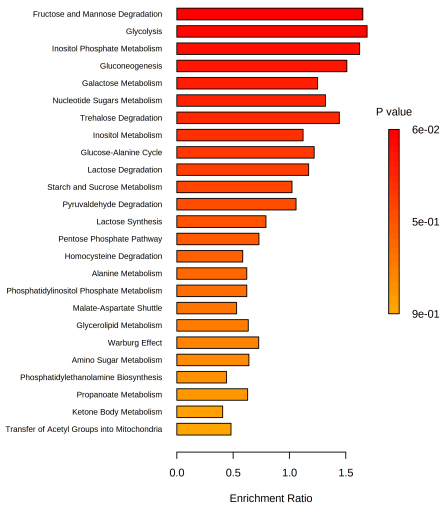
KRAS: G12D

Enrichment Overview (top 25)



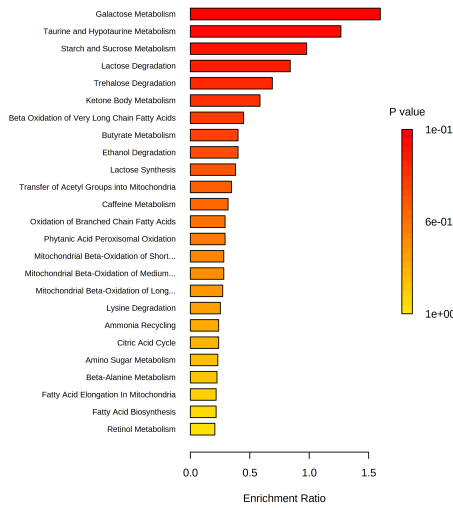
KRAS: G12V

Enrichment Overview (top 25)



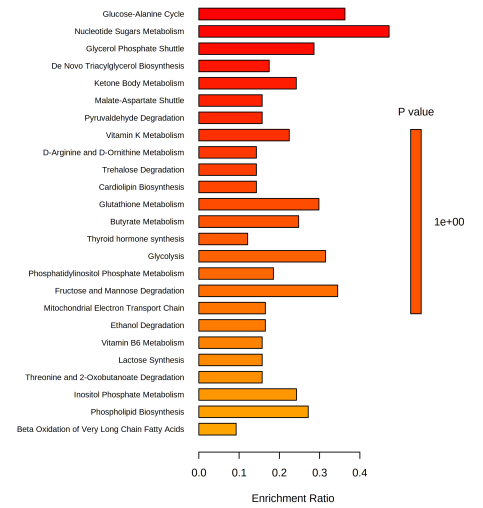
KRAS: G13C

Enrichment Overview (top 25)



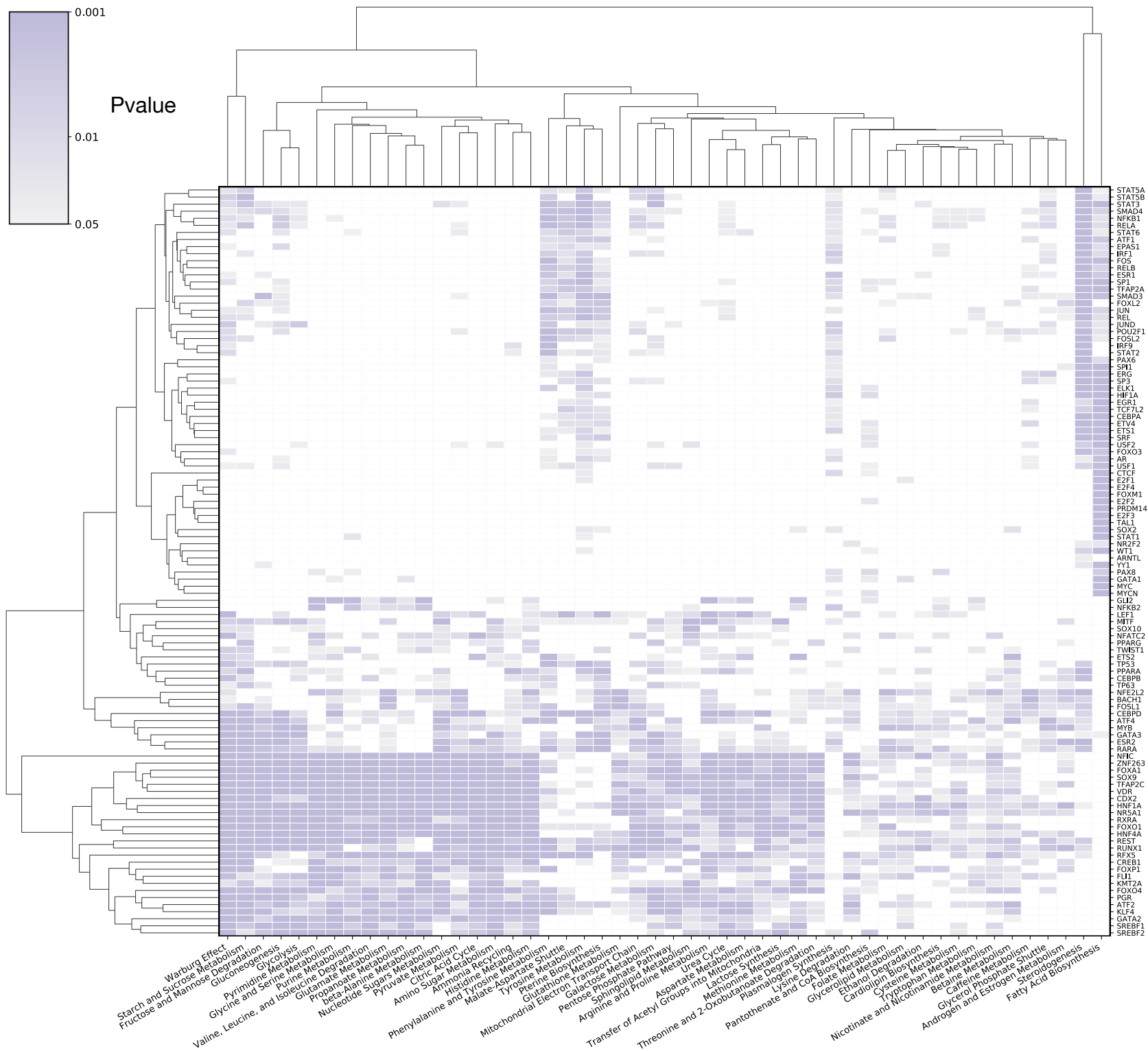
KRAS: G13D

Enrichment Overview (top 25)



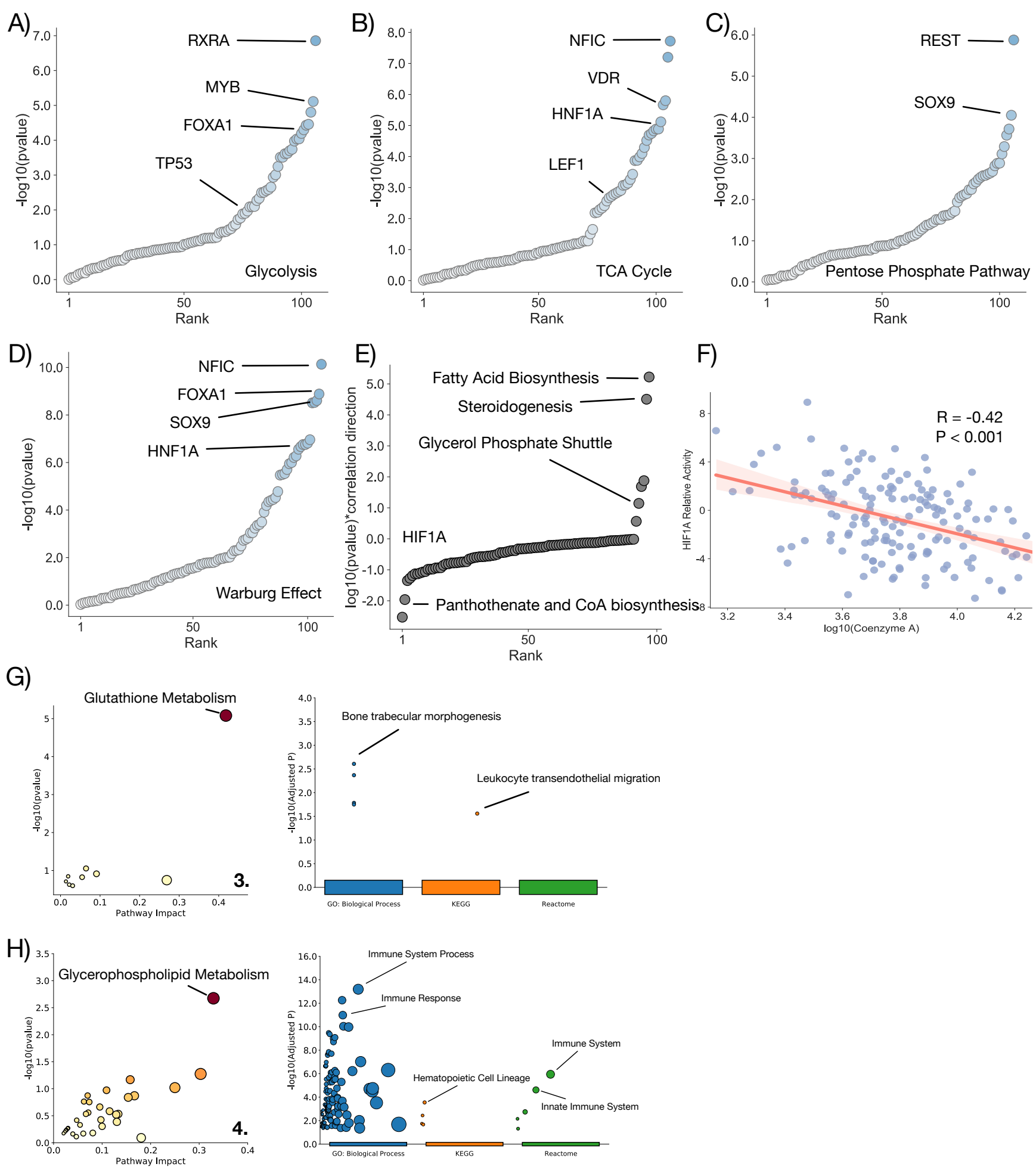
Appendix Figure S8

Enriched metabolic pathways for metabolites with a Tstat greater or lower than 5 for different KRAS mutations.



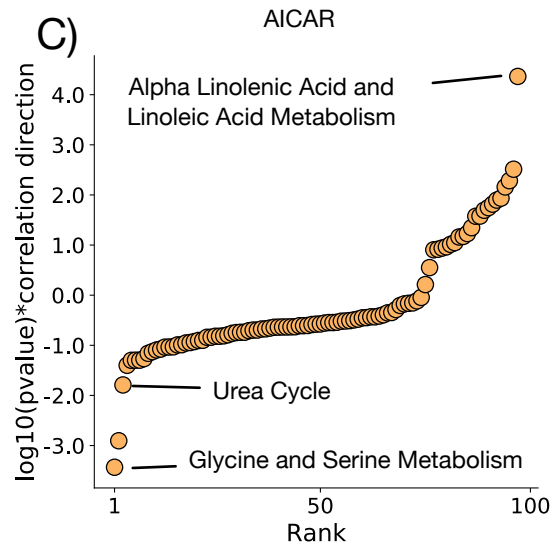
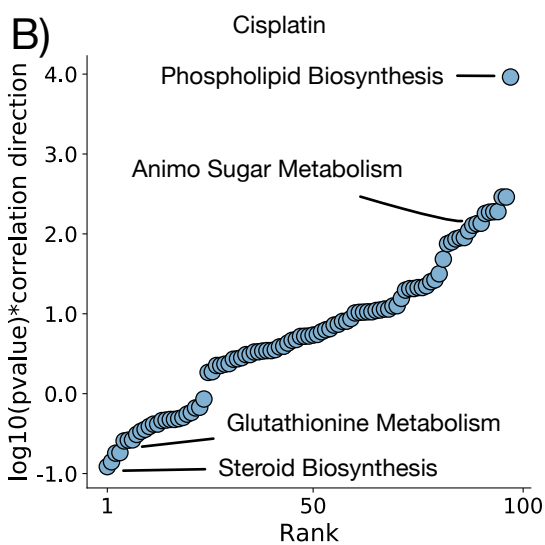
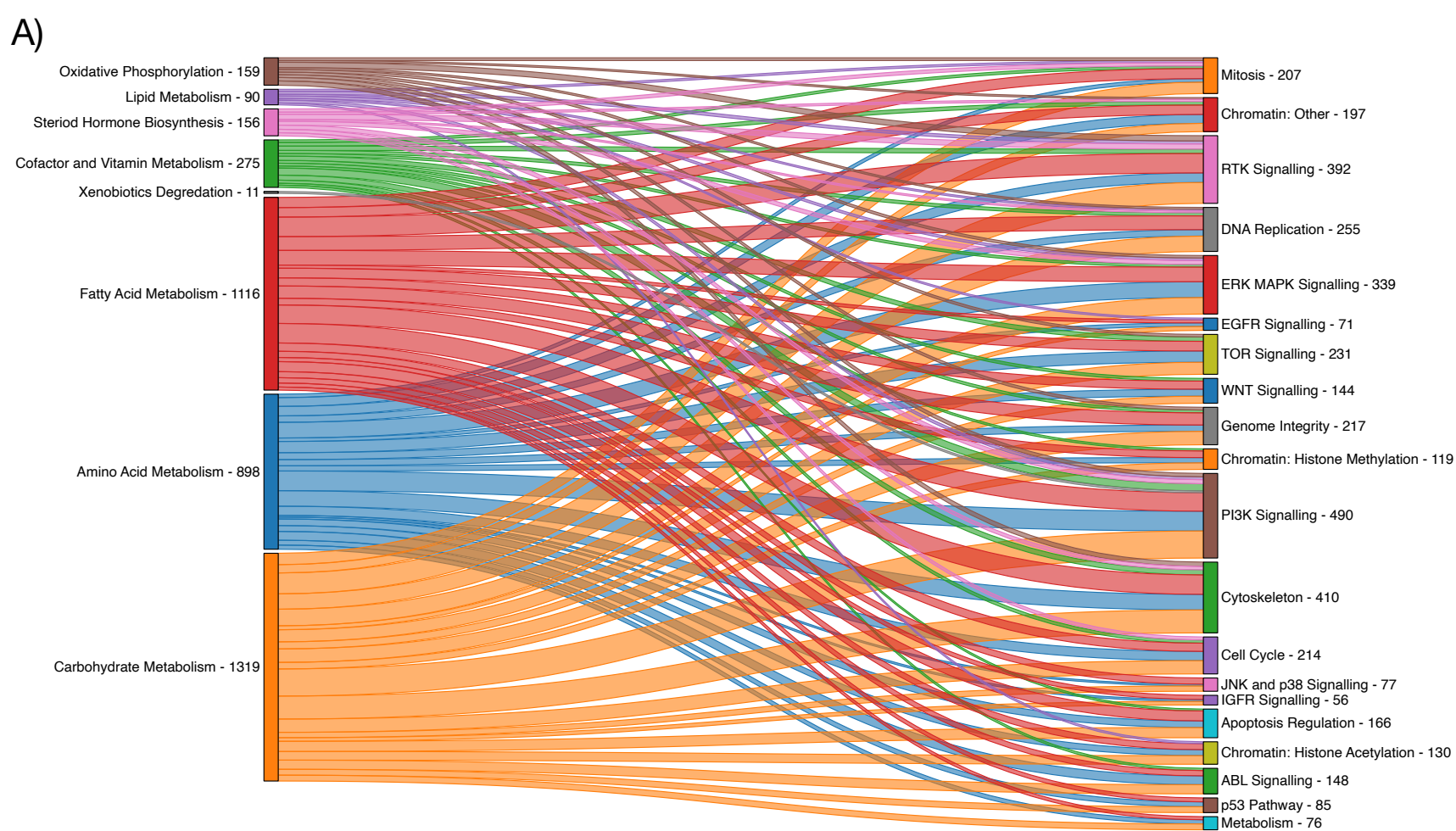
Appendix Figure S9

Heatmap of p-values (hypergeometric test) between SMPDB core metabolic pathways (x axis) and DOROTHEA (A+B significant) calculated transcriptional regulon activity (y axis)



Appendix Figure S10

- A) $-\log_{10}$ pvalue rank plot for all DOROTHEA A+B significant transcriptional regulons and their correlation against the Glycolysis SMPDB pathway.
- B) $-\log_{10}$ pvalue rank plot for all DOROTHEA A+B significant transcriptional regulons and their correlation against the TCA Cycle SMPDB pathway.
- C) $-\log_{10}$ pvalue rank plot for all DOROTHEA A+B significant transcriptional regulons and their correlation against the Pentose Phosphate SMPDB pathway.
- D) $-\log_{10}$ pvalue rank plot for all DOROTHEA A+B significant transcriptional regulons and their correlation against the Warburg Effect SMPDB pathway.
- E) $-\log_{10}$ pvalue * correlation direction plot for all SMPDB pathways against the HIF1A activity score.
- F) \log_{10} (Coenzyme A) levels against HIF1A relative activity.
- G) Metabolic pathway enrichment (left) and genetic pathway enrichment (right) for group 3 in Figure 3D.
- H) Metabolic pathway enrichment (left) and genetic pathway enrichment (right) for group 4 in Figure 3D.

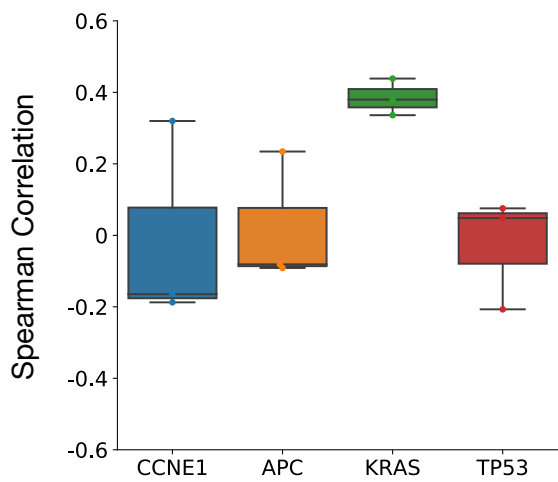


Appendix Figure S11

A) Sankey plot for significant ($p < 0.05$) associations between classes of metabolic pathways (left), and drug activities (right)

B) $\log_{10} pvalue * correlation\ direction$ for correlation between SMPDB pathways and Cisplatin.

C) $\log_{10} pvalue * correlation\ direction$ for correlation between SMPDB pathways and AICAR.



Appendix Figure S12

Boxplot of the spearman correlations between pairs of metabolite differences between mutant and WT samples for Bladder, Breast, and Lung cancer.