

Supplementary File 3 – Supplementary Figure A-D

Captions

Figure A: Gene log₂ expression values for cancer and normal samples in the *balanced* and *unbalanced* datasets for the genes in the polyamine, TCA cycle and fatty acid synthesis pathways analyzed in this study. Results are presented both for the main and validation data.

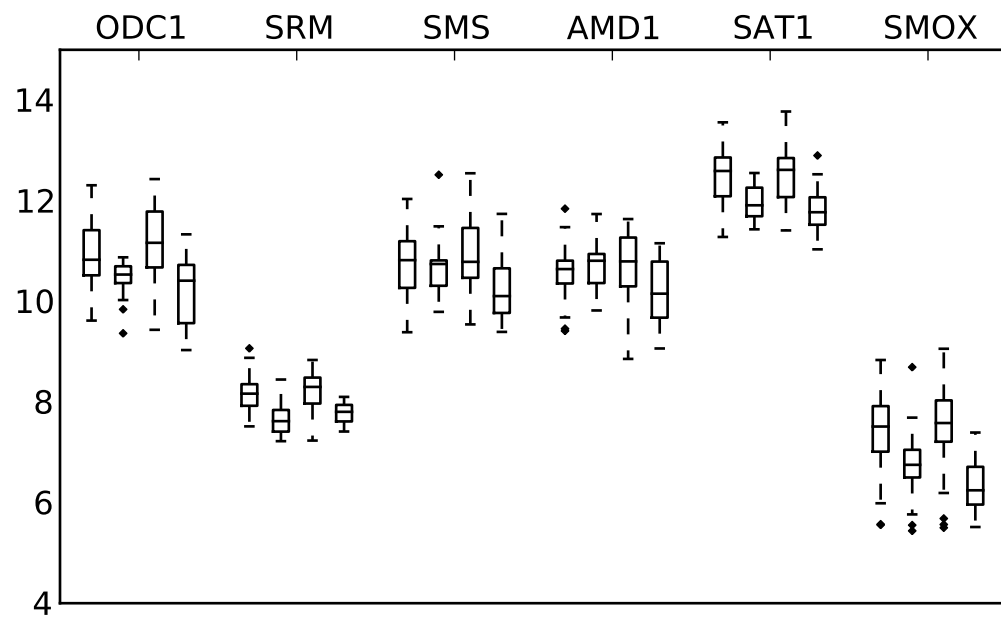
Figure B: Gene ranks for the *balanced* and *unbalanced* datasets for genes in the polyamine, TCA cycle and fatty acid synthesis pathways analyzed in this study. Results are presented both for the main and validation data.

Figure C: Validation of significant gene expression for genes commonly implicated in PCa in the *complete*, *unstratified*, *balanced* and *unbalanced* datasets. Gene names: AMACR: alpha-methylacyl-CoA racemase, KLK3: kallikrein related peptidase 3, GSTP1: glutathione S-transferase pi 1, MSMB: microseminoprotein beta, ERG: v-ets avian erythroblastosis virus E26 oncogene homolog, MYC: v-myc avian myelocytomatosis viral oncogene homolog, HPN: hepsin.

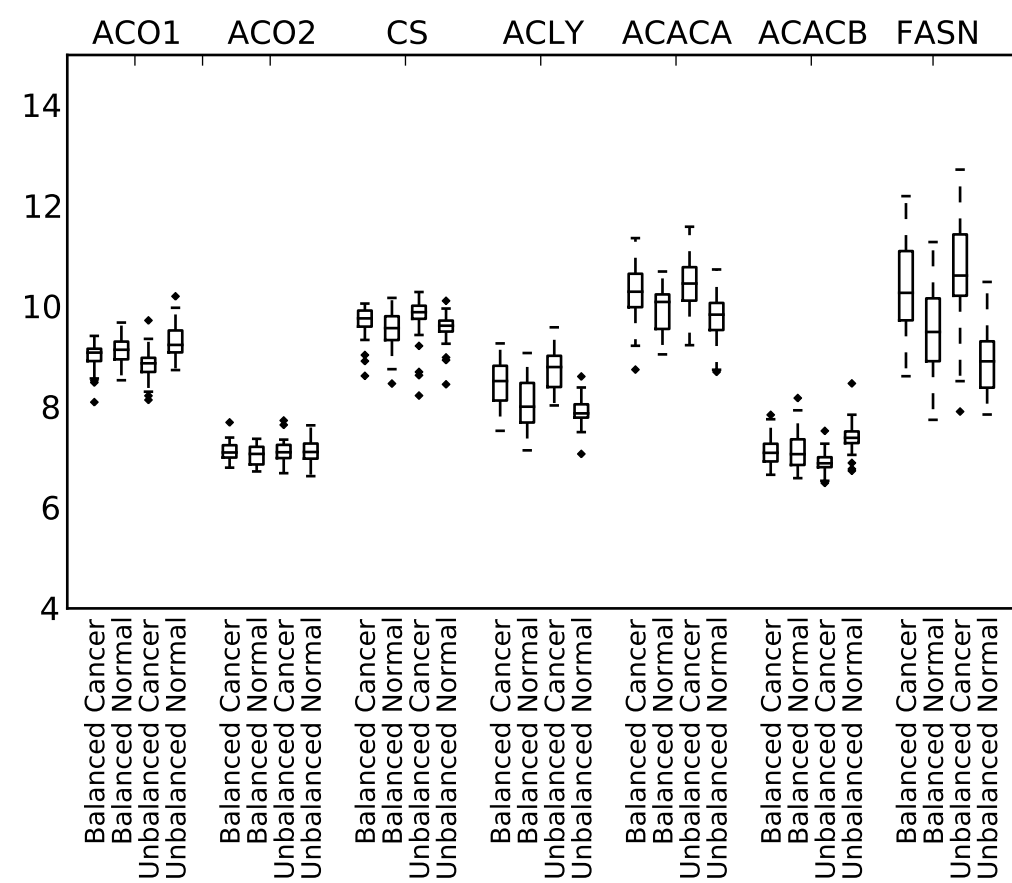
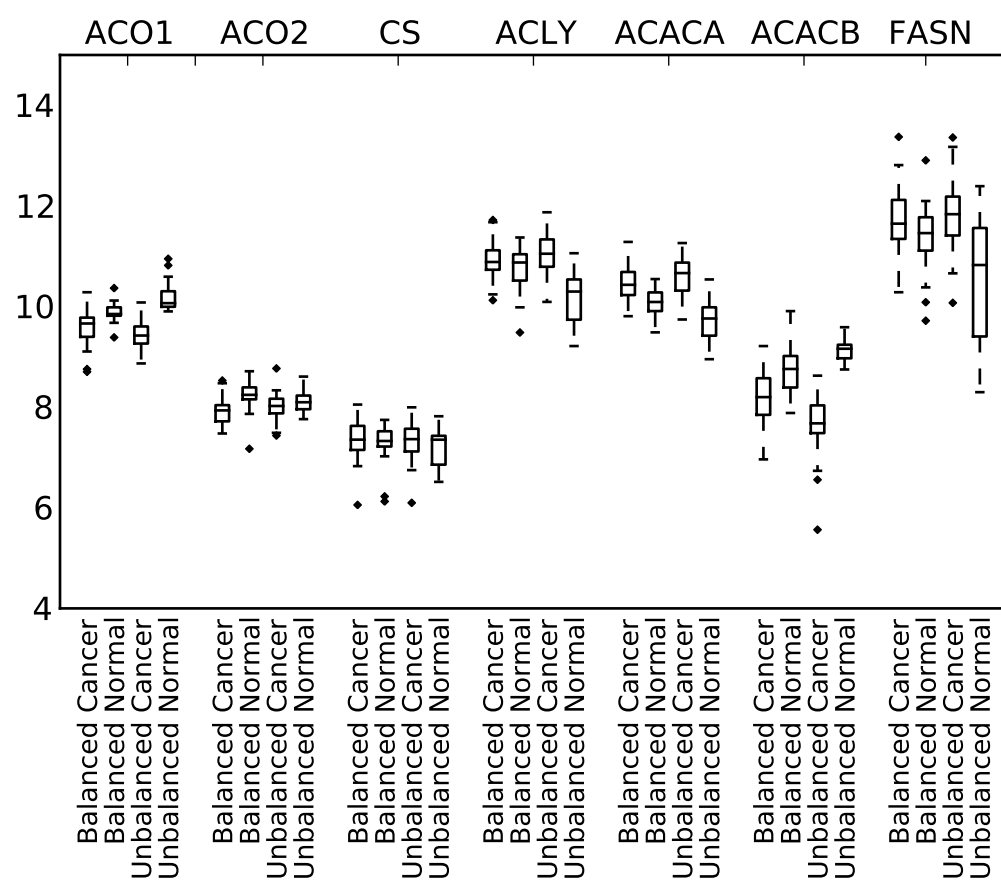
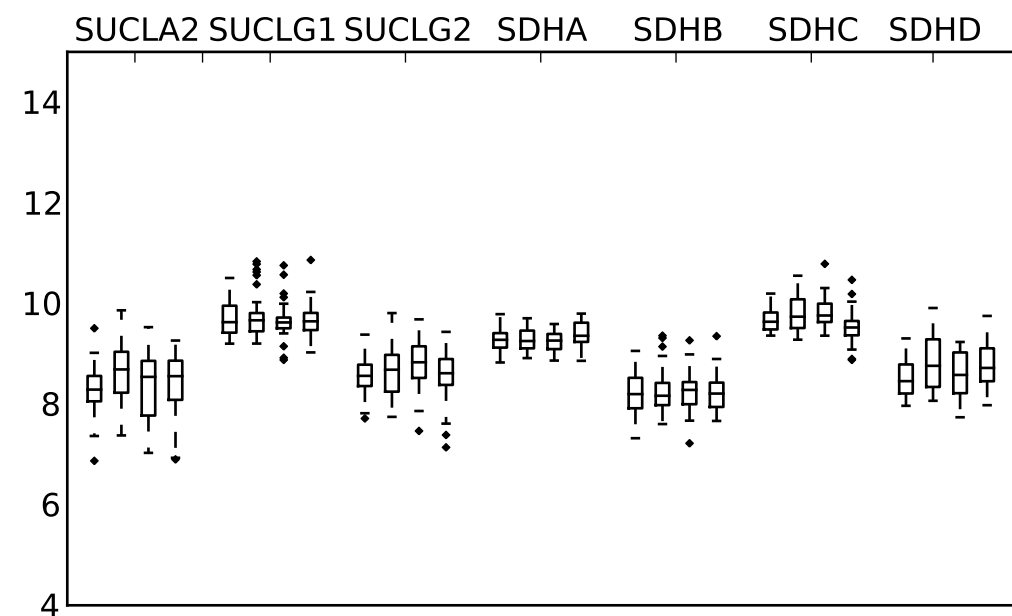
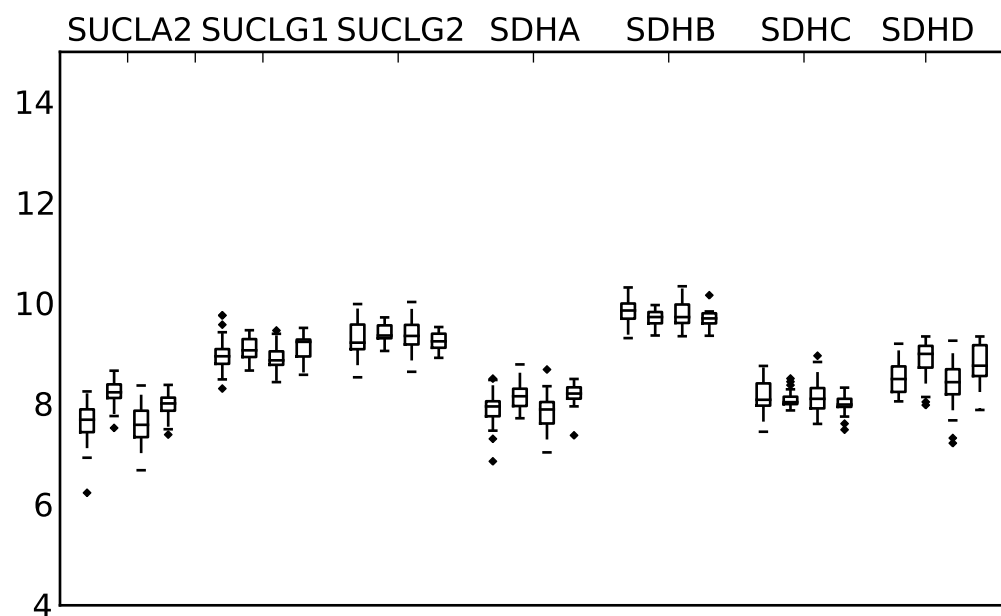
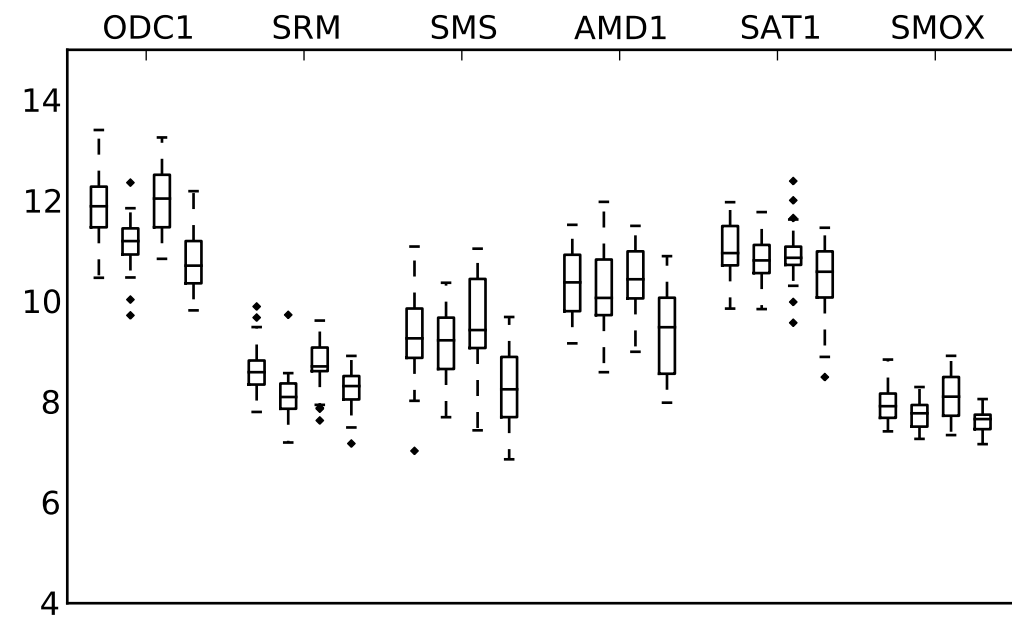
Figure D: Histogram of tissue percentage distribution for cancer, stroma and benign epithelium in the *balanced* and *unbalanced* datasets in the validation data.

log2 expression level

Main data

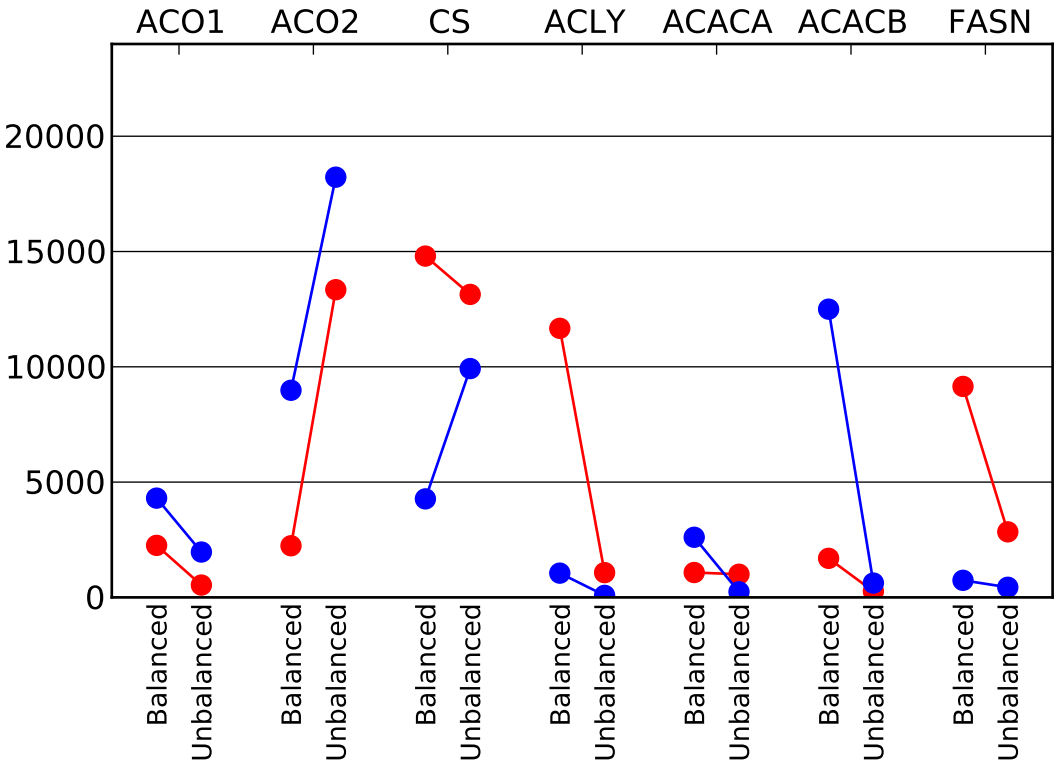
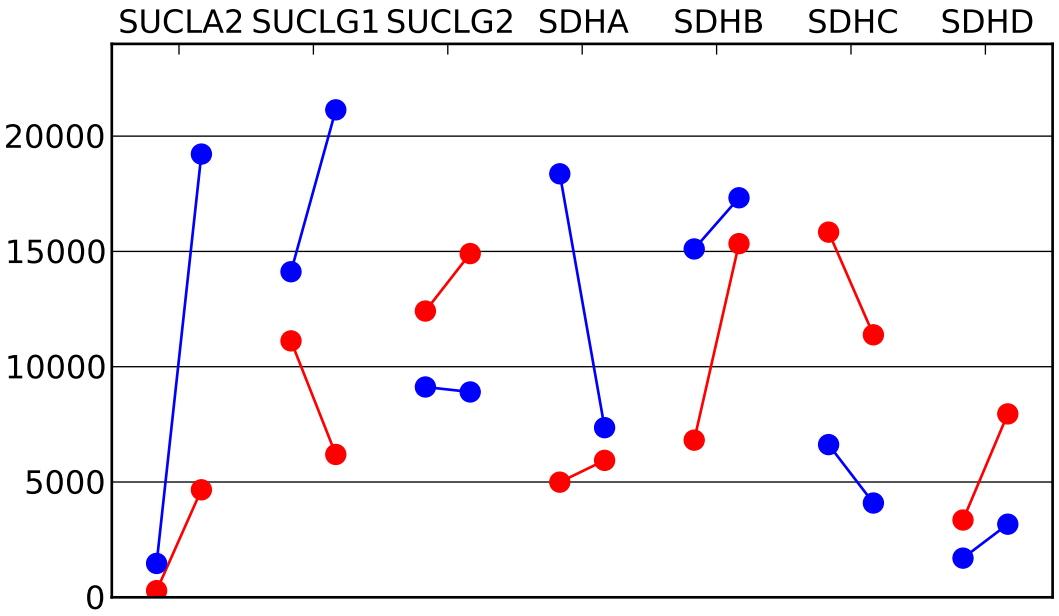
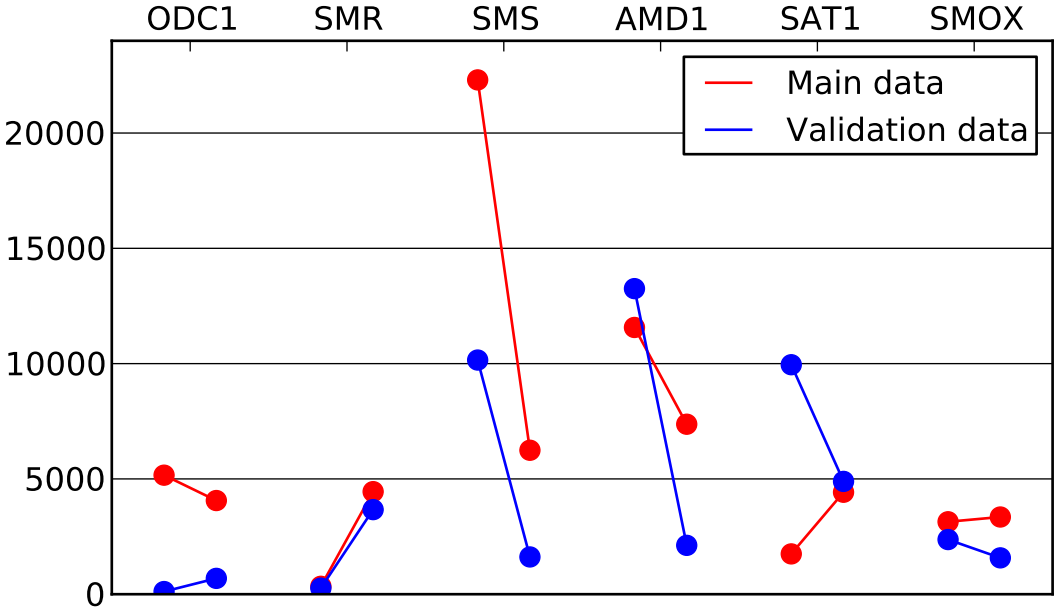


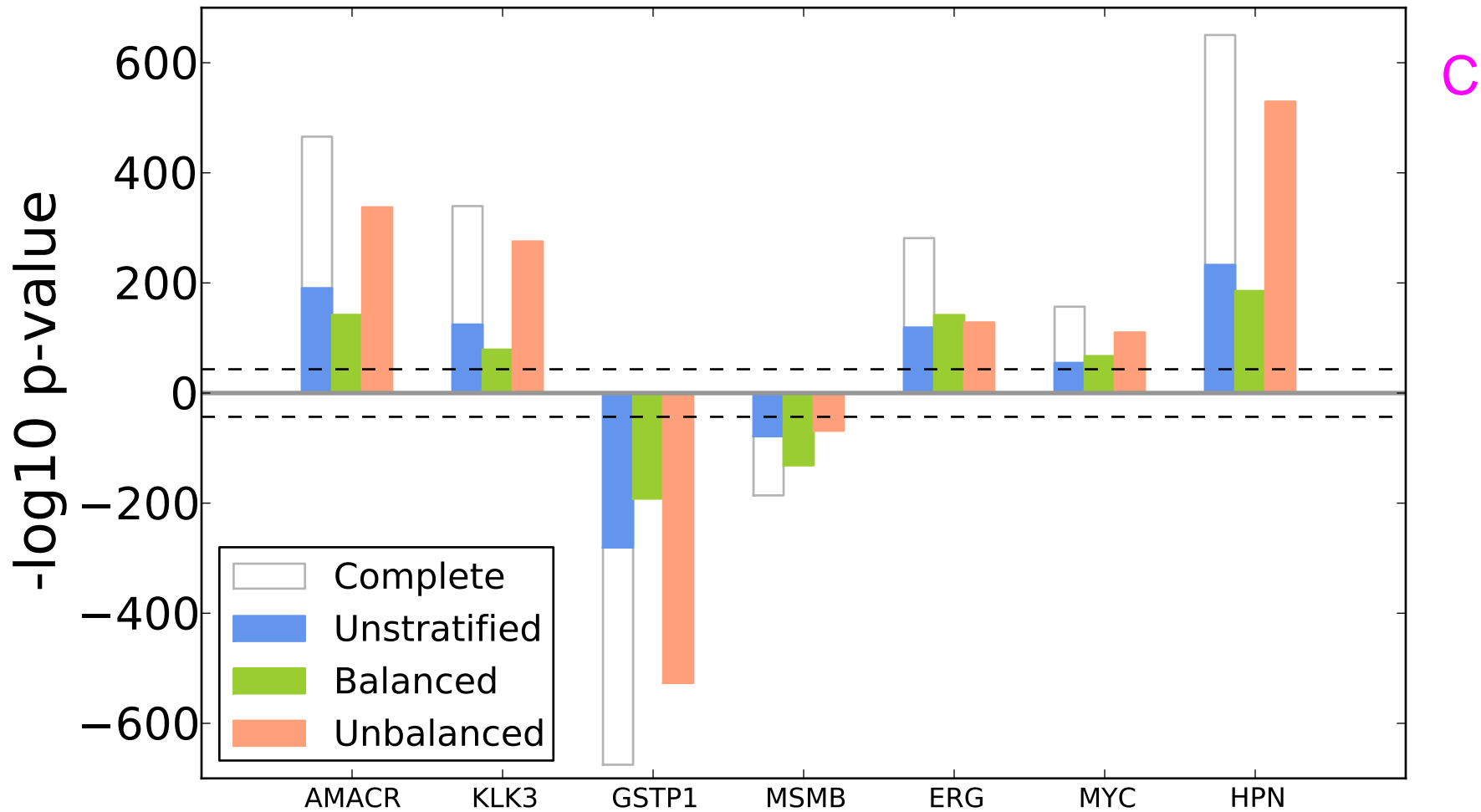
Validation data



A

B





C

D

