

Title

Bronchoalveolar Lavage Proteomics in Patients with Suspected Lung Cancer

Authors and Affiliations:

*Ana Sofia Carvalho^{1,2}, Célia Marina Cuco³, Carla Lavareda³, Francisco Miguel³,
Mafalda Ventura³, Sónia Almeida³, Paula Pinto^{3,4}, Tiago Tavares de Abreu³, Luís Vaz
Rodrigues⁵, Susana Seixas^{6,7}, Cristina Bárbara^{3,4}, Mikel Azkargorta⁷, Felix Elortza⁸,
Júlio Semedo³, John K Field⁸, Leonor Mota³, Rune Matthiesen^{1,2}*

- 1. Computational and Experimental Biology Group, Health Promotion and Chronic Diseases Department, National Institute of Health Dr Ricardo Jorge, Lisbon, Portugal.*
- 2. CEDOC, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa, Portugal*
- 3. Unidade de Técnicas Invasivas Pneumológicas, Pneumologia II, Hospital Pulido Valente, Centro Hospitalar Lisboa Norte, Lisbon, Portugal.*
- 4. Instituto de Saúde Ambiental, Faculdade de Medicina da Universidade de Lisboa*
- 5. Department of Pulmonology, Unidade Local de Saúde da Guarda, Faculty of Health Sciences, University of Beira Interior, Portugal*
- 6. Instituto de Investigação e Inovação em Saúde, Universidade do Porto (I3S), Porto, Portugal;*
- 7. Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal*
- 8. Proteomics Platform CIC bioGUNE, CIBERehd, ProteoRed-ISCI, Bizkaia Science and Technology Park, Derio, Spain.*
- 9. Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, University of Liverpool, William Duncan Building, 6 West Derby Street, Liverpool, L7 8TX, UK.*

Running Title

BAL proteomics in lung cancer diagnostics

Keywords

Lung Cancer, Bronchoalveolar Lavage, Diagnosis, Proteomics, Biomarkers

Supplementary figure legends

Figure S1. Distribution of Jaccard indexes comparing protein identifications from replica and non-replica samples.

Figure S2. Increase in the number of proteins (black bars) and protein encoding genes (grey bars) identified as sample size increase.

Figure S3. Principal component analysis using spectral count values from all abundant proteins. A) Annotated based on patients' lung cancer status as indicated in the legend. Black symbols indicate patients who died within the two years of follow-up. B) Annotated based on patients' lung cancer staging. Black symbols indicate patients who died within the two years of follow-up. C) Annotated based on patients' gender. D) Annotated based on patients' smoking status.

Figure S4. Comparison of ratios obtained from VEMS and MaxQuant analysis for biomarkers found significant regulated for non-lung cancer versus lung cancer by both software.

Figure S5. Heatmap of significant regulated proteins based on iBAQ values obtained from VEMS and MaxQuant. Left bar depicts lung cancer cases (red) versus non lung cancer controls (green).

Figure S6. Boxplot of iBAQ expression values for all significant regulated proteins between cases and controls in the consensus list. Barplot titles includes gene name, P value correct for multiple testing, M log2 fold ratio and N the sample size.

Figure S7. Boxplot of iBAQ expression values for all significant regulated proteins between cases and controls factored on histology groups. Barplot titles includes gene name, P value correct for multiple testing, M log2 fold ratio and N sample size. The indicated P values results from comparing cases and controls.

Figure S8. Heatmap of protein biomarkers identified exclusively in lung cancer samples. The ten most significantly functional enriched KEGG pathways are indicated in the top of the heatmap where a black bars depicts which functional categories the protein belong.

Figure S9. Heatmap of peptide mziXIC values for most abundant peptides. "Yes" and "No" indicates the patients status with respect to lung cancer. Two sub clusters are highlighted with only lung cancer (light red) and non-lung cancer (light green).

Figure S10. Significantly enriched GO terms within the cellular component category (CC) for all identified proteins across all patient samples. The heatmap depicts the log2 number of proteins identified in different cellular component categories across patient samples.

Figure S11. Significantly enriched GO terms within the cellular component (CC) category for proteins in the consensus list across all patient samples. The heatmap depicts the log2 number of proteins in consensus list and in different cellular component categories across patient samples.

Figure S12. Significantly enriched GO terms within the molecular function (MF) category for all identified proteins across all patient samples. The heatmap depicts the log2 number of proteins identified in different molecular function categories across patient samples.

Figure S13. Significantly enriched GO terms within the molecular function (MF) category for proteins in the consensus list across all patient samples. The heatmap depicts the log2 number of proteins in consensus list and in different functional molecular function categories across patient samples.

Figure S14. Significantly enriched GO terms within the biological process (BP) category for all identified proteins across all patient samples. The heatmap depicts the log2 number of proteins identified in different biological process categories across patient samples.

Figure S15. Significantly enriched GO terms within the biological process category (BP) for proteins in the consensus list across all patient samples. The heatmap depicts the log2 number of proteins in consensus list and in different biological process categories across patient samples.

Figure S16. Boxplot factored on lung cancer versus non-lung cancer cases of total spectral counts for all identified proteins annotated as “extracellular vesicular exosome”.

Figure S17. Venn diagram comparing identified proteins in five major cellular component categories.

Figure S18. Barplot of mean log2(iBAQ+1) for proteins specific for one of the major cellular component category in Figure S15. The numbers of the bars indicates the number of specific proteins for each of the major cellular component categories that were utilized for the calculation of mean log2(iBAQ+1) values.

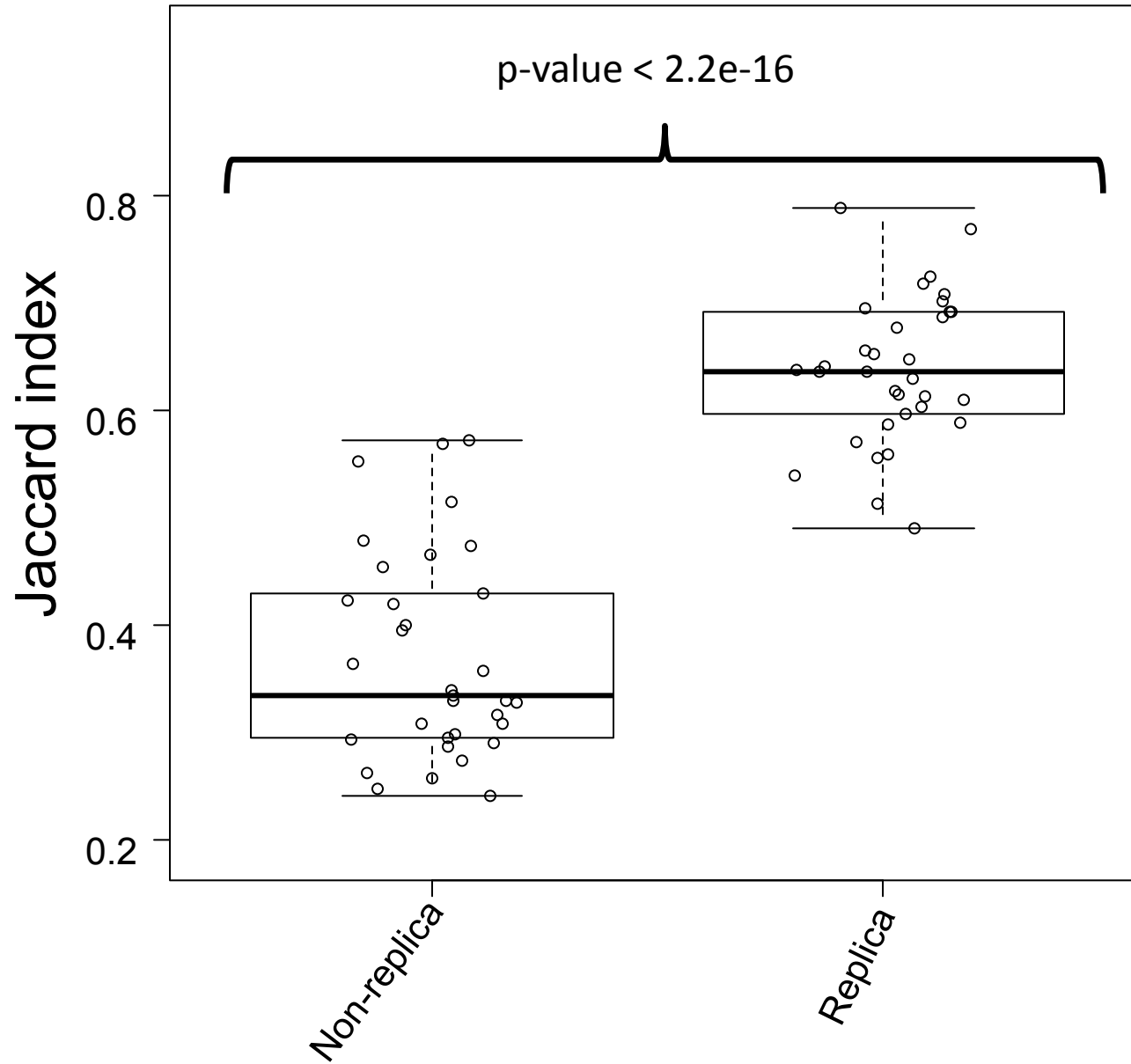
Figure S19. GO enrichment analysis (biological process) of all identified proteins (A) and significant regulated proteins between cases and controls (B). The y-axis indicates the number of proteins in each biological process category for all identified proteins and for the significantly regulated proteins. The numbers on top of the columns indicates the number of proteins in each category in total.

Figure S20. Reproducibility of VEMS/MaxQuant consensus lung cancer biomarkers in other OMICs lung cancer studies published in the literature. The proposed genes are

order with the genes reproduced most times in the top. The heatmap depicts down-regulated in green, up-regulated in red and non-significant or not detected in blue.

Figure S21. Boxplot comparing iBAQ expression of CD molecules from human leucocyte in non-lung cancer versus lung cancer (<http://www.hcdm.org>). The iBAQ expression values were compared by a Wilcox test. N indicates the number of markers that were identified by MS for the specific Leucocyte.

Distribution of Jaccard indexes



Increase in identified proteins as sample size increase

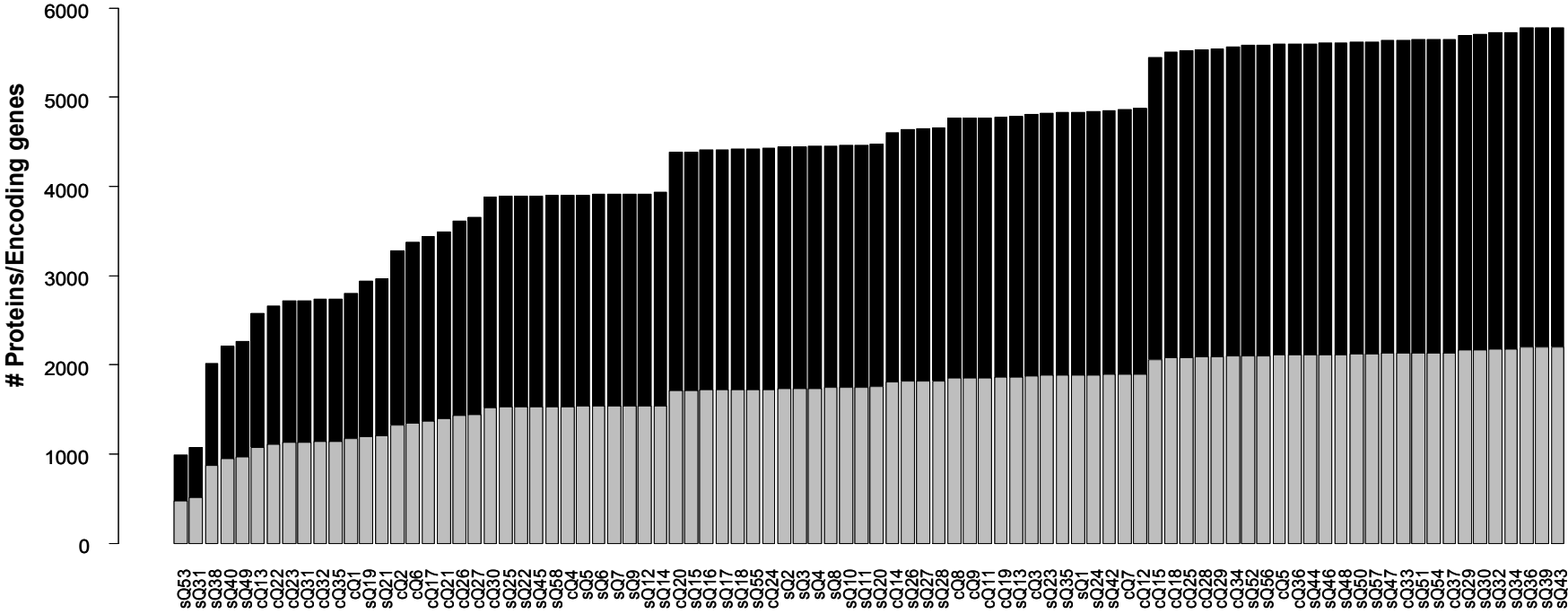


Figure S2

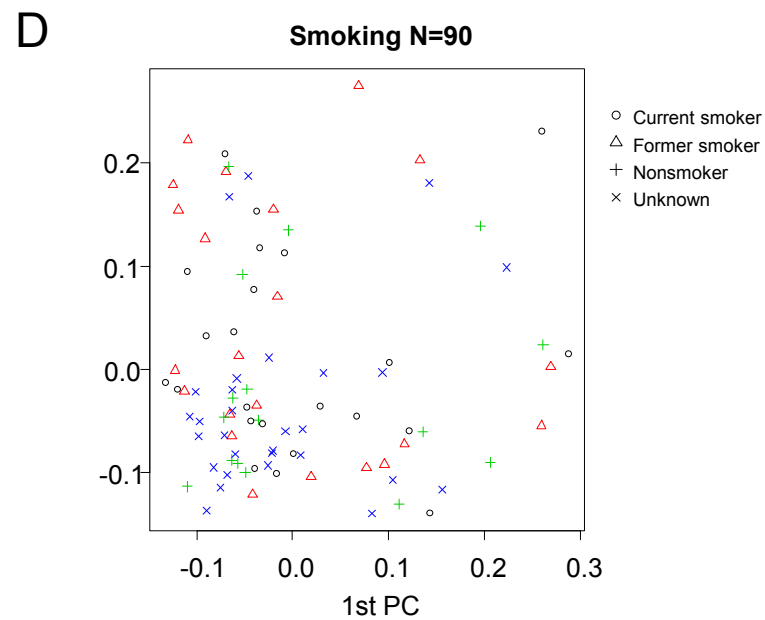
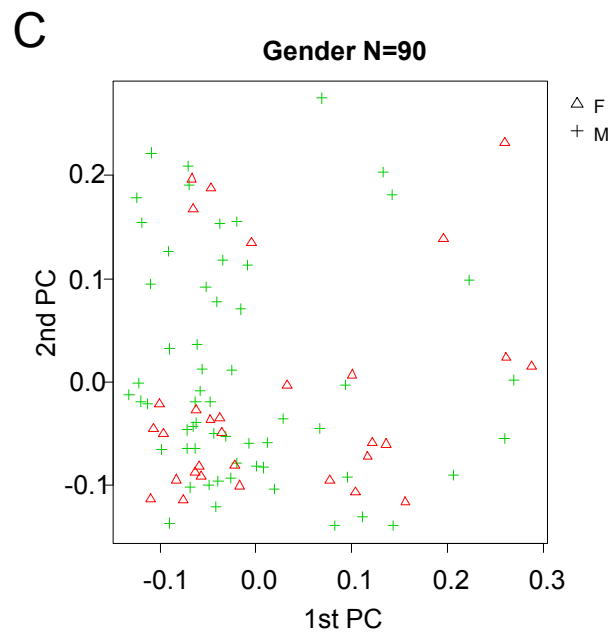
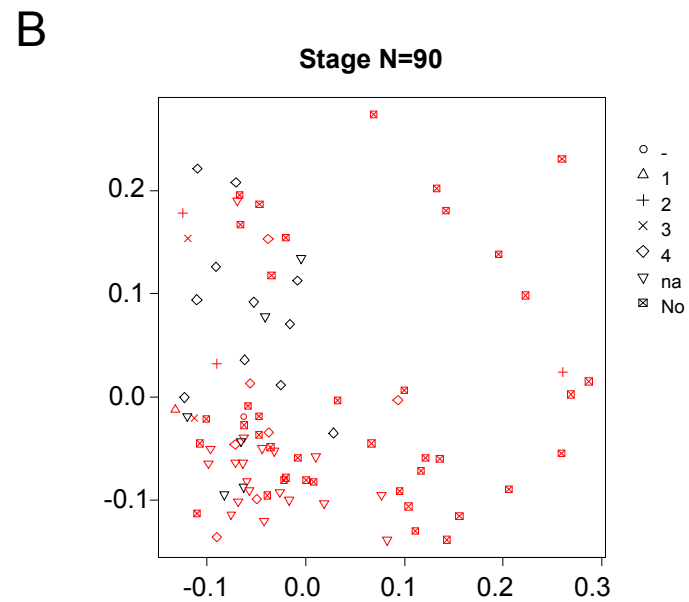
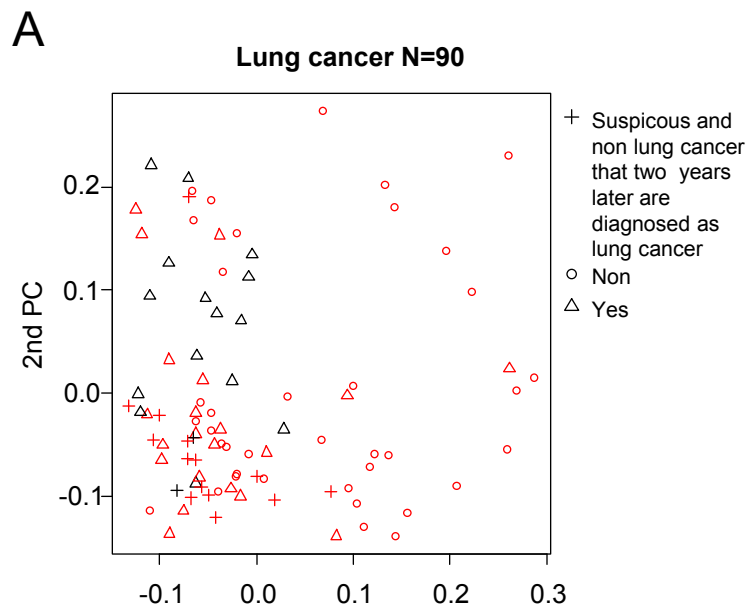


Figure S3

Comparisons of ratios obtained from VEMS and MaxQuant

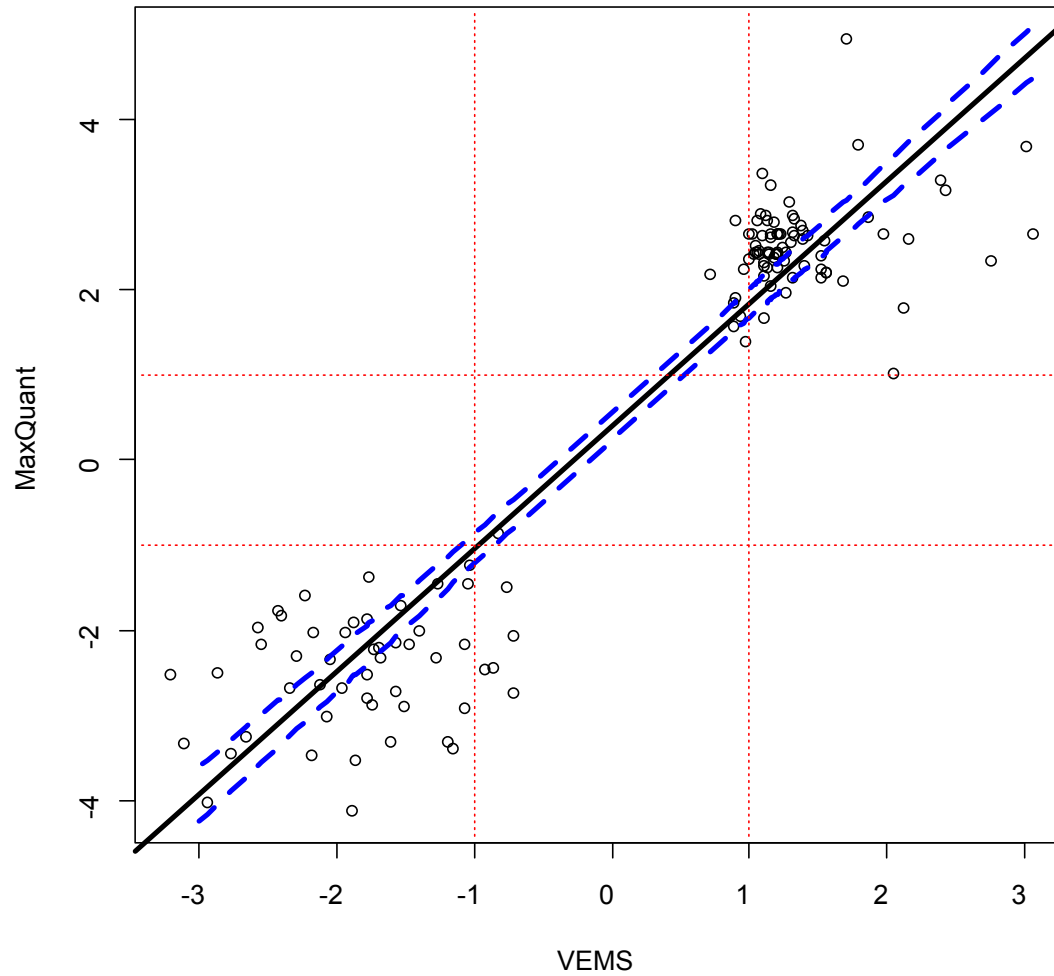
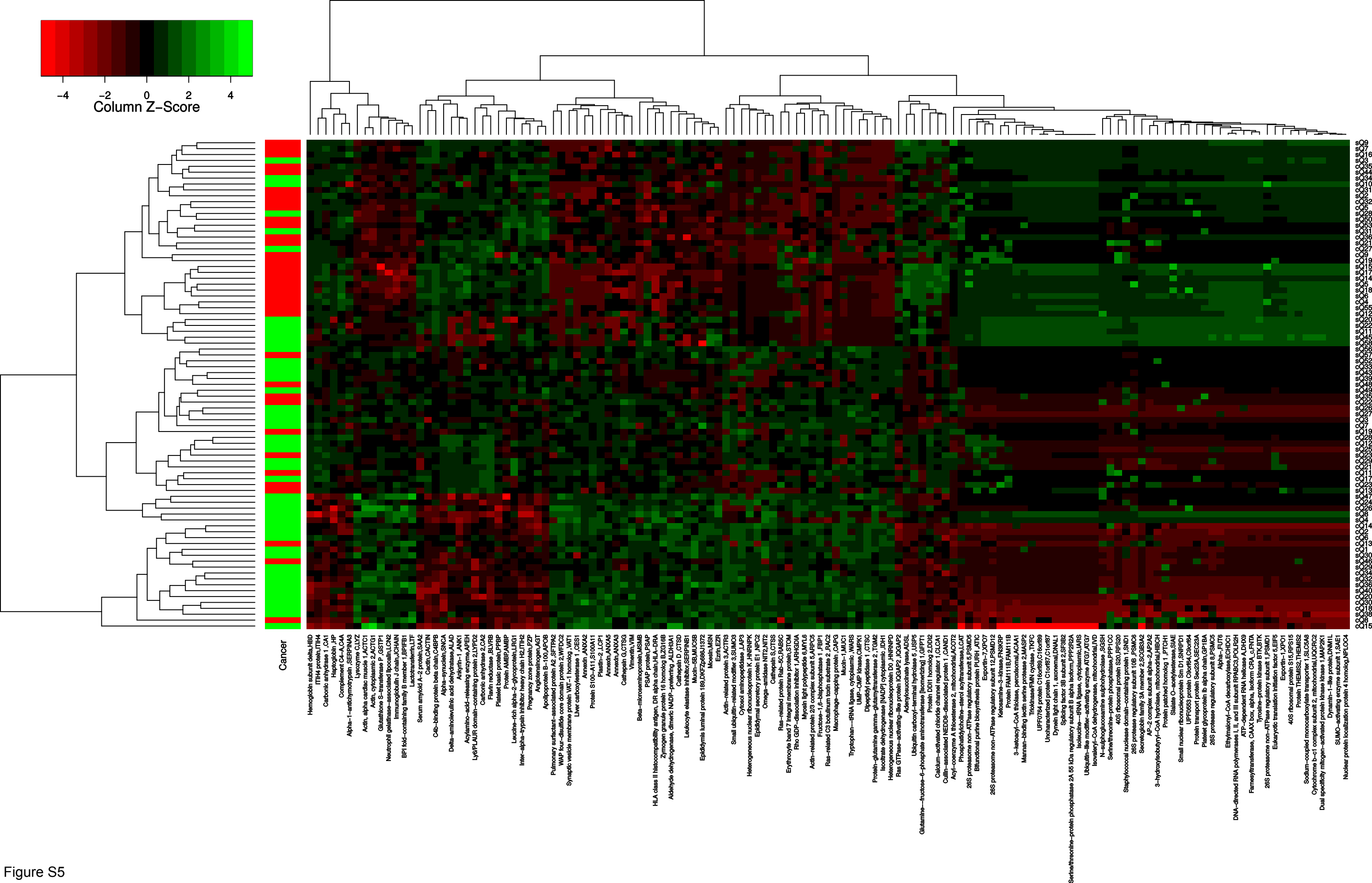


Figure S4



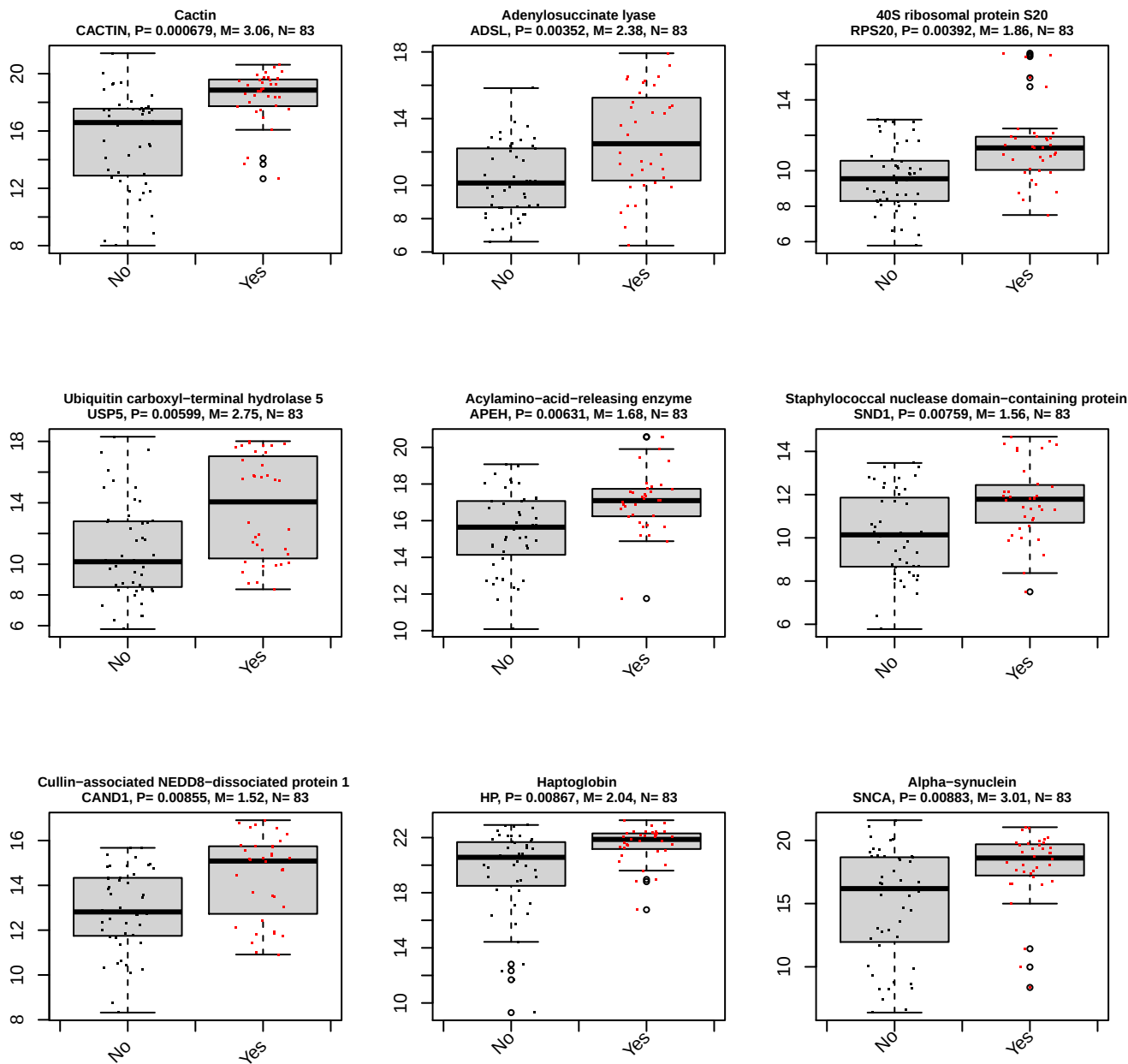
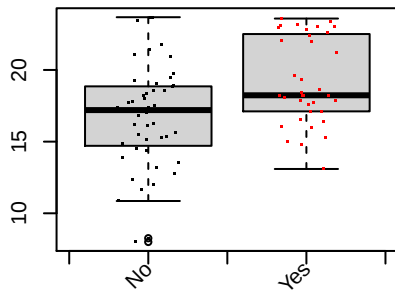
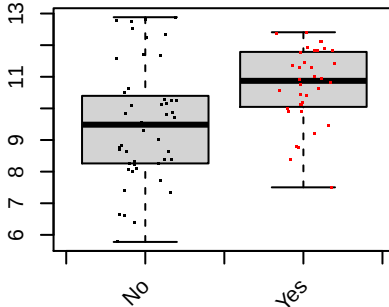


Figure S6

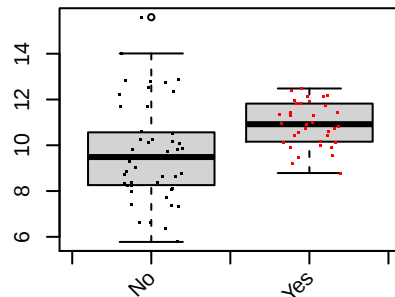
C4b-binding protein beta chain
C4BPB, P= 0.0168, M= 2.43, N= 83



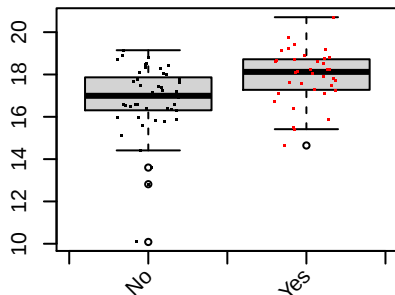
DNA-directed RNA polymerases I, II, and III subunit RPA
POLR2H, P= 0.0187, M= 1.19, N= 83



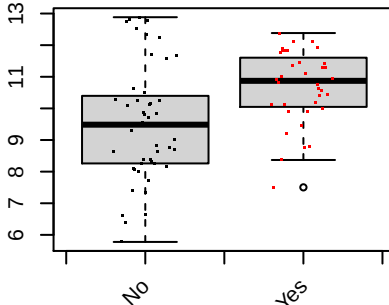
Protein transport protein Sec23A
SEC23A, P= 0.0205, M= 1.26, N= 83



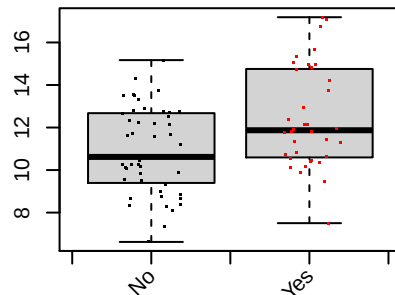
Protein AMBP
AMBP, P= 0.0209, M= 1.1, N= 83



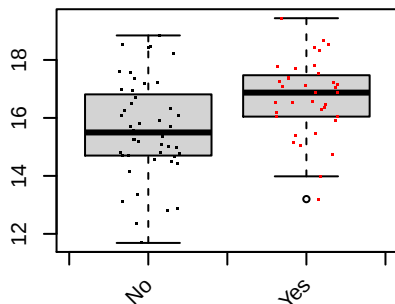
Ethylmalonyl-CoA decarboxylase
ECHDC1, P= 0.021, M= 1.15, N= 83



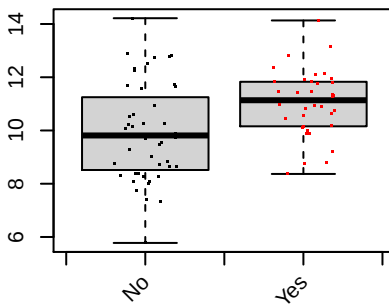
Acyl-coenzyme A thioesterase 2, mitochondrial
ACOT2, P= 0.0232, M= 1.52, N= 83



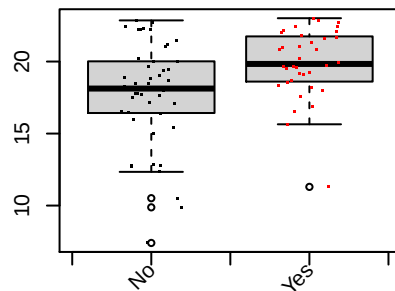
Ankyrin-1
ANK1, P= 0.0232, M= 1.1, N= 83



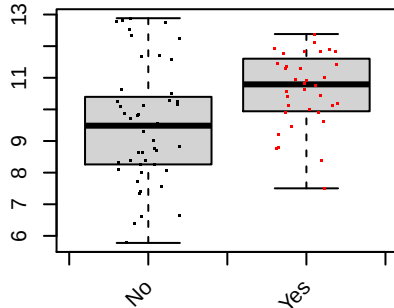
3-hydroxyisobutyryl-CoA hydrolase, mitochondrial
HIBCH, P= 0.0239, M= 1.13, N= 83



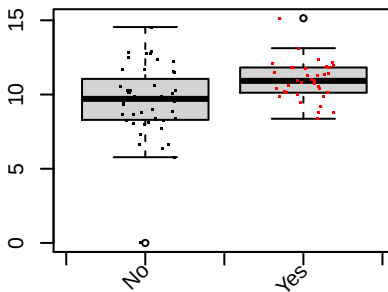
Carbonic anhydrase 2
CA2, P= 0.0243, M= 2.12, N= 83



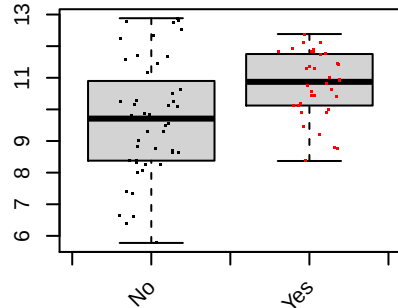
Farnesyltransferase, CAAX box, alpha, isoform CRA_
FNTA, P= 0.0249, M= 1.13, N= 83



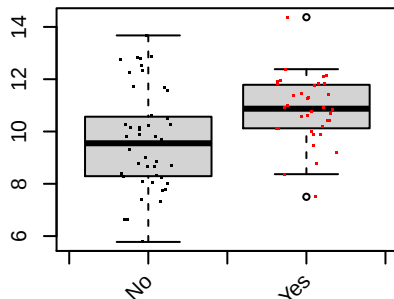
Secretoglobulin family 3A member 2
SCGB3A2, P= 0.0254, M= 1.39, N= 83



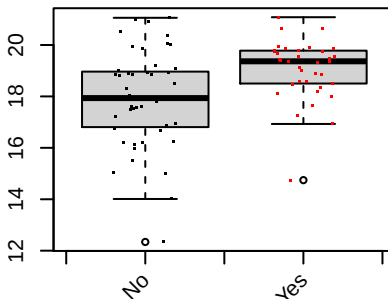
Alpha-FUT6, P= 0.0254, M= 1.09, N= 83



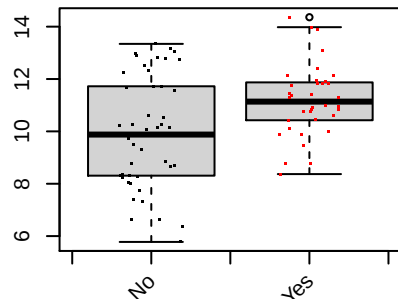
Platelet glycoprotein Ib alpha chain
GP1BA, P= 0.0254, M= 1.15, N= 83



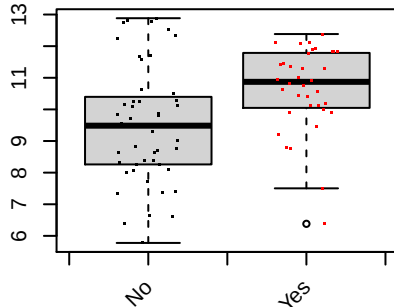
Pregnancy zone protein
PZP, P= 0.0256, M= 1.12, N= 83



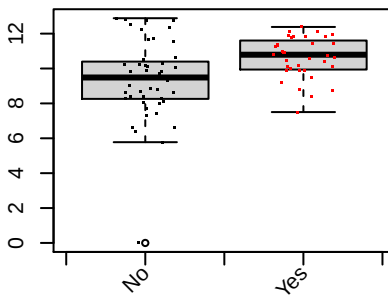
26S proteasome non-ATPase regulatory subunit 12
PSMD12, P= 0.0262, M= 1.25, N= 83



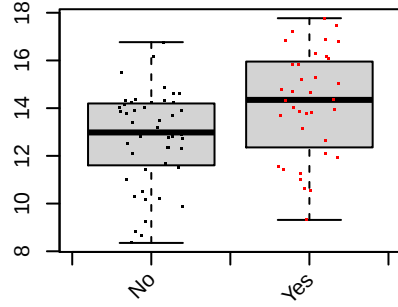
Protein THEMIS2
THEMIS2, P= 0.0264, M= 1.14, N= 83



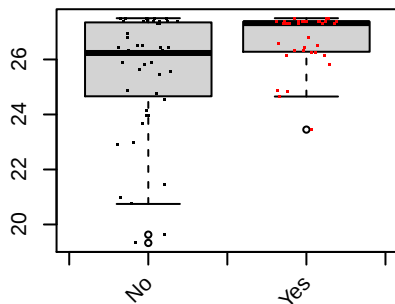
Uncharacterized protein C1orf87
C1orf87, P= 0.0265, M= 1.29, N= 83



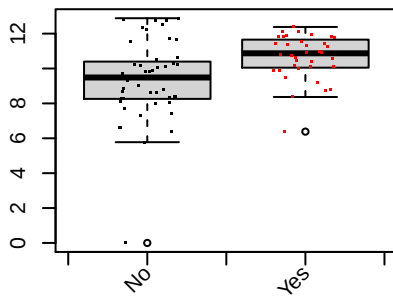
Calcium-activated chloride channel regulator 1
CLCA1, P= 0.0265, M= 1.4, N= 83



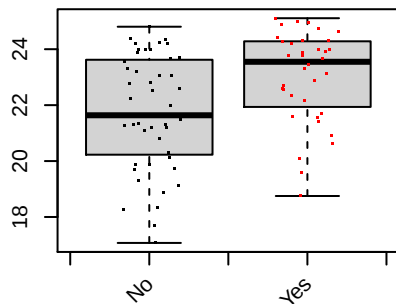
Hemoglobin subunit delta
HBD, P= 0.0265, M= 1.21, N= 83



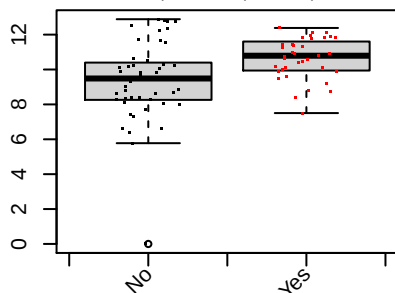
Triokinase/FMN cyclase
TKFC, P= 0.0265, M= 1.31, N= 83



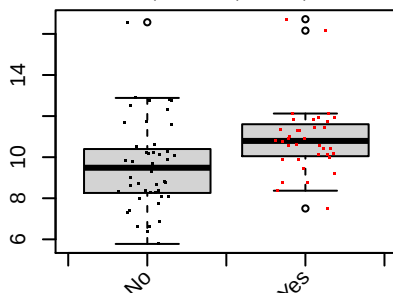
Carbonic anhydrase 1
CA1, P= 0.0267, M= 1.27, N= 83



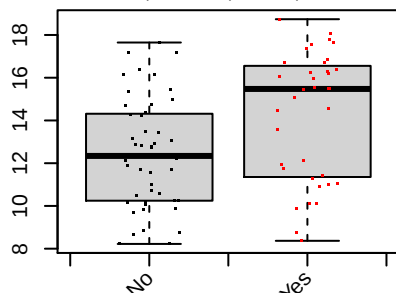
UPF0764 protein C16orf89
C16orf89, P= 0.0267, M= 1.29, N= 83



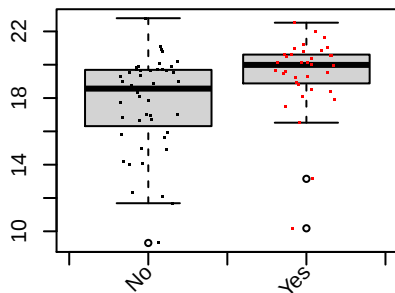
26S proteasome non-ATPase regulatory subunit 1
PSMD1, P= 0.0284, M= 1.32, N= 83



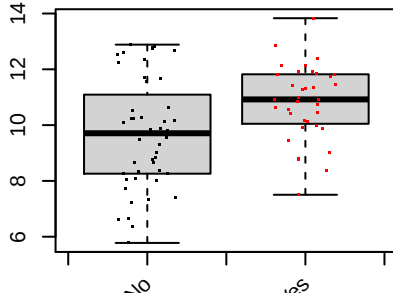
Protein DD11 homolog 2
DDI2, P= 0.0294, M= 1.79, N= 83



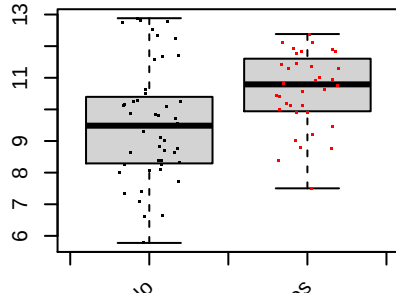
Platelet basic protein
PPBP, P= 0.0295, M= 1.7, N= 83



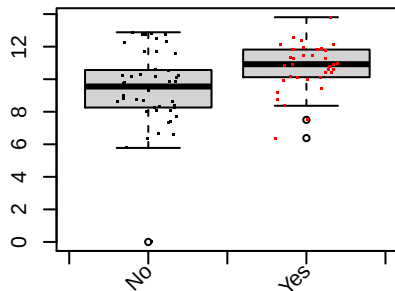
Eukaryotic translation initiation factor 5
EIF5, P= 0.0304, M= 1.15, N= 83



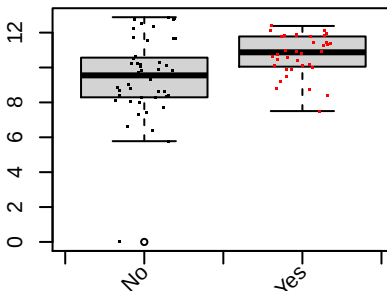
Tyrosine-protein kinase BTK
BTK, P= 0.0304, M= 1.06, N= 83



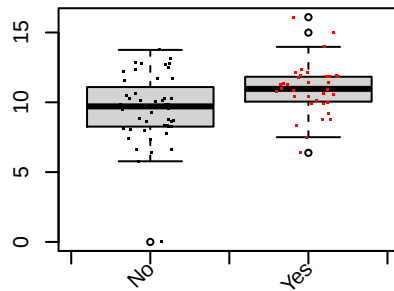
Mannan-binding lectin serine protease 2
MASP2, P= 0.0314, M= 1.32, N= 83



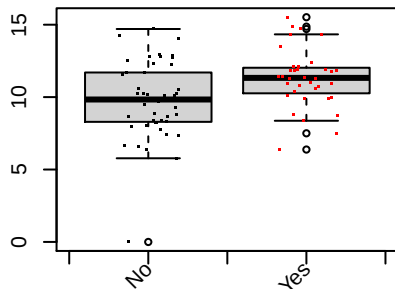
3-ketoacyl-CoA thiolase, peroxisomal
ACAA1, P= 0.0317, M= 1.24, N= 83



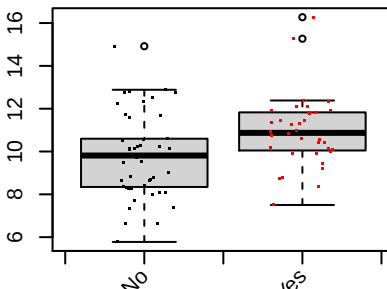
Ketosamine-3-kinase
FN3KRP, P= 0.0325, M= 1.42, N= 83



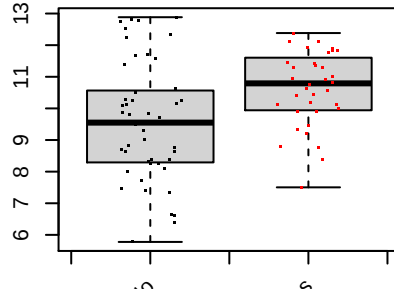
Bifunctional purine biosynthesis protein PURH
ATIC, P= 0.0341, M= 1.52, N= 83



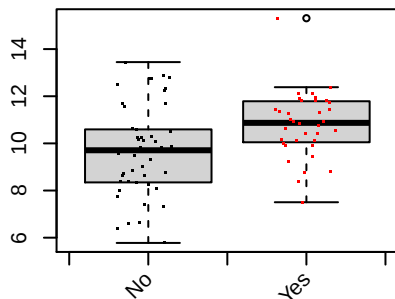
UPF0553 protein C9orf64
C9orf64, P= 0.0344, M= 1.18, N= 83



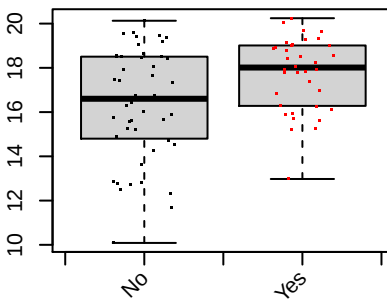
ATP-dependent RNA helicase A
DHX9, P= 0.0345, M= 1.05, N= 83



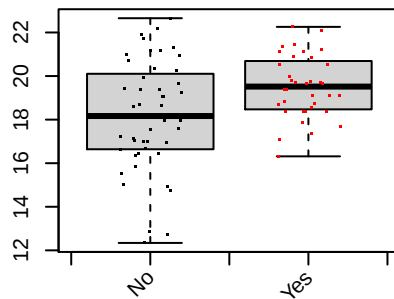
Small nuclear ribonucleoprotein Sm D1
SNRPD1, P= 0.0348, M= 1.11, N= 83



Delta-aminolevulinic acid dehydratase
ALAD, P= 0.0349, M= 1.38, N= 83

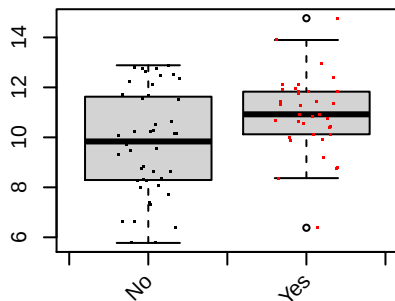


Flavin reductase
BLVRB, P= 0.0349, M= 1.33, N= 83



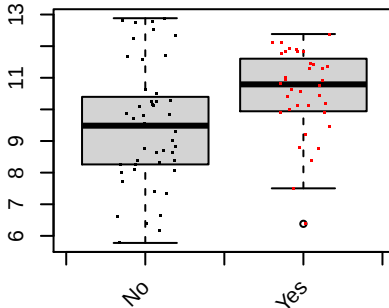
Exportin-7

XPO7, P= 0.0349, M= 1.16, N= 83



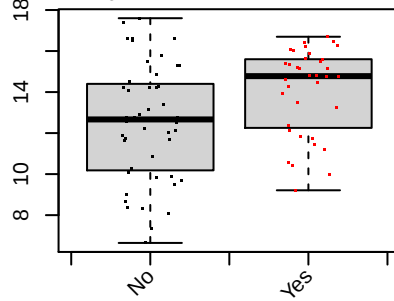
SUMO-activating enzyme subunit 1

SAE1, P= 0.0349, M= 1.07, N= 83



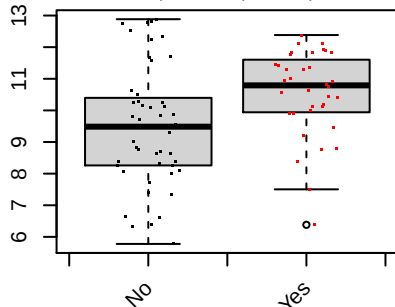
Ras GTPase-activating-like protein IQGAP2

IQGAP2, P= 0.0349, M= 1.55, N= 83



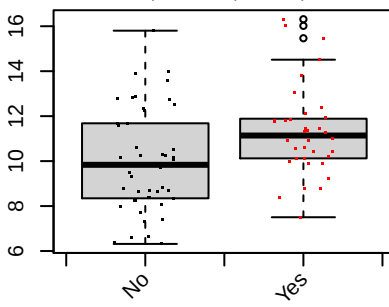
Nuclear protein localization protein 4 homolog

NPLOC4, P= 0.0349, M= 1.06, N= 83



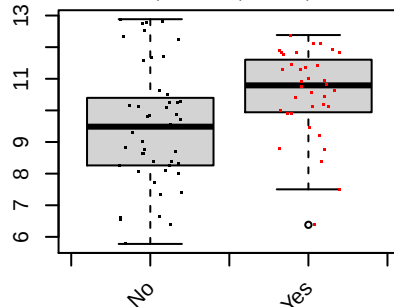
26S proteasome non-ATPase regulatory subunit 5

PSMD5, P= 0.0349, M= 1.33, N= 83



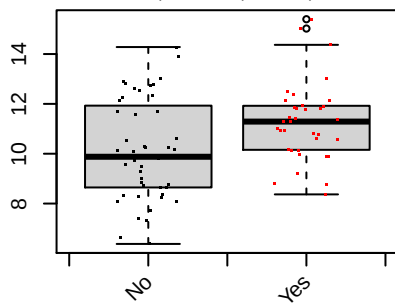
Dynammin-1-like protein

DNM1L, P= 0.0349, M= 1.06, N= 83



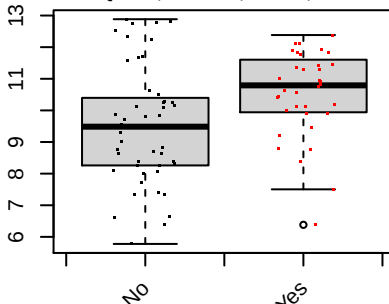
Serine/threonine-protein phosphatase

PPP1CC, P= 0.0349, M= 1.13, N= 83



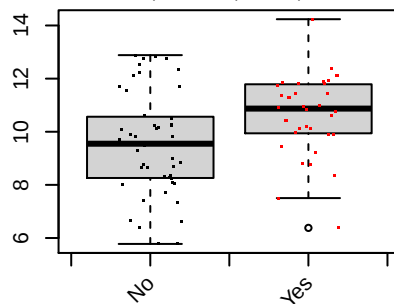
Cytochrome b-c1 complex subunit 2, mitochondrial

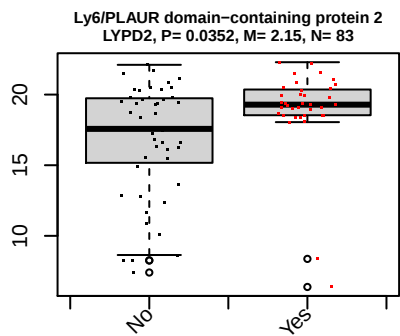
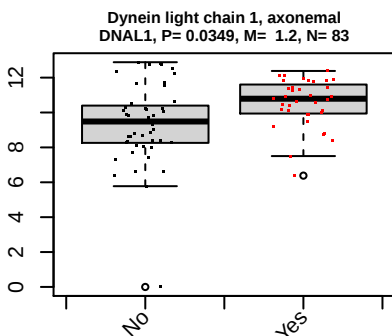
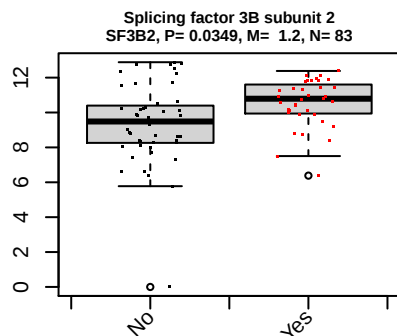
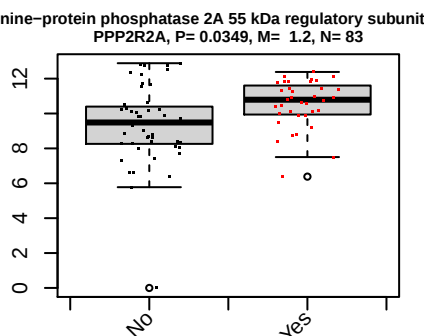
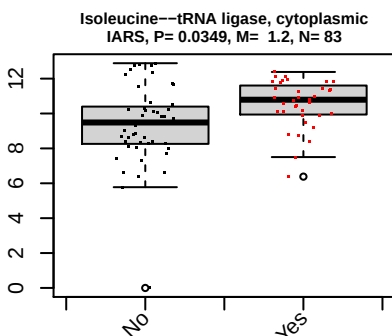
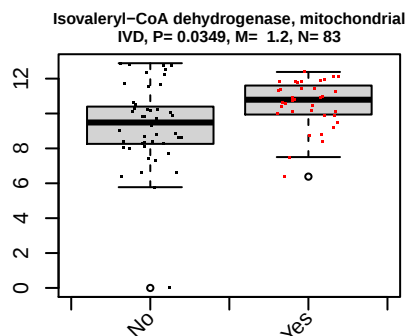
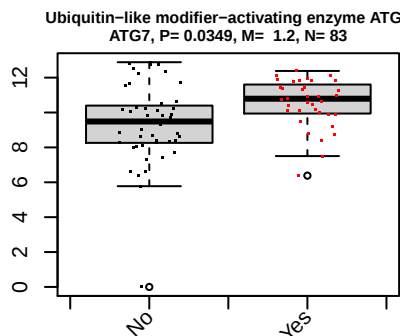
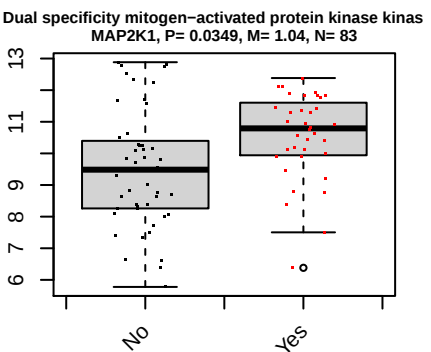
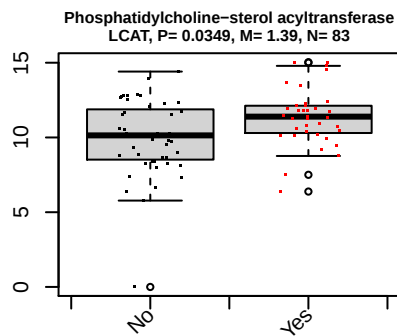
UQCRC2, P= 0.0349, M= 1.04, N= 83



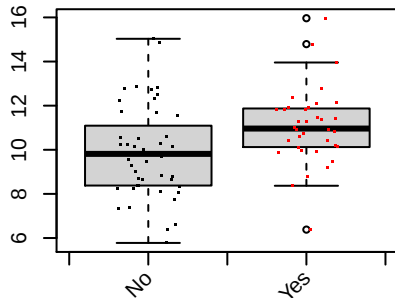
Exportin-1

XPO1, P= 0.0349, M= 1.1, N= 83

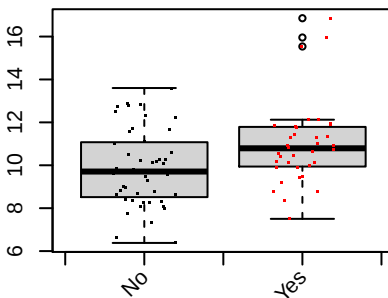




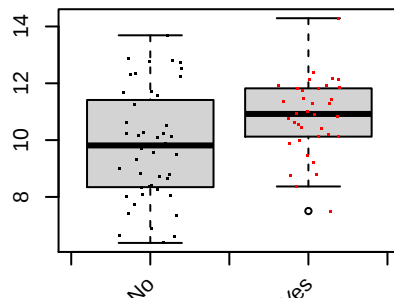
N-sulphoglucosamine sulphohydrolase
SGSH, P= 0.0368, M= 1.17, N= 83



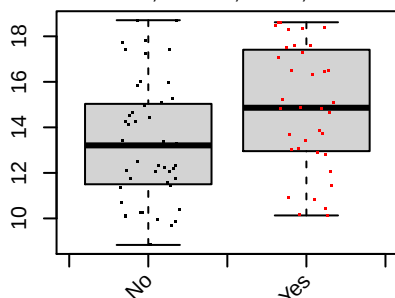
Sialate O-acetyltransferase
SIAE, P= 0.0375, M= 1.11, N= 83



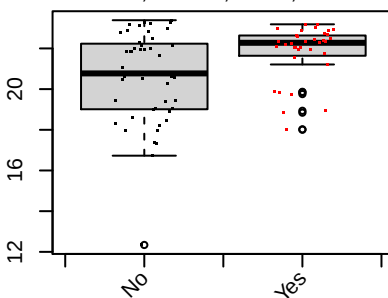
Protein patched homolog 1
PTCH1, P= 0.0382, M= 1.02, N= 83



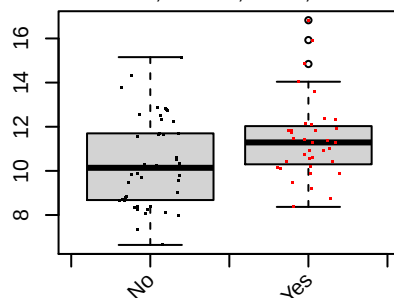
lutamine--fructose-6-phosphate aminotransferase [isome
GFPT1, P= 0.0415, M= 1.54, N= 83



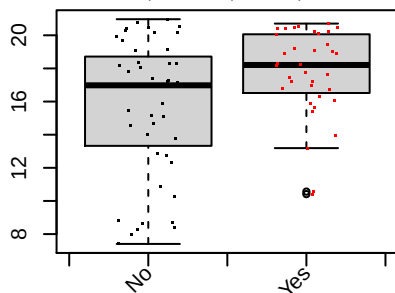
ITI4 protein
ITI4, P= 0.0417, M= 1.15, N= 83



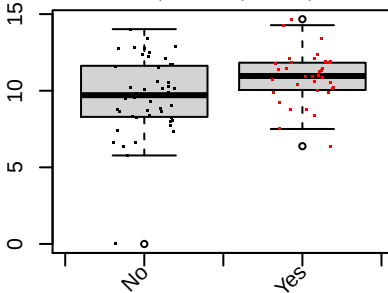
26S protease regulatory subunit 10B
PSMC6, P= 0.0438, M= 1.11, N= 83



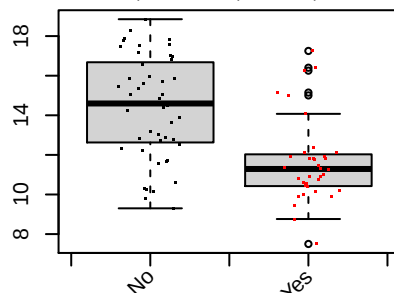
Serum amyloid A-2 protein
SAA2, P= 0.0461, M= 1.98, N= 83



Protein FAM111B
FAM111B, P= 0.0487, M= 1.23, N= 83

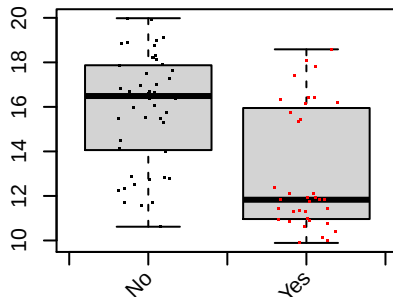


Epididymal secretory protein E1
NPC2, P= 0.000528, M= -2.77, N= 83



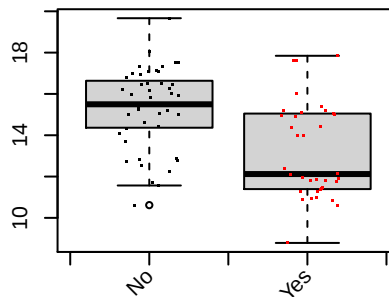
Dipeptidyl peptidase 1

CTSC, P= 0.000679, M= -2.87, N= 83



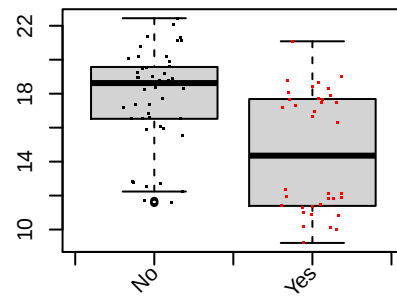
Protein-glutamine gamma-glutamyltransferase 2

TGM2, P= 0.00138, M= -2.17, N= 83



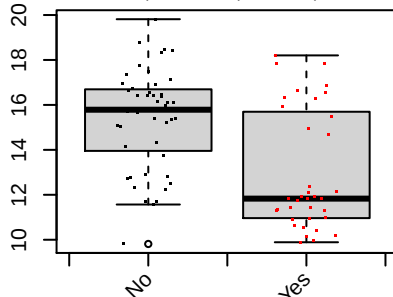
Pulmonary surfactant-associated protein A2

SFTPA2, P= 0.0014, M= -3.21, N= 83



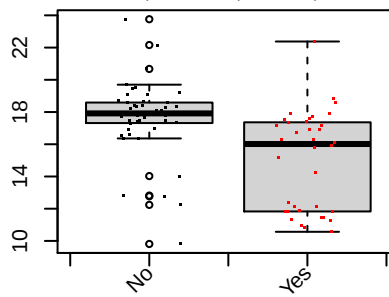
UMP-CMP kinase

CMPK1, P= 0.00177, M= -2.43, N= 83



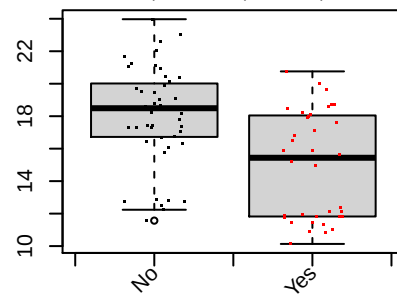
Annexin

ANXA2, P= 0.00177, M= -2.66, N= 83



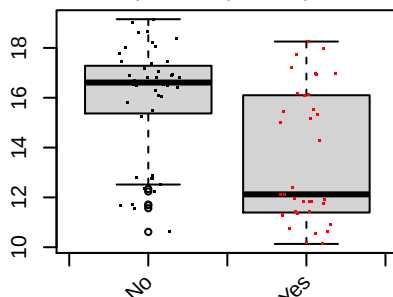
WAP four-disulfide core domain protein 2

WFDC2, P= 0.00177, M= -3.11, N= 83



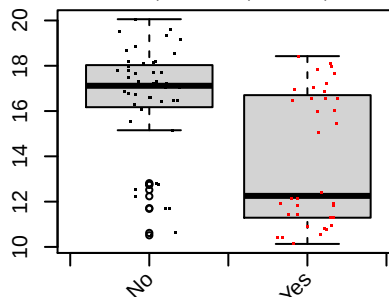
Isocitrate dehydrogenase [NADP] cytoplasmic

IDH1, P= 0.00277, M= -2.34, N= 83



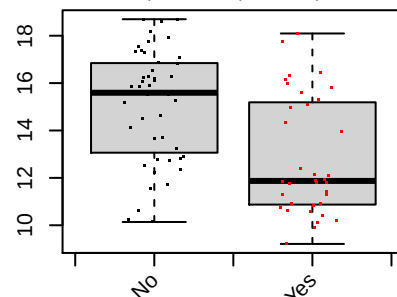
Rho GDP-dissociation inhibitor 1

ARHGDI1, P= 0.00277, M= -2.55, N= 83

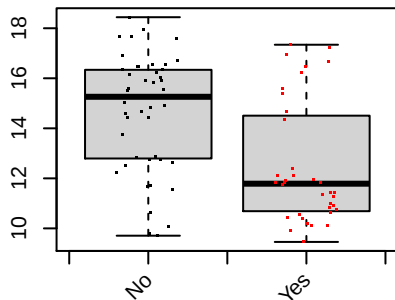


Tryptophan--tRNA ligase, cytoplasmic

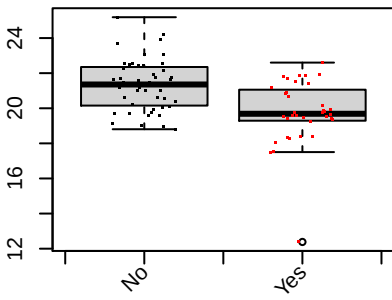
WARS, P= 0.00277, M= -2.29, N= 83



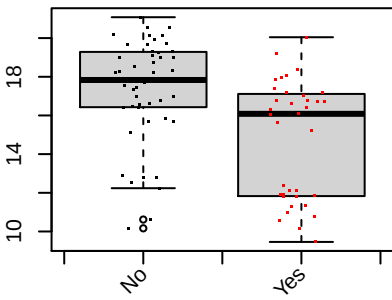
Heterogeneous nuclear ribonucleoprotein K
HNRNPK, P= 0.00288, M= -2.23, N= 83



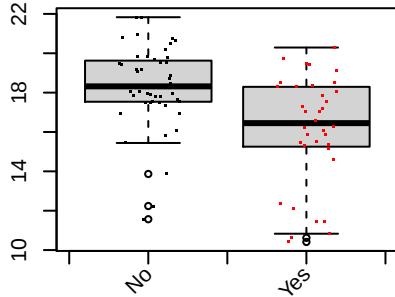
Glutathione S-transferase P
GSTP1, P= 0.00333, M= -1.57, N= 83



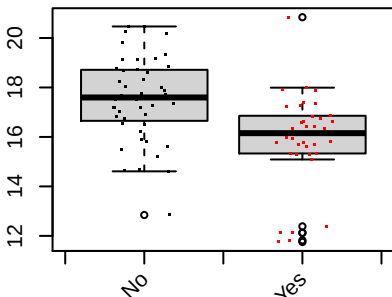
Plastin-2
LCP1, P= 0.00442, M= -2.58, N= 83



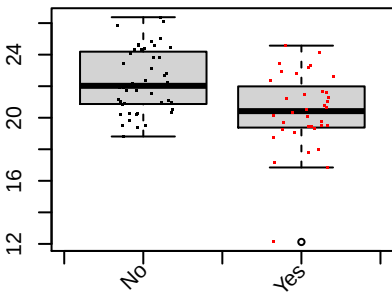
PSAP protein
PSAP, P= 0.0047, M= -2.18, N= 83



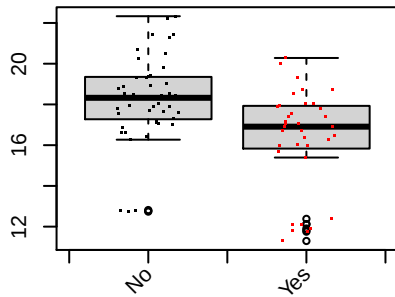
Cathepsin G
CTSG, P= 0.00498, M= -1.61, N= 83



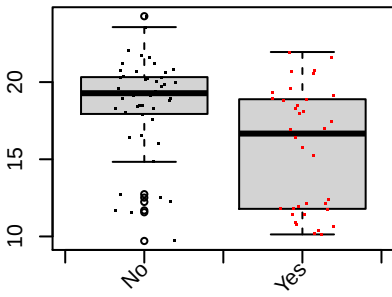
Neutrophil gelatinase-associated lipocalin
LCN2, P= 0.00599, M= -1.88, N= 83



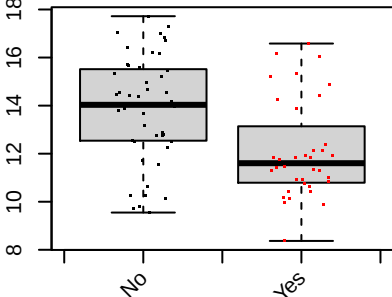
Vimentin
VIM, P= 0.00599, M= -1.96, N= 83



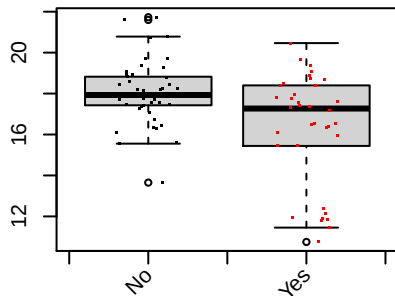
Beta-microseminoprotein
MSMB, P= 0.00651, M= -2.95, N= 83



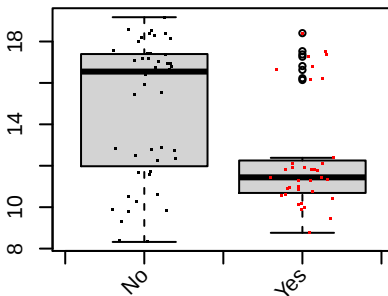
Omega-amidase NIT2
NIT2, P= 0.00804, M= -1.76, N= 83



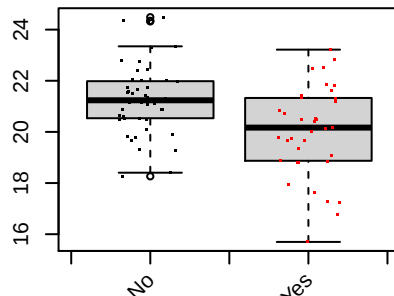
HLA class II histocompatibility antigen, DR alpha chain
HLA-DRA, P= 0.00867, M= -1.73, N= 83



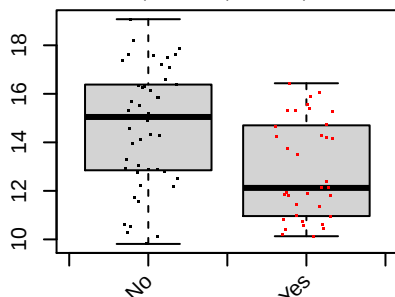
Actin-related protein 3
ACTR3, P= 0.0095, M= -2.41, N= 83



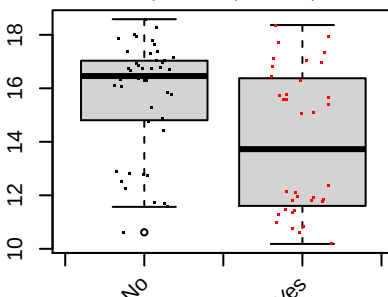
Immunoglobulin J chain
JCHAIN, P= 0.01, M= -1.27, N= 83



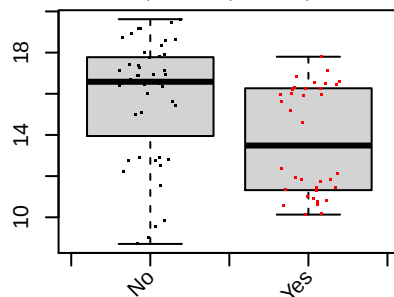
Cathepsin S
CTSS, P= 0.0126, M= -1.74, N= 83



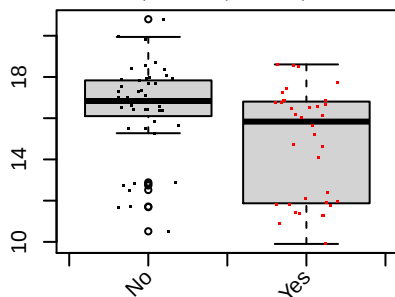
Heterogeneous nuclear ribonucleoprotein D0
HNRNPD, P= 0.0134, M= -1.78, N= 83



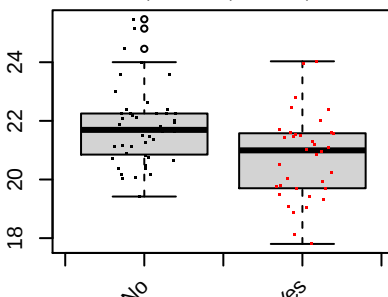
Mucin-1
MUC1, P= 0.0147, M= -2.07, N= 83



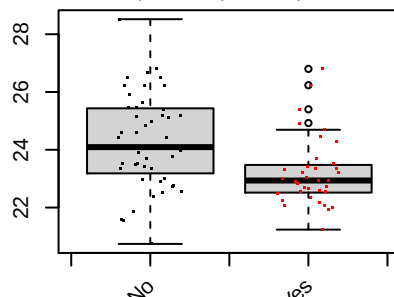
Fructose-1,6-bisphosphatase 1
FBP1, P= 0.0159, M= -1.77, N= 83



BPI fold-containing family B member 1
BPIFB1, P= 0.0192, M= -1.03, N= 83

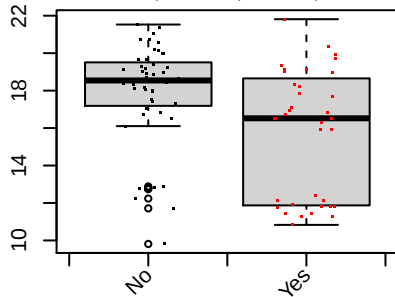


Lysozyme C
LYZ, P= 0.0207, M= -1.07, N= 83



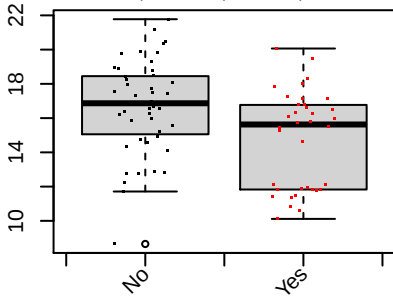
Protein S100-A11

S100A11, P= 0.0207, M= -2.12, N= 83



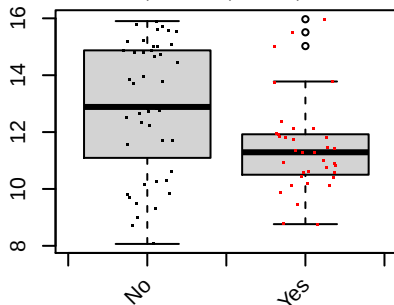
Liver carboxylesterase 1

CES1, P= 0.0207, M= -1.94, N= 83



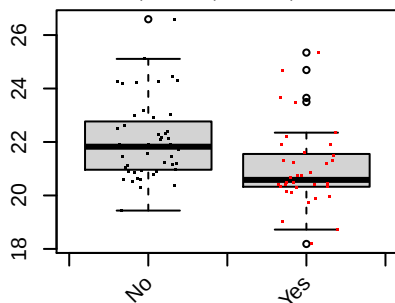
Cytosol aminopeptidase

LAP3, P= 0.0245, M= -1.4, N= 83



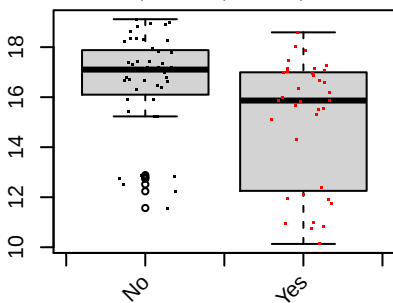
Lactotransferrin

LTF, P= 0.025, M= -1.04, N= 83



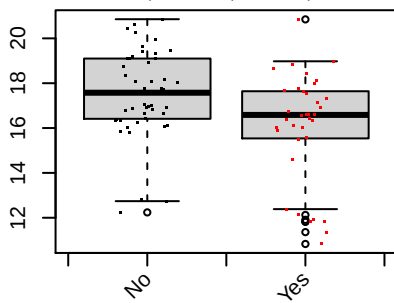
Myosin light polypeptide 6

MYL6, P= 0.0254, M= -1.51, N= 83



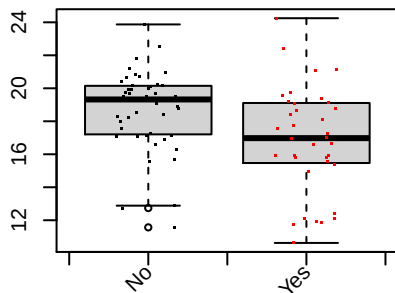
Annexin

ANXA5, P= 0.0256, M= -1.47, N= 83



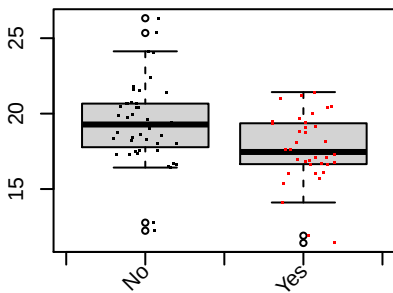
Zymogen granule protein 16 homolog B

ZG16B, P= 0.0262, M= -1.86, N= 83



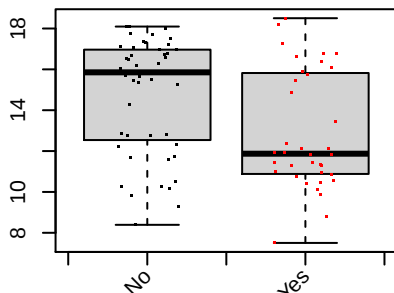
Cathepsin D

CTSD, P= 0.0267, M= -1.68, N= 83

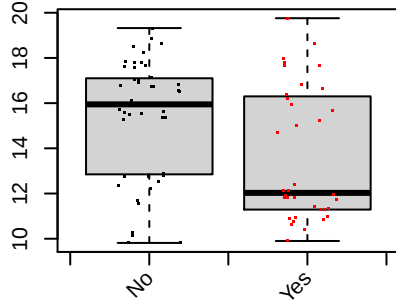


Small ubiquitin-related modifier 3

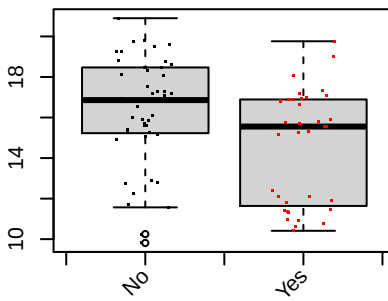
SUMO3, P= 0.0267, M= -1.87, N= 83



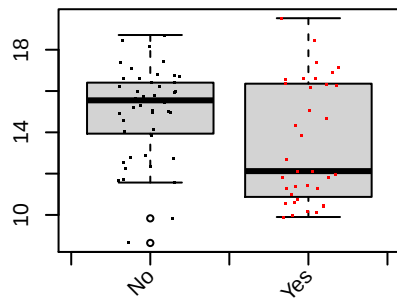
Actin-related protein 2/3 complex subunit 5
ARPC5, P= 0.0302, M= -1.78, N= 83



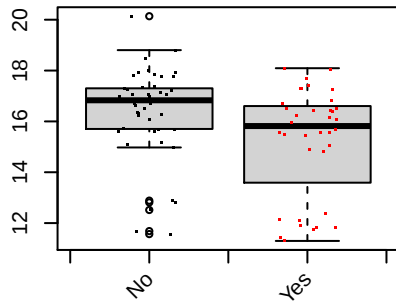
Synaptic vesicle membrane protein VAT-1 homolog
VAT1, P= 0.0349, M= -1.7, N= 83



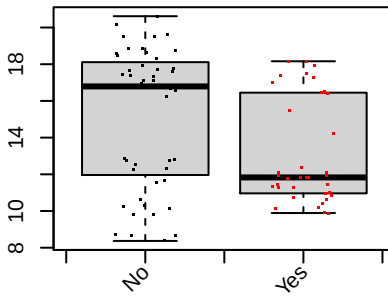
Erythrocyte band 7 integral membrane protein
STOM, P= 0.0349, M= -1.53, N= 83



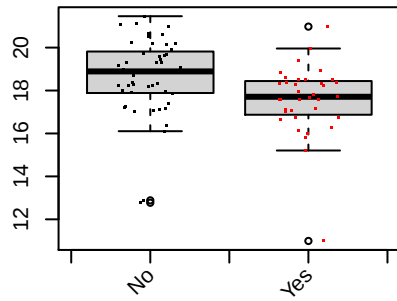
Ras-related C3 botulinum toxin substrate 2
RAC2, P= 0.0349, M= -1.19, N= 83



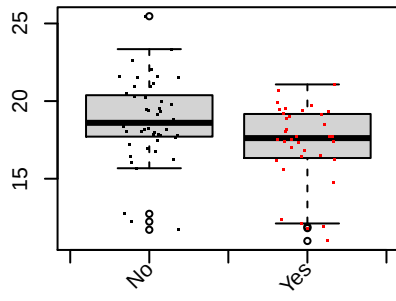
Macrophage-capping protein
CAPG, P= 0.0349, M= -2.04, N= 83



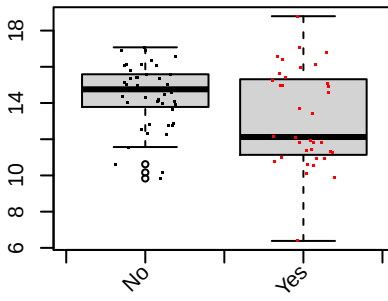
Leukocyte elastase inhibitor
SERPINB1, P= 0.0362, M= -1.07, N= 83



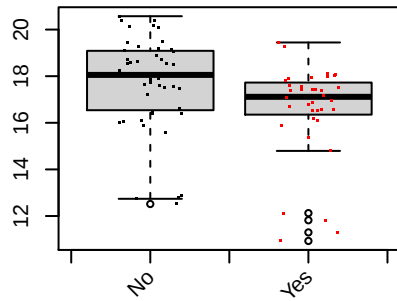
Aldehyde dehydrogenase, dimeric NADP-preferring
ALDH3A1, P= 0.0372, M= -1.57, N= 83



Ras-related protein Rab-5C
RAB5C, P= 0.0427, M= -1.28, N= 83



Annexin
ANXA3, P= 0.0479, M= -1.15, N= 83



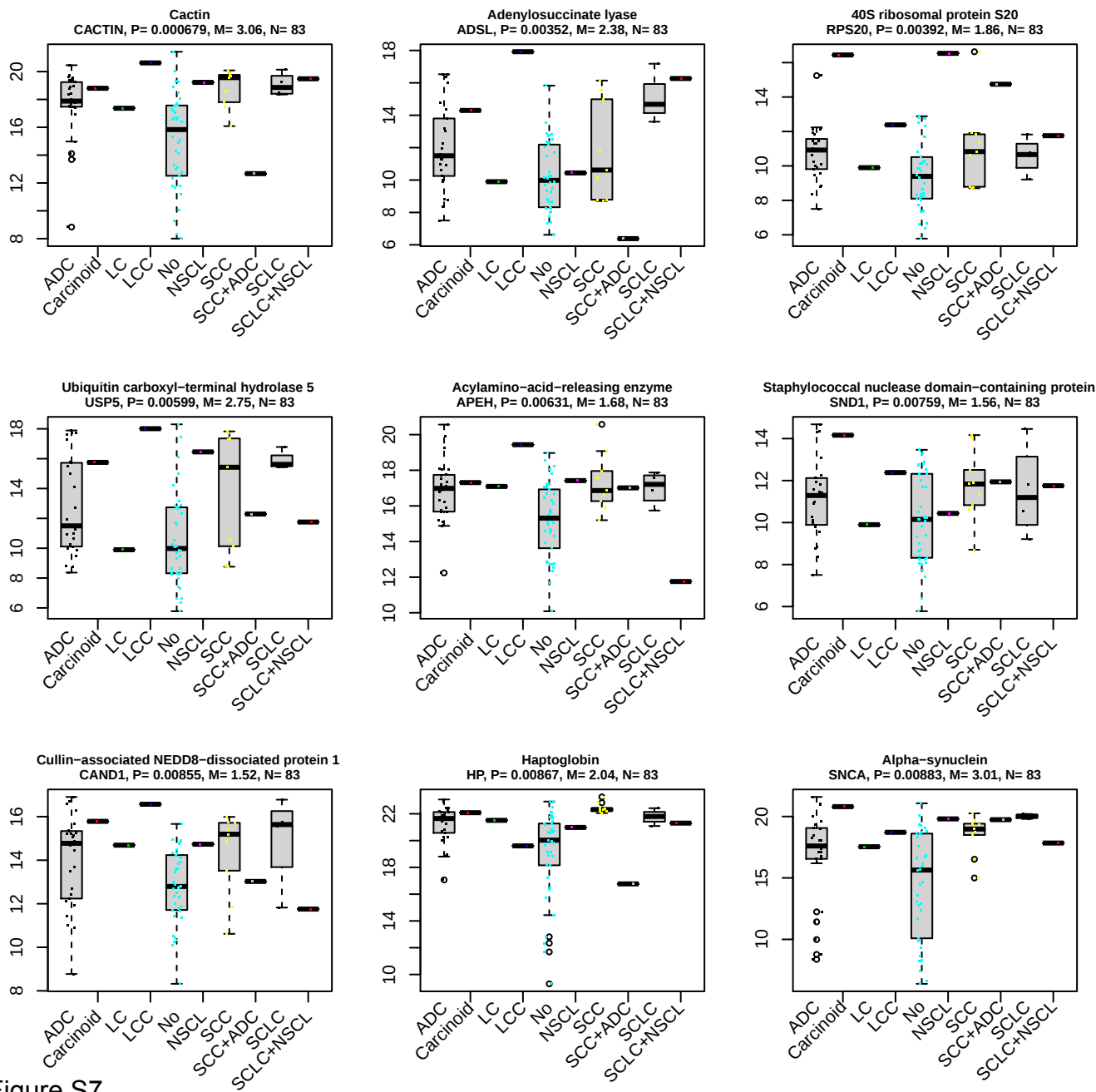
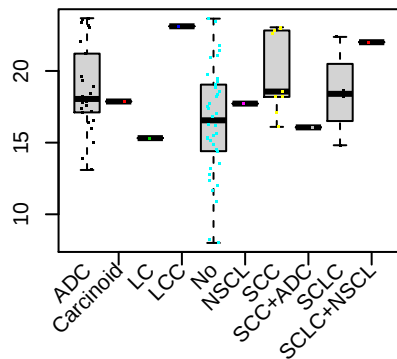
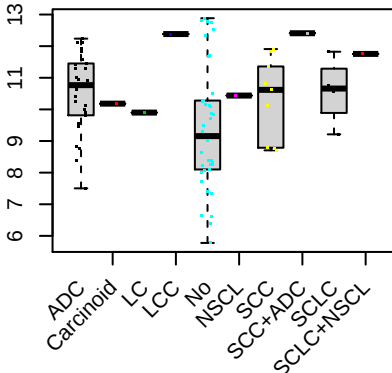


Figure S7

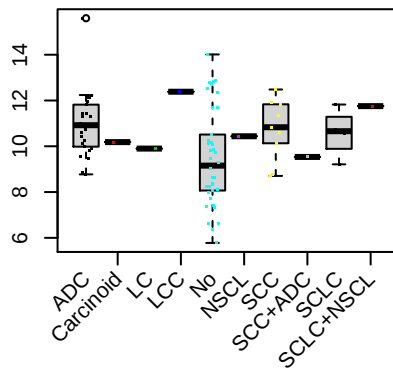
C4b-binding protein beta chain
C4BPB, P= 0.0168, M= 2.43, N= 83



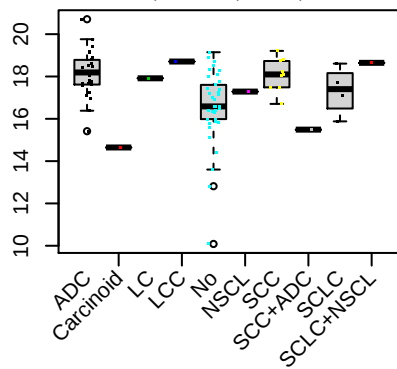
DNA-directed RNA polymerases I, II, and III subunit RPA
POLR2H, P= 0.0387, M= 1.19, N= 83



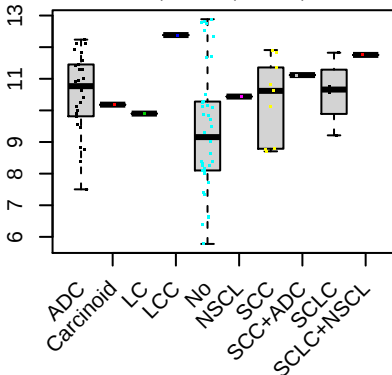
Protein transport protein Sec23A
SEC23A, P= 0.0205, M= 1.26, N= 83



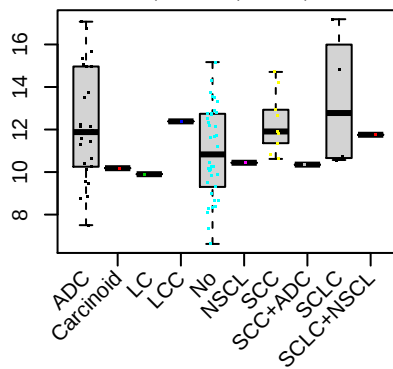
Protein AMBP
AMBP, P= 0.0209, M= 1.1, N= 83



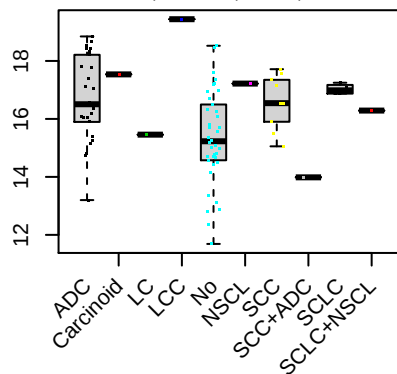
Ethylmalonyl-CoA decarboxylase
ECHDC1, P= 0.021, M= 1.15, N= 83



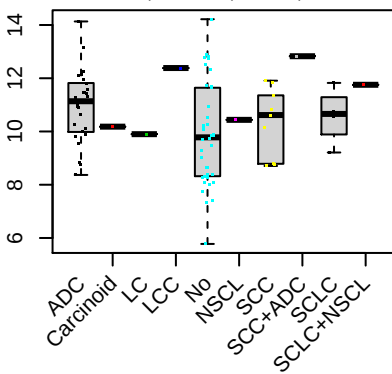
Acyl-coenzyme A thioesterase 2, mitochondrial
ACOT2, P= 0.0232, M= 1.52, N= 83



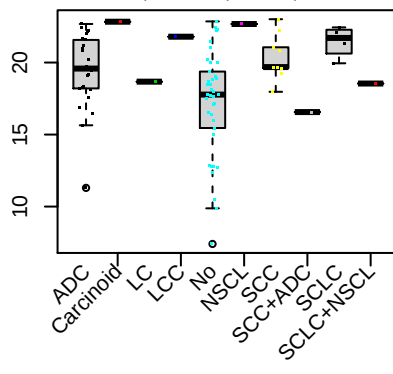
Ankyrin-1
ANK1, P= 0.0232, M= 1.1, N= 83



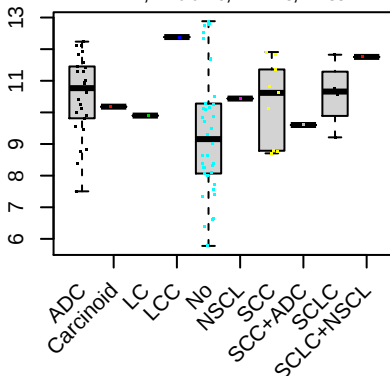
3-hydroxyisobutyryl-CoA hydrolase, mitochondrial
HIBCH, P= 0.0239, M= 1.13, N= 83



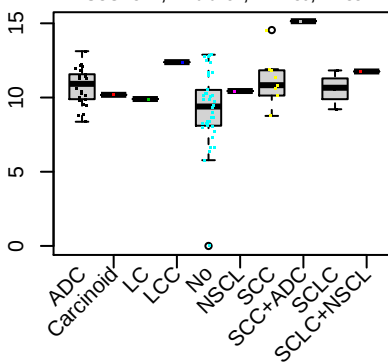
Carbonic anhydrase 2
CA2, P= 0.0243, M= 2.12, N= 83



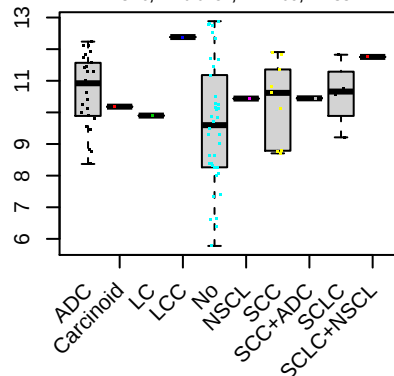
Farnesyltransferase, CAAX box, alpha, isoform CRA_1
FNTA, P= 0.0249, M= 1.13, N= 83



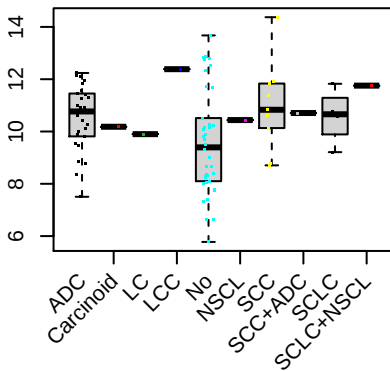
Secretoglobin family 3A member 2
SCGB3A2, P= 0.0254, M= 1.39, N= 83



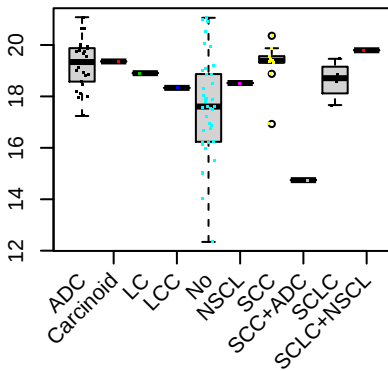
Alpha-FUT6
FUT6, P= 0.0254, M= 1.09, N= 83



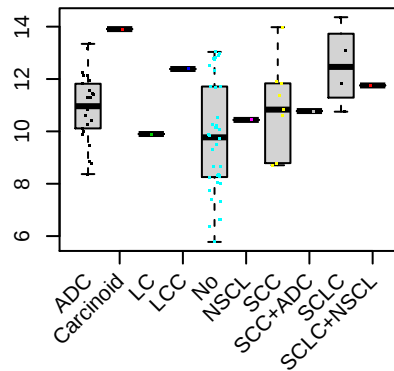
Platelet glycoprotein Ib alpha chain
GP1BA, P= 0.0254, M= 1.15, N= 83



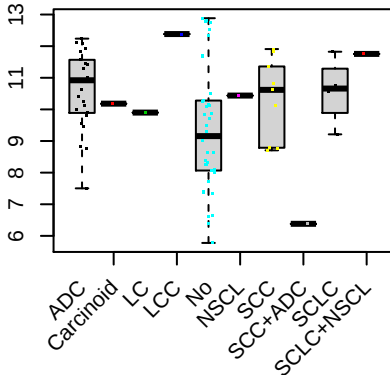
Pregnancy zone protein
PZP, P= 0.0256, M= 1.12, N= 83



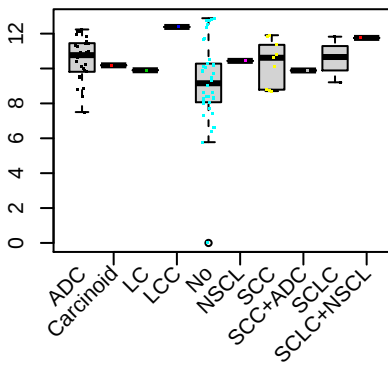
26S proteasome non-ATPase regulatory subunit 12
PSMD12, P= 0.0262, M= 1.25, N= 83



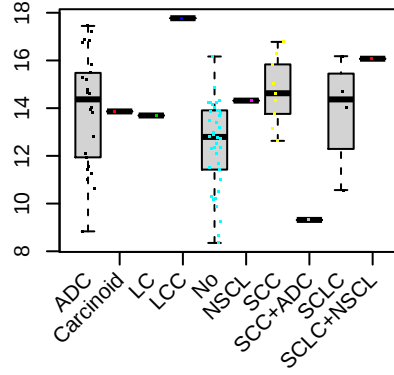
Protein THEMIS2
THEMIS2, P= 0.0264, M= 1.14, N= 83



Uncharacterized protein C1orf87
C1orf87, P= 0.0265, M= 1.29, N= 83

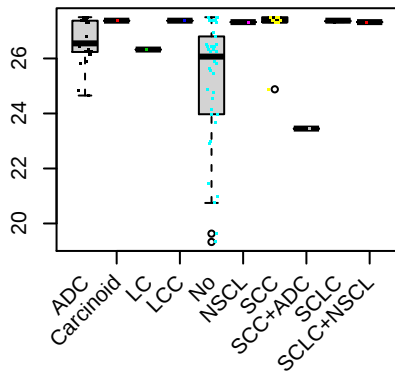


Calcium-activated chloride channel regulator 1
CLCA1, P= 0.0265, M= 1.4, N= 83



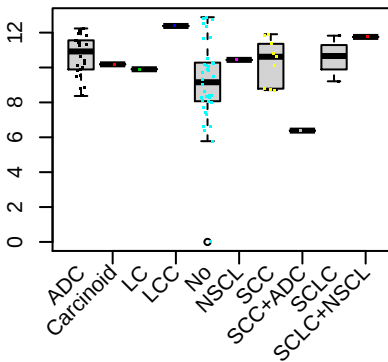
Hemoglobin subunit delta

HBD, $P=0.0265$, $M=1.21$, $N=83$



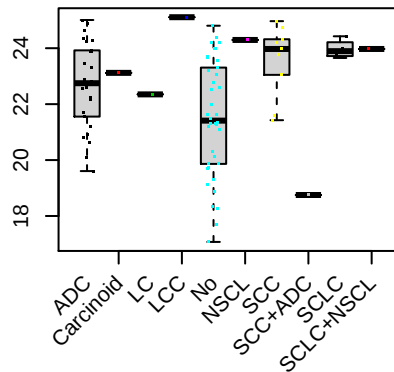
Triokinase/FMN cyclase

TKFC, $P=0.0265$, $M=1.31$, $N=83$



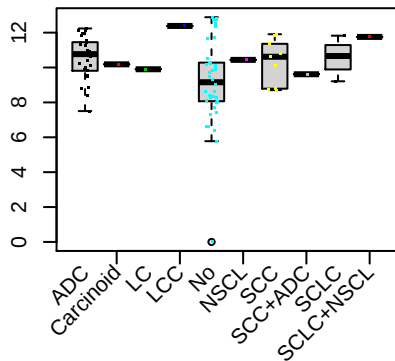
Carbonic anhydrase 1

CA1, $P=0.0267$, $M=1.27$, $N=83$



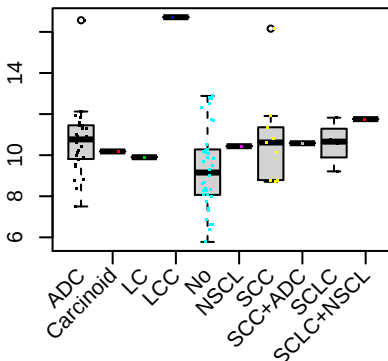
UPF0764 protein C16orf89

C16orf89, $P=0.0267$, $M=1.29$, $N=83$



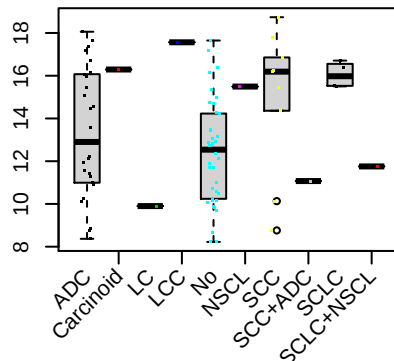
26S proteasome non-ATPase regulatory subunit 1

PSMD1, $P=0.0284$, $M=1.32$, $N=83$



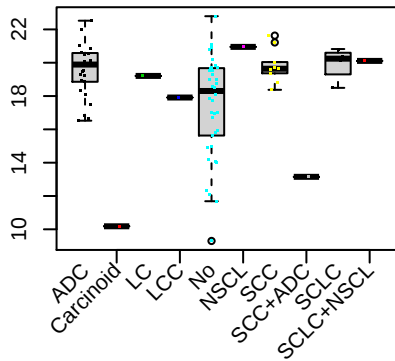
Protein DDI1 homolog 2

DDI2, $P=0.0294$, $M=1.79$, $N=83$



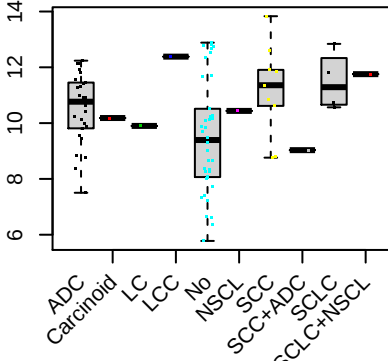
Platelet basic protein

PPBP, $P=0.0295$, $M=1.7$, $N=83$



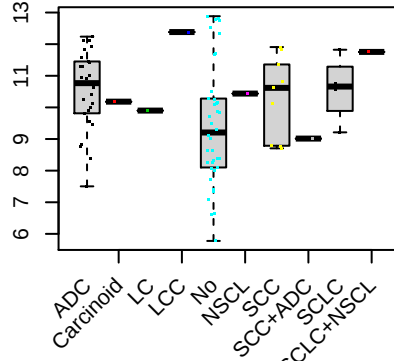
Eukaryotic translation initiation factor 5

EIF5, $P=0.0304$, $M=1.15$, $N=83$

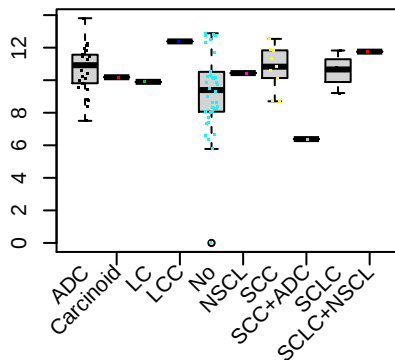


Tyrosine-protein kinase BTK

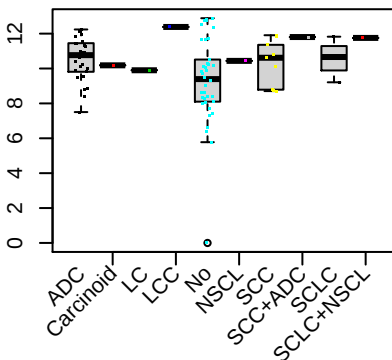
BTK, $P=0.0304$, $M=1.06$, $N=83$



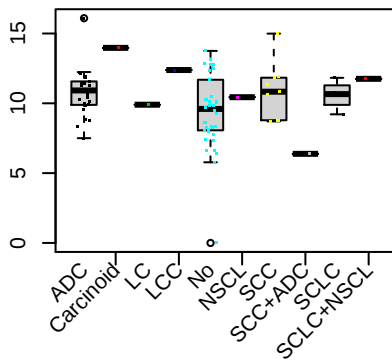
Mannan-binding lectin serine protease 2
MASP2, P= 0.0314, M= 1.32, N= 83



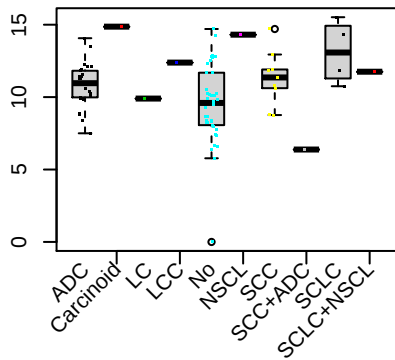
3-ketoacyl-CoA thiolase, peroxisomal
ACAA1, P= 0.0317, M= 1.24, N= 83



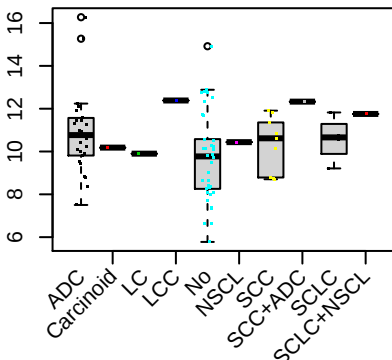
Ketosanine-3-kinase
FN3KRP, P= 0.0325, M= 1.42, N= 83



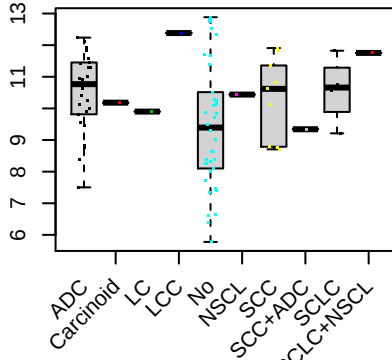
Bifunctional purine biosynthesis protein PURH
ATIC, P= 0.0341, M= 1.52, N= 83



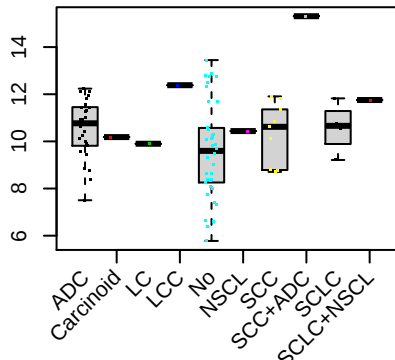
UPF0553 protein C9orf64
C9orf64, P= 0.0344, M= 1.18, N= 83



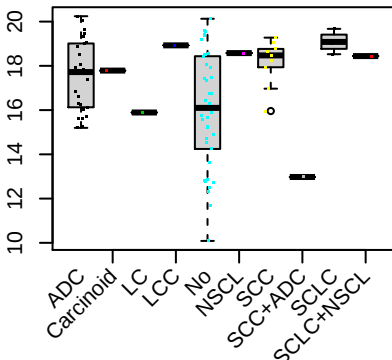
ATP-dependent RNA helicase A
DHX9, P= 0.0345, M= 1.05, N= 83



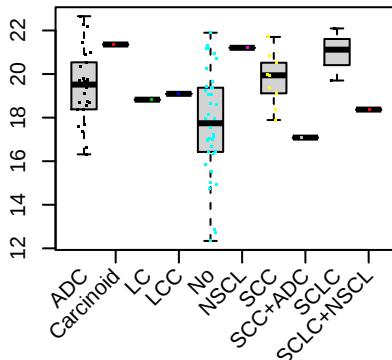
Small nuclear ribonucleoprotein Sm D1
SNRPD1, P= 0.0348, M= 1.11, N= 83



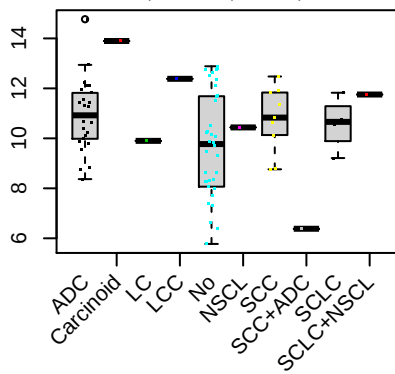
Delta-aminolevulinic acid dehydratase
ALAD, P= 0.0349, M= 1.38, N= 83



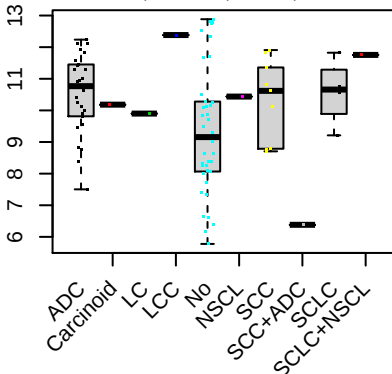
Flavin reductase
BLVRB, P= 0.0349, M= 1.33, N= 83



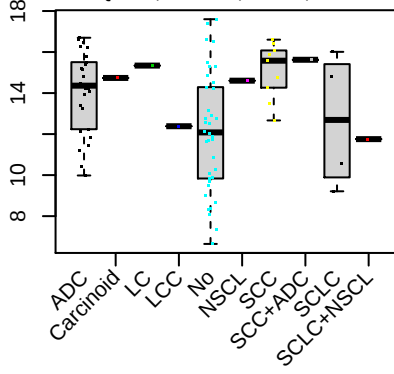
Exportin-7

XPO7, $P = 0.0349$, $M = 1.16$, $N = 83$ 

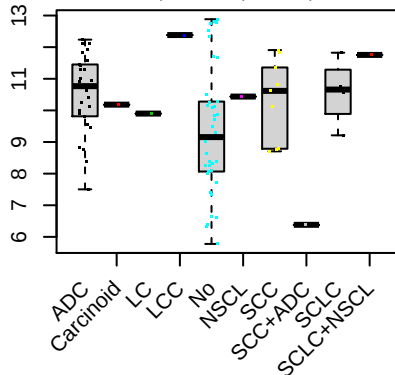
SUMO-activating enzyme subunit 1

SAE1, $P = 0.0349$, $M = 1.07$, $N = 83$ 

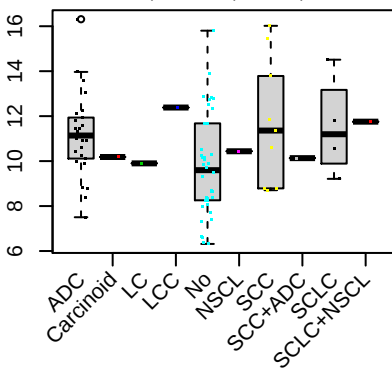
Ras GTPase-activating-like protein IQGAP2

IQGAP2, $P = 0.0349$, $M = 1.55$, $N = 83$ 

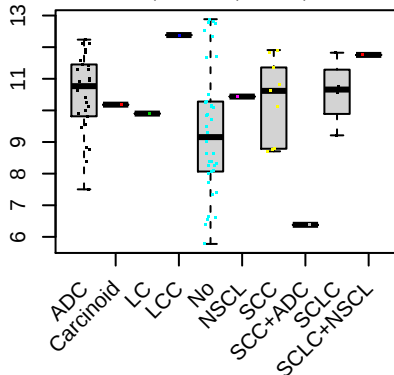
Nuclear protein localization protein 4 homolog

NPLOC4, $P = 0.0349$, $M = 1.06$, $N = 83$ 

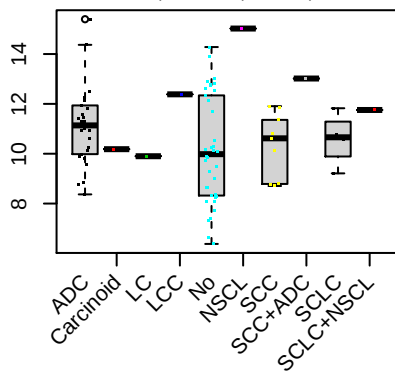
26S proteasome non-ATPase regulatory subunit 5

PSMD5, $P = 0.0349$, $M = 1.33$, $N = 83$ 

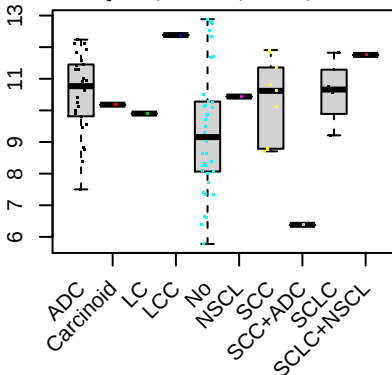
Dynamin-1-like protein

DNM1L, $P = 0.0349$, $M = 1.06$, $N = 83$ 

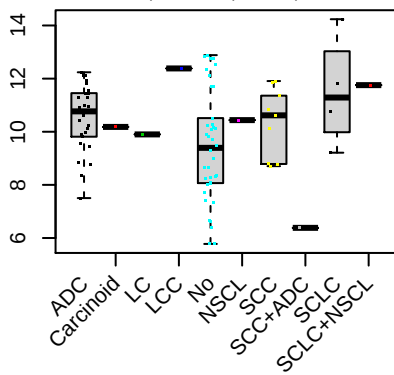
Serine/threonine-protein phosphatase

PPP1CC, $P = 0.0349$, $M = 1.13$, $N = 83$ 

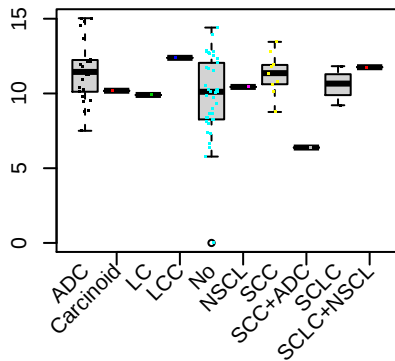
Cytochrome b-c1 complex subunit 2, mitochondrial

UQCRC2, $P = 0.0349$, $M = 1.04$, $N = 83$ 

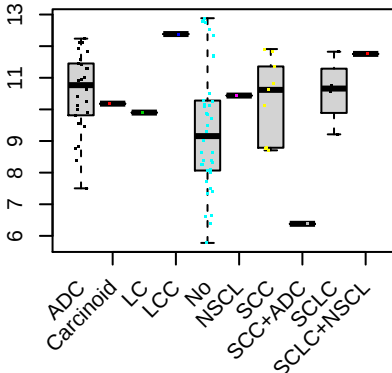
Exportin-1

XPO1, $P = 0.0349$, $M = 1.1$, $N = 83$ 

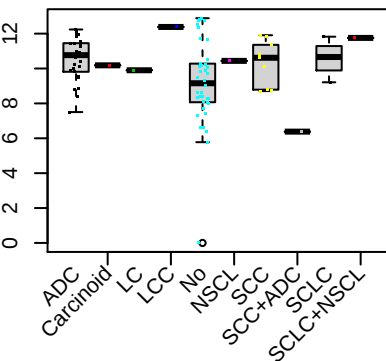
Phosphatidylcholine-sterol acyltransferase
LCAT, P= 0.0349, M= 1.39, N= 83



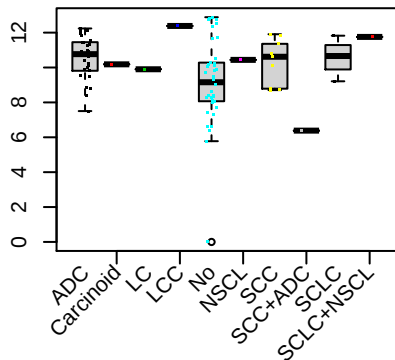
Dual specificity mitogen-activated protein kinase
MAP2K1, P= 0.0349, M= 1.04, N= 83



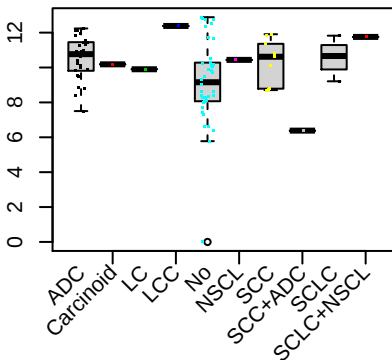
Ubiquitin-like modifier-activating enzyme ATG7
ATG7, P= 0.0349, M= 1.2, N= 83



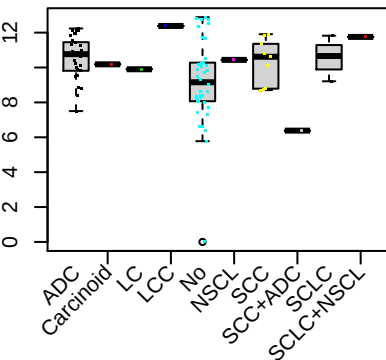
Isovaleryl-CoA dehydrogenase, mitochondrial
IVD, P= 0.0349, M= 1.2, N= 83



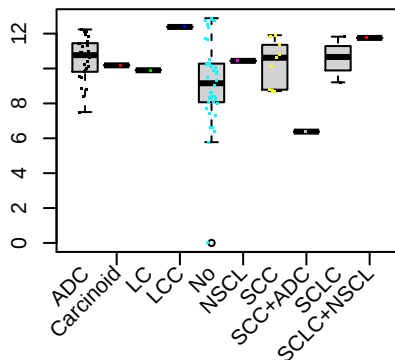
Isoleucine--tRNA ligase, cytoplasmic
IARS, P= 0.0349, M= 1.2, N= 83



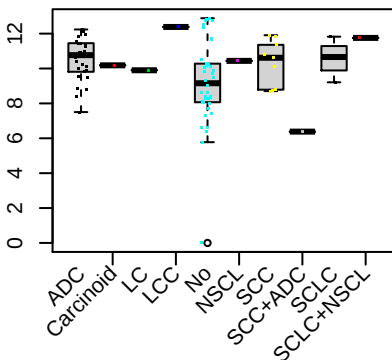
Leucine-protein phosphatase 2A 55 kDa regulatory subunit
PPP2R2A, P= 0.0349, M= 1.2, N= 83



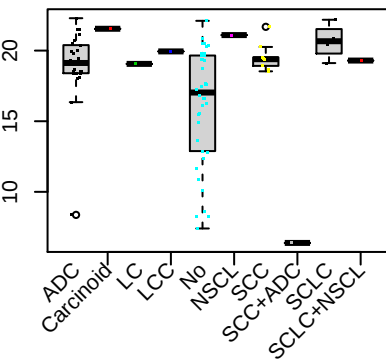
Splicing factor 3B subunit 2
SF3B2, P= 0.0349, M= 1.2, N= 83



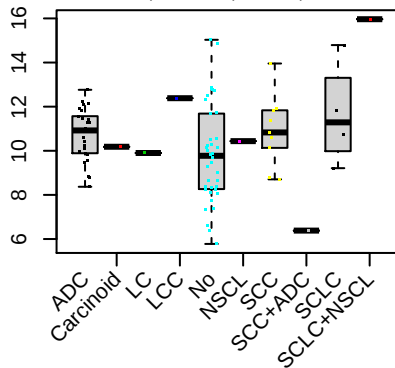
Dynein light chain 1, axonemal
DNAL1, P= 0.0349, M= 1.2, N= 83



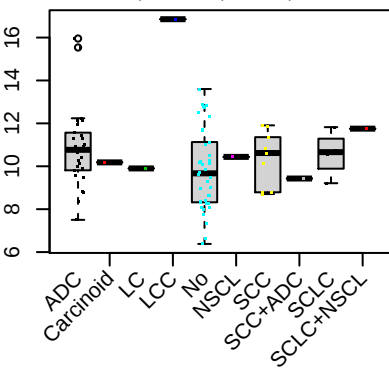
Ly6/PLAUR domain-containing protein 2
LYPD2, P= 0.0352, M= 2.15, N= 83



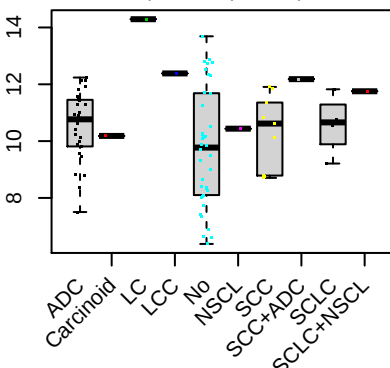
N-sulphoglucosamine sulphohydrolase
SGSH, P= 0.0368, M= 1.17, N= 83



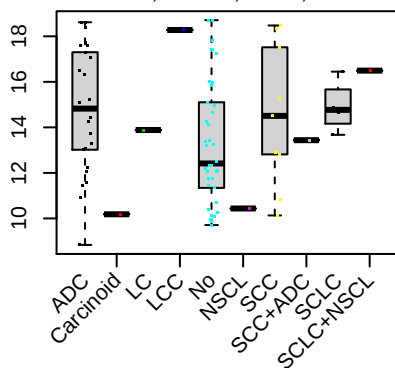
Sialate O-acetyltransferase
SIAE, P= 0.0375, M= 1.11, N= 83



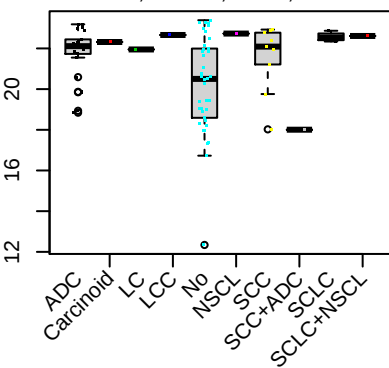
Protein patched homolog 1
PTCH1, P= 0.0382, M= 1.02, N= 83



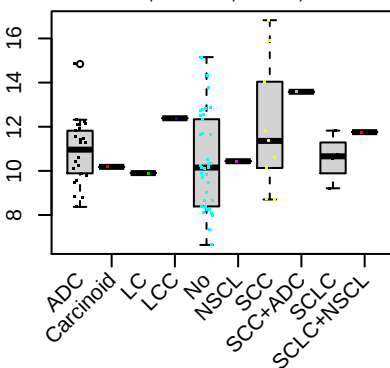
lutamine--fructose-6-phosphate aminotransferase [isome
GFPT1, P= 0.0415, M= 1.54, N= 83



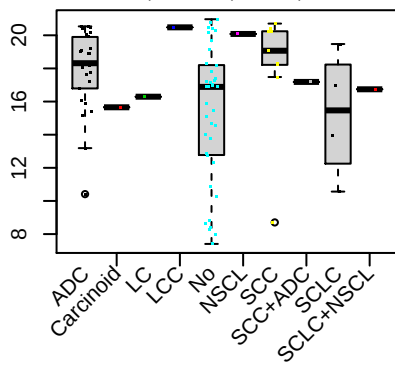
ITI4 protein
ITI4, P= 0.0417, M= 1.15, N= 83



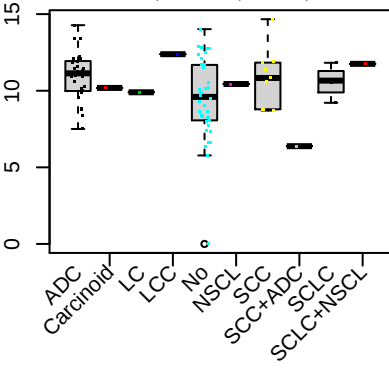
26S protease regulatory subunit 10B
PSMC6, P= 0.0438, M= 1.11, N= 83



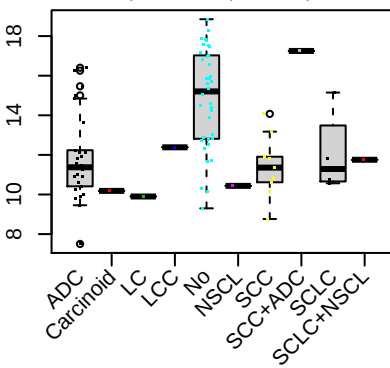
Serum amyloid A-2 protein
SAA2, P= 0.0461, M= 1.98, N= 83



Protein FAM111B
FAM111B, P= 0.0487, M= 1.23, N= 83

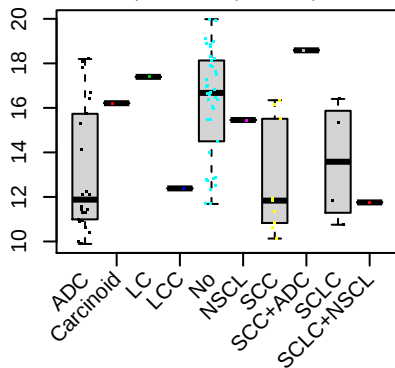


Epididymal secretory protein E1
NPC2, P= 0.000528, M= -2.77, N= 83



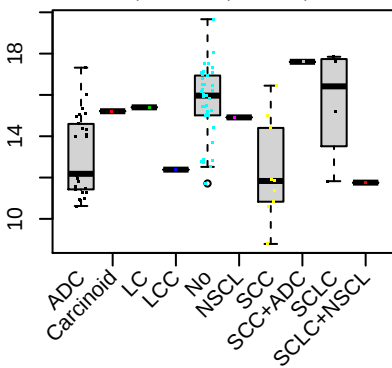
Dipeptidyl peptidase 1

CTSC, P= 0.000679, M= -2.87, N= 83



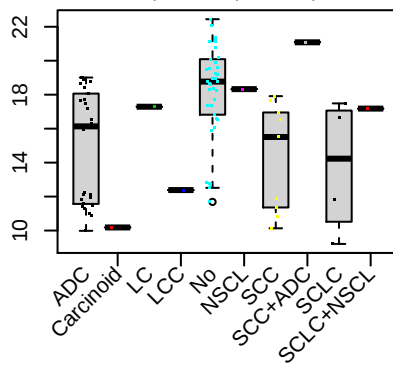
Protein-glutamine gamma-glutamyltransferase 2

TGM2, P= 0.00138, M= -2.17, N= 83



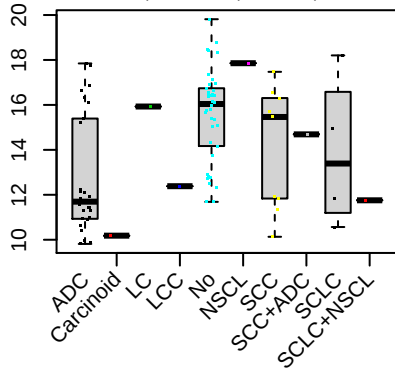
Pulmonary surfactant-associated protein A2

SFTPA2, P= 0.0014, M= -3.21, N= 83



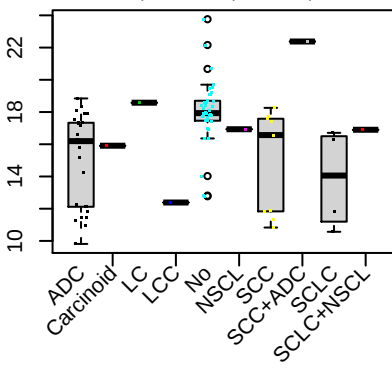
UMP-CMP kinase

CMPK1, P= 0.00177, M= -2.43, N= 83



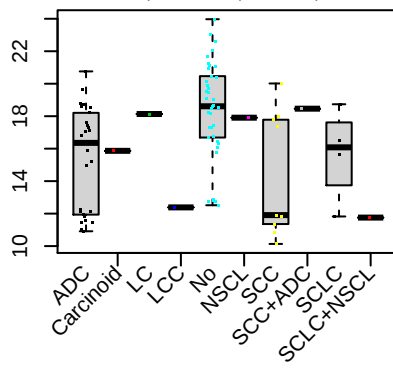
Annexin

ANXA2, P= 0.00177, M= -2.66, N= 83



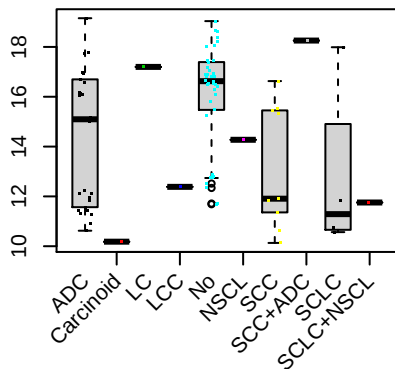
WAP four-disulfide core domain protein 2

WFDC2, P= 0.00177, M= -3.11, N= 83



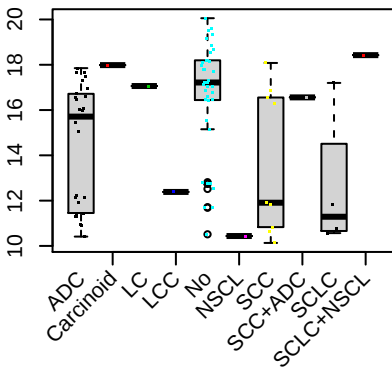
Isocitrate dehydrogenase [NADP] cytoplasmic

IDH1, P= 0.00277, M= -2.34, N= 83



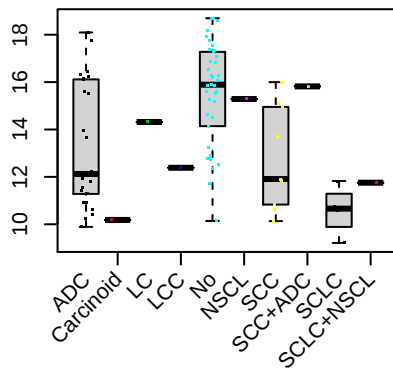
Rho GDP-dissociation inhibitor 1

ARHGDI1, P= 0.00277, M= -2.55, N= 83

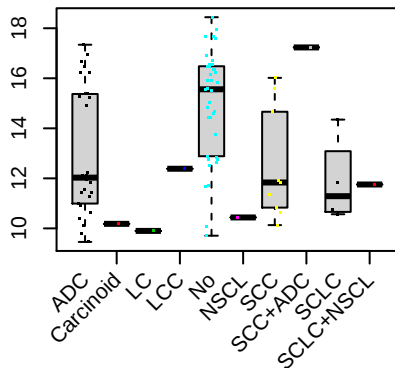


Tryptophan--tRNA ligase, cytoplasmic

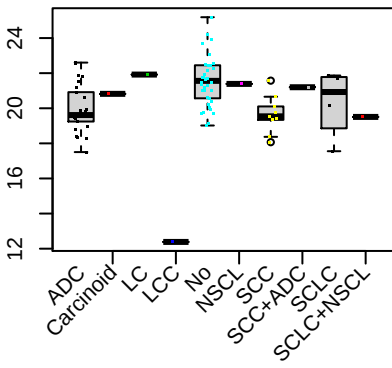
WARS, P= 0.00277, M= -2.29, N= 83



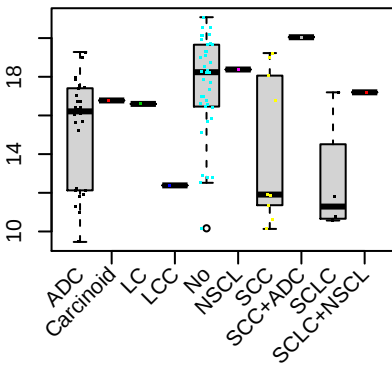
Heterogeneous nuclear ribonucleoprotein K
HNRNPK, P= 0.00288, M= -2.23, N= 83



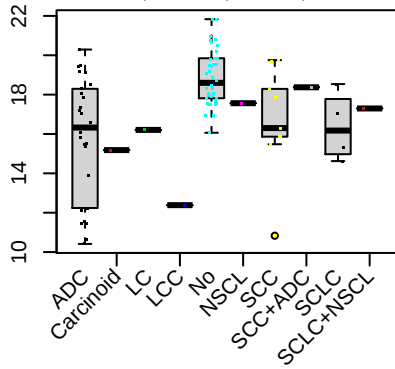
Glutathione S-transferase P
GSTP1, P= 0.00333, M= -1.57, N= 83



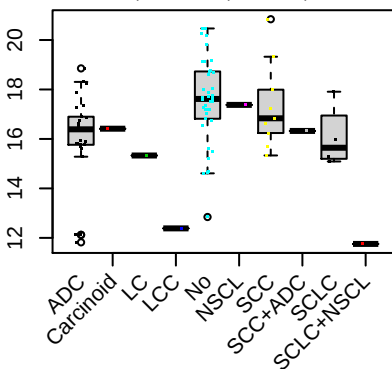
Plastin-2
LCP1, P= 0.00442, M= -2.58, N= 83



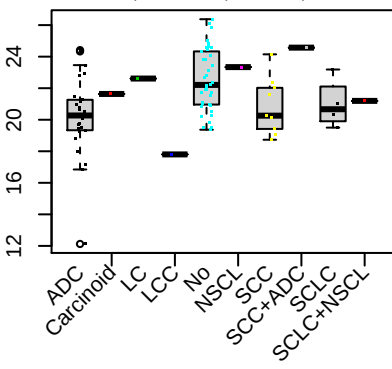
PSAP protein
PSAP, P= 0.0047, M= -2.18, N= 83



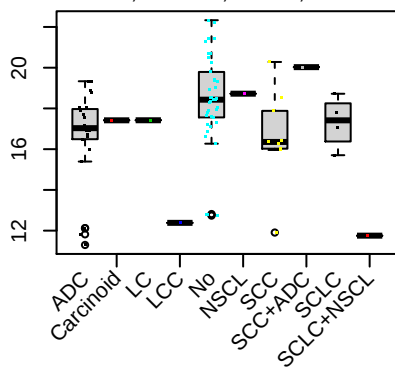
Cathepsin G
CTSG, P= 0.00498, M= -1.61, N= 83



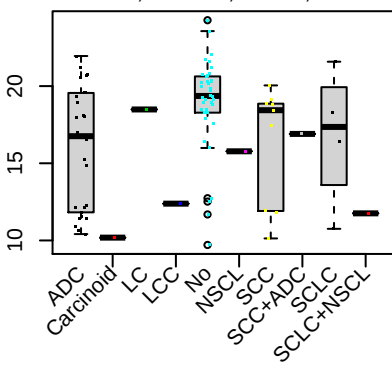
Neutrophil gelatinase-associated lipocalin
LCN2, P= 0.00599, M= -1.88, N= 83



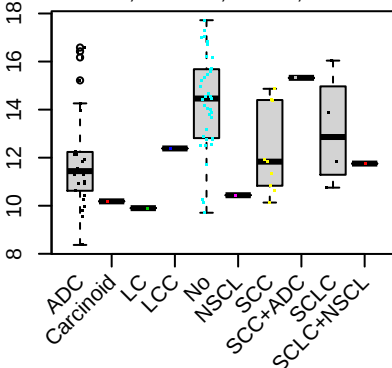
Vimentin
VIM, P= 0.00599, M= -1.96, N= 83



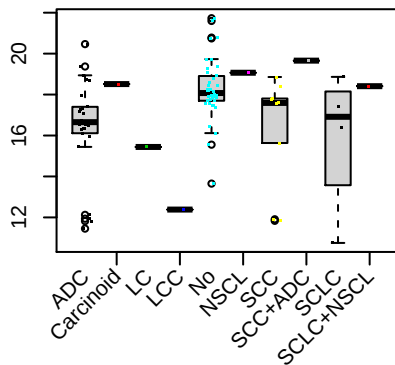
Beta-microseminoprotein
MSMB, P= 0.00651, M= -2.95, N= 83



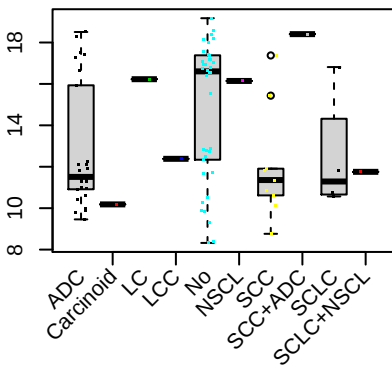
Omega-amidase NIT2
NIT2, P= 0.00804, M= -1.76, N= 83



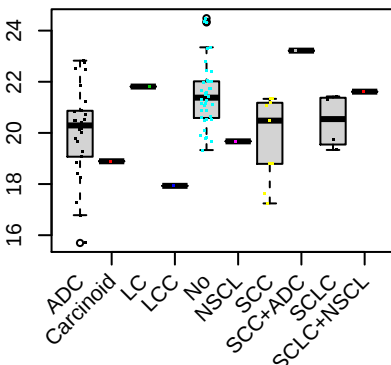
HLA class II histocompatibility antigen, DR alpha chain
HLA-DRA, P= 0.00867, M= -1.73, N= 83



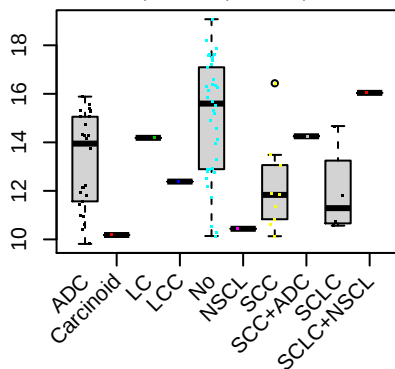
Actin-related protein 3
ACTR3, P= 0.0095, M= -2.41, N= 83



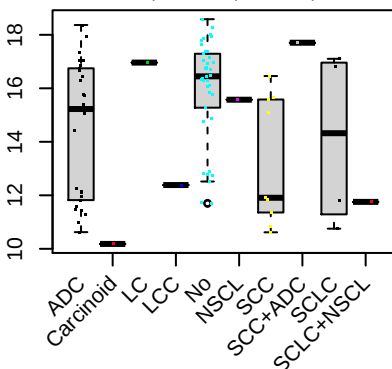
Immunoglobulin J chain
JCHAIN, P= 0.01, M= -1.27, N= 83



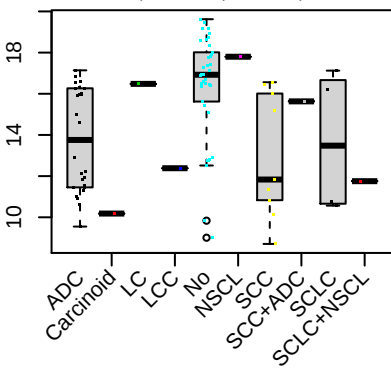
Cathepsin S
CTSS, P= 0.0126, M= -1.74, N= 83



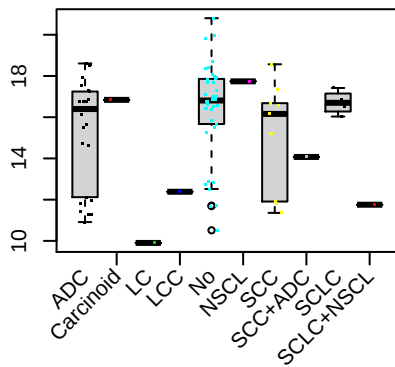
Heterogeneous nuclear ribonucleoprotein D0
HNRNPD, P= 0.0134, M= -1.78, N= 83



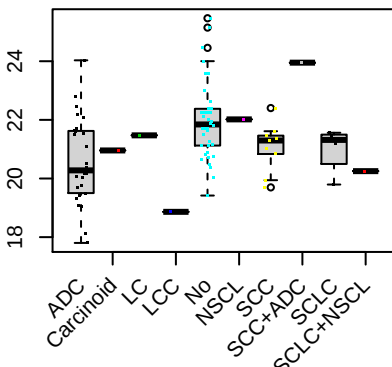
Mucin-1
MUC1, P= 0.0147, M= -2.07, N= 83



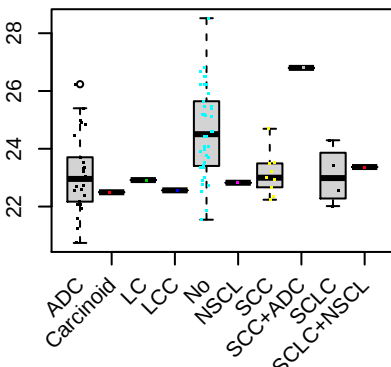
Fructose-1,6-bisphosphatase 1
FBP1, P= 0.0159, M= -1.77, N= 83



BPI fold-containing family B member 1
BPIFB1, P= 0.0192, M= -1.03, N= 83

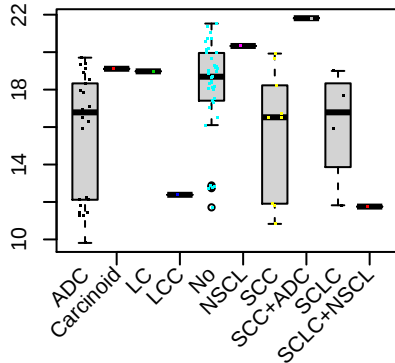


Lysozyme C
LYZ, P= 0.0207, M= -1.07, N= 83



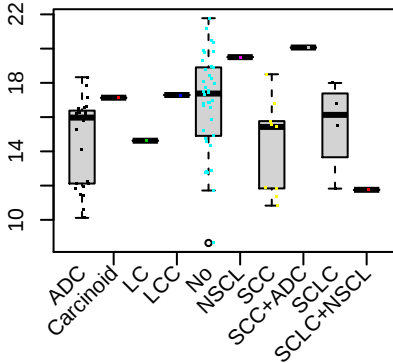
Protein S100-A11

S100A11, P= 0.0207, M= -2.12, N= 83



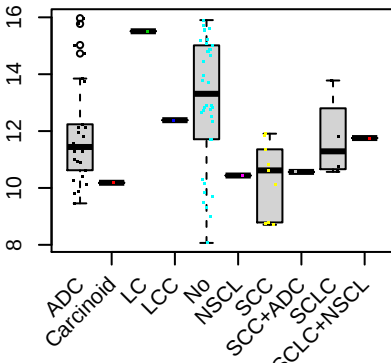
Liver carboxylesterase 1

CES1, P= 0.0207, M= -1.94, N= 83



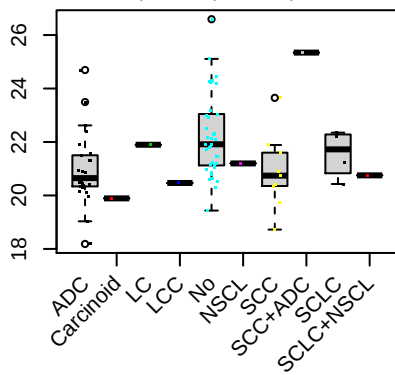
Cytosol aminopeptidase

LAP3, P= 0.0245, M= -1.4, N= 83



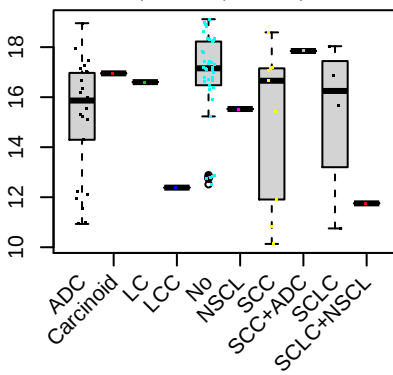
Lactotransferrin

LTF, P= 0.025, M= -1.04, N= 83



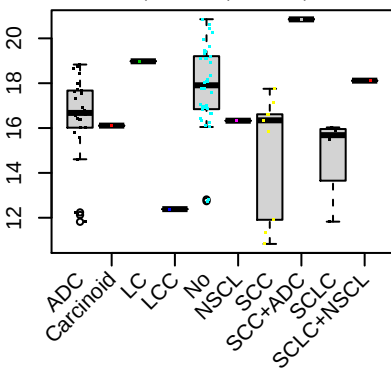
Myosin light polypeptide 6

MYL6, P= 0.0254, M= -1.51, N= 83



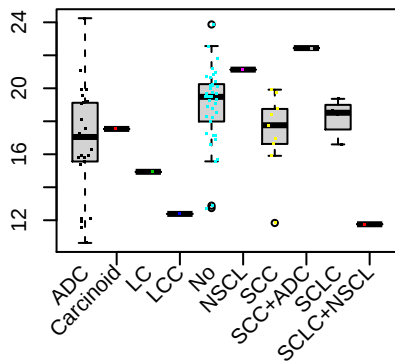
Annexin

ANXA5, P= 0.0256, M= -1.47, N= 83



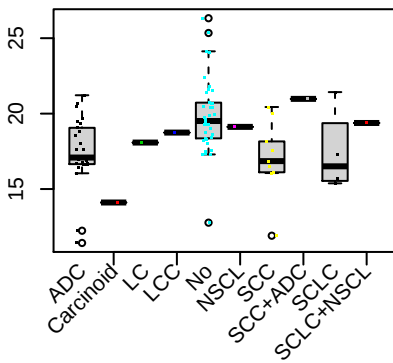
Zymogen granule protein 16 homolog B

ZG16B, P= 0.0262, M= -1.86, N= 83



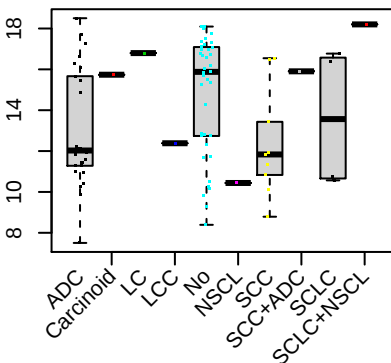
Cathepsin D

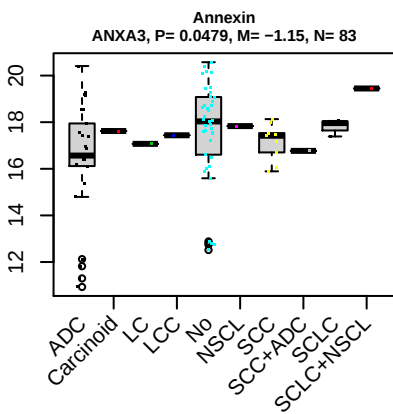
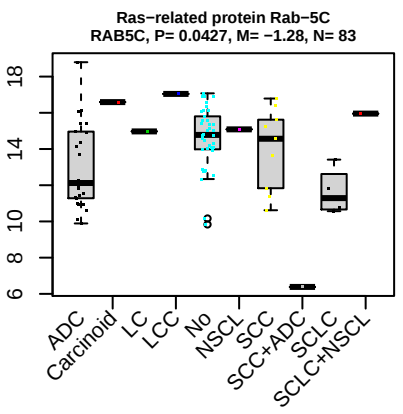
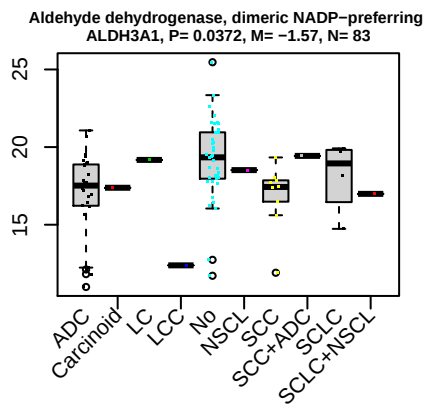
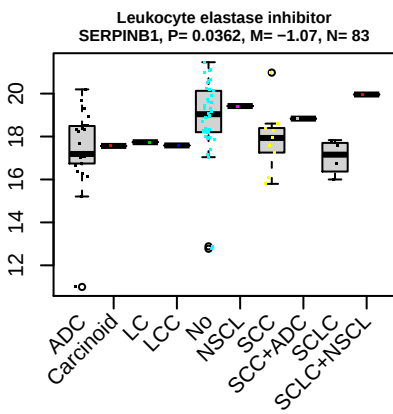
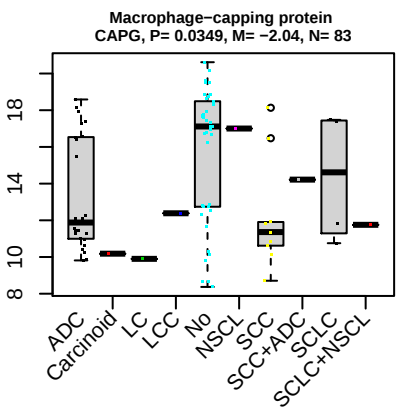
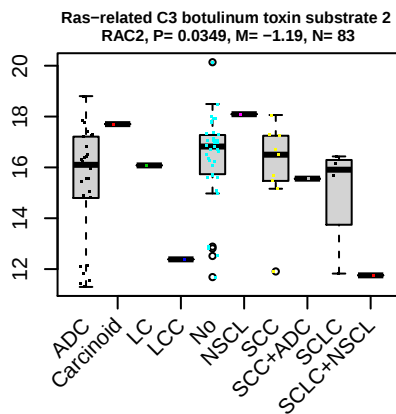
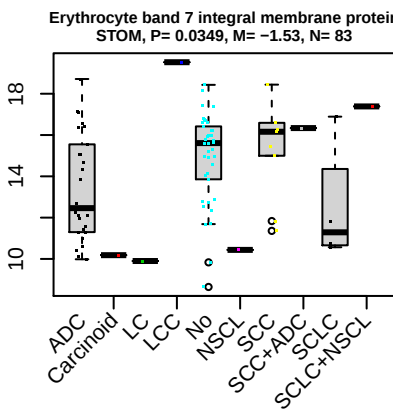
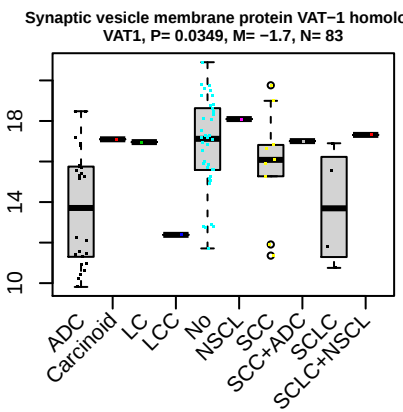
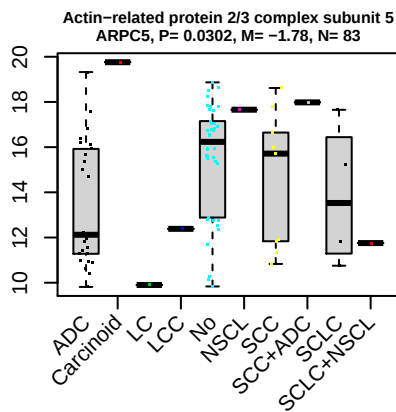
CTSD, P= 0.0267, M= -1.68, N= 83



Small ubiquitin-related modifier 3

SUMO3, P= 0.0267, M= -1.87, N= 83





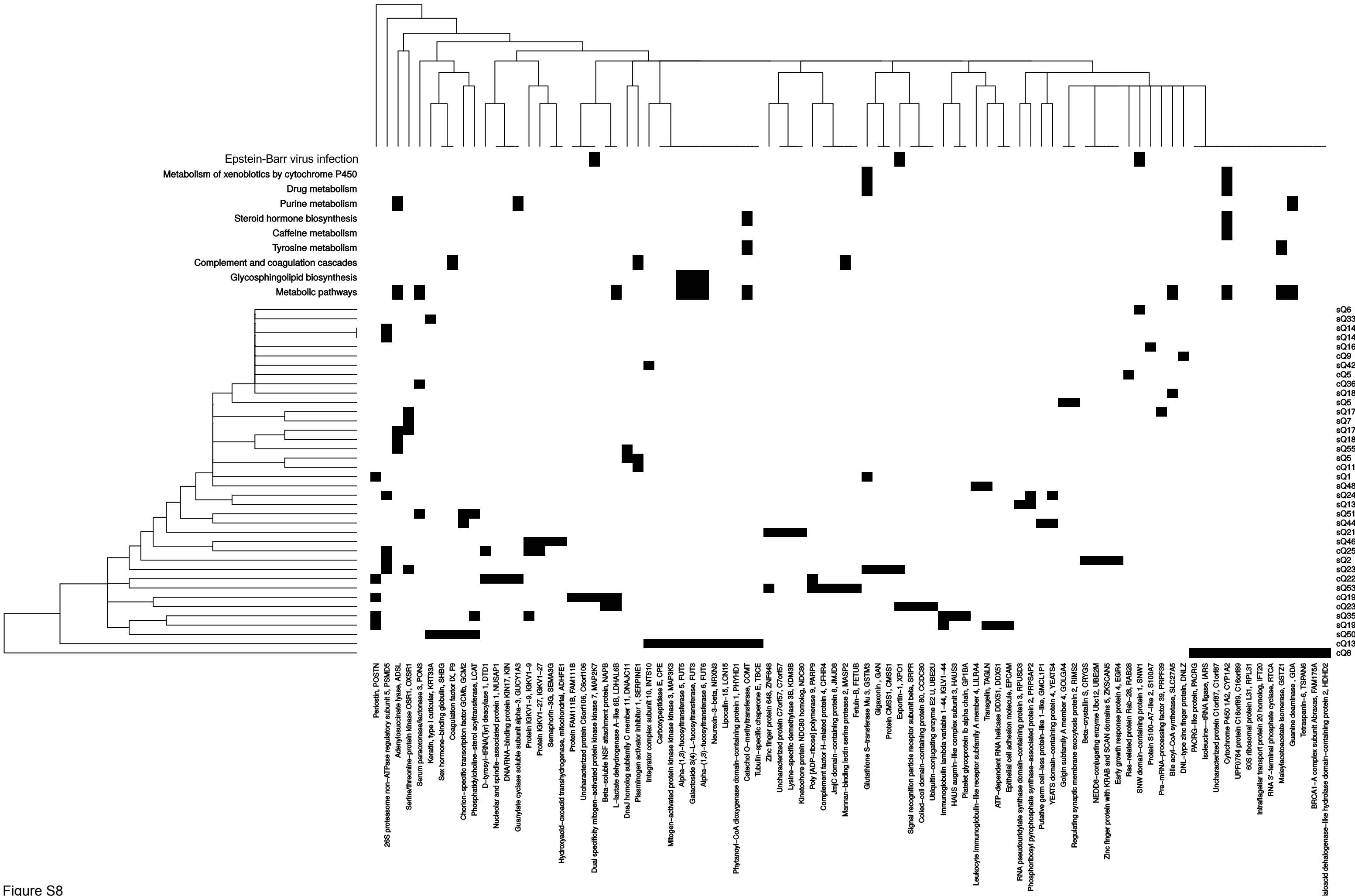
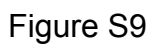


Figure S8



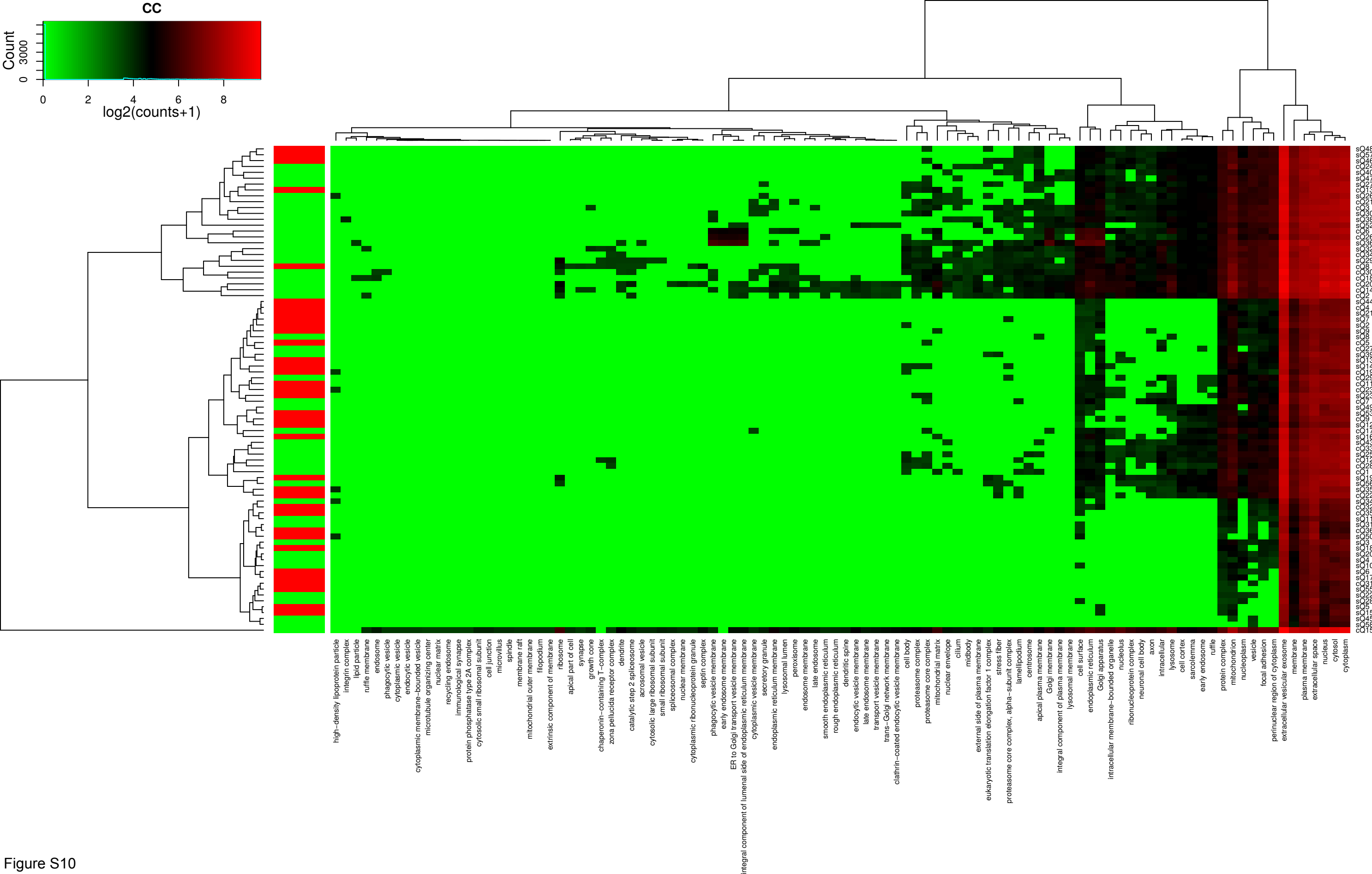


Figure S10

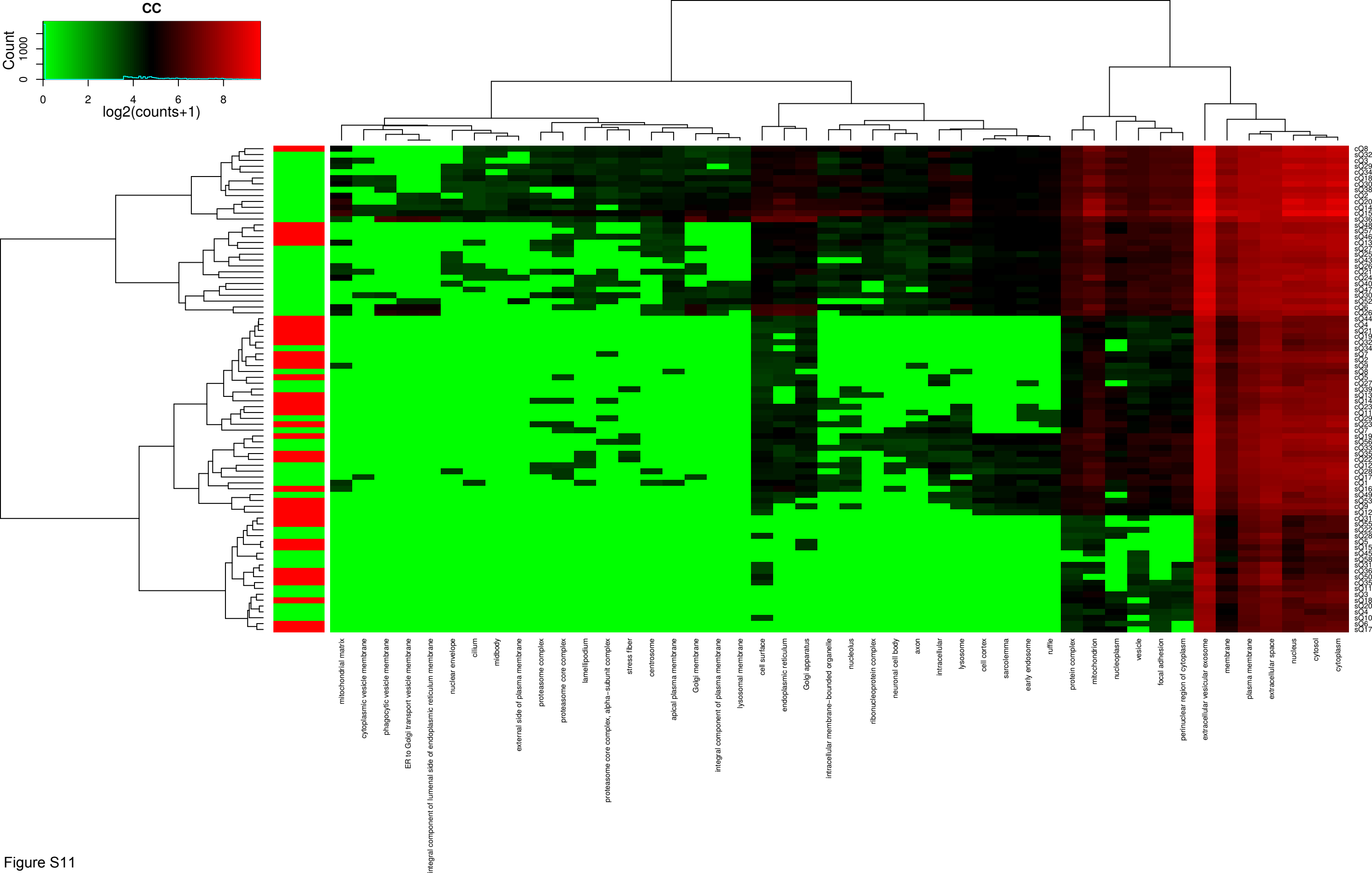


Figure S11

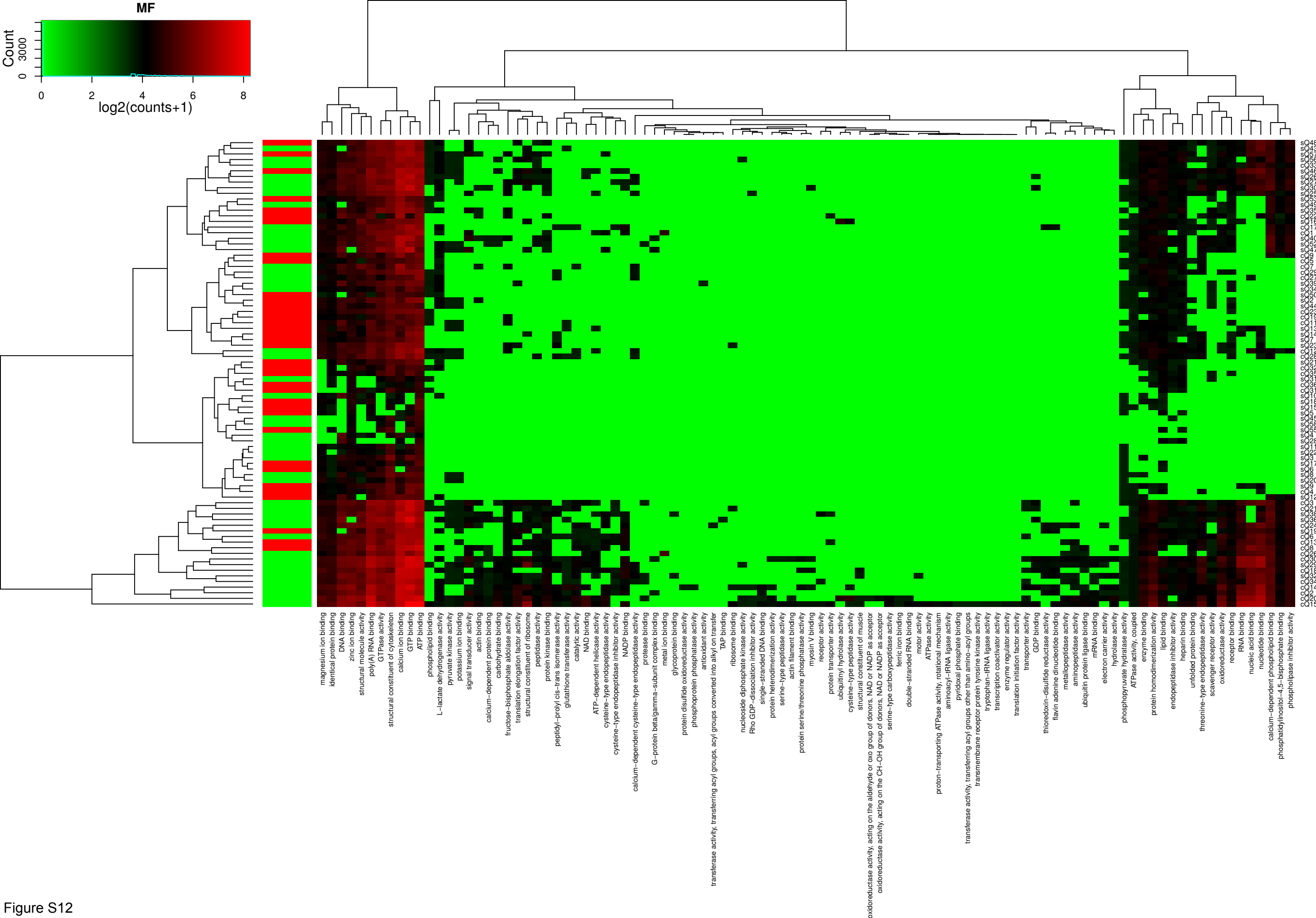


Figure S12

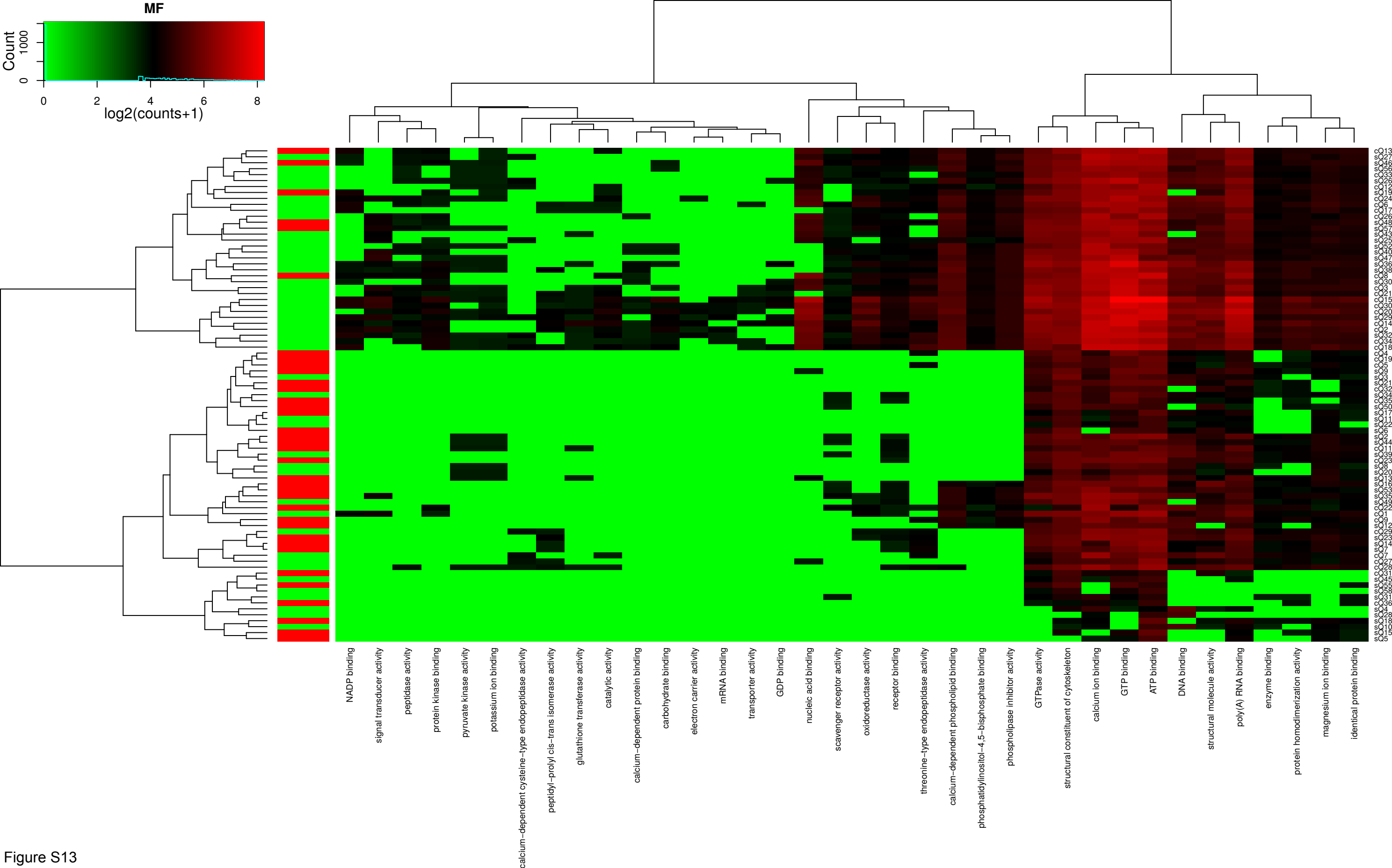
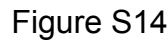
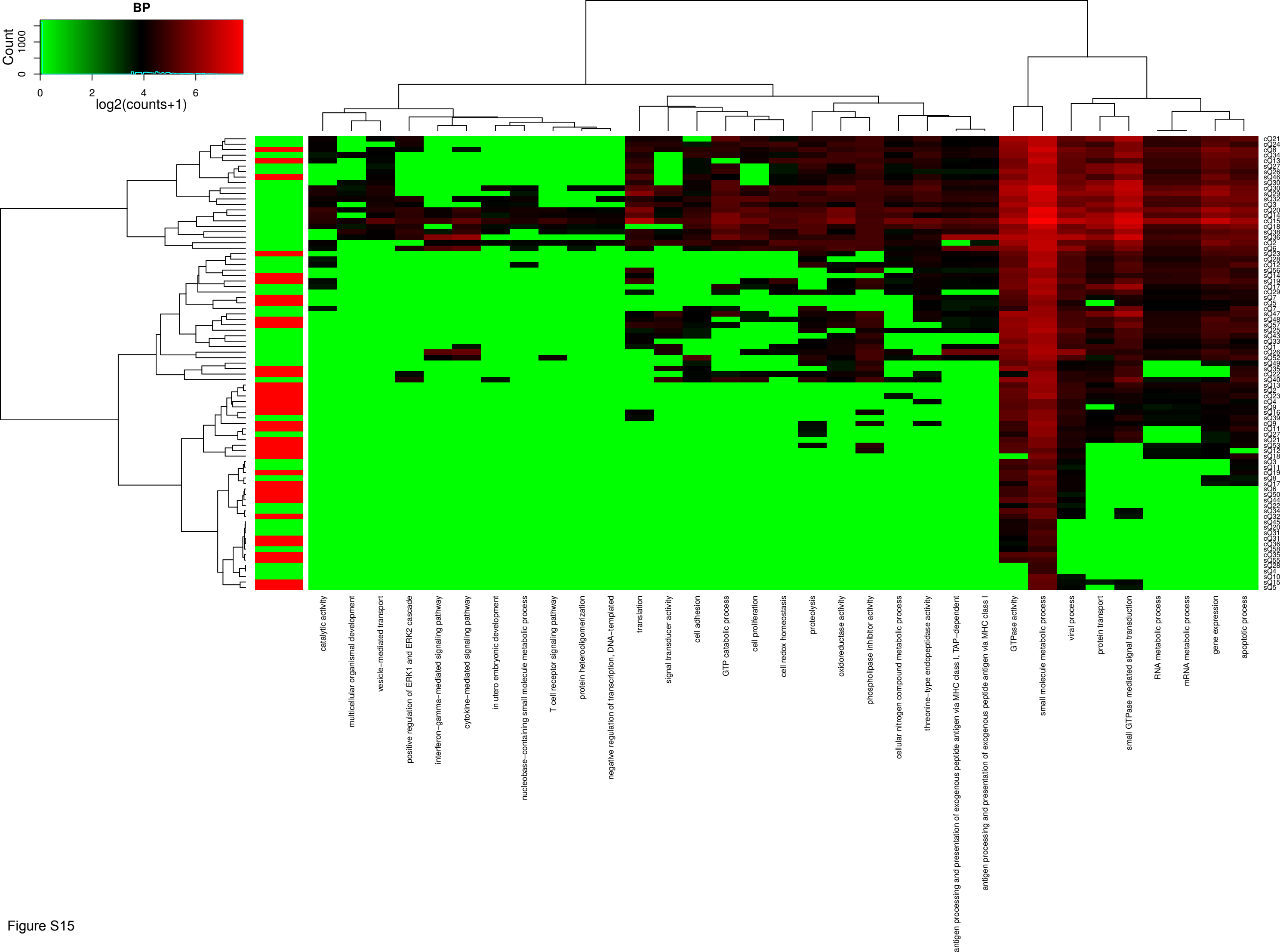


Figure S13





Spectral counts for “extracellular vesicular exosome”

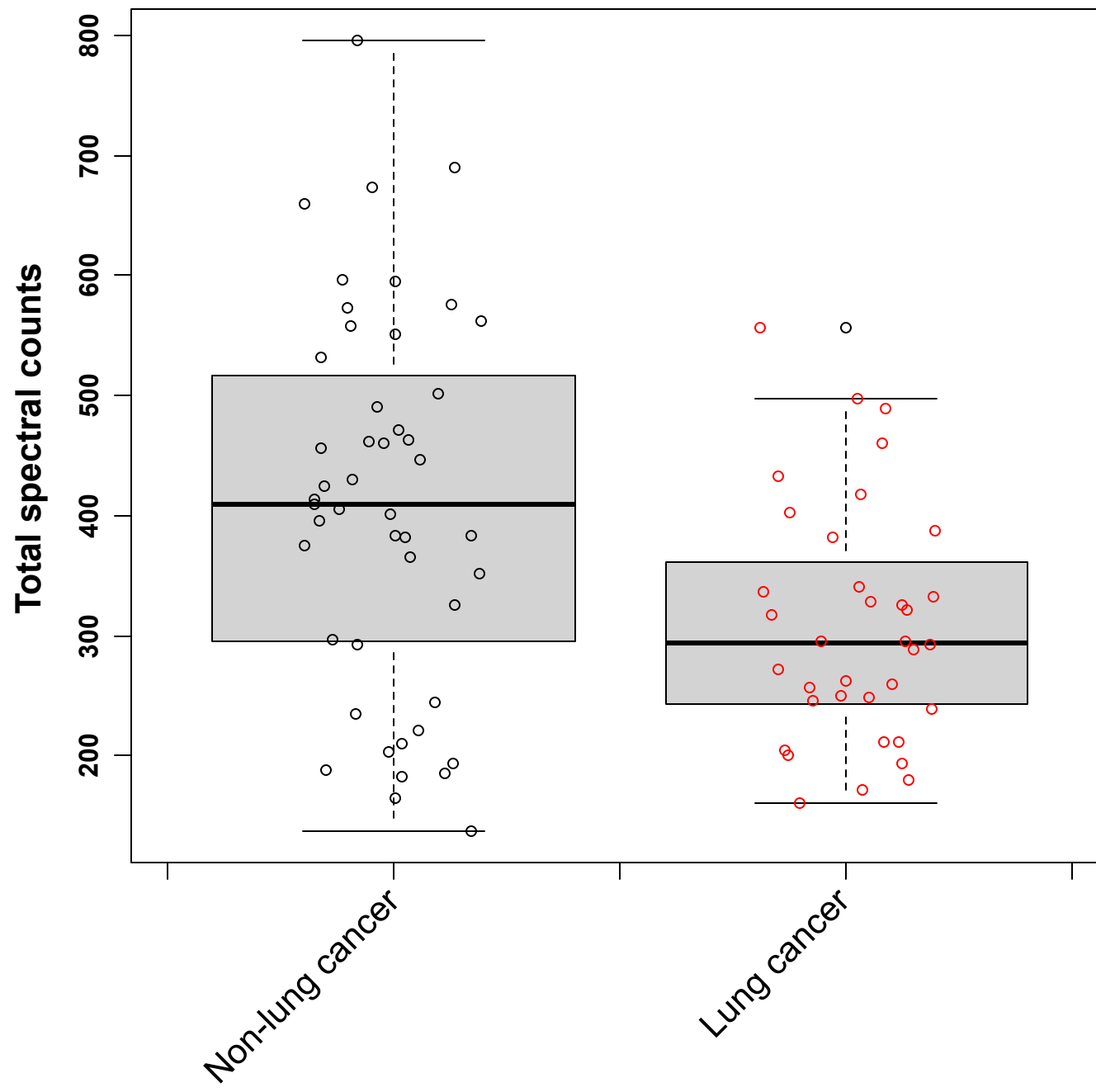


Figure S16

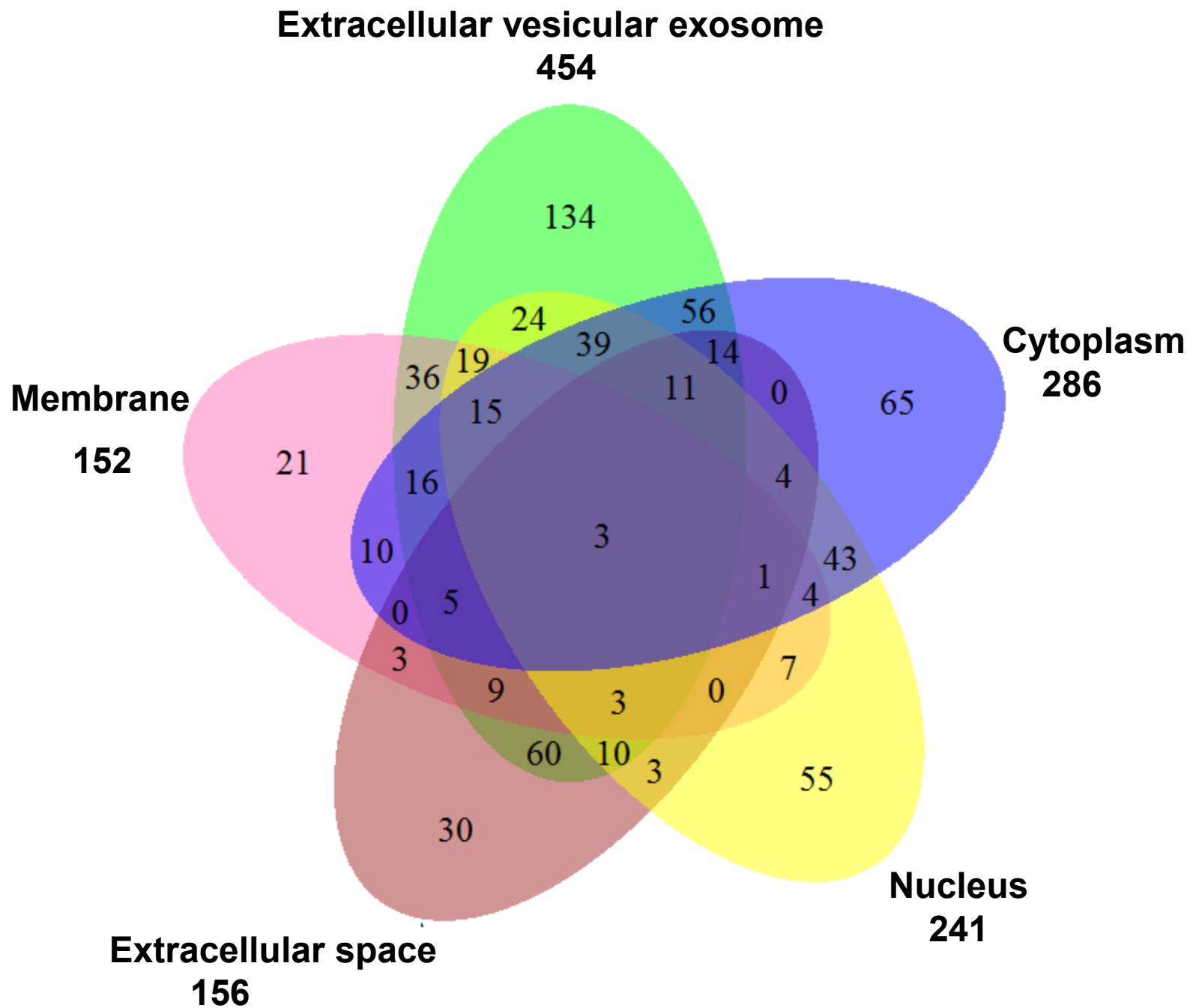
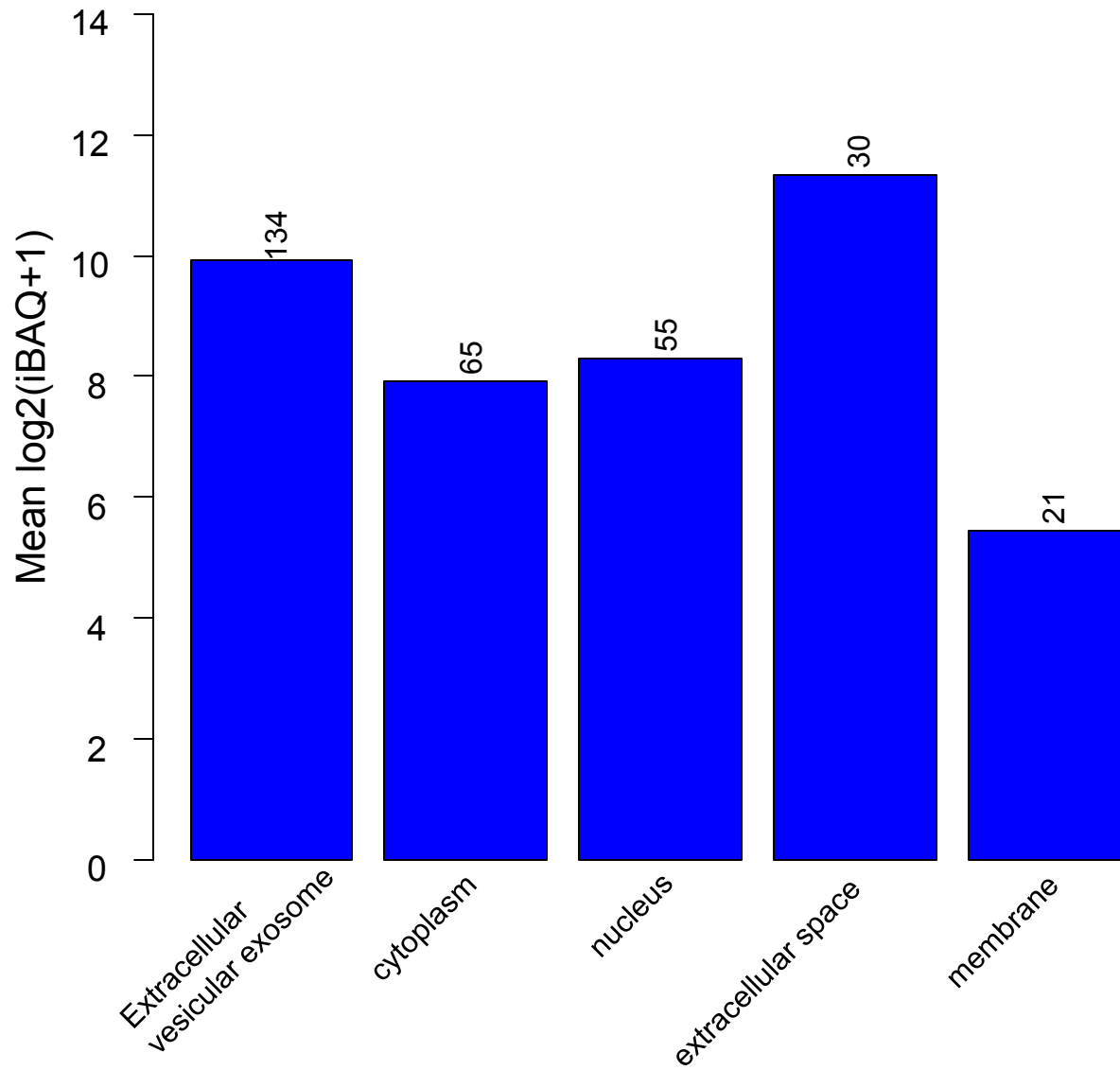
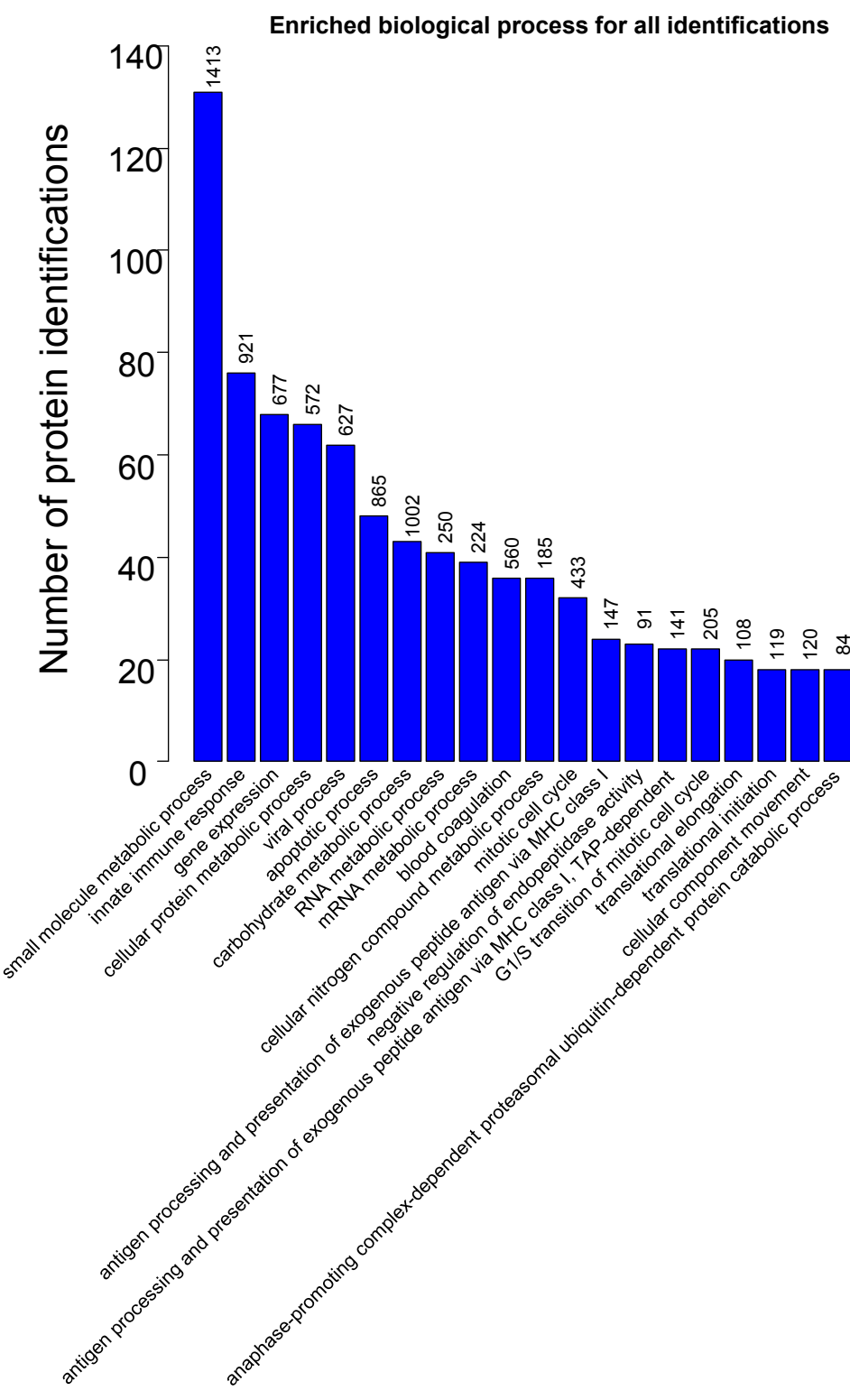


Figure S17

iBAQ for proteins specific for a subcellular localization



A



B

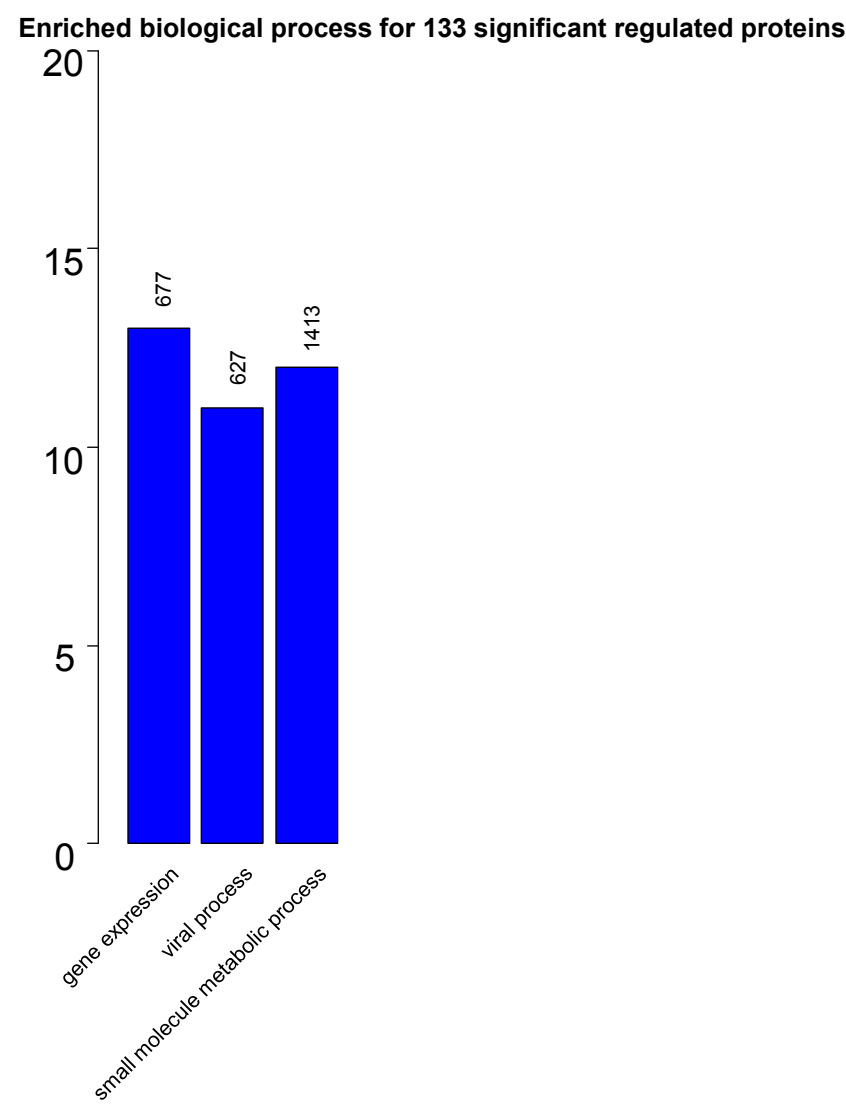


Figure S19

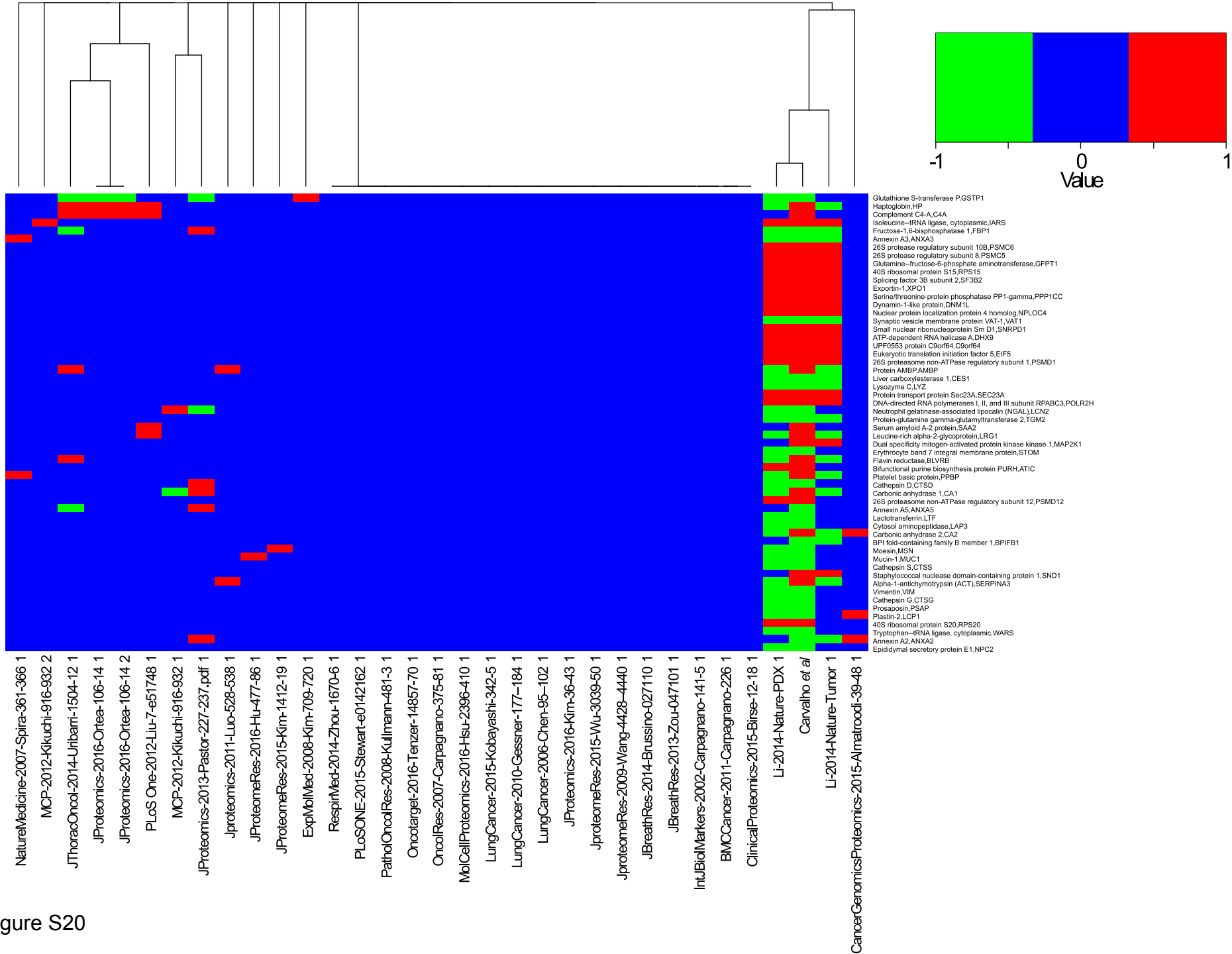


Figure S20

T.cell N= 6
P(Wilcox)= 0.214

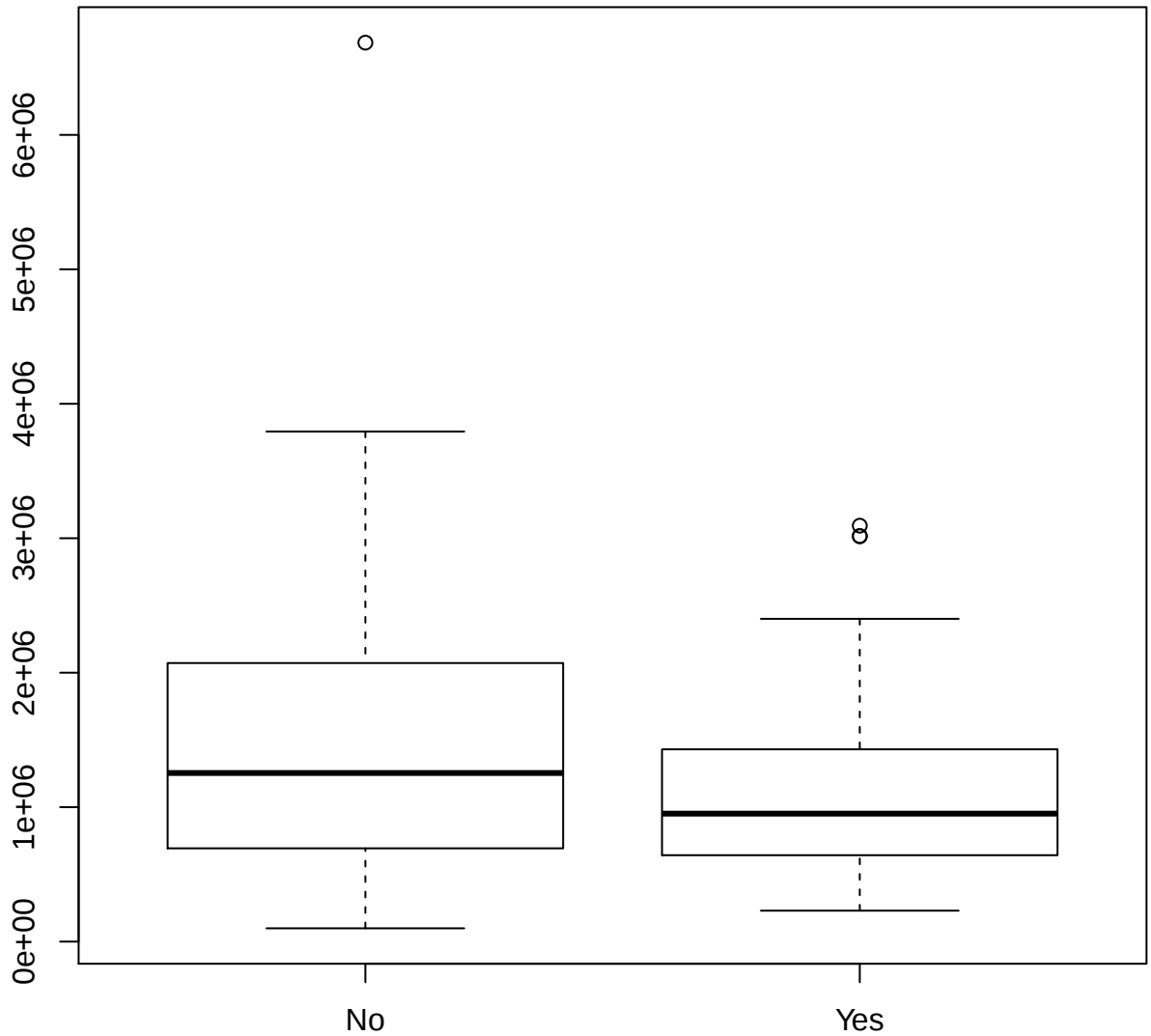
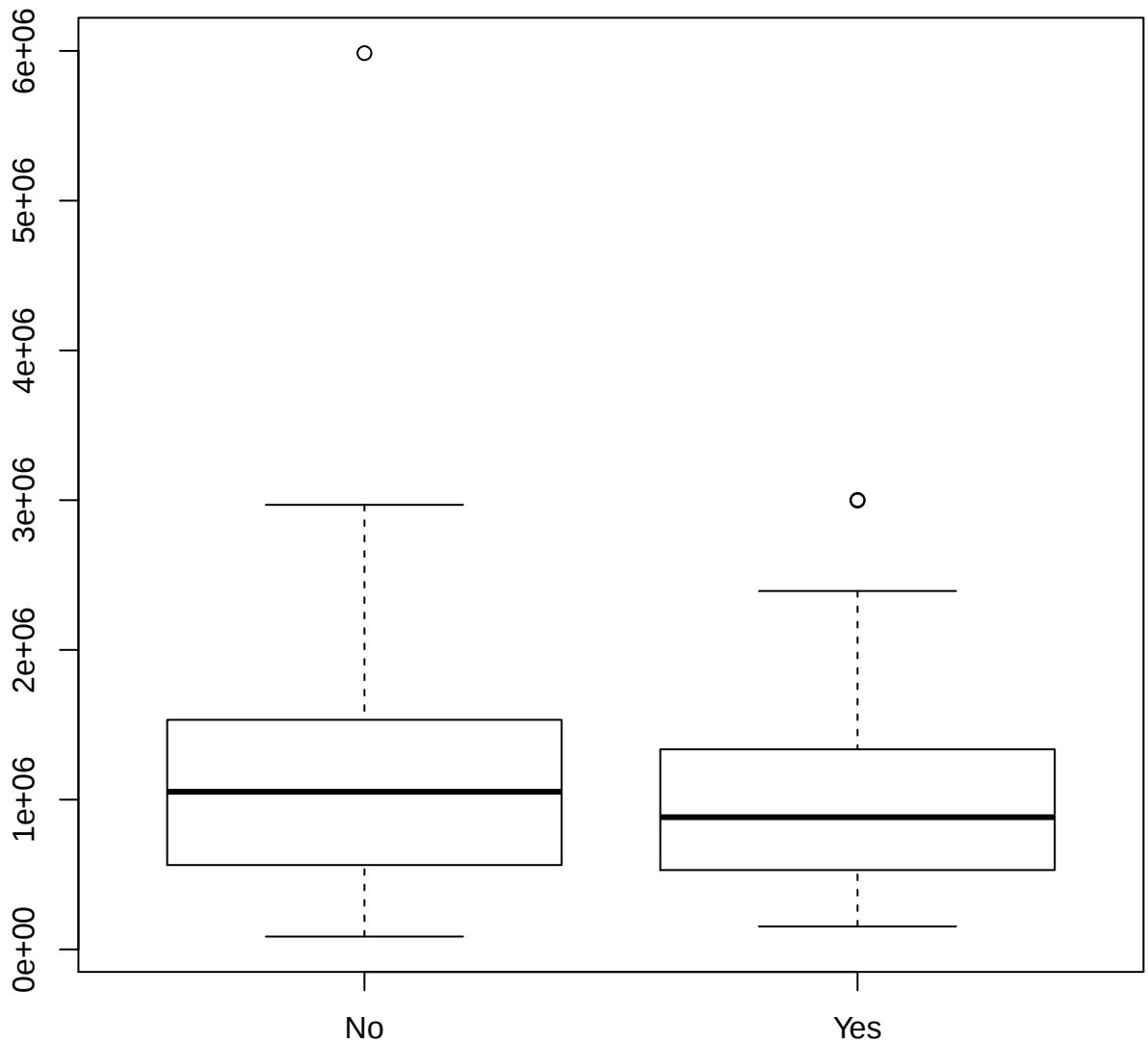


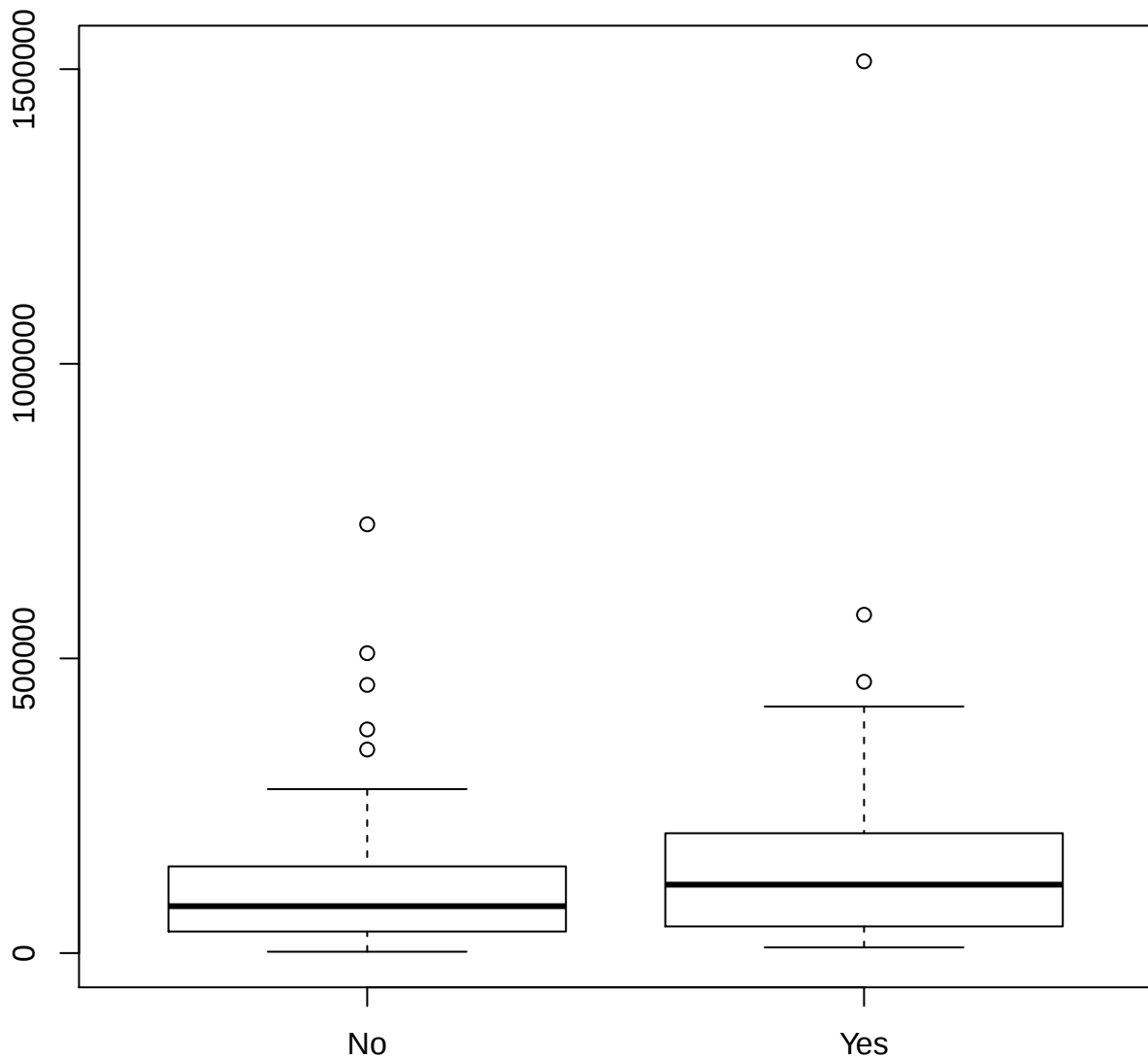
Figure S21

B.cell N= 5
P(Wilcox)= 0.519

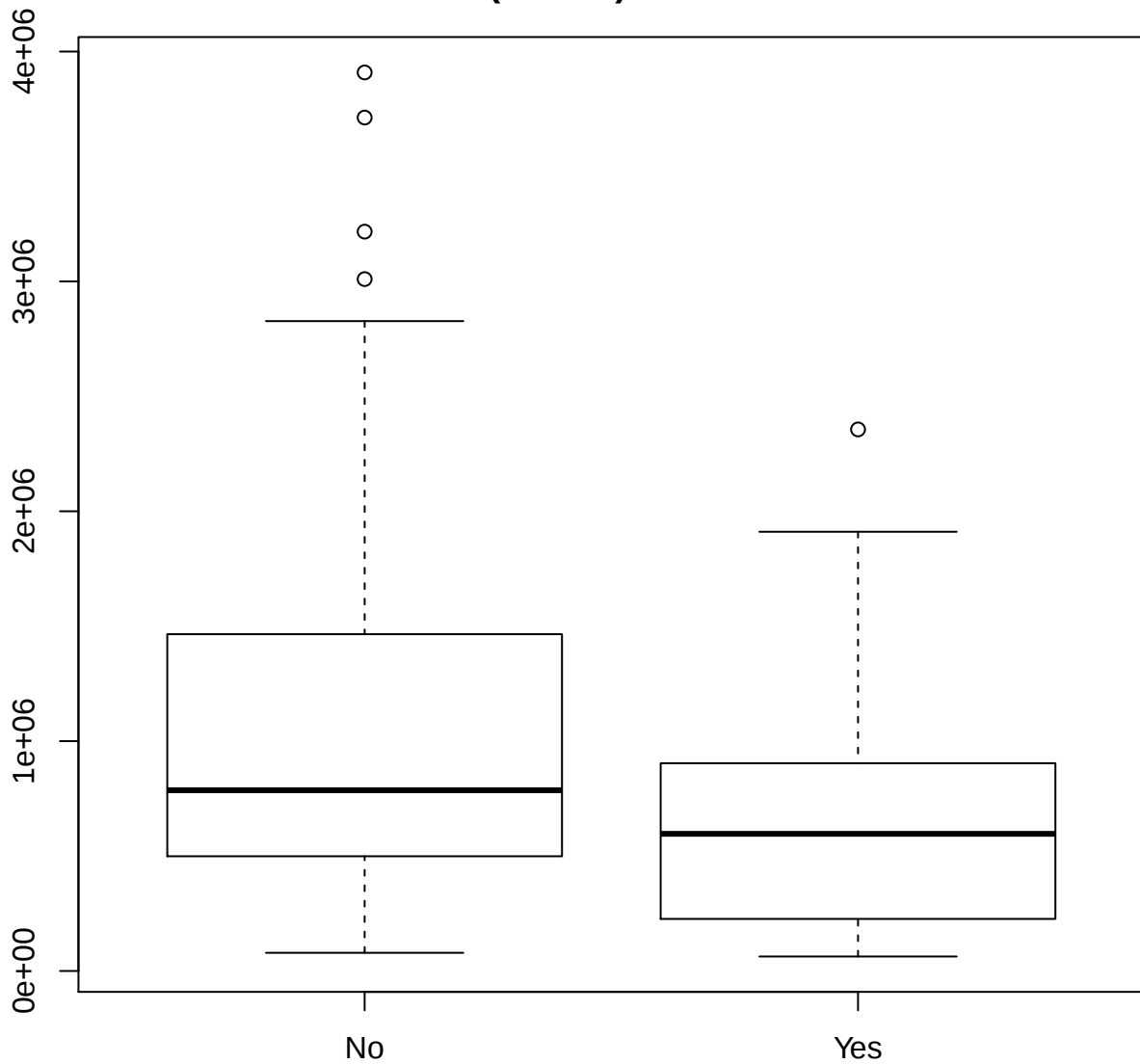


Denditric.cell N= 3

P(Wilcox)= 0.191

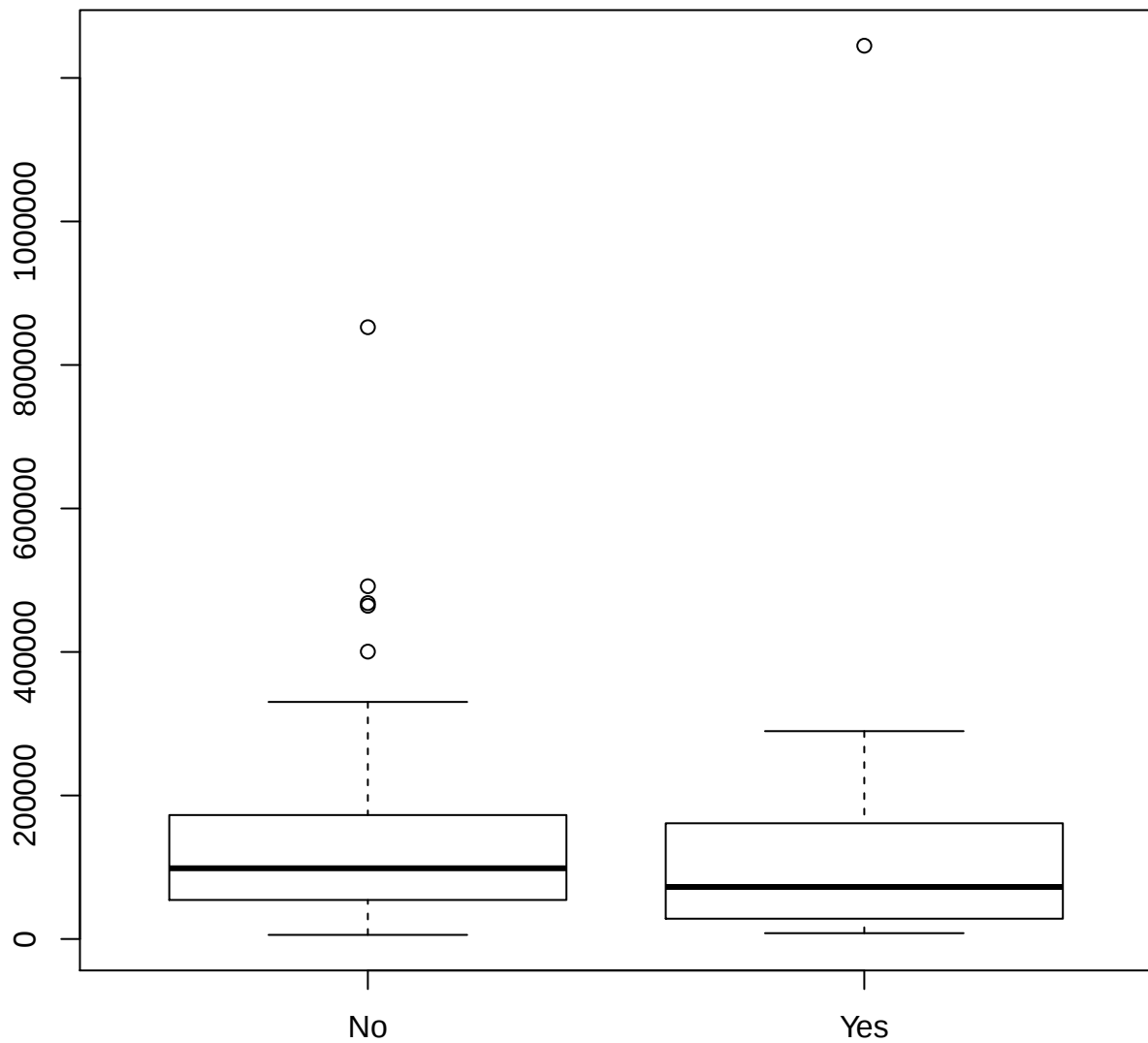


NK.cell N= 5
P(Wilcox)= 0.0433



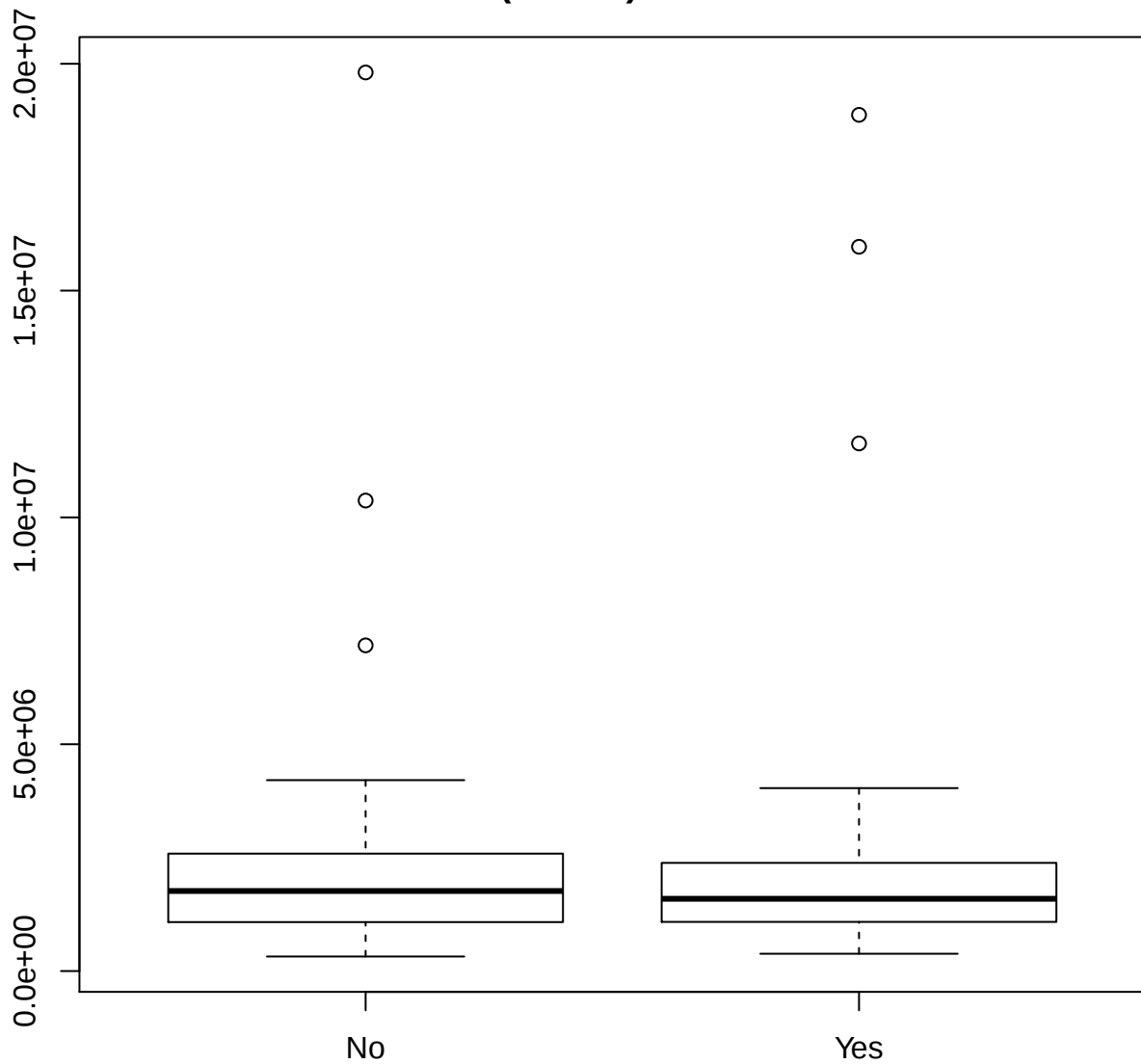
Stem.cell.Precursor N= 2

P(Wilcox)= 0.113

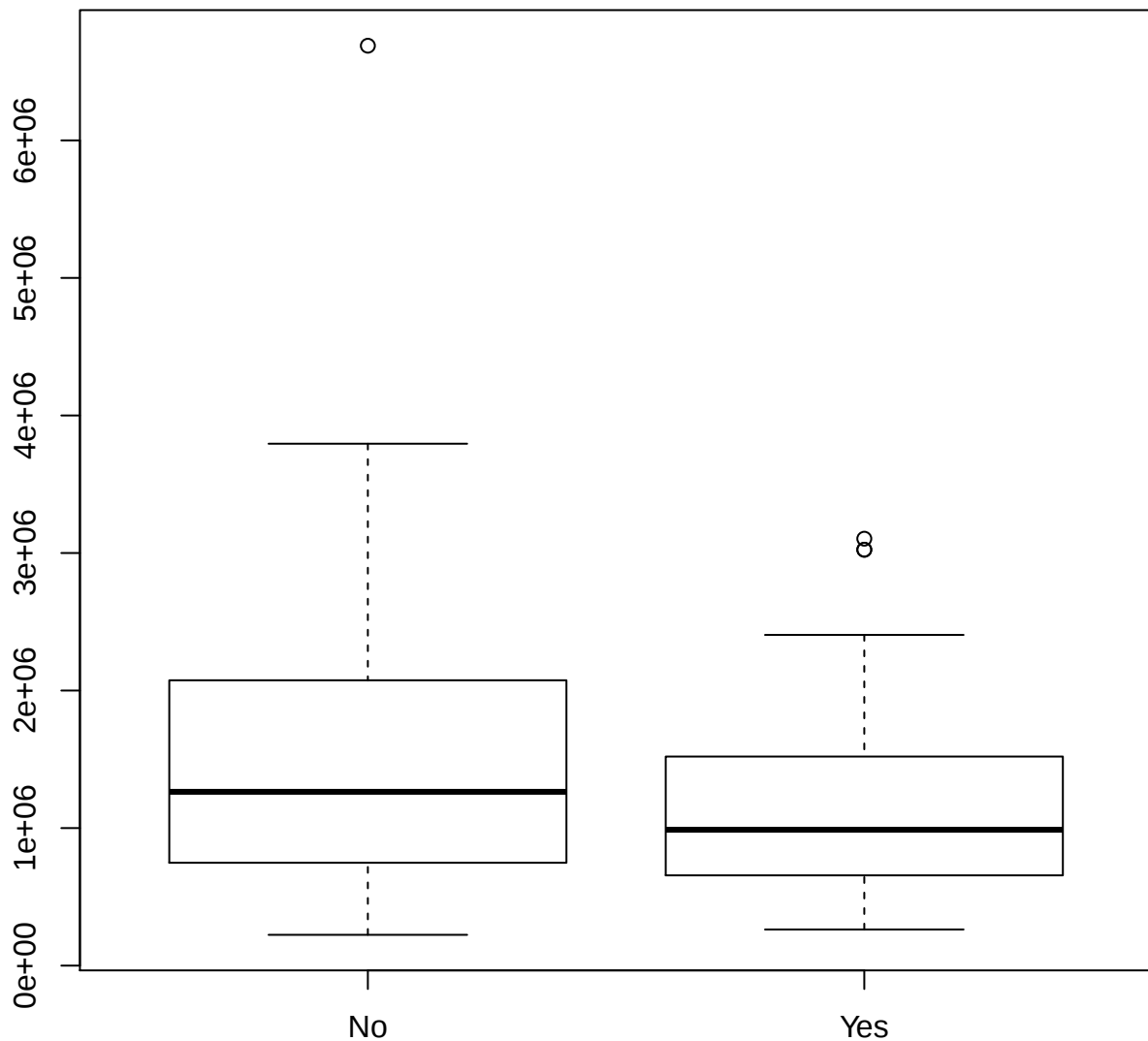


Macrophage.Monocyte N= 11

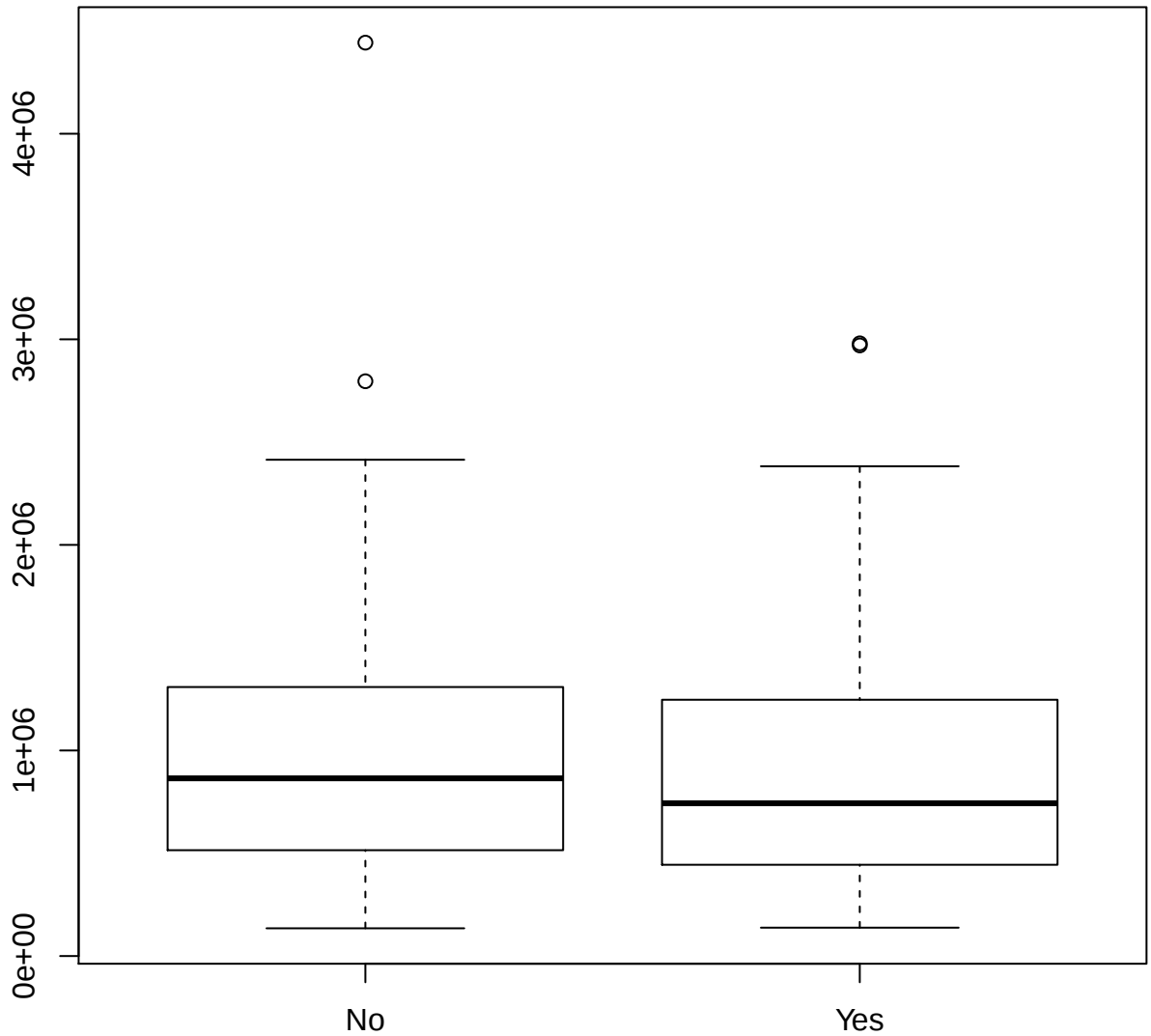
P(Wilcox)= 0.718



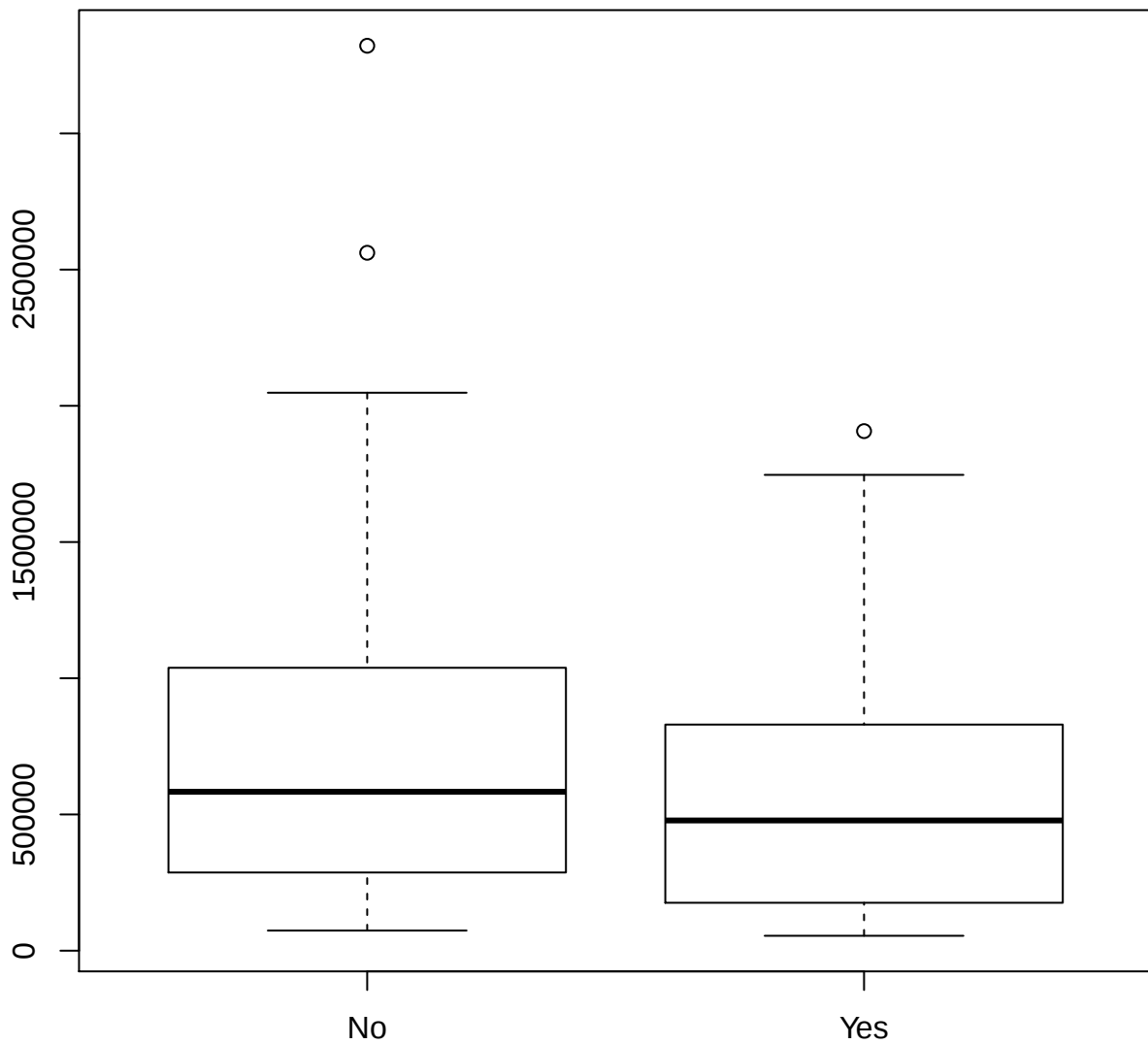
Granulocyte N= 7
P(Wilcox)= 0.235



Platelet N= 5
P(Wilcox)= 0.531

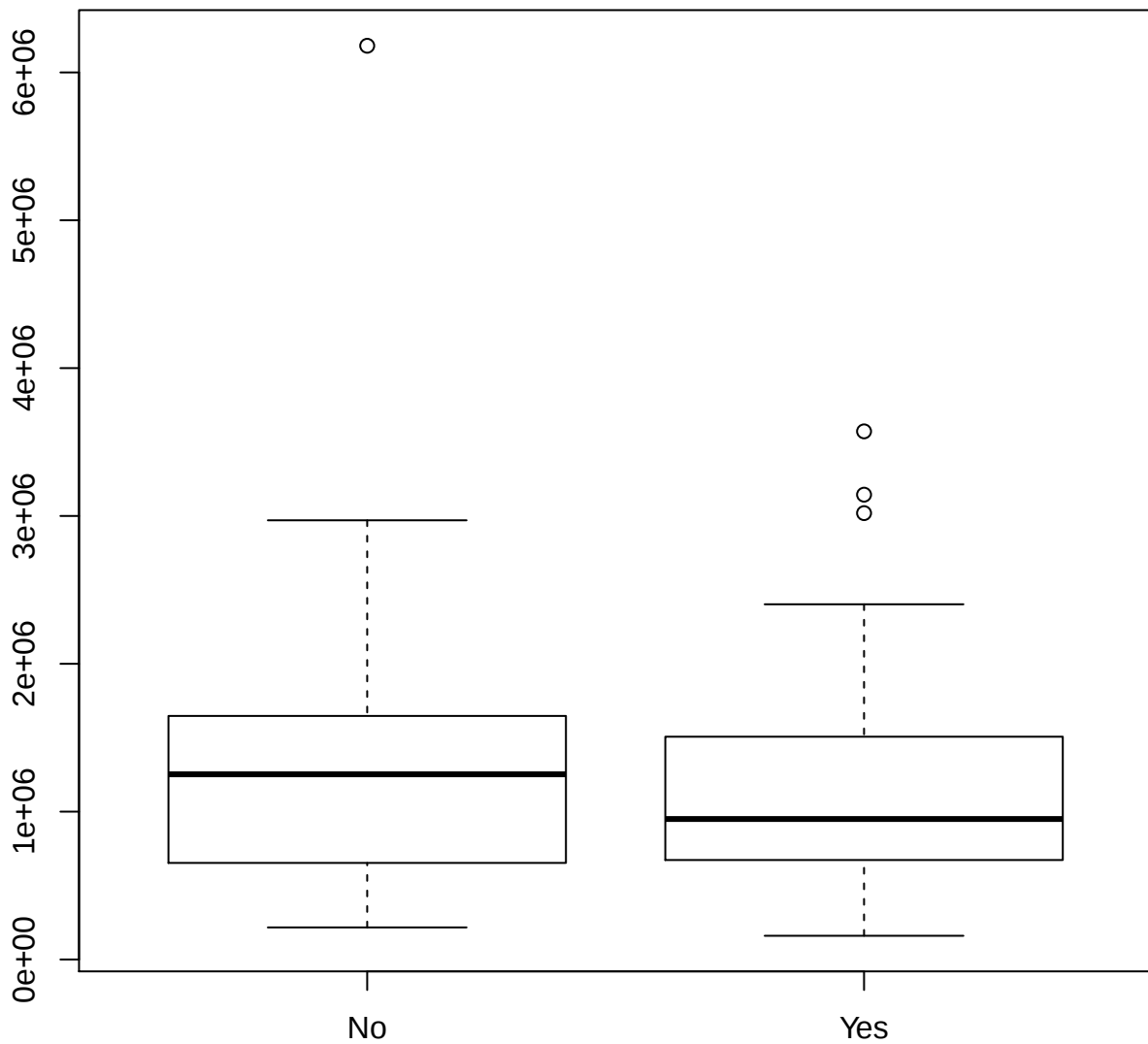


Erythrocyte N= 5
P(Wilcox)= 0.0955



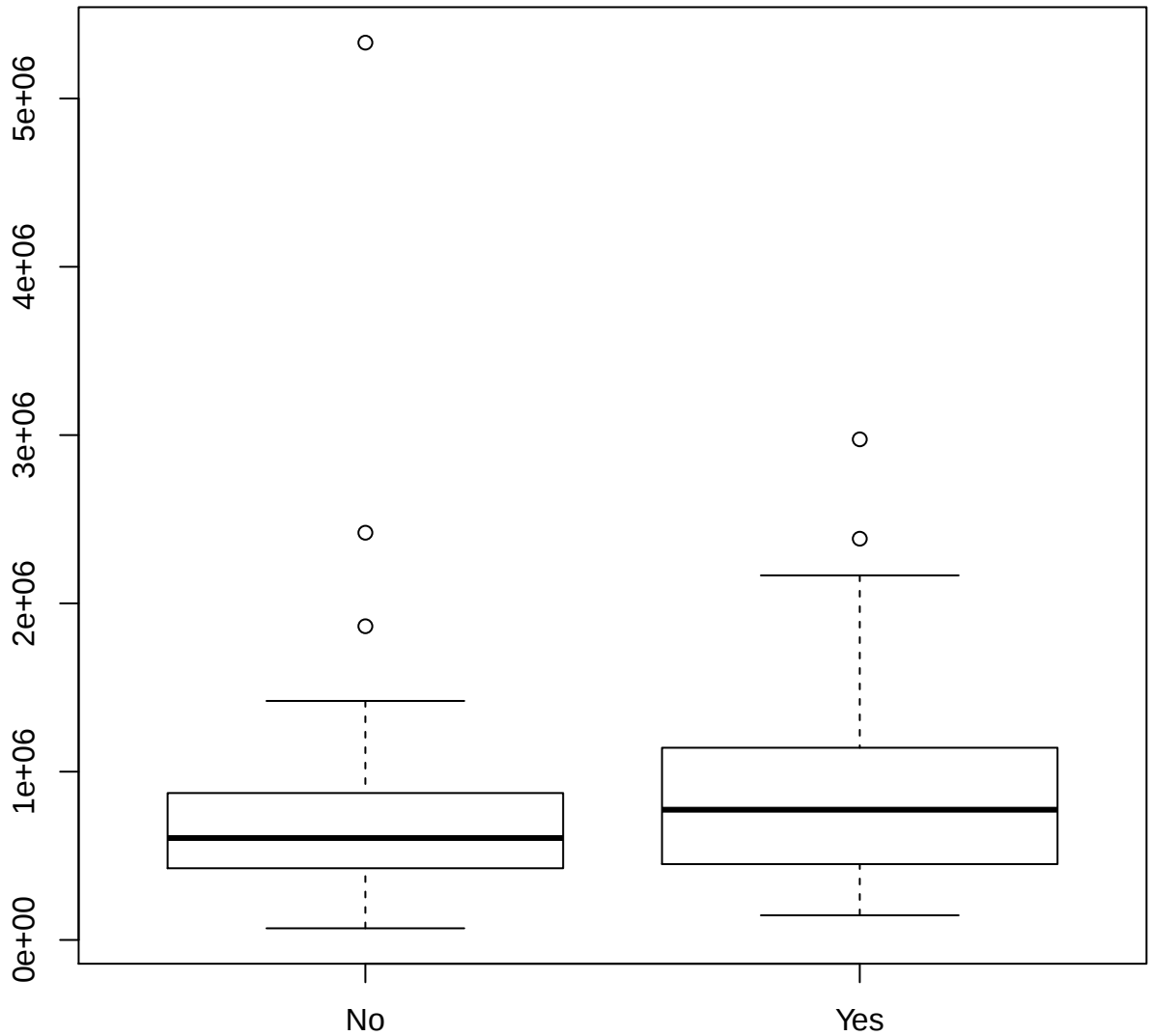
Endothelial.cell N= 7

P(Wilcox)= 0.711



Epithelial.cell N= 4

P(Wilcox)= 0.387



References for table S1

- 1 Almatroodi, S. A., McDonald, C. F., Collins, A. L., Darby, I. A. & Pouniotis, D. S. Quantitative proteomics of bronchoalveolar lavage fluid in lung adenocarcinoma. *Cancer genomics & proteomics* **12**, 39-48 (2015).
- 2 Ortea, I., Rodriguez-Ariza, A., Chicano-Galvez, E., Arenas Vacas, M. S. & Jurado Gamez, B. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. *Journal of proteomics* **138**, 106-114, doi:10.1016/j.jprot.2016.02.010 (2016).
- 3 Uribarri, M. *et al.* A new biomarker panel in bronchoalveolar lavage for an improved lung cancer diagnosis. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* **9**, 1504-1512, doi:10.1097/JTO.0000000000000282 (2014).
- 4 Pastor, M. D. *et al.* Identification of proteomic signatures associated with lung cancer and COPD. *Journal of proteomics* **89**, 227-237, doi:10.1016/j.jprot.2013.04.037 (2013).
- 5 Kim, Y. J. *et al.* Verification of the biomarker candidates for non-small-cell lung cancer using a targeted proteomics approach. *Journal of proteome research* **14**, 1412-1419, doi:10.1021/pr5010828 (2015).
- 6 Wu, H. Y. *et al.* Qualification and Verification of Serological Biomarker Candidates for Lung Adenocarcinoma by Targeted Mass Spectrometry. *Journal of proteome research* **14**, 3039-3050, doi:10.1021/pr501195t (2015).
- 7 Wang, C. L. *et al.* Discovery of retinoblastoma-associated binding protein 46 as a novel prognostic marker for distant metastasis in nonsmall cell lung cancer by combined analysis of cancer cell secretome and pleural effusion proteome. *Journal of proteome research* **8**, 4428-4440, doi:10.1021/pr900160h (2009).
- 8 Luo, X. *et al.* A high-quality secretome of A549 cells aided the discovery of C4b-binding protein as a novel serum biomarker for non-small cell lung cancer. *Journal of proteomics* **74**, 528-538, doi:10.1016/j.jprot.2011.01.011 (2011).
- 9 Chen, Y. *et al.* Elevation of serum l-lactate dehydrogenase B correlated with the clinical stage of lung cancer. *Lung Cancer* **54**, 95-102, doi:10.1016/j.lungcan.2006.06.014 (2006).
- 10 Liu, Y. *et al.* Integrative proteomics and tissue microarray profiling indicate the association between overexpressed serum proteins and non-small cell lung cancer. *PloS one* **7**, e51748, doi:10.1371/journal.pone.0051748 (2012).
- 11 Birse, C. E. *et al.* Blood-based lung cancer biomarkers identified through proteomic discovery in cancer tissues, cell lines and conditioned medium. *Clinical proteomics* **12**, 18, doi:10.1186/s12014-015-9090-9 (2015).
- 12 Kobayashi, M. *et al.* Calnexin is a novel sero-diagnostic marker for lung cancer. *Lung Cancer* **90**, 342-345, doi:10.1016/j.lungcan.2015.08.015 (2015).
- 13 Kim, Y. I. *et al.* Meta-markers for the differential diagnosis of lung cancer and lung disease. *Journal of proteomics* **148**, 36-43, doi:10.1016/j.jprot.2016.04.052 (2016).
- 14 Kikuchi, T. *et al.* In-depth proteomic analysis of nonsmall cell lung cancer to discover molecular targets and candidate biomarkers. *Molecular & cellular proteomics : MCP* **11**, 916-932, doi:10.1074/mcp.M111.015370 (2012).
- 15 Tenzer, S. *et al.* Integrated quantitative proteomic and transcriptomic analysis of lung tumor and control tissue: a lung cancer showcase. *Oncotarget* **7**, 14857-14870, doi:10.18632/oncotarget.7562 (2016).

- 16 Li, L. *et al.* Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. *Nature communications* **5**, 5469, doi:10.1038/ncomms6469 (2014).
- 17 Stewart, P. A. *et al.* A Pilot Proteogenomic Study with Data Integration Identifies MCT1 and GLUT1 as Prognostic Markers in Lung Adenocarcinoma. *PloS one* **10**, e0142162, doi:10.1371/journal.pone.0142162 (2015).
- 18 Hsu, C. H. *et al.* Identification and Characterization of Potential Biomarkers by Quantitative Tissue Proteomics of Primary Lung Adenocarcinoma. *Molecular & cellular proteomics : MCP* **15**, 2396-2410, doi:10.1074/mcp.M115.057026 (2016).
- 19 Kim, J. E., Koo, K. H., Kim, Y. H., Sohn, J. & Park, Y. G. Identification of potential lung cancer biomarkers using an in vitro carcinogenesis model. *Experimental & molecular medicine* **40**, 709-720, doi:10.3858/emm.2008.40.6.709 (2008).
- 20 Hu, R. *et al.* Quantitative Secretomic Analysis Identifies Extracellular Protein Factors That Modulate the Metastatic Phenotype of Non-Small Cell Lung Cancer. *Journal of proteome research* **15**, 477-486, doi:10.1021/acs.jproteome.5b00819 (2016).
- 21 Spira, A. *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nature medicine* **13**, 361-366, doi:10.1038/nm1556 (2007).
- 22 Brussino, L. *et al.* Inflammatory cytokines and VEGF measured in exhaled breath condensate are correlated with tumor mass in non-small cell lung cancer. *Journal of breath research* **8**, 027110, doi:10.1088/1752-7155/8/2/027110 (2014).
- 23 Carpagnano, G. E., Resta, O., Foschino-Barbaro, M. P., Gramiccioni, E. & Carpagnano, F. Interleukin-6 is increased in breath condensate of patients with non-small cell lung cancer. *The International journal of biological markers* **17**, 141-145 (2002).
- 24 Carpagnano, G. E. *et al.* IL-2, TNF-alpha, and leptin: local versus systemic concentrations in NSCLC patients. *Oncology research* **16**, 375-381 (2007).
- 25 Gessner, C. *et al.* Angiogenic markers in breath condensate identify non-small cell lung cancer. *Lung Cancer* **68**, 177-184, doi:10.1016/j.lungcan.2009.06.010 (2010).
- 26 Carpagnano, G. E. *et al.* Neutrophilic airways inflammation in lung cancer: the role of exhaled LTB-4 and IL-8. *BMC cancer* **11**, 226, doi:10.1186/1471-2407-11-226 (2011).
- 27 Kullmann, T., Barta, I., Csiszer, E., Antus, B. & Horvath, I. Differential cytokine pattern in the exhaled breath of patients with lung cancer. *Pathology oncology research : POR* **14**, 481-483, doi:10.1007/s12253-008-9046-8 (2008).
- 28 Zou, Y. *et al.* CEA, SCC and NSE levels in exhaled breath condensate--possible markers for early detection of lung cancer. *Journal of breath research* **7**, 047101, doi:10.1088/1752-7155/7/4/047101 (2013).
- 29 Carpagnano, G. E. *et al.* Cigarette smoke and increased COX-2 and survivin levels in exhaled breath condensate of lung cancer patients: how hot is the link? *Lung Cancer* **67**, 108-113, doi:10.1016/j.lungcan.2009.03.033 (2010).
- 30 Zhou, F. *et al.* Increased levels of exhaled sICAM1, sVCAM1, and sE-selectin in patients with non-small cell lung cancer. *Respiratory medicine* **108**, 1670-1676, doi:10.1016/j.rmed.2014.08.003 (2014).