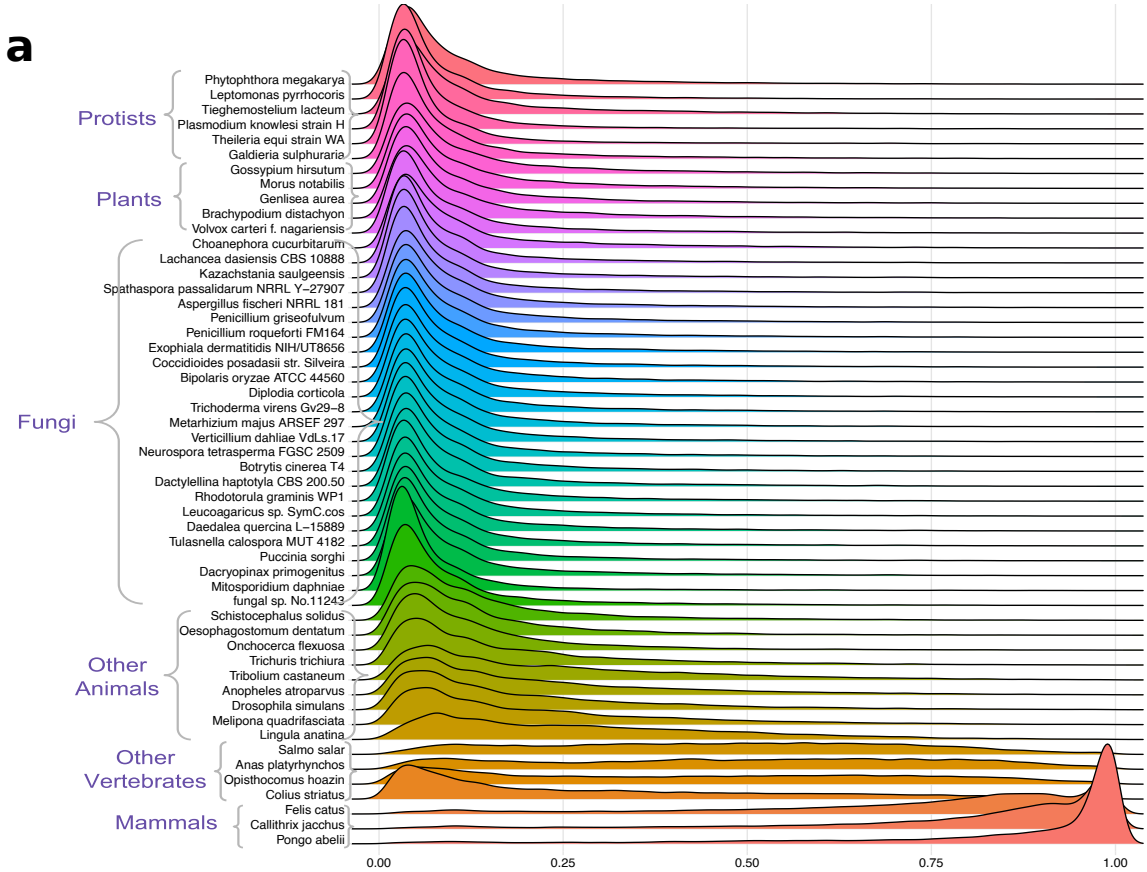

a
Protists
Phytophthora megakarya
Leptomonas pyrrhocoris
Tieghemostelium lacteum
Plasmodium knowlesi strain H
Theileria equi strain WA
Galdieria sulphuraria
Plants
Gossypium hirsutum
Morus notabilis
Genlisea aurea
Brachypodium distachyon
Volvox carteri f. nagariensis
Fungi
Choanephora cucurbitarum
Lachancea dasiensis CBS 10888
Kazachstania saulgeensis
Spathaspora passalidarum NRRL Y−27907
Aspergillus fischeri NRRL 181
Penicillium griseofulvum
Penicillium roqueforti FM164
Exophiala dermatitidis NIH/UT8656
Coccidioides posadasii str. Silveira
Bipolaris oryzae ATCC 44560
Diplodia corticola
Trichoderma virens Gv29−8
Metarhizium majus ARSEF 297
Verticillium dahliae VdLs.17
Neurospora tetrasperma FGSC 2509
Botrytis cinerea T4
Dactylellina haptotyla CBS 200.50
Rhodotorula graminis WP1
Leucoagaricus sp. SymC.cos
Daedalea quercina L−15889
Tulasnella calospora MUT 4182
Puccinia sorghi
Dacryopinax primogenitus
Mitosporidium daphniae
fungal sp. No.11243
Other Animals
Schistocephalus solidus
Oesophagostomum dentatum
Onchocerca flexuosa
Trichuris trichiura
Tribolium castaneum
Anopheles atroparvus
Drosophila simulans
Melipona quadrifasciata
Lingula anatina
Other Vertebrates
Salmo salar
Anas platyrhynchos
Opisthocomus hoazin
Colius striatus
Mammals
Felis catus
Callithrix jacchus
Pongo abelii
0.00
0.25
0.50
0.75
1.00

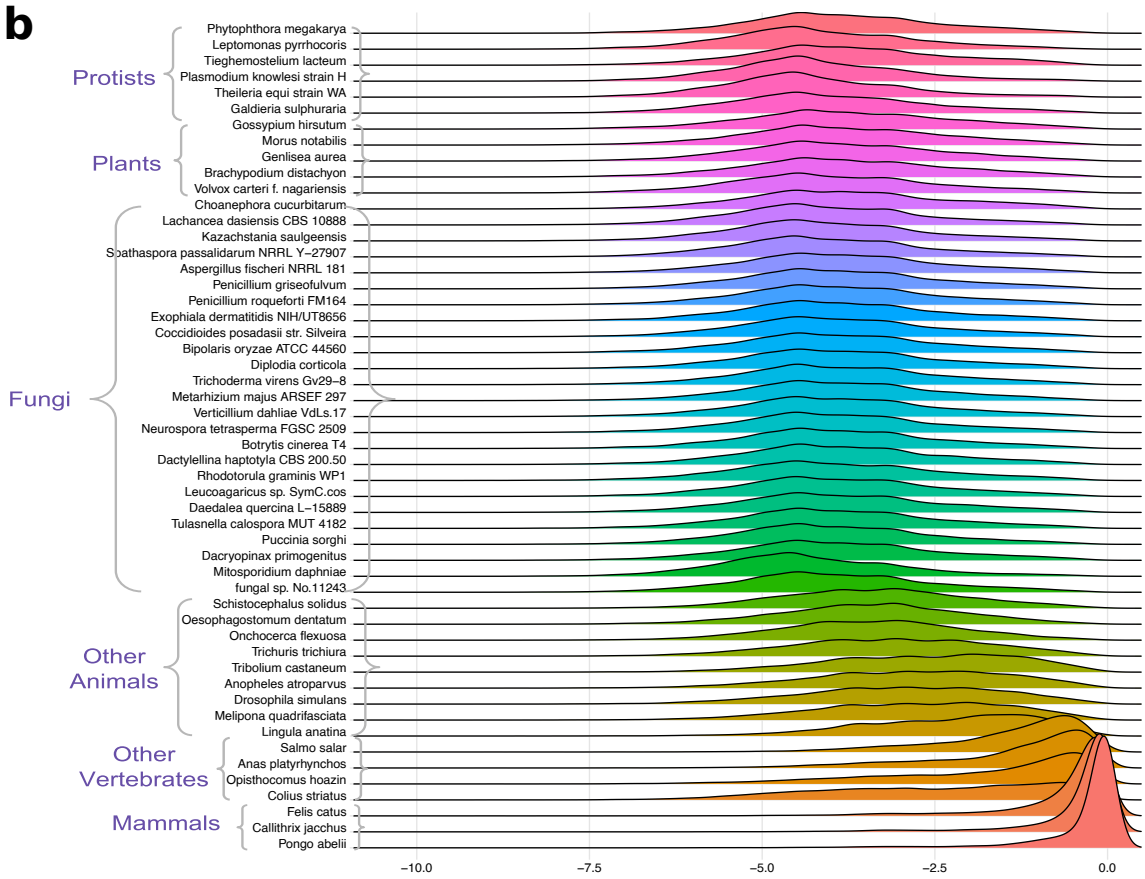

b
Protists
Phytophthora megakarya
Leptomonas pyrrhocoris
Tieghemostelium lacteum
Plasmodium knowlesi strain H
Theileria equi strain WA
Galdieria sulphuraria
Plants
Gossypium hirsutum
Morus notabilis
Genlisea aurea
Brachypodium distachyon
Volvox carteri f. nagariensis
Fungi
Choanephora cucurbitarum
Lachancea dasiensis CBS 10888
Kazachstania saulgeensis
Spathaspora passalidarum NRRL Y−27907
Aspergillus fischeri NRRL 181
Penicillium griseofulvum
Penicillium roqueforti FM164
Exophiala dermatitidis NIH/UT8656
Coccidioides posadasii str. Silveira
Bipolaris oryzae ATCC 44560
Diplodia corticola
Trichoderma virens Gv29−8
Metarhizium majus ARSEF 297
Verticillium dahliae VdLs.17
Neurospora tetrasperma FGSC 2509
Botrytis cinerea T4
Dactylellina haptotyla CBS 200.50
Rhodotorula graminis WP1
Leucoagaricus sp. SymC.cos
Daedalea quercina L−15889
Tulasnella calospora MUT 4182
Puccinia sorghi
Dacryopinax primogenitus
Mitosporidium daphniae
fungal sp. No.11243
Other Animals
Schistocephalus solidus
Oesophagostomum dentatum
Onchocerca flexuosa
Trichuris trichiura
Tribolium castaneum
Anopheles atroparvus
Drosophila simulans
Melipona quadrifasciata
Lingula anatina
Other Vertebrates
Salmo salar
Anas platyrhynchos
Opisthocomus hoazin
Colius striatus
Mammals
Felis catus
Callithrix jacchus
Pongo abelii
−10.0
−7.5
−5.0
−2.5
0.0

Figure S1 Transformation stage

Distributions of BLASTP bit scores following division by a query genome vs. itself. Fifty-two selected species in the NPP (every 20th species) are shown, from species close to *Homo sapiens* (lower rows) to distant species (upper rows). **a.** Scores before transformation. As the species become more distant from *Homo sapiens*, the distribution becomes more right-tailed because most bit scores are low, reaching close to zero following division. **b.** Scores after the log2 transformation show that the distributions of distant species become more symmetric. Distributions of close species still reflect the concentration of the high scores.

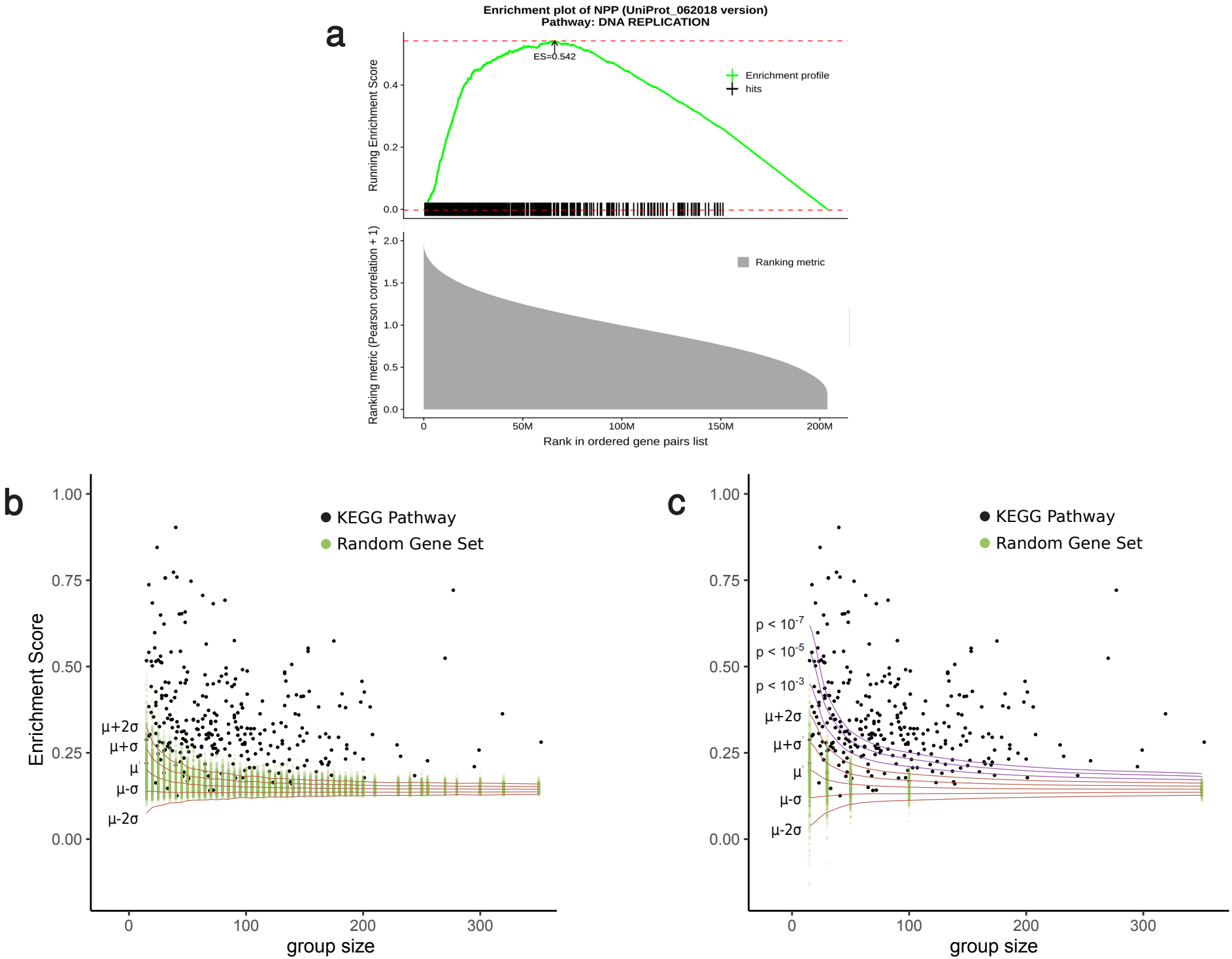
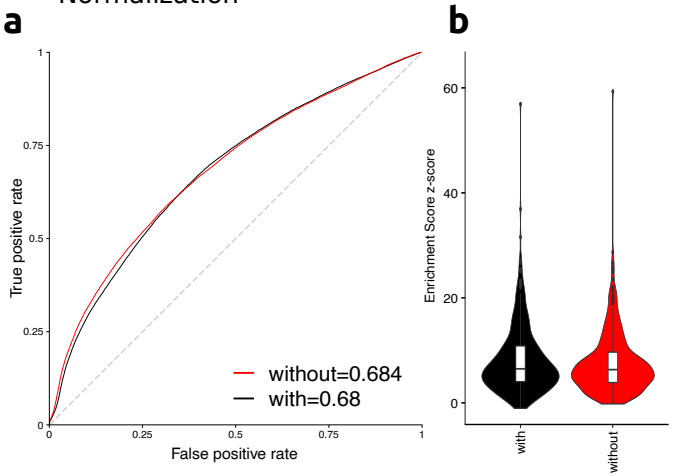

a
Enrichment plot of NPP (UniProt_062018 version)
Pathway: DNA REPLICATION
ES=0.542
Enrichment profile
hits
Running Enrichment Score
0.0
0.2
0.4
Ranking metric
Ranking metric (Pearson correlation + 1)
0.0
0.5
1.0
1.5
2.0
0
50M
100M
150M
200M
Rank in ordered gene pairs list
b
KEGG Pathway
Random Gene Set
Enrichment Score
1.00
0.75
0.50
0.25
0.00
μ+2σ
μ+σ
μ
μ-σ
μ-2σ
0
100
200
300
group size
c
KEGG Pathway
Random Gene Set
1.00
0.75
0.50
0.25
0.00
p < 10-7
p < 10-5
p < 10-3
μ+2σ
μ+σ
μ
μ-σ
μ-2σ
0
100
200
300
group size

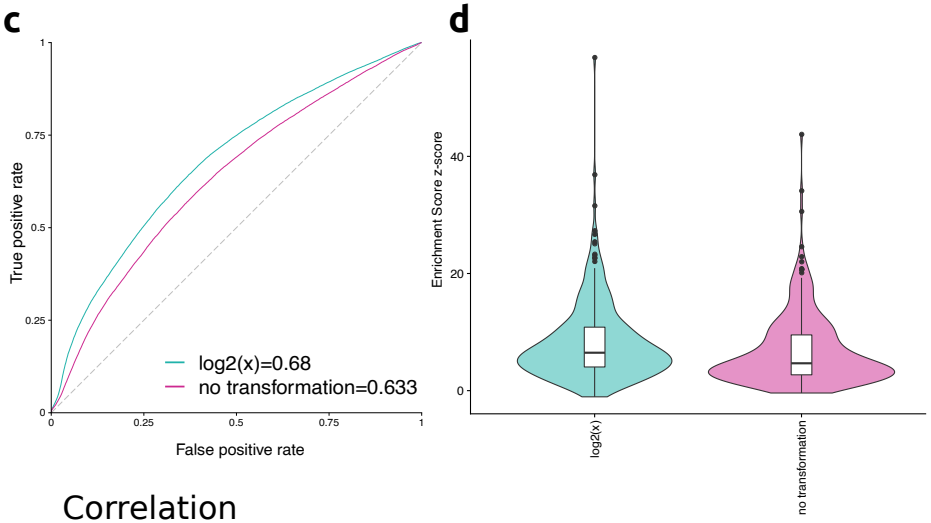
Figure S2 ES score test

For each gene set S in the KEGG pathways, we calculated the enrichment score (ES) of S gene pairs in all NPP gene pairs ranked by correlation of gene pairs phylogenetic profiles. **a.** Example of an ES calculation for the DNA replication pathway. The upper line plot demonstrates the running sum for DNA Replication through the 200M ranked gene pairs list, including the location of the maximum ES value of 0.542. The black ticks represent the hits of gene pairs that exist in the pathway along the ranked NPP gene pairs list. The lower area plot shows the Pearson correlation values of the gene pairs in the ranked list, with an offset of 1 added. **b.** ES of NPP for all KEGG pathways (black dots) as compared with the ES for 1000 random gene sets for each size in the KEGG pathway gene set sizes (green dots). The brown lines denote the mean, mean+-sd and mean+-2*sd of the ES of the random gene sets for each size. **c.** ES of NPP for all KEGG pathways (black dots) as compared with the ES for 1000 random gene sets for five representative sizes (15, 30, 50, 100, 350) (green dots). The scores for the random gene sets are “mirrored” around the mean of the values for each size. The brown lines denote the smoothed estimating lines for the mean, mean+-sd and mean+-2*sd of the mirrored random scores. The purple lines indicate the p-values of the ES values for KEGG pathways.

Length Normalization



With/Without Transformation



Correlation

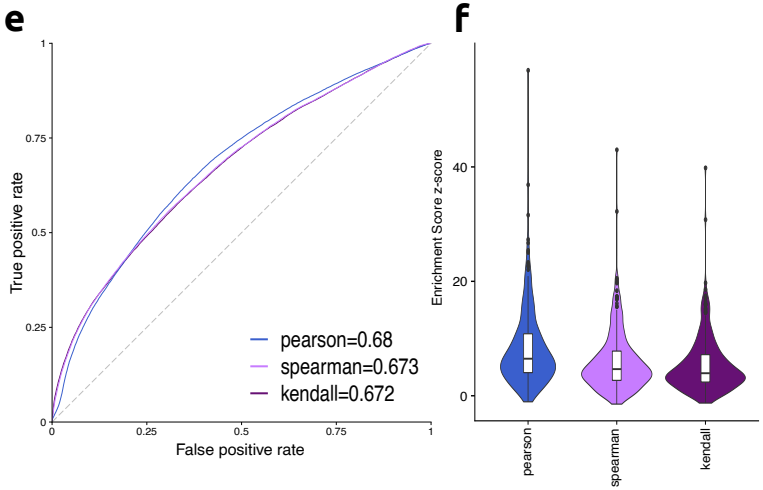


Figure S3 Test results - length normalization and correlation

The results of the two tests for length normalization and correlation. The results are also summarized in Table 1. **Panels a,c,e.** ROC curve analysis of NPP built with different parameters. NPP correlations for gene pairs from CORUM (positive) were compared to control (negative). The X and Y axes represent the False Positive Rate (FPR = 1 − specificity) and True Positive Rate (TPR, sensitivity), respectively. **Panels b,d,f.** Distributions of the ES z-score values of NPP built with different parameters for 298 KEGG pathways. **Panels a,b.** Effect of length normalization on NPP. **Panels c,d**. Effect of log2 transformation on NPP. **Panels e,f.** Different correlation methods.

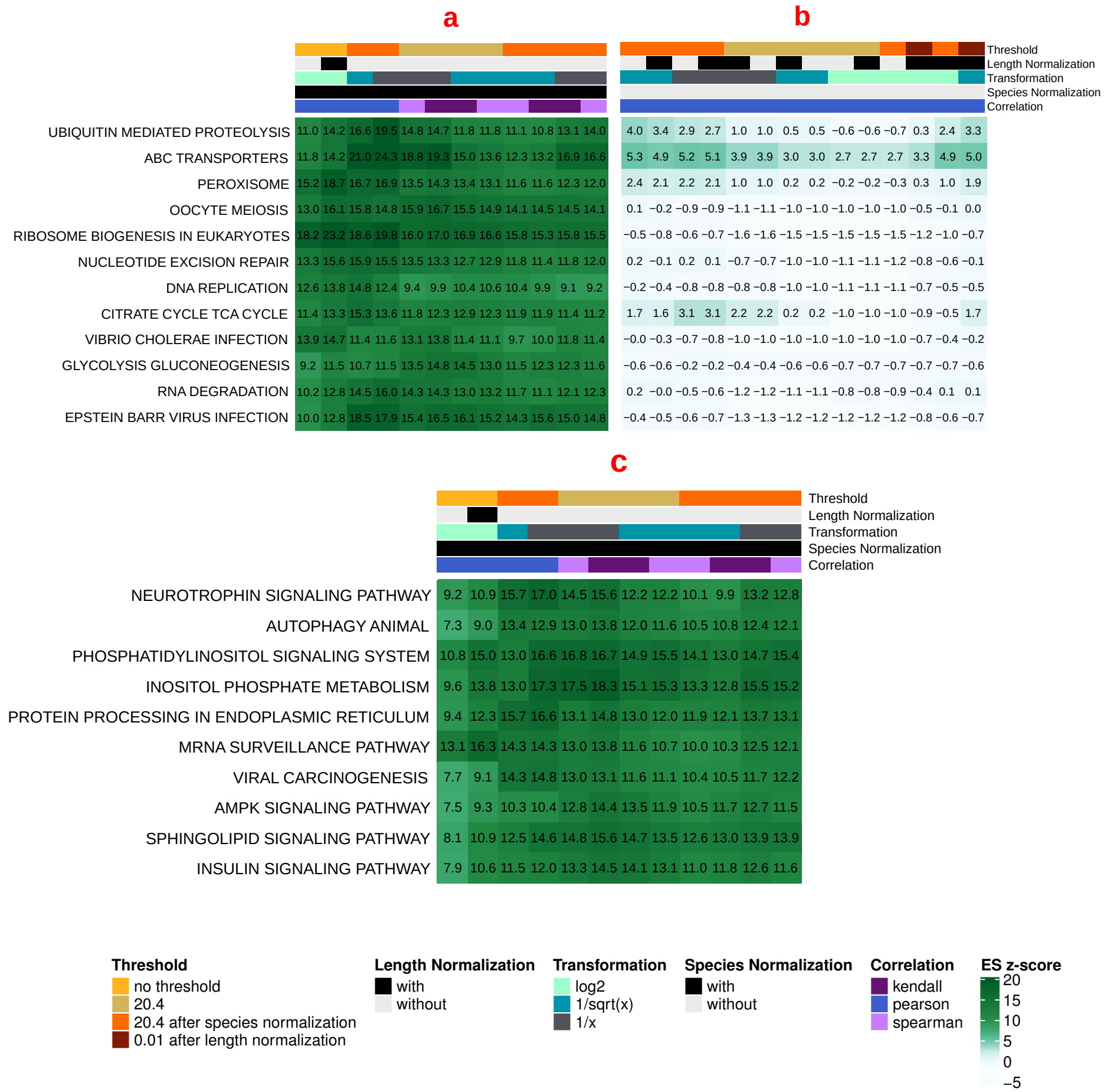

a
b
c
Threshold
Length Normalization
Transformation
Species Normalization
Correlation
UBIQUITIN MEDIATED PROTEOLYSIS
ABC TRANSPORTERS
PEROXISOME
OOCYTE MEIOSIS
RIBOSOME BIOGENESIS IN EUKARYOTES
NUCLEOTIDE EXCISION REPAIR
DNA REPLICATION
CITRATE CYCLE TCA CYCLE
VIBRIO CHOLERAE INFECTION
GLYCOLYSIS GLUCONEOGENESIS
RNA DEGRADATION
EPSTEIN BARR VIRUS INFECTION
NEUROTROPHIN SIGNALING PATHWAY
AUTOPHAGY ANIMAL
PHOSPHATIDYLINOSITOL SIGNALING SYSTEM
INOSITOL PHOSPHATE METABOLISM
PROTEIN PROCESSING IN ENDOPLASMIC RETICULUM
MRNA SURVEILLANCE PATHWAY
VIRAL CARCINOGENESIS
AMPK SIGNALING PATHWAY
SPHINGOLIPID SIGNALING PATHWAY
INSULIN SIGNALING PATHWAY
Threshold
no threshold
20.4
20.4 after species normalization
0.01 after length normalization
Length Normalization
with
without
Transformation
log2
1/sqrt(x)
1/x
Species Normalization
with
without
Correlation
kendall
pearson
spearman
ES z-score
20
15
10
5
0
−5

Figure S4 Combination heatmap zoom a-c

Heatmap of the ES z-scores of specific KEGG pathways in part of the 126 different combinations of the NPP parameters. Rows and columns represent pathways and parameter combinations, respectively. **a,b.** Several pathways that are known to be captured well by phylogenetic profiling, here receiving high ES z-scores for the combinations in (a) and low ES for the parameter combinations in (b). **c.** Pathways related to signaling and to basic cellular processing, which are usually not captured well by PP, receive high ES z-scores for specific parameter combinations.

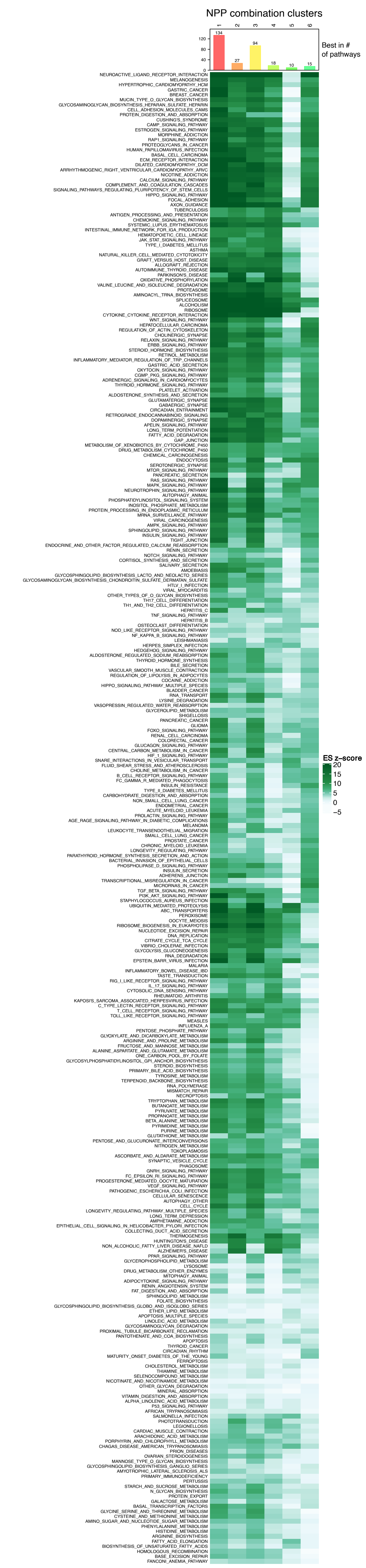

NPP combination clusters
1
2
3
4
5
6
134
27
94
18
10
15
0
40
80
120
Best in #
of pathways
NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION
MELANOGENESIS
HYPERTROPHIC_CARDIOMYOPATHY_HCM
GASTRIC_CANCER
BREAST_CANCER
MUCIN_TYPE_O_GLYCAN_BIOSYNTHESIS
GLYCOSAMINOGLYCAN_BIOSYNTHESIS_HEPARAN_SULFATE_HEPARIN
CELL_ADHESION_MOLECULES_CAMS
PROTEIN_DIGESTION_AND_ABSORPTION
CUSHING'S_SYNDROME
CAMP_SIGNALING_PATHWAY
ESTROGEN_SIGNALING_PATHWAY
MORPHINE_ADDICTION
RAP1_SIGNALING_PATHWAY
PROTEOGLYCANS_IN_CANCER
HUMAN_PAPILLOMAVIRUS_INFECTION
BASAL_CELL_CARCINOMA
ECM_RECEPTOR_INTERACTION
DILATED_CARDIOMYOPATHY_DCM
ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC
NICOTINE_ADDICTION
CALCIUM_SIGNALING_PATHWAY
COMPLEMENT_AND_COAGULATION_CASCADES
SIGNALING_PATHWAYS_REGULATING_PLURIPOTENCY_OF_STEM_CELLS
HIPPO_SIGNALING_PATHWAY
FOCAL_ADHESION
AXON_GUIDANCE
TUBERCULOSIS
ANTIGEN_PROCESSING_AND_PRESENTATION
CHEMOKINE_SIGNALING_PATHWAY
SYSTEMIC_LUPUS_ERYTHEMATOSUS
INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION
HEMATOPOIETIC_CELL_LINEAGE
JAK_STAT_SIGNALING_PATHWAY
TYPE_I_DIABETES_MELLITUS
ASTHMA
NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY
GRAFT_VERSUS_HOST_DISEASE
ALLOGRAFT_REJECTION
AUTOIMMUNE_THYROID_DISEASE
PARKINSON'S_DISEASE
OXIDATIVE_PHOSPHORYLATION
VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION
PROTEASOME
AMINOACYL_TRNA_BIOSYNTHESIS
SPLICEOSOME
ALCOHOLISM
RIBOSOME
CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION
WNT_SIGNALING_PATHWAY
HEPATOCELLULAR_CARCINOMA
REGULATION_OF_ACTIN_CYTOSKELETON
CHOLINERGIC_SYNAPSE
RELAXIN_SIGNALING_PATHWAY
ERBB_SIGNALING_PATHWAY
STEROID_HORMONE_BIOSYNTHESIS
RETINOL_METABOLISM
INFLAMMATORY_MEDIATOR_REGULATION_OF_TRP_CHANNELS
GASTRIC_ACID_SECRETION
OXYTOCIN_SIGNALING_PATHWAY
CGMP_PKG_SIGNALING_PATHWAY
ADRENERGIC_SIGNALING_IN_CARDIOMYOCYTES
THYROID_HORMONE_SIGNALING_PATHWAY
PLATELET_ACTIVATION
ALDOSTERONE_SYNTHESIS_AND_SECRETION
GLUTAMATERGIC_SYNAPSE
GABAERGIC_SYNAPSE
CIRCADIAN_ENTRAINMENT
RETROGRADE_ENDOCANNABINOID_SIGNALING
DOPAMINERGIC_SYNAPSE
APELIN_SIGNALING_PATHWAY
LONG_TERM_POTENTIATION
FATTY_ACID_DEGRADATION
GAP_JUNCTION
METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450
DRUG_METABOLISM_CYTOCHROME_P450
CHEMICAL_CARCINOGENESIS
ENDOCYTOSIS
SEROTONERGIC_SYNAPSE
MTOR_SIGNALING_PATHWAY
PANCREATIC_SECRETION
RAS_SIGNALING_PATHWAY
MAPK_SIGNALING_PATHWAY
NEUROTROPHIN_SIGNALING_PATHWAY
AUTOPHAGY_ANIMAL
PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM
INOSITOL_PHOSPHATE_METABOLISM
PROTEIN_PROCESSING_IN_ENDOPLASMIC_RETICULUM
MRNA_SURVEILLANCE_PATHWAY
VIRAL_CARCINOGENESIS
AMPK_SIGNALING_PATHWAY
SPHINGOLIPID_SIGNALING_PATHWAY
INSULIN_SIGNALING_PATHWAY
TIGHT_JUNCTION
ENDOCRINE_AND_OTHER_FACTOR_REGULATED_CALCIUM_REABSORPTION
RENIN_SECRETION
NOTCH_SIGNALING_PATHWAY
CORTISOL_SYNTHESIS_AND_SECRETION
SALIVARY_SECRETION
AMOEBIASIS
GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO_AND_NEOLACTO_SERIES
GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE_DERMATAN_SULFATE
HTLV_I_INFECTION
VIRAL_MYOCARDITIS
OTHER_TYPES_OF_O_GLYCAN_BIOSYNTHESIS
TH17_CELL_DIFFERENTIATION
TH1_AND_TH2_CELL_DIFFERENTIATION
HEPATITIS_C
TNF_SIGNALING_PATHWAY
HEPATITIS_B
OSTEOCLAST_DIFFERENTIATION
NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY
NF_KAPPA_B_SIGNALING_PATHWAY
LEISHMANIASIS
HERPES_SIMPLEX_INFECTION
HEDGEHOG_SIGNALING_PATHWAY
ALDOSTERONE_REGULATED_SODIUM_REABSORPTION
THYROID_HORMONE_SYNTHESIS
BILE_SECRETION
VASCULAR_SMOOTH_MUSCLE_CONTRACTION
REGULATION_OF_LIPOLYSIS_IN_ADIPOCYTES
COCAINE_ADDICTION
HIPPO_SIGNALING_PATHWAY_MULTIPLE_SPECIES
BLADDER_CANCER
RNA_TRANSPORT
LYSINE_DEGRADATION
VASOPRESSIN_REGULATED_WATER_REABSORPTION
GLYCEROLIPID_METABOLISM
SHIGELLOSIS
PANCREATIC_CANCER
GLIOMA
FOXO_SIGNALING_PATHWAY
RENAL_CELL_CARCINOMA
COLORECTAL_CANCER
GLUCAGON_SIGNALING_PATHWAY
CENTRAL_CARBON_METABOLISM_IN_CANCER
HIF_1_SIGNALING_PATHWAY
SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT
FLUID_SHEAR_STRESS_AND_ATHEROSCLEROSIS
CHOLINE_METABOLISM_IN_CANCER
B_CELL_RECEPTOR_SIGNALING_PATHWAY
FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS
INSULIN_RESISTANCE
TYPE_II_DIABETES_MELLITUS
CARBOHYDRATE_DIGESTION_AND_ABSORPTION
NON_SMALL_CELL_LUNG_CANCER
ENDOMETRIAL_CANCER
ACUTE_MYELOID_LEUKEMIA
PROLACTIN_SIGNALING_PATHWAY
AGE_RAGE_SIGNALING_PATHWAY_IN_DIABETIC_COMPLICATIONS
MELANOMA
LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION
SMALL_CELL_LUNG_CANCER
PROSTATE_CANCER
CHRONIC_MYELOID_LEUKEMIA
LONGEVITY_REGULATING_PATHWAY
PARATHYROID_HORMONE_SYNTHESIS_SECRETION_AND_ACTION
BACTERIAL_INVASION_OF_EPITHELIAL_CELLS
PHOSPHOLIPASE_D_SIGNALING_PATHWAY
INSULIN_SECRETION
ADHERENS_JUNCTION
TRANSCRIPTIONAL_MISREGULATION_IN_CANCER
MICRORNAS_IN_CANCER
TGF_BETA_SIGNALING_PATHWAY
PI3K_AKT_SIGNALING_PATHWAY
STAPHYLOCOCCUS_AUREUS_INFECTION
UBIQUITIN_MEDIATED_PROTEOLYSIS
ABC_TRANSPORTERS
PEROXISOME
OOCYTE_MEIOSIS
RIBOSOME_BIOGENESIS_IN_EUKARYOTES
NUCLEOTIDE_EXCISION_REPAIR
DNA_REPLICATION
CITRATE_CYCLE_TCA_CYCLE
VIBRIO_CHOLERAE_INFECTION
GLYCOLYSIS_GLUCONEOGENESIS
RNA_DEGRADATION
EPSTEIN_BARR_VIRUS_INFECTION
MALARIA
INFLAMMATORY_BOWEL_DISEASE_IBD
TASTE_TRANSDUCTION
RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY
IL_17_SIGNALING_PATHWAY
CYTOSOLIC_DNA_SENSING_PATHWAY
RHEUMATOID_ARTHRITIS
KAPOSI'S_SARCOMA_ASSOCIATED_HERPESVIRUS_INFECTION
C_TYPE_LECTIN_RECEPTOR_SIGNALING_PATHWAY
T_CELL_RECEPTOR_SIGNALING_PATHWAY
TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY
MEASLES
INFLUENZA_A
PENTOSE_PHOSPHATE_PATHWAY
GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM
ARGININE_AND_PROLINE_METABOLISM
FRUCTOSE_AND_MANNOSE_METABOLISM
ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM
ONE_CARBON_POOL_BY_FOLATE
GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS
STEROID_BIOSYNTHESIS
PRIMARY_BILE_ACID_BIOSYNTHESIS
TYROSINE_METABOLISM
TERPENOID_BACKBONE_BIOSYNTHESIS
RNA_POLYMERASE
MISMATCH_REPAIR
NECROPTOSIS
TRYPTOPHAN_METABOLISM
BUTANOATE_METABOLISM
PYRUVATE_METABOLISM
PROPANOATE_METABOLISM
BETA_ALANINE_METABOLISM
PYRIMIDINE_METABOLISM
PURINE_METABOLISM
GLUTATHIONE_METABOLISM
PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS
NITROGEN_METABOLISM
TOXOPLASMOSIS
ASCORBATE_AND_ALDARATE_METABOLISM
SYNAPTIC_VESICLE_CYCLE
PHAGOSOME
GNRH_SIGNALING_PATHWAY
FC_EPSILON_RI_SIGNALING_PATHWAY
PROGESTERONE_MEDIATED_OOCYTE_MATURATION
VEGF_SIGNALING_PATHWAY
PATHOGENIC_ESCHERICHIA_COLI_INFECTION
CELLULAR_SENESCENCE
AUTOPHAGY_OTHER
CELL_CYCLE
LONGEVITY_REGULATING_PATHWAY_MULTIPLE_SPECIES
LONG_TERM_DEPRESSION
AMPHETAMINE_ADDICTION
EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTION
COLLECTING_DUCT_ACID_SECRETION
THERMOGENESIS
HUNTINGTON'S_DISEASE
NON_ALCOHOLIC_FATTY_LIVER_DISEASE_NAFLD
ALZHEIMER'S_DISEASE
PPAR_SIGNALING_PATHWAY
GLYCEROPHOSPHOLIPID_METABOLISM
LYSOSOME
DRUG_METABOLISM_OTHER_ENZYMES
MITOPHAGY_ANIMAL
ADIPOCYTOKINE_SIGNALING_PATHWAY
RENIN_ANGIOTENSIN_SYSTEM
FAT_DIGESTION_AND_ABSORPTION
SPHINGOLIPID_METABOLISM
FOLATE_BIOSYNTHESIS
GLYCOSPHINGOLIPID_BIOSYNTHESIS_GLOBO_AND_ISOGLOBO_SERIES
ETHER_LIPID_METABOLISM
APOPTOSIS_MULTIPLE_SPECIES
LINOLEIC_ACID_METABOLISM
GLYCOSAMINOGLYCAN_DEGRADATION
PROXIMAL_TUBULE_BICARBONATE_RECLAMATION
PANTOTHENATE_AND_COA_BIOSYNTHESIS
APOPTOSIS
THYROID_CANCER
CIRCADIAN_RHYTHM
MATURITY_ONSET_DIABETES_OF_THE_YOUNG
FERROPTOSIS
CHOLESTEROL_METABOLISM
THIAMINE_METABOLISM
SELENOCOMPOUND_METABOLISM
NICOTINATE_AND_NICOTINAMIDE_METABOLISM
OTHER_GLYCAN_DEGRADATION
MINERAL_ABSORPTION
VITAMIN_DIGESTION_AND_ABSORPTION
ALPHA_LINOLENIC_ACID_METABOLISM
P53_SIGNALING_PATHWAY
AFRICAN_TRYPANOSOMIASIS
SALMONELLA_INFECTION
PHOTOTRANSDUCTION
LEGIONELLOSIS
CARDIAC_MUSCLE_CONTRACTION
ARACHIDONIC_ACID_METABOLISM
PORPHYRIN_AND_CHLOROPHYLL_METABOLISM
CHAGAS_DISEASE_AMERICAN_TRYPANOSOMIASIS
PRION_DISEASES
OVARIAN_STEROIDOGENESIS
MANNOSE_TYPE_O_GLYCAN_BIOSYNTHESIS
GLYCOSPHINGOLIPID_BIOSYNTHESIS_GANGLIO_SERIES
AMYOTROPHIC_LATERAL_SCLEROSIS_ALS
PRIMARY_IMMUNODEFICIENCY
PERTUSSIS
STARCH_AND_SUCROSE_METABOLISM
N_GLYCAN_BIOSYNTHESIS
PROTEIN_EXPORT
GALACTOSE_METABOLISM
BASAL_TRANSCRIPTION_FACTORS
GLYCINE_SERINE_AND_THREONINE_METABOLISM
CYSTEINE_AND_METHIONINE_METABOLISM
AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM
PHENYLALANINE_METABOLISM
HISTIDINE_METABOLISM
ARGININE_BIOSYNTHESIS
FATTY_ACID_ELONGATION
BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS
HOMOLOGOUS_RECOMBINATION
BASE_EXCISION_REPAIR
FANCONI_ANEMIA_PATHWAY
ES z–score
20
15
10
5
0
−5

Figure S5 Compressed combination heatmap

This heatmap shows the ES z-scores of 298 KEGG pathways (rows) in 6 representative parameter combinations out of 126 tested combinations of the NPP parameters (columns). The top bar plot represents the number of pathways in which each combination cluster was the best as compared to all other selected combinations.

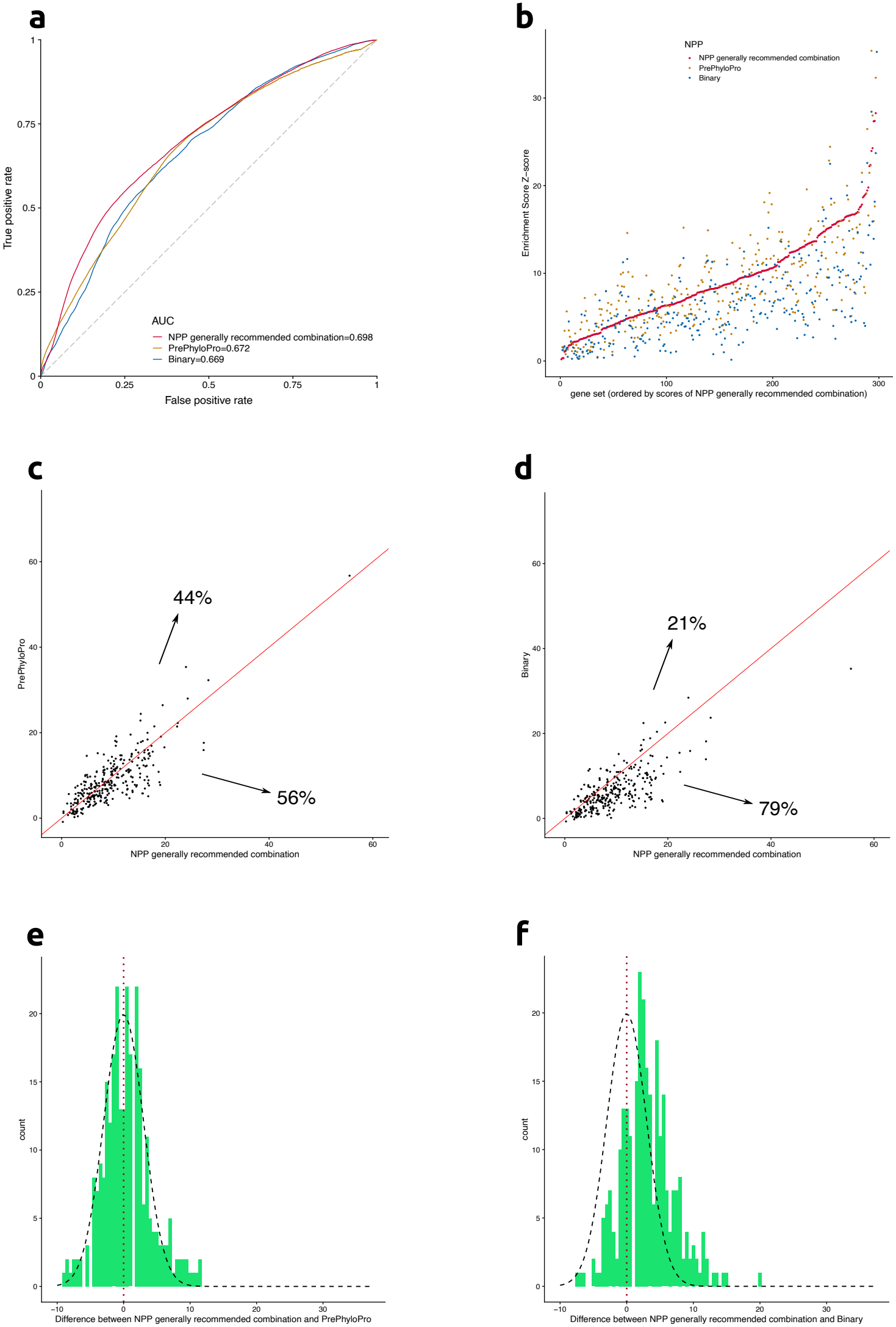

a
True positive rate
False positive rate
AUC
NPP generally recommended combination=0.698
PrePhyloPro=0.672
Binary=0.669
b
NPP
NPP generally recommended combination
PrePhyloPro
Binary
Enrichment Score Z−score
gene set (ordered by scores of NPP generally recommended combination)
c
44%
56%
PrePhyloPro
NPP generally recommended combination
d
21%
79%
Binary
NPP generally recommended combination
e
count
Difference between NPP generally recommended combination and PrePhyloPro
f
count
Difference between NPP generally recommended combination and Binary

Figure S6 Comparison of NPP to PrePhyloPro and Binary PP

a. ROC curve analysis of NPP generally recommended combination vs PrePhyloPro and "Binary Hamming". The correlations or distances for gene pairs from CORUM (positive) were compared to control (negative) for each method. The X and Y axes represent the False Positive Rate (FPR = 1 − specificity) and True Positive Rate (TPR, sensitivity), respectively. **b.** ES z-score values of NPP generally recommended combination, PrePhyloPro and "Binary Hamming" for 298 KEGG pathways, ordered by the values of NPP generally recommended combination. **c.** Scatter-plot of ES z-score values of NPP generally recommended combination vs "PrePhyloPro". **d.** Scatter-plot of ES z-score values of NPP generally recommended combination vs "Binary Hamming". **e.** Distribution of the differences between ES z-score values of NPP generally recommended combination and "PrePhyloPro". The black dashed line demonstrates a normal distribution curve for reference. **f.** Distribution of the differences between ES z-score values of NPP generally recommended combination and "Binary Hamming". The black dashed line demonstrates a normal distribution curve for reference.

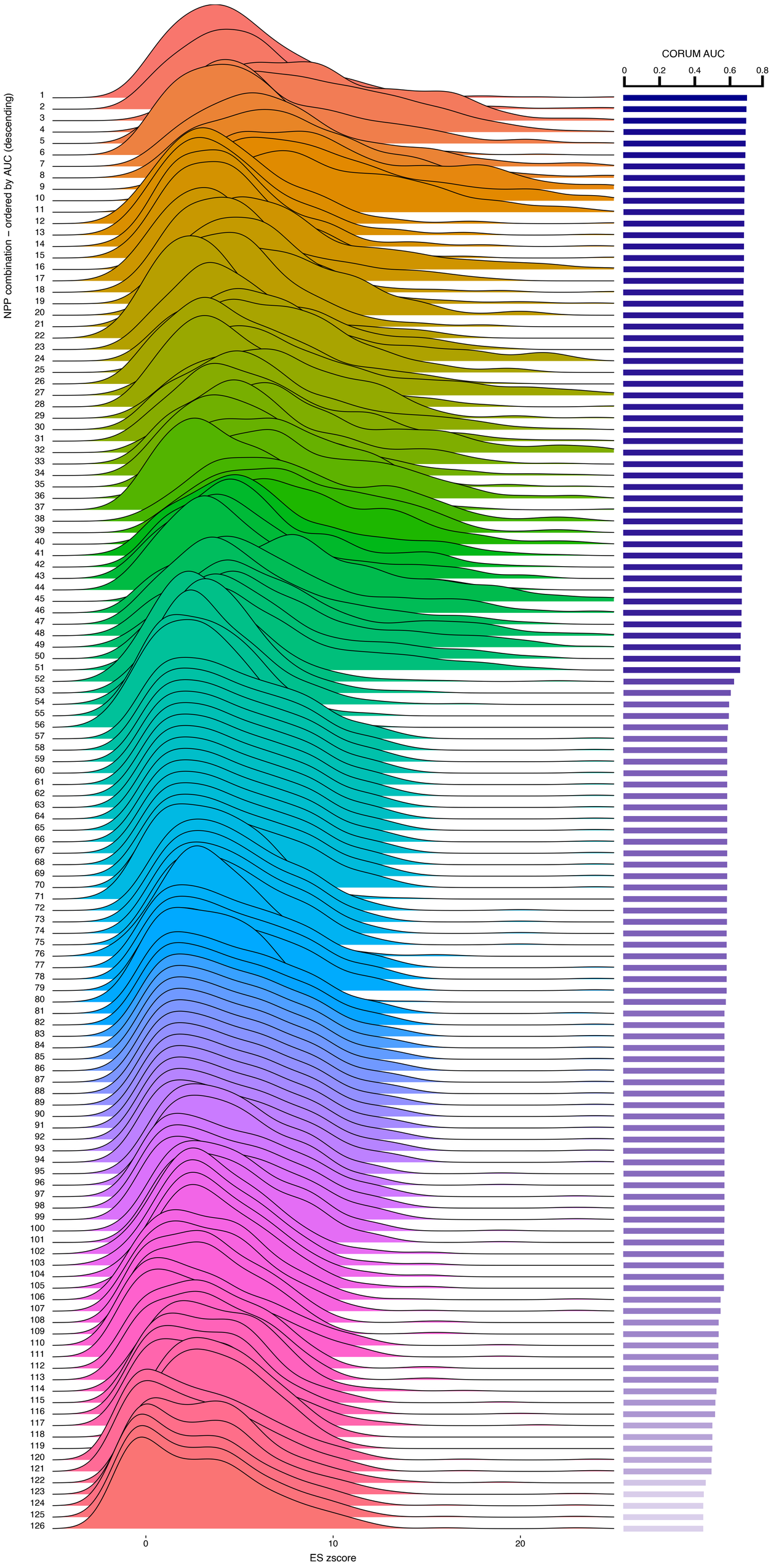

CORUM AUC
0
0.2
0.4
0.6
0.8
NPP combination – ordered by AUC (descending)
ES zscore
0
10
20

Figure S7 Distributions of KEGG ES z-scores by NPP combination

The density ridges demonstrate the distributions for the KEGG ES z-scores of the different combination ordered by the CORUM AUC scores. The bar plot on the right presents the AUC value for each combination. The figure shows the general similarity between the two methods.