

Supplementary Table S6. GO terms and KEGG pathways under and over-represented in disease vs. health. We used the R package 'goseq' to identify significantly over and under-represented terms in our results. We used a cutoff value of False Discovery Rate (FDR) of <0.05.

GO Number	GO categories	p-value	FDR
<b>Over represented</b>			
GO:0003735	structural constituent of ribosome	<0.00001	<0.00001
GO:0005840	ribosome	<0.00001	<0.00001
GO:0006412	translation	<0.00001	<0.00001
GO:0005622	intracellular	<0.00001	<0.00001
GO:0003723	RNA binding	<0.00001	<0.00001
GO:0046933	hydrogen ion transporting ATP synthase activity rotational mechanism	<0.00001	<0.00001
GO:0005525	GTP binding	<0.00001	<0.00001
GO:0030699	glycine reductase activity	<0.00001	<0.00001
GO:0046961	hydrogen ion transporting ATPase activity rotational mechanism	<0.00001	<0.00001
GO:0009055	electron carrier activity	<0.00001	<0.00001
GO:0006118	electron transport	<0.00001	<0.00001
GO:0048037	cofactor binding	<0.00001	<0.00001
GO:0004365	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) activity	<0.00001	<0.00001
GO:0004634	phosphopyruvate hydratase activity	<0.00001	<0.00001
GO:0003743	translation initiation factor activity	<0.00001	<0.00001
GO:0005506	iron ion binding	<0.00001	<0.00001
GO:0004784	superoxide dismutase activity	<0.00001	<0.00001
GO:0047553	2-oxoglutarate synthase activity	<0.00001	<0.00001
GO:0006413	translational initiation	<0.00001	<0.00001
GO:0010181	FMN binding	<0.00001	<0.00001
GO:0042254	ribosome biogenesis and assembly	<0.00001	<0.00001
GO:0006457	protein folding	<0.00001	<0.00001
GO:0003899	DNA-directed RNA polymerase activity	<0.00001	<0.00001
GO:0016966	nitric oxide reductase activity	<0.00001	<0.00001
GO:0030060	L-malate dehydrogenase activity	<0.00001	<0.00001
GO:0009306	protein secretion	<0.00001	<0.00001
GO:0006520	amino acid metabolic process	<0.00001	<0.00001
GO:0004315	3-oxoacyl-[acyl-carrier-protein] synthase activity	<0.00001	<0.00001
GO:0008565	protein transporter activity	<0.00001	<0.00001
GO:0004332	fructose-bisphosphate aldolase activity	<0.00001	<0.00001
GO:0004494	methylmalonyl-CoA mutase activity	<0.00001	<0.00001
GO:0015450	P-P-bond-hydrolysis-driven protein transmembrane transporter activity	<0.00001	<0.00001
GO:0003697	single-stranded DNA binding	<0.00001	<0.00001

Supplementary Table S6. GO terms and KEGG pathways under and over-represented in disease vs. health. We used the R package 'goseq' to identify significantly over and under-represented terms in our results. We used a cutoff value of False Discovery Rate (FDR) of <0.05.

GO Number	GO categories	p-value	FDR
<b>Over represented</b>			
GO:0008462	endopeptidase Clp activity	<0.00001	<0.00001
GO:0004322	ferroxidase activity	<0.00001	<0.00001
GO:0003711	transcription elongation regulator activity	<0.00001	<0.00001
GO:0030976	thiamin pyrophosphate binding	<0.00001	<0.00001
GO:0005515	protein binding	<0.00001	<0.00001
GO:0003746	translation elongation factor activity	<0.00001	<0.00001
GO:0047334	diphosphate-fructose-6-phosphate 1-phosphotransferase activity	<0.00001	<0.00001
GO:0015934	large ribosomal subunit	<0.00001	<0.00001
GO:0004492	methylmalonyl-CoA decarboxylase activity	<0.00001	<0.00001
GO:0050242	pyruvate phosphate dikinase activity	<0.00001	<0.00001
GO:0008097	5S rRNA binding	<0.00001	<0.00001
GO:0008121	ubiquinol-cytochrome-c reductase activity	<0.00001	<0.00001
GO:0006605	protein targeting	<0.00001	<0.00001
GO:0031072	heat shock protein binding	<0.00001	<0.00001
GO:0009245	lipid A biosynthetic process	<0.00001	<0.00001
GO:0006352	transcription initiation	<0.00001	<0.00001
GO:0016987	sigma factor activity	<0.00001	<0.00001
GO:0018522	benzoyl-CoA reductase activity	<0.00001	<0.00001
GO:0008780	acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase activity	<0.00001	<0.00001
GO:0008137	NADH dehydrogenase (ubiquinone) activity	<0.00001	<0.00001
GO:0004781	sulfate adenylyltransferase (ATP) activity	<0.00001	<0.00001
GO:0008940	nitrate reductase activity	<0.00001	<0.00001
GO:0008890	glycine C-acetyltransferase activity	<0.00001	<0.00001
GO:0044267	cellular protein metabolic process	<0.00001	<0.00001
GO:0050002	D-proline reductase (dithiol) activity	<0.00001	<0.00001
GO:0001882	nucleoside binding	<0.00001	<0.00001
GO:0006886	intracellular protein transport	<0.00001	<0.00001
GO:0003715	transcription termination factor activity	<0.00001	<0.00001
GO:0016852	sirohydrochlorin cobaltochelataase activity	<0.00001	<0.00001
GO:0042279	nitrite reductase (cytochrome ammonia-forming) activity	<0.00001	<0.00001
GO:0019538	protein metabolic process	<0.00001	<0.00001
GO:0003963	RNA-3'-phosphate cyclase activity	<0.00001	<0.00001
GO:0047850	diaminopimelate dehydrogenase activity	<0.00001	<0.00001

Supplementary Table S6. GO terms and KEGG pathways under and over-represented in disease vs. health. We used the R package 'goseq' to identify significantly over and under-represented terms in our results. We used a cutoff value of False Discovery Rate (FDR) of <0.05.

GO Number	GO categories	p-value	FDR
<b>Over represented</b>			
GO:0003824	catalytic activity	<0.00001	<0.00001
GO:0047702	beta-lysine 5 6-aminomutase activity	<0.00001	<0.00001
GO:0004612	phosphoenolpyruvate carboxykinase (ATP) activity	<0.00001	<0.00001
GO:0016829	lyase activity	<0.00001	<0.00001
GO:0008861	formate C-acetyltransferase activity	<0.00001	<0.00001
GO:0004375	glycine dehydrogenase (decarboxylating) activity	<0.00001	<0.00001
GO:0004473	malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) activity	<0.00001	<0.00001
GO:0006414	translational elongation	<0.00001	<0.00001
GO:0051082	unfolded protein binding	<0.00001	<0.00001
GO:0006353	transcription termination	<0.00001	<0.00001
GO:0006879	cellular iron ion homeostasis	<0.00001	<0.00001
GO:0005315	inorganic phosphate transmembrane transporter activity	<0.00001	<0.00001
GO:0016491	oxidoreductase activity	<0.00001	<0.00001
GO:0047330	polyphosphate-glucose phosphotransferase activity	<0.00001	<0.00001
GO:0003979	UDP-glucose 6-dehydrogenase activity	<0.00001	<0.00001
GO:0019867	outer membrane	<0.00001	<0.00001
GO:0016071	mRNA metabolic process	<0.00001	<0.00001
GO:0016469	proton-transporting two-sector ATPase complex	<0.00001	<0.00001
GO:0050660	FAD binding	<0.00001	<0.00001
GO:0043064	flagellum organization and biogenesis	<0.00001	<0.00001
GO:0003918	DNA topoisomerase (ATP-hydrolyzing) activity	<0.00001	<0.00001
GO:0015986	ATP synthesis coupled proton transport	<0.00001	<0.00001
GO:0008959	phosphate acetyltransferase activity	<0.00001	<0.00001
GO:0004239	methionyl aminopeptidase activity	<0.00001	<0.00001
GO:0004807	triose-phosphate isomerase activity	<0.00001	<0.00001
GO:0015093	ferrous iron transmembrane transporter activity	<0.00001	<0.00001
GO:0015684	ferrous iron transport	<0.00001	<0.00001
GO:0004618	phosphoglycerate kinase activity	<0.00001	<0.00001
GO:0050304	nitrous-oxide reductase activity	<0.00001	0.01
GO:0004658	propionyl-CoA carboxylase activity	<0.00001	0.01
GO:0000774	adenyl-nucleotide exchange factor activity	<0.00001	0.01
GO:0004746	riboflavin synthase activity	<0.00001	0.01
GO:0006857	oligopeptide transport	<0.00001	0.01

Supplementary Table S6. GO terms and KEGG pathways under and over-represented in disease vs. health. We used the R package 'goseq' to identify significantly over and under-represented terms in our results. We used a cutoff value of False Discovery Rate (FDR) of <0.05.

GO Number	GO categories	p-value	FDR
<b>Over represented</b>			
GO:0004655	porphobilinogen synthase activity	<0.00001	0.01
GO:0006826	iron ion transport	<0.00001	0.01
GO:0005198	structural molecule activity	<0.00001	0.01
GO:0042286	glutamate-1-semialdehyde 2 1-aminomutase activity	<0.00001	0.01
GO:0015129	lactate transmembrane transporter activity	<0.00001	0.01
GO:0008415	acyltransferase activity	<0.00001	0.01
GO:0008928	mannose-1-phosphate guanylyltransferase (GDP) activity	<0.00001	0.01
GO:0047134	protein-disulfide reductase activity	<0.00001	0.01
GO:0001539	ciliary or flagellar motility	<0.00001	0.01
GO:0006928	cell motility	<0.00001	0.01
GO:0020037	heme binding	<0.00001	0.01
GO:0005960	glycine cleavage complex	<0.00001	0.01
GO:0006935	chemotaxis	<0.00001	0.01
GO:0000104	succinate dehydrogenase activity	<0.00001	0.01
GO:0008667	2 3-dihydro-2 3-dihydroxybenzoate dehydrogenase activity	<0.00001	0.01
GO:0003842	1-pyrroline-5-carboxylate dehydrogenase activity	<0.00001	0.02
GO:0004654	polyribonucleotide nucleotidyltransferase activity	<0.00001	0.02
GO:0006817	phosphate transport	<0.00001	0.02
GO:0003989	acetyl-CoA carboxylase activity	<0.00001	0.02
GO:0008446	GDP-mannose 4 6-dehydratase activity	<0.00001	0.02
GO:0030529	ribonucleoprotein complex	<0.00001	0.02
GO:0004352	glutamate dehydrogenase activity	<0.00001	0.03
GO:0051087	chaperone binding	<0.00001	0.03
GO:0005216	ion channel activity	<0.00001	0.03
GO:0008797	aspartate ammonia-lyase activity	<0.00001	0.03
GO:0004084	branched-chain-amino-acid transaminase activity	<0.00001	0.03
GO:0004739	pyruvate dehydrogenase (acetyl-transferring) activity	<0.00001	0.03
GO:0017165	dipeptidase E activity	<0.00001	0.03
GO:0004035	alkaline phosphatase activity	<0.00001	0.03
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	<0.00001	0.03
GO:0030655	beta-lactam antibiotic catabolic process	<0.00001	0.04
GO:0008893	guanosine-3' 5'-bis(diphosphate) 3'-diphosphatase activity	<0.00001	0.04
GO:0004308	exo-alpha-sialidase activity	<0.00001	0.04

Supplementary Table S6. GO terms and KEGG pathways under and over-represented in disease vs. health. We used the R package 'goseq' to identify significantly over and under-represented terms in our results. We used a cutoff value of False Discovery Rate (FDR) of <0.05.

GO Number	GO categories	p-value	FDR
<b>Over represented</b>			
GO:0010033	response to organic substance	<0.00001	0.04
GO:0015031	protein transport	<0.00001	0.04
GO:0008998	ribonucleoside-triphosphate reductase activity	<0.00001	0.04
GO:0016740	transferase activity	<0.00001	0.04
GO:0004085	butyryl-CoA dehydrogenase activity	<0.00001	0.04
GO:0004148	dihydrolipoyl dehydrogenase activity	<0.00001	0.04
GO:0051745	4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase activity	<0.00001	0.04
GO:0003935	GTP cyclohydrolase II activity	<0.00001	0.05
GO:0008817	cob(I)yrinic acid a c-diamide adenosyltransferase activity	<0.00001	0.05
GO:0008759	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase activity	<0.00001	0.05
GO:0008829	dCTP deaminase activity	<0.00001	0.05
GO:0003747	translation release factor activity	0.01	0.05
GO:0006415	translational termination	0.01	0.05
GO:0003882	CDP-diacylglycerol-serine O-phosphatidyltransferase activity	0.01	0.05
GO Number	GO categories	p-value	FDR
<b>Under represented</b>			
GO:0005886	plasma membrane	<0.00001	<0.00001
GO:0008982	protein-N(PI)-phosphohistidine-sugar phosphotransferase activity	<0.00001	<0.00001
GO:0040007	growth	<0.00001	<0.00001
GO:0005618	cell wall	<0.00001	<0.00001
GO:0047473	D-alanine-poly(phosphoribitol) ligase activity	<0.00001	<0.00001
GO:0006813	potassium ion transport	<0.00001	<0.00001
GO:0004300	enoyl-CoA hydratase activity	<0.00001	<0.00001
GO:0008422	beta-glucosidase activity	<0.00001	<0.00001
GO:0017103	UTP:galactose-1-phosphate uridylyltransferase activity	<0.00001	<0.00001
GO:0004856	xylulokinase activity	<0.00001	<0.00001
GO:0047265	poly(glycerol-phosphate) alpha-glucosyltransferase activity	<0.00001	<0.00001
GO:0008977	prephenate dehydrogenase activity	<0.00001	<0.00001
GO:0004357	glutamate-cysteine ligase activity	<0.00001	<0.00001
GO:0005829	cytosol	<0.00001	<0.00001
GO:0003887	DNA-directed DNA polymerase activity	<0.00001	<0.00001
GO:0003985	acetyl-CoA C-acetyltransferase activity	<0.00001	<0.00001

Supplementary Table S6. GO terms and KEGG pathways under and over-represented in disease vs. health. We used the R package 'goseq' to identify significantly over and under-represented terms in our results. We used a cutoff value of False Discovery Rate (FDR) of <0.05.

<b>GO Number</b>	<b>GO categories</b>	<b>p-value</b>	<b>FDR</b>
<b>Under represented</b>			
GO:0015079	potassium ion transmembrane transporter activity	<0.00001	0.01
GO:0007049	cell cycle	<0.00001	0.01
GO:0009384	N-acylmannosamine kinase activity	<0.00001	0.01
GO:0046316	gluconokinase activity	<0.00001	0.01
GO:0000271	polysaccharide biosynthetic process	<0.00001	0.01
GO:0008728	GTP diphosphokinase activity	<0.00001	0.01
GO:0004558	alpha-glucosidase activity	<0.00001	0.01
GO:0009024	tagatose-6-phosphate kinase activity	<0.00001	0.01
GO:0000166	nucleotide binding	<0.00001	0.02
GO:0050118	N-acetyldiaminopimelate deacetylase activity	<0.00001	0.02
GO:0004003	ATP-dependent DNA helicase activity	<0.00001	0.02
GO:0003988	acetyl-CoA C-acyltransferase activity	<0.00001	0.02
GO:0050334	thiaminase activity	<0.00001	0.02
GO:0008477	purine nucleosidase activity	<0.00001	0.02
GO:0043565	sequence-specific DNA binding	<0.00001	0.03
GO:0004672	protein kinase activity	<0.00001	0.03
GO:0009044	xylan 1 4-beta-xylosidase activity	<0.00001	0.03
GO:0005975	carbohydrate metabolic process	<0.00001	0.03
GO:0006109	regulation of carbohydrate metabolic process	<0.00001	0.03
GO:0004777	succinate-semialdehyde dehydrogenase activity	<0.00001	0.03
GO:0008973	phosphopentomutase activity	<0.00001	0.03
GO:0008908	isochorismatase activity	<0.00001	0.04
GO:0008745	N-acetylmuramoyl-L-alanine amidase activity	<0.00001	0.04
GO:0008448	N-acetylglucosamine-6-phosphate deacetylase activity	<0.00001	0.04
GO:0004788	thiamin diphosphokinase activity	<0.00001	0.04
GO:0017057	6-phosphogluconolactonase activity	<0.00001	0.04
<b>KEGG pathway number</b>			
<b>Over represented</b>			
ko00190	Oxidative phosphorylation	<0.00001	<0.00001
ko00195	Photosynthesis	<0.00001	<0.00001
ko00020	Citrate cycle (TCA cycle)	<0.00001	<0.00001
ko00720	Reductive carboxylate cycle (CO2 fixation)	<0.00001	<0.00001

Supplementary Table S6. GO terms and KEGG pathways under and over-represented in disease vs. health. We used the R package 'goseq' to identify significantly over and under-represented terms in our results. We used a cutoff value of False Discovery Rate (FDR) of <0.05.

KEGG pathway number	KEGG pathway	p-value	FDR
<b>Over represented</b>			
ko00710	Carbon fixation in photosynthetic organisms	<0.00001	<0.00001
ko00910	Nitrogen metabolism	<0.00001	<0.00001
ko00620	Pyruvate metabolism	<0.00001	<0.00001
ko00061	Fatty acid biosynthesis	<0.00001	<0.00001
ko00290	Valine leucine and isoleucine biosynthesis	<0.00001	<0.00001
ko00740	Riboflavin metabolism	<0.00001	<0.00001
ko20035	Bacterial motility proteins	<0.00001	<0.00001
ko20030	Bacterial chemotaxis	<0.00001	<0.00001
ko00860	Porphyrin and chlorophyll metabolism	<0.00001	<0.00001
ko00970	Aminoacyl-tRNA biosynthesis	<0.00001	0.03
<b>Under represented</b>			
ko00053	Ascorbate and aldarate metabolism	<0.00001	<0.00001
ko00052	Galactose metabolism	<0.00001	<0.00001
ko00281	Geraniol degradation	<0.00001	<0.00001
ko00062	Fatty acid elongation in mitochondria	<0.00001	<0.00001
ko00500	Starch and sucrose metabolism	<0.00001	<0.00001
ko00905	Brassinosteroid biosynthesis	<0.00001	<0.00001
ko00401	Novobiocin biosynthesis	<0.00001	<0.00001
ko00460	Cyanoamino acid metabolism	<0.00001	<0.00001
ko00360	Phenylalanine metabolism	<0.00001	<0.00001
ko00943	Isoflavonoid biosynthesis	<0.00001	<0.00001
ko00071	Fatty acid metabolism	<0.00001	0.01
ko00561	Glycerolipid metabolism	<0.00001	0.01
ko00960	Tropane piperidine and pyridine alkaloid biosynthesis	<0.00001	0.01
ko00311	Penicillin and cephalosporin biosynthesis	<0.00001	0.02
ko00522	Biosynthesis of 12- 14- and 16-membered macrolides	<0.00001	0.02
ko00380	Tryptophan metabolism	<0.00001	0.02
ko00051	Fructose and mannose metabolism	<0.00001	0.02
ko00440	Phosphonate and phosphinate metabolism	<0.00001	0.03