SUPPLEMENT MATERIAL

Regenerative Potential of Neonatal Porcine Hearts

Running title: Cardiomyocyte Regeneration in Neonatal Pigs

Wuqiang Zhu[1,#], MD, PhD, Eric Zhang[1], BS, Meng Zhao[1] , MD, Zechen Chong[2], PhD, Chengming Fan[1], MD, Yawen Tang[1], BS, Jervaughn D. Hunter[1], BS, Anton V. Borovjagin[1],PhD, Greggory P. Walcott[1], MD, Jake Y. Chen[3], PhD, Gangjian Qin[1], MD, Jianyi Zhang[1,#], MD, PhD

1. School of Medicine and School of Engineering, the University of Alabama at Birmingham, Birmingham, AL 35294, USA
2. Department of Genetics, School of Medicine, University of Alabama at Birmingham, AL 35294, USA
3. Informatics Institute, School of Medicine, University of Alabama at Birmingham, AL 35294, USA

#, Corresponding author

Jianyi (Jay) Zhang, MD, PhD
Department of Biomedical Engineering
School of Medicine and School of Engineering
The University of Alabama at Birmingham, 1720 2nd Ave S, VH G094J
Birmingham, AL 35294-0019
Telephone: 205-934-8421
E-mail: jayzhang@uab.edu

Or

Wuqiang Zhu, MD, PhD
Department of Biomedical Engineering

School of Medicine, School of Engineering
The University of Alabama at Birmingham
1720 2nd Ave S, VH G094E, Birmingham, AL 35294-0019
Tel: 205-934-0228
Fax: 205-934-9101
E-mail: wzhu@uab.edu

SUPPLEMENTAL METHODS

RNA Sequencing

The Pig (*Sus Scrofa*) reference (GenBank Assembly ID GCA_000003025.6) assembly was downloaded from Ensembl database.[1] Non-chromosome DNA sequences were not used. The unmasked genomic DNA sequences were concatenated into a single fasta file. The reference genome sequence was indexed using STAR[2] v2.5.3a. The annotation files were also downloaded from Ensembl. Only the chromosome DNA annotation was used.

The quality control of raw sequencing data was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) v0.11.5. All the sequenced reads passed the Basic Statistics, Per base sequence quality, Per tile sequence quality, Per sequence quality scores, Per base N content, and Adapter Content. The high-quality RNA sequencing experiment was confirmed.

For each sample, paired-ends RNA-seq reads were mapped to pig reference genome using STAR v2.5.3a with the parameters "--outSAMtype BAM SortedByCoordinate Unsorted --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 20 --alignIntronMax 100000 --alignMatesGapMax 100000". SAMtools[3] v0.1.18 was used to further process the alignment BAM file. Htseq-count[4] v0.7.2 was used to count reads in gene features.

DESeq2[5] was used to identify differentially expressed genes (DEGs) between groups. Specifically, we calculated the DEGs in three comparisons: P1 vs P3, P1 vs P14, and P3 vs P14. Gene Ontology (GO) and Pathway enrichment analysis was performed using DAVID[6] v6.8.

To identify candidate genes that are involved in the regeneration process after myocardium injury, we deep sequenced the whole transcriptome of P1, P3, and P14 samples, respectively. A total of 47.9G base pairs were generated (~6.8Gb per sample). The Quality Control (QC) of the raw sequencing data indicated all sequencing experiments were performed appropriately **(Supplementary Table II, and Supplemental Figure I A – I F)**. The reads were mapped to the reference genome using STAR.[2] The average uniquely mapped reads ratio is 85.2% (ranging from 76.07% to 90.3%), indicating a high mapping rate since we only mapped to the reference excluding non-chromosome DNA sequences.

We used htseq-count[3] to count reads in features. We used default method of DESeq2[5] to identify differentially expressed genes (DEGs) between groups. The differential expression analysis in DESeq2 used generalized linear model fitting based on a negative binomial distribution and the Wald significance test to calculate p-values; then, a multiple-test correction (FDR method of Benjamini-Hochberg procedure) was used to maximize the number of adjusted p-values that were less than a given critical value (alpha). For alpha (adjusted p-value) <0.05, we found 734, 258, and 364 significant DEGs between P1 and P3, between P1 and P14, and between P3 and P14,

respectively. To identify the relevant biological functions of these DEGs, we performed Gene Ontology enrichment analysis **(Supplementary Figures I A, I C, and I E)** using DAVID[6] tool. We also performed pathway enrichment analysis **(Supplementary Figure I B, I D, and I F)**. p-values from the enrichment analysis were corrected for multiple testing using Benjamini-Hochberg method.

Mapping Pig Differentially Expressed Transcripts

The Ensemble gene ID for Sus scrofa (pig) are mapped to their gene symbols using the UniProt database as the following: 1) map Ensembl Gene ID directly to pig gene symbols, then further mapped to its human gene counterparts; 2) for the unmapped genes, we manually examined the UniProt homologous gene clusters at the 100%, 90%, and 50% identity levels to map the pig gene to its homologous human gene counterparts within the same gene cluster. After such gene processing, we obtained the following final results:

- 706 out of the 734 differentially genes between P1 and P3 conditions (adjusted p-value<0.05) are mapped to human genes. The top 25 genes that are differentially expressed were listed in Supplemental Table III.
- 205 out of the 258 differentially genes between P1 and P14 conditions (adjusted p-value<0.05) are mapped to human genes. The top 25 genes that are differentially expressed were listed in Supplemental Table IV.
- 348 out of the 364 differentially genes between P3 and P14 conditions (adjusted p-value<0.05) are mapped to human genes. The top 25 genes that are differentially expressed were listed in Supplemental Table V.

Method for Ranking Genes

The ranking of differentially expressed genes was performed adapted from a well-established network biology based gene ranking method originally described in[7]. The method depends on the use of disease-specific human protein-protein interactions (PPIs) with overall good data coverage and overall good quality. For this work, we choose the HAPPI database[8 9], currently in version 2.0, which comprehensively compiled PPI data from sources including bioGRID[10] and STRING[11] with quantitative PPI data confidence information. To assign weights to genes from a particular biological condition, we choose a weight score r_p for each gene, calculated using $r_p = k \ln(\sum_{q \in NET} conf(p,g)) - \ln(\sum_{q \in NET} N(p,q))$ where *p* and *q* are indices for proteins in the disease gene/protein association network constructed from one-layer neighborhood node expansions into the HAPPI PPI database (quality threshold >=0.75), *k* is an empirical constant (*k*=2 in this study), *conf(p, q)* is the confidence score assigned to each interaction between protein *p* and *q*, and *N(p, q)* holds the value of 1 if the protein *p* interacts with *q*.

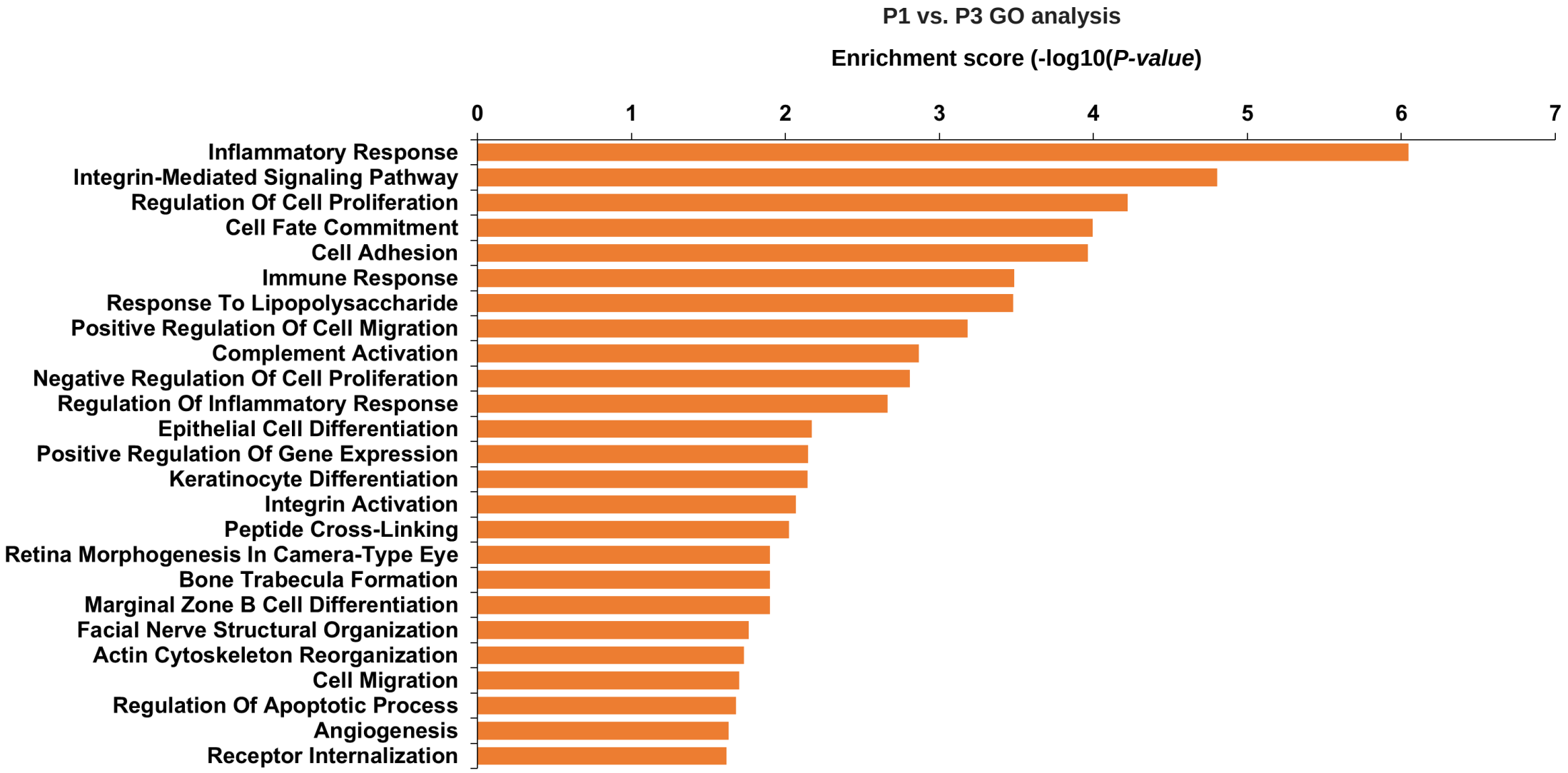
SUPPLEMENTAL REFERENCES

1. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A and Flicek P. Ensembl 2018. *Nucleic Acids Res*. 2018;46:D754-D761.
2. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15-21.
3. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078-9.
4. Anders S, Pyl PT and Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166-9.
5. Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
6. Huang da W, Sherman BT and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1-13.
7. Chen JY, Shen C and Sivachenko AY. Mining Alzheimer disease relevant proteins from integrated protein interactome data. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*. 2006:367-78.
8. Chen JY, Mamidipalli S and Huan T. HAPPI: an online database of comprehensive human annotated and predicted protein interactions. *BMC genomics*. 2009;10 Suppl 1:S16.
9. Chen JY, Pandey R and Nguyen TM. HAPPI-2: a Comprehensive and High-quality Map of Human Annotated and Predicted Protein Interactions. *BMC genomics*. 2017;18:182.
10. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M, Oughtred R, Dolinski K and Tyers M. The BioGRID interaction database: 2013 update. *Nucleic acids research*. 2013;41:D816-23.
11. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ and von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*. 2011;39:D561-8.

SUPPLEMENTAL FIGURE LEGENDS

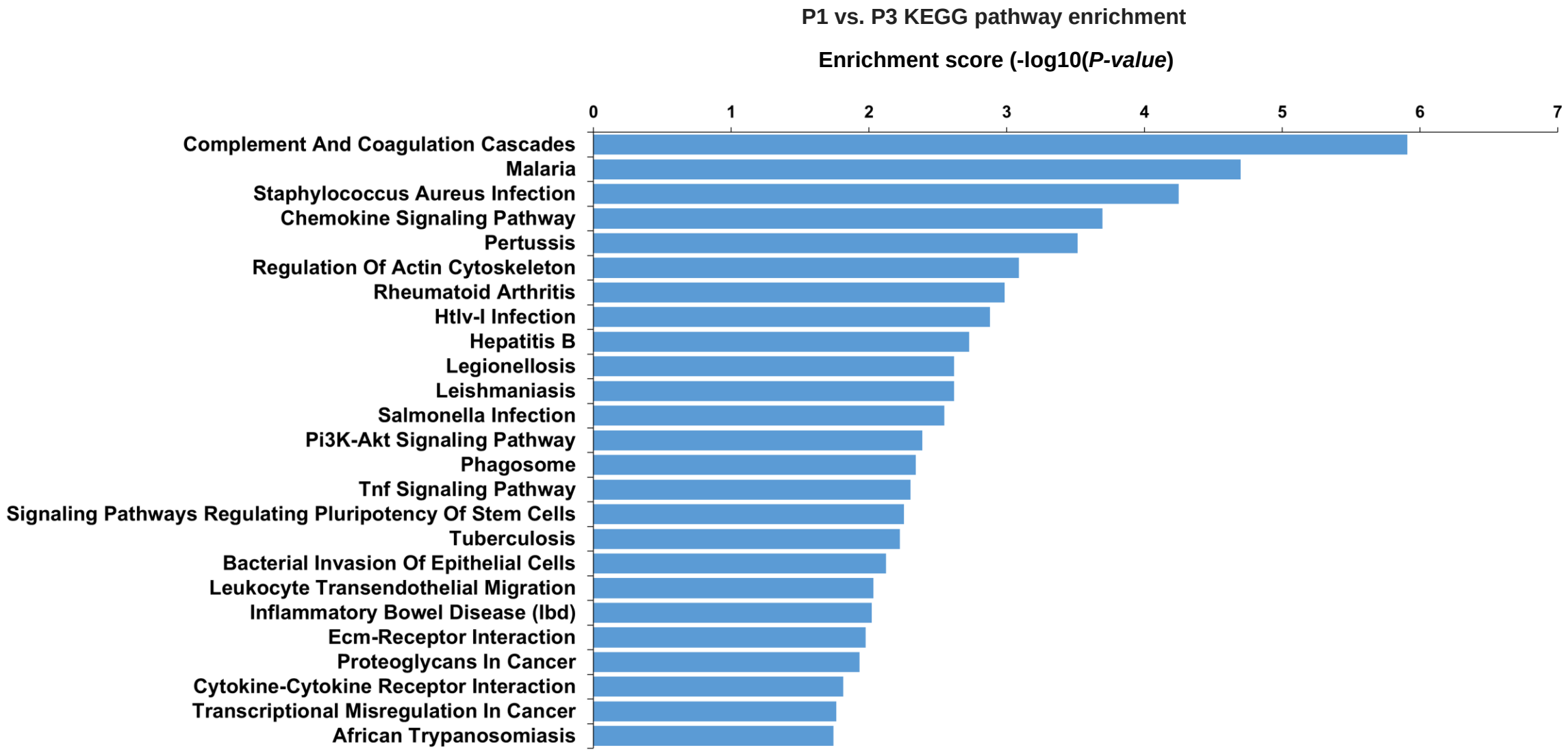
Supplemental Figure I. Gene ontology (GO) analysis and pathway analysis for the differentially expressed genes between P1, P3, and P14. GO analysis according to biological processes for (A) P1 vs. P3, (C) P1 vs. P14, and (E) P3 vs. P14. Pathway analysis for gene enrichment for (B) P1 vs. P3, (D) P1 vs. P14, and (F) P3 vs. P14. Genes involved in inflammatory and immune response, cell proliferation, apoptosis, and angiogenesis were highly enriched.

A.

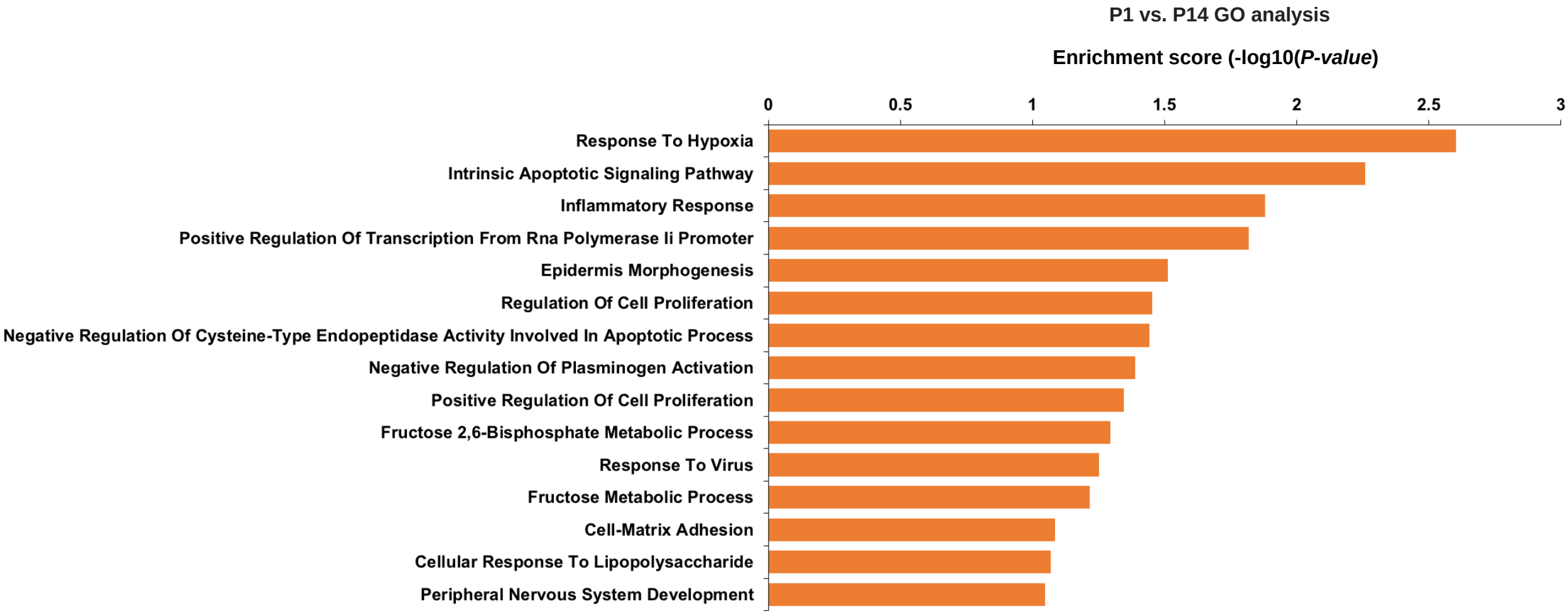


Supplementary Figure I.

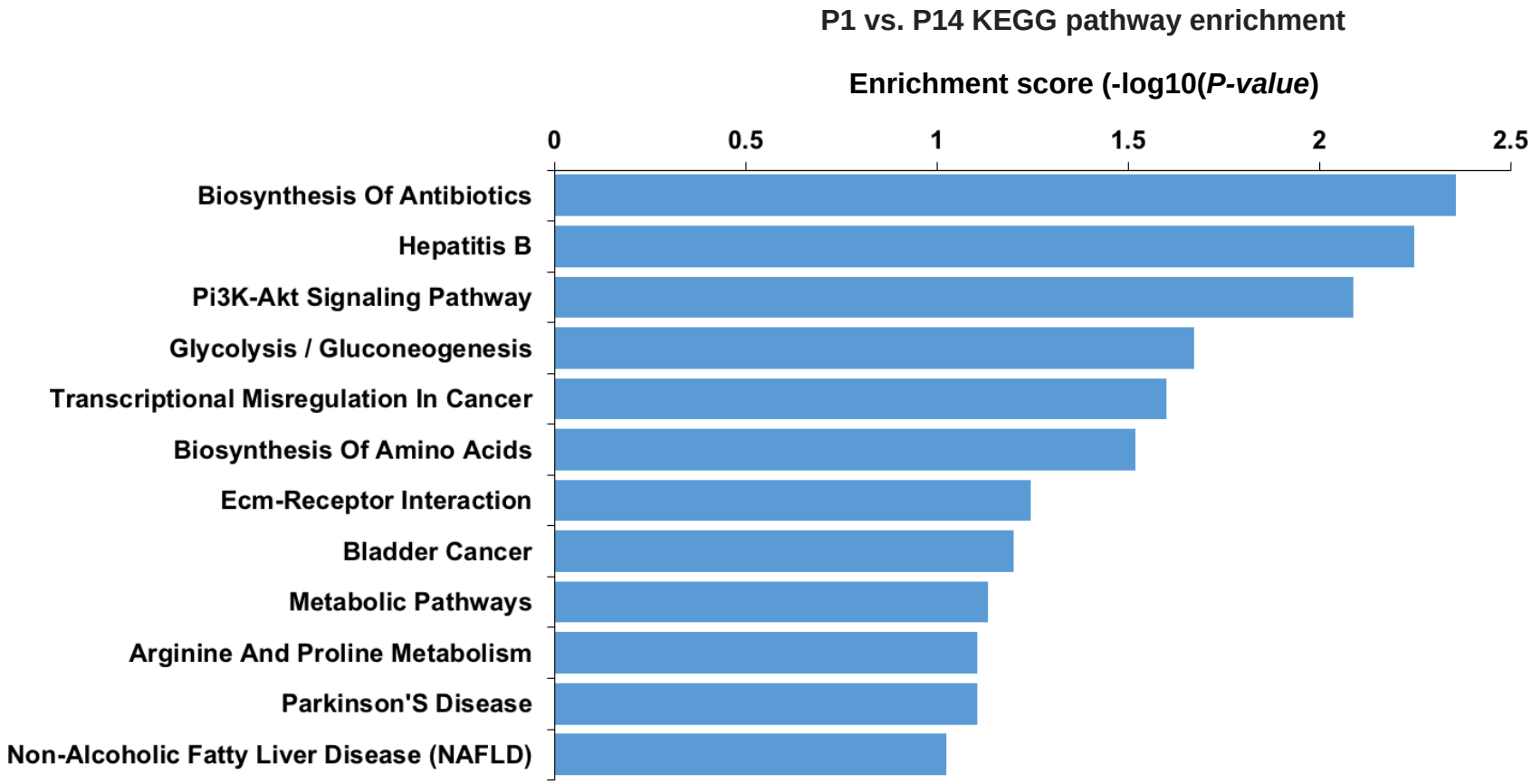
B.


P1 vs. P3 KEGG pathway enrichment
Enrichment score (-log10(*P-value*))
0
1
2
3
4
5
6
7
Complement And Coagulation Cascades
Malaria
Staphylococcus Aureus Infection
Chemokine Signaling Pathway
Pertussis
Regulation Of Actin Cytoskeleton
Rheumatoid Arthritis
Htlv-I Infection
Hepatitis B
Legionellosis
Leishmaniasis
Salmonella Infection
Pi3K-Akt Signaling Pathway
Phagosome
Tnf Signaling Pathway
Signaling Pathways Regulating Pluripotency Of Stem Cells
Tuberculosis
Bacterial Invasion Of Epithelial Cells
Leukocyte Transendothelial Migration
Inflammatory Bowel Disease (Ibd)
Ecm-Receptor Interaction
Proteoglycans In Cancer
Cytokine-Cytokine Receptor Interaction
Transcriptional Misregulation In Cancer
African Trypanosomiasis

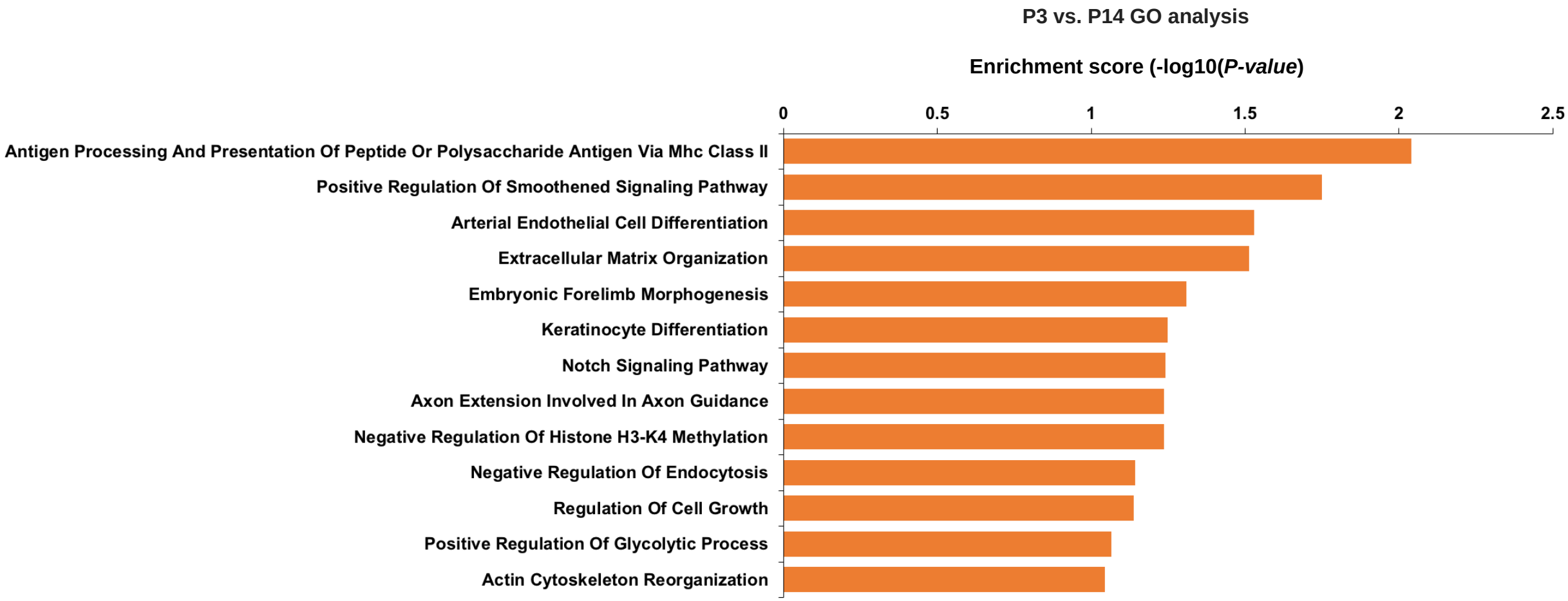
C.


P1 vs. P14 GO analysis
Enrichment score (-log10(*P-value*)
0
0.5
1
1.5
2
2.5
3
Response To Hypoxia
Intrinsic Apoptotic Signaling Pathway
Inflammatory Response
Positive Regulation Of Transcription From Rna Polymerase Ii Promoter
Epidermis Morphogenesis
Regulation Of Cell Proliferation
Negative Regulation Of Cysteine-Type Endopeptidase Activity Involved In Apoptotic Process
Negative Regulation Of Plasminogen Activation
Positive Regulation Of Cell Proliferation
Fructose 2,6-Bisphosphate Metabolic Process
Response To Virus
Fructose Metabolic Process
Cell-Matrix Adhesion
Cellular Response To Lipopolysaccharide
Peripheral Nervous System Development

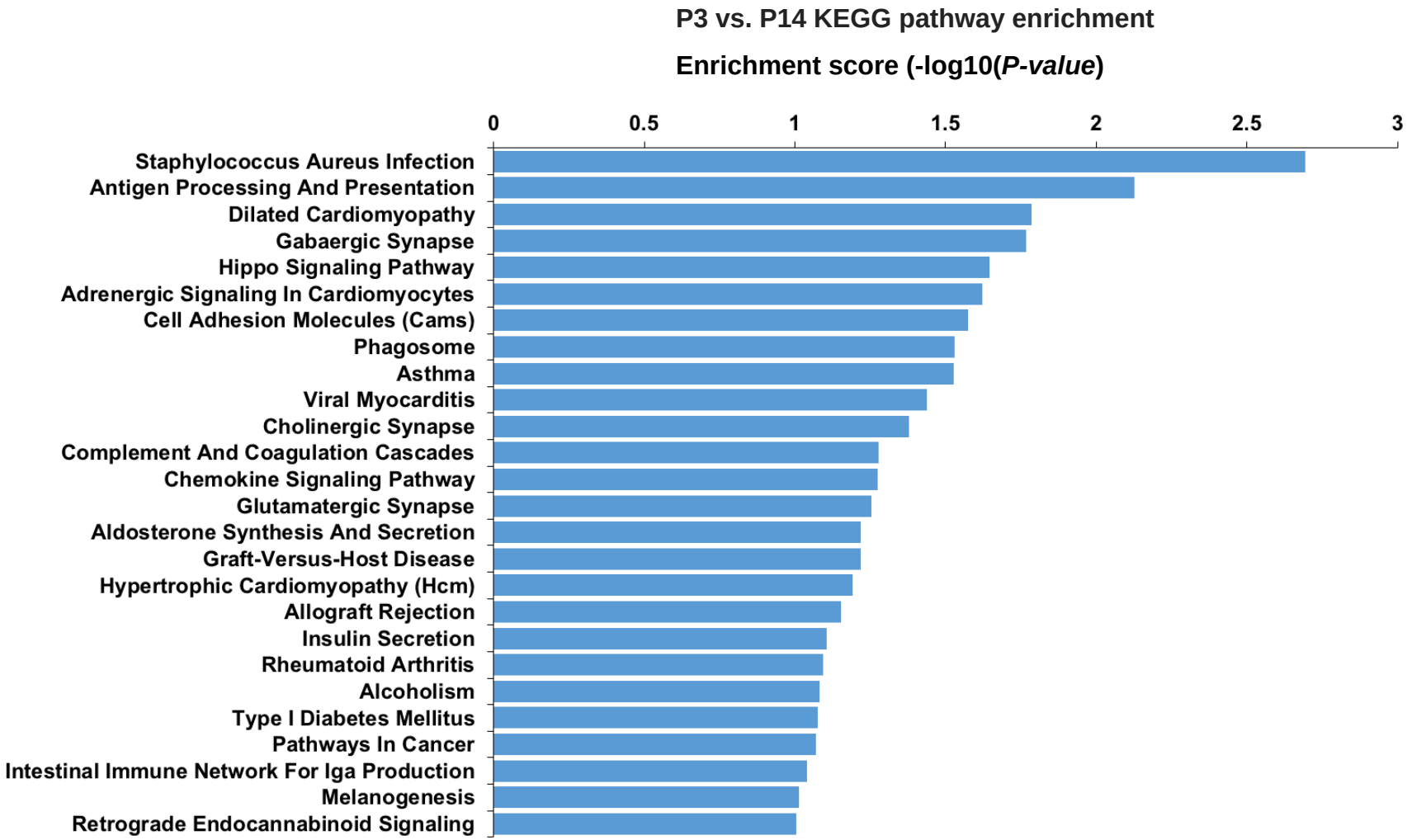
D.


P1 vs. P14 KEGG pathway enrichment
Enrichment score (-log10(*P-value*)
0
0.5
1
1.5
2
2.5
Biosynthesis Of Antibiotics
Hepatitis B
Pi3K-Akt Signaling Pathway
Glycolysis / Gluconeogenesis
Transcriptional Misregulation In Cancer
Biosynthesis Of Amino Acids
Ecm-Receptor Interaction
Bladder Cancer
Metabolic Pathways
Arginine And Proline Metabolism
Parkinson'S Disease
Non-Alcoholic Fatty Liver Disease (NAFLD)

E.


P3 vs. P14 GO analysis
Enrichment score (-log10(*P-value*))
0
0.5
1
1.5
2
2.5
Antigen Processing And Presentation Of Peptide Or Polysaccharide Antigen Via Mhc Class II
Positive Regulation Of Smoothened Signaling Pathway
Arterial Endothelial Cell Differentiation
Extracellular Matrix Organization
Embryonic Forelimb Morphogenesis
Keratinocyte Differentiation
Notch Signaling Pathway
Axon Extension Involved In Axon Guidance
Negative Regulation Of Histone H3-K4 Methylation
Negative Regulation Of Endocytosis
Regulation Of Cell Growth
Positive Regulation Of Glycolytic Process
Actin Cytoskeleton Reorganization

F.


P3 vs. P14 KEGG pathway enrichment
Enrichment score (-log10(*P-value*))
0
0.5
1
1.5
2
2.5
3
Staphylococcus Aureus Infection
Antigen Processing And Presentation
Dilated Cardiomyopathy
Gabaergic Synapse
Hippo Signaling Pathway
Adrenergic Signaling In Cardiomyocytes
Cell Adhesion Molecules (Cams)
Phagosome
Asthma
Viral Myocarditis
Cholinergic Synapse
Complement And Coagulation Cascades
Chemokine Signaling Pathway
Glutamatergic Synapse
Aldosterone Synthesis And Secretion
Graft-Versus-Host Disease
Hypertrophic Cardiomyopathy (Hcm)
Allograft Rejection
Insulin Secretion
Rheumatoid Arthritis
Alcoholism
Type I Diabetes Mellitus
Pathways In Cancer
Intestinal Immune Network For Iga Production
Melanogenesis
Retrograde Endocannabinoid Signaling

Supplemental Table I. Antibodies.

Antigen	Manufacturer	Catalog #	Type	Immuno-histochemistry	Western blot
Cardiac Troponin T	R&D Systems	MAB1874	Mouse monoclonal	1:100	
Ki67	EMD Millipore	AB9260	Rabbit polyclonal	1:100	
Phosphorylated Histone H3 (serine 10)	EMD Millipore	06-570	Rabbit polyclonal	1:300	
Aurora B	EMD Millipore	04-1036	Rabbit Monoclonal	1:100	
Nkx2.5	R&D Systems	AF2444	Goat polyclonal	1:25	
Nkx2.5	Thermo Fisher Scientific	PA5-49431	Rabbit polyclonal		1:1500
GAPDH	Abcam	ab22555	Rabbit polyclonal		1:1000
FITC-donkey anti-mouse	Jackson ImmunoResearch Laboratory	715-095-150	Polyclonal	1:200	
Cy™3 -donkey anti-rabbit	Jackson ImmunoResearch Laboratory	711-165-152	Polyclonal	1:200	
Cy5-donkey anti-mouse	Jackson ImmunoResearch Laboratory	715-175-150	Polyclonal	1:200	

Supplemental Table II. Sample information for RNA sequencing

Sample Name	File Name	Condition
P1_1	7640_S9	P1
P1_2	7690_2_S3	P1
P3_1	7654_S10	P3
P3_2	7655_S8	P3
P3_3	7657_S11	P3
P14_1	7618_S8	P14
P14_2	7619_S9	P14

Supplemental Table III. Top 25 genes that are differentially expressed between P1 and P3 conditions ranked by network centrality measures (RP Score).

Rank	Symbol	P1 vs. P3 (logFC)	RP Score
1	FN1	-3.13	270.62
2	ACTB	-1.45	195.78
3	STAT3	-0.84	174.25
4	FOS	-3.41	172.77
5	ANXA1	-1.38	141.02
6	ADCY4	-1.05	131.52
7	SHC1	-0.66	130.29
8	CCNB1	2.49	127.95
9	LPAR1	-2.03	122.39
10	VIM	-1.44	115.42
11	BDKRB1	-5.68	113.11
12	RGS19	-1.47	111.44
13	RUVBL1	0.75	108.30
14	CDH1	-4.11	102.12
15	TNFRSF1A	-0.79	98.39
16	VAV1	-2.35	97.68
17	PTGS2	-2.96	96.34
18	C3	-3.82	87.57
19	SNCA	-3.19	82.98
20	HCK	-2.99	81.18
21	RGS2	-2.76	80.76
22	JAK3	-2.26	80.30
23	PTK2B	-1.53	79.47
24	CXCR6	-3.53	79.43
25	FGR	-2.46	71.5

Supplemental Table IV. Top 25 genes that are differentially expressed between P1 and P14 conditions ranked by network centrality measures (RP Score).

Rank	Symbol	P1 vs. P14 (logFC)	RP Score
1	MYC	-1.56	291.72
2	FOS	-4.63	156.84
3	CCND1	1.34	140.85
4	CXCR2	-9.47	86.83
5	PTGS2	-2.86	81.01
6	PKM	1.24	79.54
7	F2R	0.82	76.75
8	RGS2	-2.12	76.66
9	RPS2	-2.33	76.01
10	CXCL2	-3.78	70.88
11	RGS1	-4.13	68.69
12	DYNLL1	1.34	64.08
13	SOCS3	-2.54	58.86
14	EIF1AY	24.77	56.67
15	NR4A1	-3.54	54.79
16	NXF1	-1.04	52.88
17	GAB2	-0.79	51.57
18	ARHGEF25	-0.89	51.27
19	PCF11	-1.04	48.85
20	ISG15	2.47	47.84
21	STAT6	-1.19	46.98
22	CCNA1	7.38	44.70
23	THBS1	-2.92	44.66
24	NDE1	-1.12	42.09
25	PLAU	-1.57	39.12

Supplemental Table V. Top 25 genes that are differentially expressed between P3 and P14 conditions ranked by network centrality measures (RP Score).

Rank	Symbol	P3 vs P14 (logFC)	Rp Score
1	ACTB	1.26	189.55
2	ANXA1	1.39	170.94
3	LPAR1	1.43	151.69
4	BDKRB1	5.24	141.42
5	ADCY4	1.31	139.34
6	RGS19	1.19	136.49
7	CDH1	3.40	105.52
8	GAL	5.07	90.65
9	PYY	6.39	84.34
10	HTR1D	-4.17	83.67
11	ADRA1B	-1.78	81.32
12	MET	1.77	79.49
13	PPARGC1A	-1.01	75.64
14	TAC3	6.88	74.00
15	NR4A1	-2.34	73.33
16	DNMT1	-1.35	67.40
17	ISG15	1.98	65.13
18	GAB2	-0.86	54.95
19	TSHR	5.99	54.76
20	CDC45	1.38	54.56
21	NGFR	2.74	51.51
22	CAMK2A	-1.05	49.81
23	AVPR2	2.12	48.89
24	SEC61B	1.09	47.52
25	FABP4	1.90	44.70