Supplementary Material

Transcriptomic investigation reveals donor specific gene signatures in human lung transplants

Cristina Baciu[1], Andrew Sage[1], Ricardo Zamel[1], Jason Shin[1], Xiao-Hui Bai[1], Olivia Hough[1], Mamatha Bhat[2,3], Jonathan C Yeung[1,2,4], Marcelo Cypel[1,2,4,5], Shaf Keshavjee*[1,2,4,5], Mingyao Liu*[1,2,4,5]

Affiliations:

[1]Latner Thoracic Surgery Research Laboratories, Toronto General Hospital Research Institute, University Health Network

[2]Multiorgan Transplant Program, University Health Network

[3]Division of Gastroenterology, University of Toronto

[4]Toronto Lung Transplant Program, Department of Surgery, University of Toronto

[5]Institute of Medical Science, University of Toronto

*These authors share senior authorship.

Table of contents

Supplementary Methods

Differential gene expression analysis

We have performed microarray gene expression analysis using limma package [1] in R (Version 3.5.0). For data normalization we employed the Robust Multi-array Average (RMA). After having fit the model with lmFit function (linear model), the differential gene expression was calculated using eBayes function (moderated t-test, p-value, B stats). The differential gene expression was calculated between DBD and DCD samples in the three group categories (All samples, EVLP, non-EVLP), or between EVLP and non-EVLP samples within DBD or DCD lung samples. Differentially expressed genes were defined as having an FDR<0.05 using the Benjamin-Hochberg procedure [2] first, and then by fold change, as per main text.

Pathway and network analysis

The lists of the DE genes and their statistical and experimental parameters (FDR-corrected p-value, $\log_2$FC) corresponding to each group comparison in this study, as explained above, were uploaded to the IPA (Ingenuity Systems®, www.ingenuity.com) to perform pathway and network analyses. IPA uses its own, manually curated "Knowledge Database" which gathers data from experiments already validated and published in peer-reviewed journals. A pathway is predicted to be activated or inhibited based on a calculated z-score using a specific algorithm meant to reduce the chance that random data will generate significant predictions. A z-score ≥ 2 implies high activation, and z-score ≤ -2 defines strong inhibition. Statistical p–values were also calculated for each pathway and network, based on the number of input genes and the total number of molecules

known by the IPA Knowledge Database to be present in that network, using a right-tailed Fisher's exact test [3].

Alternatively, for network analysis we also employed STRING Database version 10.5 [4]. The input to STRING was the short list of DE genes identified for each group comparison, strictly filtered by FDR-corrected p-value and fold change cut-off, FC $\geq$ 2 or FC $\leq$ 0.5. The networks created by our input molecules were used for further centrality calculations.

Network centrality analysis

The central nodes of the networks and the betweenness scores were identified and calculated using igraph package [5] (version 1.01) in R (Version 3.5.0), by computing the shortest paths between all the pairs of nodes in the network. Using the betweenness function, we calculated the centrality score of the nodes (vertices) in the corresponding network. A node with higher betweenness centrality would have more control over the network, because more information will pass through that node.

Multiple Logistic Regression and 10-fold cross validation

We investigated the correlation between the seven highly DE genes in EVLP, DBD vs. non-EVLP, DBD comparison using stepwise multiple logistic regression method with a selection of packages: dplyr [6], PerformanceAnalytics [7] and corrplot [8]. We validated the best model with 10-fold cross validation method, using caret [9] package. Area under the Curve (AUC) was calculated with ROCR [10] package.

References

1. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015: 43(7): e47-e47.

2. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Stat Soc Series B Methodol* 1995: 57(1): 289-300.

3. Fisher RA. On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. *J Royal Stat Soc Series B Methodol* 1922: 85(1): 87-94.

4. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015: 43(D1): D447-D452.

5. Gabor C, Tamas N. The igraph software package for complex network research. *IJ Comp Sys* 2006: 1695.

6. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation., 2019.

7. Peterson BG, Carl P. PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis. R package version 1.5.2. 2018.

8. Wei T, Simko V. R package "corrplot": Visualization of a Correlation Matrix., 2017.

9. Max Kuhn. Contributions from Jed Wing SW, Andre Williams, Chris Keefer, Allan Engelhardt, Tony, Cooper ZM, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem,, Luca Scrucca YT, Can Candan and Tyler Hunt. caret: Classification and Regression Training. . 2018.

10. Beerenwinkel N, Sander O, Lengauer T, Sing T. ROCR: visualizing classifier performance in R. *Bioinformatics* 2005: 21(20): 3940-3941.

Supplementary Table 1. Fold change (DBD vs. DCD lungs) of highly differentially expressed genes (FDR corrected p-value≤ 0.05 and fold change FC ≥2 or FC ≤ 0.5).

gene	all samples	non-EVLP samples	EVLP samples
CCL2	3.4	2.3	4.8
CXCL2	2.6	2.3	3.2
CXCL8	2.9	2.3	4.1
NR4A1	2.3	2.0	2.6
NR4A2	3.4	2.7	4.5
NR4A3	3.5	2.7	4.9
MT1M	2.4	2.7	2.4
MT1G	2.3	2.1	2.8
MT1X	2.3	2.6	2.2
MT1A	2.2	2.2	2.4
MT1JP	2.0	2.1	2.1
ADAMTS4	4.2	3.5	5.3
SELE	3.3	2.5	4.8
FOSB	4.8	4.1	6.3
SERPINE1	2.0	2.0	2.1
S100A12	2.8	2.6	2.9
CH25H	2.3	2.1	2.7
AREG	2.2	2.2	2.5
IL1R2	2.0	2.1	
CCL20	2.5		3.8
IL6	2.1		2.6
PTGS2	2.2		2.7
NAMPTP1	2.1		2.5
SOCS3	2.1		2.3
MYC	2.0		2.2
LOC102724428	2.0		2.2
HAS2	2.0		2.8
SLC19A2	2.0		2.0
LOC101926959		0.50	
RND1			2.6
IL1β			2.2
PTX3			2.2
NFKBIZ			2.2
IER3			2.2
PIGA			2.1
CSF3			2.1

EVLP, ex-vivo lung perfusion. DBD, donors after brain death; DCD, donation after cardiac death.

Supplementary Table 2. Pathway analysis detailed information. In orange are shown activated pathways, in blue inhibited pathways. DBD, donation after brain death; DCD, donation after circulatory death; EVLP, ex-vivo lung perfusion.

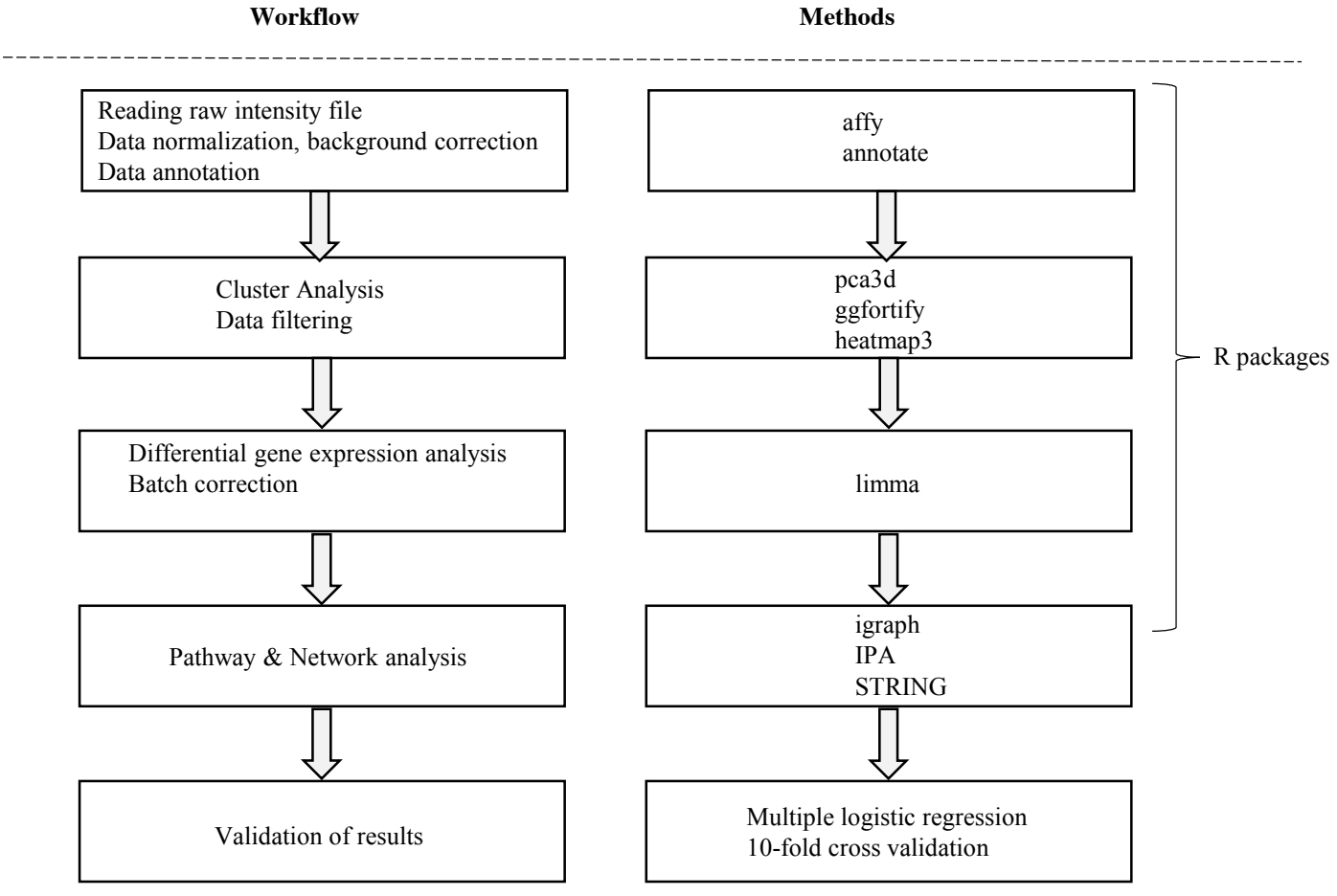
Group	Ingenuity Canonical Pathway	-log(p-value)	zScore
All	ERK5 Signaling	2.73E+00	3.00
(DBD vs DCD)	IL-6 Signaling	4.37E+00	2.65
	TREM1 Signaling	3.57E+00	2.65
	HMGB1 Signaling	3.62E+00	2.59
	Hypoxia Signaling in the Cardiovascular System	2.46E+00	2.50
	Acute Phase Response Signaling	3.27E+00	2.41
	B Cell Receptor Signaling	3.97E+00	2.21
	MIF-mediated Glucocorticoid Regulation	2.37E+00	2.14
	MIF Regulation of Innate Immunity	1.51E+00	2.14
	Pyridoxal 5'-phosphate Salvage Pathway	1.39E+00	2.06
	p38 MAPK Signaling	2.95E+00	2.06
	LXR/RXR Activation	3.83E+00	-2.54
	Complement System	1.71E+00	-2.53
	Th1 Pathway	3.32E+00	-2.33
	iCOS-iCOSL Signaling in T Helper Cells	3.99E+00	-2.06
Non-EVLP	Chondroitin Sulfate Biosynthesis	1.54E+00	3.00
(DBD vs DCD)	Dermatan Sulfate Biosynthesis	1.45E+00	3.00
	TREM1 Signaling	3.29E+00	2.84
	HMGB1 Signaling	1.42E+00	2.84
	p38 MAPK Signaling	3.09E+00	2.83
	IL-6 Signaling	4.54E+00	2.60
	MIF-mediated Glucocorticoid Regulation	1.31E+00	2.45
	AMPK Signaling	1.82E+00	2.36
	ERK/MAPK Signaling	1.38E+00	2.29
	1D-myo-inositol Hexakisphosphate Biosynthesis II	1.90E+00	2.24
	ERK5 Signaling	1.68E+00	2.11
	Valine Degradation I	1.35E+00	-2.00
EVLP	IL-6 Signaling	4.50E+00	3.77
(DBD vs DCD)	p38 MAPK Signaling	3.78E+00	3.13
	Pyridoxal 5'-phosphate Salvage Pathway	1.69E+00	3.05
	NRF2-mediated Oxidative Stress Response	2.08E+00	2.98
	Role of IL-17F in Allergic Inflammatory Airway D	2.07E+00	2.71
	ERK5 Signaling	2.84E+00	2.67
	HMGB1 Signaling	1.60E+00	2.56
	IL-1 Signaling	2.39E+00	2.52
	LPS-stimulated MAPK Signaling	1.39E+00	2.50
	PI3K Signaling in B Lymphocytes	1.97E+00	2.40
	TREM1 Signaling	3.06E+00	2.36
	Acute Phase Response Signaling	3.91E+00	2.26
	4-1BB Signaling in T Lymphocytes	1.74E+00	2.24
	Salvage Pathways of Pyrimidine Ribonucleotides	1.52E+00	2.18

Hypoxia Signaling in the Cardiovascular System	2.64E+00	2.12
iNOS Signaling	1.31E+00	2.12
Lymphotoxin Œ≤ Receptor Signaling	2.34E+00	2.11
Aryl Hydrocarbon Receptor Signaling	1.53E+00	2.06
IL-17A Signaling in Gastric Cells	1.33E+00	2.00
LXR/RXR Activation	3.34E+00	-3.13

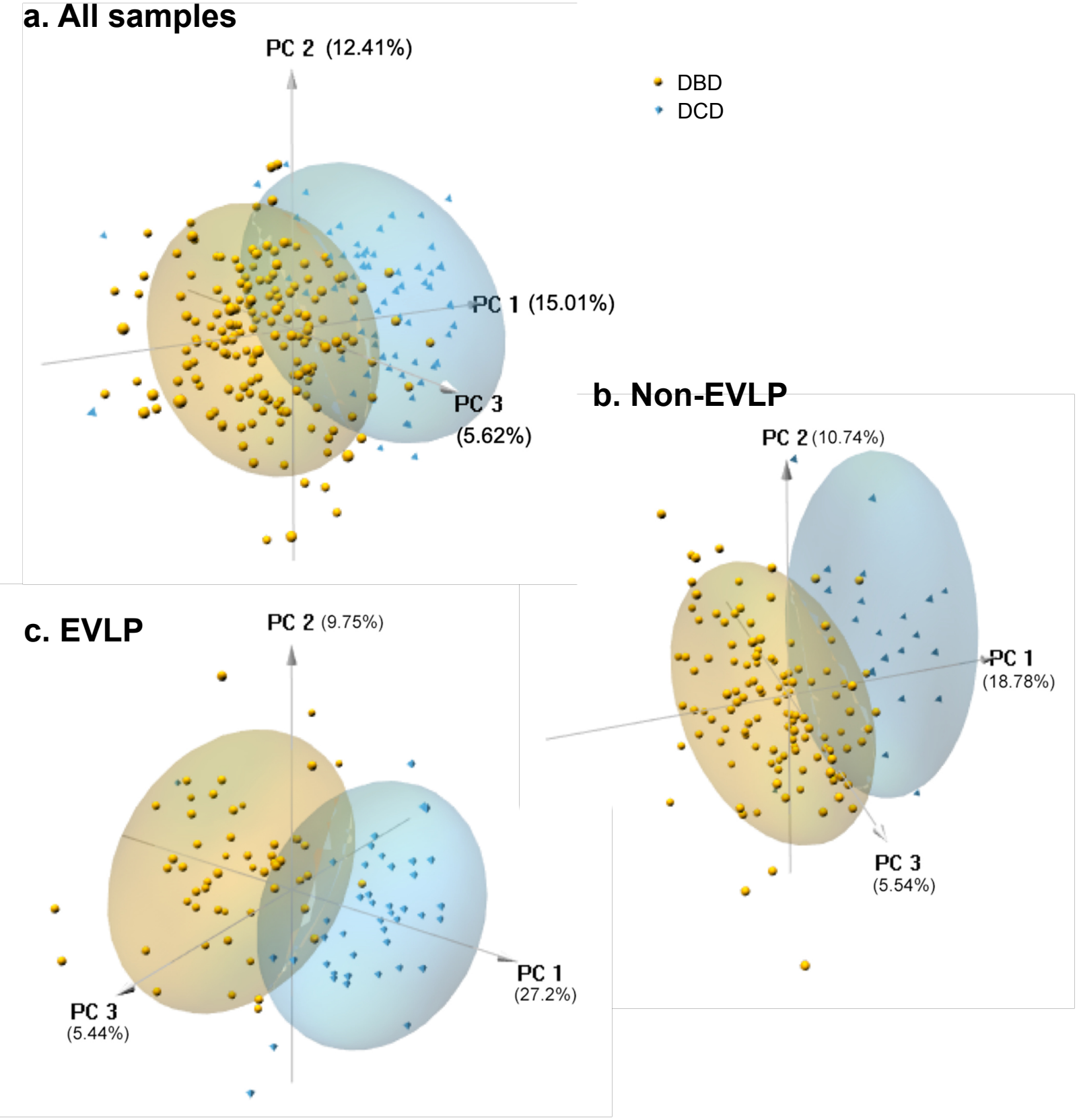
Supplementary Table 3: IPA Pathway analysis results in details, for DBD lungs, EVLP vs non-EVLP. DBD, donation after brain death; EVLP, ex-vivo lung perfusion.

Ingenuity Canonical Pathways	-log(p-value)	zScore	Molecules
TNFR2 Signaling	3.01E+00	1.00	NFKB1,NFKBIE,NFKBIA,BIRC3
TWEAK Signaling	2.76E+00	-1.00	NFKB1,NFKBIE,NFKBIA,BIRC3
MIF-mediated Glucocorticoid Regulation	2.67E+00	1.00	NFKB1,NFKBIE,NFKBIA,PLA2G5
MIF Regulation of Innate Immunity	2.32E+00	1.00	NFKB1,NFKBIE,NFKBIA,PLA2G5
Nicotine Degradation II	2.28E+00	-2.24	CYP4B1,CYP4X1,INMT,FMO2
TNFR1 Signaling	2.19E+00	1.00	NFKB1,NFKBIE,NFKBIA,BIRC3
Induction of Apoptosis by HIV1	1.89E+00	-1.00	NFKB1,NFKBIE,NFKBIA,BIRC3
Protein Kinase A Signaling	1.67E+00	1.63	HIST3H3,NFKB1,PPP1R14A,EYA1,PPP1R3C,PTP4A1,CNGA4,AKAP14, PTPRT,NFKBIE,NFKBIA,MYLK3
Antioxidant Action of Vitamin C	1.56E+00	-1.00	NFKB1,NFKBIE,NFKBIA,PLA2G5

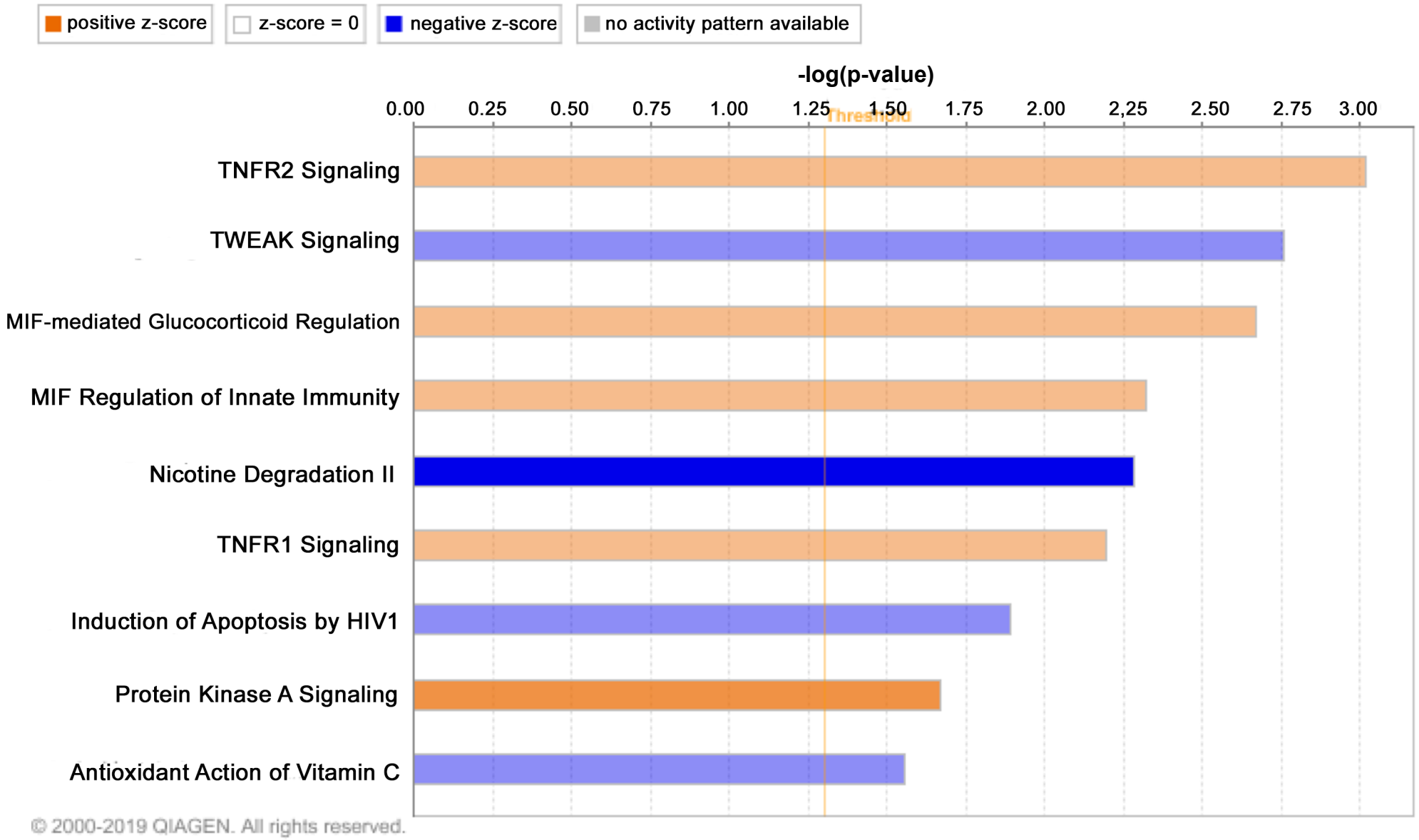
activated pathway
inactivated pathway



Supplementary Figure 1. Flow chart of bioinformatics analysis



Supplementary Figure 2. Principal Component Analysis: **a.** All samples; **b.** non-EVLP samples only; **c.** EVLP samples only. The numbers in parenthesis show the percent variance explained by the principal component. DBD, donation after brain death; DCD, donation after circulatory death; EVLP, ex-vivo lung perfusion



Supplementary Figure 3. Pathways predicted to be activated or inhibited in EVLP vs. non-EVLP, DBD samples
DBD, donation after brain death; EVLP, ex-vivo lung perfusion