Pre-transplant Transcriptomic Signature in Peripheral Blood Predicts Early Acute Rejection

Weijia Zhang Ph.D.[1], Zhengzi Yi M.S.[1] , Chengguo Wei Ph.D.[1] , Karen L. Keung M.D.[2], Zeguo Sun M.S.[1], Caixia Xi M.S.[1], Christopher Woytovich B.S.[1] , Samira Farouk M.D.[1], Lorenzo Gallon M.D.[3], Madhav C. Menon M.D.[1], Ciara Magee M.D.[4], Nader Najafian M.D.[4], Milagros D. Samaniego M.D.[5], Arjang Djamali M.D.[6], Stephen I. Alexander M.D.[2], Ivy A. Rosales M.D[7], Rex Neal Smith M.D, Ph.D.[7], Philip J. O'Connell M.D.[2] , Robert Colvin M.D.[7], Paolo Cravedi M.D. Ph.D.[1] and Barbara Murphy M.D.[1]

Running title: pre-transplant blood transcriptomic signature in kidney transplantation

1 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
2 Department of Medicine, Westmead Clinical School, The University of Sydney, Sydney, NSW, Australia
3 Department of Medicine-Nephrology and Surgery-Organ Transplantation, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
4 Department of Medicine, Brigham and Women's Hospital, 75 Francis Street, Boston, MA, USA
5 Henry Ford Hospital, Detroit, MI 48202
6 Division of Nephrology, Department of Medicine, University of Wisconsin, Madison, WI, USA
7 Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Correspondence:

Dr. Barbara Murphy, M.D.
Division of Nephrology
Department of Medicine
Icahn School of Medicine at Mount Sinai
One Gustave L Levy Place, Box 1243
New York, NY 10029
Phone: 212-241-8001
E-mail: barbara.murphy@mssm.edu

Supplementary Materials

Supplementary methods:

RNA sequencing experiments and data processing

The whole blood was drawn from kidney transplant recipients prior to transplant using PAXgene tube at 5 centers and shipped to Mount Sinai Center for RNA extraction. Total RNA was isolated using Preanalytix PAXgene blood miRNA kit that purifies both total RNA and miRNA from whole blood collected in PAXgene tube (PreAnalytiX GmbH) and mRNA sequencing was performed on an Illumina HiSeq4000 sequencer by following manufactory protocol (Illumina Inc.). The clean reads were firstly aligned to human reference databases including hg19 human genome, exon, splicing junction segment and contamination database of ribosome and mitochondria sequences using BWA alignment algorithm (*1*). After filtering out reads mapped to contamination database, the reads that were uniquely aligned to the exon and splicing-junction segments with a maximal 2 mismatches for each transcript were then counted as expression level for corresponding transcript. The read counts were log2 transformed, quantile-normalized and corrected for experimental batch using ComBat R package (*2*) in order to compare transcription levels across samples.

Identification of transcriptomic signatures associated with EAR

Differential gene expression analysis between EAR and non-EAR patients was performed by LIMMA test (*3*) with the following 8 recipient confounders: Age, Gender, Race, Clinical Center, Types of Kidney Diseases before transplant, Anti-HLA Antibodies Class I (Y or N), Anti HLA Antibodies Class II (Y or N), and Dialysis (Y or N). Differentially Expressed Genes (DEGs) were initially identified at p-value<0.05.

Biological functional/pathways enriched for DEGs were determined by fisher-exact test at p value < 0.05 using the information of biological process category in Gene Ontology (GO) (*4*) and pathways curated in the several pathway databases (KEGG, Ingenuity IPA, BIOCARTA, NABA, Panther, PID,

REACTOME, Wiki-pathway). Alternatively, Geneset Enrichment Analysis (GSEA) (*5*) was applied to the entire expression dataset to determine pathways that were enriched in samples with EAR vs non-EAR.

The immune cell types associated with EAR were investigate by fisher-exact test ($p < 0.05$) of the enrichment of immune cell specific genes in DEGs. The immune cell specific genes were identified from ImmGene databases as described previously (*6*). Briefly, we downloaded the public expression data of various immune cell types (https://www.immgen.org/) and identified highly expressed genes for each immune cell type by the rank of gene expression across cell types. We then checked which immune cell types are correlated with EAR based on the enrichment of immune cell type genes in the DEGs by Fisher exact test at $p <0.05$. The deconvolution method using CIBERSORT algorithm (*7*) was also performed to estimate the population percentage of immune cells for each sample from bulk RNA sequencing profile. Based on the assumption that the expression value for each immune cell marker in the bulk RNA-seq is the weighted sum of each cell type in the expression base matrix of 547 immune cell markers in 22 sorted pure immune cells (547 x22 matrix), CIBERSORT performs Support Vector Regression (SVR) (*8*) on the bulk expression value of marker genes to calculate the weight of each cell type which is then converted into cell population percentages. Student's t test was used to determine population change for each cell type between groups (*e.g.*, EAR vs non-EAR) at the cutoff of false discovery rate less than 0.01.

To understand how DEGs were co-regulated/co-expressed during EAR development, the expression correlation network of these DEGs within the EAR samples was built. Only connections between genes with a p-value of Pearson correlation test of less than 0.05 and absolute correlation coefficient equal to or above 0.6 were kept. Markov Cluster Algorithm (*9*) was then applied to divide the correlation matric into sub-correlation networks for which the Gene Ontology (GO) functions were determined by fisher-exact test at p less than 0.05. Genes were then ranked by number of correlation connection edges with other genes (correlation connectivity) within the network. The hub genes with the highest connectivity with other genes were further considered as potential candidates for prediction of EAR.

Identification of an optimal gene set for prediction of EAR

To identify an optimal gene set to predict EAR, a focus gene set specifically associated with EAR was firstly identified from DEGs under the following criteria: 1) log2 expression value between 6.5 and 15 to eliminate the extreme low or high expression values; 2) significantly differential expression between EAR and No-EAR at p value <0.05 after correction for recipient pre-transplant characteristics described above; 3) highly correlated with at least 40 other genes within co-expression correlation network with correlation coefficiency >0.60 and 4) maintained significance of LIMMA test ($p<0.05$) within two randomly-assigned subsets of equal size from discovery dataset in at least 2 out of 1,000 iterations using the approach described previously (*6*). The focus gene sets were ordered based on the combined significance of these criteria.

Next, by using a cumulative gene risk scoring system and forward selection approach (adding one gene each time from the focus gene set), a minimal gene set was identified with the best prediction (area under the curve (AUC) of ROC (Receive Operating Characteristic) curve) in the discovery set with all-inclusive iterations of the focus gene set. The cumulative gene risk score was computed in the following formula:

$$r = -(\log_{10}(p_1)*g_1 + \log_{10}(p_2)*g_2 + \ldots + \log10(p_i)*g_i + \ldots + \log_{10}(p_n)*g_n)$$

where p_i is the significance p value of LIMMA-test (*3*) on expression values for each focus gene i (i=1…n) between the EAR vs the non-EAR groups in the training set, g_i is a logic number (1, -1, or 0) for each focus gene i (i=1…n). The logic number was determined based on the comparison of the expression of gene i (E_i) to the median value of EAR (M_{EAR}) or non-EAR (M_{NEAR}) groups in the discovery set in the following situations:

1. 1 if $E_i > M_{EAR}$ for an upregulated gene or if $E_i < M_{EAR}$ for a downregulated gene;
2. -1 if $E_i < M_{NEAR}$ for an upregulated gene or if $E_i > M_{NEAR}$ for a downregulated gene;
3. 0 if E_i is between M_{EAR} and M_{NEAR}

The weighted cumulative score (r) was used as a gene risk score for acute rejection for each patient. ROC (Receive Operating Characteristic) curve of the true positive rate versus the false positive rate at various threshold settings of the risk score was generated in the discovery and training sets and the area under the curve (AUC) was calculated to estimate the overall prediction accuracy. Positive prediction value (PPV) of acute rejection and negative predictive value (NPV) of no-acute rejection were determined at a given risk score cutoff.

Luminex protein assay

Luminex assay (ThermalFisher Inc.) with a panel of 45 immune cell surface or cytokines proteins was performed on serum samples from 21 randomly-selected patients (10 EAR and 11 non-EAR) in the discovery set. The protein expression levels were determined and normalized based on the titration curve of the standard protein markers that were included in the assay. The cytokine or immune response proteins whose expression at the baseline were significantly associated with EAR were identified by t test at $p < 0.05$.

EAR prediction with clinical factors alone or along with gene risk score

Clinical factors that were associated with EAR were first selected via t-test for continuous variables or fisher-exact test for categorical variables. To combine the clinical factors, we re-defined binary clinical factors as 1 or -1 (1 if the factor is positively associated with EAR and -1 if a factor was negatively associated with EAR). Continuous variable such as recipient age was defined 1 or -1 (1 if above median for positive correlation or below median for negative correlation by the median value). Individual clinical scores for all factors were then summed to a composite score for each patient.

To estimate the prediction accuracy with both gene risk score and clinical score, clinical score based on the age, kidney diseases and presence of anti-HLA antibody and p-weighted cumulative gene risk score were further put into Penalized Logistic Regression model in R package logistf (*10*). The penalized logistic regression model used Firth's bias reduction method to reduce the bias of maximum

likelihood estimates due to small sample size, which will resolve the issue of overfitting from standard logistic regression method. The formula to calculate probability is $P = \frac{e^{\sum \beta X}}{1+e^{\sum \beta X}}$ where β are coefficients generated from penalized logistic regression model and X are expression values. A penalized logistic regression model was built on Discovery Set (D) and then applied on validation set (V) using same parameters to estimate prediction performance.

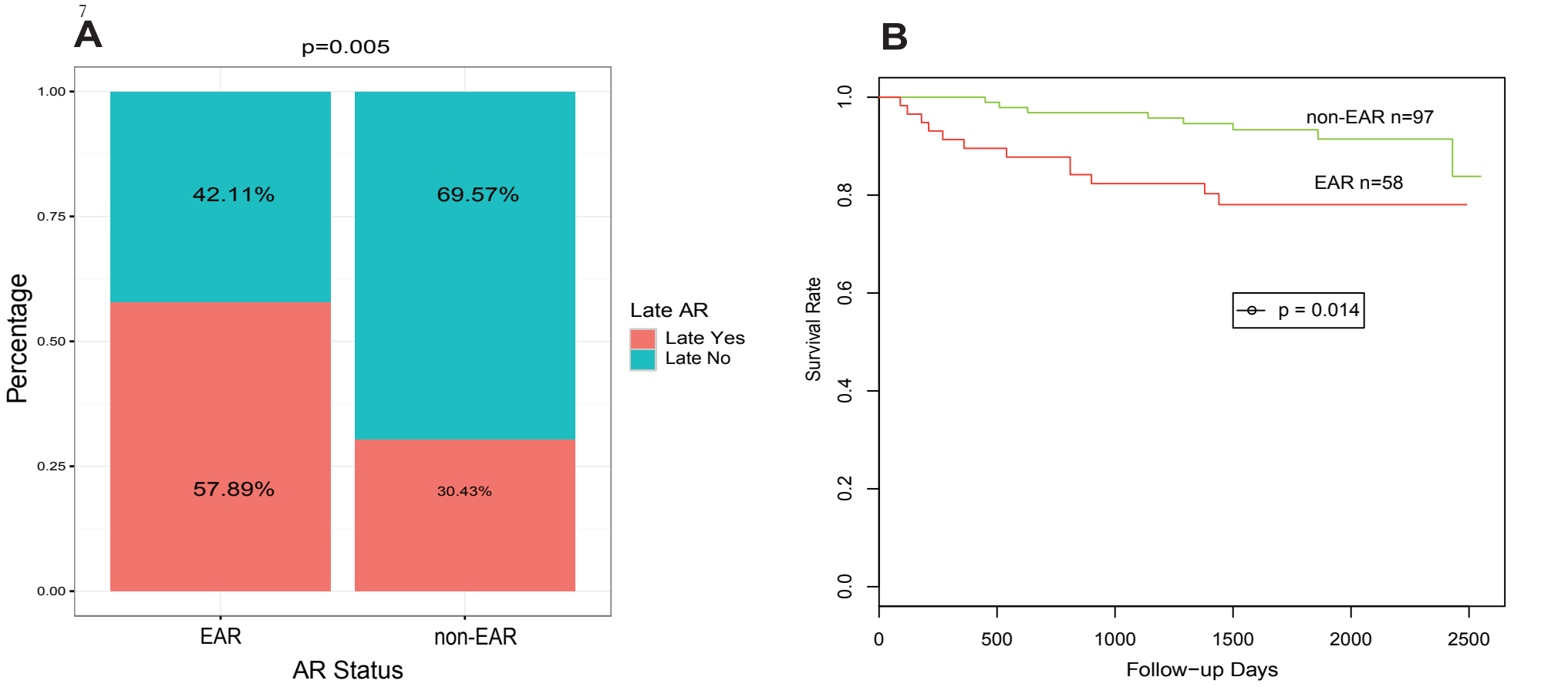


Figure S1. Association of EAR with late acute rejection and graft loss. (**A**) The bar chart shows the percentage of subjects with late acute rejection (after 6 months post-transplant) within patients with or without EAR (t-test p=0.005); (**B**) Kaplan-Meier curve for graft loss in patients with EAR (n=58) or non-EAR (n=97) (log rank test p=0.014); (**C**) ROC curve for prediction of EAR based on demographic and clinical factors (recipient age, kidney diseases and the presence of anti-HLA antibody) (n=155, AUC=0.59).

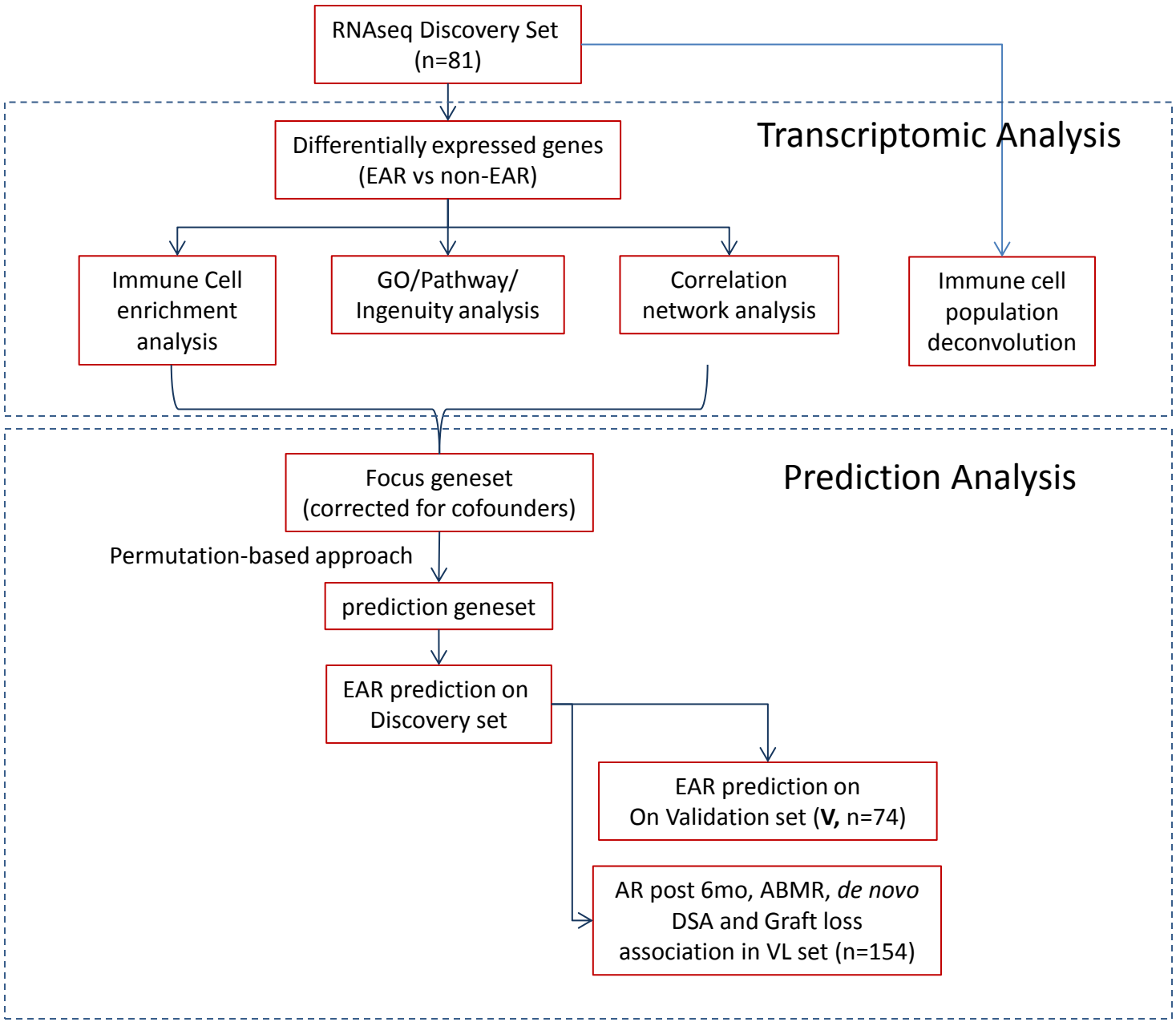


Figure S2. Data analysis workflow. The data analysis workflow includes transcriptomic and prediction analyses. The transcriptomic analysis identified the transcriptomic signatures in pre-transplant blood associated with EAR and revealed the cell functions, pathways and interaction networks for these gene signatures. The prediction analysis was to identify an optimal gene set from the discovery set that predicted EAR. EAR prediction with the gene risk score developed from the gene set was further validated on independent validation set. The association of acute rejection post 6 months, antibody mediated rejection (ABMR), *de novo* DSA and graft loss with the gene risk score was also investigated.

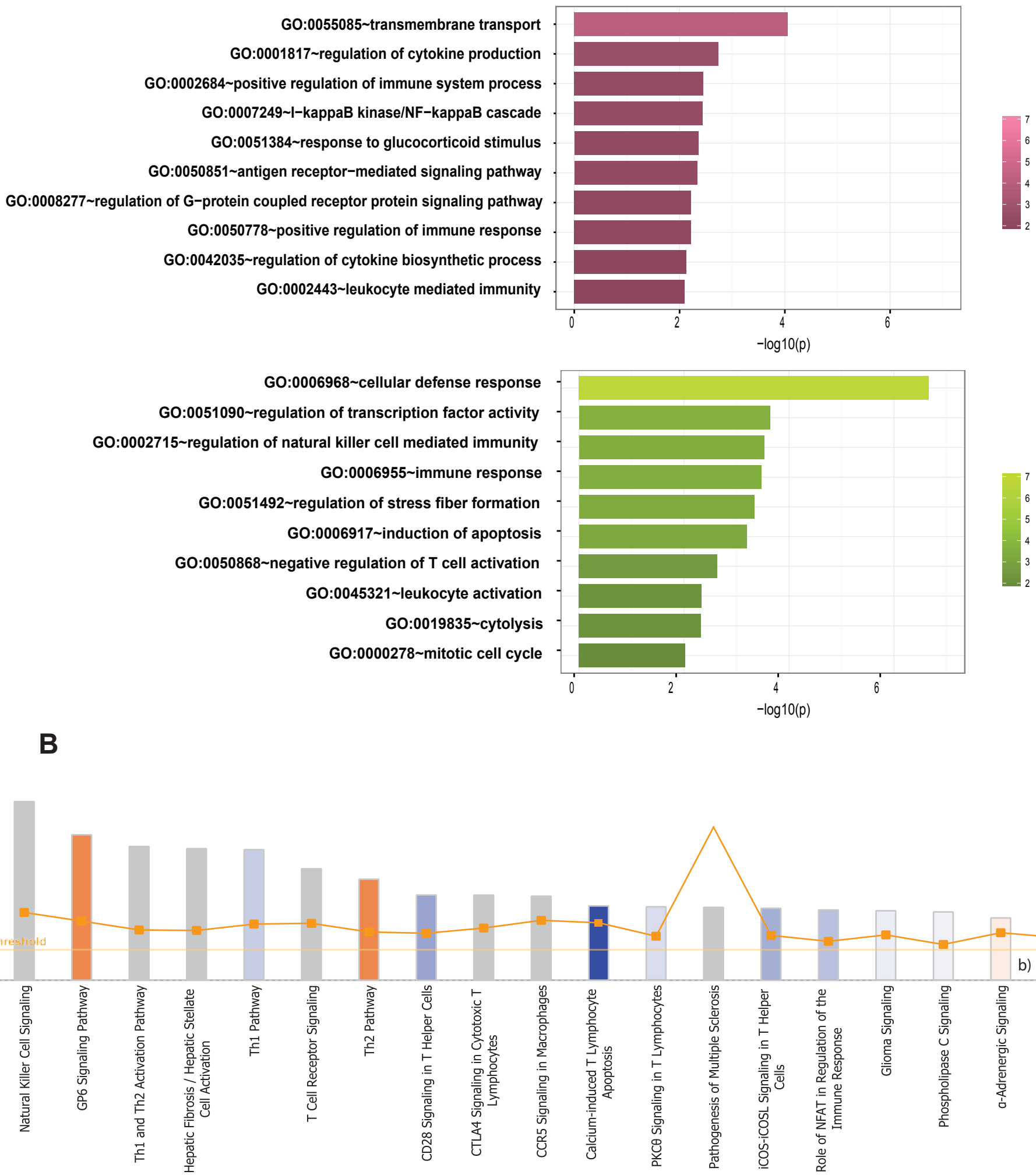
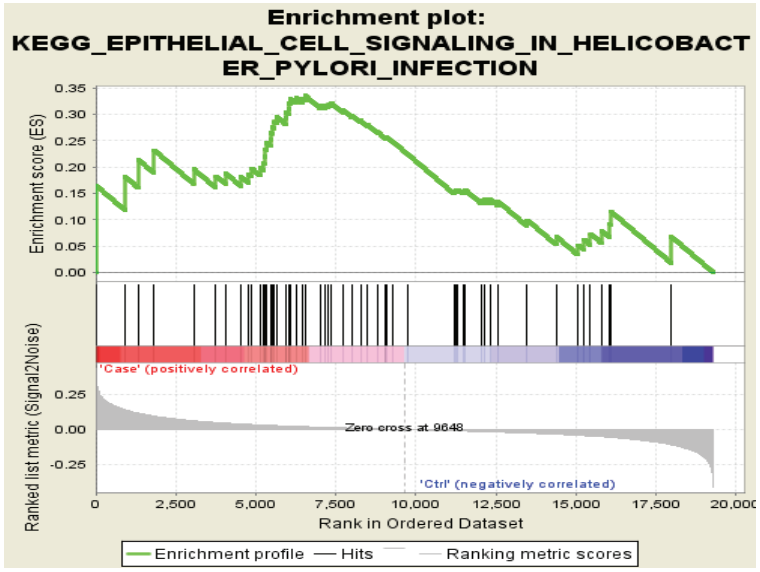
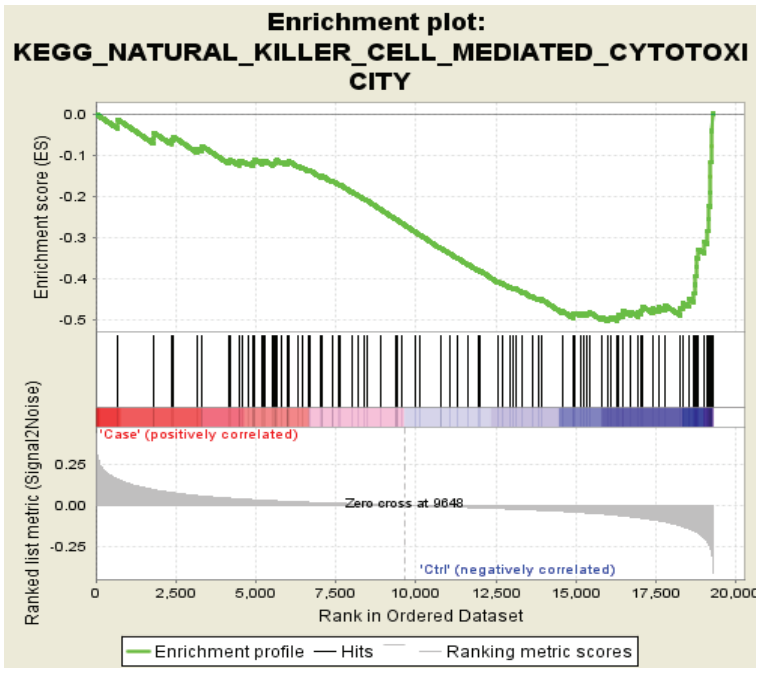


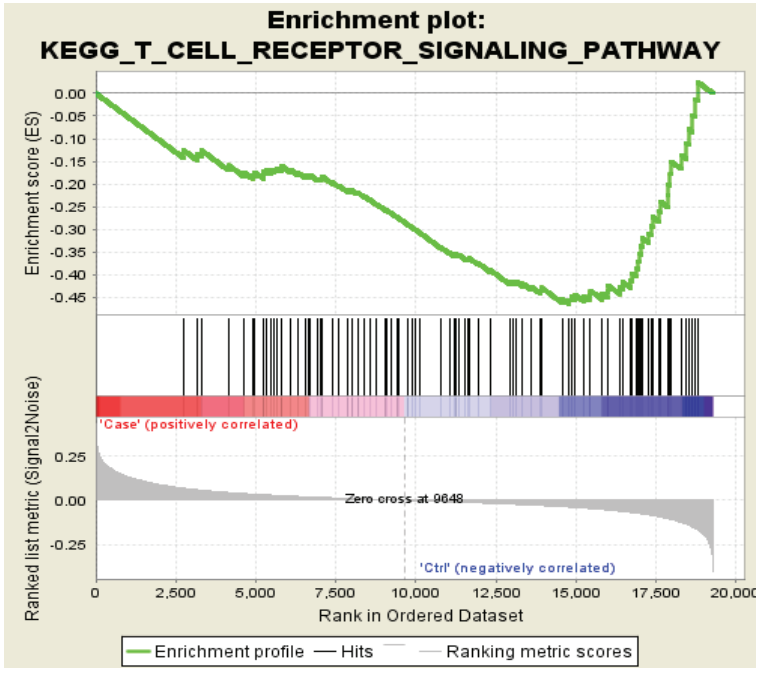
Figure S3. Gene function and pathways for differentially expressed genes (DEGs) associated with EAR. (**A**) The bar chart of Gene Ontology enrichment analysis for up- **(upper panel)** or down- **(lower panel)** regulated genes in EAR vs No-EAR. The bar represents –log10 p value of enrichment significance for Gene Ontology terms by Fisher exact test; (**B**) The bar chart of enriched canonical pathways for DEGs from Ingenuity Pathway Analysis (IPA) tool. The Y-axis represents –log10 p value of enrichment significance of IPA pathways by Fisher exact test.

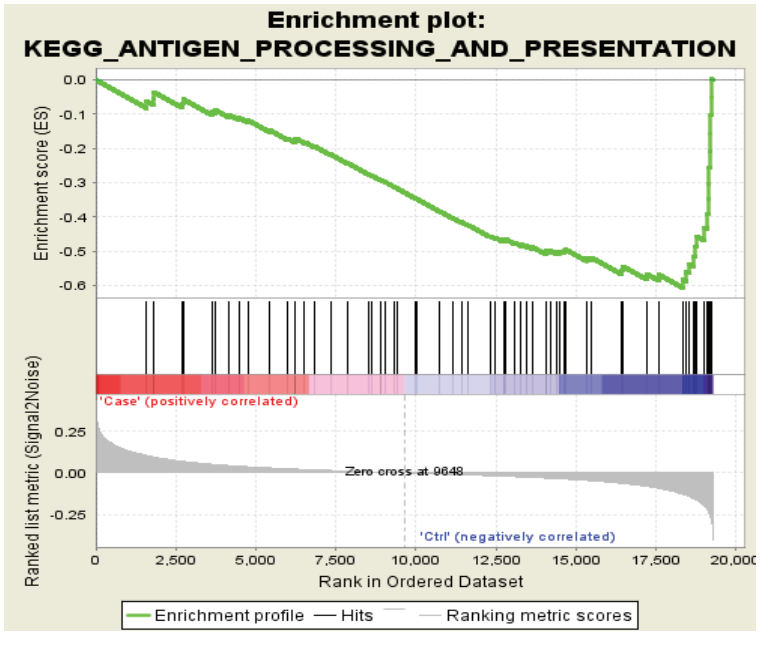
A



B







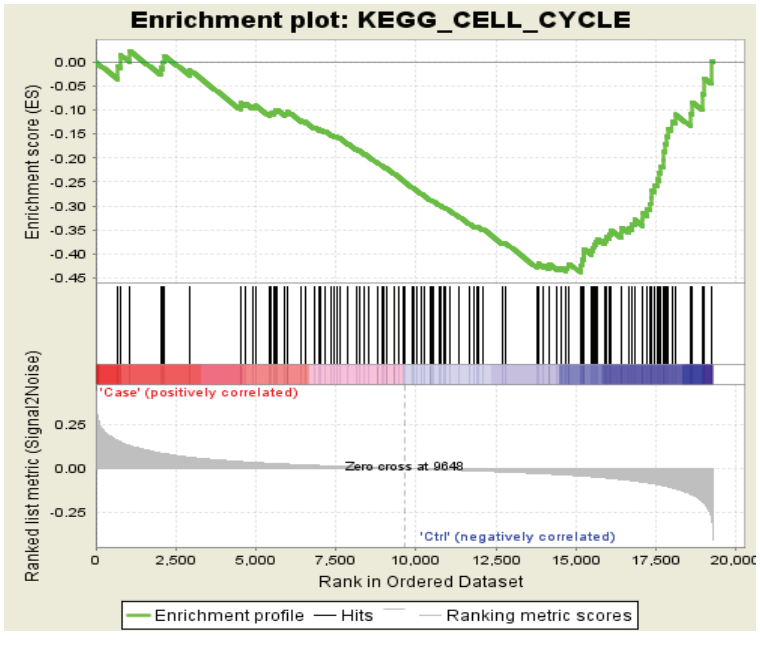


Figure S4. Dysregulated KEGG pathways associated with EAR by Gene Set Enrichment Analysis (GSEA). (**A**) The plot of upregulated pathways from GSEA analysis; (**B**) The plot of down-regulated pathways from GSEA analysis.

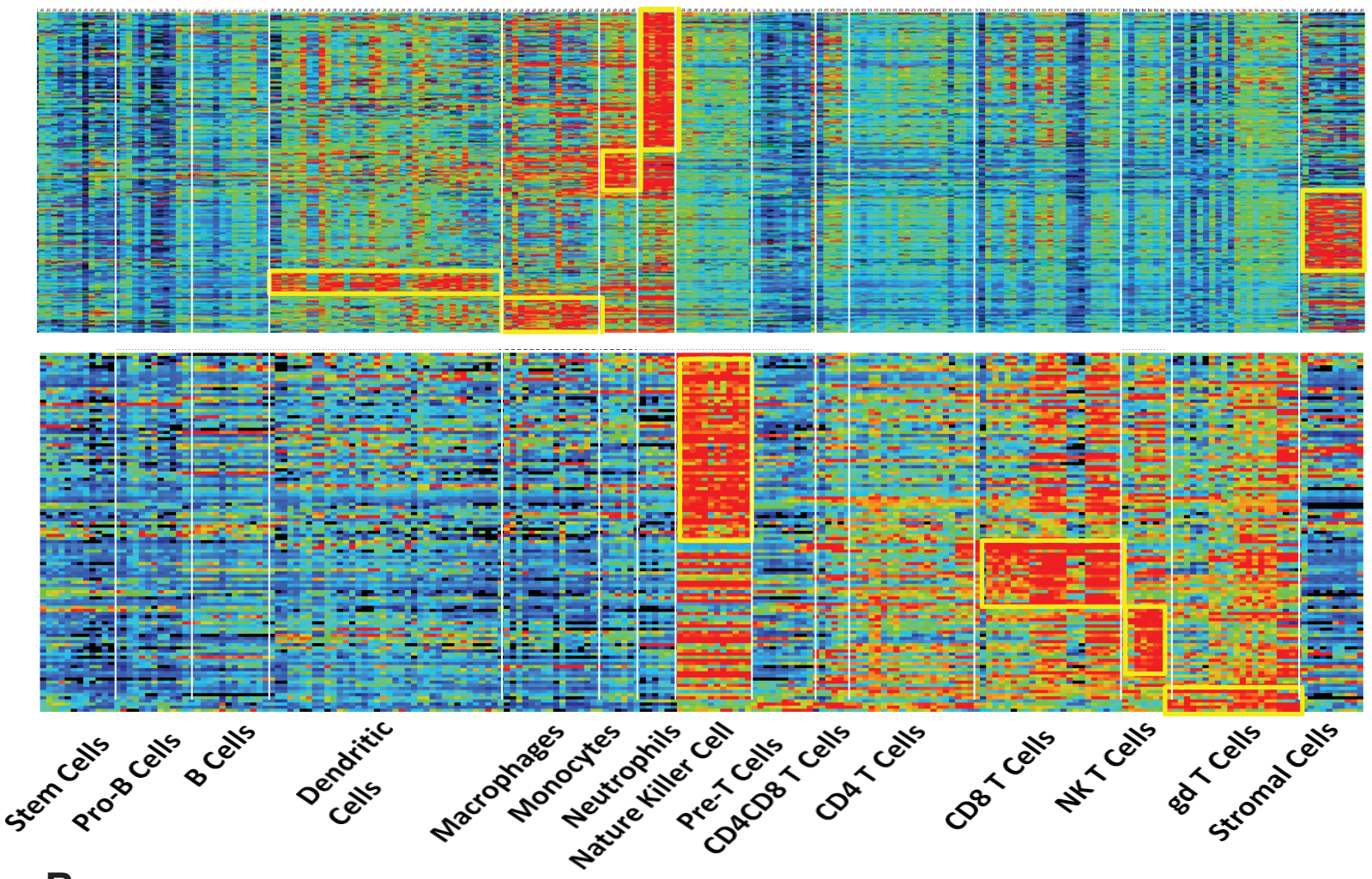


Figure S5. Immune cell enrichment analysis of DEGs associated with EAR in Discovery Set (n=81). (**A**) The heatmap shows expression of up- (upper panel) and down- (lower panel) regulated genes that were significantly enriched for immune cell types; (**B**) The bar chart shows the difference in NK and CD8+ T cell populations between EAR and non-EAR recipients based on gene expression. The population percentages of NK and CD8+ T cells were deconvoluted from the RNA sequencing using the expression profiles of sorted immune cells.

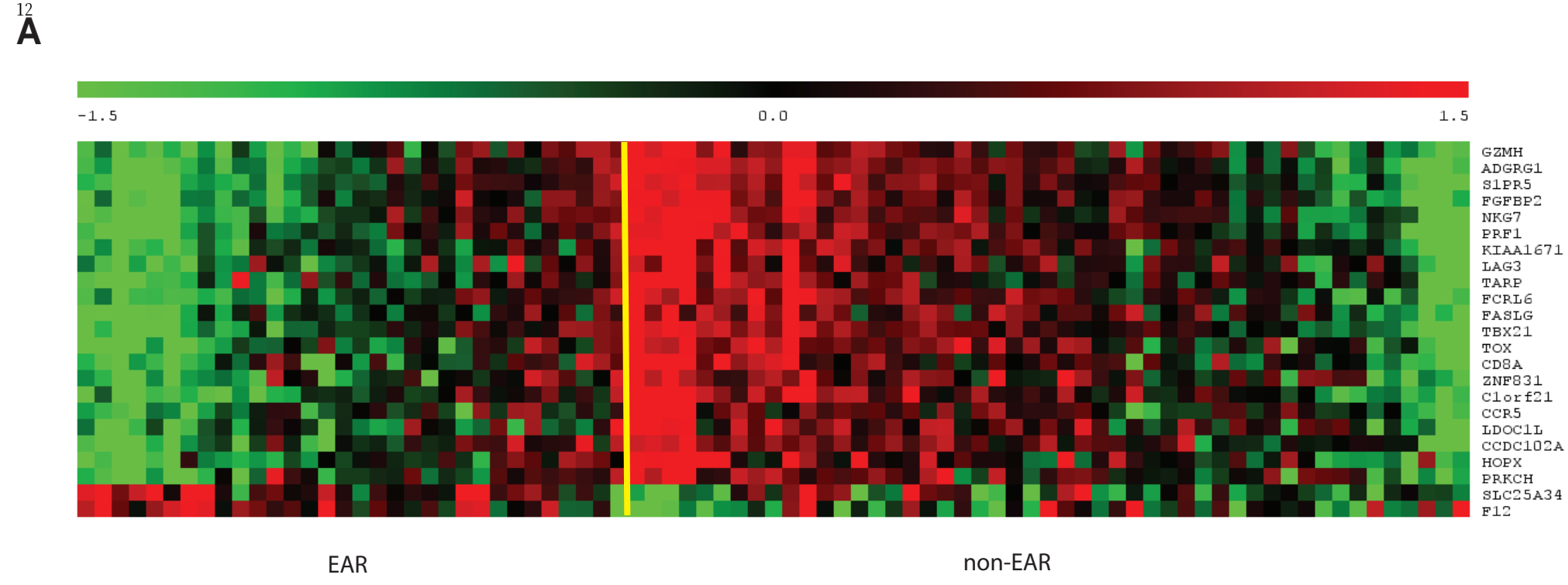


Figure S6. Identification of 23-gene set for EAR prediction from the discovery (D) set (n=81). (**A**) The heatmap for expression of the 23 genes in EAR and non-EAR recipients; (**B**) The dot plot of the gene risk score derived for the 23-gene set in the discovery set. EAR is shown in red color.

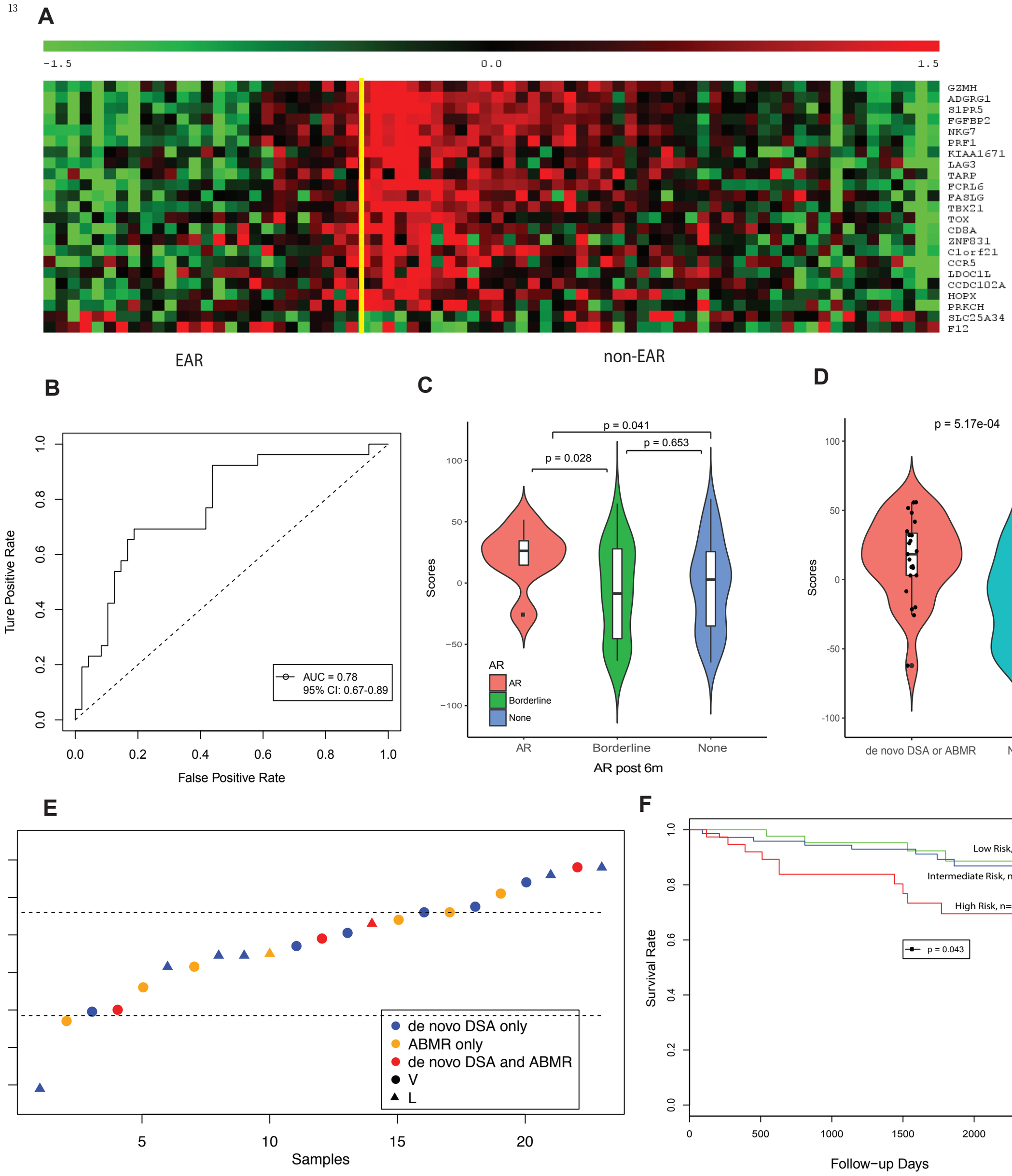


Figure S7. Association of gene risk score with clinical outcomes. (**A**) The heatmap of expression of 23 genes in EAR and non-EAR in validation (V) set (n=74); (**B**) The ROC curve for prediction of EAR with gene set and demographic/clinical characteristics in V set (n=74, AUC=0.78); (**C**) The violin plot of distribution of risk scores among the patents with the acute rejection borderline, 1A and above, and no-AR after 6 months post-transplant in VL (V+L) set (n=154); P values are significant between AR and non-AR group (t-test p=0.041) and between AR and borderline groups (t-test p=0.028); (**D**) The violin plot of distribution of risk scores between the patents with the antibody mediated rejection (ABMR) or de novo DSA and the patents without clinical events in VL cohorts (n=154, t-test p=5.17e-04); (**E**) The dot plot of the gene risk scores of the patents who developed ABMR (n=7), de nono DSA (n=12) or both (n=4) in VL set; (**F**) Kaplan-Meier curve of graft loss for high, intermediate or low gene-risk groups in the VL set (n=154, log rank test p=0.043).

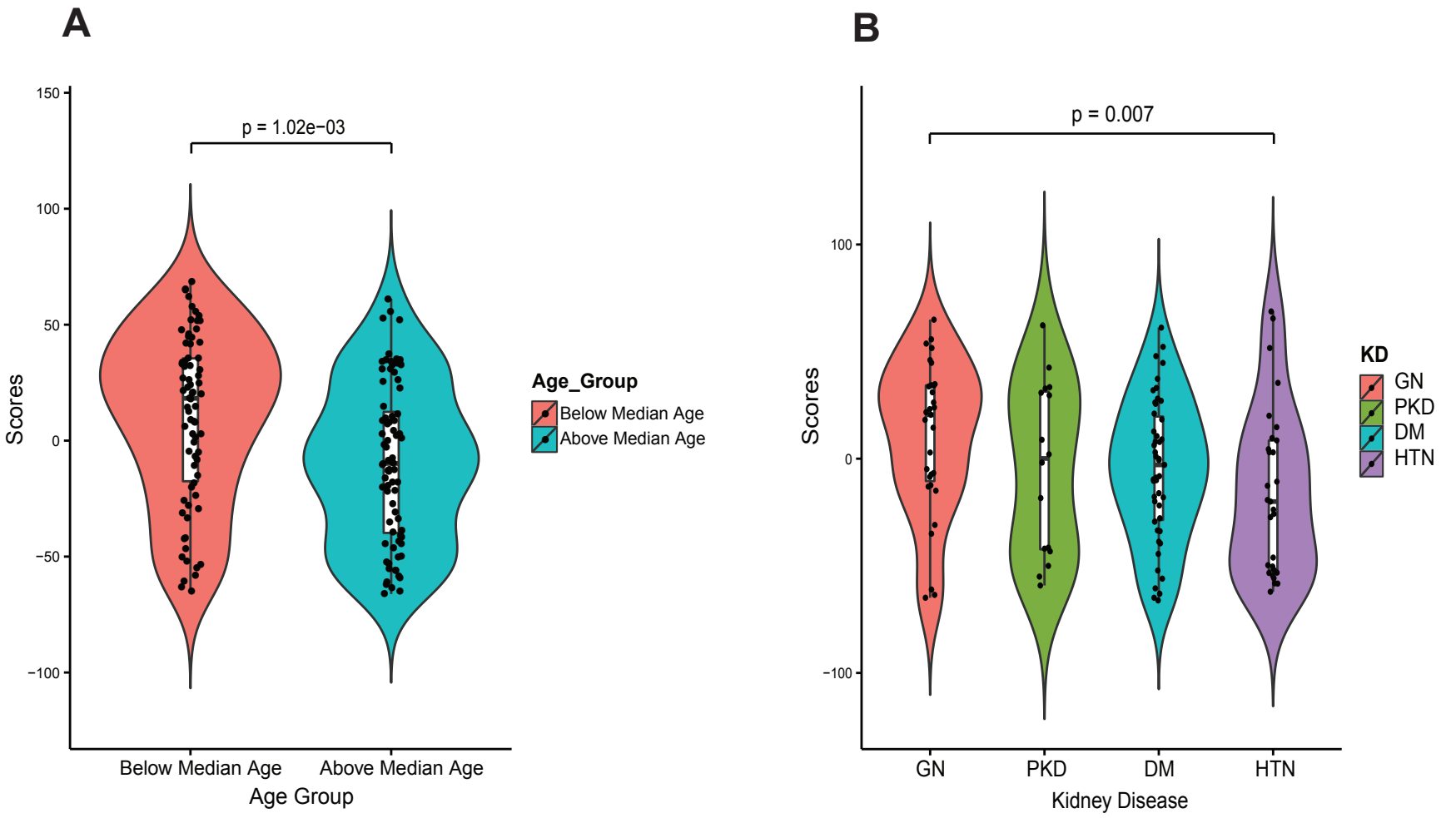


Figure S8. (**A**) The violin plot of risk score distribution of the recipients with young (below median age) or old (above median age) ages in VL set (n=154, t-test p = 1.02 x 10-3); (**B**) The violin plot of risk score distribution of the recipients with kidney diseases in VL set (n=154, t-test p=0.007 for Glomerular Nephropathy (GN) vs hypertension (HTN)).

Table S1. Statistics of clinical events in EAR discovery and validation cohorts and late biopsy cohort

Biopsy Time	Discovery (n=81)	Validation (n=74)	Late Biopsy (n=80)
Acute Cellular Rejection	1A or above/borderline/None (*Percentage %*)	1A or above/borderline/None (*Percentage %*)	1A or above/borderline/None Count (*Percentage %*)
Any time before 6m (EAR)	8/24/49 *(10/30/60)*	7/19/48 *(9/26/65)*	
12m surveillance	1/13/52	6/11/18	5/14/48
24m surveillance	3/15/27	1/6/18	1/7/33
Clinical indication after 6m	1/1/0	1/4/1	1/0/1
Antibody Mediated Rejection *(anytime)*	4	9	2
De novo DSA* *(anytime)	3	9	7
Death Censored Graft Loss	6	14	10
CADI 3m (mean ± sd)	1.77 ± 1.63	1.82 ± 1.93	-
CADI 12m (mean ± sd)	1.89 ± 1.78	2.51 ± 2.42	1.93 ± 2.35

Table S2. Univariate and multivariate association analysis of baseline characteristics with EAR

Characteristics	Correlation Coefficient	Pvalue	Lower 95%CI	Upper 95%CI
Kidney Disease	-0.18337	0.022383 *	-0.33143	-0.02648
Age	-0.19459	0.015255 *	-0.34176	-0.03811
Gender	-0.12314	0.12688	-0.27544	0.035188
Race	0.044044	0.586327	-0.1144	0.200301
Anti HLA Ab Class I	0.28196	0.000379 *	0.130094	0.420899
Anti HLA Ab Class II	0.246645	0.001976 *	0.092597	0.389161
Dialysis	-0.00507	0.95005	-0.16259	0.152698
Donor Age	0.121585	0.131794	-0.03677	0.273982
Donor Race	-0.01811	0.823027	-0.17526	0.139938
Donor Gender	-0.01157	0.886415	-0.16891	0.146348
Race mismatch	-0.11147	0.16732	-0.26447	0.047004
HLA mismatch	0.141071	0.079969	-0.01695	0.29222
Induction Type	0.195783	0.014629 *	0.03935	0.34285
Deceased Donor	-0.0183	0.821171	-0.17544	0.139749
CIT min	-0.0834	0.302209	-0.23792	0.075238
Baseline DSA	0.105306	0.535062	-0.22644	0.41516

Multivariate analysis

Characteristics	Estimate	Pvalue	Lower 95%CI	Upper 95%CI
Kidney Disease	-0.156521	0.044092 *	-0.30884746	-0.004194476
Age	-0.005425	0.061926 .	-0.01112437	0.000274093
Anti_HLA_Ab_Class_I	0.200429	0.076493 .	-0.02159867	0.422457022
Anti_HLA_Ab_Class_II	0.059473	0.645666	-0.19560043	0.314546052
Induction Type	0.1228	0.125596	-0.0347319	0.280331011

Table S3 70 focus gene set

Symbol	Refseq	Name	p	Log2Rat
UNC5A	NM_133369	unc-5 netrin receptor A	0.000564	0.744594
SLC22A1	NM_003057	solute carrier family 22 member 1	0.00184	0.721141
LOC100128770	NR_047572	uncharacterized LOC100128770	0.009685	0.695806
TNFRSF9	NM_001561	TNF receptor superfamily member 9	0.004618	0.582904
F12	NM_000505	coagulation factor XII	0.001584	0.578932
LOC101927759	NR_132750	uncharacterized LOC101927759	0.005349	0.547646
SLC25A34	NM_207348	solute carrier family 25 member 34	0.000146	0.502699
TLR9	NM_017442	toll like receptor 9	0.0072	0.472316
PNPLA1	NM_173676	patatin like phospholipase domain containing 1	0.004332	0.456271
PRR7-AS1	NR_038916	PRR7 antisense RNA 1	0.006132	0.446602
TLL2	NM_012465	tolloid like 2	0.00265	0.411318
EFCAB2	NM_032328	EF-hand calcium binding domain 2	0.003917	0.409008
FAM71F2	NM_001290254	family with sequence similarity 71 member F2	0.000262	0.405794
PRKCH	NM_006255	protein kinase C eta	0.002781	-0.39483
OSBPL3	NR_104112	oxysterol binding protein like 3	0.00196	-0.40196
TAF5L	NM_001025247	TATA-box binding protein associated factor 5 like	0.002858	-0.40219
RALGDS	NM_001042368	ral guanine nucleotide dissociation stimulator	0.007487	-0.40731
OSBPL5	NM_020896	oxysterol binding protein like 5	0.004351	-0.41255
KLRC4-KLRK1	NM_001199805	KLRC4-KLRK1 readthrough	0.007978	-0.41722
CXCR6	NM_006564	C-X-C motif chemokine receptor 6	0.006575	-0.42696
KLRK1	NM_007360	killer cell lectin like receptor K1	0.007653	-0.43276
SAMD3	NM_001258275	sterile alpha motif domain containing 3	0.000293	-0.43517
HOPX	NM_001145460	HOP homeobox	0.001553	-0.43995
RUNX3	NM_001031680	runt related transcription factor 3	0.006633	-0.44917
NCALD	NM_032041	neurocalcin delta	0.008782	-0.45435
DUSP14	NM_007026	dual specificity phosphatase 14	0.000271	-0.46375
S1PR1	NM_001320730	sphingosine-1-phosphate receptor 1	0.007227	-0.4643
TGFBR3	NM_001195683	transforming growth factor beta receptor 3	0.005018	-0.47309
SOX13	NM_005686	SRY-box 13	0.001239	-0.47423
ATP1A3	NM_152296	ATPase Na+/K+ transporting subunit alpha 3	0.008461	-0.48659
IL12RB2	NR_047583	interleukin 12 receptor subunit beta 2	0.002396	-0.50308
CTSW	NM_001335	cathepsin W	0.009033	-0.50945
CRY1	NM_004075	cryptochrome circadian clock 1	0.008473	-0.51275
RAB11FIP5	NM_015470	RAB11 family interacting protein 5	0.002967	-0.51774
CHST10	NM_004854	carbohydrate sulfotransferase 10	0.00368	-0.51957
MYO6	NM_004999	myosin VI	0.002548	-0.53286
PDCD1	NM_005018	programmed cell death 1	0.007274	-0.53372
CCDC102A	NM_033212	coiled-coil domain containing 102A	0.002596	-0.53815
PPP2R2B	NM_001271899	protein phosphatase 2 regulatory subunit Bbeta	0.002096	-0.53989
LDOC1L	NM_032287	LDOC1 like	0.000533	-0.56501

Symbol	Refseq	Name	p	Log2Rat
CCR5	NM_000579	C-C motif chemokine receptor 5 (gene/pseudogene)	0.004226	-0.5786
CCL4	NM_002984	C-C motif chemokine ligand 4	0.006202	-0.58038
CCL5	NM_001278736	C-C motif chemokine ligand 5	0.006226	-0.59069
C1orf21	NM_030806	chromosome 1 open reading frame 21	9.98E-05	-0.59152
PDGFD	NM_025208	platelet derived growth factor D	0.001433	-0.59698
ZNF831	NM_178457	zinc finger protein 831	0.000768	-0.60595
JAKMIP1	NM_144720	janus kinase and microtubule interacting protein 1	0.004344	-0.60809
CD8A	NM_001145873	CD8a molecule	0.003693	-0.61079
JAKMIP2	NM_014790	janus kinase and microtubule interacting protein 2	0.000476	-0.62183
ASCL2	NM_005170	achaete-scute family bHLH transcription factor 2	0.004651	-0.63453
SLAMF7	NM_001282589	SLAM family member 7	0.006602	-0.63742
GNLY	NM_006433	granulysin	0.005048	-0.66176
CD8B	NM_172101	CD8b molecule	0.008459	-0.66572
TOX	NM_014729	thymocyte selection associated high mobility group box	0.000157	-0.67501
TBX21	NM_013351	T-box 21	0.002089	-0.6909
FASLG	NM_001302746	Fas ligand	0.000245	-0.73934
FCRL6	NM_001004310	Fc receptor like 6	0.001577	-0.74821
TARP	NM_001003799	TCR gamma alternate reading frame protein	0.002415	-0.75633
LAG3	NM_002286	lymphocyte activating 3	0.002004	-0.8103
KIAA1671	NM_001145206	KIAA1671	0.00025	-0.81243
EOMES	NM_001278182	eomesodermin	0.001642	-0.85472
PRF1	NM_001083116	perforin 1	0.002216	-0.85621
NKG7	NM_005601	natural killer cell granule protein 7	0.001314	-0.86597
RGS9	NM_001081955	regulator of G-protein signaling 9	0.002093	-0.91064
FGFBP2	NM_031950	fibroblast growth factor binding protein 2	0.001386	-0.9483
S1PR5	NM_001166215	sphingosine-1-phosphate receptor 5	0.00091	-1.0053
SPON2	NM_001199021	spondin 2	0.004025	-1.00941
MSC	NM_005098	musculin	0.00931	-1.01379
ADGRG1	NM_005682	adhesion G protein-coupled receptor G1	0.000952	-1.12477
GZMH	NM_001270780	granzyme H	0.002051	-1.34764

Table S4. Association of recipient baseline characteristics with the gene risk score in VL cohort

Characteristics	Low Risk Group (n=43)	Intermediate Riks Group (n=73)	High Risk Group (n=38)	Pvalue
Recipient Age	54 ±13.09	49.2 ±13.69	45.1 ±12.87	0.0113
Recipient Gender				0.0760
Male	28 (65.12)	60 (82.19)	26 (68.42)	
Female	15 (34.88)	13 (17.81)	12 (31.58)	
Recipient Race				0.3268
RaceWhite / Caucasian	24 (55.81)	49 (67.12)	28 (73.68)	
RaceBlack or African American	10 (23.26)	13 (17.81)	3 (7.89)	
RaceOthers	9 (20.93)	11 (15.07)	7 (18.42)	
Dialysis (Y/N)				0.2337
Y	37 (86.05)	53 (72.6)	28 (73.68)	
N	6 (13.95)	20 (27.4)	10 (26.32)	
Anti_HLA_Ab_Class_I (Y/N)				0.2117
Y	7 (16.28)	14 (19.18)	12 (31.58)	
N	36 (83.72)	59 (80.82)	26 (68.42)	
Anti_HLA_Ab_Class_II (Y/N)				0.2526
Y	8 (18.6)	10 (13.7)	10 (26.32)	
N	35 (81.4)	63 (86.3)	28 (73.68)	
Induction_Type				0.3209
Lymphocyte Non-depletion	12 (27.91)	20 (27.4)	12 (31.58)	
Lymphocyte Depletion	23 (53.49)	38 (52.05)	13 (34.21)	
None	8 (18.6)	15 (20.55)	13 (34.21)	
Kidney Disease				0.0241
Diabetes Mellitus	13 (30.23)	27 (36.99)	7 (18.42)	
Glomerulonephritis	5 (11.63)	16 (21.92)	10 (26.32)	
Hypertension	14 (32.56)	11 (15.07)	4 (10.53)	
Polycystic Kidney Disease	6 (13.95)	6 (8.22)	4 (10.53)	
Reflux	0 (0)	2 (2.74)	5 (13.16)	
Other	5 (11.63)	11 (15.07)	8 (21.05)	
CDC B cell				0.2097
Negative	28 (96.55)	52 (98.11)	15 (88.24)	
Positive	1 (3.45)	1 (1.89)	2 (11.76)	
Number of Transplants	0.2 ±0.45	0.2 ±0.42	0.2 ±0.56	0.9954
Baseline DSA (Y/N)				0.1445
Y	1 (2.33)	6 (8.45)	5 (13.89)	
N	42 (97.67)	65 (91.55)	31 (86.11)	

References

1. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
2. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
3. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
4. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
5. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
6. P. J. O'Connell *et al.*, Biopsy transcriptome expression profiling to identify kidney transplants at risk of chronic injury: a multicentre, prospective study. *Lancet* **388**, 983-993 (2016).
7. A. M. Newman *et al.*, Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457 (2015).
8. B. Scholkopf, A. J. Smola, R. C. Williamson, P. L. Bartlett, New support vector algorithms. *Neural Comput* **12**, 1207-1245 (2000).
9. S. van Dongen, Graph Clustering by Flow Simulation. *PhD thesis, University of Utrecht*, (2000).
10. G. Heinze, Ploner, M.,, Dunkler, D., Southworth, H. (2016)