Supplemental Materials

Supplementary methods:

Clinical Data Collection

Donor and recipient data were collected at baseline. Donor information included age, race, gender, HLA genotype, cause of death, and graft status (i.e. SCD, ECD, or DCD). Recipient data included age, gender, race, cause of ESRD, HLA genotype, PRA, presence and type of anti-HLA antibodies, cross match status, cold ischemia time (CIT), delayed graft function (DGF), immunosuppression regimen, CMV status, HCV and HBV status, dialysis vintage, dialysis modality, transfusion history, pregnancy history, and previous transplants We have adhered to the STROBE checklist throughout the manuscript. The clinical data from 5 different clinical centers were deposited into a central clinical eRAP database developed at Mount Sinai.

Histopathology:

2 tissue cores were taken from each of the 3-month and 1 year protocol renal biopsies of the Genomics of Chronic Allograft Rejection (GoCAR) cohort. One core was processed for histology and the other core was processed for mRNA. When only one core could be obtained priority was given to mRNA at 3-months and to histology at 12-months. Renal biopsies were processed and read centrally. Formalin-fixed, paraffin-embedded sections were processed for histologic stains (hematoxylin and eosin, periodic acid Schiff, trichrome and Weigerts elastic stains). Immunohistochemistry for C4d was done on an automated stainer on paraffin sections stained with a rabbit polyclonal antibody (American Research Products, Inc.). All slides were scanned with a whole slide scanner (Aperio CS) and high-resolution digital images and archived in an image database.

Biopsies were evaluated and scored separately by 2 renal pathologists, without knowledge of the clinical data, using the Revised Banff 2007 Classification for Renal Allograft Pathology [1] (SIS reference). Where diagnoses were discordant, a meeting was held with a third pathologist for a consensus diagnosis.

Scoring was done on the whole slide images for all cases. Scores were entered into a custom Filemaker Pro database that calculated the Banff categories and Chronic Allograft Damage Index (CADI). The CADI-score is a composite score that includes six histologic components – vascular intimal sclerosis (cv), tubular atrophy (ct), interstitial fibrosis (ci), interstitial inflammation (i), mesangial matrix increase (mm) and glomerusclerosis (g). Each component is scored between 0 & 3, giving a maximum possible score of 18. CADI-scores in protocol biospies has been validated to directly correlate with outcomes by several authors[2-3].

Definition of "protocol biopsy" is that the biopsy is performed for the study or as part of a standard protocol at a given site in which a biopsy is performed based on time post-transplant and not based on an increase in creatinine or change in clinical status.

Treatment of Censored data: For analysis for histological outcomes, patients without 12-month biopsies were excluded from analysis. For analysis of functional outcomes or eGFR - we imputed a GFR of 10 ml/min for all death-censored graft losses. For analysis of graft survival via Kaplan-Meier, we censored patients whose follow up data became unavailable at respective time points, whereas patients who had "events" (ie graft-loss) were censored at the time of the event.

Microarray experiments

The graft biospies from 5 participating centers were sent to Mount Sinai and were stabilized with RNA-later(Qiagen, Inc). Total RNA was extracted from percutaneous graft biopsy samples obtained at 3 month after transplantation using All prep kit (QIAGEN-ALLprep kit, Valencia, CA USA) by one technician at Mount Sinai. RNA quality was assessed using Bioanalyzer 2100 (Agilent Technologies). Samples with an RNA Integrity Number greater than eight were used in subsequent microarray experiments. Affymetrix humanexon 1.0 ST arrays were used following standard protocol provided by the manufacturer (Affymetrix Inc.). In brief, ENCORE amplification and labeling kit (NuGen, San Carlos, CA) was applied to the first batch of samples starting with approximately 100ng of total RNA to generate

biotin-labeled RNA fragments for hybridization to the chip. For samples with a low RNA concentration, the Nugen Ovation PICO amplification kit (NuGen, San Carlos, CA) was applied. The chips were scanned using GeneChip Scanner 7G (Affymetrix Inc.). The Affymetrix genechip experiments were performed at Mount Sinai Microarray Core.

Microarray data processing

The intensity data of microrray experiments at the gene level were extracted and summarized with the RMA algorithm[4]. Data quality was assessed using the Affymetrix Expression Console (Affymetrix Inc). The Affymetrix control probesets and probesets with a low intensity (less than 20% quantile among all the data points) across all samples were excluded from downstream analysis. Batch effects were adjusted using the ComBat R package[5].

Bioinformatic analyses

The workflow of bioinformatic analysis is depicted in Suppl. Fig 3. and the analysis was performed with statistical R packages. The goal of analyses was to derive a relatively robust set of genes (~10-20) that predicts the development of chronic graft nephropathy.

Identification of the graft transcriptional signature: Spearman correlation analyses were performed on the 3-month graft gene expression data for 3-month graft CADI score (CADI-3) as well as 12-month CADI score (CADI-12). The correlation coefficient and the p-value for the relationship between the level of expression and CADI score were calculated for each gene. The slope of gene expression against the CADI score was also computed using a linear regression model. Genes with a p value of < 0.05 were selected. Two lists of genes with $p<0.05$ were generated corresponding to either the 3 month or 12 month CADI scores. Annotated functional and molecular mechanisms of these 2lists of genes were determined by Gene Ontology (GO) enrichment analysis based on Fisher-exact test. To investigate which immune cell type genes are associated with CADI-3 or CADI-12, 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 CADI-3 or CADI-12 based on the enrichment of immune cell type genes.

Alternatively, the gene expression dataset was analyzed to determine biological functions that are enriched in biopsies with higher CADI scores. To accomplish this, we applied Geneset Enrichment Analysis (GSEA)[6-7] to the entire microarray dataset and determined gene functions that are enriched in samples with a high CADI score (CADI≥2) versus those with a low CADI score (CADI<2). Top GO terms associated with both the high and low CADI groups were determined, and compared to the results of GO enrichment analysis derived from the analyses of correlation between gene expression level and CADI score described above.

Prediction analysis:

To derive a more significant and focused geneset from the large list of genes that have statistically significant association with CADI-3 or CADI-12 scores, we filtered the gene list by applying various statistical prediction models. First, the whole cohort of patients was randomly assigned to 2 groups in a 1:1 ratio. Spearman correlation analysis was applied to determine the genes with expressions levels that correlated with the severity of CADI score at 3 and 12 months. The 1:1 randomization was repeated 100 times and correlation analysis of gene expression with CADI score at 3 and 12 month was performed for each of the 100 iterations. We considered genes that occurred more than twice in the 100 iterations of randomization with a correlation at a $P<0.05$ with CADI in both groups as a focused geneset from which a minimal prediction set was identified for predicting kidney fibrosis. Genes that were exclusive to the CADI-12 focus geneset (i.e. genes not shared with the CADI-3 focus geneset) were derived and further filtered by correction for clinical confounders (donor age, living vs deceased donor, donor gender and race, cold ischemia time(CIT, minutes), induction therapy, anti HLA class I, and II antibodies) using multiple linear regression analysis, as well as exclusion of genes with a low median log2 intensity of less than 5.

Finally, we performed iterative logistic model fitting (5000 iterations) in order to identify an optimal and minimal geneset for prediction of future kidney fibrosis. We started by randomly selecting 20 genes from the filtered CADI-12 focus geneset. The expression data of the 20-gene group was fitted into Firth-type bias-reduced logistic regression model which panelizes the maximum likelihood [8 9] in logistf

R package for prediction of high (CADI-12 ≥2) and low (CADI<2) CADI-12. We used a CADI-12 cutoff of ≥ 2.0 to derive our prediction geneset. The paper we referenced by Yilmaz[2] also evaluated protocol biopsies at 1 year and divided their biopsies into 3 groups <2, 2-3.9 and >4. Whilst a CADI score >4 had the strongest association with graft loss, those with a CADI score of 2-3.9, were also associated with graft loss where as those with a score <2 did not. We selected a CADI-12≥2 based on this prior publication[3]. The genes with significant association with high/low CADI-12 ($p<0.05$) were identified from the regression model for each of the 20-gene group. The steps above were repeated 5000 times. Statistically significant genes ($P<0.05$) were identified from each iterative operation. The occurrence of significant genes from the 5000 iterations was calculated. Finally, the top 40 genes ranked by the number of occurrences were applied back to the penalized logistic regression model for high vs. low CADI-12 prediction. Statistically significant genes ($P<0.05$) using this model were considered the final optimal geneset. The AUC score and sensitivity and specificity were calculated from logistic regression model using the final optimal geneset. To investigate the overfitting issues of prediction of training set with the geneset, cross-validated prediction AUC was also calculated using a 3-fold cross-validation method. Briefly, the patients were randomly divided into 3 groups of equal size and equal number of high and low CADI-12 patients and the data for any 2groups were used as the training set with the third as the prediction set. The penalized logistic regression model that was built on the training set was applied on the prediction set to predict the outcome and the true and false positive rates. Prediction accuracy was calculated from the prediction data set and then averaged from three possible permutations. We repeated the steps over 100 times. The overall true or false positive rates and prediction accuracy were computed. The distribution of AUCs on the testing set based on the model derived using the training set for 100 iterations was plotted.

To assess if the geneset we identified was an optimal geneset for high/low CADI-12 prediction, we compared the original prediction AUC with prediction AUCs from the genesets that were identified from high/low CADI-12 groups after 2000-time random re-shuffling of CADI-12 scores. Briefly, CADI-12 scores were randomly re-assigned to the patients, the prediction geneset from re-shuffled high/low

CAD-12 groups was identified in the same approach as for original CADI-12 groups and the AUC was therefore calculated. These steps were repeated 2000 times and the original AUC was compared to the 1000 AUCs from randomly-assigned CADI-12 groups.

Our geneset identification and prediction approach was further fully cross-validated with leave-one-out cross-validation algorithm. Briefly, we took one patient as validation set and the remaining patients as training set and identified new geneset from training set in the same approach as for original cohort. The logistic model built on training set with new geneset was then applied to the patient that was left out and the prediction probability was calculated. These steps were repeated in all the possible ways of selecting training and testing sets and ROC curve was drawn based on the probabilities of testing sets from all iterations.

Prediction of high/low CADI-12 at a different threshold (high CADI-12≥3 or high CADI-12 ≥4) was also performed to assess the robustness of 13 geneset prediction. To check if the inflammation was the driver of the 13 geneset, we evaluated the prediction of Banff criteria (Ci+Ct scores), IFTA and high CADI-12 on the patients without acute rejection. Lastly we investigated if the geneset can predict the kidney function and we calculated prediction AUC of eGFR at12 month or 24 month with geneset.

To investigate whether prediction by the geneset is superior to prediction by clinical variables, we performed the multivariate logistic regression for prediction of high/low CADI-12 by including the following demographic/clinical variables: donor age, recipient race and gender, deceased donor status, extended-criteria donor kidney, cold ischemia time (CIT, minutes), induction therapy, presence of Anti-HLA antibodies, delayed graft function and HLA mismatch. After step-wise selection, the variables that remained significant were used in final model. The AUC for the ROC curve of the final model was then calculated and compared to CADI-12 prediction with the geneset.

We also applied our optimal geneset to predict the progressors and non-progressors using the same approach described above. Patients who had CADI-3≤3 and demonstrated a ΔCADI≥2 by 12 month were considered as progressors, and those who had ΔCADI ≤1 were considered non-progressors. A

similar assessments were done for those with CADI score at 24 months and also for the patients with CADI-3≤2.

To test if our geneset could predict early graft loss post-transplant for the original 159 patients, we applied logistic regression prediction model with our geneset among only those patients who either had graft loss within 3yr or had been followed-up for at least three years without graft loss and calculated the AUC. Secondly, survival analysis on all 159 patients was performed to examine if our geneset was associated with graft loss: to assess the association of the whole geneset with clinical outcome, the major principle components of the expression data of the geneset rather than individual genes can be used to fit an association statistic model along with clinical parameters, especially in the case where the geneset contains many genes[10-11]. In this study, we initially performed Principle Components Analysis (PCA) on expression data for the 13 genes and the top 10 principle components (PC) were applied to Cox proportional hazard model of time to graft loss. The principle components (PC) that were significantly associated with graft loss were selected ($p<0.05$) and the linear combination of eigenvalues of significant components multiplied by the coefficiencies of corresponding PCs from Cox model was used as the geneset risk score (GR-score). The patients were then stratified into 2populations based on geneset risk score (GR-score) for Kaplan-Meier survival analysis. Finally the time-dependent ROC for graft loss prediction within 2 or 3 yrs post-transplant was plotted and the AUCs calculated. The demographic and clinical variables, including 3-month estimated glomerular filtration rate (m3_eGFR), acute cellular rejection at- or before 3-months (pre_or_m3_ACR), CADI-3, cold ischemia time (CIT, minutes), deceased donor status, the presence of Anti-HLA antibodies, induction therapy, recipient race,donor age, delayed graft function and HLA mismatch were also fitted in Cox proportional hazard model of time to graft loss to investigate if the demographic or clinical variables were associated with graft loss.

Validation of Geneset:

We also validated our final optimal geneset on 2independent public datasets. Both public datasets were on the Affymetrix GeneChip platform HU430plus2 (GSE21374[12], GSE25902[13]). The raw data of these public datasets were processed separately in Affymetrix Expression Console using the RMA normalization approach. The expression data for each of the genes in our final optimal geneset was extracted and a prediction model with the geneset was built on each individual dataset. Predictions of clinical data (graft loss post biopsy at any time for GSE21374, and progressor/non-progressor based on CADI score for GSE25902) was performed using the penalized logistic regression model. AUC scores for each of these 2 datasets were calculated from the ROC curves for prediction specificity over sensitivity. We also performed time to graft loss analysis on dataset1 (GSE21374) using the same approach as that for GOCAR dataset.

qPCR experiments

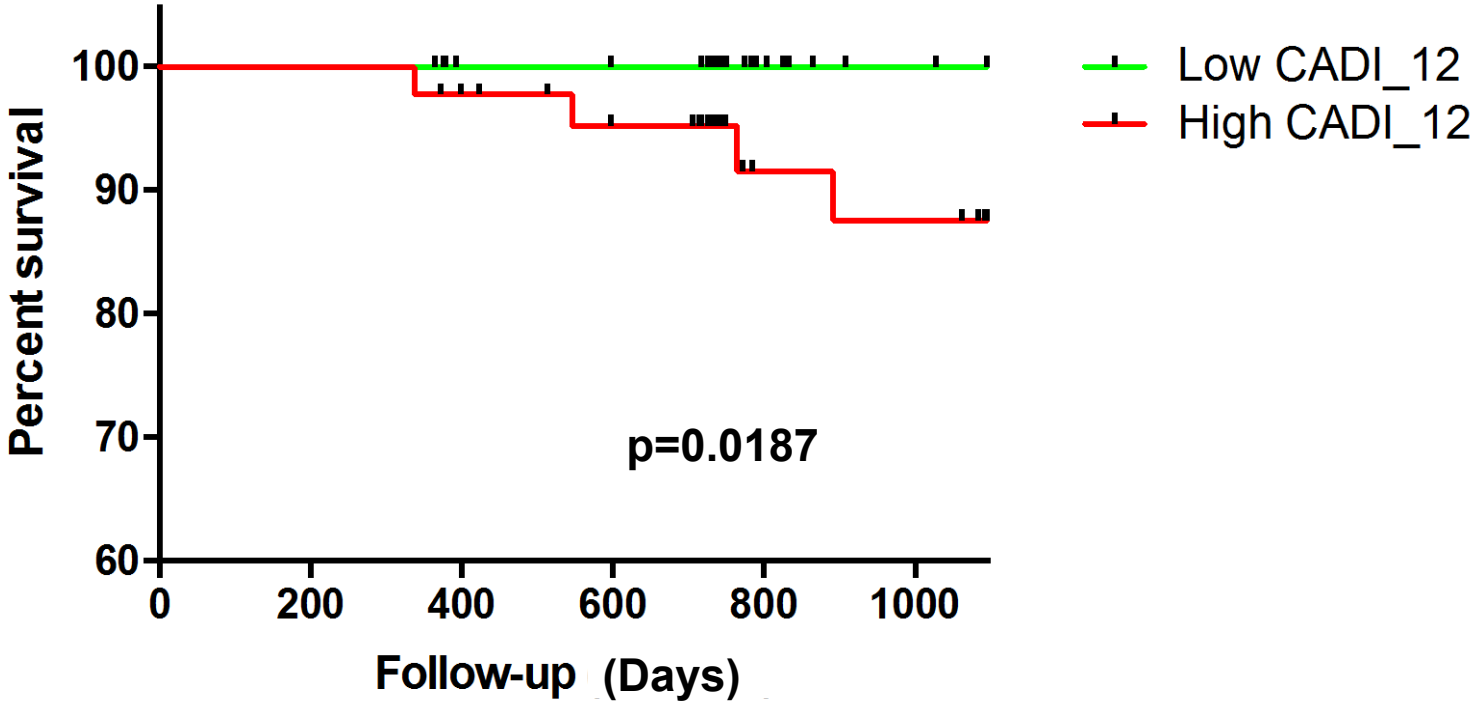
Total RNA was extracted from graft biopsy samples of 45 independent GOCAR patients (N=18: CADI-12 ≥2, and N=27:CADI-12 <2) using Allprep kit (QIAGEN-ALLprep kit, Valencia, CA USA). cDNA was synthesized using AffinityScript RT kit with oligo DT primers (Agilent Inc. Santa Clara, CA). TaqMan qPCR assays for the 13 geneset, 2 house-keeping genes (ACTB, GAPDH) and 18s were purchased from ABI Life Technology (Grand Island, NY). qPCR experiments were performed on cDNA using TAQMAN universal mix and PCR reaction was monitored and acquired using an ABI7900HT system. Samples were measured in triplicates. Cycle Times (CT)values for the prediction geneset as well as the 3 housing genes were generated. The ΔCT value of each gene was calculated by subtracting the average CT value for the house-keeping genes from the CT value of each gene and penalized logistic regression fitting model was then applied on ΔCT values for prediction of the high and low CADI-12 in 45 patients and AUC score was then calculated as described above.

Supplementary results

Intragraft molecular phenotype is time dependent.

Gene expression profiles from m3 biopsies were analyzed by correlation analysis and Geneset Enrichment Analysis (GSEA) to understand molecular mechanisms of IF/TA (Figure S2) (n=159). We identified 1127 genes significantly correlated with CADI-3 (716 positively and 411 negatively) and 1,143 genes with CADI-12 (914 positively and 229 negatively) at a cutoff unadjusted $p < 0.05$, (Figure S3a, S3b and S3c). Only 230 genes (20.4%) correlated with both CADI-3 and CADI-12 (Figure S3a). Gene Ontology enrichment indicated that the transcripts specifically associated with CADI-3 alone were related to alloimmunity, including T-cell activation; while genes involved in programmed cell death/apoptosis and cell adhesion were associated with CADI-12 alone (Figure S4a). By enrichment analysis of immune cell type genes, we observed that dendritic cell genes were specifically associated with CADI-3; however stromal cell genes (mostly fibroblast cells) were the most significantly associated with CADI-12 in addition to macrophage, dentritic and CD4 T cell related genes (Figure S4b). Biological functions were further confirmed by GSEA method in which gene expression data in GO categories were compared between patients with high (≥2) and low (<2) CADI at 3- or 12-months (Figure S5).

Figure S1



High CADI-12:	44	44	42	38	24	22
Low CADI-12:	57	57	54	52	34	28

Figure S2

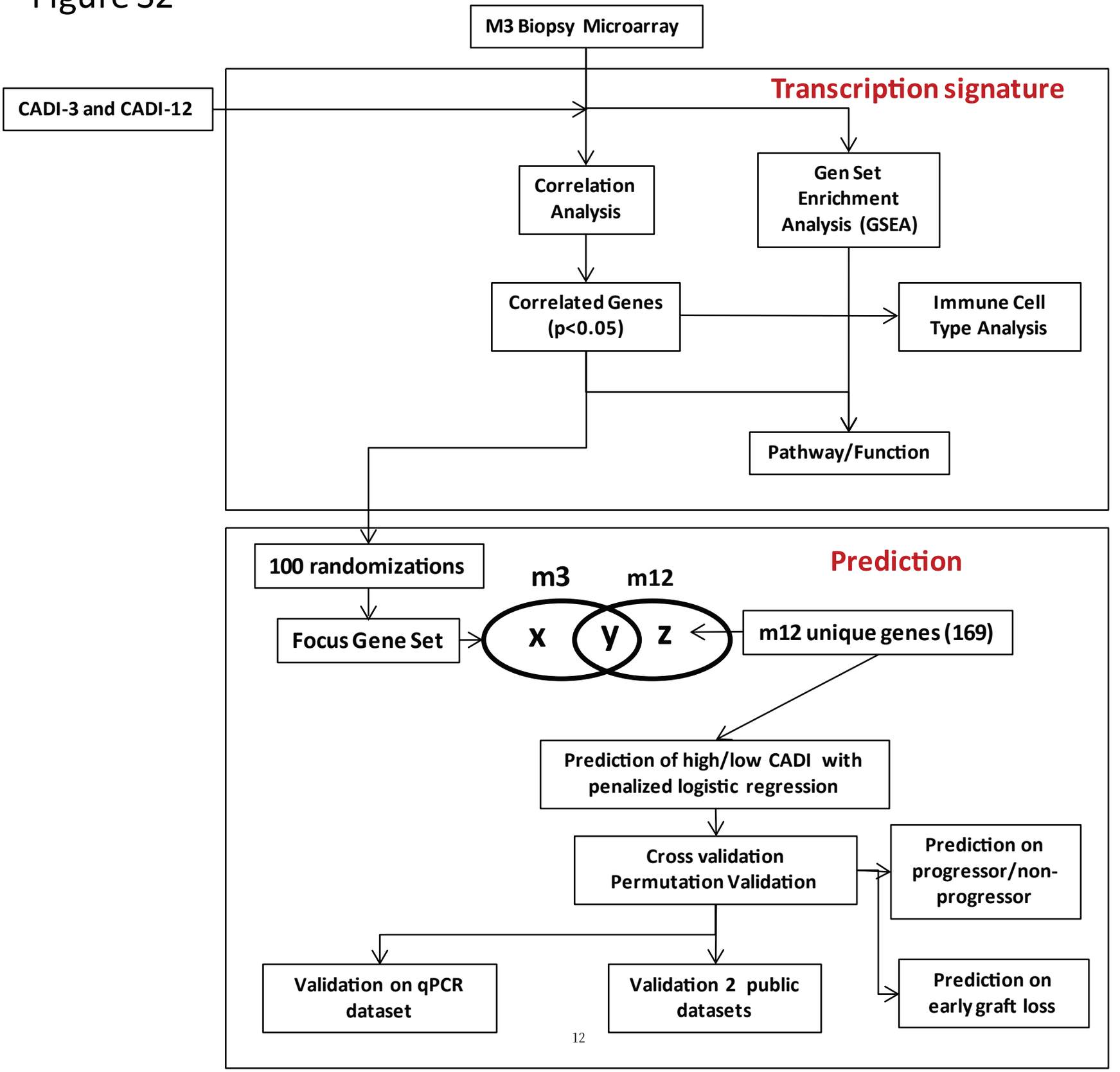

M3 Biopsy Microarray
CADI-3 and CADI-12
Transcription signature
Correlation Analysis
Gen Set Enrichment Analysis (GSEA)
Correlated Genes (p<0.05)
Immune Cell Type Analysis
Pathway/Function
100 randomizations
Prediction
m3
m12
Focus Gene Set
x
y
z
m12 unique genes (169)
Prediction of high/low CADI with penalized logistic regression
Cross validation Permutation Validation
Prediction on progressor/non-progressor
Validation on qPCR dataset
Validation 2 public datasets
Prediction on early graft loss

Figure S3a

CADI	Spearman Correlation P<0.05	Genes	Top 25 Genes
3 month CADI	Positive	716	IRF8, ALOX5, GABRP, GAPT, IKZF1, NUF2, PPM1M, ARNTL2, TOR1A, CPA3, BTK, CD180, ARL4C, CD200R1, PTPN22, ABCA12, APBB1IP, ITGB2, IGJ, C17orf87, ARG1, GNA15, GCNT1, AIM2, CLSPN
	Negative	411	USP2, FLJ42875, GAD1, B3GNT9, NAPB, LOC100130232, CA12, EGF, SLC4A9, TTC18, ADHFE1, ZNF793, NCOA6, SERPINA5, F11, PER3, ZNF385B, RALGPS1, FREM1, FRAS1, SALL1, ATXN7L1, WNK1, NEDD4L, CYP4Z1
12 month CADI	Positive	914	MACC1, LAMC2, CHCHD10, CCL2, GABRP, DUSP6, PROM1, CDC42SE2, ARL4C, SLC34A2, ARPC5, GCNT3, DLGAP1, KLHL13, FJX1, LIX1, TES, HS3ST1, SFN, MET, ITGB6, CLU, HPGDS, CYR61, FAM110B
	Negative	229	NHLH1, GPRC6A, TRPM6, LCE2D, SLC22A23, SORCS1, C11orf53, RXRA, PDK4, C7orf53, KANK1, KRT9, IRGC, C9orf84, NCKAP5, GOLGA2B, PCSK2, MGC13053, KREMEN1, TACC2, OR10A5, CCDC90A, PRM1, MT1B, C5orf45

Figure S3b

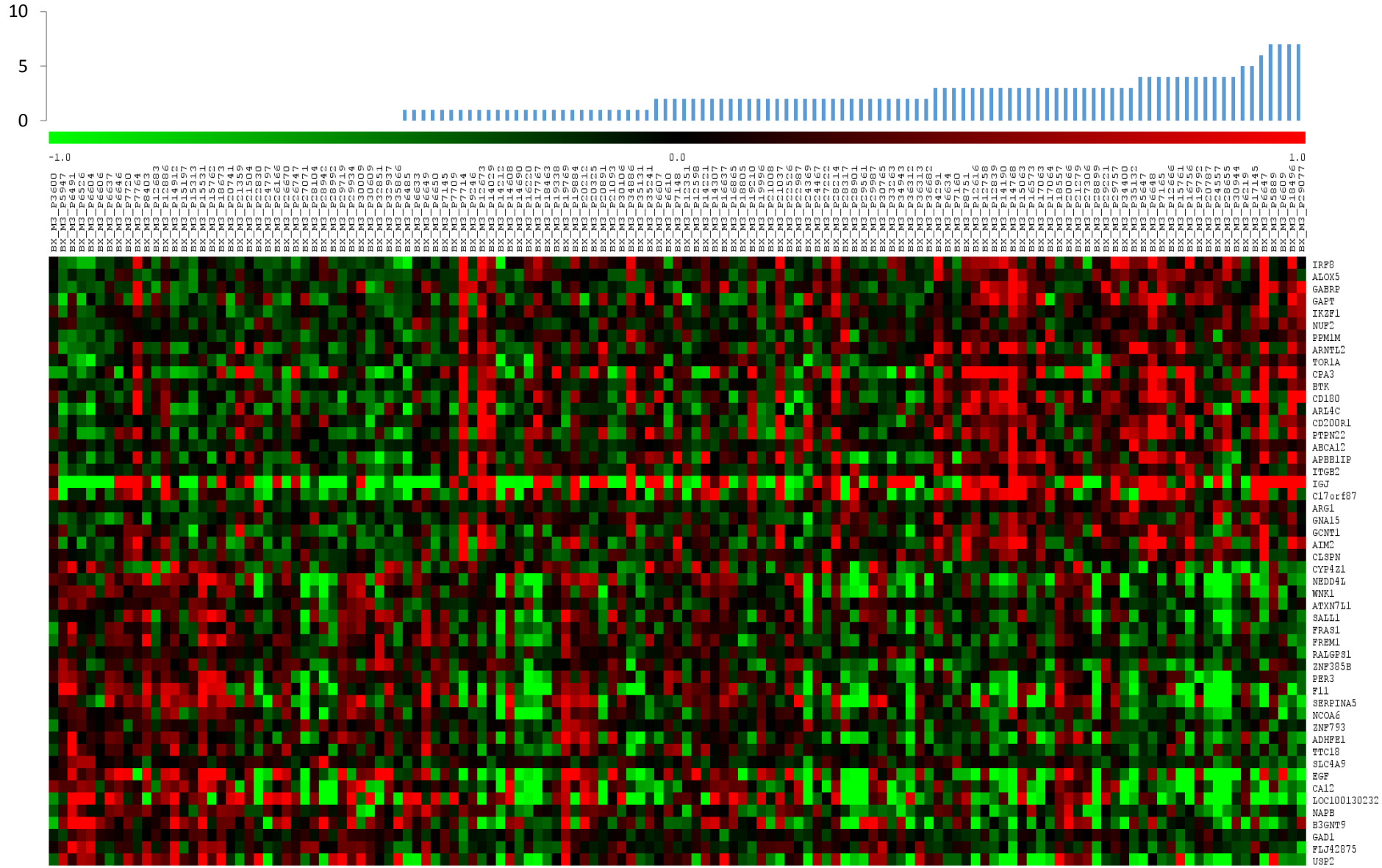

10
5
0
-1.0
0.0
1.0
IRF8
ALOX5
GABRP
GAPT
IKZF1
NUF2
PPM1M
ARNTL2
TOR1A
CPA3
BTK
CD180
ARL4C
CD200R1
PTPN22
ABCA12
APBB1IP
ITGB2
IGJ
C17orf87
ARG1
GNA15
GCNT1
AIM2
CLSPN
CYP4Z1
NEDD4L
WNK1
ATXN7L1
SALL1
FRAS1
FREM1
RALGPS1
ZNF385B
PER3
F11
SERPINA5
NCOA6
ZNF793
ADHFE1
TTC18
SLC4A9
EGF
CA12
LOC100130232
NAPB
B3GNT9
GAD1
FLJ42875
USP2

Figure S3c

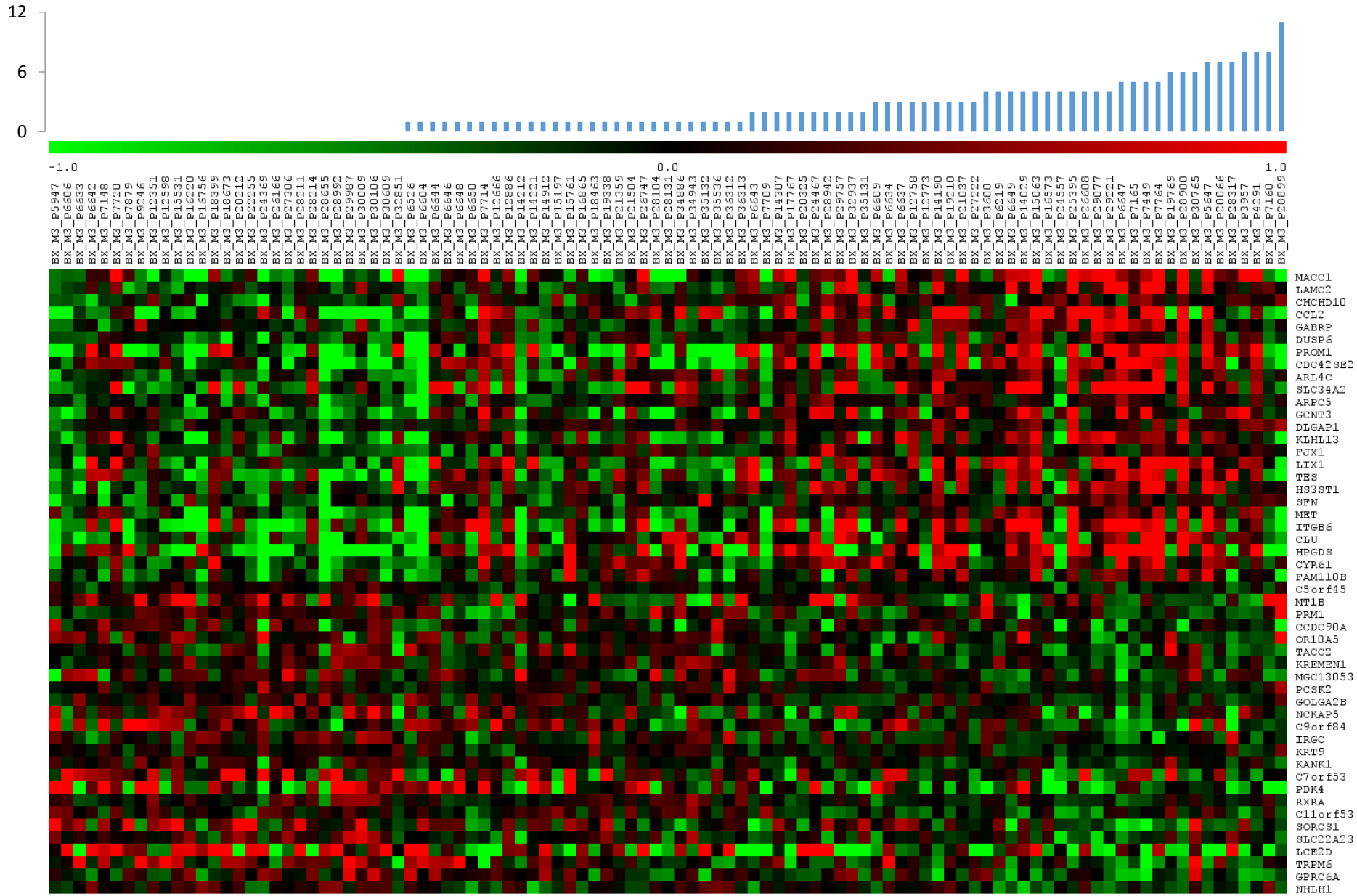

12
6
0
-1.0
0.0
1.0
BX_M3_P5947
BX_M3_P6606
BX_M3_P6633
BX_M3_P6642
BX_M3_P7148
BX_M3_P7720
BX_M3_P7879
BX_M3_P9246
BX_M3_P12351
BX_M3_P12598
BX_M3_P15531
BX_M3_P16220
BX_M3_P16756
BX_M3_P18399
BX_M3_P18673
BX_M3_P20212
BX_M3_P22255
BX_M3_P24369
BX_M3_P26166
BX_M3_P27306
BX_M3_P28211
BX_M3_P28214
BX_M3_P28655
BX_M3_P28992
BX_M3_P29987
BX_M3_P30009
BX_M3_P30106
BX_M3_P30609
BX_M3_P32851
BX_M3_P65c6
BX_M3_P6604
BX_M3_P6644
BX_M3_P6646
BX_M3_P6648
BX_M3_P6650
BX_M3_P7714
BX_M3_P12666
BX_M3_P12886
BX_M3_P14212
BX_M3_P14221
BX_M3_P14912
BX_M3_P15197
BX_M3_P15761
BX_M3_P16865
BX_M3_P18463
BX_M3_P19338
BX_M3_P21359
BX_M3_P21504
BX_M3_P26747
BX_M3_P28104
BX_M3_P28131
BX_M3_P34886
BX_M3_P34943
BX_M3_P35132
BX_M3_P35536
BX_M3_P36312
BX_M3_P36313
BX_M3_P6643
BX_M3_P7709
BX_M3_P14307
BX_M3_P17767
BX_M3_P20325
BX_M3_P24467
BX_M3_P28942
BX_M3_P29757
BX_M3_P32937
BX_M3_P35131
BX_M3_P6609
BX_M3_P6634
BX_M3_P6637
BX_M3_P12758
BX_M3_P12773
BX_M3_P14190
BX_M3_P19210
BX_M3_P21037
BX_M3_P27222
BX_M3_P3600
BX_M3_P6219
BX_M3_P6649
BX_M3_P14029
BX_M3_P15063
BX_M3_P16573
BX_M3_P24557
BX_M3_P25395
BX_M3_P26608
BX_M3_P29077
BX_M3_P29221
BX_M3_P6647
BX_M3_P7165
BX_M3_P7449
BX_M3_P7764
BX_M3_P19769
BX_M3_P28900
BX_M3_P30765
BX_M3_P5647
BX_M3_P20066
BX_M3_P28317
BX_M3_P3957
BX_M3_P4291
BX_M3_P7160
BX_M3_P28899
MACC1
LAMC2
CHCHD10
CCL2
GABRP
DUSP6
PROM1
CDC42SE2
ARL4C
SLC34A2
ARPC5
GCNT3
DLGAP1
KLHL13
FJX1
LIX1
TES
HS3ST1
SFN
MET
ITGB6
CLU
HPGDS
CYR61
FAM110B
C5orf45
MT1B
PRM1
CCDC90A
OR10A5
TACC2
KREMEN1
MGC13053
PCSK2
GOLGA2B
NCKAP5
C9orf84
IRGC
KRT9
KANK1
C7orf53
PDK4
RXRA
C11orf53
SORCS1
SLC22A23
LCE2D
TRPM6
GPRC6A
NHLH1

Figure S4a

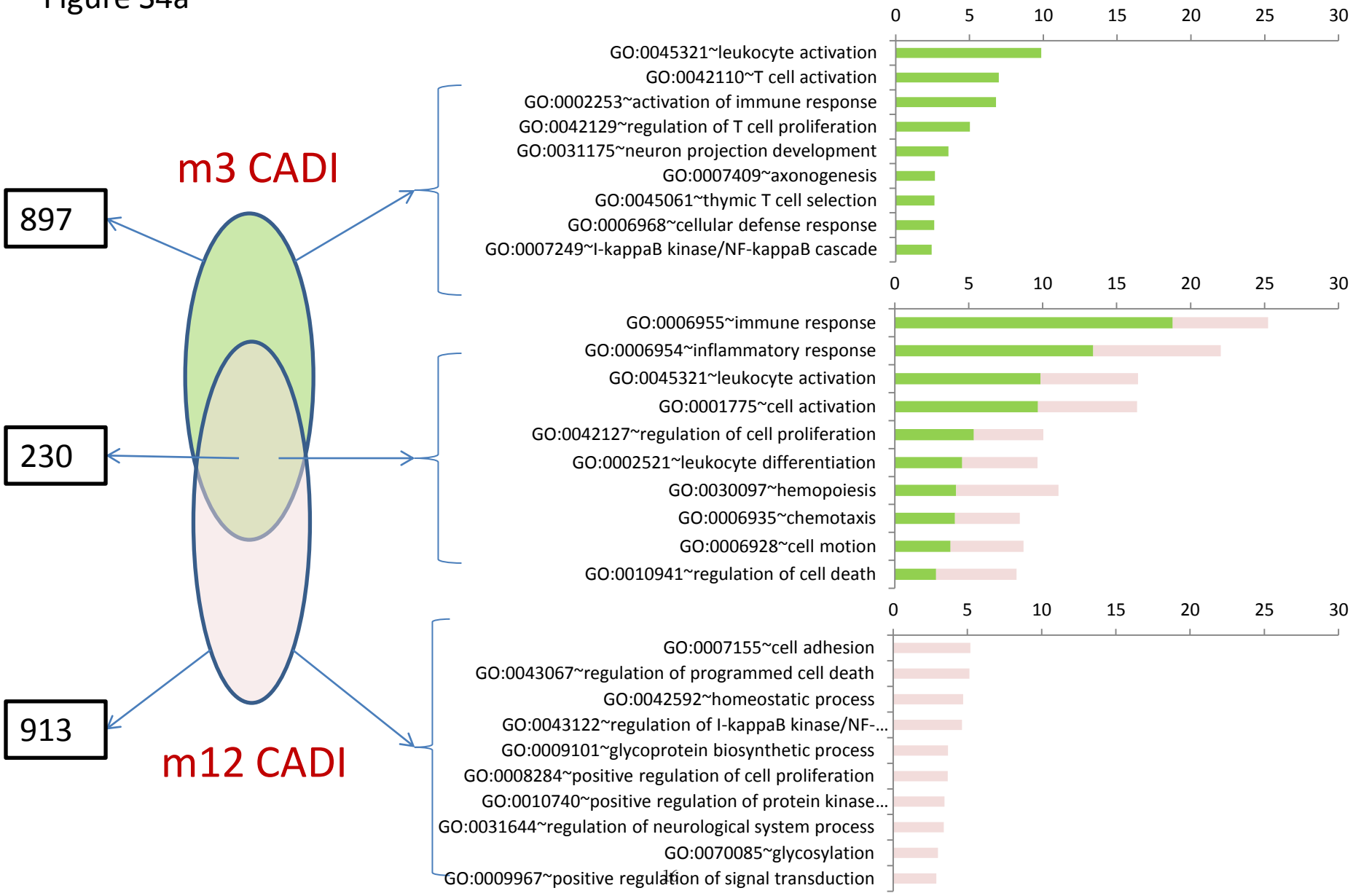
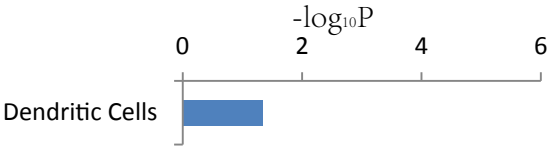
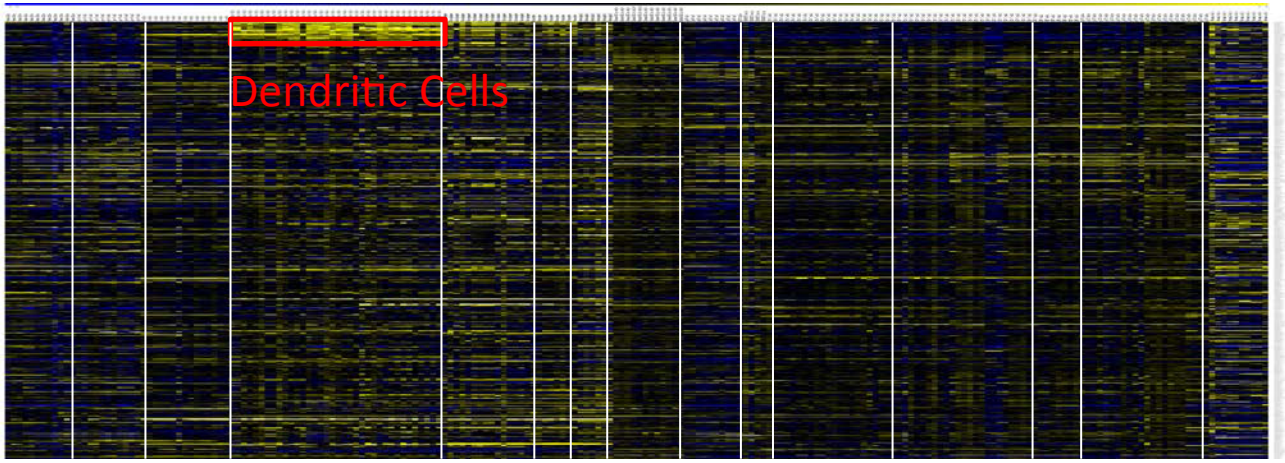

m3 CADI
897
230
913
m12 CADI
0
5
10
15
20
25
30
GO:0045321~leukocyte activation
GO:0042110~T cell activation
GO:0002253~activation of immune response
GO:0042129~regulation of T cell proliferation
GO:0031175~neuron projection development
GO:0007409~axonogenesis
GO:0045061~thymic T cell selection
GO:0006968~cellular defense response
GO:0007249~I-kappaB kinase/NF-kappaB cascade
GO:0006955~immune response
GO:0006954~inflammatory response
GO:0045321~leukocyte activation
GO:0001775~cell activation
GO:0042127~regulation of cell proliferation
GO:0002521~leukocyte differentiation
GO:0030097~hemopoiesis
GO:0006935~chemotaxis
GO:0006928~cell motion
GO:0010941~regulation of cell death
GO:0007155~cell adhesion
GO:0043067~regulation of programmed cell death
GO:0042592~homeostatic process
GO:0043122~regulation of I-kappaB kinase/NF-...
GO:0009101~glycoprotein biosynthetic process
GO:0008284~positive regulation of cell proliferation
GO:0010740~positive regulation of protein kinase...
GO:0031644~regulation of neurological system process
GO:0070085~glycosylation
GO:0009967~positive regulation of signal transduction

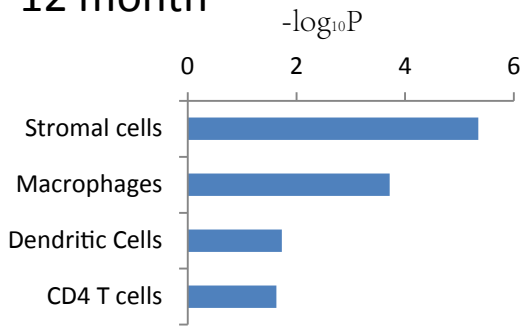
Figure s4b

3 month





12 month



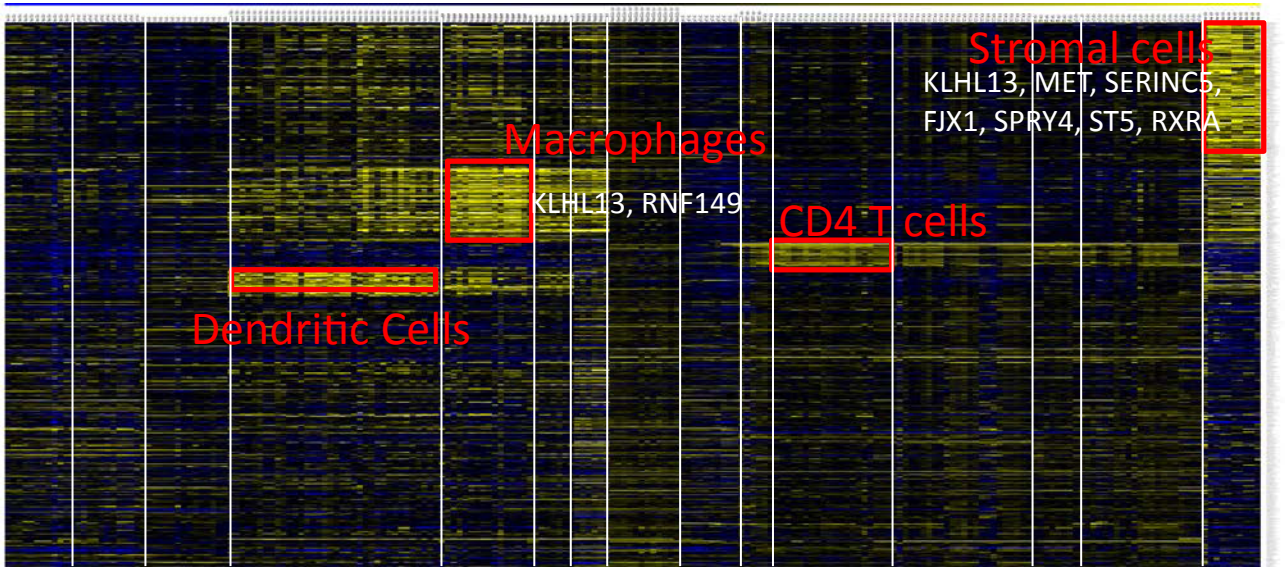
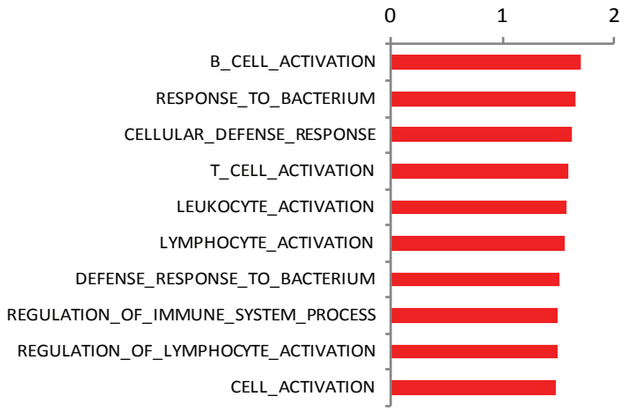
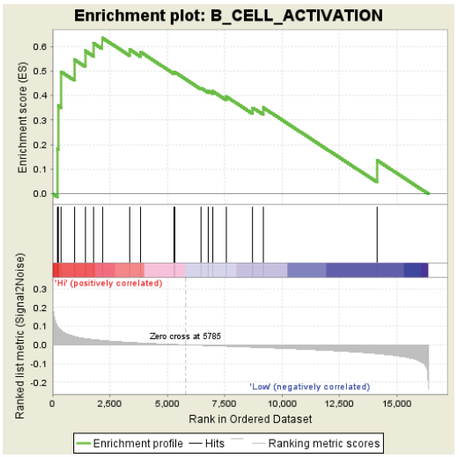
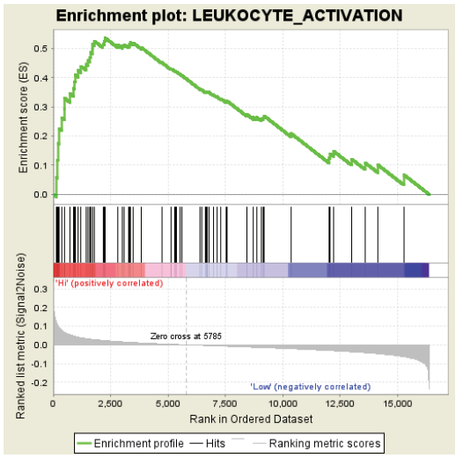


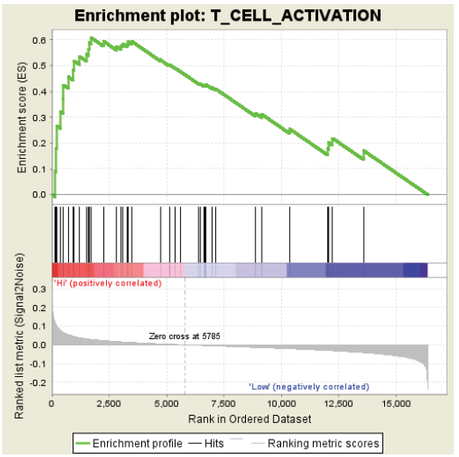
Figure S5

a)

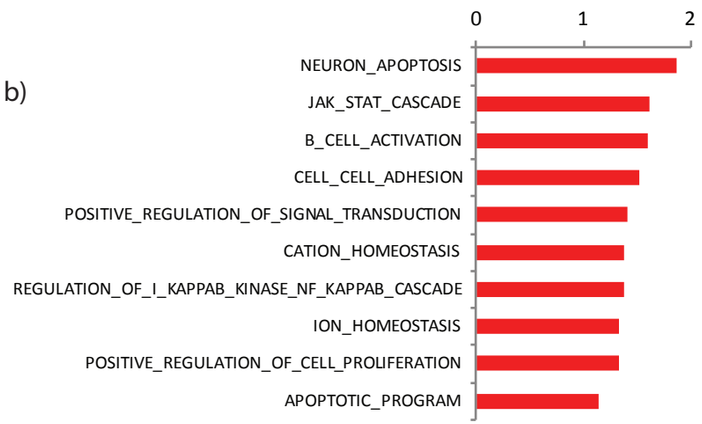

0
1
2
B_CELL_ACTIVATION
RESPONSE_TO_BACTERIUM
CELLULAR_DEFENSE_RESPONSE
T_CELL_ACTIVATION
LEUKOCYTE_ACTIVATION
LYMPHOCYTE_ACTIVATION
DEFENSE_RESPONSE_TO_BACTERIUM
REGULATION_OF_IMMUNE_SYSTEM_PROCESS
REGULATION_OF_LYMPHOCYTE_ACTIVATION
CELL_ACTIVATION

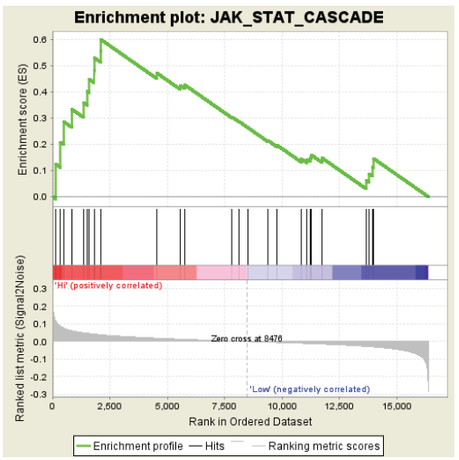

Enrichment plot: B_CELL_ACTIVATION
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'Hi' (positively correlated)
Zero cross at 5785
'Low' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

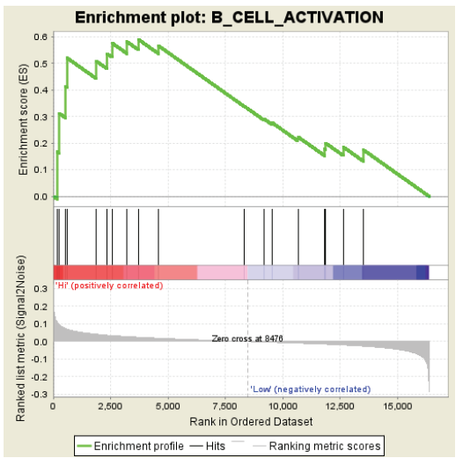

Enrichment plot: LEUKOCYTE_ACTIVATION
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'Hi' (positively correlated)
Zero cross at 5785
'Low' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores


Enrichment plot: T_CELL_ACTIVATION
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'Hi' (positively correlated)
Zero cross at 5785
'Low' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

b)


0
1
2
NEURON_APOPTOSIS
JAK_STAT_CASCADE
B_CELL_ACTIVATION
CELL_CELL_ADHESION
POSITIVE_REGULATION_OF_SIGNAL_TRANSDUCTION
CATION_HOMEOSTASIS
REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE
ION_HOMEOSTASIS
POSITIVE_REGULATION_OF_CELL_PROLIFERATION
APOPTOTIC_PROGRAM


Enrichment plot: JAK_STAT_CASCADE
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'Hi' (positively correlated)
Zero cross at 8476
'Low' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores


Enrichment plot: B_CELL_ACTIVATION
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'Hi' (positively correlated)
Zero cross at 8476
'Low' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

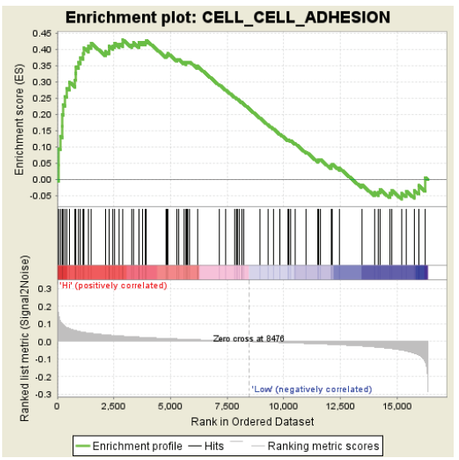
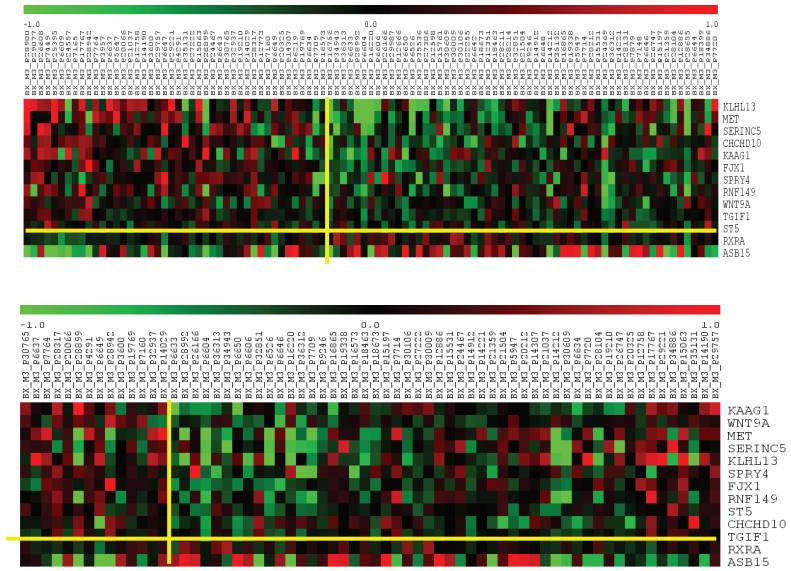
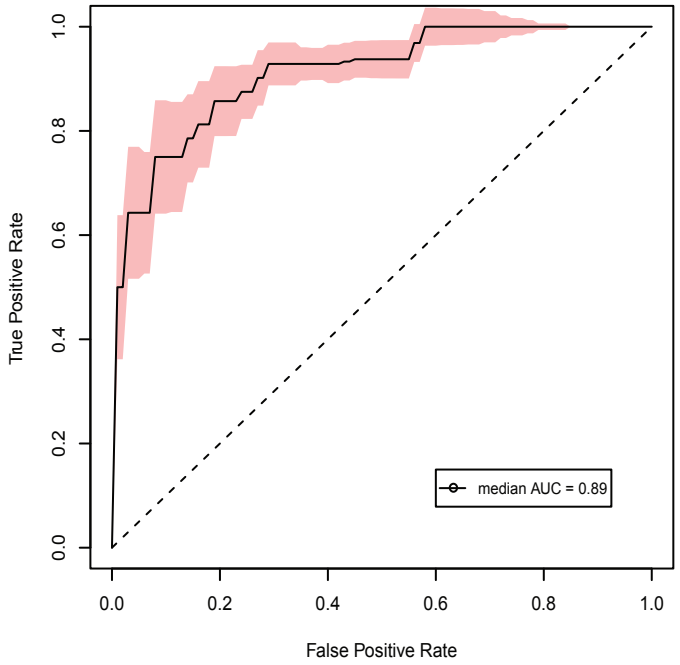

Enrichment plot: CELL_CELL_ADHESION
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'Hi' (positively correlated)
Zero cross at 8476
'Low' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

Figure S6

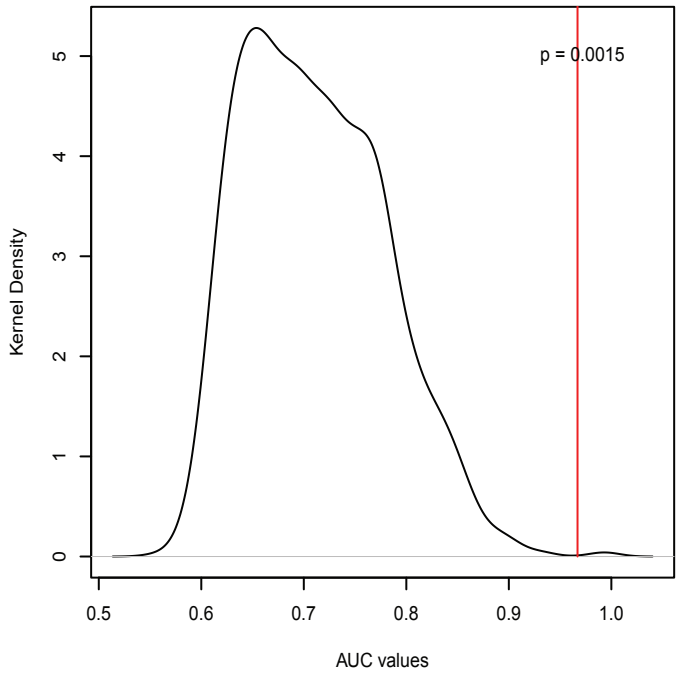
a)



b)



c)



d)

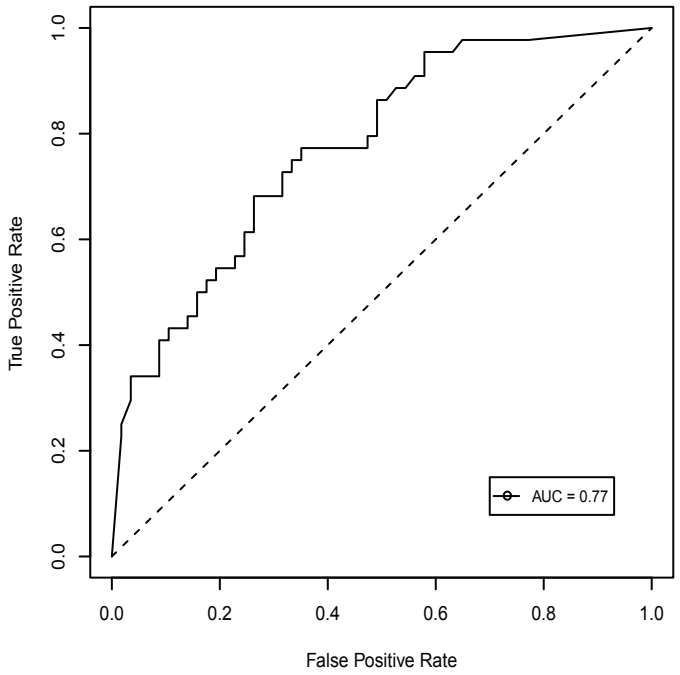
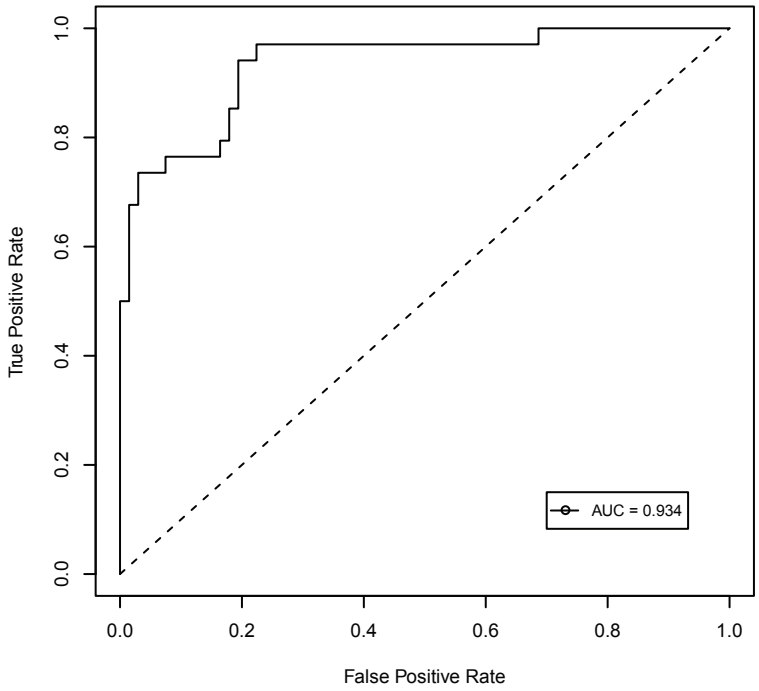


Figure S7

a)


True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.934

b)

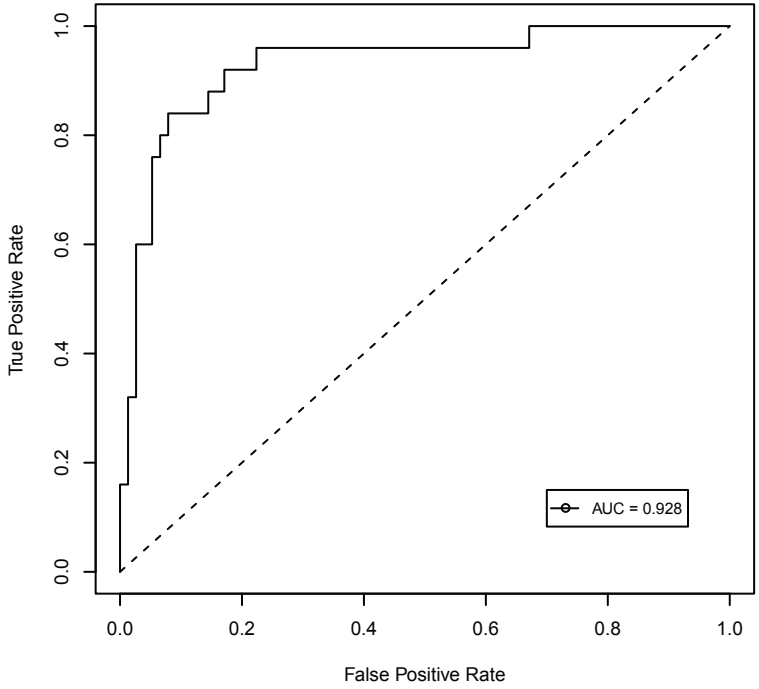

True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.928

Figure S8

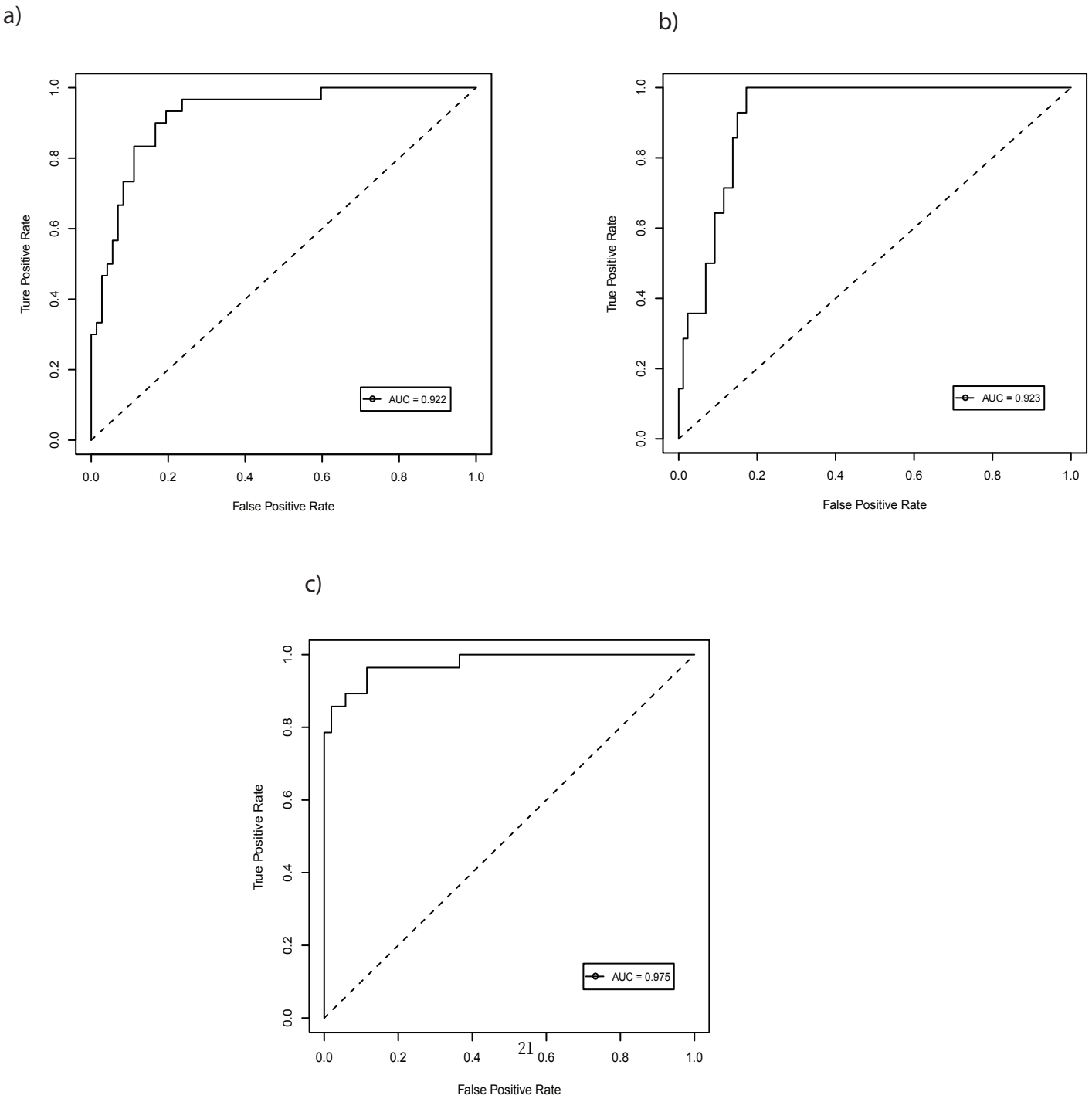
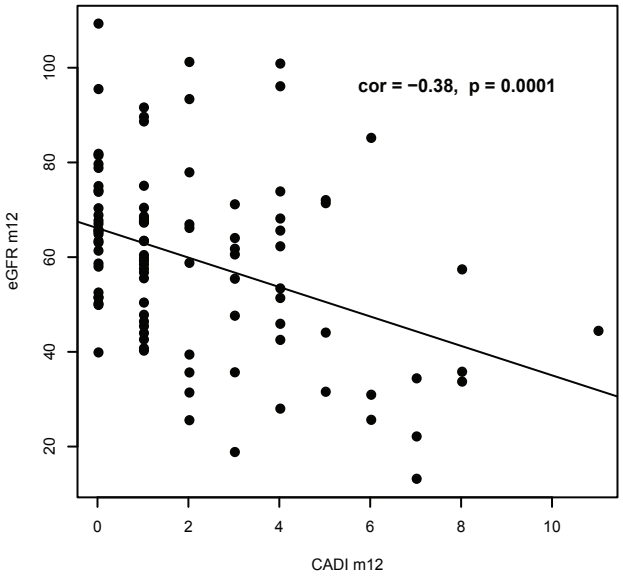
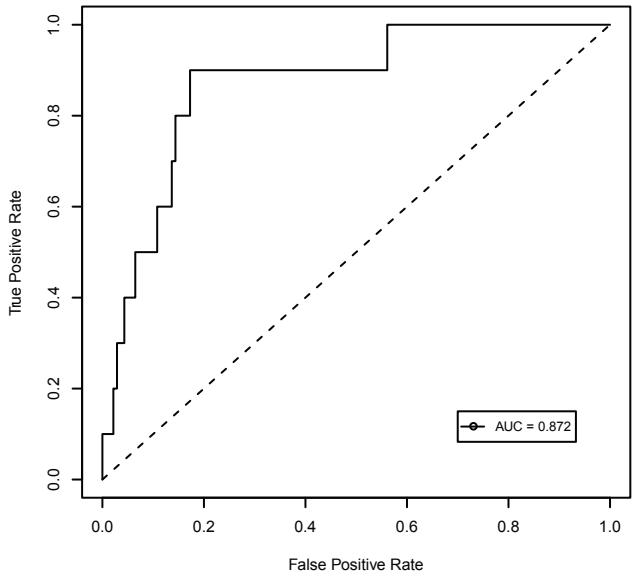

a)
Ture Positive Rate
False Positive Rate
AUC = 0.922
b)
True Positive Rate
False Positive Rate
AUC = 0.923
c)
True Positive Rate
False Positive Rate
AUC = 0.975
0.0
0.2
0.4
0.6
0.8
1.0

Figure S9

a)



b)



c)

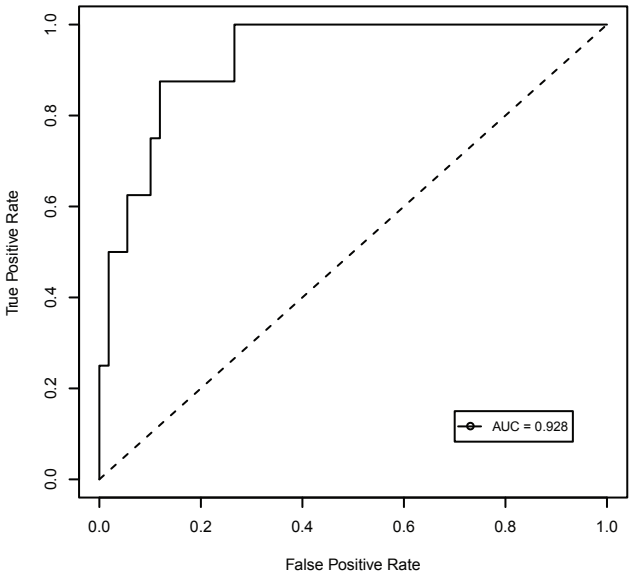
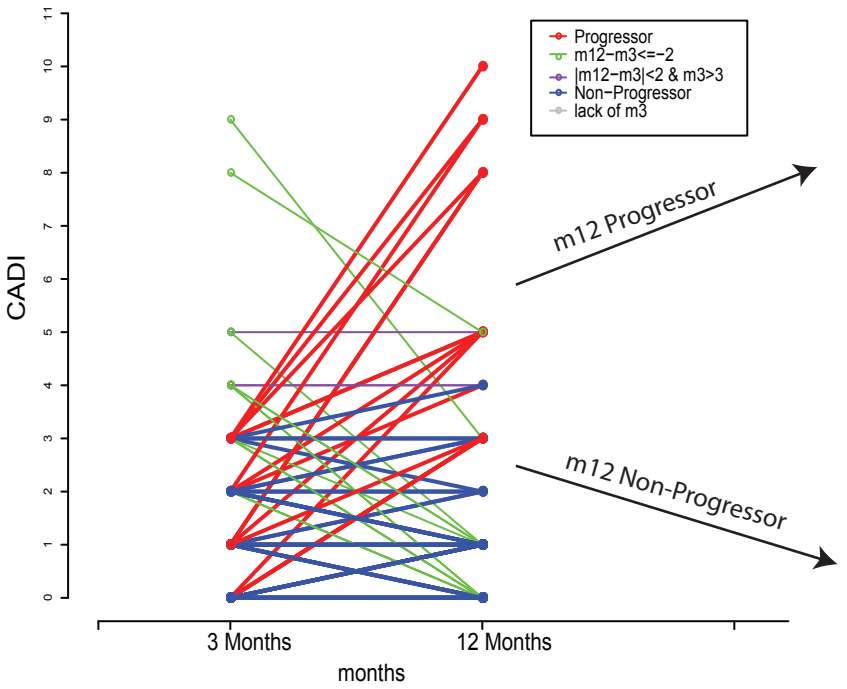
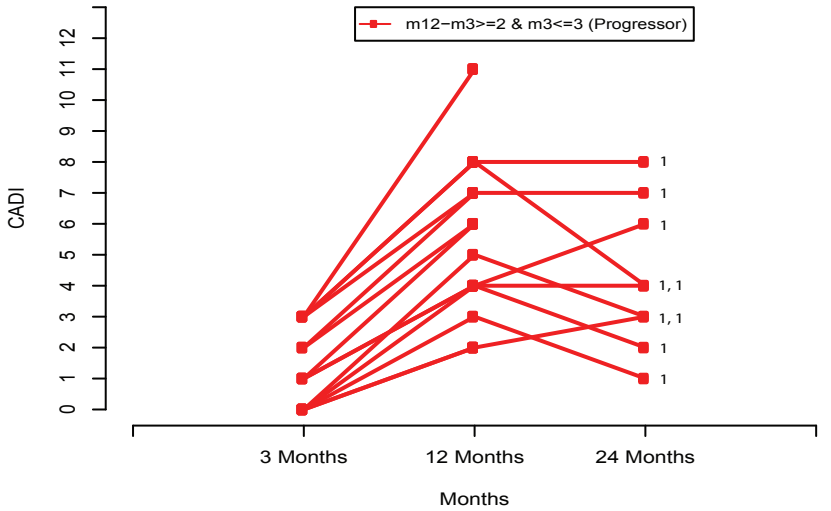
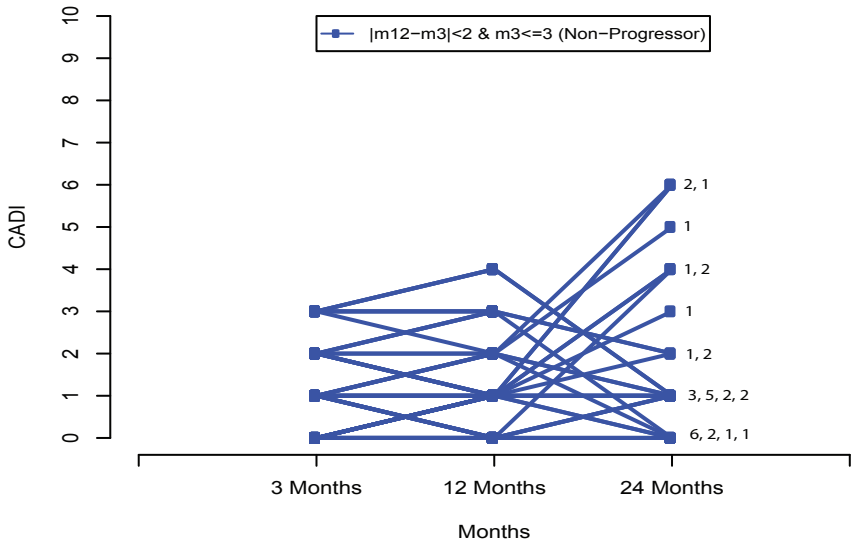


Figure S10

a)







b)

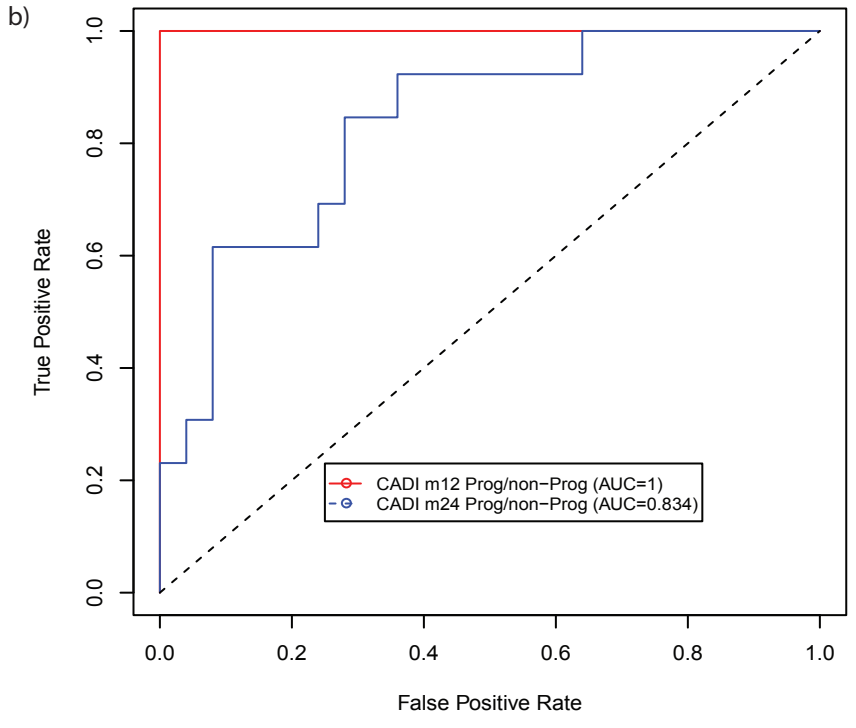


Figure S11

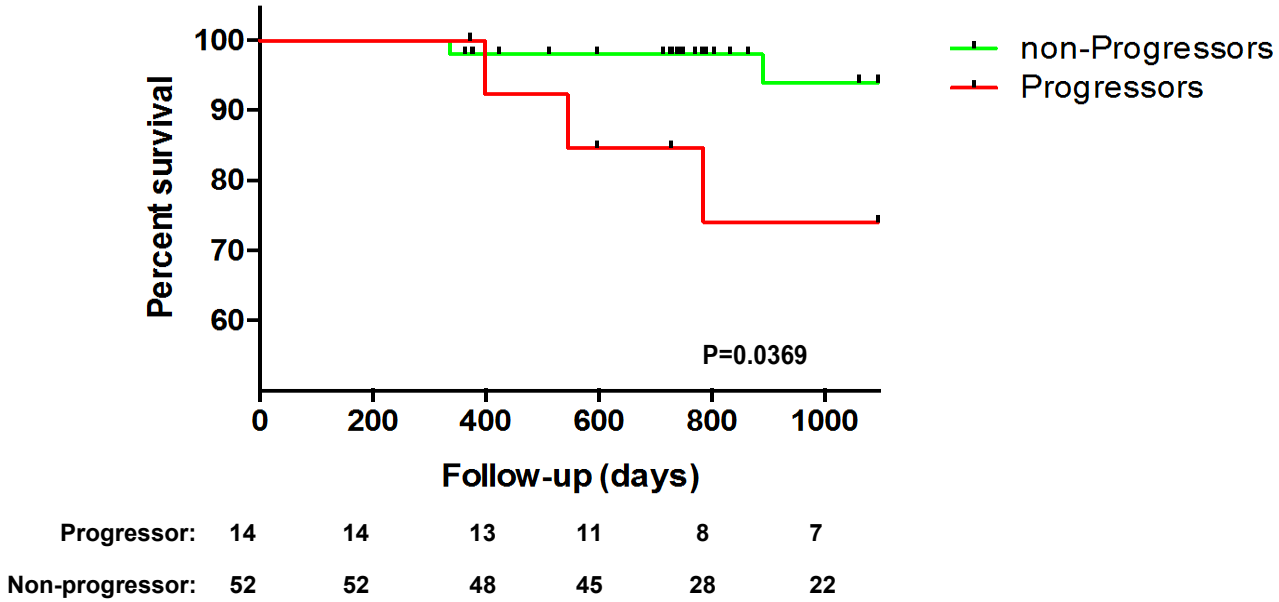

Percent survival
100
90
80
70
60
P=0.0369
0
200
400
600
800
1000
Follow-up (days)
non-Progressors
Progressors
Progressor: 14 14 13 11 8 7
Non-progressor: 52 52 48 45 28 22

Table S1 : Baseline clinical and demographic characteristics for GoCAR patient cohorts.

Characteristics:	Microarray patients (n=101) Mean ± SD (%)	RT-PCR patients (n=45) Mean ± SD (%)	*P-value
Recipient age	46.90±12.38	46.81±11.52	0.69
Recipient race			0.96
Caucasian	66 (65.35)	33 (73.33)	
African American	15 (14.85)	3 (6.67)	
Other	20 (19.80)	9 (20.0)	
Recipient ESRD diagnosis			0.63
Diabetic nephropathy	33 (32.67)	10 (22.22)	
Hypertension	17 (16.83)	6 (13.33)	
Glomerulonephritis	22 (21.78)	15 (33.33)	
Polycystic Kidney	13 (12.87)	5 (11.11)	
Other	16 (15.84)	9 (20.00)	
Donor age	40.73±16.80	44.87±14.68	0.16
Donor race			0.72
Caucasian	80 (79.21)	42 (93.33)	
Non-Caucasian	21 (20.79)	3 (6.67)	
h/o Delayed graft function	9 (8.91)	5 (11.11)	0.54
Anti-HLA antibodies**	26 (26.8)	19 (42.2)	0.08
Class-I	26 (26.8)	19 (42.2)	
Class-II	10 (9.9)	12 (26.7)	
3-month eGFR-creatinine	59.48±18.11	59.27±18.91	0.94
3-month Acute rejection[#]	22 (21.7)	11 (28.94)	0.50
High/low CADI-12	44/57	18/27	0.72
High CADI-12			0.68
- Mean±SD	4.21±2.09	4.00±2.09	
- Median(IQR)	3 (2.75-5.0)	3 (2.75-5.0)	
Low CADI-12			0.99
- Mean±SD	0.49±0.50	0.48±0.51	
- Median(IQR)	0.0 (0.0-1.0)	0.0 (0.0-1.0)	

*P-value by Unpaired T-test (or non-parametric) and, Chi-square/Fisher's exact test.
** 97/101 & 38/45 patients had HLA antibodies measured.

Table S2: The 84 focus geneset

ProbeID	Gene Symbc	Gene Description	Cytoband	mRNA Accessio	CADI Corr	Pvalue
3040518	MACC1	metastasis associated in colon cancer 1	7p21.1	NM_182762	0.411	1.95E-05
3954887	CHCHD10	coiled-coil-helix-coiled-coil-helix domain containing 10	22q11.23	NM_213720	0.404	2.85E-05
3464860	DUSP6	dual specificity phosphatase 6	12q22-q23	NM_001946	0.391	5.20E-05
2761842	PROM1	prominin 1	4p15.32	NM_00114584	0.380	9.03E-05
2721959	SLC34A2	solute carrier family 34 (sodium phosphate), member 2	4p15.3-p15	NM_00117799	0.375	1.11E-04
3596147	GCNT3	glucosaminyl (N-acetyl) transferase 3, mucin type	15q21.3	NM_004751	0.373	1.22E-04
3796620	DLGAP1	discs, large (Drosophila) homolog-associated protein 1	18p11.3	NM_004746	0.369	1.44E-04
4019160	KLHL13	kelch-like 13 (Drosophila)	Xq23-q24	NM_00116830	0.369	1.49E-04
3326826	FJX1	four jointed box 1 (Drosophila)	11p13	NM_014344	0.367	1.60E-04
2868265	LIX1	Lix1 homolog (chicken)	5q15	NM_153234	0.359	2.29E-04
3020192	TES	testis derived transcript (3 LIM domains)	7q31.2	NM_015641	0.357	2.51E-04
3020343	MET	met proto-oncogene (hepatocyte growth factor receptor)	7q31	NM_00112750	0.352	3.01E-04
2583465	ITGB6	integrin, beta 6	2q24.2	NM_000888	0.352	3.09E-04
3129065	CLU	clusterin	8p21-p12	NM_001831	0.349	3.46E-04
2344888	CYR61	cysteine-rich, angiogenic inducer, 61	1p31-p22	NM_001554	0.342	4.60E-04
3167110	ANXA2P2	annexin A2 pseudogene 2	9p13	NR_003573	0.340	5.02E-04
2602770	DNER	delta/notch-like EGF repeat containing	2q36.3	NM_139072	0.340	5.02E-04
2825629	TNFAIP8	tumor necrosis factor, alpha-induced protein 8	5q23.1	NM_014350	0.338	5.45E-04
2974413	MOXD1	monooxygenase, DBH-like 1	6q23.1-q23	NM_015529	0.328	8.11E-04
2864449	SERINC5	serine incorporator 5	5q14.1	NM_00117407	0.318	0.0012
3108489	LAPTM4B	lysosomal protein transmembrane 4 beta	8q22.1	NM_018407	0.318	0.0012
3024025	MEST	mesoderm specific transcript homolog (mouse)	7q32	NM_002402	0.304	0.0020
3662041	OGFOD1	2-oxoglutarate and iron-dependent oxygenase domain containing 1	16q12.2	NM_018233	0.303	0.0021
3605395	ADAMTSL3	ADAMTS-like 3	15q25.2	NM_207517	0.300	0.0023
2876361	PITX1	paired-like homeodomain 1	5q31	NM_002653	0.294	0.0028
3224087	TTLL11	tubulin tyrosine ligase-like family, member 11	9q33.2	NM_00113944	0.287	0.0036
3872335	ZNF416	zinc finger protein 416	19q13.4	NM_017879	0.287	0.0037
3332913	TMEM216	transmembrane protein 216	11q13.1	NM_016499	0.286	0.0037
3888383	SLC9A8	solute carrier family 9 (sodium/hydrogen exchanger), member 8	20q13.13	NM_015266	0.286	0.0037
2669979	CX3CR1	chemokine (C-X3-C motif) receptor 1	3p21|3p21	NM_00117117	0.284	0.0040
2486927	ARHGAP25	Rho GTPase activating protein 25	2p13.3	NM_014882	0.284	0.0040
2435218	TDRKH	tudor and KH domain containing	1q21	NM_00108396	0.283	0.0041

2933392	SYNJ2	synaptojanin 2	6q25.3	NM_003898	0.281	0.0044
3431892	SH2B3	SH2B adaptor protein 3	12q24	NM_005475	0.281	0.0044
2672140	LTF	lactotransferrin	3p21.31	NM_002343	0.281	0.0045
2567583	RNF149	ring finger protein 149	2q11.2	NM_173647	0.280	0.0046
3734648	SLC16A5	solute carrier family 16, member 5 (monocarboxylic acid transporte	17q25.1	NM_004695	0.277	0.0050
3726154	ITGA3	integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)	17q21.33	NM_002204	0.272	0.0060
3850445	CDKN2D	cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4)	19p13	NM_001800	0.272	0.0060
2899437	BTN2A1	butyrophilin, subfamily 2, member A1	6p22.1	NM_078476	0.270	0.0062
2879105	SPRY4	sprouty homolog 4 (Drosophila)	5q31.3	NM_030964	0.270	0.0062
3623717	FLJ10038	hypothetical protein FLJ10038	15q21.2	NR_026891	0.267	0.0070
3168938	POLR1E	polymerase (RNA) I polypeptide E, 53kDa	9p13.2	NM_022490	0.266	0.0072
2714132	PDE6B	phosphodiesterase 6B, cGMP-specific, rod, beta	4p16.3	NM_000283	0.263	0.0078
2356142	LIX1L	Lix1 homolog (mouse)-like	1q21.1	NM_153713	0.263	0.0080
3232349	PFKP	phosphofructokinase, platelet	10p15.3-p1	NM_002627	0.261	0.0084
2931391	MTHFD1L	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-li	6q25.1	NM_015440	0.259	0.0089
3960061	RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP bi	22q13.1	NM_002872	0.259	0.0089
3261009	KAZALD1	Kazal-type serine peptidase inhibitor domain 1	10q24.31	NM_030929	0.255	0.0099
2315918	ATAD3C	ATPase family, AAA domain containing 3C	1p36.33	NM_00103921	0.254	0.0103
3820443	ICAM1	intercellular adhesion molecule 1	19p13.3-p1	NM_000201	0.253	0.0108
2374982	RNPEP	arginyl aminopeptidase (aminopeptidase B)	1q32	NM_020216	0.252	0.0109
3405587	GPRC5A	G protein-coupled receptor, family C, group 5, member A	12p13-p12	NM_003979	0.250	0.0115
3270270	PTPRE	protein tyrosine phosphatase, receptor type, E	10q26	NM_006504	0.247	0.0126
2359885	SLC27A3	solute carrier family 27 (fatty acid transporter), member 3	1q21.3	NM_024330	0.247	0.0127
3415320	KRT7	keratin 7	12q12-q13	NM_005556	0.246	0.0130
2414958	TACSTD2	tumor-associated calcium signal transducer 2	1p32-p31	NM_002353	0.246	0.0130
3868998	NKG7	natural killer cell group 7 sequence	19q13.41	NM_005601	0.245	0.0136
2361342	SEMA4A	sema domain, immunoglobulin domain (Ig), transmembrane domai	1q22	NM_022367	0.245	0.0137
3776504	TGIF1	TGFB-induced factor homeobox 1	18p11.3	NM_170695	0.244	0.0140
3028217	---	---	---	AK303101	0.243	0.0143
2881187	CSF1R	colony stimulating factor 1 receptor	5q33-q35	NM_005211	0.242	0.0146
2898441	KAAG1	kidney associated antigen 1	6p22.1	NM_181337	0.240	0.0154
3056264	ABHD11	abhydrolase domain containing 11	7q11.23	NR_026912	0.240	0.0156
2621881	P4HTM	prolyl 4-hydroxylase, transmembrane (endoplasmic reticulum)	3p21.31	NM_177938	0.237	0.0168
3185593	BSPRY	B-box and SPRY domain containing	9q32	NM_017688	0.237	0.0169

2369484	TOR3A	torsin family 3, member A	1q25.2	NM_022371	0.237	0.0171
2787902	GYPE	glycophorin E (MNS blood group)	4q31.1	NM_198682	0.237	0.0172
3738471	RAC3	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP bi	17q25.3	NM_005052	0.234	0.0186
2692319	ADCY5	adenylate cyclase 5	3q13.2-q21	NM_183357	0.233	0.0192
3361971	ST5	suppression of tumorigenicity 5	11p15	NM_005418	0.232	0.0197
3865998	PNMAL1	PNMA-like 1	19q13.32	NM_018215	0.220	0.0270
2407985	HEYL	hairy/enhancer-of-split related with YRPW motif-like	1p34.3	NM_014571	0.219	0.0275
2323899	UBXN10	UBX domain protein 10	1p36.12	NM_152376	0.216	0.0300
2459352	WNT9A	wingless-type MMTV integration site family, member 9A	1q42	NM_003395	0.212	0.0332
3051655	VOPP1	vesicular, overexpressed in cancer, prosurvival protein 1	7p11.2	NM_030796	0.210	0.0353
3635903	LOC388152	hypothetical LOC388152	15q25.2	BC054509	-0.213	0.0327
3394412	THY1	Thy-1 cell surface antigen	11q22.3-q2	NM_006288	-0.234	0.0184
3907507	C20orf165	chromosome 20 open reading frame 165	20q13.12	NM_080608	-0.239	0.0160
3021696	ASB15	ankyrin repeat and SOCS box-containing 15	7q31.31	NM_080928	-0.263	0.0079
3645901	NAT15	N-acetyltransferase 15 (GCN5-related, putative)	16p13.3	NM_024845	-0.264	0.0078
2942578	CCDC90A	coiled-coil domain containing 90A	6p24.3-p23	NM_00103171	-0.272	0.0059
3193339	RXRA	retinoid X receptor, alpha	9q34.3	NM_002957	-0.300	0.0023
3305801	SORCS1	sortilin-related VPS10 domain containing receptor 1	10q23-q25	NM_052918	-0.306	0.0019

Table S3a: Multivariate analysis of high CADI-12 prediction with clinical/pathological variables only

Terms	OR	lower	upper	pvalue
Donor_Age	1.023	-0.015	0.061	0.2331
Recipient Race (Caucasian vs Non-Caucasian)	1.486	-0.742	1.534	0.4901
Recipient Gender	0.601	-1.627	0.583	0.3574
Deceased_donor	1.323	-1.233	1.773	0.7116
ECD_kidney	0.321	-3.548	1.238	0.3397
CIT_min	1.001	-0.001	0.004	0.1794
Induction_Therapy	0.774	-1.432	0.904	0.6631
Anti_HLA_Ab	0.856	-1.387	1.022	0.7973
Delayed_Graft_Function	1.719	-1.473	2.669	0.5912
HLA_Mismatch	1.202	-0.111	0.500	0.2242
m3_eGFR	0.971	-0.065	0.001	0.0552
pre_or_m3_ACR	2.977	-0.085	2.342	0.0690
CADI-3	1.212	-0.171	0.590	0.3058

*N=83 patients have complete demographic, clinical and pathological data

Table S3b: Multivariate analysis of high CADI-12 prediction with clinical/ pathological variables and the geneset

Terms	OR	lower	upper	pvalue
Donor_Age	0.997	0.931	1.066	0.9341
Recipient Race (Caucasian vs Non-Caucasian)	0.464	0.041	3.186	0.4433
Recipient Gender	0.709	0.113	6.838	0.7079
Deceased_donor	1.231	0.099	15.878	0.8621
ECD_kidney	0.319	0.003	14.644	0.5771
CIT_min	1.001	0.998	1.006	0.4830
Induction_Therapy	1.214	0.168	10.472	0.8444
Anti_HLA_Ab	0.215	0.010	1.652	0.1463
Delayed_Graft_Function	0.213	0.001	5.963	0.4057
HLA_Mismatch	1.021	0.575	1.989	0.9400
m3_eGFR	0.960	0.871	1.016	0.1696
pre_or_m3_ACR	2.732	0.345	31.021	0.3319
CADI-3	1.466	0.721	3.608	0.2921
Geneset	621.774	49.955	9.74E+04	1.06E-10

*N=83 patients have complete demographic, clinical and pathological data

Table S4: CADI subscore comparisons between 12-month progressors/non-progressors

	3-month			12-month		
Parameter	Progressors Mean±SD	Non-Progressors Mean±SD	*p-value	Progressors Mean±SD	Non-Progressors Mean±SD	*p-value
CADI	1.36±1.28	1.07±1.06	0.45	5.50±2.56	1.25±1.17	<0.0001
ct-score	0.38±0.50	0.46±0.50	0.75	1.71±0.91	0.56±0.50	<0.0001
cv-score	0.22±0.67	0.22±0.59	0.95	0.91±1.09	0.19±0.55	0.0025
ci-score	0.15±0.38	0.15±0.41	0.99	1.93±1.07	0.21±0.45	<0.0001
i-score	0.07±0.27	0.02±0.14	0.36	0.35±0.74	0.07±0.43	0.0439
mm-score	0.0±0.0	0.02±0.1	0.99	0.14±0.53	0.04±0.28	0.39
g-score	0.45±0.82	0.17±0.61	0.09	0.35±0.74	0.08±0.44	0.0478

*Mann-Whitney test

Table S5a: Multivariate analysis of m12 progressor prediction with clinical parameters and geneset

Terms	OR	lower	upper	pvalue
Donor_Age	1.020	-0.025	0.073	0.3869
Race:Caucasian vs Non-Caucasian	2.686	-0.698	2.860	0.2477
Gender	0.421	-2.607	0.708	0.2797
Deceased_donor	4.036	-0.609	3.910	0.1791
ECD_kidney	0.188	-5.168	1.446	0.2938
CIT_min	1.001	-0.001	0.004	0.3131
Induction_Therapy	0.983	-1.658	1.761	0.9833
Anti_HLA_Ab	0.245	-3.581	0.358	0.1227
Delayed_Graft_Function	0.217	-4.557	0.878	0.2236
HLA_Mismatch	0.992	-0.508	0.463	0.9731
m3_eGFR	0.951	-0.116	-0.002	0.0425
pre_or_m3_ACR	5.795	0.118	3.691	0.0353
m3_CADI	0.626	-1.346	0.244	0.2040

*N=63 patients have complete demographic, clinical and pathological data

Table S5b: Multivariate analysis of m12 progressor prediction with clinical parameters and geneset

Terms	OR	lower	upper	pvalue
Donor_Age	0.993	0.911	1.068	0.8191
Race:Caucasian vs Non-Caucasian	0.529	0.021	6.518	0.6238
Gender	0.449	0.026	8.293	0.5410
Deceased_donor	2.093	0.066	148.439	0.6640
ECD_kidney	0.233	0.003	9.497	0.4440
CIT_min	1.001	0.998	1.005	0.6260
Induction_Therapy	0.992	0.102	18.526	0.9942
Anti_HLA_Ab	0.349	0.028	3.138	0.3361
Delayed_Graft_Function	0.079	0.000	2.887	0.2004
HLA_Mismatch	0.836	0.306	1.558	0.5506
m3_eGFR	0.934	0.809	0.991	0.0200
pre_or_m3_ACR	2.133	0.254	30.003	0.4591
m3_CADI	0.530	0.114	1.432	0.2261
Generisk	5.84E+03	23.237	3.25E+09	0.0002

*N=63 patients have complete demographic, clinical and pathological data

Table S5c: Multivariate analysis of m24 progressor prediction with clinical parameters and geneset

Terms	OR	lower	upper	pvalue
Donor_Age	0.993	-0.057	0.042	0.7850
Race:Caucasian vs Non-Caucasian	0.525	-2.106	0.740	0.3601
Gender	1.315	-1.050	1.642	0.6826
Deceased_donor	0.606	-3.013	1.872	0.6784
ECD_kidney	0.649	-4.509	2.914	0.8043
CIT_min	1.000	-0.002	0.003	0.6818
Induction_Therapy	0.441	-2.256	0.541	0.2389
Anti_HLA_Ab	1.186	-1.493	1.783	0.8340
Delayed_Graft_Function	2.057	-1.780	3.125	0.5504
HLA_Mismatch	1.086	-0.304	0.498	0.6746
m3_eGFR	0.980	-0.068	0.021	0.3413
pre_or_m3_ACR	2.062	-0.795	2.260	0.3428
m3_CADI	1.058	-1.033	1.146	0.9179

*N=50 patients have complete demographic, clinical and pathological data

Table S5d: Multivariate analysis of m24 progressor prediction with clinical parameters and geneset

Terms	OR	lower	upper	pvalue
Donor_Age	0.941	0.785	1.050	0.2601
Race:Caucasian vs Non-Caucasian	0.127	0.001	1.534	0.1051
Gender	4.554	0.395	1.38E+04	0.2425
Deceased_donor	0.012	4.26E-07	3.046	0.1184
ECD_kidney	0.034	6.85E-08	99.226	0.4065
CIT_min	1.003	1.000	1.011	0.0889
Induction_Therapy	0.198	0.001	2.772	0.2090
Anti_HLA_Ab	7.356	0.219	740.253	0.2903
Delayed_Graft_Function	3.797	0.001	3.42E+04	0.7457
HLA_Mismatch	1.341	0.533	3.168	0.4341
m3_eGFR	0.974	0.848	1.093	0.6251
pre_or_m3_ACR	2.934	0.080	674.514	0.4715
m3_CADI	3.393	0.311	251.299	0.2903
Generisk	7.51E+03	56.185	3.98E+12	2.22E-06

*N=50 patients have complete demographic, clinical and pathological data

Table S6: Association of 10 principle components of 13 geneset with graft loss in Cox proportional hazard model

	coef	exp(coef)	se(coef)	z	p	
PC1	-5.592	3.73E-03	29.95	-0.1867	0.8500	
PC2	-7.413	6.03E-04	6.37	-1.1642	0.2400	
PC3	-2.874	5.65E-02	8.23	-0.3494	0.7300	
PC4	-14.75	3.93E-07	5	-2.9512	0.0032	*
PC5	0.783	2.19E+00	4.02	0.1948	0.8500	
PC6	13.701	8.92E+05	5.45	2.5141	0.0120	*
PC7	6.738	8.44E+02	4.11	1.6389	0.1000	
PC8	0.224	1.25E+00	3.72	0.0601	0.9500	
PC9	-0.749	4.73E-01	4.95	-0.1513	0.8800	
PC10	4.418	8.29E+01	4.07	1.0853	0.2800	

Likelihood ratio test=20.1 on 10 df, p=0.0287 n= 159, number of events= 11

Table S7: Association of demographic or clinical variables with graft loss in Cox proportional hazard mode

Variable	Coef	Exp(coef)	SE(coef)	Z	P
m3_eGFR	-0.016	0.984	0.027	-0.598	0.5500
pre_or_m3_ACR	2.475	11.877	1.186	2.086	0.0370
CADI-3	-0.329	0.720	0.264	-1.245	0.2100
CIT_min	2.6E-04	1.000	0.001	0.263	0.7900
Deceased_donor (yes vs no)	1.439	4.218	1.434	1.003	0.3200
Anti_HLA_Ab	-0.719	0.487	1.162	-0.619	0.5400
Induction_Type (yes vs no)	0.233	1.263	1.207	0.193	0.8500
Recipient Race (Caucasian vs non-Caucasian)	-1.150	0.317	1.004	-1.146	0.2500
Donor Age	5.3E-03	1.005	0.024	0.217	0.8300
HLA mismatch	-0.051	0.950	0.470	-0.108	0.9100
Delayed Graft Function	1.028	2.797	0.979	1.051	0.2900

Anti_HLA_Ab : Yes (donor specific antigen or non dsa antibody), No : no antibody

Induction type: Yes (Lymphocyte Depletion or Lymphocyte Non-Depletion); No; Induction

Likelihood ratio test=11.5 on 11 df, p=0.4010 n= 120, number of events= 7

Table S8: Validation of the GoCAR gene set in other kidney transplant cohorts.

Data Set	Genechip / Platform	Sample Size	Outcome	AUC	Ref.
Dataset 1	Affymetrix U133 Plus 2.0	282	Allograft loss	0.83	9
Dataset 2	Affymetrix U133 Plus 2.0	24	CADI	0.972	10

Reference

1. Solez K, Colvin RB, Racusen LC, Haas M, Sis B, Mengel M, et al. Banff 07 classification of renal allograft pathology: updates and future directions. Am J Transplant. 2008 Apr;8(4):753-60.

2. Yilmaz S, Tomlanovich S, Mathew T, Taskinen E, Paavonen T, Navarro M, et al. Protocol core needle biopsy and histologic Chronic Allograft Damage Index (CADI) as surrogate end point for long-term graft survival in multicenter studies. J Am Soc Nephrol. 2003 Mar;14(3):773-9.

3. Yilmaz S, McLaughlin K, Paavonen T, Taskinen E, Monroy M, Aavik E, et al. Clinical predictors of renal allograft histopathology: a comparative study of single-lesion histology versus a composite, quantitative scoring system. Transplantation. 2007 Mar 27;83(6):671-6.

4. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003 Feb 15;31(4):e15.

5. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007 Jan;8(1):118-27.

6. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003 Jul;34(3):267-73.

7. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

8. Firth D. Bias Reduction of Maximum Likelihood Estimates. Biometrika. 1993;80(1):12.

9. Heinze G, Ploner M. Fixing the nonconvergence bug in logistic regression with SPLUS and SAS. Comput Methods Programs Biomed. 2003 Jun;71(2):181-7.

10. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol. 2004 Apr;2(4):E108.

11. Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell. 2013 Apr 25;153(3):707-20.

12. Einecke G, Reeve J, Sis B, Mengel M, Hidalgo L, Famulski KS, et al. A molecular classifier for predicting future graft loss in late kidney transplant biopsies. J Clin Invest. 2010 Jun;120(6):1862-72.

13. Naesens M, Khatri P, Li L, Sigdel TK, Vitalone MJ, Chen R, et al. Progressive histological damage in renal allografts is associated with expression of innate and adaptive immunity genes. Kidney Int. 2011 Dec;80(12):1364-76.