**SUPPLEMENTARY MATERIAL**

## RESULTS

### Functional annotation and biological relevance

To further explore the functional relationships between specific PBMC subsets and tolerance-related expression patterns we dissected the molecular pathways contained in the microarray differential gene expression data set employing both *Gene set enrichment analysis* (GSEA) and *Ingenuity Pathway Analysis* (IPA). Applying GSEA to manually curated gene set databases (C2 MsigDB; Table S5 in Supplementary Material) a number of canonical pathways comprising inflammatory and immune stimulatory genes were significantly associated with the non-tolerant phenotype, while only the propanoate pathway was significantly enriched in the tolerant phenotype. We also investigated whether gene expression measurements could be linked to conserved regulatory motifs (C3 MSigDB) but this yielded no significant results. In contrast, exploration of computed gene expression compendiums (C4 MsigDB) revealed that the tolerant phenotype was highly enriched in 3 overlapping gene sets (neighbourhood of IL2RB, PTPN4, CD7; Table S1) altogether comprising 76 genes known to be preferentially expressed by NK and other cytotoxic lymphocytes such as CD8 and γδTCR+ T cells (ref. S1). To exclude the effect of HCV infection on the functional profiling of operational tolerance, we then applied GSEA to compare HCV-neg TOL and Non-TOL recipients (Table S1). The use of C2 MsigDB to identify canonical pathways resulted in the detection of 3 pro-inflammatory gene sets significantly enriched in HCV-neg Non-TOL samples (CTLA4, CMAC and hypertrophy model pathways; Table S1). However, genes included within these three pathways (e.g. *CD28, ICOS, CTLA4, JUN, TNF, IFNG, PIK3CA, ITK*) were not present among the genes

discriminating between HCV-neg TOL and Non-TOL as assessed by SAM at FDR<5%. No clear functional differences were noted between HCV-neg TOL and Non-TOL recipients when employing C3 MSigDB (regulatory motifs) databases, while the use of computed gene expression databases (C4 MsigDB) showed again enrichment in HCV-neg TOL samples of gene sets (neighbourhood of IL2RB, PTPN4, CD97; Table S1) commonly expressed by NK and other cytotoxic lymphocytes (ref. S1). The use of IPA on the complete TOL and Non-TOL differential expression data set identified SAPK/JNK signalling pathway and NK cell signalling pathway as the most significant canonical pathways associated with tolerance (Figure S2). The stress-activated SAPK/JNK pathway included a number of pro-inflammatory genes (*CDK4, CDK8, CSNK1A1, DAXX, DUSP10, MAP4K4, MAPK9, SOS1, TRA@*) that were differentially expressed between TOL and Non-TOL samples at FDR<5% only in HCV-pos recipients (data not shown). In contrast, NK signalling pathway comprised genes (*CD244, CD300A, KLRC3, KLRD1, KLRK1, SH2D1B,* and *SOS1*) significantly up-regulated in TOL samples at FDR<5% regardless of HCV infection (data not shown). Next, to understand the potential biological relevance of the most informative set of genes, we used IPA to functionally analyse the 45 genes differentially expressed by qPCR between TOL and either Non-TOL or CONT samples. IPA identified 3 partially overlapping networks connecting 33 out of the 45-gene list (Figure S2). The first network, which was built from 14 genes and received the highest IPA score, was centred on IL-8, NFkB and Akt and associated with cancer, cellular movement and immune and lymphatic system function. The second network, incorporating 13 out of the 45 genes, was centred on TP53 and CDKN1A and associated with cancer, cell death and immunological disease. The third network built on 5 genes was mostly centred around IL-4 and associated with cell-to-cell signalling and cellular development. Taken

together, functional profiling reveals that tolerance-related expression signatures are highly enriched in genes involved in the regulation of innate immune cell function. While a number of pro-inflammatory pathways are over-represented in Non-TOL recipients, this appears to be mainly attributable to the effect of chronic HCV infection and not directly related to operational tolerance.

## MATERIAL AND METHODS

### Peripheral blood immunophenotyping on sorted PBMC subsets

The expression at the protein level of 7 of the most discriminative genes identified by microarray and qPCR experiments (ILRB2, KLRB1, CD244, CD9, KLRF1, CD160, SLAMF7) was assessed on sorted PBMC subpopulations from a subset of 6 TOL, 6 Non-TOL and 5 CONT patients. CD160 fluorescent monoclonal antibodies were purchased from Beckman Coulter, SLAMF7 and KLRF1 from R&D Systems. All remaining antibodies were purchased from BD Biosciences.

### Functional annotation

Gene Set Enrichment Analysis (GSEA) was employed to identify biological pathways significantly associated with the tolerant state (ref. S2). In comparison to other strategies for analysis of molecular profiling data that focus on high scoring individual genes, GSEA does not employ a significance threshold and evaluates microarray data at the level of gene sets defined based on prior biological knowledge. This approach has been reported to yield robust results even when dealing with heterogeneous samples with subtle sample class differences. For the current analysis (incorporating all probes collapsed by genes with at least one log2-expression measurement >5) gene sets were extracted from Molecular Signature Database (MSigDB v.2-0) C2 (manually curated

canonical pathways), C3 (gene sets containing genes that share transcription factor or microRNA binding motifs) and C4 (computational gene sets generated in previous gene expression experiments) of MSigDB. Analysis were based on a *t*-test and a weighted scoring scheme with 1000 permutations on gene sets. Only gene sets with more than 15 genes were included in the analysis. Functional profiling was also performed on differentially expressed genes (SAM FDR<1%) employing the computational gene network prediction tool Ingenuity Pathway Analysis (IPA; www.ingenuity.com). This commercial application maps the uploaded gene identifiers into a global molecular network developed from a literature-supported Ingenuity Pathways Knowledge Base (IPKB), and then generates networks that represent the molecular relationships between the genes and their products. The biological functions significantly associated with the genes in the networks are provided and scored employing Fischer's exact test.

**FIGURE LEGENDS**

**Figure S1: Differences in protein expression in peripheral mononuclear between TOL, Non-TOL and CONT recipients.** A) Expression of ILRB2, KLRB1, CD244, CD9, KLRF1, CD160 and SLAMF7 on peripheral blood mononuclear cells. Representatitve flow cytometry histograms showing protein expression on TOL, Non-TOL and CONT samples. B) Differences in protein expression levels between TOL, Non-TOL and CONT samples. Bar plots represent mean expression (% of positive cells or mean fluorescence intensity (MFI) depending on the marker analysed) +/- SD from 6 TOL, 6 Non-TOL and 5 CONT samples. (*) = *P*-value <0.05 (*t*-test) between TOL and Non-TOL; (**) = *P*-value <0.05 (*t*-test) between TOL and CONT.

**Figure S2: Functional analysis of tolerance-related gene expression patterns.** A) Identification of the canonical pathways from Ingenuity Pathways Knowledge Base (IPKB) most significantly associated with the genes differentially expressed between TOL and Non-TOL samples. Genes selected by SAM at FDR <1% were considered for the analysis. The significance of the association was measured on the basis of the ratio of the number of genes from the data set that map to the pathway divided by the total number of genes that map to the canonical pathway (as displayed); and a *P*-value determining the probability that the association between the genes in the data set and the canonical pathway is explained by chance alone (Fischer's exact test). B-D) Gene and protein interaction networks defined by the 45 gene classifiers validated by qPCR. Three networks were built using Ingenuity Pathway Analysis (IPA) from 14 (B), 13 (C), and 5 (D) genes. Genes or gene products are represented as nodes and the biological relationship between two nodes is represented as an edge (line). The intensity of the node colour corresponds to up- (red) or down- (green) regulation.

**REFERENCES**


S1. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 101:6062-6067.

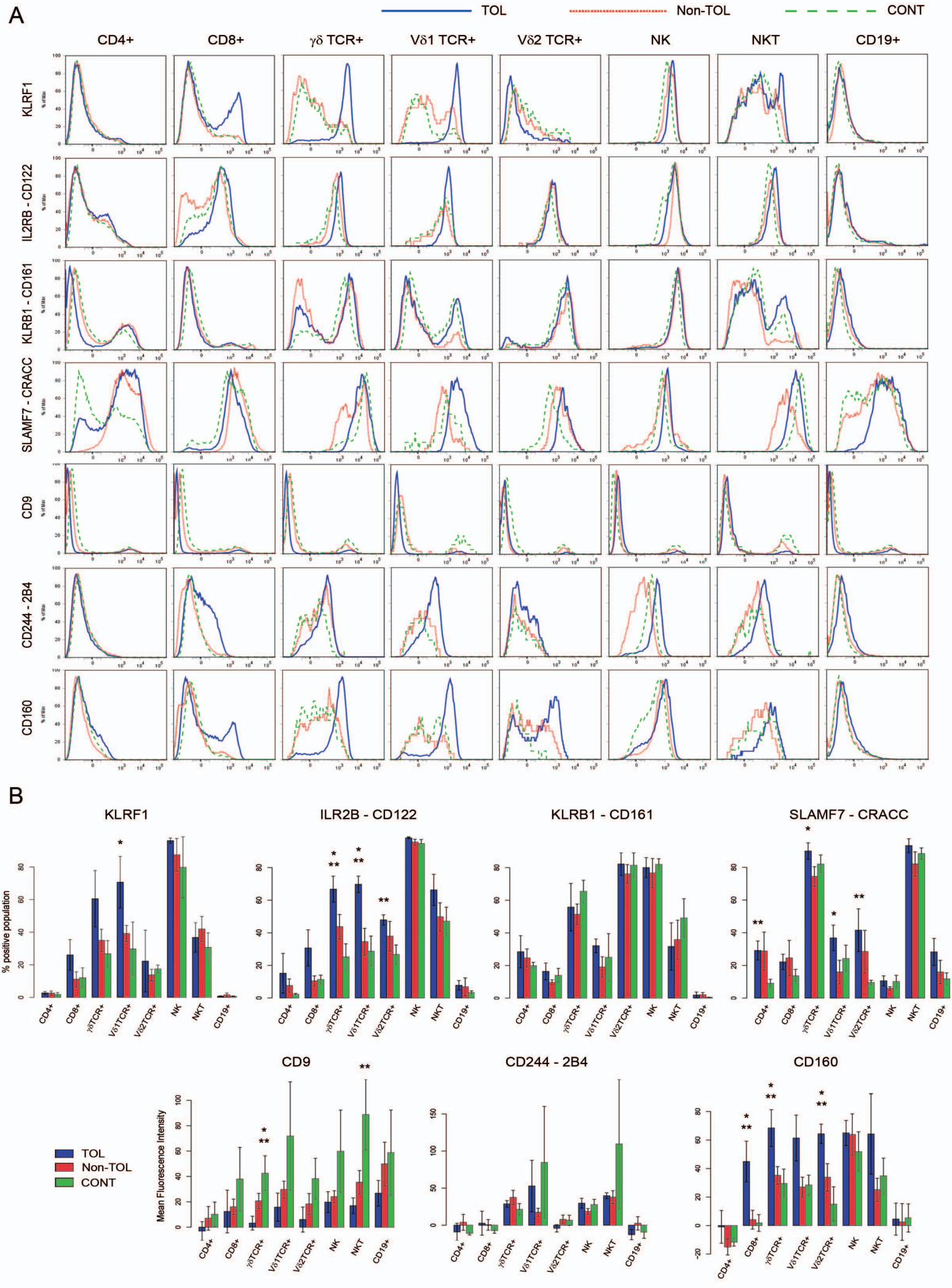
S2. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550.

Table S1: Functional gene set enrichment in tolerance-related differentially expressed gene lists (SAM; FDR<0.05) as assessed by gene-set enrichment analysis (GSEA)

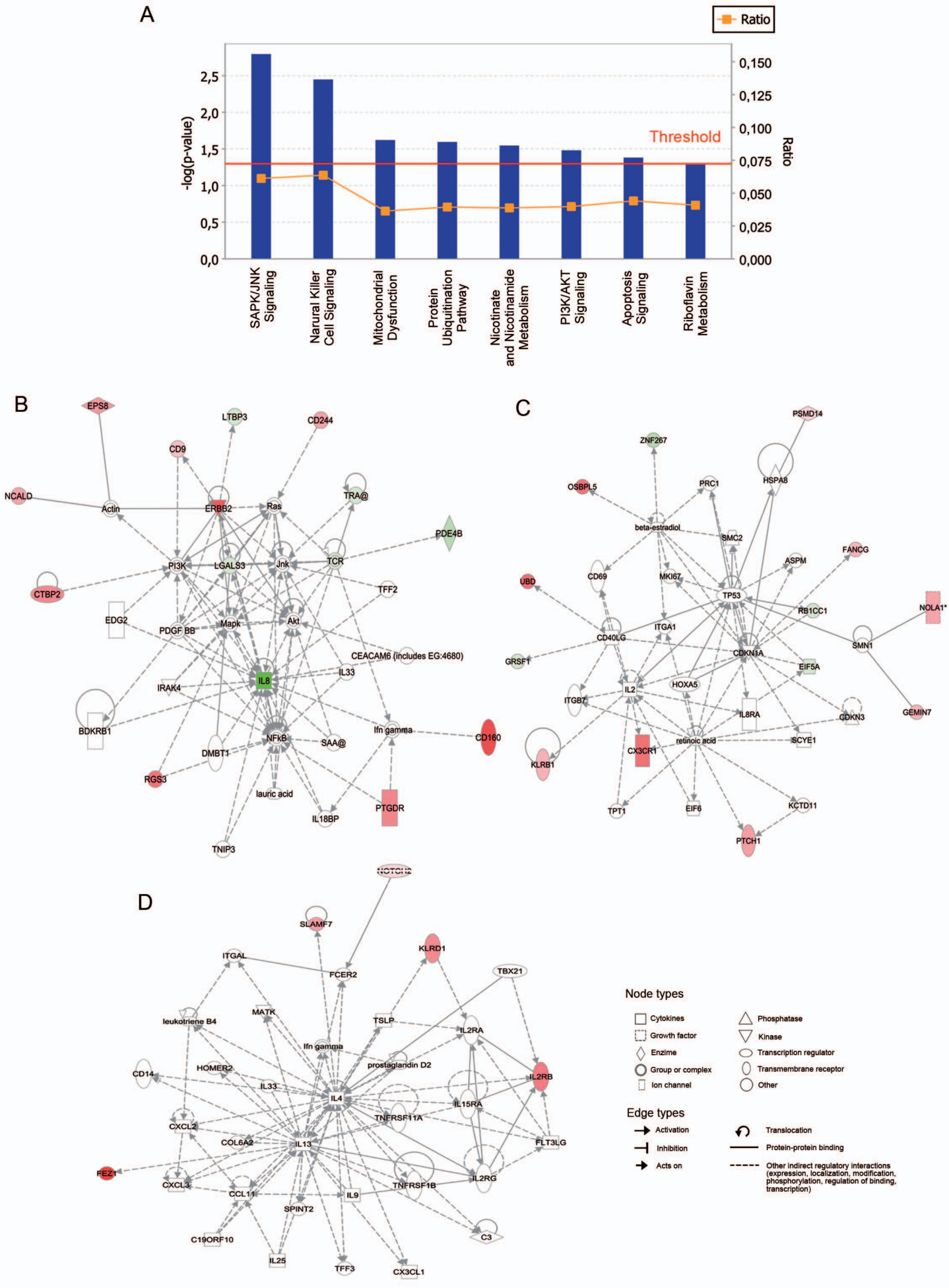
| **TOL / Non-TOL differential gene expression data set** | | | |
|---|---|---|---|
| **Enriched in TOL samples** | ***P*-value** | **FDR *q*-value** | **Genes with highest enrichment scores** |
| **Canonical pathways** (C2 MSigDB) | | | |
| Propanoate metabolism | 0.000 | 0.31 | *MCEE, ECHS1, ALDH9A1, PCCA, ALDH3A2* |
| **Computational gene sets** (C4 MSigDB) | | | |
| Neigborhood of PTPN4 | 0.000 | 0.000 | *Il2RB, CD160, PTGDR, KLRF1, KLRD1, KLRK1, KLRC3* |
| Neighborhood of IL2RB | 0.000 | 0.000 | *Il2RB, CD160, PTGDR, CD244, CX3CR1, PRF1, KIR3DL1* |
| Neigborhood of CD7 | 0.000 | 0.000 | *Il2RB, CD160, PTGDR, CD244, CX3CR1, PRF1, SPON2* |
| Neigborhood of MATK | 0.000 | 0.001 | *IL2RB, PTGDR, ARHGEF3, KLRD1, PRF1, GZMA, ZAP70* |
| Neigborhood of RAB7L1 | 0.001 | 0.004 | *IL2RB, KLRD1, KLRF1, APOBEC3G, PTGER2, BIN2, NCR3* |
| Neigborhood of BMPR2 | 0.001 | 0.005 | *WDR67, DFFB, IPO8, HDAC9, SMYD2, C22ORF9* |
| | | | |
| **Enriched in Non-TOL samples** | | | |
| **Canonical pathways** (C2 MSigDB) | | | |
| IL1R pathway | 0.000 | 0.000 | *TRAF6, IRAK3, IL1A, IL1R1, IL1R1, NFKBIA* |
| Hypertrophy model | 0.000 | 0.000 | *IFNG, IFDR1, VEGF, IL1A, IL1R1, ATF3* |
| Brest cancer estrogen signaling | 0.000 | 0.000 | *ITGA6, CDKN2A, FOSL1, SLC7A5, CCNE1, VEGF* |
| NFAT signaling | 0.000 | 0.009 | *IFNG, ITK, RELA, NFKBIB, SLA, FOS, IL8RA* |
| Tumor necrosis factor pathway | 0.000 | 0.011 | *NFKB1, CFLAR, FADD, TNFAIP3, NFKB2, JUN* |
| NTHI pathway | 0.000 | 0.014 | *IL8, TNF, IL1B, NFKBIA, DUSP1, RELA, MAP2K6* |
| CTLA4 pathway | 0.006 | 0.013 | *CD28, CTLA4, ICOS, PIK3R1, ITK, CD3E, TRA@* |
| CMAC pathway | 0.000 | 0.012 | *TNF, JUN, NFKBIA, FOS, RELA, RAF1, MAPK3* |
| NFKB pathway | 0.000 | 0.015 | *TNF,IL1R1, IL1A, TNFAIP3, TRAF6, RELA, FADD* |
| **Computational gene sets** (C4 MSigDB) | | | |
| Neigborhood of MMP1 | 0.000 | 0.31 | *CXCL1, TNFAIP6, IL6, PTX3, IL1B, CXCL3* |

| **HCV-negative TOL / Non-TOL differential gene expression data set** | | | |
|---|---|---|---|
| **Enriched in TOL samples** | ***P*-value** | **FDR *q*-value** | **Genes with highest enrichment scores** |
| **Canonical pathways** (C2 MSigDB) | | | |
| VIPP pathway | 0.000 | 0.05 | *ERG2, ERG3, PRKAR1B* |
| **Computational gene sets** (C4 MSigDB) | | | |
| Neigborhood of PTPN4 | 0.000 | 0.000 | *XCL2, Il2RB, KLRC3, PTGER2, PTGDR, CD160, TUSC4* |
| Neighborhood of CD97 | 0.000 | 0.000 | *PTGER2, DOK2, RIN3, CD300A, CD244, BIN2, CX3CR1* |
| Neighborhood of IL2RB | 0.000 | 0.001 | *XCL2, Il2RB, KLRC3, PTGER2, PTGDR, CD160,ASCL2* |
| Neigborhood of CD7 | 0.000 | 0.005 | *XCL2, Il2RB, KLRC3, PTGER2, PTGDR, CD160, TUSC4* |
| Neigborhood of RAB7L1 | 0.000 | 0.005 | *XCL2, Il2RB, KLRC3, PTGER2,BIN2, PRF1, NCR3, KLRF1* |
| Neigborhood of MATK | 0.000 | 0.004 | *IL2RB, PTGDR, ARHGEF3,PRF1, MATK, KLRK1, KLRD1* |
| Neigborhood of RAP1B | 0.000 | 0.007 | *BIN2, CD97, ELF4, JARID1A, VPS16, RAP2B,* |
| Neigborhood of JAK1 | 0.000 | 0.017 | *PTGER2, BIN2, ARHGEF3, CD97, NCR3, LOC54103* |
| | | | |
| **Enriched in Non-TOL samples** | | | |
| **Canonical pathways** (C2 MSigDB) | | | |
| CTLA4 pathway | 0.000 | 0.008 | *ICOS, CD28, CTLA4, TRA@, PIK3CA, PÎK3R1* |
| CMAC pathway | 0.004 | 0.074 | *JUN, TNF, FOS, RAF1, PLCB1, MAPK3, RELA* |
| Hypertrophy model pathway | 0.010 | 0.082 | *NR4A3, IFNG, TCF8, IL1R1, HBEGF, ADAM10* |
| **Computational gene sets** (C4 MSigDB) | | | |
| Neighborhood of EIF3S6 | 0.000 | 0.005 | *RPL27A, RPS29, FAU, EIF3S7, RPL11, RPS8* |
| Neighborhood of TPT1 | 0.000 | 0.012 | *RPL27A, RPS29, FAU, RPL1, RPS5, RPS8, EEF2* |
| Neighborhood of GLTSCR2 | 0.000 | 0.011 | *RPS29, FAU, EEF1B, RPS9, RPL13A, RPS16* |
| Neighborhood of MAX | 0.000 | 0.020 | *C14ORF11, CSDE1, FAM76D, ABT1, COPS2, CCDC117* |
| Neighborhood of NPM1 | 0.000 | 0.063 | *RPL27A, RPS29, FAU, NCL, EIF3S7, RPL11, RPS5, RPS8* |
| Neighborhood of ACTG1 | 0.000 | 0.092 | *RPL27A, SSR2, RPS29, FAU, NCL, RPL11, RPS5, RPS8* |
| Neighborhood of CEBPA | 0.069 | 0.17 | *CYP1A2, HP, ORM1, CYP27A1, CYP2D6, GSTM1, CES1* |

# Figure S1


A
TOL
Non-TOL
CONT
CD4+
CD8+
γδ TCR+
Vδ1 TCR+
Vδ2 TCR+
NK
NKT
CD19+
KLRF1
IL2RB - CD122
KLRB1 - CD161
SLAMF7 - CRACC
CD9
CD244 - 2B4
CD160
% of Max
B
KLRF1
ILR2B - CD122
KLRB1 - CD161
SLAMF7 - CRACC
CD9
CD244 - 2B4
CD160
% positive population
Mean Fluorescence Intensity
TOL
Non-TOL
CONT
CD4+
CD8+
γδTCR+
Vδ1TCR+
Vδ2TCR+
NK
NKT
CD19+

# Figure S2


A
Ratio
-log(p-value)
Ratio
Threshold
2,5
2,0
1,5
1,0
0,5
0,0
0,150
0,125
0,100
0,075
0,050
0,025
0,000
SAPK/JNK Signaling
Narural Killer Cell Signaling
Mitochondrial Dysfunction
Protein Ubiquitination Pathway
Nicotinate and Nicotinamide Metabolism
PI3K/AKT Signaling
Apoptosis Signaling
Riboflavin Metabolism
B
EPS8
LTBP3
CD244
CD9
NCALD
Actin
ERBB2
Ras
TRA@
PDE4B
PI3K
LGALS3
Jnk
TCR
CTBP2
TFF2
EDG2
PDGF BB
Mapk
Akt
CEACAM6 (includes EG:4680)
IL33
IL8
IRAK4
BDKRB1
NFkB
SAA@
Ifn gamma
CD160
DMBT1
RGS3
lauric acid
PTGDR
IL18BP
TNIP3
C
PSMD14
ZNF267
HSPA8
OSBPL5
PRC1
beta-estradiol
SMC2
FANCG
ASPM
CD69
MKI67
UBD
TP53
RB1CC1
NOLA1*
ITGA1
CD40LG
SMN1
GRSF1
CDKN1A
EIF5A
HOXA5
IL2
ITGB7
IL8RA
CDKN3
GEMIN7
SCYE1
retinoic acid
CX3CR1
KLRB1
TPT1
EIF6
KCTD11
PTCH1
D
SLAMF7
KLRD1
ITGAL
FCER2
TBX21
MATK
leukotriene B4
TSLP
IL2RA
Ifn gamma
prostaglandin D2
CD14
HOMER2
IL33
IL4
IL2RB
IL15RA
TNFRSF11A
CXCL2
COL6A2
IL13
FLT3LG
FEZ1
CXCL3
CCL11
IL9
TNFRSF1B
IL2RG
SPINT2
C19ORF10
C3
IL25
TFF3
CX3CL1
Node types
Cytokines
Growth factor
Enzime
Group or complex
Ion channel
Phosphatase
Kinase
Transcription regulator
Transmembrane receptor
Other
Edge types
Activation
Inhibition
Acts on
Translocation
Protein-protein binding
Other indirect regulatory interactions (expression, localization, modification, phosphorylation, regulation of binding, transcription)