Metagene Projection Model based on Non negative matrix factorization

The metagene projection model us based in the **Non Negative Matrix Factorization** technique, which is a matrix factorization algorithm originally introduced by Lee *et al*. to the analysis of facial images. Formally, the non-negative matrix decomposition can be described as follow:

$$V \approx WH$$

where $V \in \mathbb{R}^{m \times n}$ is a positive data matrix with *m* variables and *n* objects, $W \in \mathbb{R}^{m \times k}$ are the reduced *k* basis vectors or factors, and $H \in \mathbb{R}^{k \times n}$ contains the coefficients of the linear combinations of the basis vectors needed to reconstruct the original data (also known as encoding vectors). Additionally, we have the following conditions: $k \leq m$, all matrices V, W, H are non-negative, and the columns of W (the basis vectors) are normalized (sum up to 1). The main difference between NMF and other classical factorization models relies in the non-negativity constraints imposed on both the basis W and encoding vectors H . In this way, only additive combinations are possible:

$$(V)_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{k} W_{ia} H_{a\mu}$$

The objective function, based on the Poisson likelihood, can be defined using the following divergence function, which we need to minimize:

$$D(V, WH) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(V_{ij} \ln \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij} \right)$$

For the **metagene projection model** a factorization rank *k* is defined and NMF is applied to the training data. From the factoring of the original matrix, it is constructed a mapping to project a data set into the space of the metagenes. Mathematically, this can be accomplished by using the Moore–Penrose generalized pseudoinverse of W , so that

$$\widehat{H} = (W)^{-1} V \text{ , where } \widehat{H} \approx H$$

The computation of the pseudoinverse of W_M is done using “ginv” from R's MASS package. Using the projected data a one-versus-all SVM classifier is used to predict classes by using the *k* metagenes as the input features. This method provides a predicted class and a predictive confidence by using a modified Brier score.

$$C_P = 1 - \frac{\left((1 - P_1)^2 + \sum_{i=2}^{K} P_i^2 \right)}{(1 - 1/k)^2 + (k-1)(1/k)^2}$$, where $P_1 > P_2 > \ldots > P_k$ is the sorted list of *k* output probabilities for a given sample.

ADDITIONAL Figures an tables: Carmona-Saez et al. "Metagene projection characterizes GEN2.2 and CAL-1 as relevant human plasmacytoid dendritic cell models"

1. DATASETS USED IN THE ANALYSIS

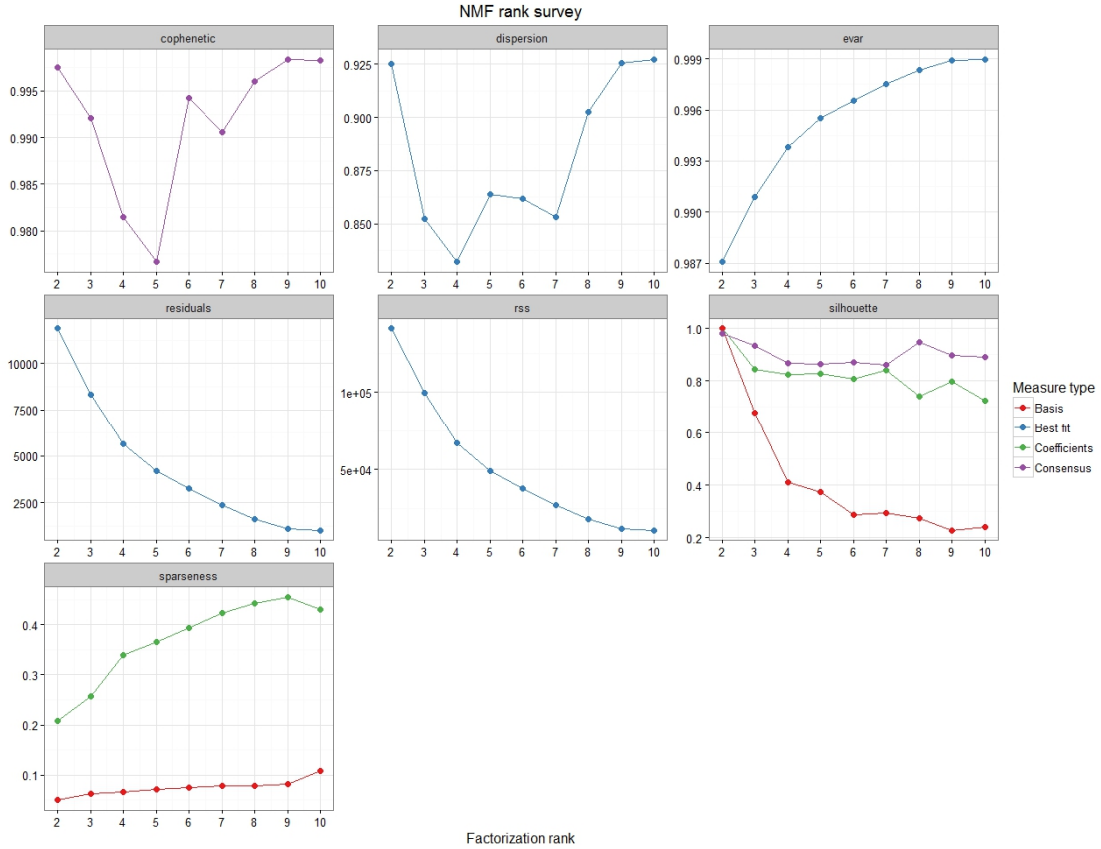
Table S1 –Public gene expression datasets used in the study. This table contains the GEO ID, cell types, number of samples, platform and reference of all datasets used in our analysis.

GEO ID	Cell types	Samples	Platform	Ref.
GSE28490	monocytes, B cells, $CD4^+$ T cells, $CD8^+$ T cells, NK cells, eosinophils, cDCs (mDC), neutrophils and pDC	47	Affymetrix HG U133 Plus 2.0	(Allantaz *et al.*, 2012)
GSE28491	Monocytes, B cells, $CD4^+$ T cells, $CD8^+$ T cells, NK cells, eosinophils and neutrophils	33	Affymetrix HG U133 Plus 2.0	(Allantaz *et al.*, 2012)
GSE15215	pDC $CD2^-$ and pDC $CD2^+$	4	Affymetrix HG U133A	(Matsui *et al.*, 2009)
GSE35457	pDC, $CD14^-cd16^+$ monocytes, $CD14^+$ monocytes, $CD141^+$ dendritic cells, CD16 monocytes, $CD1c^+$ dendritic cells	33	Illumina HumanHT-12 V4.0	(Haniffa *et al.*, 2012)
GSE12507	CAL-1 cell line	2	Affymetrix HG U133 Plus 2.0	(Cisse *et al.*, 2008)
GSE30849	CAL-1 (untreated and CpG treated samples)	9	NCI/ATC Hs-OperonV3	(Steinhagen *et al.*, 2012)
GSE55467	pDC (untreated and CpG treated samples)	5	Affymetrix Human Gene 2.0 ST Array	(Steinhagen *et al.*, 2012)

2. EVALUATION OF FACTORIZATION RANK

We applied different metrics implemented in the NMF R package (Gaujoux and Seoighe, 2010), to evaluate the performance of the factorization for different values of k. A good value of rank will properly reduce the dimensionality but still preserving the main features of the data. We find that at k=9 the cophenetic correlation coefficient (Brunet *et al.*, 2004) showed the highest value, which is in agreement with the previous information about the cell types in the dataset. Results are provided in figure 1

Figure S1. Quality and performance measures from factorization at ranks different ranks. The plots show (from left to right and top to bottom): Cophenetic correlation coefficient, dispersion, evar (explained variance), residuals, rss (residual sum of squares), silhouette and sparseness. More details about these metrics can be found in the NMF R package documentation and associated references. We found that correlation cophenetic coefficient showed the highest value at rank = 9.



3. GENE SELECTION AND GENE SET ENRICHMENT ANALYSIS OF METAGENES

To get insights into the biological meaning of the extracted metagenes we used the GSEA method (Subramanian *et al.*, 2005). GSEA evaluates the distribution of genes belonging to the same gene set across a ranked list of genes and determines whether this gene set is statistically over-represented in the top (or bottom) of the list. GSEA analysis was performed in each metagene, sorting all genes by their metagene coefficients and evaluating the enrichment of gene sets related to immunologic signatures from in the MSigDB database (www.broadinstitute.org/gsea/msigdb). To examine which genesets were enriched in each metagene we tested gene sets from a compendium of 1910 gene signatures that represent cell states and perturbations within the immune system (MSigDB, collection C7). The results from GSEA analysis are provided in additional file 2. Attending to the enrichment results in metagene 7, which was specific of pDC, we found that the top six significant gene sets are from gene sets containing genes up-regulated in pDC Vs other cell lines, or what is similar, down-regulated in the comparison of cell lines Vs pDC. This fact reinforces the notion that this metagene captured patterns associated to pDC.

Table S2 –Top six genesets enriched in metagene 7

Gene Set Name[a]	Brief Description[b]	Size[c]	NES[d]
GSE29618_BCELL_VS_PDC_DN	Genes down-regulated in comparison of B cells versus plasmacytoid dendritic cells (pDC)	103	2.63
GSE29618_MONOCYTE_VS_PDC_DAY7_FLU_VACCINE_DN	Genes down-regulated in comparison of monocytes from influenza vaccinee at day 7 post-vaccination versus plasmacytoid dendritic cells (pDC) at day 7 post-vaccination.	95	2.59
GSE29618_PDC_VS_MDC_UP	Genes up-regulated in comparison of plasmacytoid dendritic cells (DC) versus myeloid DCs	106	2.56
GSE29618_BCELL_VS_PDC_DAY7_FLU_VACCINE_DN	Genes down-regulated in comparison of B cells from influenza vaccinee at day 7 post-vaccination versus plasmacytoid dendritic cells (pDC) at day 7 post-vaccination	102	2.54
GSE29618_MONOCYTE_VS_PDC_DN	Genes down-regulated in comparison of monocytes versus plasmacytoid dendritic cells (pDC)	94	2.50
GSE29618_PDC_VS_MDC_DAY7_FLU_VACCINE_UP	Genes up-regulated in comparison of plasmacytoid dendritic cells (pDC) from influenza vaccinee at day 7 post-vaccination versus myeloid DCs at day 7 post-vaccination	101	2.46

[a] Gene set name from the Molecular Signature Database (MSigDB), [b] brief description of the gene set, [c] number of genes contained in the gene set and [d] normalized enrichment score.

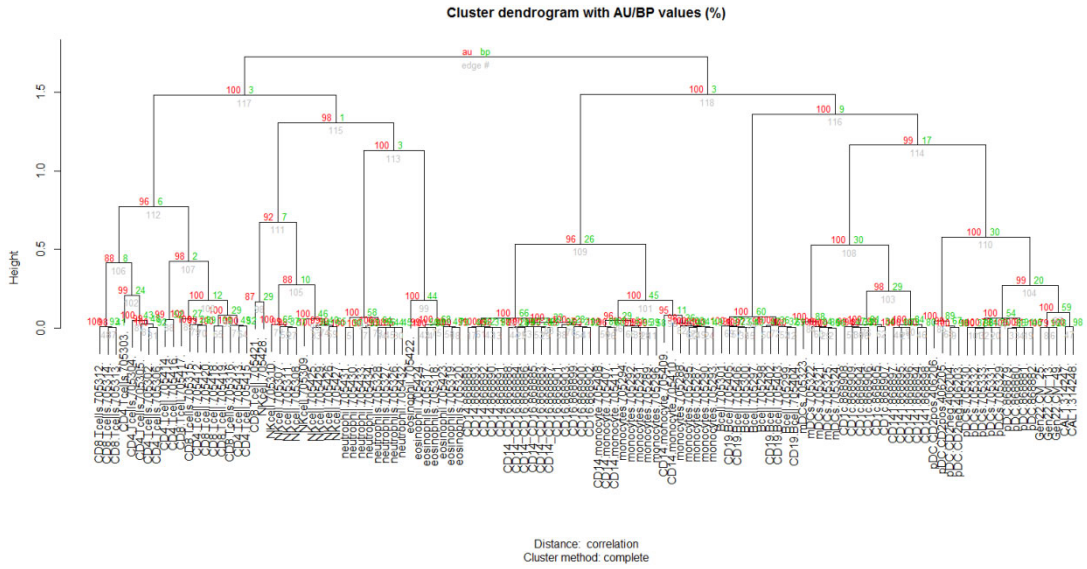
In addition, we defined the set of most relevant genes in each metagene using the methodology developed by Kim and Tidor (Kim and Park, 2007) and implemented in the nmf package. Results are provided in additional file 3. Due to the nature of nmf these genes do not necessarily represent the most differentially expressed genes in pDC with respect to the rest of cell lines but are genes that are co-expressed in a subset of cell lines, in this case pDCs. For more details about the meaning of nmf factors see (Carmona-Saez *et al.*, 2006; Kim and Tidor, 2003)

4. EVALUATION OF CLUSTERING RESULTS

We assessed the certainty of the clusters using a boot-strapping approach (Suzuki and Shimodaira, 2006). This method computes *p*-values for each cluster by multiscale bootstrap resampling. These p-values indicate how strongly the cluster is supported by data. It provides two types of p-values: AU (computed by multiscale bootstrap resampling) p-value and BP (computed by normal bootstrap resampling). Results from cluster stability analysis are shown in the next figures.

Figure S2. Cluster stability in the projected data. The figure shows results using correlation (A) and euclidean (B) distances. Red values are AU p-values and green values are BP values. Clusters with AU larger than 95% are strongly supported by data

A



B

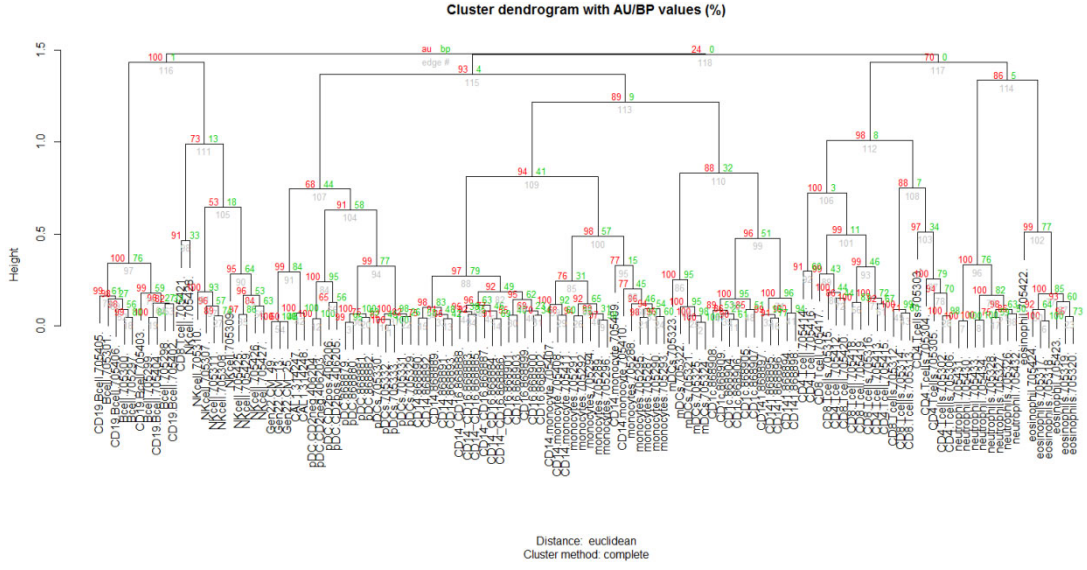
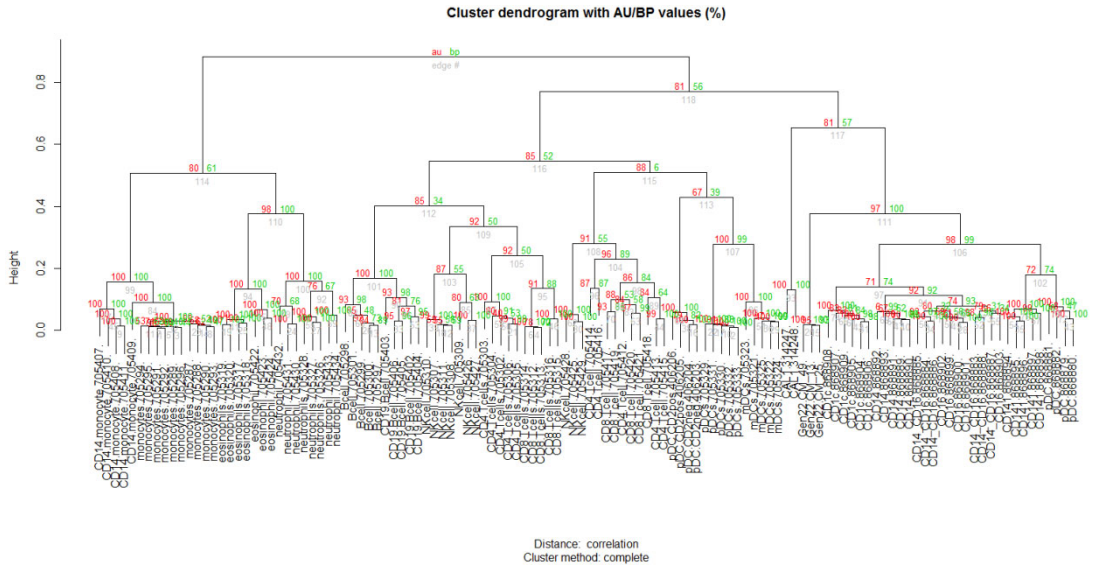
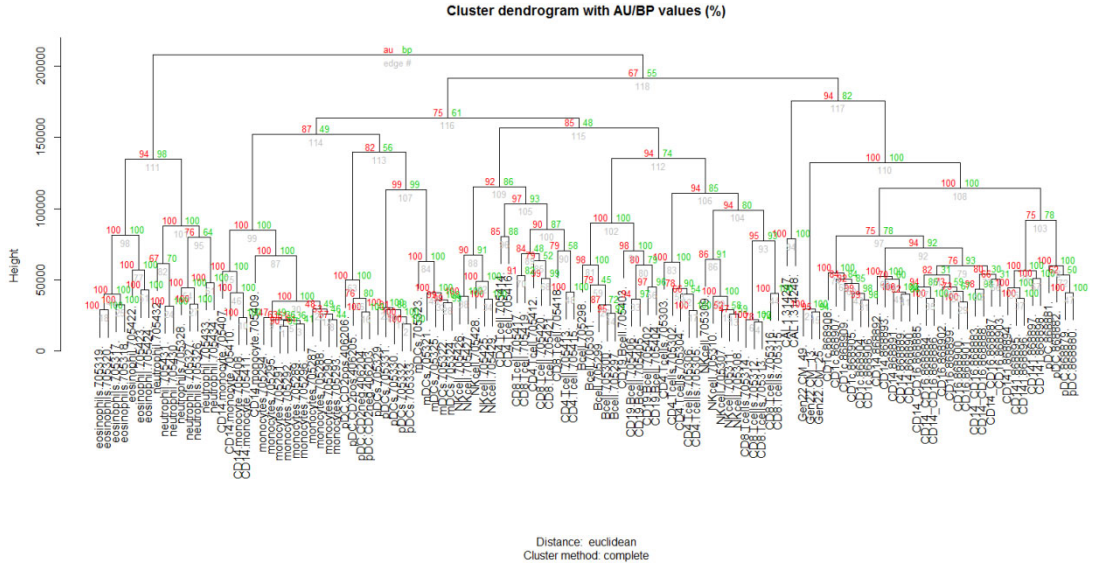


Figure S3. Cluster stability in the original data. The figure shows results using correlation (A) and euclidean (B) distances of clustering results from the original data

A



B



5. COMMON GENE SIGNATURE

Figure S4. Heatmap showing expression profiles of common signature. The figure shows the normalized gene expression of the most differentially expressed genes. Red represents high expression and green low expression.

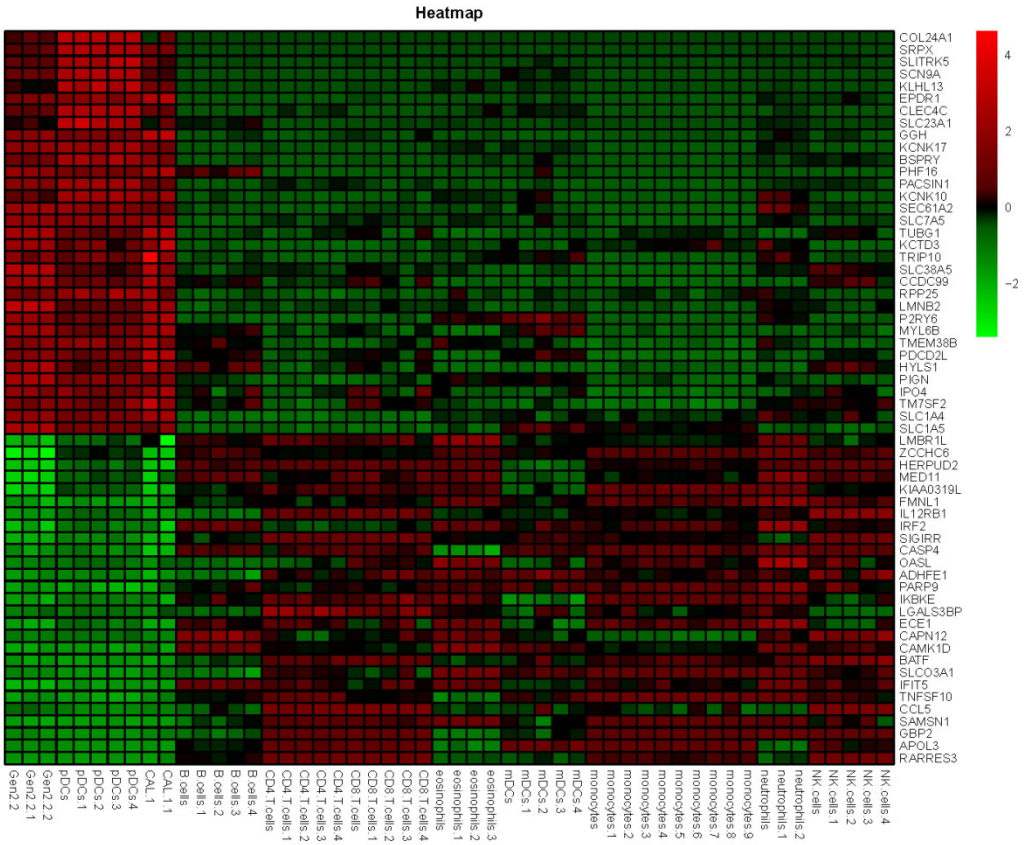


Table S3. Enrichment of GO Terms in over- and under-expressed genes from the common signature

Enriched annotations (in at least 3 genes and corrected pvalue <0.05) in over-expressed genes

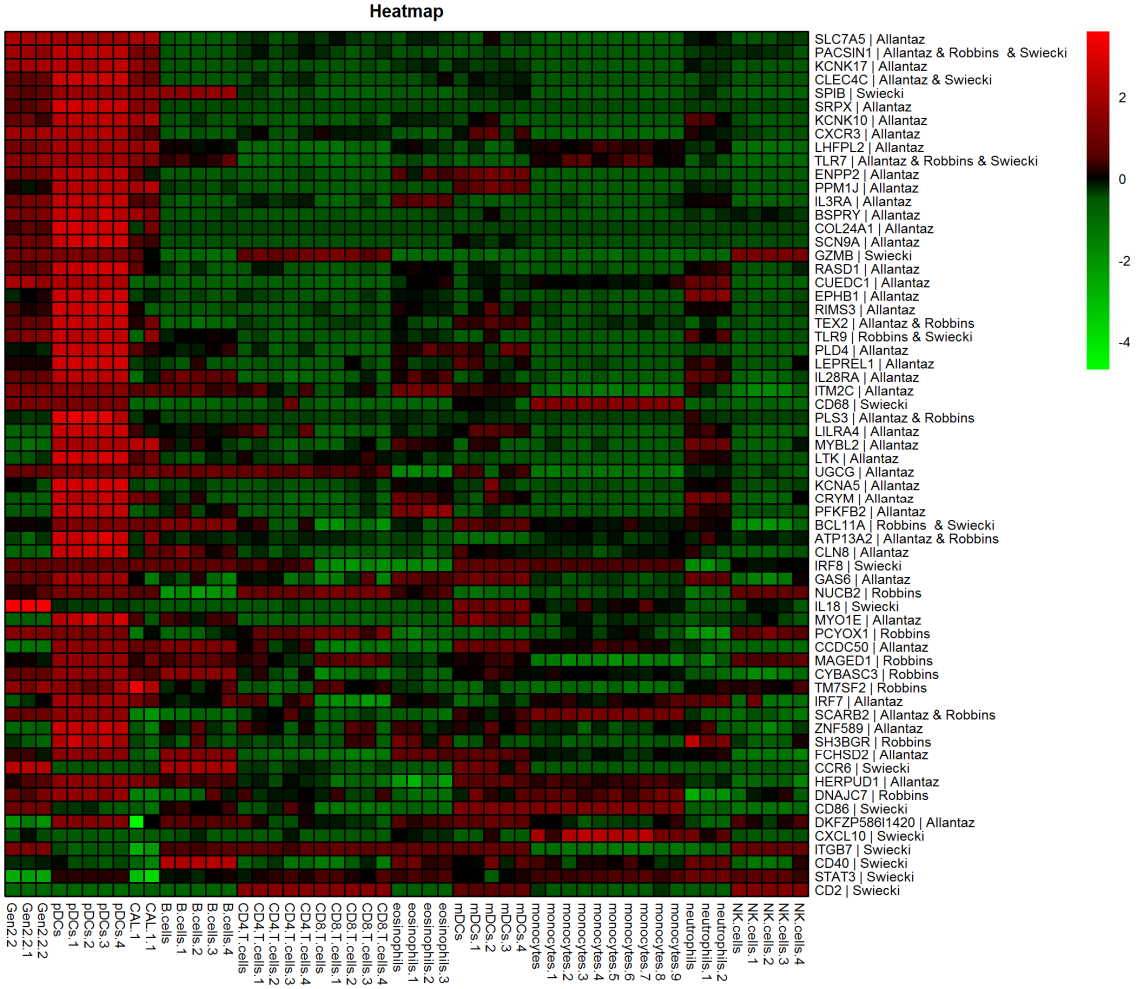
GO Description	corrected pval	Genes
ion transport (BP)	6.74E-09	SLC23A1,SLC1A4,SCN9A,BSPRY,SLC1A5,KCNK10,KCNK17,SLC38A5,TMEM38B,SLC7A5
amino acid transport (BP)	3.44E-06	SLC1A4,SLC1A5,SLC38A5,SLC7A5
potassium ion transmembrane transport (BP)	0.000150696	KCNK10,KCNK17,TMEM38B
potassium ion transport (BP)	0.000297985	KCNK10,KCNK17,KCTD3,TMEM38B
transmembrane transport (BP)	0.000382895	SLC23A1,SLC1A4,SCN9A,SLC1A5,SLC38A5,SLC7A5
cellular nitrogen compound metabolic process (BP)	0.000437838	SLC1A4,SLC1A5,SLC38A5,SLC7A5

Enriched annotations (in at least 3 genes and corrected pvalue <0.05) in under-expressed genes

GO Description	corrected pval	Genes
cytokine-mediated signaling pathway (BP)	0.00110088	OASL,IRF2,IL12RB1,GBP2
interferon-gamma-mediated signaling pathway (BP)	0.00169484	OASL,IRF2,GBP2
type I interferon-mediated signaling pathway (BP)	0.0017294	OASL,IRF2,GBP2
immune response (BP)	0.00203029	IKBKE,OASL,TNFSF10,CCL5,GBP2
positive regulation of I-kappaB kinase/NF-kappaB cascade (BP)	0.00553898	APOL3,IKBKE,TNFSF10
inflammatory response (BP)	0.00998262	APOL3,CAMK1D,CCL5
signal transduction (BP)	0.0143186	SIGIRR,APOL3,LGALS3BP,IL12RB1,TNFSF10
proteolysis (BP)	0.0201934	ECE1,CASP4,CAPN12
apoptotic process (BP)	0.0226044	ECE1,TNFSF10,CASP4

6. EXPRESSION PATTERNS OF PREVIOUSLY REPORTED pDC RELATED GENES

Figure S5. Heatmap showing expression profiles of previously reported pDC markers. The figure shows the normalized gene expression of genes indentified by Robbins et al. (Robbins *et al.*, 2008), Allantaz et al. (Allantaz *et al.*, 2012) and Swiecki and Colonna (Swiecki *et al.*, 2010) that were also present in the combined dataset generated in this study. From the study of Swiecki and Colonna we selected genes from the categories “Diverse functions of pDCs”, “Factors influencing pDC migration”, “Phenotypes of human pDCs” and “Regulation of pDC development and function” in different sections and figures of the paper.



7. DIFFERENTIAL EXPRESSION ANALYSIS OF CAL-1 AND GENE2.2 Vs pDC

A comparative analysis of the different pDC at steady state was carried out. Genes differentially expressed in each comparison can be found in the additional file 4. Enrichment analysis of GO terms in the set of differentially expressed genes revealed, as expected, that processes related to proliferative capacity were over-represented in the list of over-expressed genes when comparing leukemia counterparts models GEN2.2 and CAL-1 with primary pDCs.

Table S4. GO terms from processed significantly enriched (corrected p-value < 0.01) in over-expressed genes in (A) GEN2.2 Vs Pdc and (B) CAL-1 Vs pDC

B

GO Description	adjusted Pvalue
cell division (BP)	1.39E-17
mitosis (BP)	9.76E-15
mitotic cell cycle (BP)	9.29E-14
DNA replication (BP)	2.44E-08
M phase of mitotic cell cycle (BP)	8.24E-08
cell cycle (BP)	2.56E-07
mitotic spindle organization (BP)	6.07E-05
spindle organization (BP)	6.91E-05
G2/M transition of mitotic cell cycle (BP)	7.61E-05
regulation of cyclin-dependent protein kinase activity (BP)	0.000325436
DNA repair (BP)	0.000407465
meiosis (BP)	0.000822227
RNA metabolic process (BP)	0.00121659
cell cycle checkpoint (BP)	0.00133582
microtubule-based movement (BP)	0.00224695
mRNA metabolic process (BP)	0.00296169
chromosome segregation (BP)	0.005029
S phase of mitotic cell cycle (BP)	0.0053098
phosphatidylinositol-mediated signaling (BP)	0.00934407

B

GO Description	adjusted Pvalue
synaptic transmission (BP)	0.00362436
cell differentiation (BP)	0.00858681
meiosis (BP)	0.00881772
M phase of mitotic cell cycle (BP)	0.00937211
multicellular organismal development (BP)	0.00999253
muscle filament sliding (BP)	0.0102228
mitotic cell cycle (BP)	0.0114304
mitotic prometaphase (BP)	0.0118771
response to amphetamine (BP)	0.0127033
chromosome segregation (BP)	0.0141564
G-protein signaling, coupled to cyclic nucleotide second messenger (BP)	0.0195861

8. DIFFERENTIAL EXPRESSION AND FUNCTIONAL ANALYSIS IN CpG STIMULATED CELLS

Table S5. Over-expressed Genes (rank products pfp <0.05) in treated Vs untreated cells

GEN2.2				GSE30849 (CAL-1)		GSE55467 (pDC)	
LOC728835	MGC3020	ATF3	RNU4-2	IFIT2	EPSTI1	IFNA7	NT5C3
CCL3L3	NFKBIA	GJB2	MGC4677	CCL4	PRKAG3	IFNA22P	CD38
CCL3	GADD45B	FOXA1	SEMA7A	CCL2	MASTL	IFNA21	EAF2
CCL3L1	TNF	RPPH1	MCL1	IFIT1	UBE2L6	IFNA4	MIR223
CCL4L1	RND1	RNU4ATAC	C7orf40	LOC643549	SP110	IFNA1	MACC1
IFNB1	BHLHB2	SNORD13	CCND1	IL1B	BTG2	IFNA10	CD40
IL1B	TNFSF14	IFIT2	MARCKSL1	SAMD9L	ACSL1	IFNA17	FLJ10489
PLAU	PDGFB	KLF2	ETS2	NFKBIZ	CTSZ	GPR34	IDO1
IL8	SNORD3D	ZDHHC18	LOC649143	IL23A	FAM49A	IFNA2	LOC284648
CD69	HLA-DQA2	C16orf67	HBEGF	SAMD9	OR13F1	IFNA8	IFNA6
CCL4L2	PTGER2	TNFSF18	HLX	IFIT5	HERC5	IFNA16	EBI3
RGS16	MIR221	IRAK2	ISG20	IFIH1	CCL3	IFNW1	FOXC1
CCL5	HNRNPA3P1	CCR7	IL28RA	BIRC3	CLEC4C	IFNA5	CXorf21
MIR155HG	C14orf181	TMEM107	P2RX4	WNT10A	RAB9A	IFNB1	HERC6
IL23A	ZC3H12C	C15orf48	TNFRSF4	IL6	IFI44L	IFNE	FAM111A
LOC728830	PIM3	ELL3	SIRT1	NFKBIA	IFI16	GPR171	SEMA3A
EGR1	SNORD84	MIR302C	PLA1A	MX1	EIF2AK2	MIR31HG	NEXN
TRIB1	CSF2	OTUD1	B4GALT5	GPR183	INSR	IL29	DDX60
LOC643930	STX11	GPR183	CLCF1	IL18R1	NOD2	RANBP3L	POGLUT1
SGK1	SNORD3C	OBFC2A	ATP1B1	CCL1	GPR119	IFNA14	CDK6
TNFAIP2	LOC338758	NLRP3	PTGIR	IL10RA	IGHV3-48	CXCL9	CSRP2
CD83	CDKN1A	LOC100130229	OXCT2	CHRM1	IGFBP4	CXCL11	NMB
TNFSF9	MAFF	FAM100B	RNF19B	TNFSF10	MCTP1	IL28B	LOC100499405
SGK	IGFBPL1	RNU1-3	INTS9	CD83	NEXN	IL12A	SLC1A2
TMEM88	HOXC9	CLLU1OS	LOC644641	OAS2	PDRG1	IL18RAP	TEX14
RGS1	RAB11FIP1	SOCS1	OASL	ISG15	CD70	MMP1	CHML
LOC642093	ZFP36	LOC652616	CCL14	STAT1	CD300C	KIAA1377	LOC100287562
EDN1	MARCKS	TRAF1	MRPL39	ASB2	IFI6	TNFSF4	BMP4
TNFAIP3	BTG1	TNIP2	IFNGR2	RGS1	SNAI1	IL36G	GBP4
DUSP5	NCOA7	RNU4-1	CHST7	OR51T1	REL	IGFBP4	CMPK2
DUSP2	SNORD43	RCAN1	REL	SOCS1	APOBEC3A	CLDN1	EPSTI1
PTGS2	DUSP1	HOXC8	TGIF2	TNFAIP3	SMPD3	IL2RA	TMPRSS13
CCL1	EBI2	LHX2	C6orf150	RGL1	RAX	CLECL1	LAMP3
ZC3H12A	KDM6B	OSGIN2	KMO	CXorf21	CCL3L1	CD274	PGAP1
PTGER4	IER5	ZMYND15	BIRC2	RELB	ICAM1	CXCL10	MYO1A
ICAM1	ARL5B	LONRF1	TAPBP	IFIT3	IRF7	DEPDC7	HIST2H2BC
FFAR2	WNT10A	IL10RA	BID	B3GNT2	CXCL8	PRRG4	POLR3C
EGR2	SNORD3A	TRIP10	ZBTB43	MIR155HG	PI4K2B	TAL1	EXOSC9
NFKBIZ	NFKBID	IMAA	C17orf96	TNFRSF4	ISG20	NRG4	TRAFD1
IL18RAP	RNU1-5	SOCS3	CBX3	TRIM21	TAP1	ASPHD2	ITGB8
BTG2	CXCR4	LOC650832	MLLT11	IER3	TGIF2LY	MIR222	SPINK1
BIRC3	RNU5A	HERPUD1	TCEB3	PARP9	TICAM1	CALCRL	MMP7
LTA	NFE2L2	IL4I1	C10orf10	GPR84		LOC100132707	BET1
MAP3K8	NR4A2	NCF1	PNRC1	TAGAP		RGS13	TRANK1
KLF6	C6orf222	ZFP36L1	GPR35	PIM2		CLEC2D	NXT2
CCL2	DRAM1	ETV3	C9orf38	LTA		HERC5	RGL1
IER3	RAPGEF1	NCF1C	GBP2	SLAMF7		FOXA1	LOC100131541
PIM1	RELB	LILRA5	SPRED2	GIMAP7		LOC344887	IL12B
CD40	NFKB2	LOC199800	SNORA3	PLEK		TFPI2	OVOS2
CYB5D1	IL6	IL24	LOC153684	JUNB		IKZF2	HAPLN3
PLEK	LOC652479	PPTC7	RANBP6	ARHGEF3		EXT1	IFIH1
FOSB	FTH1	GRASP	RNY5	MARS2		KIF14	IFI6
PPP1R15A	EHD1	IL18R1	SLC25A34	HIVEP2		SLC7A11	SAMD9
GPR84	CFLAR	IER2	RNU1A3	SNX20		SAMD9L	MMP10
SNORA9	LOC650557	LRRC50	CHST11	RGS16		RSAD2	GPR180
JUNB	RNU11	FOXC1	SERPINB8	PTGER4		LTA	XAF1
CD70	RASGEF1B	SLAMF7	RASSF5	TRAF1		HESX1	
SDC4	CCL22	CCNL1	GEM	BLZF1		C5orf28	
PMAIP1	TICAM1	LOC653610	P2RY10	RPL13AP17		NMI	
TAGAP	PIM2	STX12	ADAM8	NFKB1		LINC00158	
LOC648998	TRAF4	NFKBIE	BTG3	MCL1		DKK4	
IL1RN	BCL3	PLAUR	RGL1	DDX58		IFI44L	
RIPK2	GPR132	NEU4	CA2	OR8K3		PUS10	
NEK8	NINJ1	LOC728175	MIR939	IRF1		USP18	
C1orf61	SNORA42	NFKB1	SNORD95	CD69		PLA2G4A	
JUN	EGR3	SNORA28	RNU1F1	FGL2		PNPT1	
TLCD1	LRRC32	RAB9A		ONECUT2		IFIT1	
AXUD1	IL20	IRF4		OR4A13P		SNORD102	
PHLDA1	RNU1G2	ITPKC		BCL2A1		TNFSF10	

Figure S6. Gene Set Enrichment Analysis (GSEA) of stimulated signatures. We evaluated the distribution of genes that were found differentially expressed in stimulated CAL-1 (A) and GEN2.2 (B) across the ranked list of differentially expressed genes in the stimulated pDC signature. The green curve represents the evolution of the density of the genes identified in the pDC signature. Tables on the right had side contain the summary statistics of each geneset.

A

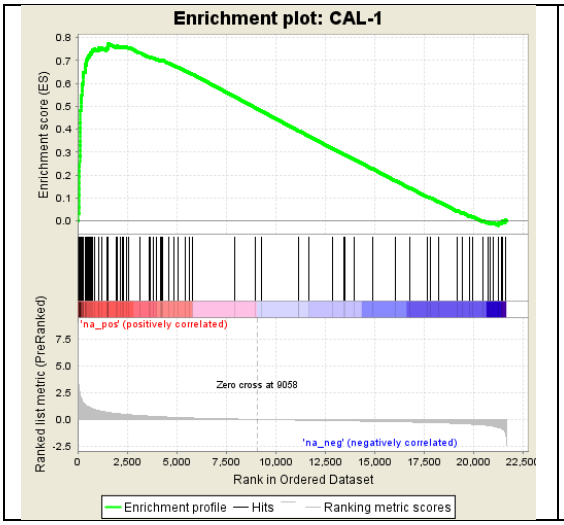


Table: GSEA Results Summary

Dataset	GSE55467.1000upFC
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_pos
GeneSet	CAL-1
Enrichment Score (ES)	0.7741329
Normalized Enrichment Score (NES)	2.5327556
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

B

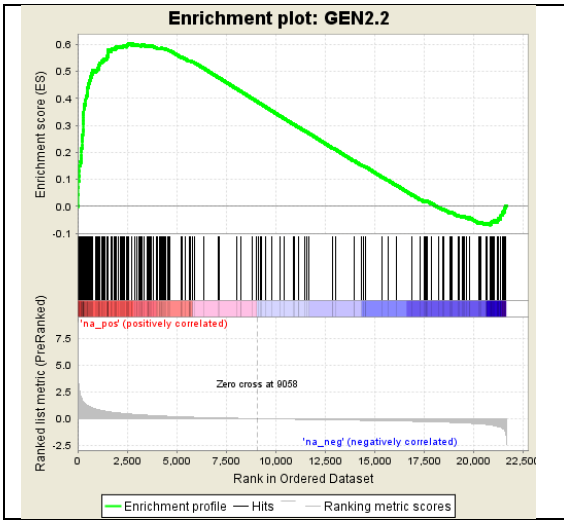
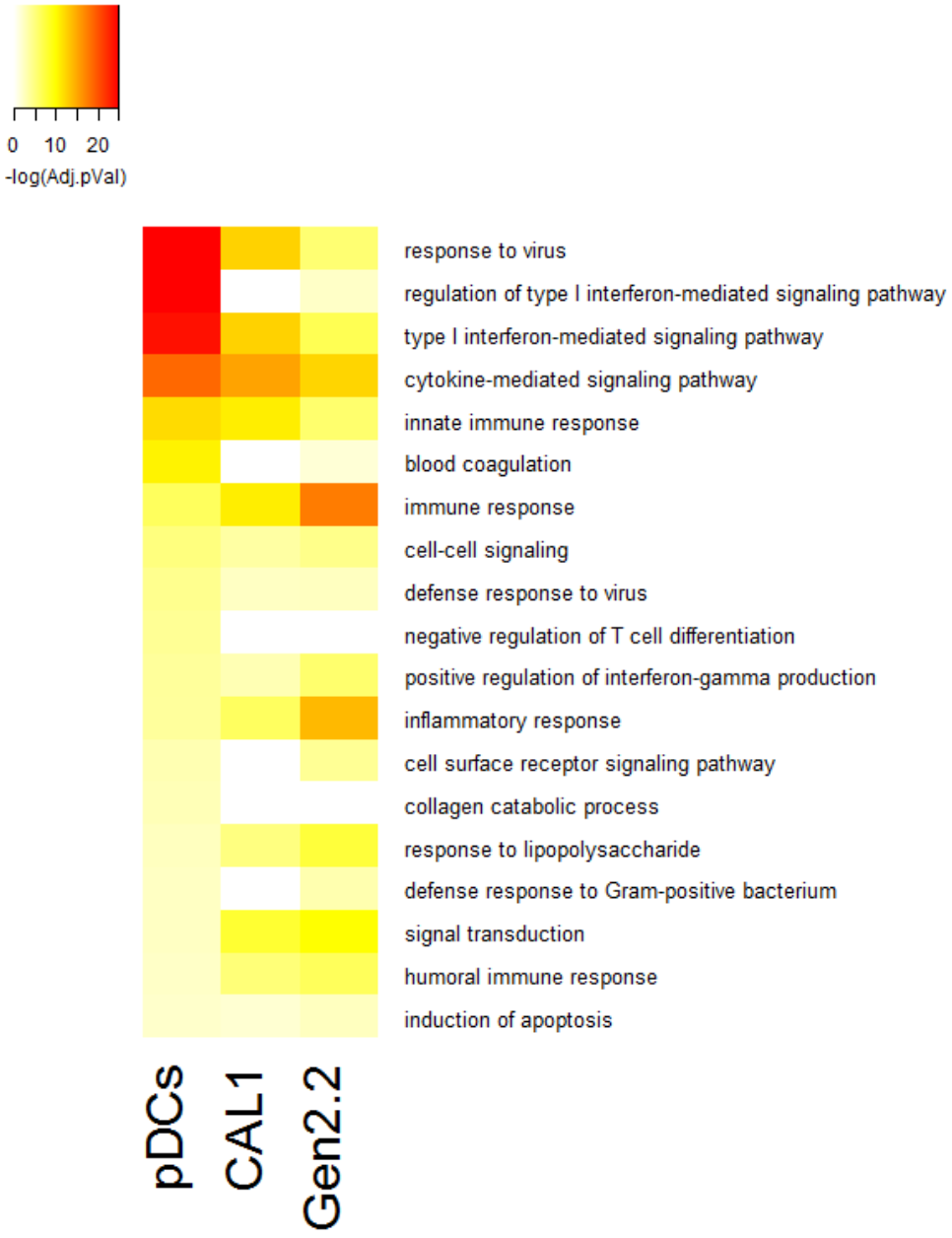


Table: GSEA Results Summary

Dataset	GSE55467.1000upFC
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_pos
GeneSet	GEN2.2
Enrichment Score (ES)	0.60368663
Normalized Enrichment Score (NES)	2.0900996
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

Figure S7. Heatmap representing biological processes significantly enriched in pDCs after stimulation. All GO terms that were found significantly enriched with a p-value < 0.01 in over-expressed genes in pDC after stimulation are shown. The color scale indicates the significance of the enrichment, from white (not enriched) to red (enriched).



9. REFERENCES

Allantaz,F. *et al.* (2012) Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PloS One*, **7**, e29979.

Brunet,J.-P. *et al.* (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci.*, **101**, 4164–4169.

Carmona-Saez,P. *et al.* (2006) Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics*, **7**, 78.

Cisse,B. *et al.* (2008) Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell*, **135**, 37–48.

Gaujoux,R. and Seoighe,C. (2010) A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, **11**, 367.

Haniffa,M. *et al.* (2012) Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. *Immunity*, **37**, 60–73.

Kim,H. and Park,H. (2007) Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinforma. Oxf. Engl.*, **23**, 1495–1502.

Kim,P.M. and Tidor,B. (2003) Subsystem identification through dimensionality reduction of large-scale gene expression data. *Genome Res.*, **13**, 1706–1718.

Matsui,T. *et al.* (2009) CD2 distinguishes two subsets of human plasmacytoid dendritic cells with distinct phenotype and functions. *J. Immunol. Baltim. Md 1950*, **182**, 6815–6823.

Robbins,S.H. *et al.* (2008) Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol.*, **9**, R17.

Steinhagen,F. *et al.* (2012) Activation of type I interferon-dependent genes characterizes the 'core response' induced by CpG DNA. *J. Leukoc. Biol.*, **92**, 775–785.

Subramanian,A. *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.

Suzuki,R. and Shimodaira,H. (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, **22**, 1540–1542.

Swiecki,M. *et al.* (2010) Plasmacytoid dendritic cell ablation impacts early interferon responses and antiviral NK and CD8(+) T cell accrual. *Immunity*, **33**, 955–966.