Single-cell RNA sequencing reveals developmental heterogeneity among early lymphoid progenitors

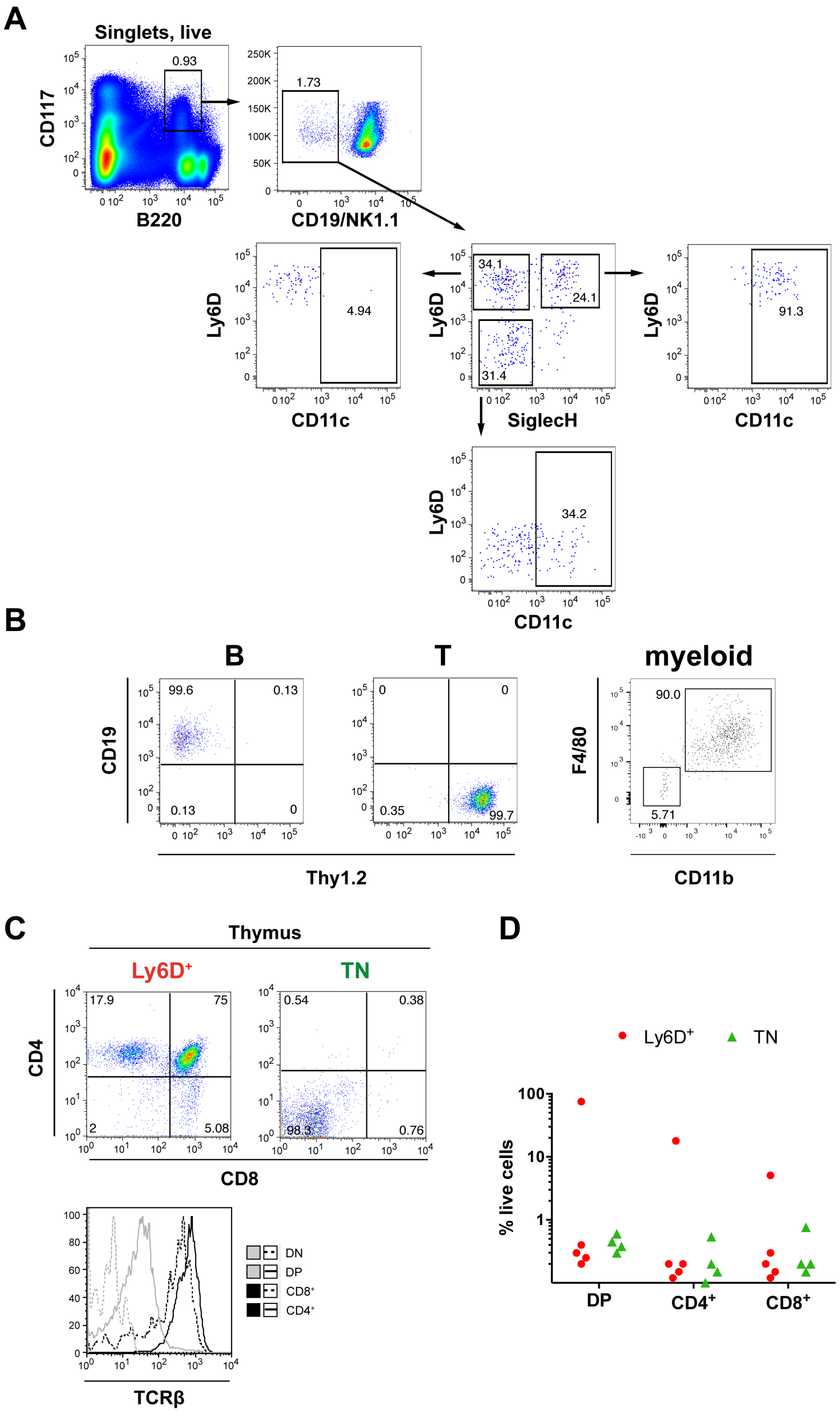
Table of contents

Appendix Figures S1-5……………………………………………… pages 2-8

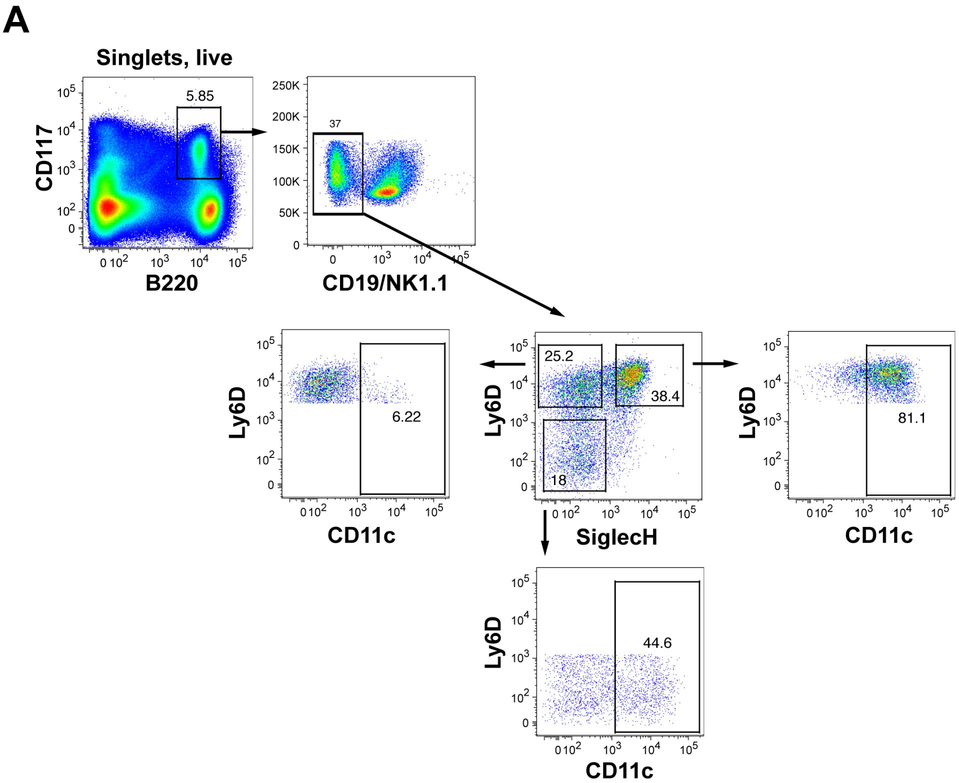
Appendix Supplementary Methods………………………………... pages 9-12

Appendix Supplementary References……………………………... page 13

Appendix Figures S1-5

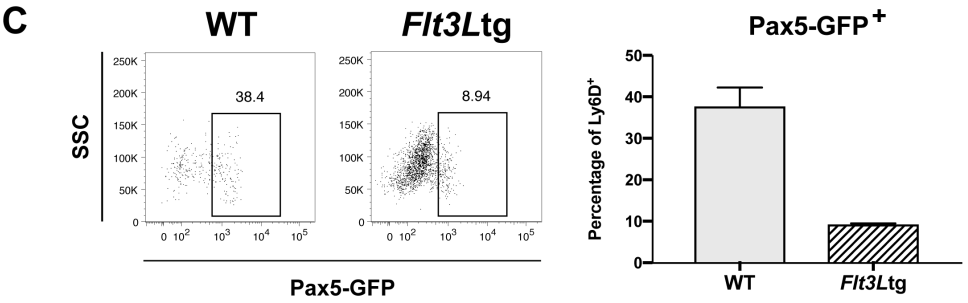


Appendix figure S1. (A) EPLM (B220$^+$CD117^{int}CD19$^-$NK1.1$^-$) FACS staining in WT mice (n=5) with the addition of Ly6D and SiglecH identifying three fractions (middle panel). CD11c expression for each of the fractions is shown. **(B)** Representative FACS staining of a B-cell clone (Ly6D$^+$ on OP9), a T-cell clone (Ly6D$^+$ on OP9-DL1) and a myeloid clone (TN on ST2). **(C,D)** Reconstitution of the T-cell compartment of sub-lethally irradiated *Rag2*-deficient mice with 4x10^3 Ly6D$^+$ (n=5) or TN (n=4) cells from WT. **(C)** CD4 and CD8 expression on thymocytes and TCRβ expression on DN, DP, CD8$^+$ and CD4$^+$ gated thymocytes 3 weeks after transfer. **(D)** Quantification of CD4$^+$ and CD8$^+$ thymocytes. Shown is mean ± SEM.

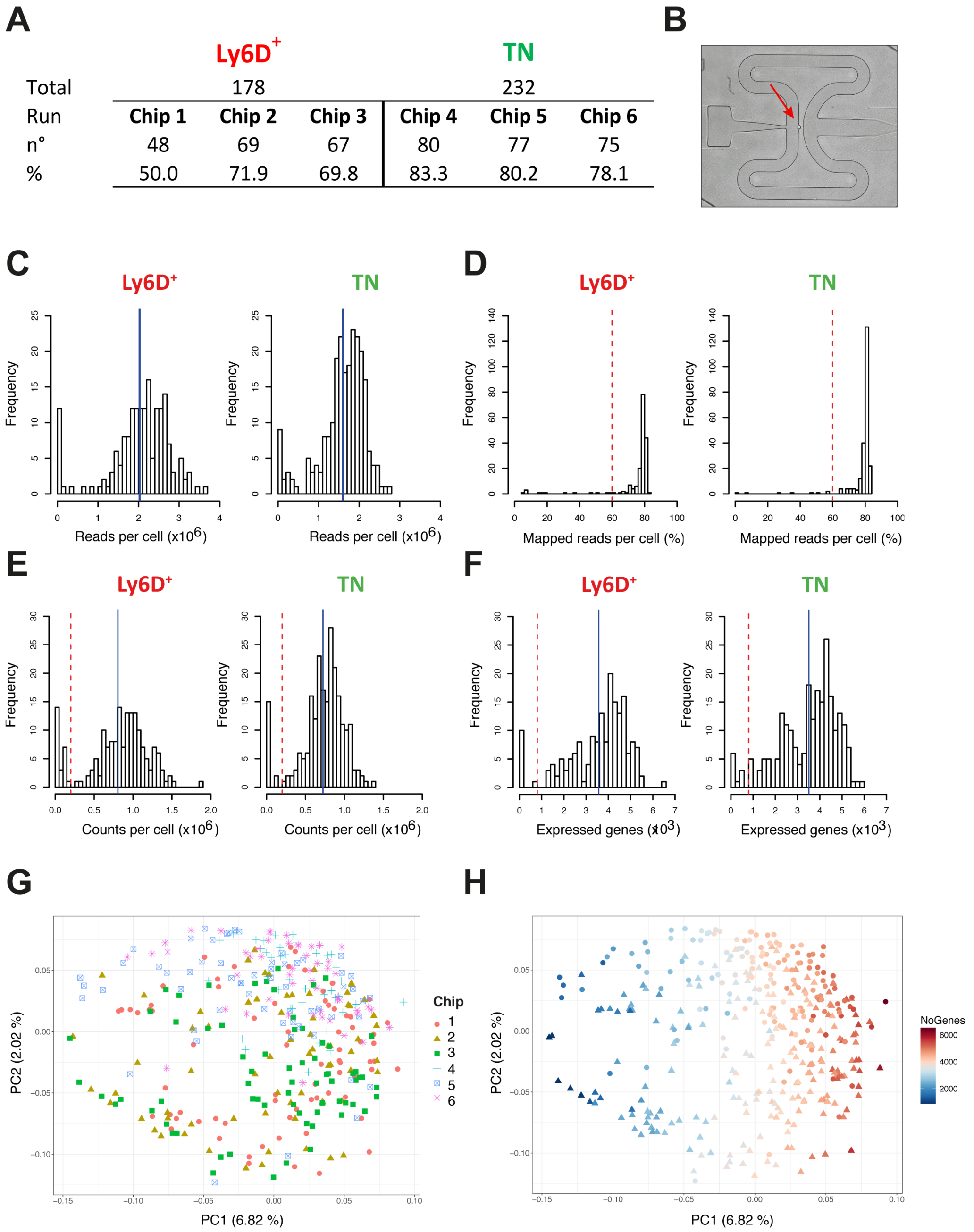


B

		WT			*Flt3L*tg		
		B-cell potential	T-cell potential	Myeloid potential	B-cell potential	T-cell potential	Myeloid potential
Ly6D$^+$	n	11	6	7	8	4	5
	mean freq	1 in 6.4	1 in 34	< 1 in 99	1 in 11	1 in 5.2	< 1 in 500
SiglecH$^+$	n	4	3	6	3	2	3
	mean freq	< 1 in 500	< 1 in 500	1 in 14	< 1 in 500	< 1 in 500	1 in 30
TN	n	7	3	3	6	3	4
	mean freq	1 in 34	1 in 55	1 in 25	1 in 70	1 in 8.6	1 in 15
CD11c$^+$	n	1	1	1	3	3	3
	mean freq	< 1 in 500	< 1 in 500	1 in 5	< 1 in 500	< 1 in 500	1 in 20

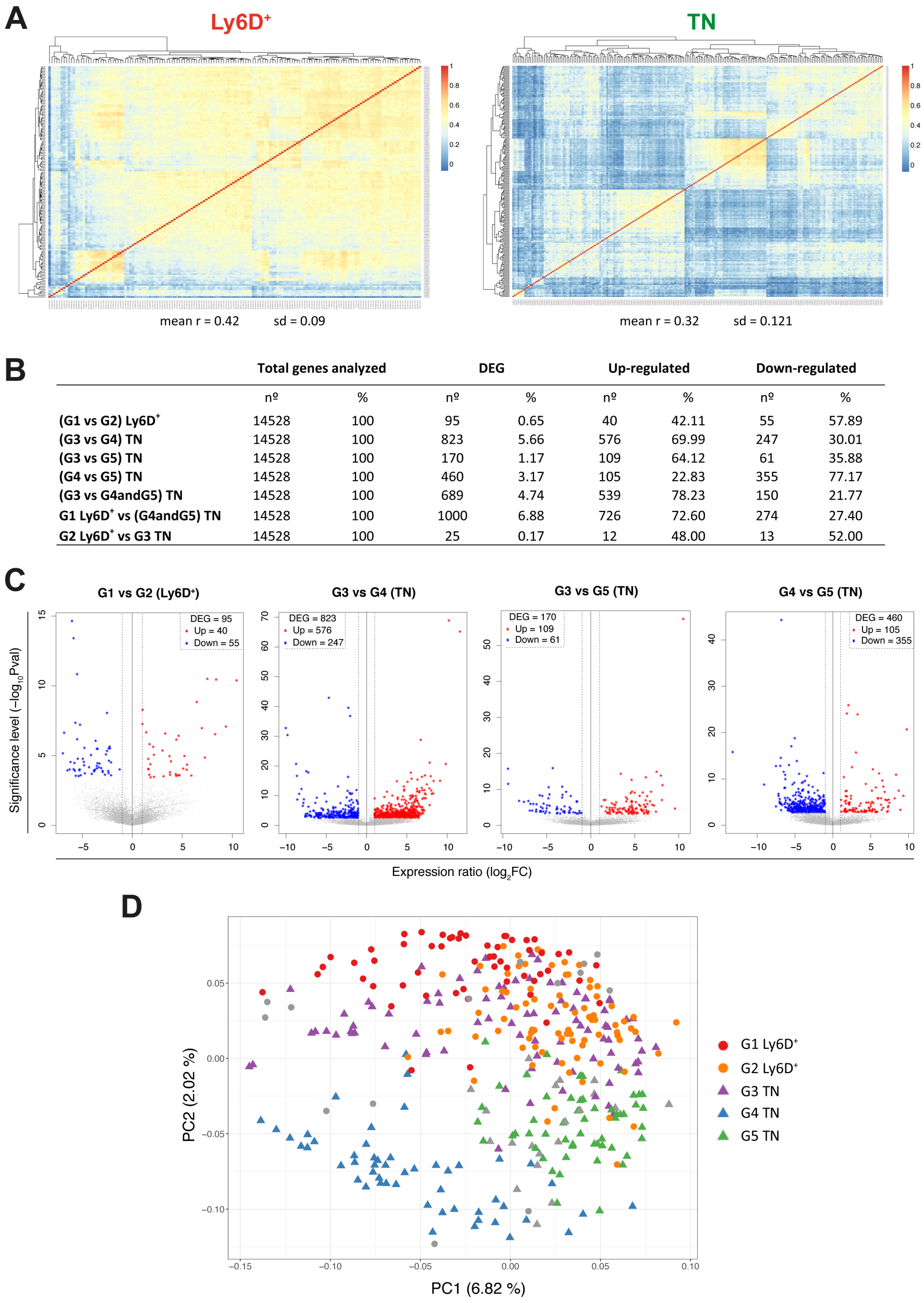


Appendix figure S2. (A) EPLM (B220$^+$ CD117^{int} CD19$^-$ NK1.1$^-$) FACS staining in *Flt3L*tg mice with the addition of Ly6D and SiglecH identifying three fractions (middle panel). CD11c expression for each of the fractions is shown. **(B)** Summary table of the mean frequencies of WT or *Flt3Lt*g EPLM subpopulations (indicated on the left) able to give rise to B-cells, T-cells and myeloid cells *in vitro*. Number of independent experiments is shown (n) for the different populations. **(C)** Pax5 expression within Ly6D$^+$ cells from Pax5-reporter mice (WT background or crossed to *Flt3L*tg mice). Representative FACS staining (left) and summary of percentages (n=3, right). Shown is mean ± SEM.



	Ly6D+			TN		
Total	178			232		
Run	**Chip 1**	**Chip 2**	**Chip 3**	**Chip 4**	**Chip 5**	**Chip 6**
n°	48	69	67	80	77	75
%	50.0	71.9	69.8	83.3	80.2	78.1

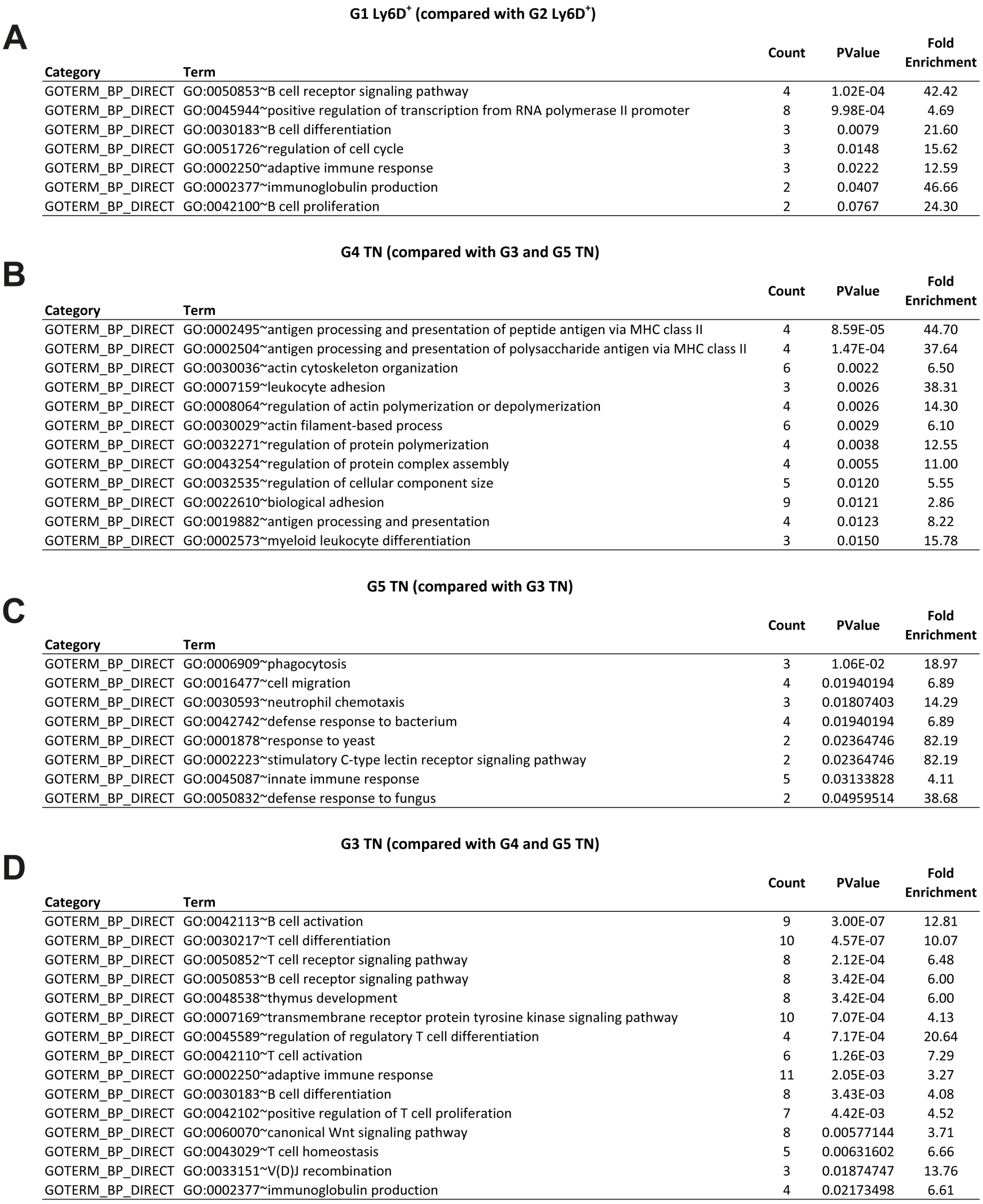
Appendix figure S3. (A) Single-cell capturing efficiency**.** Number (n°) and percentage (%) of single-cells captured per chip (run) or per population (total). **(B)** Representation of a single-cell captured in one of the 96 chambers of the C1 Fluidigm small chip. Picture taken with phase-contrast microscope in the second chip run of TN cells. **(C-F)** Quantification and quality control of raw sequenced data for Ly6D^{+} (left panels) and TN (right panels) single cells**.** Per cell distribution of **(C)** number of sequenced reads, **(D)** percentage of reads mapped to the mouse genome (mm9) out of the total number of reads, **(E)** counts (library size) considering reads mapped to genes (exons only) and, **(F)** total number of genes detected (with at least 1 count). Ly6D^{+} n=178; TN n=213. Blue line: mean; dotted red line: thresholds applied to the data (any cells not meeting these thresholds failed the quality control and were excluded from the analysis). **(G,H)** Principal component analysis using all detected genes (14,528) of the 152 Ly6D^{+} and 213 TN cells that passed the quality control. Cells are coloured according to the chip they were captured **(G)** or the number of detected genes (NoGenes) per cell **(H)**.



	Total genes analyzed		DEG		Up-regulated		Down-regulated	
	nº	%	nº	%	nº	%	nº	%
(G1 vs G2) Ly6D$^+$	14528	100	95	0.65	40	42.11	55	57.89
(G3 vs G4) TN	14528	100	823	5.66	576	69.99	247	30.01
(G3 vs G5) TN	14528	100	170	1.17	109	64.12	61	35.88
(G4 vs G5) TN	14528	100	460	3.17	105	22.83	355	77.17
(G3 vs G4andG5) TN	14528	100	689	4.74	539	78.23	150	21.77
G1 Ly6D$^+$ vs (G4andG5) TN	14528	100	1000	6.88	726	72.60	274	27.40
G2 Ly6D$^+$ vs G3 TN	14528	100	25	0.17	12	48.00	13	52.00

Appendix figure S4. (A) Heatmap with cell-to-cell Pearson's transcriptome correlation of Ly6D$^+$ (left) and TN (right) single cells using the 1008 differentially expressed genes from the bulk RNA-seq experiment when comparing Ly6D$^+$ with TN populations. Mean correlation value (mean r) and standard deviation (sd) is shown. **(B)** Summary table of differential expression analysis of the indicated transcriptome comparisons. The differentially expressed, up-regulated and down-regulated genes are shown in number and percentage. DEG are considered when absIlog$_2$(FoldChange)I >1 and false discovery rate (FDR) <0.05. **(C)**

Volcano plot (plotted significance against expression ratio) of the indicated pair-wise transcriptome comparisons. Each dot/star represents a gene. Grey dots: not DEGs; red star: up-regulated genes; blue stars: down-regulated genes. **(D)** PCA generated as in Sup. Fig. 3G showing that the subgroups revealed by the PAM clustering method are still well separated (mainly by PC2) in a PCA plot using all detected genes. G1 Ly6D$^+$ (n=56), G2 Ly6D$^+$ (n=82), G3 TN (n=85), G4 TN (n=52), G5 TN (n=56)



A

G1 Ly6D$^+$ (compared with G2 Ly6D$^+$)

Category	Term	Count	PValue	Fold Enrichment
GOTERM_BP_DIRECT	GO:0050853~B cell receptor signaling pathway	4	1.02E-04	42.42
GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	8	9.98E-04	4.69
GOTERM_BP_DIRECT	GO:0030183~B cell differentiation	3	0.0079	21.60
GOTERM_BP_DIRECT	GO:0051726~regulation of cell cycle	3	0.0148	15.62
GOTERM_BP_DIRECT	GO:0002250~adaptive immune response	3	0.0222	12.59
GOTERM_BP_DIRECT	GO:0002377~immunoglobulin production	2	0.0407	46.66
GOTERM_BP_DIRECT	GO:0042100~B cell proliferation	2	0.0767	24.30

B

G4 TN (compared with G3 and G5 TN)

Category	Term	Count	PValue	Fold Enrichment
GOTERM_BP_DIRECT	GO:0002495~antigen processing and presentation of peptide antigen via MHC class II	4	8.59E-05	44.70
GOTERM_BP_DIRECT	GO:0002504~antigen processing and presentation of polysaccharide antigen via MHC class II	4	1.47E-04	37.64
GOTERM_BP_DIRECT	GO:0030036~actin cytoskeleton organization	6	0.0022	6.50
GOTERM_BP_DIRECT	GO:0007159~leukocyte adhesion	3	0.0026	38.31
GOTERM_BP_DIRECT	GO:0008064~regulation of actin polymerization or depolymerization	4	0.0026	14.30
GOTERM_BP_DIRECT	GO:0030029~actin filament-based process	6	0.0029	6.10
GOTERM_BP_DIRECT	GO:0032271~regulation of protein polymerization	4	0.0038	12.55
GOTERM_BP_DIRECT	GO:0043254~regulation of protein complex assembly	4	0.0055	11.00
GOTERM_BP_DIRECT	GO:0032535~regulation of cellular component size	5	0.0120	5.55
GOTERM_BP_DIRECT	GO:0022610~biological adhesion	9	0.0121	2.86
GOTERM_BP_DIRECT	GO:0019882~antigen processing and presentation	4	0.0123	8.22
GOTERM_BP_DIRECT	GO:0002573~myeloid leukocyte differentiation	3	0.0150	15.78

C

G5 TN (compared with G3 TN)

Category	Term	Count	PValue	Fold Enrichment
GOTERM_BP_DIRECT	GO:0006909~phagocytosis	3	1.06E-02	18.97
GOTERM_BP_DIRECT	GO:0016477~cell migration	4	0.01940194	6.89
GOTERM_BP_DIRECT	GO:0030593~neutrophil chemotaxis	3	0.01807403	14.29
GOTERM_BP_DIRECT	GO:0042742~defense response to bacterium	4	0.01940194	6.89
GOTERM_BP_DIRECT	GO:0001878~response to yeast	2	0.02364746	82.19
GOTERM_BP_DIRECT	GO:0002223~stimulatory C-type lectin receptor signaling pathway	2	0.02364746	82.19
GOTERM_BP_DIRECT	GO:0045087~innate immune response	5	0.03133828	4.11
GOTERM_BP_DIRECT	GO:0050832~defense response to fungus	2	0.04959514	38.68

D

G3 TN (compared with G4 and G5 TN)

Category	Term	Count	PValue	Fold Enrichment
GOTERM_BP_DIRECT	GO:0042113~B cell activation	9	3.00E-07	12.81
GOTERM_BP_DIRECT	GO:0030217~T cell differentiation	10	4.57E-07	10.07
GOTERM_BP_DIRECT	GO:0050852~T cell receptor signaling pathway	8	2.12E-04	6.48
GOTERM_BP_DIRECT	GO:0050853~B cell receptor signaling pathway	8	3.42E-04	6.00
GOTERM_BP_DIRECT	GO:0048538~thymus development	8	3.42E-04	6.00
GOTERM_BP_DIRECT	GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway	10	7.07E-04	4.13
GOTERM_BP_DIRECT	GO:0045589~regulation of regulatory T cell differentiation	4	7.17E-04	20.64
GOTERM_BP_DIRECT	GO:0042110~T cell activation	6	1.26E-03	7.29
GOTERM_BP_DIRECT	GO:0002250~adaptive immune response	11	2.05E-03	3.27
GOTERM_BP_DIRECT	GO:0030183~B cell differentiation	8	3.43E-03	4.08
GOTERM_BP_DIRECT	GO:0042102~positive regulation of T cell proliferation	7	4.42E-03	4.52
GOTERM_BP_DIRECT	GO:0060070~canonical Wnt signaling pathway	8	0.00577144	3.71
GOTERM_BP_DIRECT	GO:0043029~T cell homeostasis	5	0.00631602	6.66
GOTERM_BP_DIRECT	GO:0033151~V(D)J recombination	3	0.01874747	13.76
GOTERM_BP_DIRECT	GO:0002377~immunoglobulin production	4	0.02173498	6.61

Appendix figure S5. Selection of enriched biological processes (BP) in up-regulated genes of the G1 Ly6D$^+$ **(A)**, G2 TN **(B)**, G3 TN **(C)**, G1 TN **(D)** groups of cells compared with the subgroups indicated in brackets. Complete lists in **Table EV3**.

Appendix Supplementary Methods

Flow cytometry and cell sorting

Bone marrow cell suspensions were obtained from both femurs of individual mice. Bones were flushed with a syringe in PBS containing 0.5% BSA and 5mM EDTA. Afterwards, single-cell suspensions were subjected to ACK treatment for erythrocyte depletion, stained with the appropriate combination of antibodies for 30 minutes at 4°C, and washed for subsequent flow cytometry or cell sorting. The following antibodies were used (from BD Pharmingen, eBioscience, BioLegend, or produced in house): anti-B220 (RA3-6B2), anti-CD117 (2B8), anti-CD19 (1D3), anti-NK1.1 (PK136), anti-SiglecH (551), anti-CD11c (HL3), anti-Ly6D (49-H4), conjugated with FITC, PE, PE/Cy7, APC, BV421 or Biotin. Biotin-labelled antibodies were revealed using streptavidin-BV650. Analytical flow cytometry was performed using a BD LSR Fortessa (BD Biosciences) and data were analyzed using FlowJo v9.8 Software (Treestar). For cell sorting, a FACS Aria IIu (BD Biosciences) was used and in all instances, sorted bulk cells were >98% pure.

In vitro limiting dilution assay

ST2 (Ogawa et al, 1988), OP9 (Nakano et al, 1994) and OP9-DL1 (Schmitt & Zuniga-Pflucker, 2002) were plated at 3000 cells per well in a 96-well flat-bottom plate one day prior to co-culture. The following day, semi-confluent stromal cells were γ-irradiated with 3000 rad using a Cobalt source (Gammacell 40, Atomic Energy of Canada, Ltd) at 100 rad/min and co-cultured with graded numbers of sorted hematopoietic progenitors in 48 replicates (or as indicated). Cells were maintained as a monolayer in IMDM supplemented with $5x10^{-5}$M β-mercaptoethanol, 1mM glutamine, 0.03% w/v Primatone (Quest Naarden, The Netherlands), 100U/mL Penicillin, 100 µg/mL Streptomycin and 5% FBS (Amimed) at 37°C in a humidified atmosphere containing 10% CO_2 in the air. OP9 and OP9-DL1 co-cultures where additionally supplemented with 100U/ml IL-7. After 10 days (for OP9 cell cultures) or 15 days (for OP9-DL1 and ST2 cell cultures), wells were inspected using an inverted microscope and the colony morphology was confirmed by phenotype with flow cytometry stainings (Appendix Fig S1B). Wells containing colonies of more than 50 cells were scored as positive. For each experiment, minimal estimates of precursor frequencies were obtained by the minimum chi-square method from the Poisson distribution relationship between the responding cell number and the logarithm of the percentage of non-responding cultures.

Bulk RNA sequencing

RNA extraction and quality

Total RNA was extracted from *ex-vivo* sorted samples using TRIzol-based method (Chomczynski & Sacchi, 1987; Chomczynski & Sacchi, 2006). Briefly, $1x10^5$ to $3x10^5$ cells were lysed in 0.5ml of TRIzol reagent and 0.1ml of chloroform was added per 0.5ml TRI reagent. After incubation and centrifugation for phase separation, the aqueous phase containing the RNA was recovered and mixed with isopropanol in a 1:1 ratio for RNA precipitation. Following 15min incubation and centrifugation, the supernatant was discarded while the RNA pellet was first washed with 75% ethanol and subsequently resuspended with 20ul of DEPC treated water. Concentration and 260/280 purity ratio was initially determined using NanoDrop 1000 Spectrophotometer (Witec AG). Selected RNA samples with 500ng were sent for quality control, library preparation and sequencing to the Genomics Facility (D-

BSSE, Basel). Quality and level of degradation of the extracted RNA was assessed with RNA integrity number (RIN) assigned by the Agilent 2100 Bioanalyzer instrument using either the Nano or the Pico Agilent RNA 6000 kit (Agilent Technologies). Samples with a RIN value over 8 and presenting clean peaks were considered for further analysis. The RNA quantity was measured by the Infinite M1000 PRO - Tecan instrument using the Quant-iT RiboGreen RNA Assay Kit.

Library preparation and sequencing

For the preparation of sequencing libraries, the TruSeq Stranded mRNA LT Sample Preparation kit was used following the manufacturer's guide (Borodina et al, 2011). Size and purity of the library fragments was assessed by the Fragment Analyzer using the NGS Fragment 1-6000bp method, while quantification was done with Quant-iT PicoGreen dsDNA Assay Kit; TEcan instrument. Indexed DNA libraries were pooled in equal volumes and loaded on one NextSeq 500 High Output flow cell (Illumina). Single-end sequencing was performed on the Illumina NextSeq™ 500 Sequencing System (D-BSSE, Basel) for 81 cycles yielding in 21 to 35 millions of reads, 81-mers, per sample**.**

Pre-processing of sequencing data

A quality control of the sequenced data was performed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, version 0.11.3) and all samples showed good sequencing quality. The downstream analysis was performed using the open source software R accessed via RStudio server (R version 3.2.0). Sequencing reads were aligned to the mouse genome assembly, version mm9 (downloaded from UCSC http://genome.ucsc.edu), with SpliceMap (Au et al, 2010; Langmead et al, 2009), included in the R/Bioconductor package QuasR, version 1.14.0. Splice-map was also capable of mapping reads that cover exon junctions. More than 80% of total reads were successfully mapped for each sample. Subsequently, a count table with gene expression levels was generated using the qCount function from QuasR package and coordinates of RefSeq mRNA genes (http://hgdownload.soe.ucsc.edu/downloads.html#mouse, downloaded in December 2013). The expression level was defined as a number of reads that started within any annotated exon of a gene (exon-union model). Total counts per sample ranged from 13 to 22 millions, the so-called library size. Genes with no counts across all samples were filtered out from the analysis. For 17,290 genes at least 1 read was detected across all samples, corresponding to ~14,800 genes per sample.

Raw counts were normalized between samples with the TMM method (weighted trimmed mean of M-values (Robinson & Oshlack, 2010)), expressed as counts per million mapped reads (CPM), and transformed to the log2-scale ($\log_2$CPM).

Data analysis

Differential expression analysis was performed using edgeR v3.12.1 (Robinson et al, 2010). A prior count of 8 was used in order to minimize the large log-fold changes for genes with small number of counts. Genes with a false discovery rate (FDR)<0.05 and abslog2(FoldChange)l>1 were considered as differentially expressed. The heatmap with sample pair-wise Pearson's correlation coefficients was generated with the top 50% of genes with highest variance across analysed dataset (calculated as inter-quartile range) and visualized with the NMF v0.20.6 R package. The MA plot was produced using custom R scripts.

Single-cell RNA sequencing

Capture of single cells

Single cells were captured from *ex-vivo* sorted hematopoietic progenitors on a small-sized (5-10µm) "C1 Single-Cell Auto Prep IFC for mRNA sequencing" (Fluidigm) using the Fluidigm C1 system. Cell diameter of the Ly6D$^+$ and TN cells, imaged on Leica DMI 4000 microscope and measured with ImageJ software, was similar and homogeneous, namely 8.54µm and 8.77µm respectively. Therefore, no bias due to cell size or morphology was expected during the capturing procedure. Cells were loaded onto the chip at a concentration of ~300 cells/µl as recommended by the manufacturer and the 96 chambers were inspected by phase-contrast microscopy to determine the number of captured single cells. A total of 3 chips per population were used yielding to 178 Ly6D$^+$ and 232 TN single cells captured (Appendix Fig S3A and B). Subsequently, cells were lysed, the polyA containing mRNA molecules were hybridized to oligo-dT and whole-transcriptome full-length amplified cDNA was prepared by template switching on the C1 fluidigm chip using the SMARTer Ultra Low RNA kit for the Fluidigm C1 System (Clontech). Quantification of cDNA was done with Quant-iT PicoGreen dsDNA Assay Kit; TEcan instrument.

Library preparation and sequencing

Illumina single-cell libraries were constructed in 96-well plates using the Nextera XT DNA Library Preparation Kit (Illumina) following the protocol supplied by Fluidigm ("Using C1 to Generate Single-Cell cDNA Libraries for mRNA Sequencing"). Briefly, 0.1-0.3ng of harvested cDNA was subjected to tagmentation, a process in which the DNA fragmentation and sequencing adapter ligation occurs in a single step performed by the Nextera XT transposome, followed by purification with AMPure XP beads. Indexed DNA libraries originated from single cells captured in 3 different chips (288 libraries) were pooled in equal volumes and loaded on one NextSeq 500 High Output flow cell (Illumina). Single-end sequencing was performed on the Illumina NextSeq™ 500 Sequencing System (D-BSSE, Basel) for 76 cycles. Only FastQ files corresponding to C1 chambers with a single cell, previously determined by phase-contrast microscopy, were selected for downstream analysis, thus excluding doublets, debris or empty chambers. We obtained a total of 360 and 371 millions of reads for the Ly6D$^+$ and TN cells, respectively. The average number of reads per cell was 2x10^6 for the Ly6D$^+$ and 1.6x10^6 for the TN **(Appendix Fig S3C)**.

Pre-processing of sequencing data

All downstream analysis was performed using the open source R software accessed via RStudio server (R version 3.2.0). Sequencing reads were aligned and count table generated as in the bulk RNA-seq experiment explained above. Approximately 80% of total reads were successfully mapped for each sample **(Appendix Fig S3D).** Total counts per cell were approximately 8.1x10^5 for the Ly6D$^+$ and 7.2x10^{5}for the TN **(Appendix Fig S3E)**, the so-called library size. Genes with no counts across all samples were excluded from the analysis. At least one read was detected for a total of 14,814 genes across all 410 captured cells, corresponding to approximately 3,500 expressed genes per cell in both Ly6D$^+$ and TN **(Appendix Fig S3F)**. During the quality control, cells having less than 60% of mapped reads, less than 2x10^5 counts, or less than 800 detected genes were filtered out from further analysis (dotted red lines in **Appendix Fig S3**). In total, 89% of the cells (152 Ly6D$^+$ and 213 TN) passed these criteria. Raw counts were normalized between cells and genes, expressed as fragments per kilobase of transcript per million mapped reads (FPKM). For visualization purposes, 1 was added to FPKM values and transformed to the log2-scale (log$_2$FPKM).

Data analysis

If not otherwise specified, the downstream analysis was performed using the 1008 DEG (false discovery rate (FDR)<0.05 and abslllog$_2$(FoldChange)l>1) from the bulk RNA-seq experiment when comparing Ly6D$^+$ with TN populations.

Dimensionality reduction was performed with PCA analysis. Average gene expression was centered to zero and PCA plots were generated with the ggplot2 v2.1.0 R package. To visualize the degree of cell-to-cell heterogeneity, an annotated heatmap of sample pair-wise Pearson's correlation coefficients was produced using the NMF v0.20.6 R package. Eight Ly6D$^+$ cells were not considered for subsequent clustering because of their very low transcriptome correlation to any other cell, on average less than 0.3 (**Appendix Fig S4A** left). Cell clustering was performed using the Partitioning Around Medoids (PAM) method implemented in the cluster v2.0.4 R package (Reynolds et al, 2006). Gene expression was first centered (mean=0) and distances were calculated from cell-to-cell Pearson's correlation values using the Euclidean method. The optimal number of clusters was selected based on silhouette plot, which for Ly6D$^+$ corresponded to K=2 (with average silhouette width of 0.10) and K=3 for the TN (with average silhouette width of 0.13). Cells with negative silhouette width values were excluded while the other 331 cells were assigned to one of the 5 groups. Average expression across all detected genes was calculated for each of the five clusters of single cells, and a heatmap with Pearson's correlation coefficients was generated with the top 50% of genes with highest variance across analyzed datasets (calculated as inter-quartile range) and visualized with the NMF v0.20.6 R package.

Differential gene-expression analysis to compare the clustered groups of cells was performed using the 14,528 detected expressed genes across the 331 single cells with edgeR v3.12.1 (Robinson et al, 2010). A prior count of 0.5 was added to all gene counts in order to minimize the large log-fold changes for genes with a small number of counts. Genes with a FDR<0.05 and abslllog$_2$(FoldChange)l>1 were considered as differentially expressed. Volcano, violin and scatter plots were produced using custom R scripts.

Gene-ontology enrichment analysis was performed with the DAVID 6.8 bioinformatics database, based on Fisher's Exact method (Huang da et al, 2009a; Huang da et al, 2009b).

Appendix Supplementary References

Au KF, Jiang H, Lin L, Xing Y, Wong WH (2010) Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res* **38:** 4570-4578

Borodina T, Adjaye J, Sultan M (2011) A strand-specific library preparation protocol for RNA sequencing. *Methods in enzymology* **500:** 79-98

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry* **162:** 156-159

Chomczynski P, Sacchi N (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc* **1:** 581-585

Huang da W, Sherman BT, Lempicki RA (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37:** 1-13

Huang da W, Sherman BT, Lempicki RA (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44-57

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25

Nakano T, Kodama H, Honjo T (1994) Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265:** 1098-1101

Ogawa M, Nishikawa S, Ikuta K, Yamamura F, Naito M, Takahashi K, Nishikawa S (1988) B cell ontogeny in murine embryo studied by a culture system with the monolayer of a stromal cell clone, ST2: B cell progenitor develops first in the embryonal body rather than in the yolk sac. *EMBO J* **7:** 1337-1343

Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ (2006) Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms* **5:** 475-504

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139-140

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11:** R25

Schmitt TM, Zuniga-Pflucker JC (2002) Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity* **17:** 749-756