SUPPLEMENTARY APPENDIX

Cell-intrinsic depletion of Aml1-ETO-expressing pre-leukemic hematopoietic stem cells by K-Ras activating mutation

Cristina Di Genua,[1] Ruggiero Norfo,[1] Alba Rodriguez-Meira,[1] Wei Xiong Wen,[1,2] Roy Drissen,[1] Christopher A.G. Booth,[1] Benjamin Povinelli,[1] Emmanouela Repapi,[3] Nicki Gray,[3] Joana Carrelha,[1] Laura M. Kettyle,[1] Lauren Jamieson,[1] Wen Hao Neo,[1] Supat Thongjuea,[1,2] Claus Nerlov,[1] and Adam J. Mead[1]

[1]MRC Molecular Haematology Unit; [2]WIMM Centre for Computational Biology and [3]Computational Biology Research Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

CN and AJM contributed equally to this work.

 doi:10.3324/haematol.2018.205351

Received: September 19, 2018.

Accepted: April 9, 2019.

Pre-published: May 10, 2019.

Correspondence: *ADAM J. MEAD* - adam.mead@imm.ox.ac.uk

CLAUS NERLOV - claus.nerlov@imm.ox.ac.uk

Supplemental information for:

Cell-intrinsic depletion of Aml1-ETO-expressing pre-leukemic hematopoietic stem cells by K-Ras activating mutation

Cristina Di Genua,[1] Ruggiero Norfo,[1] Alba Rodriguez-Meira,[1] Wei Xiong Wen,[1,2] Roy Drissen,[1] Christopher A.G. Booth,[1] Benjamin Povinelli,[1] Emmanouela Repapi,[3] Nicki Gray,[3] Joana Carrelha,[1] Laura M. Kettyle,[1] Lauren Jamieson,[1] Wen Hao Neo,[1] Supat Thongjuea,[1,2] Claus Nerlov,[1] and Adam J. Mead[1]

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.
[2]MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[3]Computational Biology Research Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

C.N and A.J.M contributed equally to this work

Supplementary material and methods

Competitive transplantation

Congenic CD45.1 WT C57Bl/6J adult mice (7-13 weeks old) were utilized as recipients. Competitive transplantations were performed using $2.5x10^5$ CD45.2 FL cells and $1x10^6$ CD45.1 WT BM competitor cells into lethally irradiated recipients (2 x 500rads). Recombination was induced by subcutaneous injection of poly(I:C) four weeks post-transplantation, and mice culled eight weeks post-poly(I:C) and hematopoietic phenotype analyzed.

RNA-sequencing

cDNA synthesis and PCR amplification were performed based on the published Smart-seq 2 protocol[1] with some modifications. SMARTScribe RT enzyme (Takara Bio USA, Inc) was used in the RT mix (50 U) and SeqAMP enzyme (Takara Bio USA, Inc, 50 U) was used for the PCR step for 18 cycles. cDNA traces were bead-purified using Ampure XP beads (Beckman Coulter, California, USA) and evaluated using a High Sensitivity DNA kit (Agilent, California, USA) on an Agilent 2100 Bioanalyzer. Normalized cDNA traces were used for library preparation using a miniaturized version of the Nextera XT Kit (Illumina, California, USA). After tagmentation and 12 cycles of barcoding PCR, tagmented libraries were purified using AmpureXP beads, evaluated using a High Sensitivity DNA kit on an Agilent 2100 Bioanalyzer and quantified using Qubit (Invitrogen, California, USA). Finally, libraries were pooled and sequenced on four lanes on a NextSeq 500 (Illumina), using 75 bp single-end reads, with a sequencing depth of approximately 17 million reads per sample.

Following QC analysis with the fastQC package (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), reads were aligned using STAR[2] against the mouse genome assembly (mouse NCBI build38 (mm10) UCSC transcripts). Gene expression levels were quantified as read counts using the featureCounts function from the Subread package[3] with default parameters. The quality of the data was assessed using the

software RNA-SeQC[4] and multiQC.[5] The read counts were used for the identification of global differential gene expression between genotypes using the edgeR package.[6] Reads per kilobase of transcript per million (RPKM) values were also generated using the edgeR package. Genes were considered differentially expressed between populations if they had an adjusted p-value (FDR) of less than 0.05. Gene set enrichment analysis (GSEA) was performed using GSEA software.[7, 8] HSC signature genes from Ivanova *et al*. [9] was retrieved from class c2 in MSigDB[10] and used as input for GSEA comparing AKM vs. AM. Microarray dataset for GMPs lacking *Gata1* expression and HSCs were similarly retrieved from a previous publication (GEO accession number: GSE49241).[11] Pair-wise comparison was performed for each gene using *t*-test and genes with FDR < 0.05 were subsequently used as input for GSEA comparing AKM vs. AM. Venn-diagrams were generated using Venny (http://bioinfogp.cnb.csic.es/tools/venny/). The pheatmap function was used to generate a heatmap in R statistical programming environment (www.r-project.org). Highly variable genes from each genotype were used in the principal component analysis (PCA). For each genotype, the coefficient of variance (CV), variance/square of mean, for each gene was computed and the top 10% of genes with the highest CV with mean of RPKM ≥ 1 were tabulated. In total, 2,495 unique genes were used in the PCA. Genes identified to be differentially expressed across the different genotypes were used in the PCA. F-test from the genefilter R package was used to identify genes that were differentially expressed across the genotypes and genes with false discovery rate (FDR) < 0.05 were considered to be differentially expressed. In total, 379 genes were identified to be differentially expressed. The same gene set was used in the unsupervised hierarchical clustering. The dendrogram was drawn using pvclust R package with the options “euclidean” and “complete” for the distance measure and agglomerative method, respectively.[12] AU (Approximately unbiased) p-values were calculated using multi scale bootstrap re-sampling while BP (bootstrapping probability) p-values were calculated using ordinary bootstrapping re-sampling. Functional annotation clustering was performed on differentially up- and down-regulated genes respectively using GOstats.[13] Functional clusters with p-values < 0.05 after Bonferroni correction were

considered to be statistically significant. RNA sequencing data are available through the GEO repository by accession number GSE112024.

Table S1. Anti-mouse antibodies and viability stains used for FACS staining.

Antigen	Conjugate	Company	Catalogue number	Staining Panel
7-Aminoactinomycin D (7AAD)	PE-Cy5	Cayman Chemical	11397	Lineage, myelo-erythroid progenitors, HSC analysis
CD4	APC-eF780	eBioscience	47-0042-82	Lineage
CD8a	APC-eF780	eBioscience	47-0081-82	Lineage
NK1.1	PB	BioLegend	108722	Lineage
Gr1	PO	Caltag Medsystems	RM3030	Lineage
CD19	PE-Cy7	eBioscience	25-0193-82	Lineage
Mac1	APC	BioLegend	101212	Lineage
CD45.1	PE	eBioscience	12-0453-83	Lineage
CD45.2	AF700	BioLegend	109822	Lineage, myelo-erythroid progenitors, HSC analysis, HSC cell cycle, HSC RNA sequencing/LSK sort
FcγR	PE	eBioscience	12-0161-81	Myelo-erythroid progenitors
CD4	PE-Cy5	BioLegend	100514	Myelo-erythroid progenitors
CD8a	PE-Cy5	BioLegend	100710	Myelo-erythroid progenitors
Ter119	PE-Cy5	BioLegend	116210	Myelo-erythroid progenitors, HSC analysis
Mac1	PE-Cy5	BioLegend	101210	Myelo-erythroid progenitors
CD19	PE-Cy5	eBioscience	15-0193-83	Myelo-erythroid progenitors, HSC analysis
Gr1	PE-Cy5	BioLegend	108410	Myelo-erythroid progenitors, HSC analysis
CD150	APC	BioLegend	115910	Myelo-erythroid progenitors
cKit	APC-eF780	eBioscience	47-1171-82	Myelo-erythroid progenitors, HSC analysis, HSC cell cycle, HSC RNA sequencing/LSK sort
Sca1	BV786	BD	563991	Myelo-erythroid progenitors
CD105	BV421	BD	562760	Myelo-erythroid progenitors
CD41	PE-Cy7	eBioscience	25-0411-82	Myelo-erythroid progenitors
CD45.1	BV650	BioLegend	110736	Myelo-erythroid progenitors, HSC analysis, HSC RNA sequencing/LSK sort
CD48	APC	BioLegend	103412	HSC analysis
F4/80	PE-Cy5	BioLegend	123112	HSC analysis
CD3e	PE-Cy5	BioLegend	100310	HSC analysis
NK1.1	PE-Cy5	BioLegend	108716	HSC analysis
CD150	PE-Cy7	BioLegend	115914	HSC analysis, HSC cell cycle, HSC RNA sequencing/LSK sort
Sca1	PB	BioLegend	108120	HSC analysis, HSC RNA sequencing/LSK sort
CD34	Biotin	eBioscience	13-0341-85	HSC analysis
Streptavidin	PE-TxRed	BD	551487	HSC analysis
Flt3	PE	BioLegend	135306	HSC analysis, HSC RNA sequencing/LSK sort
CD45.1	FITC	BioLegend	110706	HSC cell cycle
Ki-67	PE	BD	556027	HSC cell cycle
Flt3	APC	BioLegend	135309	HSC cell cycle
Sca1	PerCP-Cy5.5	eBioscience	45-5981-80	HSC cell cycle

4',6-diamidino-2-phenylindole (DAPI)	BV421	Thermo Fisher Scientific	D3571	HSC cell cycle
F4/80	APC	BioLegend	123116	HSC RNA sequencing/LSK sort
CD3e	APC	eBioscience	17-0031-83	HSC RNA sequencing/LSK sort
NK1.1	APC	eBioscience	17-5941-82	HSC RNA sequencing/LSK sort
Ter119	APC	eBioscience	17-5921-81	HSC RNA sequencing/LSK sort
Gr1	APC	BioLegend	108412	HSC RNA sequencing/LSK sort
CD19	APC	eBioscience	17-0193-82	HSC RNA sequencing/LSK sort

Table S1. Antibodies used for FACS staining. Antibodies used for each staining panel are indicated in the staining panel column. HSC, hematopoietic stem cell; LSK, Lineage$^-$ Sca1$^+$cKit$^+$.

Table S2. Hematopoietic phenotype.

Genotype	RBC (x10^{12}/L)	HCT	MCV (fL)	MCH (pg)	MCHC (g/dL)
CON	9.80 ± 1.24	0.48 ± 0.13	51.03 ± 2.73	16.03 ± 0.92	31.49 ± 2.88
AM	9.70 ± 1.15	0.49 ± 0.06	50.56 ± 3.12	16.27 ± 1.82	32.05 ± 4.35
KM	7.18 ± 1.85	0.34 ± 0.11	52.90 ± 6.60	16.46 ± 2.43	31.23 ± 3.61
AKM	9.42 ± 2.19	0.46 ± 0.10	49.24 ± 4.07	15.84 ± 2.63	32.61 ± 5.37

Table S2. Hematopoietic phenotype in the PB eight weeks post-poly(I:C) from recipients of CON (n=13), AM (n=14), KM (n=12) and AKM FL (n=14). Average ± standard deviation for RBC, red blood cell; HCT, hematocrit; MCV; Mean cell volume = HCT/RBC; MCH, average amount of hemoglobin (HGB) per red cell = HGB/RBC; and MCHC, mean cell hemoglobin concentration = HGB/HCT.

Table S3. Enrichment results for the hallmark gene set for AM vs. CON, AKM vs. AM, AKM vs. CON, and KM vs. CON.

Table S4: GO term results for up-regulated and down-regulated biological processes for AKM vs AM.

Figure S1

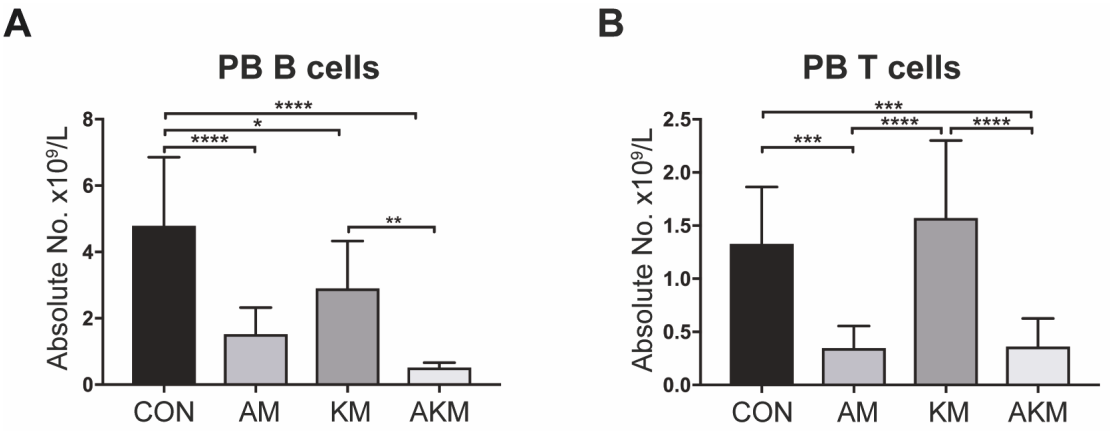

Figure S1. B and T cells in the PB. (A, B) Absolute number of CD45.2 B (A) and T cells (B) in the PB from recipients of CON (n=8), AM (n=10), KM (n=7) and AKM FL (n=10) in the PB. Results were generated in two independent experiments. The results were analyzed using multiple comparison ANOVA. The results are presented as the mean ± SD. *P<0.05 **P<0.01 ***P<0.001 ****P<0.0001.

Figure S2

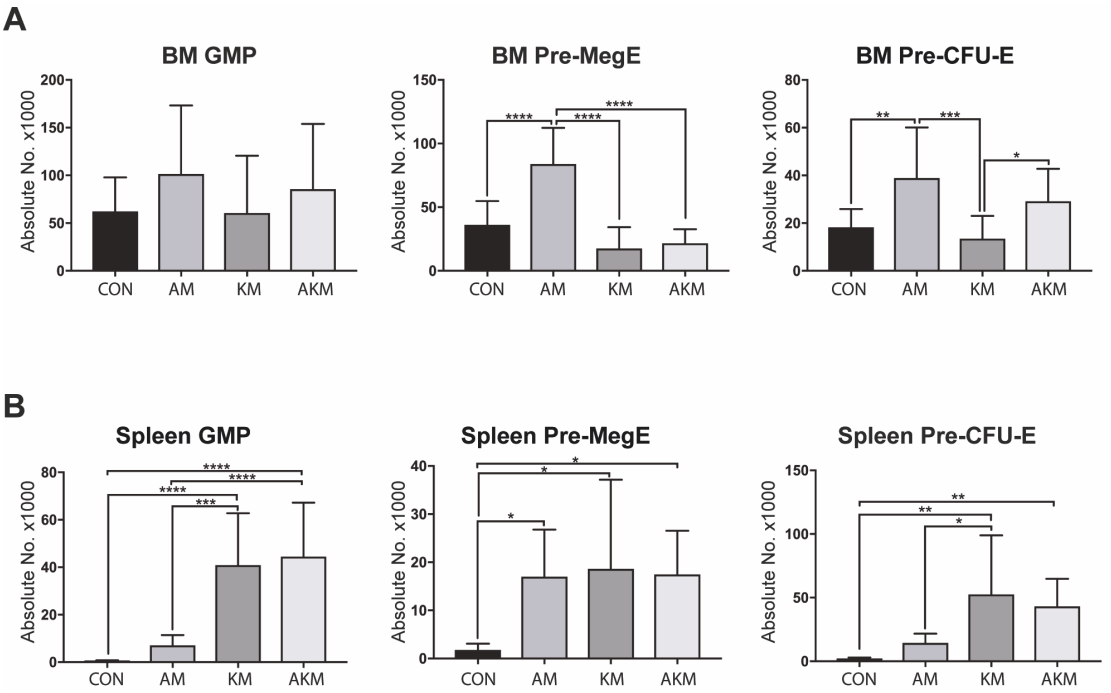

Figure S2. Myelo-erythroid progenitors in the BM and spleen. (A) Absolute numbers of CD45.2 GMP, Pre-MegE and Pre-CFU-E in the BM from recipients of CON (n=12), AM (n=14), KM (n=12) and AKM FL (n=14). Results were generated in three independent experiments; (B) Absolute numbers of CD45.2 GMP, Pre-MegE and Pre-CFU-E in the spleen from recipients of CON (n=8), AM (n=10), KM (n=8) and AKM FL (n=10). Results were generated in two independent experiments. The results were analyzed using multiple comparison ANOVA. The results are presented as the mean ± SD. *P<0.05 **P<0.01 ***P<0.001 ****P<0.0001. GMP, granulocyte-monocyte progenitor; pre-MegE, pre-megakaryocyte-erythroid; pre-CFU-E, pre-colony forming unit-erythrocyte.

Figure S3

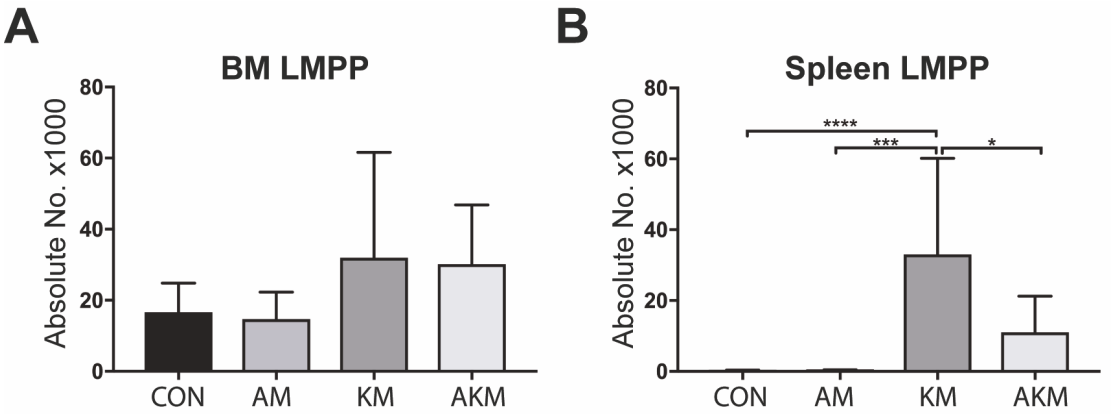

Figure S3. LMPP in the BM and spleen. (A) Absolute numbers of CD45.2 LMPP in the BM from recipients of CON (n=12), AM (n=14), KM (n=13) and AKM FL (n=14). Results were generated in three independent experiments; (B) Absolute numbers of LMPP in the spleen from recipients of CON (n=8), AM (n=10), KM (n=9) and AKM FL (n=10). Results were generated in two independent experiments. The results were analyzed using multiple comparison ANOVA. The results are presented as the mean ± SD. *P<0.05 **P<0.01 ***P<0.001 ****P<0.0001. LMPP, lymphoid-primed multipotent progenitors.

Figure S4

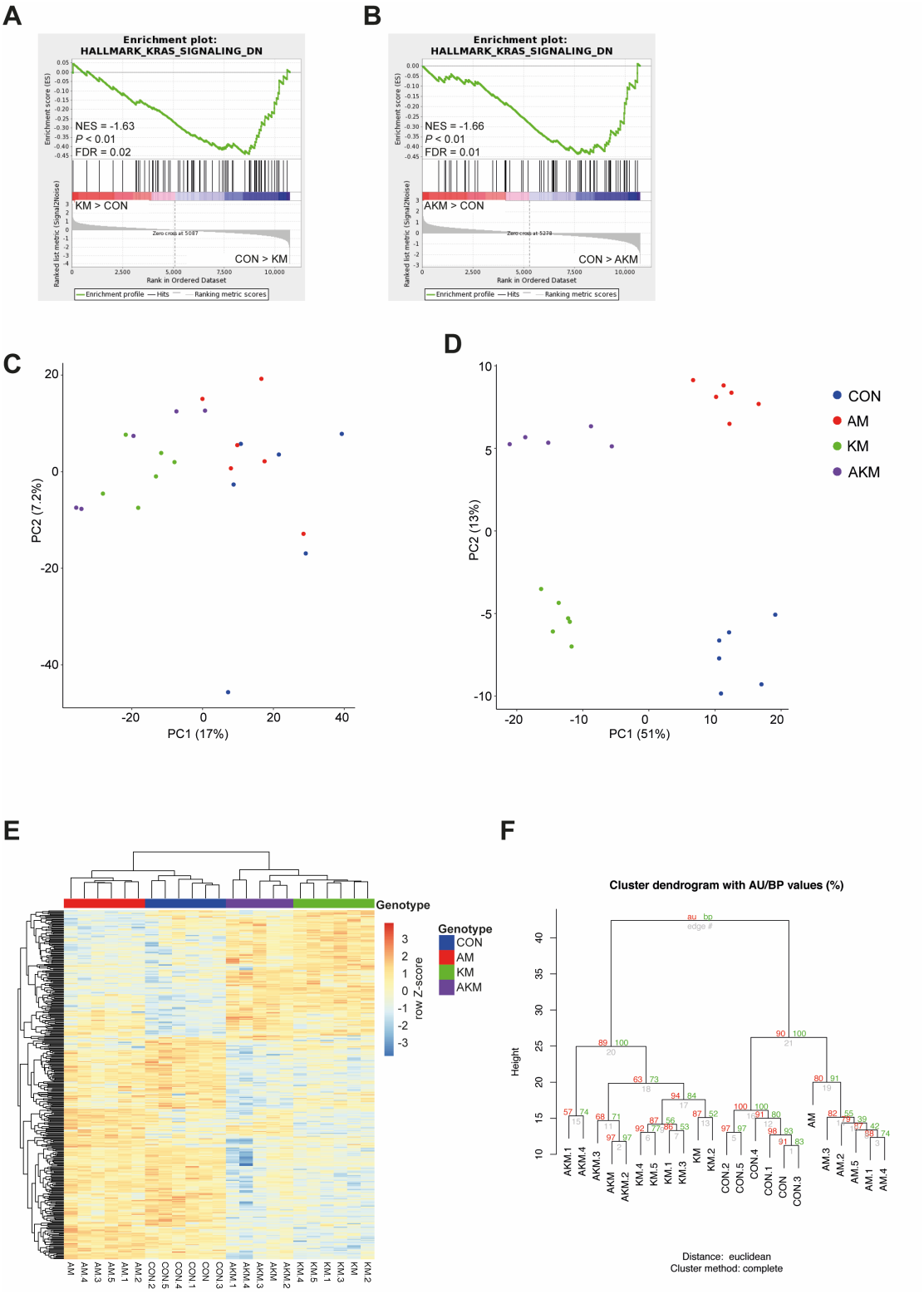

Figure S4. K-RasG12D drives transcriptional separation in AKM HSCs. (A, B) GSEA analysis of genes down-regulated by K-Ras activation in KM vs CON (A) and AKM vs CON (B); (C) Principal component analysis (PCA) illustrating the clustering of the different genotypes using 2,495 highly variable

genes; (D) PCA illustrating the clustering of the different genotypes using 379 genes that were differently expressed across the different genotypes; (E) Heat map illustrating the unsupervised hierarchical clustering of the different genotypes using 379 genes that were differently expressed across the different genotypes. Color scale represents log_2(FPKM) values centered and scaled in row direction; (F) Dendrogram illustrating hierarchical clustering of different genotypes. Values at branches are AU (approximately unbiased) p-values (left), BP (bootstrap probability) values (right), and cluster labels (bottom). Fragment per kilobase of transcript per million, FPKM.

References

1. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature methods. 2013;10(11):1096-1098.

2. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

3. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic acids research. 2013;41(10):e108.

4. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012;28(11):1530-1532.

5. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-3048.

6. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-140.

7. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-15550.

8. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics. 2003;34(3):267-273.

9. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. A stem cell molecular signature. Science. 2002;298(5593):601-604.

10. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739-1740.

11. Drissen R, Buza-Vidas N, Woll P, et al. Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. Nat Immunol. 2016;17(6):666-676.

12. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22(12):1540-1542.

13. Beissbarth T, Speed TP. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics. 2004;20(9):1464-1465.