Supplementary Information:

The EMT modulator SNAI1 contributes to AML pathogenesis via its interaction with LSD1.

Catherine L Carmichael[*1$], Jueqiong Wang[1], Thao Nguyen[1], Oluseyi Kolawole[1], Aissa Benyoucef[2,3], Charlotte De Mazière [1,4], Anna Milne[1], Sona Samuel[1], Kevin Gillinder[1], Soroor Hediyeh-zadeh[5], Anh Vo[1], Yizhou Huang[6], Kathy Knezevic[6], William R L McInnes[1], Benjamin Shields[1], Helen Mitchell[1], Matthew E Ritchie[5], Tim Lammens[7,8], Beatrice Lintermans[7,8], Pieter Van Vlierberghe[8,9], Nick Wong[1,10], Katharina Haigh[1,2,3], Julie A I Thoms[6], Emma Toulmin[1], David Curtis[1], Ethan P Oxley[1], Ross A Dickins[1], Dominik Beck[6], Andrew Perkins[1], Matthew McCormack[1], Melissa J Davis[5,11,12], Geert Berx[4,8], Johannes Zuber[13], John E Pimanda[6,14], Benjamin Kile[15], Steven Goossens[*1,4,8,9,16] and Jody J Haigh[*1,2,3$]

1. Australian Centre for Blood Diseases, Monash University, Melbourne, Australia
2. Department of Pharmacology and Therapeutics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada
3. Research Institute in Oncology and Hematology, CancerCare Manitoba, Winnipeg, Manitoba, Canada
4. Department for Biomedical Molecular Biology, Ghent University, Ghent, Belgium
5. Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia
6. Lowy Cancer Research Centre, School of Medical Sciences, Faculty of Medicine, UNSW, Sydney Australia
7. Department of Pediatric Hematology-Oncology and Stem Cell Transplantation, Ghent University Hospital, Ghent, Belgium
8. Cancer Research Institute Ghent (CRIG), Ghent University, Ghent, Belgium
9. Department of Biomolecular Medicine, Ghent University, Ghent, Belgium
10. Monash Bioinformatics Platform, Monash University, Melbourne, Australia
11. Department of Medical Biology, faculty of Medicine, Dentistry and Health Sciences, the University of Melbourne
12. Department of Biochemistry and Molecular Biology, Faculty of Medicine, Dentistry and Health Sciences, the University of Melbourne
13. Research Institute of Molecular Pathology (IMP), Vienna, Austria
14. Department of Haematology, Prince of Wales Hospital, Randwick, NSW, Australia
15. Department of Anatomy and Developmental Biology, Monash Biomedicine Discovery Institute, Monash University, Australia.
16. Department of Diagnostic Sciences, Ghent University, Ghent, Belgium

*These authors made equal intellectual contributions to this work

Address for correspondence:

Associate Professor Jody Jonathan Haigh
Department of Pharmacology and Therapeutics
Rady Faculty of Health Sciences
University of Manitoba
Senior Scientist
Research Institute in Oncology and Hematology
CancerCare Manitoba, ON5029
675 McDermot Ave
Winnipeg, MB
Canada, R3E 0V9
E-mail: jody.haigh@umanitoba.ca

Dr. Catherine Carmichael
Australian Centre for Blood Diseases
Monash University
99 Commercial Road, Prahran
Melbourne, Victoria
Australia, 3004
E-mail: catherine.carmichael@monash.edu

Contains:

SUPPLEMENTARY METHODS:

Western Blotting and Immunoprecipitation:

Cell lysis was performed using RIPA lysis buffer (150mM NaCl, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS and 25mM Tris pH 7.4) containing the complete protease inhibitor cocktail (Thermo Fisher Scientific). For human AML and $CD34^+$ samples, cells were first treated with the reversible proteasome inhibitor MG132 (10uM) for 24 hours prior to cell lysis in order to stabilise the SNAI1 protein and allow western blot visualisation.

For immunoprecipitations, LSD1-containing protein complexes were pulled down overnight at 4°C on a rotating wheel in the presence of antibody conjugated beads. Beads were then washed 5 times and eluted 2 times with SDS-lysis buffer (1% SDS, 50mM Tris0HCL ph 8.1, 10mM EDTA pH 8.0) for 15 minutes at 95 degrees. Western blotting was performed as above using rabbit polyclonal α-SNAI1 (C15D3, Cell Signalling Technology #3879), and secondary protein-A-HRP (Invitrogen #101023).

Mice and Animal Procedures:

For bone marrow transplantation studies, between 300,000 and 500,000 transduced wild type, *RosaCreERT2*$^{+/tg}$*;Snai1*$^{fl/fl}$ or *RosaCreERT2*$^{+/tg}$*;Snai1*$^{+/+}$ (C57BL/6) fetal liver cells were intravenously injected into lethally irradiated (2X550Rads) C57BL/6 (CD45.1) recipient mice. For competitive transplant studies, 500,000 CD45.1 competitor cells were transplanted alongside 500,000 CD45.2 test (*Snai1*$^{+/+,}$ *Snai1*$^{tg/+}$or *Snai1*$^{tg/tg}$) cells into lethally irradiated (2X550Rads) C57BL/6 (CD45.2) mice. For secondary MLL-AF9 or AML-ETO/NRAS transplantation studies, 500,000 bone marrow cells from moribund primary transplanted mice were intravenously injected into lethally irradiated C57BL/6 (CD45.1) recipients. To induce Cre-activity in transplanted bone marrow cells, Tamoxifen (Sigma) was given by oral gavage

once daily for 3 consecutive days (5mg tamoxifen in 50ul vehicle per 25g (10% ethanol, 90% peanut oil), and tumour latency monitored.

Chromatin Immunoprecipitation (ChIP)

Briefly, cells were harvested and washed with phosphate-buffered saline (PBS), followed by incubation of 1% (w/v) formaldehyde for 10 minutes at room temperature. To terminate the cross-link, cells were incubated with 0.125M glycine for 5 minutes. Cells were washed with PBS and lysed on ice in cell lysis buffer (10mM Tris [pH 8.0], 10mM NaCl, 0.2% NP-40) for 10 minutes to recover nuclei. After centrifugation at 1500g for 5 minutes, nuclei were lysed in nucleus lysis buffer (50mM Tris [pH 8.0], 10mM EDTA, 1% SDS) on ice for 10 minutes. The lysate was diluted in IP dilution buffer (20mM Tris [pH 8.0], 2mM EDTA, 150mM NaCl, 1% Triton-X100, 0.01% SDS) and sonicated (Settings: High, 30sec pulses, 15 cycles) using a BioRuptor® Pico sonicator (Diagenode, Liège, Belgium) to yield an average fragmentation size of approximately 200bp. The chromatin was pre-cleared with 100µg rabbit IgG per condition for 1 hour followed by incubation with 100µl protein-G agarose beads (Roche Applied Science, Penzberg, Germany) for 2 hours. Precleared chromatin was incubated with antibody (5µg/condition) for 18 hours at 4°C. To collect immune complexes, 50µl protein-G agarose beads was added to the chromatin and incubated for an additional 2 hours at 4°C. Unbound chromatin was kept as input for the subsequent sonication band check. Protein-G agarose pellets were washed at 5000g; twice with 500µl IP wash buffer 1 (20mM Tris [pH 8.0], 2mM EDTA, 50mM NaCl, 1% Triton-X100, 0.1% SDS), once with IP wash buffer 2 (10mM Tris [pH 8.0], 1mM EDTA, 0.25M LiCl, 1% NP-40, 1% sodium deoxycholate) and twice with TE (10mM Tris [pH8.0], 1mM EDTA). Immunoprecipitated chromatin was eluted in 300µl Elution Buffer (100mM NaHCO3, 1% SDS) and reverse cross-link was obtained by incubation with RNase A (1µg/condition) and NaCl (0.3M final concentration) at 67°C for 18 hours followed by treatment with Proteinase K at 45°C for 2 hours. Input DNA was treated with RNase A and

Proteinase K simultaneously. DNA was extracted twice using phenol-chloroform followed by ethanol precipitation. Purified DNA was re-suspended in 20µl nuclease-free water, and sent on dry-ice for sequencing.

ChIP-qPCR

For LSD1 ChIP-PCR analysis, chromatin-immunoprecipitated DNA was quantified by real-time qPCR using SYBR Green and a standard curve generated from normal genomic DNA. To calculate the enrichment of LSD1 protein to a particular target DNA, values obtained (using the standard curve method) for each target were divided by the amount of the corresponding target in the input fraction. Enrichments obtained from mock IPs performed in parallel using normal IgG were then subtracted from the enrichment values obtained using specific Abs. All the enrichments are expressed as a function of the enrichment obtained on the IgG from control MIG-EV cells. qPCR primers are shown in Supplementary Table T7.

RNA-sequencing analysis:

RNA-seq data analysis was carried out using R (version 3.3.0) (1) and Bioconductor (2). Reads were aligned to the *Mus musculus* genome (GRCm38) using the *Rsubread* package (version 1.22.2) (3) and assigned to genes by the featureCounts function (4) using the in-built annotation and the *org.Mm.eg.db* package (version 3.3.0) (5). Filtering and normalization used the *edgeR* package (version 3.15.0) (6). Genes with a count per million (CPM) of at least 1 in 3 or more samples were retained for further analysis. Compositional differences between libraries were normalized using the trimmed mean of *M*-values (TMM) method (7). Subsequent differential expression analysis was performed using the *limma* package (version 3.29.12) (8). Counts were transformed to log2-CPM values (with an offset of 0.5) with associated observational and sample-specific weights obtained from the voomWithQualityWeights method (9) assuming a linear model (10) with effects for genotype. Contrasts between the different genotypes were

estimated and differential expression was tested relative to an absolute fold-change of 1 with treat (11) and a false discovery rate (12) (FDR) cut-off of 0.05. FDR adjusted p-values were calculated using the Benjamini and Hochberg method (13). Mean-difference plots of the results were generated using the *Glimma* package (version 1.1.1) (14). GSEA was performed using the publicly available Broad Institute database and GSEA software (http://software.broadinstitute.org/gsea/index.jsp).

Geneset Enrichment Analysis:

Unbiased Gene Set Enrichment Analysis was performed using the Broad GSEA online analysis tool (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). Significantly differentially regulated genes in HPC7 cells expressing Snai1 (MIG-wtSnai1) compared to the Empty Vector (MIG-EV) control cells were chosen based on FDR<0.05. Significantly enriched gene sets were ranked based on FDR adjusted p-values, which were calculated using the Benjamini Hochberg method(13). Individual Gene Set Enrichment Analyses were performed using the GSEA v4.0.3 desktop application using default analysis settings. Gene Sets were selected from the Molecular Signature Database (MSigDB) and analysed against the complete HPC7 RNA-sequencing dataset. Genesets taken from GSE68348 were generated using the GEO2R function in the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/), and significantly upregulated genes were chosen as having an FDR>0.05.

To identify a gene set of genes upregulated in AML samples expressing high levels of SNAI1, we again utilized the GEO2R function within the Gene Expression Omnibus. Briefly, AML patient samples from GSE10358 were ranked based on *SNAI1* expression (n=304), and then allocated as either *SNAI1* high (top 20%, n=30), *SNAI1* low (bottom 20%, n=30) or *SNAI1* intermediate (middle 60%, n=244). Differential gene expression analysis was then performed between the SNAI1 high and SNAI1 low groups of patients using the GEO2R function, and significantly differentially expressed genes chosen as those with an FDR <0.05. This gene set

of differentially expressed genes in SNAI1 high AMLs was then compared to the complete *Snai1* expressing HPC7 dataset using the GSEA v4.0.3 desktop application.

ChIP-sequencing Data Analysis

Paired-End reads 35-37bp reads (Snai1 ChIP-sequencing) and single-end 49bp reads (H3K4me1/2 ChIP-sequencing) were assessed for sequencing quality and adapter contamination using fastqc (15). Reads were aligned to GRCm38/mm10 build of the Mus musculus genome using subread in DNA alignment model (3) and duplicate reads were marked using Picard Tools (Broad Institute Picard tools available online http://broadinstitute.github.io/picard). The BAM files were then sorted and indexed using SAMtools (16).

Snai1 ChIP-sequencing included 1 pull-down replicate and 1 input replicated per Empty Vector (EV), Snai1 mutant and WT Snai1 conditions. Peaks were called using MACS2 with the setting --nomodel (17). The called peaks were annotated to closest gene using ChIPpeakAnno (18). The H3K4me1/2 experiment included 3 biological replicates per experimental condition. The replicates were obtained in two separate batches. The first biological replicate was obtained in a different batch than the second and third replicates. One input replicate was available from each batch. For analysis of differential histone modification, the aligned H3K4me1 and H3K4me2 reads were quantified +/- 5kb around TSS for all genes using featureCounts (4). Histone counts were TMM normalized (7). Genes with Counts-Per-Million (CPM) more than 5 in at least 3 samples were retained in the analysis. The R/Bioconductor package limma (8) was used to perform a Surrogate Variable Analysis to estimate and correct for the unwanted variation when integrating data from different batches. Differential Histone Modifications (DHM) between WT and Empty Vector, and Mutant vs Empty Vector were identified following the voom-limma workflow using batch-corrected data(19).

ATAC-sequencing data analysis:

For each sequenced library, paired-end reads were aligned using Bowtie2(20) to the UCSC reference genome (mm10), filtered for duplicates (Picard Tools; **http://broadinstitute.github.io/picard/**) and ENCODE blacklisted regions(21) were removed resulting in a Uniquely Mapping Useable Fragment Depth of >41 M per library. Differential enrichment of promoter regions (-1kb to 100bp around the TSS) was performed using DiffBind(22).

REFERENCES:

1. Team RC. R: A Language and Environment for Statistical Computing (version 3.3.0). Vienna, Austria: R Foundation for Statistical Computing; 2016.
2. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS*, et al.* Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods **2015**;12(2):115-21 doi 10.1038/nmeth.3252.
3. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res **2013**;41(10):e108 doi 10.1093/nar/gkt214.
4. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **2014**;30(7):923-30 doi 10.1093/bioinformatics/btt656.
5. Carlson M. org.Mm.eg.db: Genome wide annotation for Mouse. R package version 3.7.0. 2016.
6. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **2010**;26(1):139-40 doi 10.1093/bioinformatics/btp616.
7. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol **2010**;11(3):R25 doi 10.1186/gb-2010-11-3-r25.
8. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res **2015**;43(7):e47 doi 10.1093/nar/gkv007.
9. Liu R, Holik AZ, Su S, Jansz N, Chen K, Leong HS*, et al.* Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. Nucleic Acids Res **2015**;43(15):e97 doi 10.1093/nar/gkv412.
10. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **2004**;3:Article3 doi 10.2202/1544-6115.1027.
11. McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics **2009**;25(6):765-71 doi 10.1093/bioinformatics/btp053.
12. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) **1995**;57(1):289-300.

13. Benjamini YH, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B **1995**;57(1):289–300.
14. Su S, Law CW, Ah-Cann C, Asselin-Labat ML, Blewitt ME, Ritchie ME. Glimma: interactive graphics for gene expression analysis. Bioinformatics **2017**;33(13):2050-2 doi 10.1093/bioinformatics/btx094.
15. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc
16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* The Sequence Alignment/Map format and SAMtools. Bioinformatics **2009**;25(16):2078-9 doi 10.1093/bioinformatics/btp352.
17. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE*, et al.* Model-based analysis of ChIP-Seq (MACS). Genome Biol **2008**;9(9):R137 doi 10.1186/gb-2008-9-9-r137.
18. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS*, et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics **2010**;11:237 doi 10.1186/1471-2105-11-237.
19. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol **2014**;15(2):R29 doi 10.1186/gb-2014-15-2-r29.
20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods **2012**;9(4):357-9 doi 10.1038/nmeth.1923.
21. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep **2019**;9(1):9354 doi 10.1038/s41598-019-45839-z.
22. Stark R. BGD. DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data. Bioconductor. **2011**.

SUPPLEMENTARY FIGURE LEGENDS:

Supplementary Figure S1:

A) Log2 expression plot showing that *SNAI1* is expressed significantly higher in human AML patients compared to normal hematopoietic stem and progenitor cells. Source (http://servers.binf.ku.dk/bloodspot/). **B)** Kaplan-Meier graphs showing worse overall survival of cytogenetically normal AML patients (i) or AML patients with normal/intermediate risk cytogenetics (ii) that have higher levels of *SNAI1*. Graphs modified from http://genomics.jefferson.edu/proggene/index.php (i) and https://xenabrowser.net/ (ii), with survival and expression data taken from the GSE12417 dataset: Prognostic Gene Signature for Normal Karyotype AML (i) or the GDC TCGA Acute Myeloid Leukemia LAML combined datasets (ii). **C**) qRT-PCR analysis of shRNA knockdown of *SNAI1* in MOLM13-R cells using 3 independent shRNAs (shSNAI1.770, shSNAI1.1577 and shSNAI1.1663) results in 60-90% loss of mRNA compared to the non-silencing shRen.713 shRNA. **D)** Western blot analysis also showing knockdown of SNAI1 protein in MOLM13-R cells by shSNAI1.770 and shSNAI1.1577, but not the shRen.713 control shRNA. **E)** Representative cytometry histogram plots showing shRNA mediated SNAI1 knockdown in HL60-R, NB4-R and Kasumi-R cell lines does not alter the cell surface expression of CD11b. **F)** Quantification of the percentage of $CD11b^+$ cells in shRen.713 infected cells (black bars) compared to shSNAI1 infected cells (red and blue bars).

Supplementary Figure S2:

A) Schematic showing generation of the Rosa26Snai1 transgenic mouse model. **B)** Flow cytometric analysis of a $Snai1^{tg/tg}$ mouse with MPN, showing aberrant and increased population of $CD11b^+GR1^+$ cells. **C)** Flow cytometric analysis of a $Snai1^{tg/tg}$ mouse with AML (erythroid), showing aberrant and increased population of $CD71^+TER119^+$ erythroblast cells in the bone

marrow and spleen. D) Histological bone marrow and/or spleen sections from the same mice shown in C); i) *Snai1*$^{tg/tg}$ mouse with MPN, showing increased erythropoiesis and increased myeloid cells, ii) *Snai1*$^{tg/tg}$ mouse with erythroid AML showing increased immature blast cells and loss of splenic architecture. NB: a bone marrow section was not available for this mouse. E) Histological bone marrow, spleen and liver sections from 3 NSG mice transplanted with MPN or AML cells from 3 independent primary *Snai1* transgenic mice. Cell infiltration into liver was only seen for mice transplanted with AML cells.

Supplementary Figure S3

A) Hemavet® blood cell analysis showing peripheral blood cell counts in *Snai1* transgenic mice do not differ to wild type littermate controls at 8 months of age. Data shown is mean + SEM of N=9 (*Snai1*$^{+/+}$), N=8 (*Snai1*$^{tg/+}$) and N=8 (*Snai1*$^{tg/tg}$) biological replicates. **B)** Representative hematoxylin & eosin staining of histology sections of bone marrow and spleen from a *Snai1*$^{+/+}$ and a *Snai1*$^{tg/tg}$ mouse showing normal cellularity and organ architecture. **C)** Total cellularity of bone marrow and spleen were normal in *Snai1* transgenic mice. Data shown in **C)** is the mean + SEM of N=13 (*Snai1*$^{+/+}$), N=8 (*Snai1*$^{tg/+}$) and N=7 (*Snai1*$^{tg/tg}$) biological replicates.

Supplementary Figure S4:

A) Competitive transplant assay design schematic. **B)** A significant increase in CD45.2 contribution to myeloid cells within the bone marrow from *Snai1*$^{tg/+}$ cells was observed at 12 weeks but not at 20 weeks. No difference in contribution was observed from *Snai1*$^{tg/tg}$ cells. **C)** *Snai1*$^{tg/+}$ and *Snai1*$^{tg/tg}$ cells generated significantly more myeloid cells within the spleen at 12 weeks. This was still apparent at 20 weeks post-transplant for the *Snai1*$^{tg/tg}$ cells, but not the *Snai1*$^{tg/tg}$ cells. **D)** Contribution of donor CD45.2 cells to the stem and progenitor compartment within the bone marrow at 12 weeks post-transplant was not significantly different between the

three genotypes. However, *Snai1*$^{tg/+}$ cells did contribute significantly more GMP cells compared to *Snai1*$^{+/+}$ and *Snai1* $^{tg/tg}$ cells. **E)** No difference in contribution to stem and progenitor cells was observed at 20 weeks post-transplant. Data shown is mean + SEM of N=3 (Snai1$^{+/+}$), N=3 (Snai1$^{tg/+}$) and N=3 (Snai1$^{tg/tg}$) biological replicates. * $p<0.05$ student's unpaired t-test. **F)** May-Grunwald-Giemsa staining of cytocentrifuge preparations of cells washed out of methylcellulose after the second round of replating revealed predominantly mast cells in the *Snai1*$^{+/+}$ cultures (due to the IL3 cytokine in the methylcellulose) (example: green arrows), whereas in the *Snai1*$^{tg/+}$ cultures there remained a significant proportion of immature myeloid cells (example: red arrows).

Supplementary Figure S5:

A) Mean-Difference plot showing differential expression in MIG-Snai1 cells compared to MIG-EV cells (significantly upregulated genes in red and downregulated genes in blue, False Discovery Rate < 0.05). **B)** Mean-Difference plot showing no differentially expressed genes in MIG-mutSnai1 cells compared to MIG-EV cells, except for ectopic mutSnai1 itself. **C**) Differentially expressed genes fall into several key biological pathways related to malignancy, including cytokine signalling, migration/adhesion/invasion, transcriptional regulation and tumour suppressors (upregulated genes in red and downregulated genes in blue).

Supplementary Figure S6:

A) HPC7 cells expressing WT-Snai1 (blue histograms) show increased levels of i) CD11b, ii) Ly6C and ii) Ly6G myeloid markers compared to Empty Vector (black) or Mut5Snai1 (purple) expressing HPC7 cells. **B)** H3K4me1/2 methylation marks were found to be predominantly localised around the transcriptional start site (TSS) of genes. **C**) FC-FC plot of differential histone

methylation (Y-axis) for the H3K4me2 mark versus the differential gene expression (X-axis) for all annotated genes within the genome is shown. This analysis shows a positive correlation between the level of DHM and the amount of differential expression for this histone mark. Genes with an increased methylation also tend to have an increased gene expression (top right quadrant), and genes with decreased methylation also tend to have a decreased gene expression (bottom left quadrant). Genes with both significant DHM and significant gene expression fold change are highlighted. **D)** Motif enrichment analysis identified an E-box motif within 92% of Snai1 binding sites (DREME analysis, MEME-suite). **E)** SNAI1 peak distribution in HPC7 cells expressing MIG-Snai1. **F**) Venn diagram showing overlap between significantly upregulated and significantly downregulated genes and Snai1 binding sites identified by ChIP-sequencing. **G)** FC-FC plot of differential ATAC-seq enrichment (Y-axis) versus differential gene expression (X-axis). This analysis shows a correlation between reduced ATAC-seq enrichment and reduced gene expression. Snai1 bound genes are indicated in green, and again shows correlation between SNAI1 binding, reduced ATAC-seq enrichment and reduced gene expression.

SUPPLEMENTARY TABLES:

Supplementary Table T1: AML patient sample information.

Supplementary Table T2: Differentially expressed genes (FD<0.05) in HPC7 cells expressing WT-Snai1 compared to cells expressing Empty Vector.

Supplementary Table T3: Supplementary Table T3: GSEA analysis of the most significantly enriched gene sets from the GSEA Molecular Signatures Database “Hallmarks” gene set list in HPC7 cells expressing Wt-Snai1.

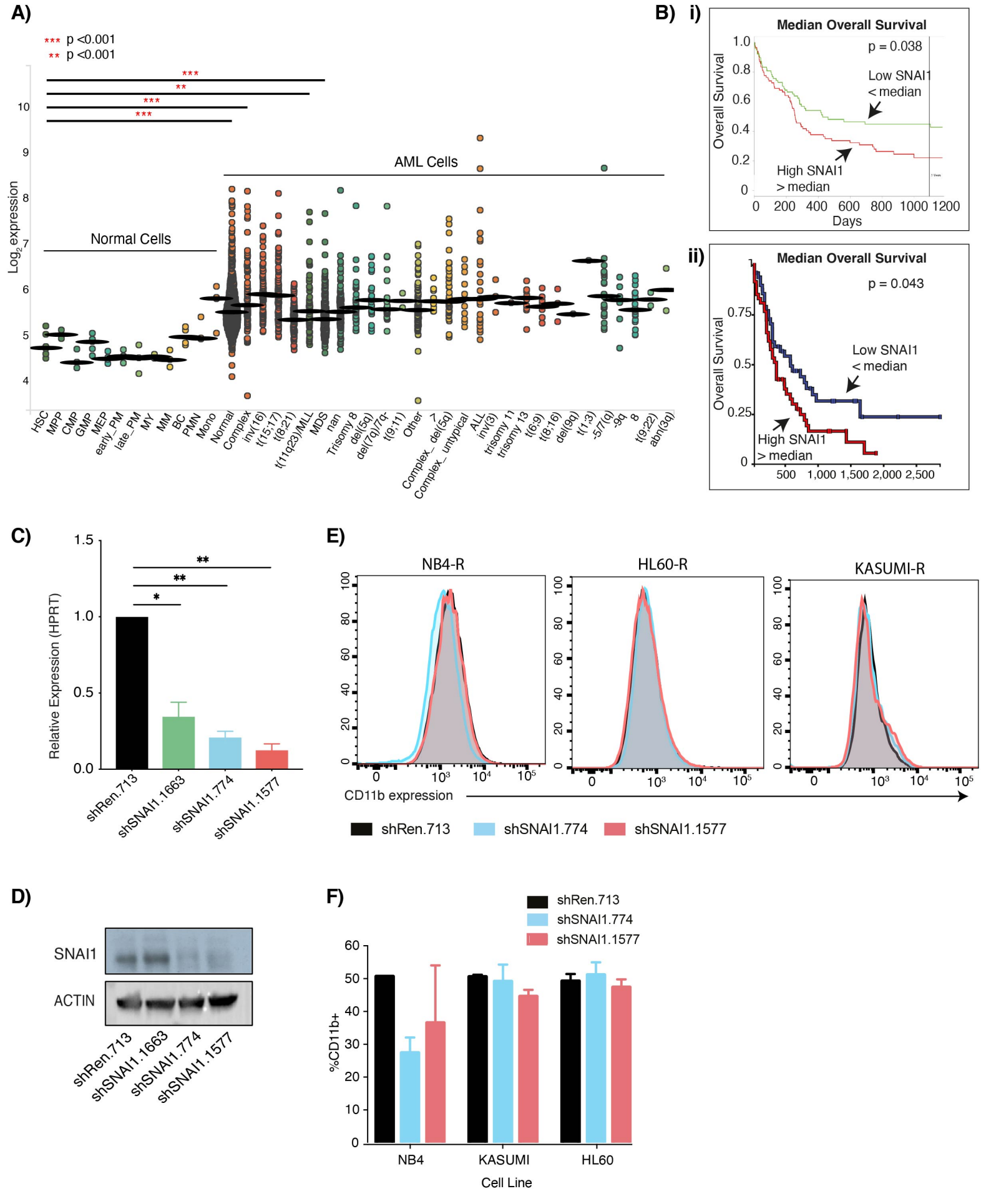
Supplementary Table T4: Differentially expressed genes showing associated Snai1 binding site and/or significant change in methylation at H3K4me1 or H3K4me2 (FDR <0.05).

Supplementary Table T5: ChIP-sequencing peaks identified in WT-Snai1, Mut-Snai1 and EV HPC7 cells.

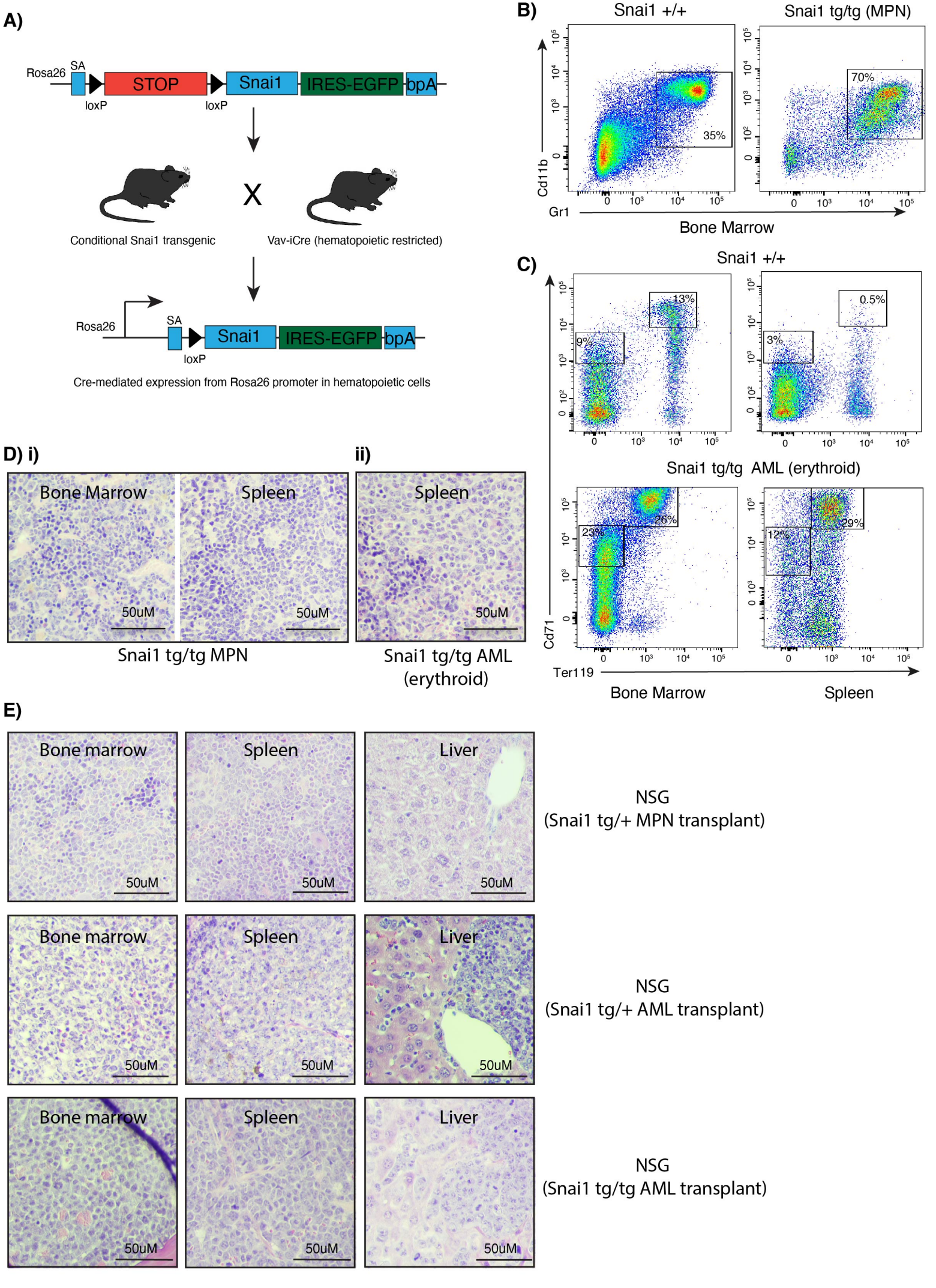
Supplementary Table T6: Antibodies used in flow cytometry and western blotting.

Supplementary Table T7: Primer sequences used for qPCR analyses.

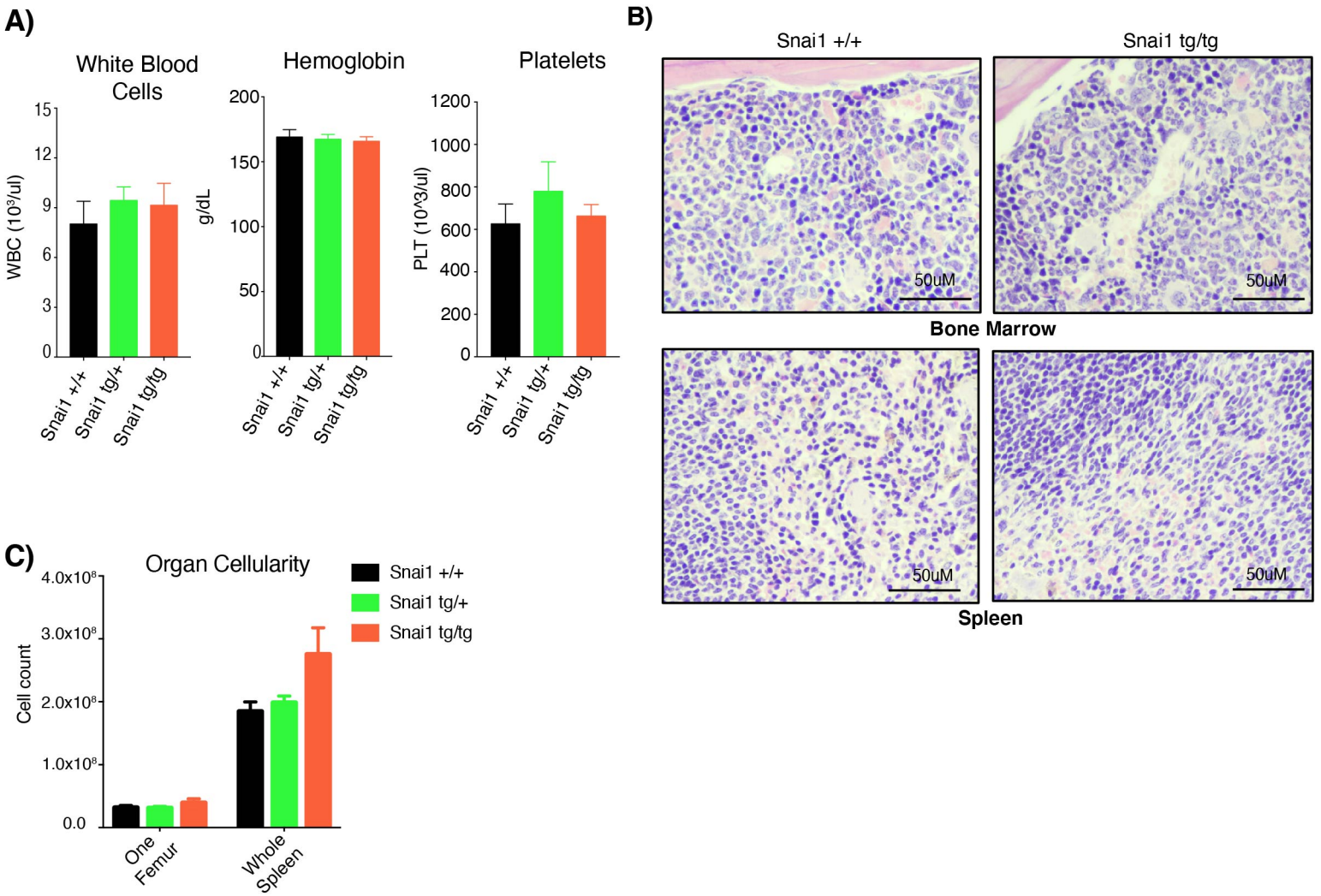
Supplementary Figure S1



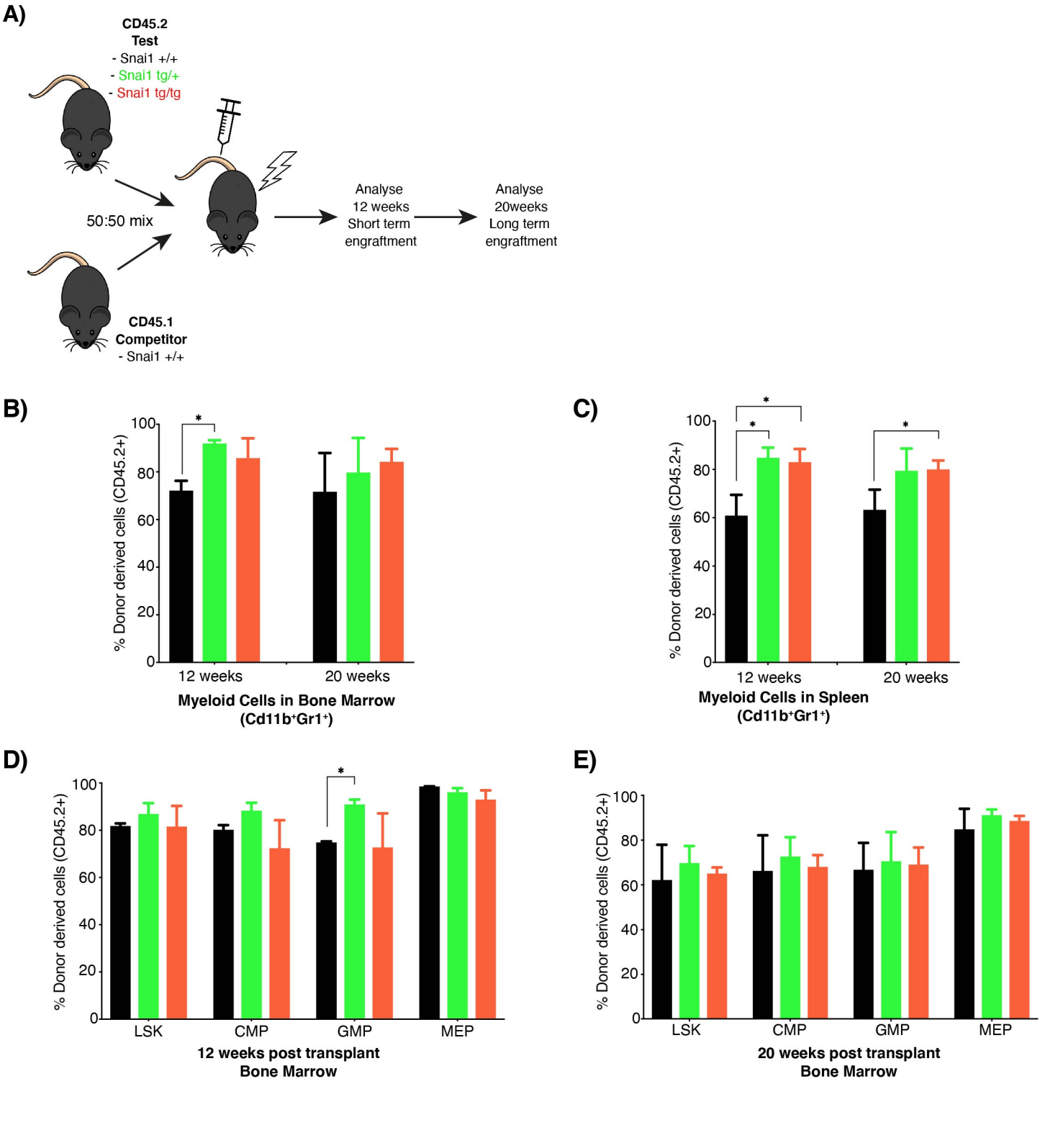
Supplementary Figure S2



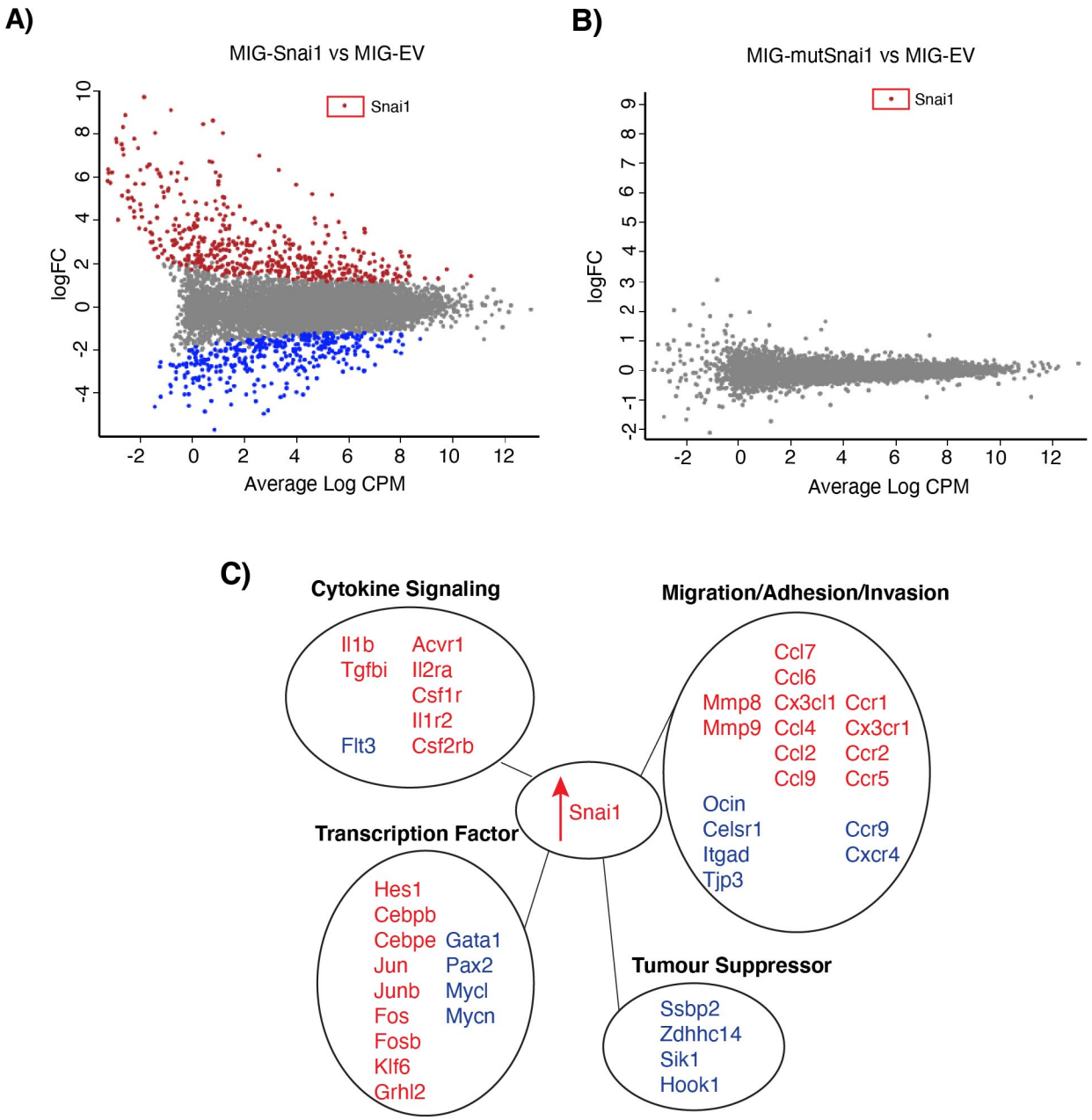
Supplementary Figure S3



Supplementary Figure S4



Supplementary Figure S5



Supplementary Figure S6