Supplemental Methods

Flow cytometry and cell selection

Progenitor populations were defined as follows;

HSCs ($CD45^+$, $EPCR^+$, $CD48^-$, $CD150^+$)

Short term-long term-progenitor-HSCs ($ST/LT^{prog/HSC}$) (Lin^-, c-kit^+, Sca-1^+, $Flt3^-$)

Multipotent progenitors (MPP) (Lin^-, c-kit^+, Sca-1^+, $Flt3^+$)

Lymphoid primed multipotent progenitors (LMPP) (Lin^-, c-kit^+, Sca-1^+, $Flt3^{hi}$)

Common lymphoid progenitors (CLP) (Lin^-, $Flt3^{hi}$, Il-7rα^+, c-kit^{int}, Sca-1^{int})

Granulocyte-monocyte progenitors (GMP) (Lin^-, Il-7rα^-, c-kit^+, Sca-1^-, $CD34^+$, $CD16/32^+$)

Common myeloid progenitors (CMP) (Lin^-, Il-7rα^-, c-kit^+, Sca-1^-, $CD34^+$, $CD16/32^-$)

Megakaryocyte-erythroid progenitors (MEP) (Lin^-, Il-7rα^-, c-kit^+, Sca-1^-, $CD34^-$, $CD16/32^-$)

Antibodies were obtained from eBiosciences or BioLegend unless stated. Markers for lineage depletion (Lin^-) included B220 (RA3-6B2), CD3ε (145-2C11), Mac-1 (M1/70), Gr-1 (RB6-8C5) and Ter119 (TER-119). E-PCR (RMEPCR1560, Stem Cell Technologies), CD45 (30-F11), CD150 (TC15-12F12.2), CD48 (HM48-1), CD34 (RAM34), CD16/32 (2.4G2), FLT3 (A2F10), c-kit (2B8), Sca-1 (E13-161.7) and Il7-rα (A7R34).

For MPP gene expression, sorting was defined as: Lin^-, c-kit^+, Sca-1^+, $CD34^+$, $CD48^+$ and $CD150^-$. Samples were flow-sorted after removal of lineage positive cells using a magnetic activated cell sorting (MACS) mouse lineage depletion kit containing antibodies for CD5, B220, CD11b, Gr-1 (Ly-6G/C), 7-4, and Ter-119 following the manufacturers' instructions (Miltenyi Biotec).

Analysis of activated ERK1/2

Selected samples were studied for total ERK1/2 (p44/42 MAPK, clone 137F5, Cell Signalling) and pERK1/2 (phosphor-p44/42 MAPK, clone 197G2, Cell Signalling).

Gene expression profiling, microarrays

Gene expression profiles of lineage negative (lin^-) or 1,000 sorted multipotent progenitors (MPPs) from mutant and wildtype controls were compared using the Illumina MouseWG-6 v2 Expression BeadChip platform (Illumina). Lin^- populations were separated from whole bone marrow using magnetic activated cell sorting (MACS, Miltenyi Biotec) and RNA isolated using a standard Trizol (Thermofisher) protocol. Flow sorted MPP populations were sorted directly into Trizol LS (Thermofisher) using a Mo-Flow^{TM} XDP (Beckman Coulter) and RNA extracted according to the manufacturer. Extracted RNA was prepared for array hybridization using the

TargetAmp™-Nano Labeling Kit (Epicentre). Global profiling was done using Illumina MouseWG-6 v2.0 Expression BeadChip. Data were quantile normalized and analyzed using the Bioconductor, *lumi* and *limma* packages with *P* values adjusted for multiple testing (Bioconductor, http://www.bioconductor.org/; lumi, http://www.bioconductor.org/packages/2.0/bioc/html/lumi.html; RTCGD, http://rtcgd.ncifcrf.gov/).[1,2]

Adjusted *P* value (<0.05) was used to identify significantly differentially expressed genes. Gene set enrichment analysis was carried out using GSEA v2.1.0 (Broad Institute). [3,4]

Comparative gene expression analysis of NPM1+ve and NPM1-ve AML samples from the Cancer Genome Atlas (TCGA) , generated using the Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix) was performed using the GCRMA (http://www.bioconductor.org/repository/devel/vignette/gcrma.pdf), *limma* and *affy* packages in Bioconductor .[2,5,6] (.CEL file sample IDs are listed in Supplemental Table 1.1, datasets and samples are listed in Supplemental Table 2, note that samples positive for *MLL* gene fusions were removed from the control sample set.)

Copy number variation, comparative genomic hybridisation (aCGH)

DNA copy number variation in leukemic samples was assessed using the Mouse Genome Comparative Genomic Hybridization 244K Microarray (Agilent Technologies). DNA was labeled with Cy3 or Cy5 according to Agilent aCGH genomic labeling protocol (Agilent Technologies). Raw data was extracted using Agilent Feature Extraction and normalised using R Package aCGH Spline. Subsequent data analysis was performed in R using aCGH Bioconductor packages (http://www.bioconductor.org).[6]

Mouse AML mutation calling and validation.

Sequence reads were aligned against the reference mouse genome (GRCm38) using the Burrows-Wheeler algorithm (BWA; specifically, *aln* for HiSeq--paired-end exome sequencing data and *mem* for MiSeq-250bp-paired-end sequencing data). For the detection of the *Flt3*^{ITD} and *Npm1*$^{flox\text{-}cA}$ or *Npm1*^{cA} alleles, a fasta entry containing these sequences was appended to the reference genome. Sam/bam files were sorted and indexed using SAMTOOLS.[7] Where necessary we also performed PCR duplicate marking using PICARD tools (http://picard.sourceforge.net) and local realignment around indels using GATK.[8] The in-house software RAMSES, was used to detect somatic mutations and indels identified using PINDEL.[9] Functional consequences of mutations were predicted using an in-house script employing Ensembl Perl API.[10] All potential transcript annotations were calculated and recorded. The most deleterious of all potential annotations was reported for each mutation.

Further details of the exome sequencing and amplicon specific validation workflow are outlined in Supplemental Methods Figure S1.

Retroviral transduction

Cloning of mouse *Nkx2-3 and Hoxa9* into MSCV-GFP/CFP retroviral backbones.

mRNA extracted from homozygous wildtype C57BL/6N mouse bone marrow cells was reverse transcribed using SuperScript III (Invitrogen) and the subsequent cDNA was used as template to amplify full length *Nkx2-3* or *Hoxa9* cDNA using high fidelity *taq polymerase* (KAPA HiFi HotStart ReadyMix, Kapa Biosystems) using the manufacturer's instructions and the following primers;

EcoRI-*mNkx2-3*-XhoI Fwd:

gaattcgccaccatgatgttaccaagcccggtcacctccacccctttctc

EcoRI-*mNkx2-3*-XhoI Rev:

tcgagtcacttgtcgtcatcgtctttgtagtcaatgtcatgatccttgtaatcgccgtcgtgccaagccctgatgccctgcaaagtcccctgcgtgcacg

This fragment was cloned into an EcoRI/XhoI linearized fragment obtained from the MSCV-IRES-GFP (Addgene plasmid # 20672) retroviral backbone using standard molecular biology techniques.

MluI-*Hoxa9*-XhoI Fwd:

aattcacgcgtatggccaccaccggggccctgggcaactactatgtggac

MluI- *Hoxa9*-XhoI Rev:

ctcgagttaagcgtaatctggaacatcgtatgggtagccgtcgtgctcgtcttttgctcggtccttgttgattttcttcattttcatcctgcggttctgg

This fragment was cloned into a MluI/XhoI linearized fragment of MSCV-IRES-CFP, a kind gift from Dr Brian Huntley.

Confirmation of CAS9 activity in *Rosa26-EF1-Cas9* AML cell lines and guide RNA (gRNA) cloning.

Activity of CAS9 in mouse AML cell lines was validated by transducing *Rosa26*$^{Cas9/+}$expressing cells with a reporter construct expressing BFP/GFP and a guide RNA targeted to GFP. In the presence of active CAS9 double positive BFP/GFP cells are reduced to single positive expressing BFP cells through CRISPR-CAS9 mediated disruption of GFP as previously described.[11]

Previously validated gRNA sequences (obtained from a previously published mouse library)[11] were annealed and cloned into the pKLV-U6gRNA(BbsI)-PGKpuro2ABFP construct as described by Koike-Yusa et al. [12] Competitive assays were performed over a 21 day period post gRNA transduction by measuring the relative number of transduced (BFP+) and non-transduced (BFP-) cells using FACS. gRNA sequences are presented in Supplemental Table S15.

RNAseq analysis

For each of the 2 cell lines, *Rosa26-EF1-Cas9*;*Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$; and *Rosa26-EF1-Cas9*;*Npm1*^{cA};*Flt3*^{ITD} mice and wildtype lineage negative hematopoietic progenitors, two RNA-seq data sets were generated and considered as biological replicates. 75bp paired-end Illumina reads were aligned to the GRCm38 mouse genome using Tophat2. The aligned reads were then filtered to retain only those that aligned as a *bone fide* read pair with an alignment quality score better than 10, using SAMtools [ref: https://www.ncbi.nlm.nih.gov/pubmed/19505943]. The total number of reads that align to the exons of each gene in the mouse genome as defined by Ensembl Genes 84 [ref: https://academic.oup.com/database/article/doi/10.1093/database/baw093/2630475/The-Ensembl-gene-annotation-system] were computed using HTSeq [ref: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu638]. These gene counts were then used to identify differentially regulated genes using DESeq2. The FPKM [ref. http://www.nature.com/nprot/journal/v7/n3/full/nprot.2012.016.html] values for each gene per condition were computed using DESeq2. A pseudo-count of 0.001 was added to the FPKM values, transformed to base 2 logarithm and further quartile normalised in order to obtain the gene expression values.

The correlation between the datasets were obtained as Pearson correlation coefficients and these correlations were then hierarchically clustered using Euclidean distances. The clustered heat map of the biological replicates is shown in Supplemental Figure S9A. Select gene expression values are also depicted.

Supplemental Table Legends

Supplemental Table S1. Comparatively altered gene expression in lineage negative bone marrow aspirates compared to wildtype (significantly differentially expressed genes, adj. p<0.05).

Supplemental Table S2. TCGA datasets[5] used for comparative gene expression analysis of human AML.

Supplemental Table S3. Comparatively altered gene expression in human AML based on NPM1 mutation status (for significantly differentially expressed genes adj.p<0.05).

Supplemental Table S4a. Comparative gene (probes) expression in multipotent progenitors, *Nras*$^{G12D/+}$ and *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ compared to wildtype.

Supplemental Table S4b. Comparative gene (probes) expression in multipotent progenitors, *Flt3ITD*$^{/+}$ and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$compared to wildtype.

Supplemental Table S5a. Kegg Pathways enriched in *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ multipotent progenitor cells (DAVID[12]).

Supplemental Table S5a. Functional annotational clustering of gene ontology terms (GO-term) enriched in *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ multipotent progenitor cells (DAVID[13]).

Supplemental Table S6. Overlap of differentially expressed genes in *Tet2*$^{-/-}$;*Flt3*$^{ITD/+}$ LSK cells and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ MPP cells compared to wildtype.

Supplemental Table S7a. 1st Round MiSeq amplicon specific primer sequences.

Supplemental Table S7b. 2nd Round MiSeq iPCR-tag primer sequences.

Supplemental Table S8. Validation of exome sequencing: MiSeq amplicon sequencing results (mVAF).

Supplemental Table S9. Combined Single Nucleotide Variant (SNV) and insertions/deletions (Indels) detected by the Exome sequencing pipeline (detailed in Methods and outlined in Supplemental Methods Figure S1)

Supplemental Table S10. aCGH results summary.

Supplemental Table S11. Mouse-Human synteny of chromosome regions with altered copy number, identified by aCGH, in murine AMLs. (Only genes identified as mutated in the TCGA AML data-set or hits in our Npm1$^{cA/+}$ Sleeping Beauty insertional mutagenesis screen are included.)

Supplemental Table S12. Comparative RNAseq in *Rosa26*$^{Cas9/+}$*;Npm1*$^{cA/+}$*;Flt3*$^{ITD/+}$ compared to wildtype lineage negative cells (n=2, logFC>2, padj <0.01).

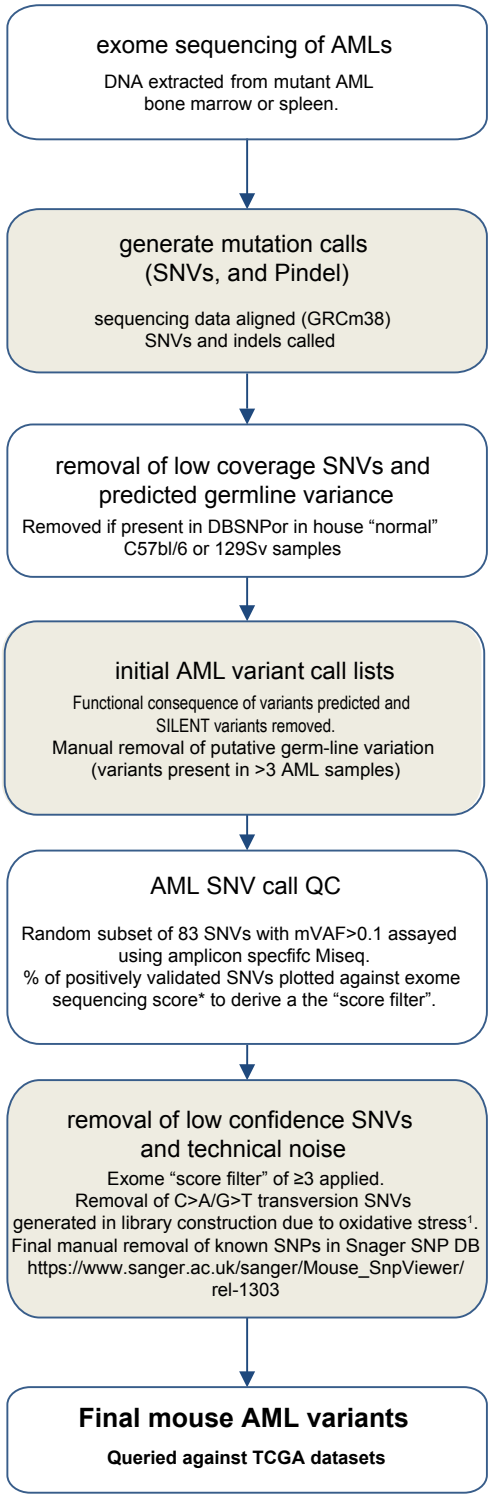
Supplemental Table S13. Comparative RNAseq in *Rosa26*$^{Cas9/+}$*;Npm1*$^{cA/+}$*;Nras*$^{G12D/+}$ compared to wildtype lineage negative cells (n=2, logFC>2, padj <0.01).

Supplemental Table S14. Comparative RNAseq in *Rosa26*$^{Cas9/+}$*;Npm1*$^{cA/+}$*;Flt3*$^{ITD/+}$ compared to *Rosa26*$^{Cas9/+}$*;Npm1*$^{cA/+}$*;Nras*$^{G12D/+}$ cell lines (n=2, logFC>2, padj <0.01).

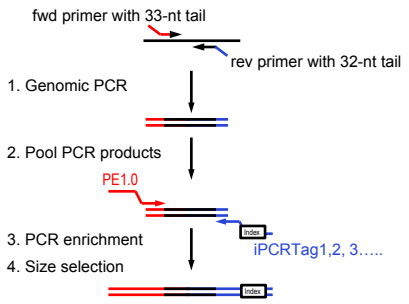
Supplemental Table S15. Guide RNA (gRNA) sequences used in CRISPR-CAS9 mediated loss of function studies.

Supplemental Methods Figure S1

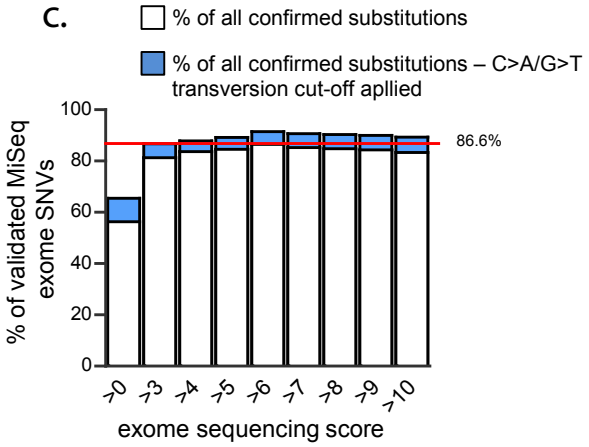
a.



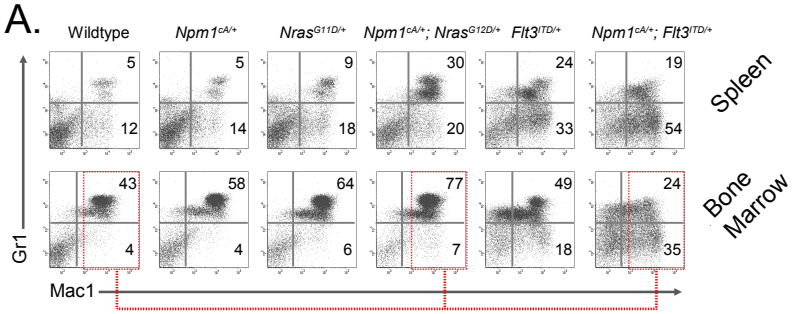
b.



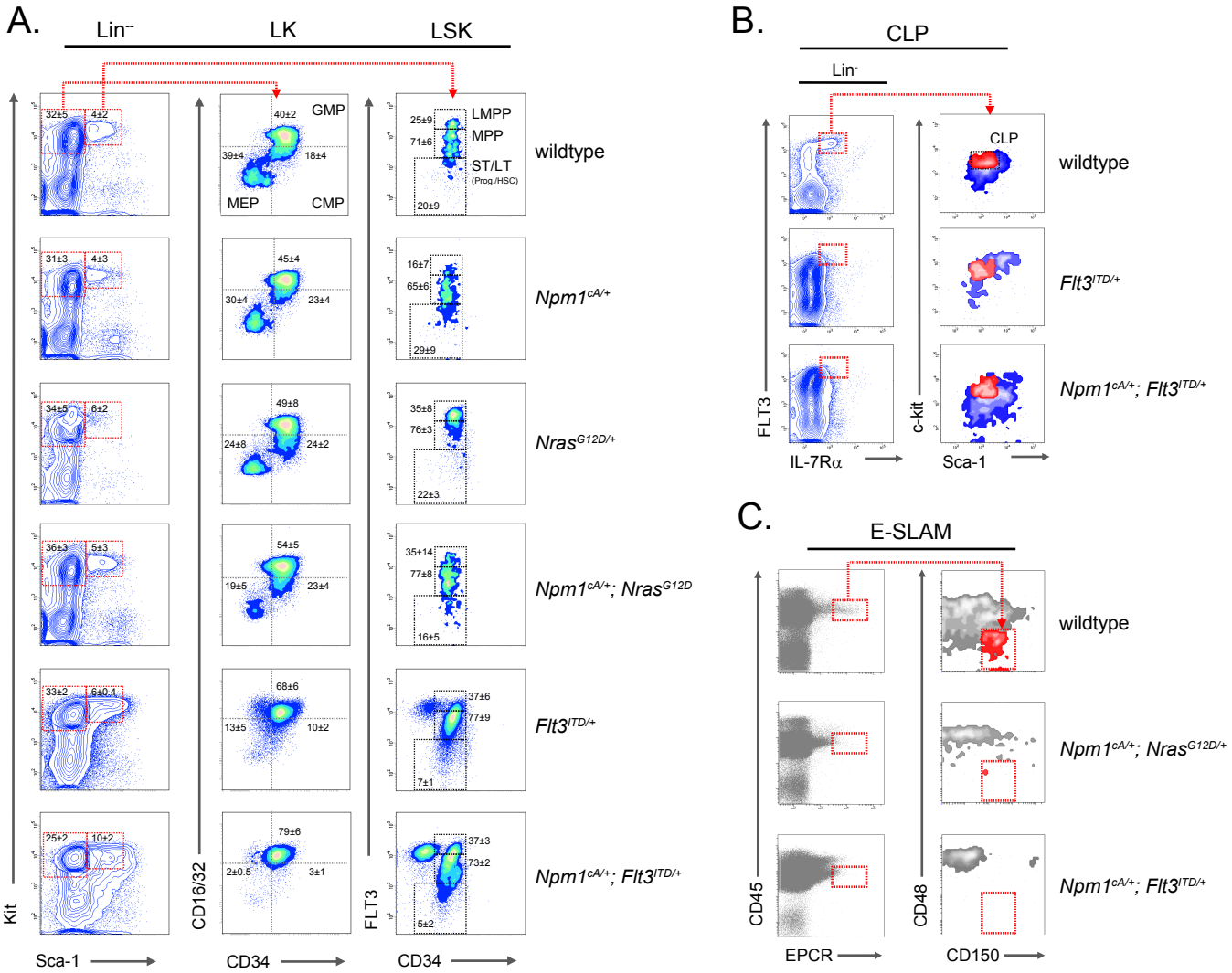
c.



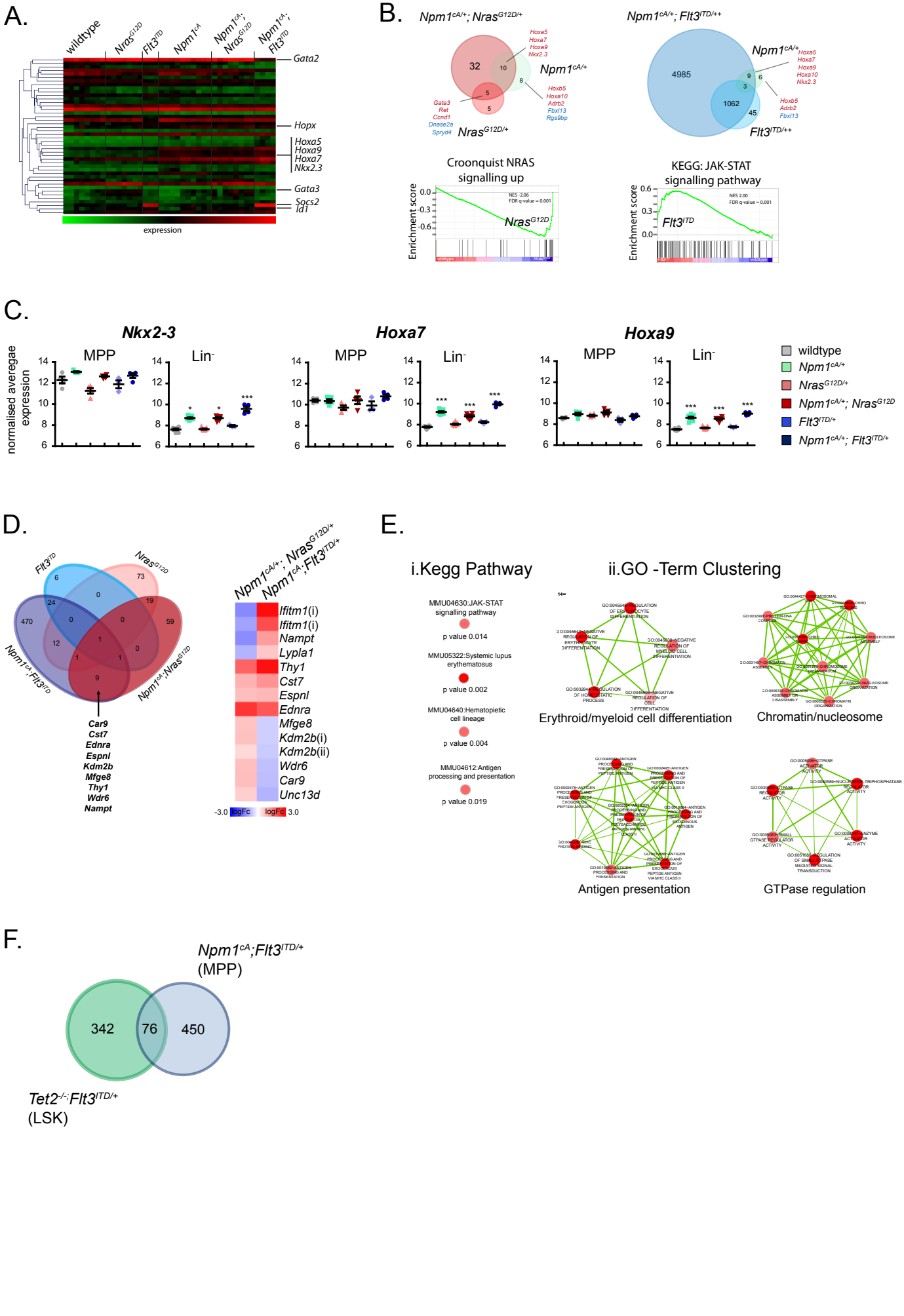
Supplemental Figure S1



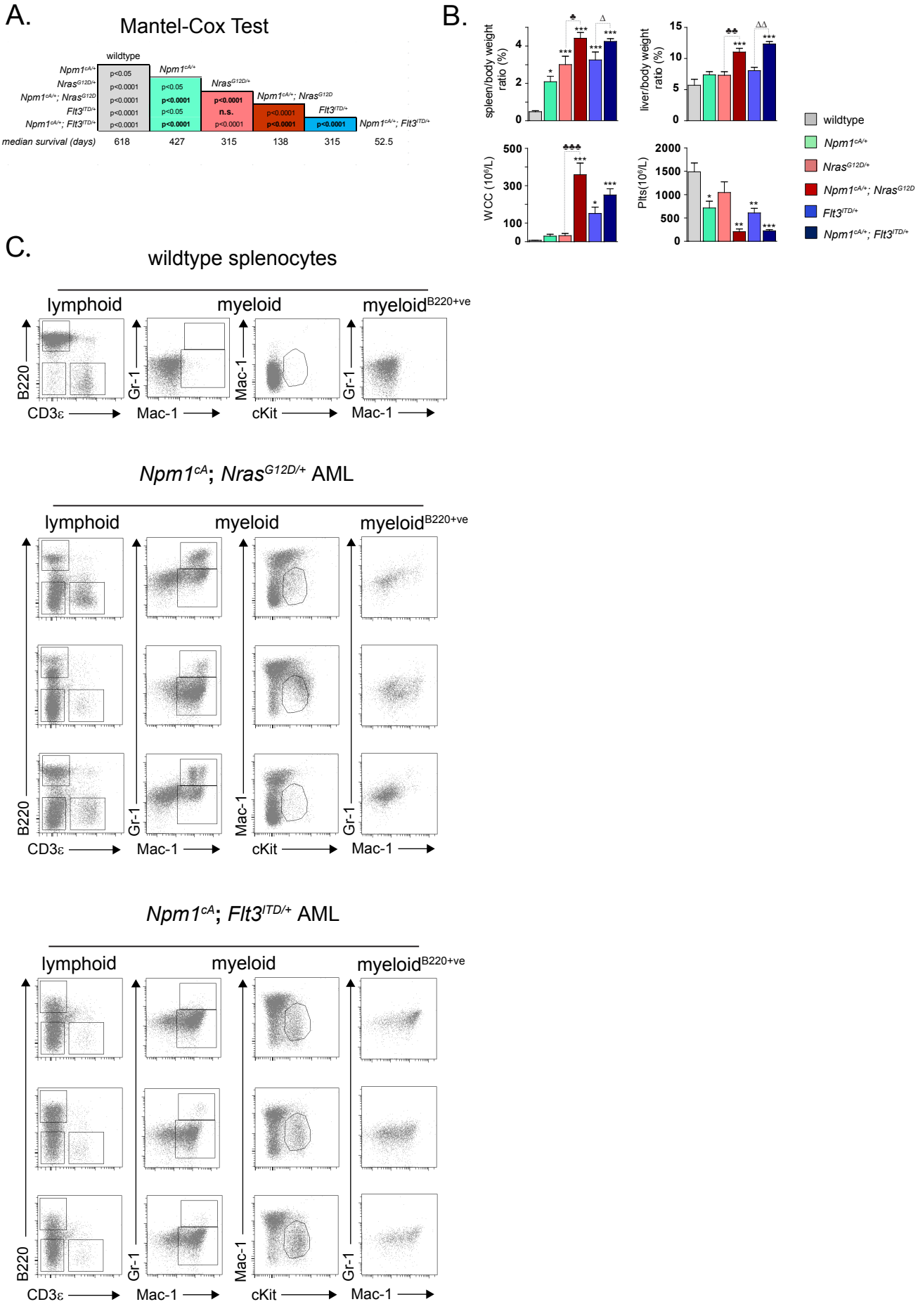
Supplemental Figure S2



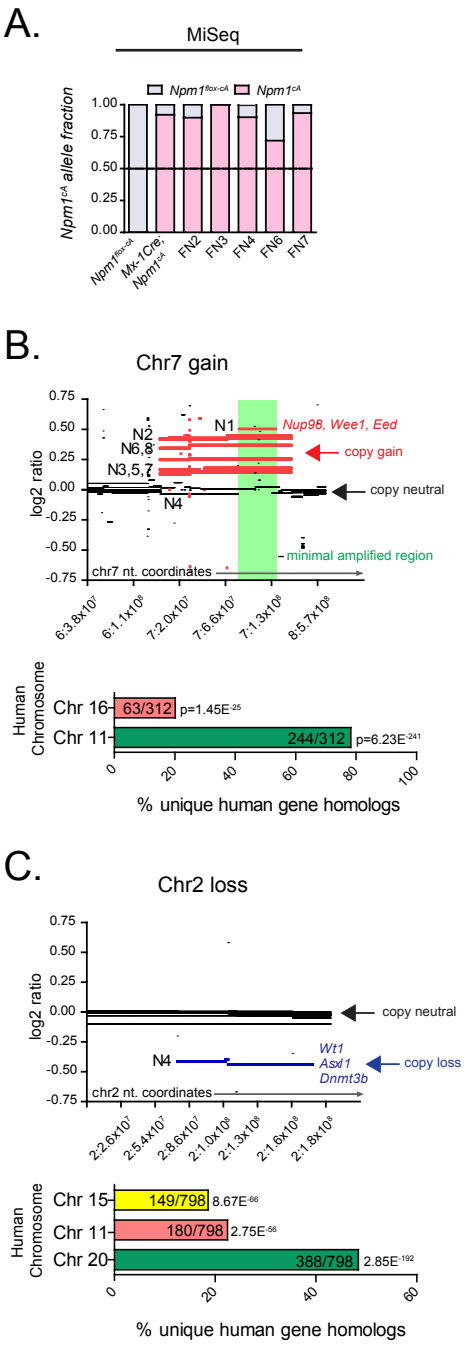
Supplemental Figure S3



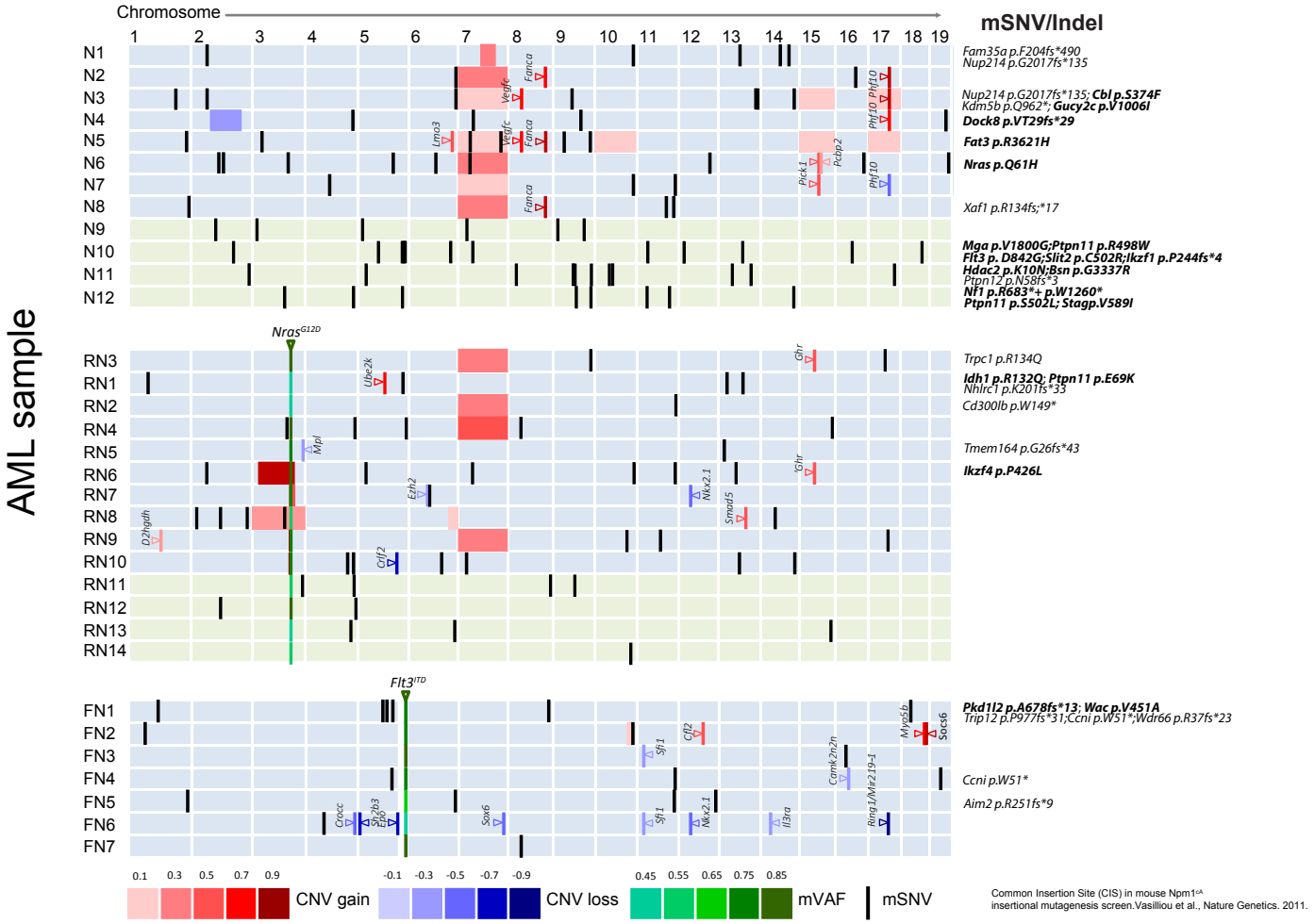
Supplemental Figure S4



Supplemental Figure S5

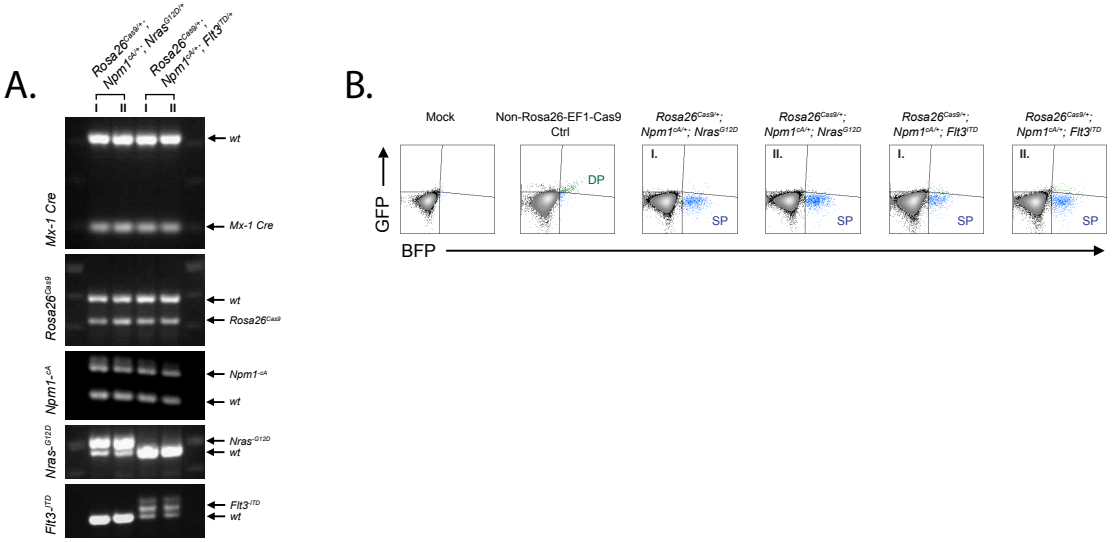


Supplemental Figure S6

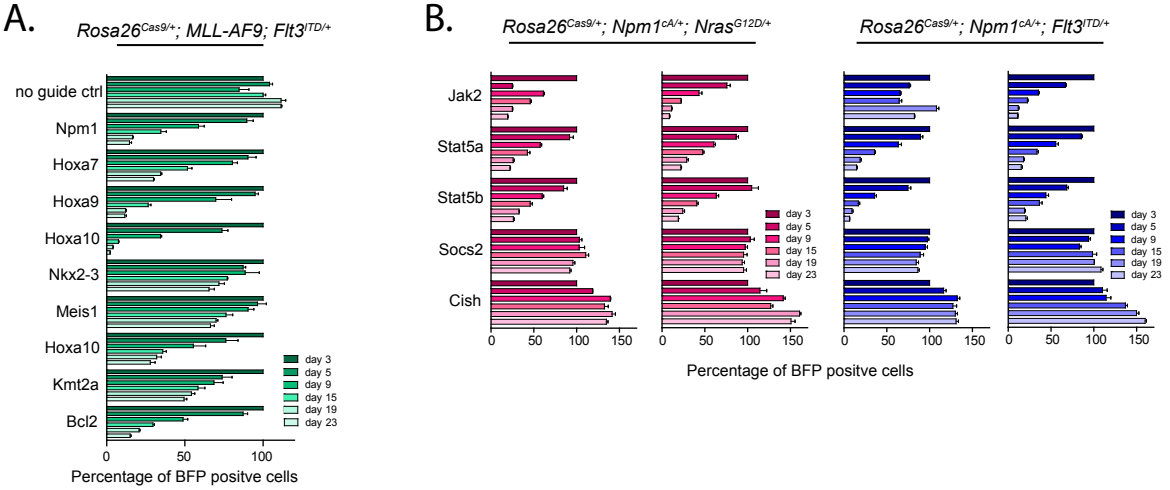


Common Insertion Site (CIS) in mouse Npm1[LA] insertional mutagenesis screen.Vasilliou et al., Nature Genetics. 2011.

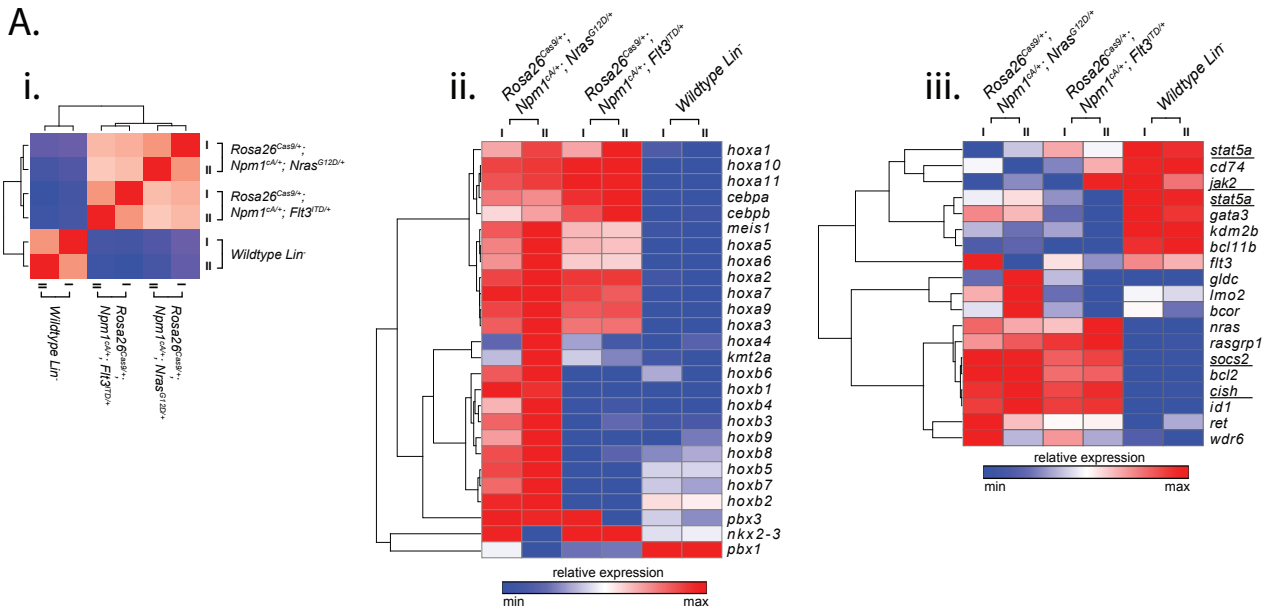
Supplementary Figure S7



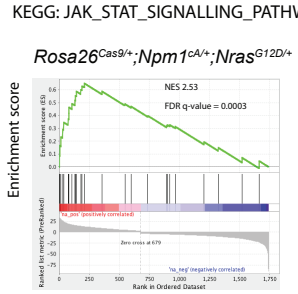
Supplementary Figure S8



Supplementary Figure 9



B.



Supplemental Figure Legends

Supplemental Figure S1. Pre-leukemic phenotypes of hematopoietic tissues. (**A**) Mac-1/Gr-1 staining of bone marrow and splenocytes from all genotypes shows an increase in myeloid commitment in *Npm1*$^{cA/+}$; *Nras*$^{G12D/+}$ (predominantly Mac-1^{+}/Gr-1^{+}granulocytic) and *Npm1*$^{cA/+}$; *Flt3*$^{ITD/+}$ (predominantly Mac-1^{+}/Gr-1^{-} monocytic) compared to singular mutants.

Supplemental Figure S2. Representative plots of pre-leukemic progenitor FACS of *Npm1*^{cA}, *Nras*$^{G12D/+}$, *Flt3*$^{ITD/+}$ and compound *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ or *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ mice. (**A**) Lin^{-}, LK, LSK. Percentages of parent populations are shown for LK, LSK, GMP, MEP, CMP, LMPP, MPP and ST/LT$^{prog-HSC}$ populations, mean ± SEM (n=4-8). Representative plots of (**B**) CLP and (**C**) E-SLAM, HSC FACS plots and gates. Note an absence of the double positive FLT3/Il7-Rα in the CLP stain. LK (Lin^{-}/Kit^{+}), LSK (Lin^{-}/Kit^{+}/Sca-1^{+}), CMP (common myeloid progenitor), MEP (megakaryocyte-erythroid progenitor), GMP (granulocyte-monocyte progenitor), MPP (multi-potent progenitor), LMPP (lymphoid primed multi-potent progenitor), CLP (common lymphoid progenitor) and HSC (hematopoietic stem cell).

Supplemental Figure S3. Global gene expression analysis of lineage negative and LSK-progenitors. (**A**) Heat map of *Hox* gene expression in Lin^{-} bone marrow from singular and compound *Npm1*$^{cA/+}$, *Nras*$^{G12D/+}$, *Flt3*$^{ITD/+}$ mice (normalised average expression values are used to generate heat map values). (**B**) Venn diagrams of overlapping differentially expressed genes in *Npm1*$^{cA/+}$, *Nras*$^{G12D/+}$, *Flt3*$^{ITD/+}$, *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ Lin^{-} bone marrow. Select over-expressed (red font) or under-expressed (blue font) are displayed. GSEA of differentially expressed genes in *Nras*$^{G12D/+}$ or *Flt3*$^{ITD/+}$ only mutants reveal enrichments for NRAS and JAK-STAT signalling pathways respectively. (**C**) Box-whisker plots of normalised average expression of *Nkx2-3*, *Hoxa7* and *Hoxa9*, as detected by microarrays in MPP and Lin^{-} populations. n=4-10 (Lin^{-}) or n=3-5 (MPP) for all genotypes (Mean ± Min-Max). (**D**) Venn diagram and heat map of overlapping and distinct differentially expressed genes in sorted MPP populations from *Npm1*$^{cA/+}$, *Nras*$^{G12D/+}$, *Flt3*$^{ITD/+}$, *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ reveals only a small sub-set of 12 deregulated genes shared in compound *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ mice (log fold change, logFC, Adj. p<0.05 was used to identify significantly differentially expressed genes). (**E**) Results of gene-annotation enrichment analysis and functional annotation of differentially expressed genes in *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ compared to wildtype MPPs (using DAVID). Statistically significant enriched Kegg pathways and enriched Gene Ontology term (GO-Term) clusters are shown (as depicted using Cytoscape 3.3.0). (**F**) A number of differentially expressed genes in *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ multipotent progenitors (MPPs) are also deregulated in compound Tet2$^{-/-}$; *Flt3*$^{ITD/+}$ lineage negative/Sca-1^{+}/c-Kit^{+} (LSK) progenitors when compared to wildtype controls.

Supplemental Figure S4. *Npm1*^{cA} and oncogenic *Nras*^{G12D} co-operate to drive AML. (**A**) Comparative survival statistics (Median survival and Mantel-Cox Test p values) of data presented in Figure 3a, Kaplan Meier. (**B**) Spleen and liver weights, blood leukocyte (WCC) and platelet (Plts) counts of wildtype (n=13), *Npm1*$^{cA/+}$ (n=17), *Nras*$^{G12D/+}$ (n=22), *Flt3*$^{ITD/+}$ (n=30), *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ (n=15) and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ (n=29). At time of sacrifice *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ display increased blood leukocyte (32.6±12, *Nras*$^{G12D/+}$ compared to 359±62 x10^{6}/L, *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$; and

151±34, *Flt3*$^{ITD/+}$ compared to 250±33 x10^6/L, *Npm1*$^{cA/+}$; *Flt3*$^{ITD/+}$) and reduced platelet counts (1046±227, *Nras*$^{G12D/+}$ compared to 504.9±209 x10^6/L, *Npm1*$^{cA/+}$; *Nras*$^{G12D/+}$; and 607±99, *Flt3*$^{ITD/+}$ compared to 225±25.8 x10^6/L, *Npm1*$^{cA/+}$; *Flt3*$^{ITD/+}$). Mean ±SEM are plotted. Significant values are reported for one-way analysis of variance (ANOVA, Bonferroni adjusted); (***** P<0.05 vs wildtype, ****** P<0.01 vs wildtype, ******* P<0.001 vs wildtype), (Δ P<0.05 vs *Flt3*$^{ITD/+}$, ΔΔ P<0.01 vs *Flt3*$^{ITD/+}$, ΔΔΔ P<0.001 vs *Flt3*$^{ITD/+}$), (♣ P<0.05 vs *Nras*$^{G12D/+}$, ♣♣ P<0.01 vs *Nras*$^{G12D/+}$, ♣♣♣ P<0.001 vs *Nras*$^{G12D/+}$). (**C**) FACS analysis of three of *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ AMLs confirms myeloid infiltration in secondary lymphoid tissue (splenocytes); lymphoid (CD3ε/B220), myeloid (Mac-1/Gr-1, Mac-1/Kit) and B220$^+$ myeloid (B220$^+$/Mac-1/Gr-1).

Supplemental Figure S5. Array comparative hybridisation (aCGH) of *Npm1*$^{cA/+}$,*Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and *Npm1*^{cA};*Flt3*^{ITD} murine AML. (**A**) To determine the extent of recombination of the *Npm1*$^{flox\text{-}cA}$ allele in FN-AMLs, we quantified the fraction of *Npm1*$^{flox\text{-}cA}$ and *Npm1*^{cA} allele reads using targeted amplicon specific MiSeq (see Materials and methods). As controls we used *Mx-1 Cre*;*Npm1*$^{cA/+}$ and *Npm1*$^{flox\text{-}cA}$ gDNA 4 months post pIpC injection, almost complete recombination is observed in the 5 samples tested. (**B**) Normalised Log2 ratio plots show gains (whole chromosome or smaller regions) of chr 3 in *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$. Green highlighted region denotes minimally mapped region of common chromosomal gain or loss (chr3: 102743581-103470550). A commonly amplified region of chr7 (ch7: 91838150-131492236) is detected in 7/8 *Npm1*$^{cA/+}$ and 4/9 *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ (not represented). 31 of the 312 genes in this region, syntenic to human chr11, are mutated in the TCGA AML data-set and include *Nup98*, *Wee1* and *Eed*. (**C**) A region of chr2 (chr2: 77889234-171131931) is deleted in *Npm1*$^{cA/+}$ AML, N4. Of the 741 genes within this region 57 are in the TCGA AML dataset and include genes deleted in AML, *Asxl1*, *Wt1* and *Dnmt3b*. Black smoothed line indicates copy neutral regions. Red or blue smoothed line denotes gain or loss, respectively, of a chromosomal region defined on the x-axis for a particular sample. For (**B**) and (**C**) enrichment of syntenic human-mouse genes are shown (enrichment p-values as determined using DAVID[12]).

Supplemental Figure S6. Combined copy number and somatic variants for *Npm1*$^{cA/+}$, *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and *Npm1*$^{cA/+}$;*Flt3*$^{ITD/+}$ AMLs. (**A**) Combined copy number (aCGH) and somatic variants (while exome sequencing) in *Npm1*$^{cA/+}$ (N-AML) compared to murine *Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ (RN-AML) and *Npm1*^{cA};*Flt3*^{ITD} (FN-AML) AML samples.

Supplemental Figure S7. Confirmation of CAS9 activity in *Rosa26-EF1-Cas9* AML cell lines. (**A**) Results of PCR genotyping *of Rosa26*$^{Cas9/+}$;*Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ and Rosa26 $^{Cas9/+}$;*Npm1*^{cA};*Flt3*^{ITD} AML cell lines. (**B**) Cas9 activity of mouse AML cell lines was validated by transducing *Rosa26* $^{Cas9/+}$expressing cells with a reporter construct expressing BFP/GFP and a guide RNA targeted to GFP. In the presence of active CAS9 double positive BFP/GFP cells are reduced to single positive expressing BFP cells through CRISPR-CAS9 mediated disruption of GFP.

Supplemental Figure S8. Genetic disruption by CRISPR-Cas9 does not reveal mutation specific dependencies in genes of the JAK/STAT signalling pathway in AML cell lines. (**A**) *Rosa26* $^{Cas9/+}$;*Flt3*$^{ITD/+}$ lineage negative bone marrow cells were transformed via transduction with *MLL-AF9* retrovirus.

Competitive assays were performed over a 23 day period using a gRNAs-BFP lentivirus targeted to the genes indicated. The BFP-positive fraction was compared with the non-transduced population. Results were normalized to day 3 for each gRNA. (**B**) Results from AML cell lines transduced with gRNAs targeting and JAK/STAT related genes.

Supplemental Figure S9. RNA sequencing reveals up-regulation of genes involved in JAK/STAT signalling, indicative of IL-3 activation. (**A**) (**i**) Clustered heat maps generated from gene expression profiles of Rosa26 $^{Cas9/+}$ AML cell lines and wildtype lineage negative (Lin^-) hematopoietic cells (n=2 technical replicates for each cell type). Heat maps for Hoxa (**ii**) and non-Hoxa (**iii**) genes. (**B**) Gene Set Enrichment Analysis of *Rosa26* $^{Cas9/+}$cell lines reveal enrichment in the JAK/STAT signalling pathway in *Rosa26* $^{Cas9/+}$;*Npm1*$^{cA/+}$;*Nras*$^{G12D/+}$ cell lines, a signature not present in our original analysis of isolated mouse progenitors.

Supplemental Methods Figure S1. Exome sequencing and mutant somatic variant validation. (**A**) Exome sequencing and MiSeq validation "pipe-line" for detecting non-synonymous mouse AML variants. (**B**) Representation of MiSeq amplicon sequencing protocol. (1) Genomic PCR was performed with genome specific/MiSeq adapter primer sequences, Supplemental Table 2.1. (2) Pooled PCR products were then (3) amplified by PCR enrichment using a universal PE1.0 forward and a unique iPCRTag reverse primer, Supplemental Table 2.2. Samples were further purified and sequenced on an Illumina MiSeq. (**C**) The percentage of SNVs detected by exome sequencing and validated by MiSeq amplicon specific sequencing increases to 83% when using an "exome sequencing score" ≥3. This is further increased to 86% upon removal of C>A/G>T trans-version SNVs with mVAF<0.3. Note, the exome sequencing score (generated by RAMSES) is a confidence value derived from the following criteria for each SNV within a given sample; (**i**) the presence of mutations in both forward and reverse reads, (**ii**) unique or multiple genomic loci alignment (BLAT) and (**iii**) read quality and depth.

References

1. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 2002;30(4):e15.
2. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article3.
3. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34(3):267-273.
4. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-15550.
5. Cancer Genome Atlas Research N, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059-2074.
6. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80.
7. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.
8. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
9. Martinez N, Almaraz C, Vaque JP, et al. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. *Leukemia.* 2014;28(6):1334-1340.
10. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics.* 2009;25(21):2865-2871.
11. Tzelepis K, Koike-Yusa H, De Braekeleer E, et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep.* 2016;17(4):1193-1205.
12. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol.* 2014;32(3):267-273.
13. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-57.