Epigenetic and functional changes imposed by NUP98-HOXA9 in a genetically engineered model of chronic myeloid leukemia progression

Ivan Sloma,[1-3°] Philip Beer,[1] Christophe Desterke,[4] Elizabeth Bulaeva,[1] Misha Bilenky,[5] Annaïck Carles,[6] Michelle Moksa,[6] Kamini Raghuram,[1] Cedric Brimacombe,[1] Karen Lambie,[1] Ali G. Turhan,[7] Orianne Wagner-Ballon,[2,3] Philippe Gaulard,[2,8] Keith Humphries,[1] Martin Hirst[5,6] and Connie J. Eaves[1,9]

[1]Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada; [2]Université Paris Est Creteil, INSERM, IMRB, Creteil, France; [3]Assistance Publique des Hôpitaux de Paris (AP-HP), Hopital Henri Mondor, Departement d'Hematologie et Immunologie, Creteil, France; [4]Université Paris-Sud, Faculté de Médecine Kremlin Bicêtre, INSERM UMS 33 Villejuif, France; [5]Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada; [6]Department of Microbiology and Immunology, Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada; [7]Service d'Hématologie Biologique, Hôpital Universitaire Paris-Saclay, INSERM U935, AP-HP, Kremlin Bicêtre, France; [8]INSERM UMRS955, Pathology Department, Hôpital Henri Mondor, AP-HP, Faculté de Médecine de Créteil, Université Paris-Est Créteil, Créteil, France and [9]University of British Columbia, Department of Medical Genetics, Vancouver, British Columbia, Canada

°Present address: Université Paris Est Creteil, INSERM, IMRB and AP-HP, Hopital Henri Mondor, Departement d'Hematologie et Immunologie, Creteil, France

Correspondence: IVAN SLOMA - ivan.sloma@aphp.fr

doi:10.3324/haematol.2020.249243

Epigenetic and functional changes imposed by NUP98-HOXA9 in a genetically engineered model of chronic myeloid leukemia progression

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY MATERIALS AND METHODS

Primary cells

Heparin-treated blood or leukapheresis cells were obtained from 3 CP CML patients with elevated WBC counts (Supplementary table 1). None had been previously treated with tyrosine kinase inhibitors. $CD34^+$ cell-enriched populations (88±10% $CD34^+$) were isolated either by positive selection or by removal of lineage-marker-positive (Lin^+) cells (using the EasySep kit or StemSep kit, respectively, from STEMCELL Technologies Vancouver, BC, Canada). These patients' samples were selected based on their predetermined elevated proportion of *BCR-ABL1*$^+$ 6-week LTC-derived CFCs (44%, 100% and 100% *BCR-ABL1*$^+$, respectively). All protocols using human cells were carried out according to procedures approved by the Research Ethics Board of the University of British Columbia (UBC).

Virus production

The *NA9* cDNA was excised from a MSCV-NA9–IRES-GFP retroviral vector originally obtained from Guy Sauvageau (IRIC, Montreal, PQ, Canada).[1] It was then cloned into the MND-PGK-GFP lentiviral vector and the construct sequence-verified to exclude coding mutations. The same lentivirus construct without the *NA9* cDNA, and containing the coding sequence of YFP instead of GFP, was used to generate control-transduced cells. Virus preparations were concentrated to equivalent titres as previously described.[2]

Transduction protocol

The protocol followed was the same as previously described.[2,3] Briefly, primary cells were pre-stimulated overnight (~16 hours) at a concentration of $5x10^5$ viable $CD34^+$ cells/ml in serum-free Iscove's medium containing bovine serum albumin, insulin and transferrin (BIT, STEMCELL Technologies) and 40 µg/ml low-density lipoproteins, 10^{-4} M 2-mercaptoethanol (Sigma, St Louis, MO, USA) (SFM) and the following 5 purified recombinant human GFs: 20 ng/ml each of IL-3, IL-6 and G-CSF, and 100 ng/ml SF (referred to as SFM+5GFs). Pre-stimulated cells were then transferred into human fibronectin (Sigma)-precoated tissue-culture wells containing SFM+5GFs and virus. After 6 hours at 37°C, the cells were removed, washed with PBS and used as described. K562 cells were similarly transduced in Iscove's medium plus 10% fetal bovine serum (FBS) and sorted 6 days later to obtain lysates that were then examined for NA9 expression by Western blotting.

***In vitro* assays**

Granulopoietic (CFU-GM), erythroid (CFU-E and BFU-E) and multi-potent myeloid (CFU-GEMM) progenitors were assessed in 12 day methylcellulose cultures containing 20 ng/ml each of GM-CSF and IL-3, IL-6 and G-CSF, 50 ng/ml SF and 3 units/mL erythropoietin, and outputs of these progenitors in 6-week LTC assays were performed as previously described.[2] CFC replating assays were performed by harvesting 2 replicate 1 ml methylcellulose cultures, diluting the contents in 40 ml PBS, and then plating the pelleted cells in duplicate in new (secondary or tertiary) methylcellulose assays containing the same growth factors (GFs) as the primary CFC assays. Suspension cultures consisted of cells added to SFM + 5GFs.

Colony genotyping

Individual 12 to 14 day-old colonies were removed with a fine pipette from CFC assays generally containing <100 colonies per 1 ml culture. For *BCR-ABL1* and *NA9* transcript

analyses, harvested colonies were washed individually in PBS and RNA was then extracted using a Picopure kit (Life Technologies, Carlsbad, CA, USA). Reverse transcription (RT) was performed using Superscript III and random hexamers (Life Technologies). Quantitative-PCR (q-PCR) measurements of *BCR-ABL1*, *NA9* and *GAPDH* transcripts were carried out using SYBR Green master mix (Life Technologies) and the following primers: forward *BCR*, ACTGTCCACAGCATTCCGC; reverse *ABL1*, GAGCGGCTTCACTCAGACC; forward-*NUP98*, GACAGCCACTTTGGGCTTT; reverse *HOXA9*, CGCTCTCATTCTCAGCATTG; forward *GAPDH,* ACGTACTCAGCGCCAGCATC; reverse *GAPDH,* ACCGTCAAGGCTGAGAACGG.

Xenotransplant analyses

Eight week-old NSG-3GS mice[4] were irradiated with 3.15 Gy of ^{137}Cs γ-rays and then each was injected intravenously with an equal mixture of $CD34^+$ CP CML cells separately exposed to control and *NA9* virus (total of $2x10^6$ cells/mouse, 3-4 mice per patient). Samples of blood cells were collected 3 weeks later and bone marrow aspirated cells thereafter. All procedures involving mice were carried out in accordance with Canadian Council on Animal Care guidelines and UBC Animal Care Committee-approved protocols. Blood and bone marrow cells were incubated in ammonium chloride to lyse the RBCs prior to being washed and stained with the following anti-human antigen-specific antibodies to: GlyA-PE (10F7MN labeled by contract), CD45-APC (2D1), CD34-APC (8G12), CD19-PE (4G7), CD14-PE (M0P9), CD15-PE (HI98) and CD3-APC (SK7) from BD, and CD20-PE (L27) and CD33-PE (P67.6) from STEMCELL Technologies, after incubation with a blocking reagent consisting of PBS with 2% FBS, 5% human AB serum, anti-human CD32 blocking antibody (clone IV.3 from STEMCELL Technologies). Immunohistochemistry was performed on deparaffinized tissue sections after appropriate antigen retrieval using a three-step immunoperoxidase technique on the BOND-III Autostainer (Leica Microsystems, Newcastle-upon-Tyne, UK) with monoclonal antibodies

against CD45 (clone 2B11-PD7/26), CD68 (clone KP1), mast cell tryptase (clone AA1) (all 3 from DakoCytomation, Glostrup, Denmark), and CD34 (clone QBEnd/10), CD33 (clone PWS44) (both from Leica Microsystems).

RNA-seq analyses

RNA was extracted from FACS-sorted GFP^+ ($NA9^+$) or YFP^+ (control) $CD34^+$ cells cultured for 2 days post-transduction in SFM+5GFs, using an Influx FACS (Becton Dickinson, San Jose, CA, USA). RNA was assessed for quality and quantified using an Agilent Bioanalyzer (Life Technologies). 10 ng of total RNA was rRNA depleted using a NEBNext rRNA Depletion Kit (New England BioLabs, E6310L). First strand cDNA was generated using a Maxima H minus First Strand cDNA Synthesis Kit (Thermo Scientific, K1652) with the addition of 1 µg of Actinomycin D (Sigma, A9415). The product was purified using in-house prepared 20% PEG, 1M NaCL Sera-Mag bead solution at 1.8X ratio and eluted in 35 µl of Qiagen EB buffer. Second strand cDNA was synthesized in a 50 µl volume using SuperScript Choice System for cDNA Synthesis (Life Technologies, 18090-019) with 12.5 mM GeneAmp dNTP Blend with dUTP. Double stranded cDNA was then purified with 20% PEG, 1M NaCL Sera-Mag bead solution at 1.8X ratio, eluted in 40 µL of Qiagen EB buffer, and fragmented using Covaris E220 (55 seconds, 20% duty factor, 200 cycles per burst). Sheared cDNA was End Repaired/Phosphorylated, single A-tailed, and Adapter Ligated using custom reagent formulations (New England BioLabs, E6000B-10) and in-house prepared Illumina forked adapter. 20% PEG, 1M NaCl Sera-Mag bead solution was used to purify the template in-between each of the enzymatic steps. To complete the process of generating strand directionality, adapter-ligated template was digested with 5 U of AmpErase Uracil N-Glycosylase (Life Technologies, N8080096). Libraries were then PCR-amplified and indexed using Phusion Hot Start II High Fidelity Polymerase (Thermo Scientific, F 549-L). An equal molar pool of each library was sequenced on HiSeq2000 (Illumina) PE75 and PE125.

For transcript quantification, paired-end 75 bp raw reads were aligned to a custom reference-containing genome (GRCh37-lite) and exon-exon junctions for Ensembl v75 gene models (JAGUAR[5]) with the Burrows-Wheeler Aligner (BWA-MEM; version 0.7.6a, with default parameters and '-P –M' options).[6] Reads aligned to the exon-exon junctions were repositioned to assign genomic coordinates. Duplicated fragments were marked with Picard 1.92 (MarkDuplicates) tool (https://broadinstitute.github.io/picard/). For every sample >98% of reads were aligned. For variant calling, JAGUAR generated bam files were sorted and merged using Samtools version 0.1.19.0 and converted to Fastq files with Bedtools version 2.17.0. Raw merged reads were then realigned with Tophat2 using the hg19 GRCh37.74 genome reference and the Refseq hg19 GTF transcripts file. Duplicates were removed with Samtools, and Read Groups were added to bam headers with Picard tools version 1.95. Post alignment processing was carried out according to RNAseq best practices with GATK version 3.1.1. Variant calling was performed using HaplotypeCaller (version 3.1.1) with the option "don't use soft clip bases". Transcript fusions were called with Tophat2 and option fusion.[7] Variant calls were annotated with Annovar (Version 2015-06-17),[8] and filtered to exclude polymorphisms when variant allele frequencies in the Exac0.3 database were above 0.01 and to exclude variants at positions showing a sequencing depth (DP) below 20X.

Transcriptome meta-analysis

Transcriptome public datasets for cells from CML patients were downloaded from Gene Expression Omnibus website: GSE47927 normalized matrix annotated with platform GPL6244 for the different hematopoietic compartments analysis,[9] and GSE4170 normalized matrix annotated with GPL2029 platform for the analysis of the different phases of the disease.[10] Transcriptome analyses were performed in R software environment 3.2.3. Expression heatmaps were drawn with made4[11] and gplots R-packages. Unsupervised PCA was performed with the FactoMineR R-package. Boxplots were drawn with ggplot2 R-package and GSEA was

performed with GSEA version 2.2.2.0 standalone software on MSigDb version 5.1.[12] Functional enrichment was performed with Go-Elite standalone software version 1.5.[13] Functional enrichment network was drawn with Cytoscape 3.0.2.[14]

H3K27ac ChIP-sequencing

Nucleosome density ChIP-seq was performed as described at https://www.jove.com/video/56085/generation-native-chromatin-immunoprecipitationsequencing-libraries.[15,16]

Libraries were sequenced on Illumina HiSeq 2500 using one lane for each (paired-end 125nt). Read sequences were aligned to NCBI Build 37 (hg19) human reference genome using BWA-backtrack aln sampe v0.5.7. [17] Bam file sorting and duplicates marking were performed using Sambamba v0.5.5.[18] The number of mapped reads produced was between 49 x 10^6 and 62 x 10^6 across H3K27ac libraries and between 34.1 x 10^6 and 44.4 x 10^6 across input libraries. Enriched regions were called using MACS2 v2.1.1 'callpeak' in narrow mode (default) with FDR cutoff 0.01. Input samples were used as control background.[19] Super-enhancers were called with ROSE[20,21] using the H3K27ac enriched regions identified using MACS2. The signal and rank values from the output files were used to generate the plots. Chip-Seq tertiary analysis methods are described in the supplementary information file.

Tertiary analysis of H3K27ac peaks

The first step following MACS2 (for proximal enhancers) or ROSE (for SE) peak calling was to analyze for each patient the differential peaks present in CML $CD34^+$ cells engineered with the NA9 vector as compared to the parental CML $CD34^+$ cells engineered with the empty control vector. Interval lists of NA9 and control CML $CD34^+$ cells were joined and intervals considered common when overlaps were ≥ 1bp. For each patient, 3 list of intervals were thus

generated with peak intervals found in NA9 cells only, in control cells only, and those common to both.

The second step was aimed at identifying intervals that were reproducibly found among the 3 patients of the study. To this end, the interval lists of peaks found in the NA9 cells only, in the controls only or in both were joined two by two and intervals considered common if overlaps were ≥25bp. For each category (e.g., for peaks found in NA9 CML CD34+ cells only, in control CML CD34+ cells only, or in both) the lists of overlapping intervals between the 3 patients were concatenated to obtain the final lists of intervals considered as reproducible if detected in at least two patients. Enhancer gene prediction was carried out with Great 3.0.0[22] with TSS ± 2kb for proximal enhancers and up to 100kb for distal enhancers). Functional enrichments were done with GREAT 3.0.0 using peak intervals and Enrichr[23] using either peak interval bed files or the gene list obtained following GREAT 3.0.0 enhancer prediction.

Statistical analysis

Results are shown as the arithmetic or geometric mean±SEM or SD as indicated. Differences between groups were assessed with the Student t-test or the Whitney test or Wilcoxon matched-pairs signed rank test using GraphPad Prism 6.0h (GraphPad Software, Inc.). A Spearman ranked Wilcoxon test was used for the correlation analyses. P values <0.05 were considered significant. The Survdiff function was used for survival analysis performed with the survival R-package.

SUPPLEMENTARY FIGURES AND DATASETS

Supplementary Figure 1: Forced expression of NA9 in CML cells.

(A) Expression of CD34 and CD38 in the CP CML samples used. From top to bottom are shown FACS profiles of the cells from patients 1, 2 and 3. **(B)** Representative FACS profiles showing the efficiency of gene transfer into cells from patient 1 (upper panel), 2 (middle panel) and 3 (lower panel). **(C)** Detection of *NA9* transcripts by RT-PCR in K562 cells engineered with control virus (ctrl) or the NA9 lentivirus. **(D)** Western blots of lysates of control and *NA9*-transduced K562 cells using anti-ABL (upper panel), anti-NUP98 (middle panel), and anti-β-actin (lower panel) antibodies. **(E)** Comparison of *BCR-ABL1* transcript levels as a function of *NA9* transcript levels (quantified relative to *GAPDH* transcript levels) in extracts of individual colonies produced from cells first cultured for 2 days following transduction of CP CML $CD34^+$ cells. Results are expressed as $2^{(Ct\ BCRABL1-Ct\ GAPDH)}$ for *BCR-ABL1* transcripts and $2^{(Ct\ NA9-Ct\ GAPDH)}$ for each colony (Spearman rank correlation test showed no significant correlation, P=0.65). **(F)** Comparison of CFC yields from replating assays of control- (white bars) or *NA9*-transduced $CD34^+$ CP CML cells (black bars) after 12 days in 1° and then 2° and 3° culture. Data are represented as mean of colonies counted in culture duplicates. Colonies numbers were not statistically different between control and *NA9*-transduced cells (Wilcoxon matched-pairs signed rank test). Black and white arrows indicate the corresponding detection limit when no colonies were seen for control- or *NA9*-transduced cells, respectively. **(G)** Outputs of patient 3's $CD34^+$ cells, (BFU-Es and CFU-GMs) in suspension cultures containing SFM+5 GFs. Data for control-transduced $CD34^+$ cells are shown as open symbols with dotted lines and for *NA9*-transduced cells as black symbols with solid lines. **(H)** Number of BFU-Es and CFU-GMs in 6-week LTCs per 10^4 input $CD34^+$ cells (n=3; p=0.03 ratio paired t-test). **(I)** Representative micrograph showing a cytospin of monocytic and granulocytic cells at various stage of differentiation and stained with Wright-Giemsa. These cells were derived from *NA9*-transduced

CP CML $CD34^+$ cells grown in suspension culture in SFM supplemented with 5 GFs for 21 days. Scale bar : 20µm. GFP, green fluorescent protein; MW molecular weight; bp, base pair; YFP, yellow fluorescent protein.

Supplementary Figure 2: Repopulation of NSG-3GS mice by transplanted *NA9*- and control-transduced CP CML $CD34^+$ cells.

(A) FACS-isolated $NA9^+$ (GFP^+) cells in the BM of mice showing eosinophilic, granulocytic and monocytic cells (upper micrograph) or predominantly monocytic (lower micrograph) cells stained with May Grunwald-Giemsa. **(B)** Hematoxylin eosin saffron (HES) stained section of a decalcified femur of a transplanted mouse. **(C)** CD45, CD68 and Tryptase immunostained sections of bone marrow (left), spleen (middle) and liver (right) samples from the same mouse as in **(B)**. On the left, is a dense marrow infiltrate of human hematopoietic $CD45^+$ cells (left) without megakaryocytes and erythroid cells. Myeloid infiltrates included scattered $CD68^+$ mature monocytic cells (spleen) together with tryptase-positive mast cells (liver). Micrographs were taken at x10 (HES), x20 ($CD45^+$ cells in bone marrow) and x40 (spleen CD68 and liver tryptase). HES, Hematoxylin eosin saffron; GFP, green fluorescent protein; YFP, yellow fluorescent protein.

Supplementary Figure 3: Analysis of differentially expressed transcripts in NA9-transduced CP CML $CD34^+$ cells.

(A) Functional annotation of differentially expressed genes common to all transcriptomes (NA9 signature from Figure 3D and Supplementary dataset 1) using Gene Ontology Biological Process data base and GoElite V1.2.[13] Red bars correspond to the number of genes and white bars the corresponding z-score. **(B)** Network of differentially expressed genes common to all transcriptomes.[14] **(C)** GSEA of whole transcriptomes showing significant enrichment of a

previously reported set of genes found up-regulated in $CD34^+$ cells 3 days following viral transduction with *NA9* (upper diagram, GSE57194).[24]

Supplementary Figure 4: Correlation of the NA9 signature of *NA9*-transduced CP CML $CD34^+$cells with transcriptional features of CML BP cells.

The NA9 signature identified in Figure 2B discriminates all except one of the transcriptomes of 42 CP and 28 BP CML patients' $CD34^+$ cells (GSE4170).[10]

Supplementary Figure 5: Differential expression of NA9 signature transcripts in CML patients according to their disease phase. Relative levels of expression of the genes of the NA9 signature were examined in the GSE4170 dataset[10] for samples obtained from CML patients' cells from different disease phases (CP in red, accelerated phase in blue, and BP in green) and compared by one-way Anova Fisher test. The most significantly differentially expressed genes with indicated P-values are shown. CP: Chronic Phase. AP: Accelerated Phase. BC: Blast Crisis.

Supplementary Figure 6: Gene set enrichment (GSE) analysis of the NA9 signature in AML patients' cells.

(A) GSE analysis of differentially expressed genes following NA9 forced expression in CML $CD34^+$ cells showing significant enrichment of genes from indicated AML transcriptome gene sets. *NPM1, MLL, FLT3, CEBPA and RUNX1-RUNX1T1 correspond to VERHAAK_AML_WITH_NPM1_MUTATED_UP,*[25]*ROSS_LEUKEMIA_WITH_MLL_FUSIONS,*[26] *VALK_AML_WITH_FLT3_ITD,*[27] *VALK_AML_WITH_CEBPA*[27] *and TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_UP*[28] gene sets, respectively. GSE was performed with GSEA version 2.2.2.0 standalone software on MSigDb version 5.1.[12]

(B) Venn diagram showing overlap of differentially expressed genes following NA9 forced

expression in CML CD34$^+$ cells found to be significantly enriched and positively correlated with genes from indicated gene sets. NA9 corresponds to the NA9 signature identified here. The Venn diagram was obtained with Venny 2.1.0. NES: Normalized Enrichment Score. FDR: False Discovery Rate.

Supplementary Figure 7: Oncoprint of the NA9 signature in AML cases from the TCGA cohort.

Oncoprint of the NA9 gene signature in AML cases from the TCGA[29] cohort was obtained from cBioPortal for Cancer and Genomics website.[30,31]

Supplementary Figure 8: NA9$^+$ CP CML CD34$^+$ signature predicts overall survival of TCGA AMLs.

The NA9 signature identified in Figure 3D and analyzed in AML samples from the TCGA cohort correlates with reduced overall survival (P = 0.025 as tested by univariate analysis using a log rank test, survival Package R 3.2.3).

Supplementary Figure 9: Integrative analysis of RNA-seq data with NA9-specific peaks of H3K27ac on chromatin.

(A) Unsupervised principal component analysis (PCA) performed with transcript candidates in RNA-seq which had specific proximal peaks for H3K27ac. **(B)** Expression heatmap of transcripts found to be predictive of NA9 during integrative analysis. **(C)** Bar-plot of transcripts predicted from an integrative analysis of the NA9-induced experimental state (negative $\log_{10}$ of p-value and R Pearson coefficient of NA9 experimental group data are presented). **(D)** H3K27ac peaks (blue) or NA9 binding sites from GSE62587 (orange) in indicated hg19 *loci* in control- (top), or *NA9*- (bottom) transduced CML CD34$^+$ cells. The grey bar in *FOXP1* indicates a SE called by ROSE.

Supplementary Dataset 1: Coding genes showing rapidly up-regulated expression in NA9-transduced CD34+ cells from different CP CML patients.

Table showing the lists of genes from the Venn diagram in Figure 1C. The Venn diagram was generated with Venny 2.1.0 using the lists of up-regulated transcripts (all transcripts from a single gene were collapsed to a single ENSG) resulting from pairwise comparison of transcript counts found in *NA9*-transduced $CD34^+$ cells compared with matched control-transduced cells.

Supplementary Datasets 2-5: Differentially expressed genes identified in NA9-transduced $CD34^+$ cells isolated from CML patients 1, 2 and 3.

Results of 3 pairwise comparisons of differentially expressed genes (collapsed transcripts) analyzed with DEfine v1 (unpublished). FDR control at 5% was used. Only genes that have a minimum RPKM=0.01 (or at least 25 reads mapped) in at least one of the conditions was considered in differential expression analysis.

Column headers are the following:

<HUGO name> <ENSEMBL ID> <corrected RPKM in NA9-GFP> < corrected RPKM in Ctrl-YFP> <p-value>.

Supplementary Dataset 2: List of coding genes upregulated in $NA9^+$ CP CML $CD34^+$ cells compared with matched control-transduced cells.

Supplementary Dataset 3: List of coding genes downregulated in $NA9^+$ CP CML $CD34^+$ cells compared with matched control-transduced cells.

Supplementary Dataset 4: List of non-coding genes upregulated in $NA9^+$ CP CML $CD34^+$ cells compared with matched control-transduced cells.

Supplementary Dataset 5: List of non-coding genes downregulated in NA9$^+$ CP CML CD34$^+$ cells compared with matched control-transduced cells.

Supplementary Dataset 6: Comparison of the set of genes upregulated in NA9$^+$ mPB CD34$^+$ cells[24] and in NA9$^+$ CML CD34$^+$ cells.

Supplementary Dataset 7: List of genomic intervals (hg19) showing H3K27ac peaks in NA9$^+$ CP CML CD34$^+$ cells of at least two different patients and absent in control CML CD34$^+$ cells engineered with an empty vector.

Supplementary Dataset 8: Functional annotation of H3K27ac marks of the Supplemental Dataset 7 using GREAT 3.0.

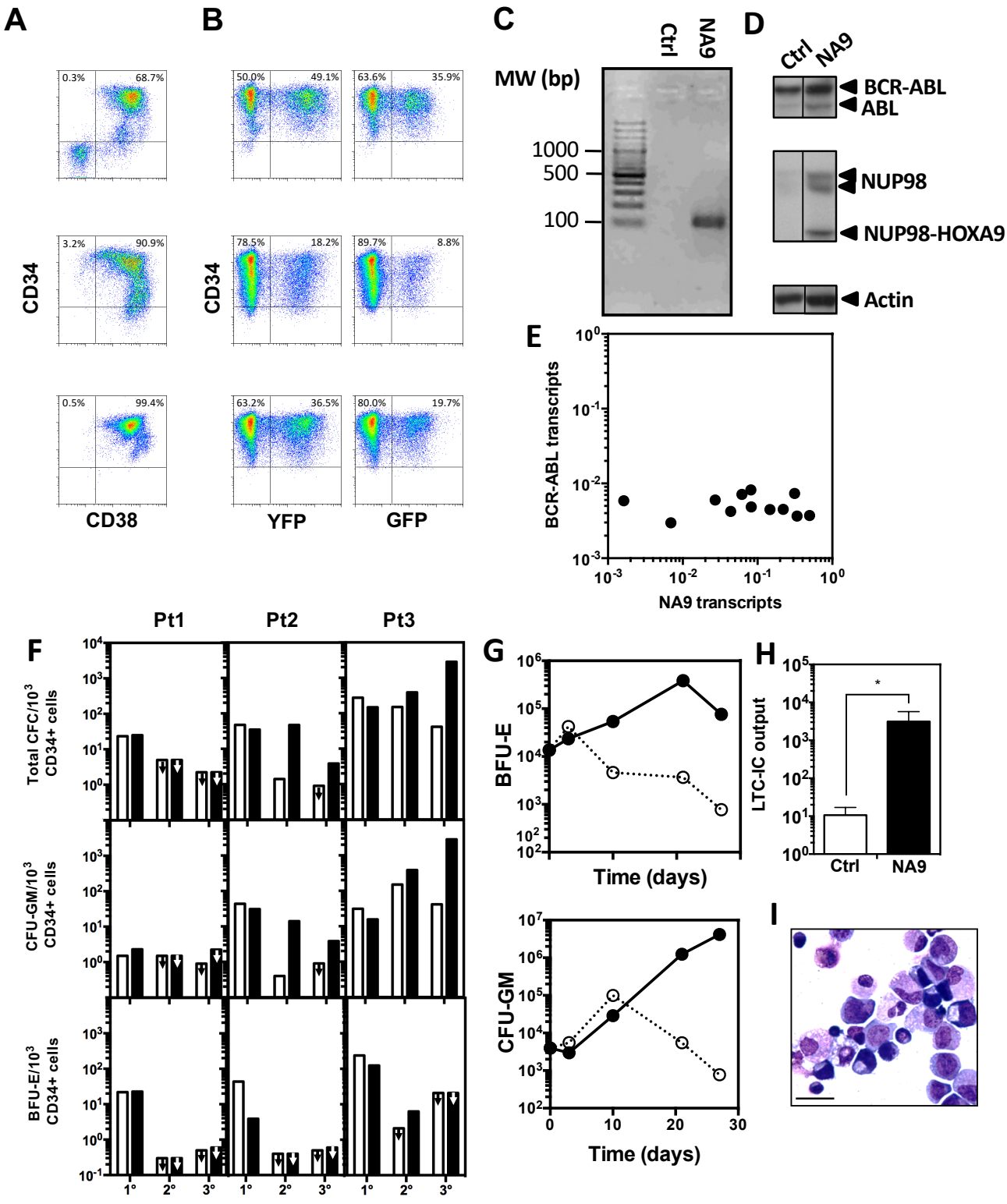
Supplementary Dataset 9: List of genomic intervals of super-enhancers in NA9$^+$ CP CML CD34$^+$ cells and absent in control CML CD34$^+$ cells engineered with an empty vector and their annotation with GREAT 3.0.

REFERENCES

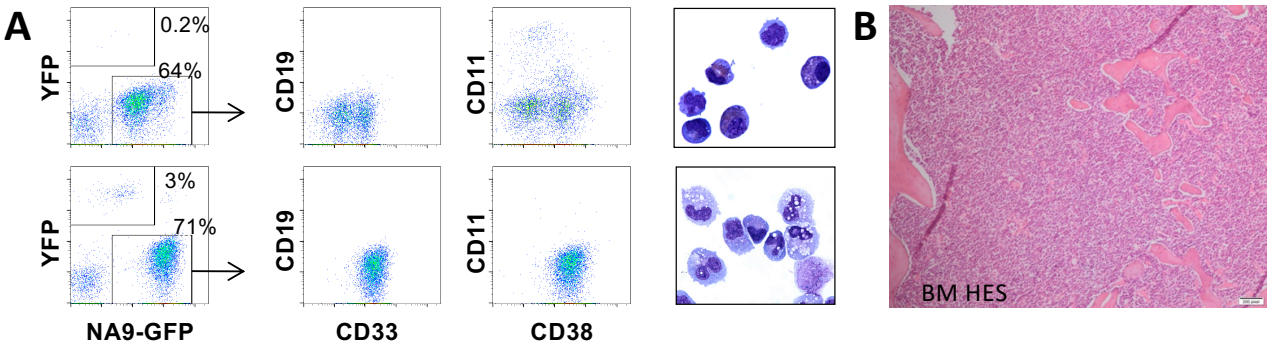
1. Kroon E, Thorsteinsdottir U, Mayotte N, Nakamura T, Sauvageau G. NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice. EMBO J. 2001;20(3):350-361.
2. Sloma I, Imren S, Beer PA, et al. Ex vivo expansion of normal and chronic myeloid leukemic stem cells without functional alteration using a NUP98HOXA10homeodomain fusion gene. Leukemia. 2013;27(1):159-169.
3. Beer PA, Knapp DJ, Miller PH, et al. Disruption of IKAROS activity in primitive chronic-phase CML cells mimics myeloid disease progression. Blood. 2015;125(3):504-515.
4. Miller PH, Cheung AM, Beer PA, et al. Enhanced normal short-term human myelopoiesis in mice engineered to express human-specific myeloid growth factors. Blood. 2013;121(5):e1-4.
5. Butterfield YS, Kreitzman M, Thiessen N, et al. JAGuaR: junction alignments to genome for RNA-seq reads. PLoS One. 2014;9(7):e102398.
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-1760.
7. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 2011;12(8):R72.
8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
9. Cramer-Morales K, Nieborowska-Skorska M, Scheibner K, et al. Personalized synthetic lethality induced by targeting RAD52 in leukemias identified by gene mutation and expression profile. Blood. 2013;122(7):1293-1304.
10. Radich JP, Dai H, Mao M, et al. Gene expression changes associated with progression and response in chronic myeloid leukemia. Proc Natl Acad Sci U S A. 2006;103(8):2794-2799.
11. Culhane AC, Thioulouse J, Perrière G, Higgins DG. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics. 2005;21(11):2789-2790.
12. 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.
13. Zambon AC, Gaj S, Ho I, et al. GO-Elite: a flexible solution for pathway and ontology over-representation. Bioinformatics. 2012;28(16):2209-2210.
14. Cline MS, Smoot M, Cerami E, et al. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007;2(10):2366-2382.
15. Lorzadeh A, Bilenky M, Hammond C, et al. Nucleosome Density ChIP-Seq Identifies Distinct Chromatin Modification Signatures Associated with MNase Accessibility. Cell Rep. 2016;17(8):2112-2124.
16. Lorzadeh A, Lopez Gutierrez R, Jackson L, Moksa M, Hirst M. Generation of Native Chromatin Immunoprecipitation Sequencing Libraries for Nucleosome Density Analysis. J Vis Exp. 2017(130).
17. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-595.
18. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015;31(12):2032-2034.
19. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

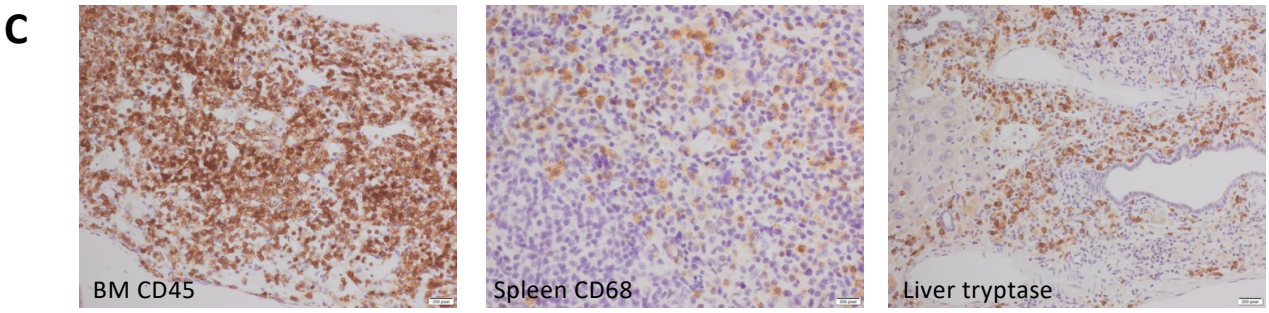
20. Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307-319.
21. Lovén J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320-334.
22. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495-501.
23. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.
24. Takeda A, Goolsby C, Yaseen NR. NUP98-HOXA9 induces long-term proliferation and blocks differentiation of primary human CD34+ hematopoietic cells. Cancer Res. 2006;66(13):6628-6637.
25. Verhaak RG, Goudswaard CS, van Putten W, et al. Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. Blood. 2005;106(12):3747-3754.
26. Ross ME, Mahfouz R, Onciu M, et al. Gene expression profiling of pediatric acute myelogenous leukemia. Blood. 2004;104(12):3679-3687.
27. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med. 2004;350(16):1617-1628.
28. Tonks A, Pearn L, Musson M, et al. Transcriptional dysregulation mediated by RUNX1-RUNX1T1 in normal human progenitor cells and in acute myeloid leukaemia. Leukemia. 2007;21(12):2495-2505.
29. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-2074.
30. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-404.
31. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.

Supplementary Figure 1



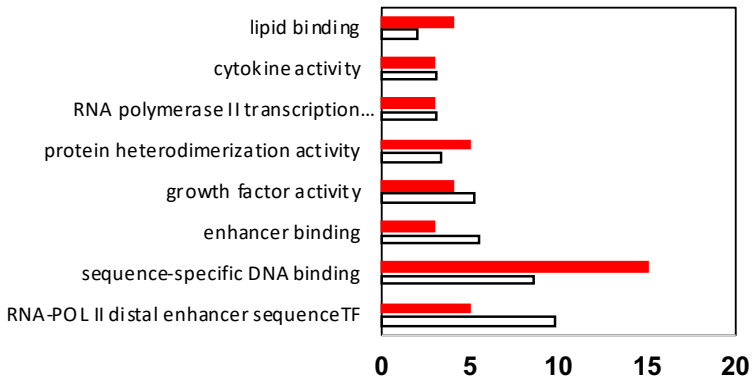
Supplementary Figure 2



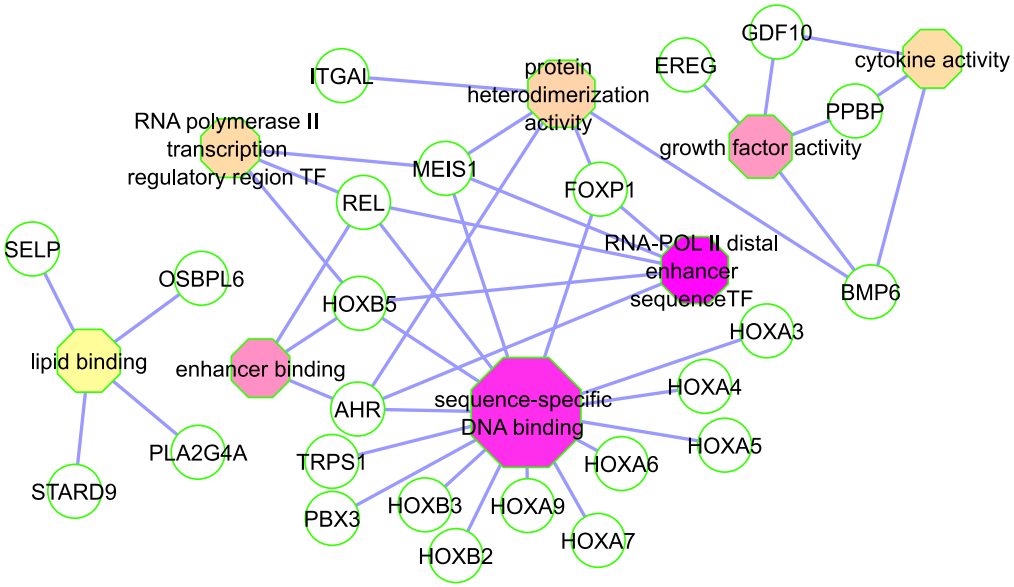


Supplementary Figure Figure 3

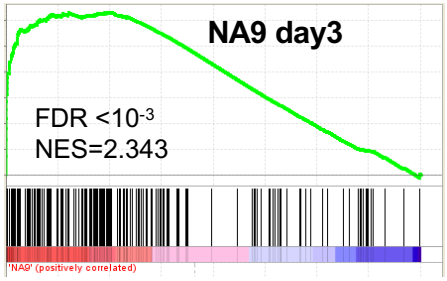
A



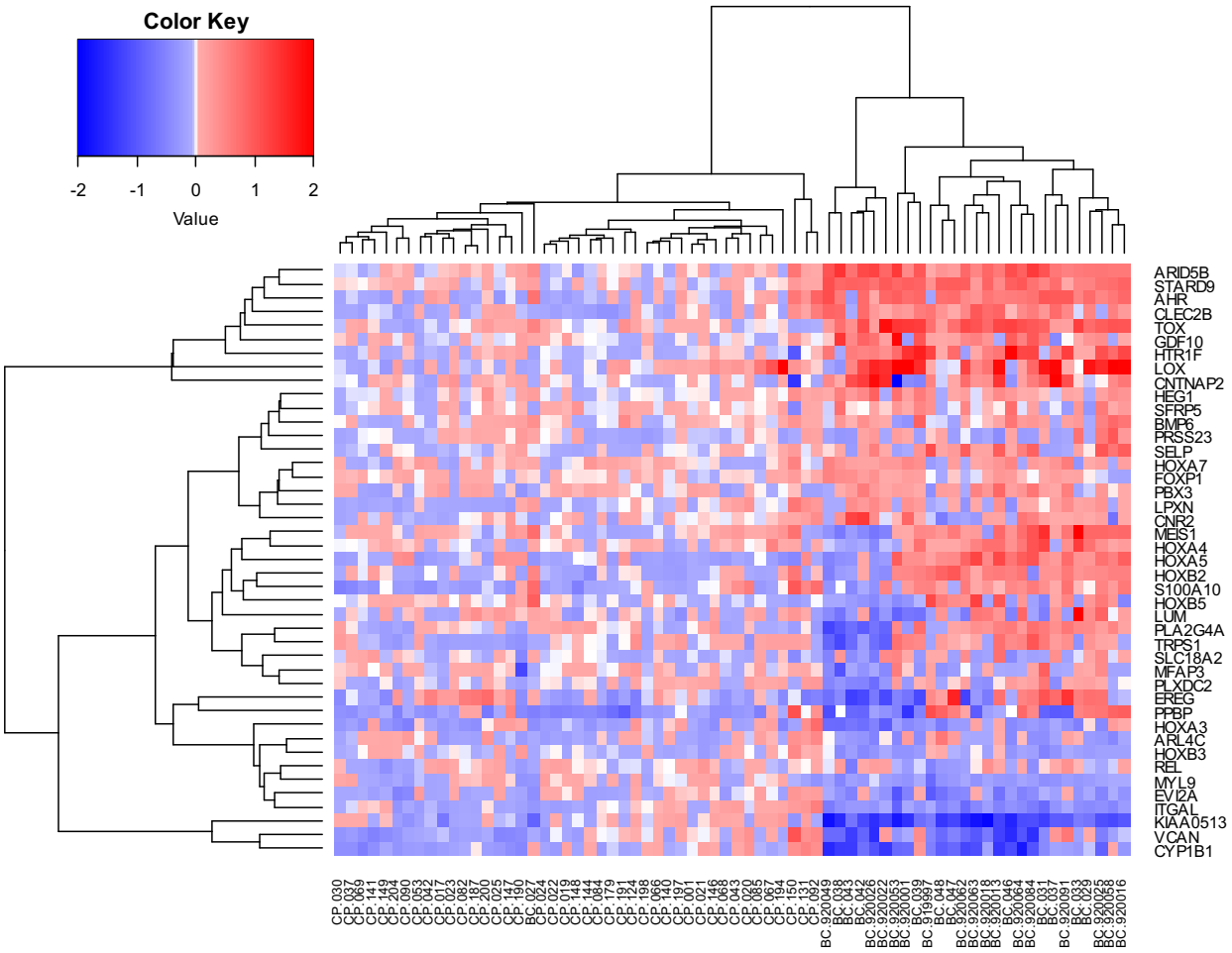
B



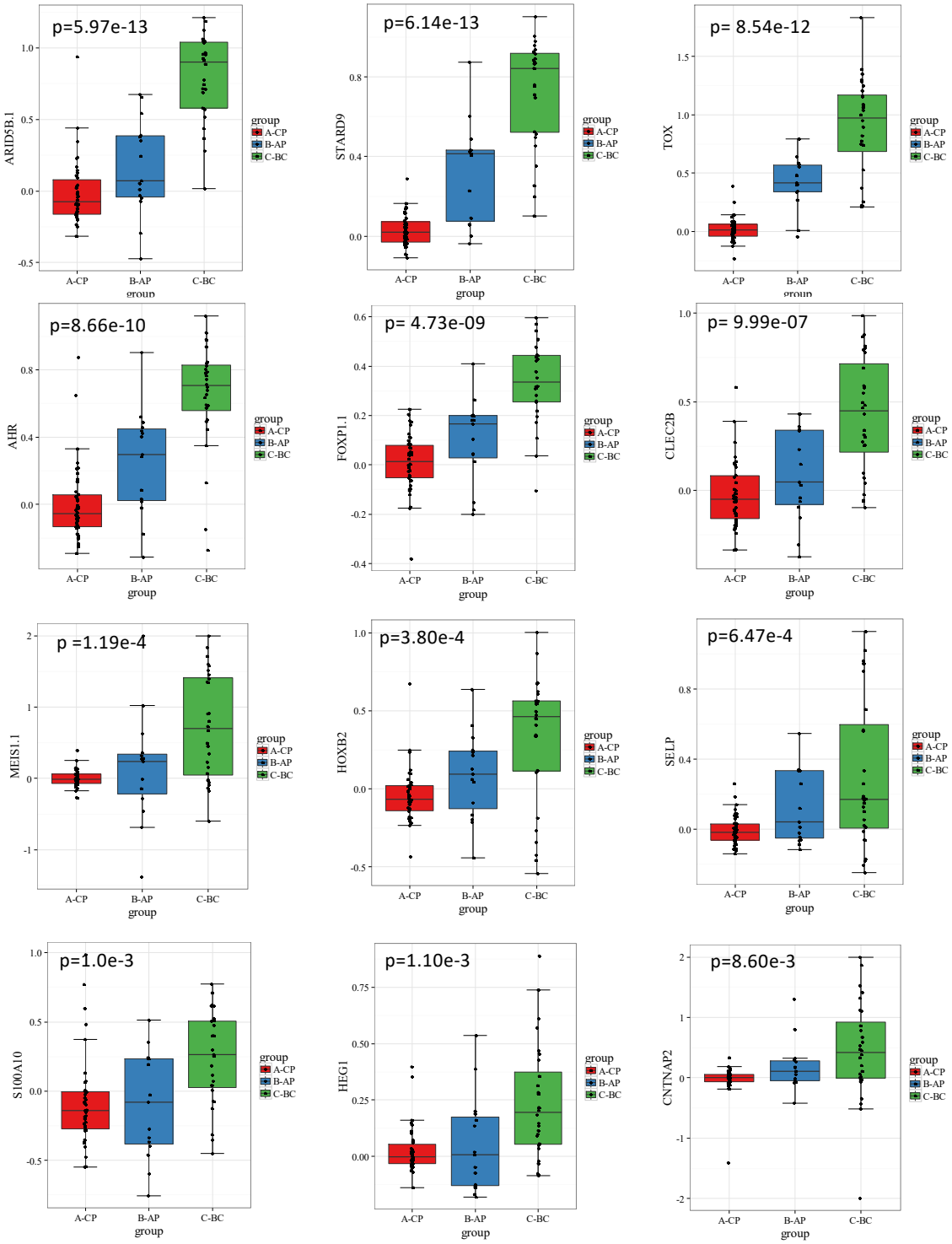
C



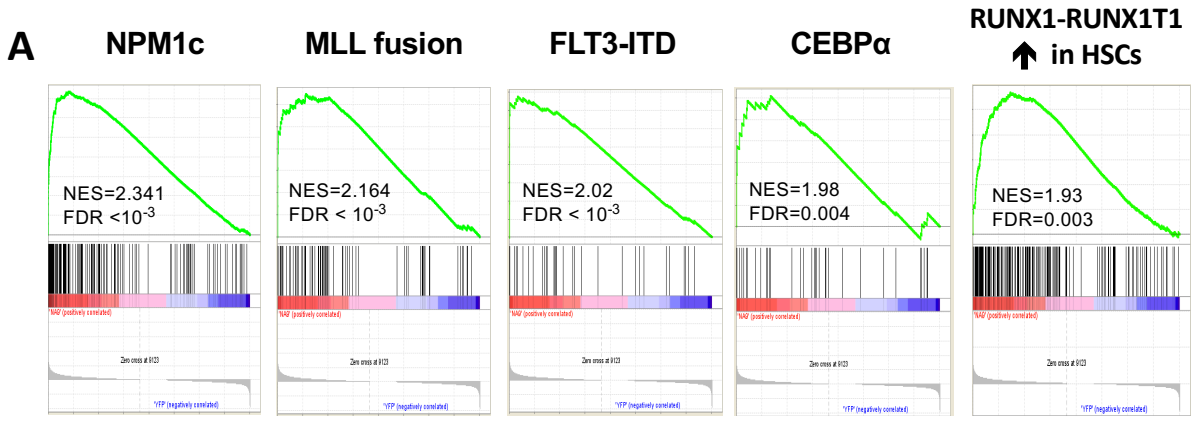
Supplementary Figure 4

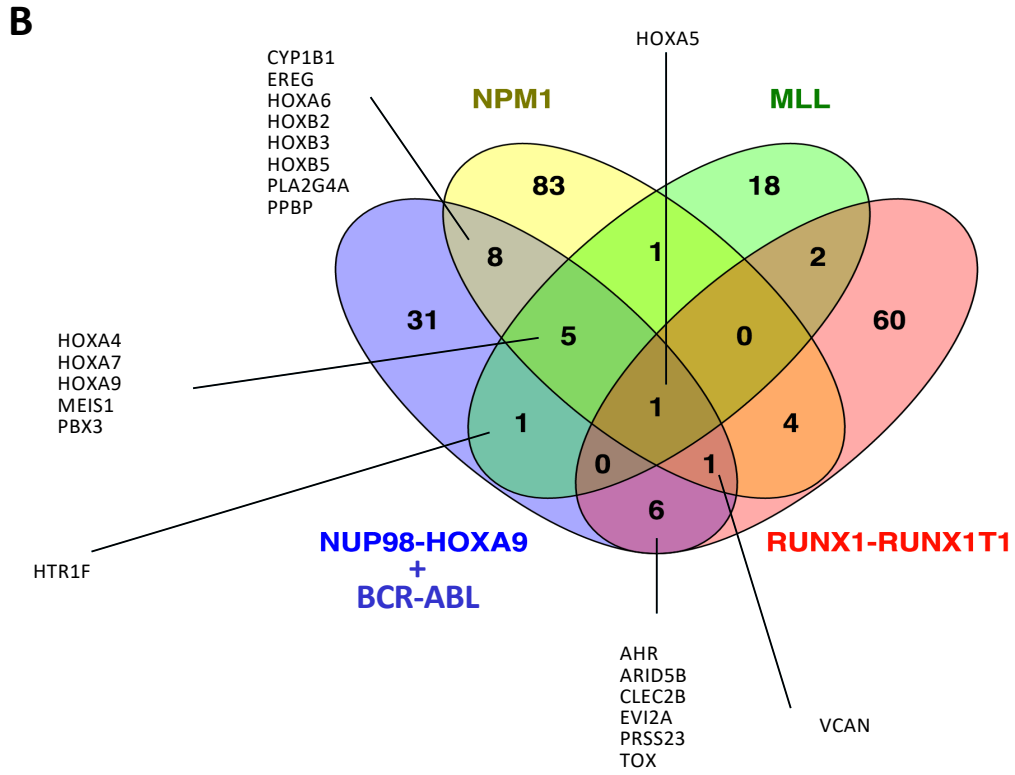


Supplementary figure 5



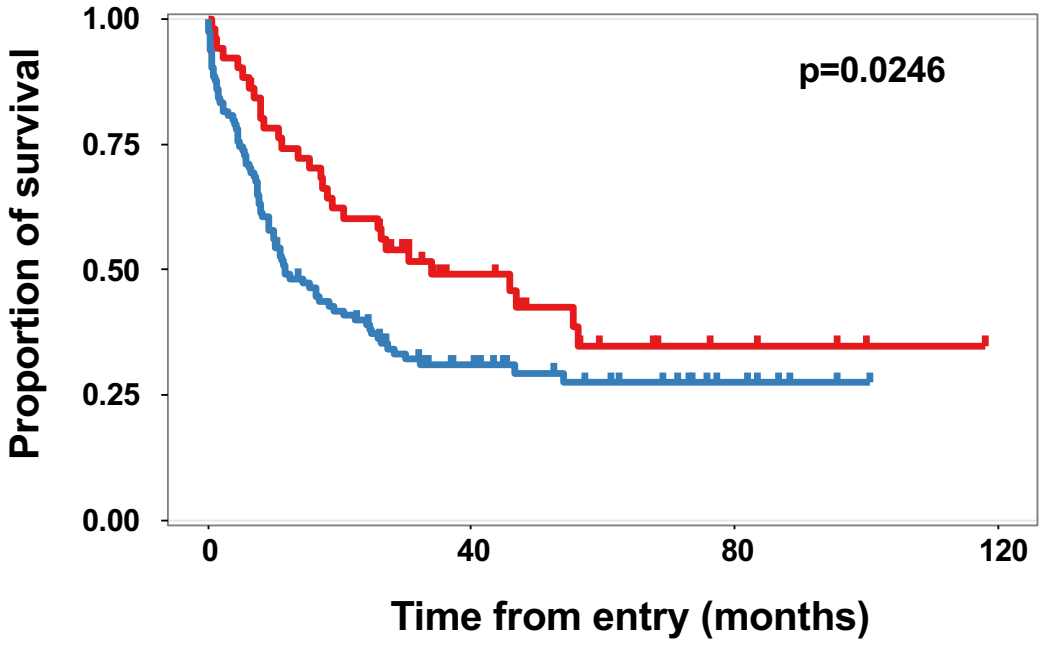
Supplementary figure 6



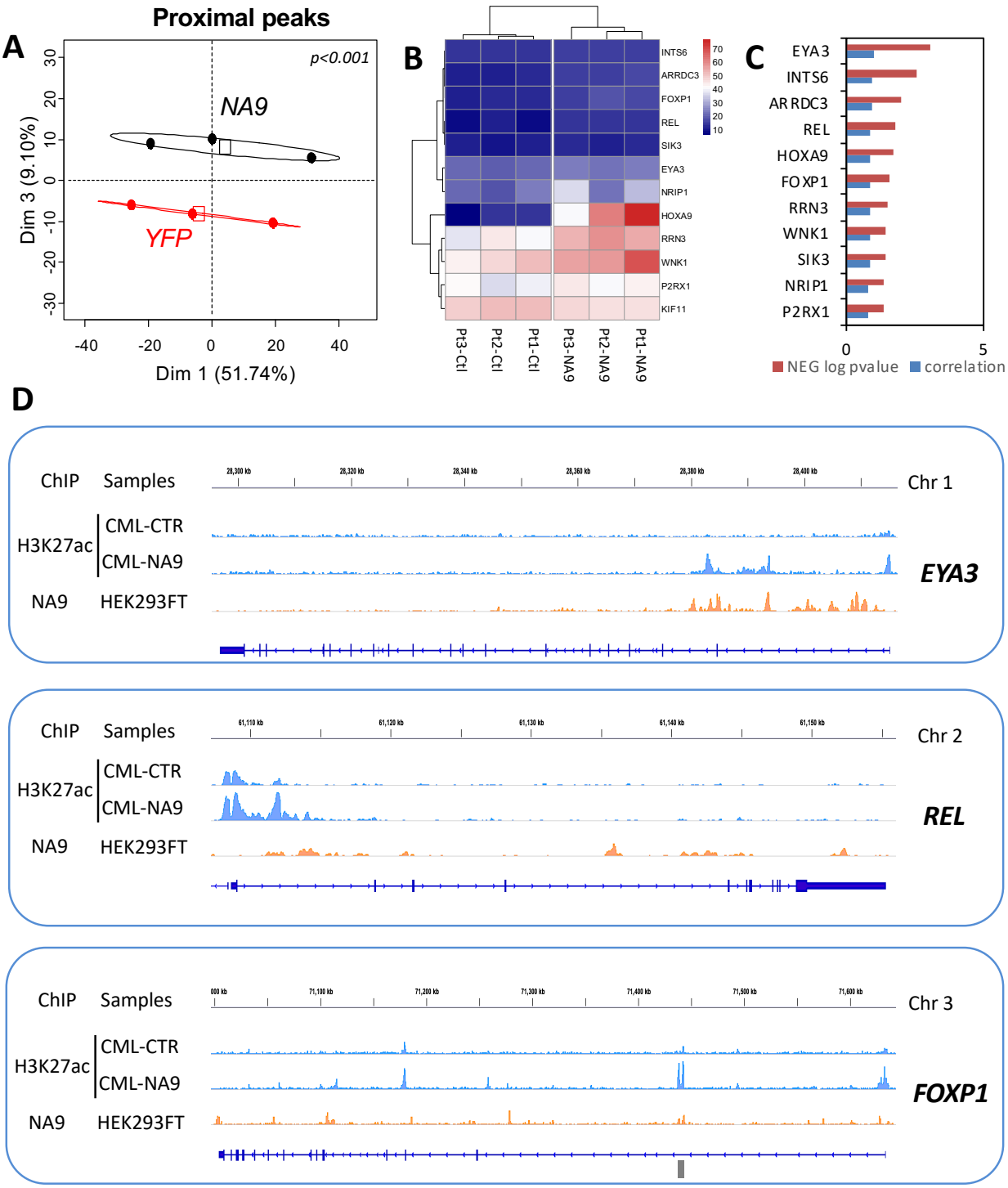


Supplementary figure 7

Supplementary figure 8



Supplementary Figure 9



Patient No.	Origin	Sex	Age	Time post Dx	WBC ($x10^9$/L)	Treatment	%B/A$^+$ LTC-ICs
1	PB	M	58	2 weeks	38	IM	>96%
2	PB*	M	n.a.	10 years	n.a.	HU	>97%
3	PB	M	57	3 weeks	457	IM	44%

Supplementary table 1. CML patient data

Dx: diagnosis. *: Leucopheresis. B/A: BCR-ABL1. IM: Imatinib. LTC-ICs: Long-term culture-initiating cells. M: male. n.a: not available. PB : peripheral blood.

Patient	Mouse	Week	Sample	BM cytology	Spleen cytology	BM histology	Spleen histology	Liver histology	Disease
1	23L	23	Necropsy			Focal sheets of eosinophils; patchy fibrosis with MC infiltrate	Normal	Normal	**MPN**
	24L	16	hCD45+	Mo+++					
		28	Necropsy			Slightly reduced E/Mk, marked increase in mature myeloid cells	Increased E/Mk; focal myeloid infiltrates	patchy myeloid infiltrate	**MPN**
	24R	28	Necropsy			Focal patches of left-shifted eosinophils	Normal	Normal	**MPN**
2	23B	8	Necropsy	MC++ Eo++ Mo++	MC++		Peripheral necrosis	Peripheral necrosis; inflammatory infiltrate	**MPN**
	24B	16	hCD45+	Mo+++					
		28	Necropsy			Decreased cellularity, normal appearances	Normal	Normal	**Normal**
	25R	8	Necropsy	Dead cells	Dead cells	Marrow effaced by sheets of eosinophils and monocytes; E/Mk near absent	Widespread patchy necrosis; eosinphil infiltrate	mild patchy inflammation; patchy necrosis	**MPN**
3	23N	10	hCD45+	Eo Mo					
		28	Necropsy			Mk/erythroid normal; fat spaces present; increase in monocytes	increased myeloid cells: monocytes and mast cells	mild patchy eosinophil infiltrate	**MPN**
	23R	28	Necropsy			Mk/erythroid normal; fat spaces present; patchy fibrosis; MC infiltrate		mild patchy eosinophil and mast cell infiltrate	**MPN**
	24N	10	hCD45+	Mo++ Eo+ MC+					
		27	Necropsy	n.d.	n.d.	Increased cellularity, no fat space, Focal sheets of eosinophils; scattered monocytes and mast cells	Massive infiltrate of red pulpe by eosinophils, monocytes, mast cells	periportal myeloid infiltrate	**MPN**
	25L	10	hCD45+	Eo++ Mo++					
		12	Necropsy	Mo++ MC++	MC++	E reduced; patchy infiltrate of eosinophils or monocytes	Increased erythroid; focal monocyte infiltrates	mild patchy mono infiltrate	**MPN**

Supplementary table 2: Transplanted mouse data

n.d.: Not done. Eo: Eosinophiles. E: Erythrocytes. MC: Mast cells. Mk: Megacaryocytes. Mo: Monocytes. MPN: Myeloproliferative neoplasm.