Blood First Edition Paper, prepublished online May 10, 2019; DOI 10.1182/blood.2018862383

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Hepatic leukemia factor is a novel leukemic stem cell regulator in DNMT3A, NPM1, and FLT3-ITD triple-mutated AML.

Tracking no: BLOOD/2018/862383R3

Swati Garg (Department of Medicine V, Hematology, Oncology and Rheumatology, University Hospital Heidelberg), Armando Reyes-Palomares (EMBL), Lixiazi He (Department of Medicine V, Hematology, Oncology and Rheumatology, University Hospital Heidelberg), Anne Bergeron (Universite Laval), Vincent-Philippe Lavallée (Laboratory of molecular genetics of stem cells, Institute for Research in Immunology and Cancer, Uni), Sébastien Lemieux (Université de Montréal), Patrick Gendron (Université de montréal), Christian Rohde (Department of Medicine V, Hematology, Oncology and Rheumatology, University Hospital Heidelberg), Jianglong Xia (Department of Medicine V, Hematology, Oncology and Rheumatology, University Hospital Heidelberg), Prarabdha Jagdhane (Department of Medicine V, Hematology, Oncology and Rheumatology, University Hospital Heidelberg), Carsten Müller-Tidow (University of Halle), Daniel Lipka (German Cancer Research Center (DKFZ)), Suzan Imren (Fred Hutchinson Cancer Research Center), R. Humphries (British Columbia Cancer Agency), Claudia Waskow (Leibniz-Institute on Aging - Fritz-Lipmann-Institute (FLI)), Binje Vick (Helmholtz Zentrum München), Irmela Jeremias (Helmholtz Zentrum München), Guillaume Richard-Carpentier (The Quebec Leukemia Cell Bank and Division of Hematology-Oncology), Josée Hébert (Hopital Maisonneuve-Rosemont), Guy Sauvageau (Université de Montréal), Judith Zaugg (EMBL), Frédéric Barabé (Université Laval), and Caroline Pabst (Department of Medicine V, Hematology, Oncology and Rheumatology, University Hospital Heidelberg)

Abstract:
FLT3, DNMT3A, and NPM1 are the most frequently mutated genes in cytogenetically normal acute myeloid leukemia (AML), but little is known about how these mutations synergize upon co-occurrence. Here we show that triple-mutated AML is characterized by high leukemia stem cell (LSC) frequency, an aberrant leukemia specific $GPR56^{high}CD34^{low}$ immunophenotype, and synergistic upregulation of Hepatic Leukemia Factor (HLF). Cell sorting based on the LSC marker GPR56 allowed isolation of triple mutated from DNMT3A/NPM1 double-mutated subclones. Moreover, in DNMT3A R882 mutated patients, CpG hypomethylation at the HLF transcription start site correlated with high HLF mRNA expression, which was itself associated with poor survival. Loss of HLF 3 via CRISPR/Cas9 significantly reduced the CD34+GPR56+ LSC compartment of primary human triple-mutated AML cells in serial xenotransplantation assays. HLF knockout cells were more actively cycling when freshly harvested from mice, but rapidly exhausted when re-introduced in culture. RNA-sequencing (RNA-Seq) of primary human triple-mutated AML cells after shRNA mediated HLF knockdown revealed the NOTCH target Hairy And Enhancer Of Split 1 (HES1) and the cyclin-dependent kinase inhibitor CDKN1C/p57 as novel targets of HLF potentially mediating these effects. Overall our data establish HLF as a novel LSC regulator in this genetically defined high-risk AML subgroup.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: S.G. performed experiments, analyzed data, generated figures, and wrote the manuscript. A.R.P. performed computational analyses, generated figures, and wrote the manuscript. S.G. and A.R.P. contributed equally to this work. L.H. contributed to in vivo experiments and edited the manuscript, A.B. performed sorting experiments, V.P.L., S.L., and P.G. performed RNA-Sequencing and mutational analyses of the Leucegene cohort samples, P.J. and J.X. helped perform in vivo experiments and lentiviral transductions, C.R. and C.M.T. performed and analyzed 450K arrays on AML samples, D.B.L. analyzed 450K arrays, S.I. and K.H. generated model leukemia cells, C.W. provided NSGW41 mice and supported in vivo experiments, B.V. and I.J. generated and characterized PDX AML cells, G.R.C. performed the survival analyses in the Leucegene cohort, J.H. and G.S. provided AML samples, RNA-Seq and clinical data of the Leucegene sample cohort, J.Z. performed bioinformatics analyses, revised the manuscript and co-supervised the project, F.B. and C.P. co-directed the project and wrote the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Sequencing data are available through accession numbers GSE49642, GSE52656, GSE62190, GSE66917, GSE67039, GSE48843, GSE48846, GSE51984, GSE68623, GSE129094, and GSE128848. HLF ChIP data was downloaded from GSE69817.

Clinical trial registration information (if any):

Hepatic leukemia factor is a novel leukemic stem cell regulator in DNMT3A, NPM1, and FLT3-ITD triple-mutated AML

Swati Garg[1,2]*, Armando Reyes-Palomares[3,4]*, Lixiazi He[1,2], Anne Bergeron[5], Vincent-Philippe Lavallée[6], Sébastien Lemieux[7], Patrick Gendron[7], Christian Rohde[1,2], Jianglong Xia[1,2], Prarabdha Jagdhane[1,2], Carsten Müller-Tidow[1,2], Daniel B. Lipka[8], Suzan Imren[9], R. Keith Humphries[10], Claudia Waskow[11], Binje Vick[12,13], Irmela Jeremias[12,13], Guillaume Richard-Carpentier[14], Josée Hébert[14,15,16], Guy Sauvageau[6,15,16], Judith Zaugg[2,3+], Fréderic Barabé[5,17+], and Caroline Pabst[1,2+**]

[1]Department of Medicine V, Hematology, Oncology and Rheumatology, University Hospital Heidelberg, Heidelberg, Germany;
[2]Molecular Medicine Partnership Unit (MMPU), University of Heidelberg and European Molecular Biology Laboratory (EMBL), Heidelberg, Germany;
[3]European Molecular Biology Laboratory (EMBL), Heidelberg, Germany;
[4]Department of Biochemistry and Molecular Biology, Complutense University of Madrid, 28040 Madrid, Spain;
[5]Centre de recherche du CHU de Québec, Centre de recherche en infectiologie du CHUL, Quebec City, Quebec, Canada;
[6]Laboratory of molecular genetics of stem cells, Institute for Research in Immunology and Cancer, University of Montreal, Montreal, Quebec, Canada;
[7]Department of Computer Science and Operations Research, University of Montreal, Montreal, Quebec, Canada;
[8]Regulation of Cellular Differentiation Group, Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany;
[9]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
[10]Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada;
[11]Regeneration in Hematopoiesis, Leibniz-Institute on Aging – Fritz-Lipmann-Institute (FLI), Jena, Germany;
[12]Research Unit Apoptosis in Hematopoietic Stem Cells, Helmholtz Zentrum München, German Research Center for Environmental Health (HMGU), Munich, Germany;
[13]German Cancer Consortium (DKTK), partner site Munich, Germany;

[14]The Quebec Leukemia Cell Bank, Research Centre, Maisonneuve-Rosemont Hospital, Montréal, Canada;
[15]Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal, Canada;
[16]Division of Hematology-Oncology, Maisonneuve-Rosemont Hospital, Montreal, Quebec, Canada;
[17]Department of Medicine, Université Laval, Quebec City, Quebec, Canada

*Authors contributed equally to this work
[+]Co-corresponding authors
**Correspondence:
Caroline.Pabst@med.uni-heidelberg.de

Running Title

HLF maintains slowly cycling LSCs in triple AML

Key points

1. *HLF* is overexpressed in normal karyotype AML triple-mutated for NPM1, DNMT3A, and FLT3-ITD.
2. Loss of *HLF* reduces the CD34+GPR56+ compartment, accelerates cell cycle progression and decreases *HES1* and *CDKN1C* expression.

Abstract (196 words), Main Text (3,992), 6 Figures, 8 Supplemental Figures, 12 Supplemental Tables

Abstract

FLT3, DNMT3A, and *NPM1* are the most frequently mutated genes in cytogenetically normal acute myeloid leukemia (AML), but little is known about how these mutations synergize upon co-occurrence. Here we show that triple-mutated AML is characterized by high leukemia stem cell (LSC) frequency, an aberrant leukemia specific *GPR56*^{high}CD34^{low} immunophenotype, and synergistic upregulation of Hepatic Leukemia Factor *(HLF)*. Cell sorting based on the LSC marker GPR56 allowed isolation of triple-mutated from *DNMT3A/NPM1* double-mutated subclones. Moreover, in *DNMT3A* R882 mutated patients, CpG hypomethylation at the *HLF* transcription start site correlated with high *HLF* mRNA expression, which was itself associated with poor survival. Loss of *HLF*

via CRISPR/Cas9 significantly reduced the CD34+GPR56+ LSC compartment of primary human triple-mutated AML cells in serial xenotransplantation assays. *HLF* knockout cells were more actively cycling when freshly harvested from mice, but rapidly exhausted when re-introduced in culture. RNA-sequencing (RNA-Seq) of primary human triple-mutated AML cells after shRNA mediated *HLF* knockdown revealed the NOTCH target Hairy And Enhancer Of Split 1 (*HES1*) and the cyclin-dependent kinase inhibitor *CDKN1C/p57* as novel targets of HLF potentially mediating these effects. Overall our data establish *HLF* as a novel LSC regulator in this genetically defined high-risk AML subgroup.

Introduction

For a considerable number of patients suffering from Acute Myeloid Leukemia (AML) current anti-leukemic therapies fail to permanently eradicate the disease. Hence allogeneic stem cell transplantation (SCT) often remains the only curative approach, but is itself associated with high treatment-related morbidity and mortality[1]. Numerous efforts have been taken to precisely predict therapy outcome and established a clear association between the cytogenetic background of the disease and prognosis[2]. Patients with no genetic aberrations detectable by standard cytogenetics (cytogenetically normal (CN)-AML represent approximately 40% of AML patients[2]. Knowledge about molecular genetic aberrations has therefore become crucial in these patients not only for our understanding of the underlying pathomechanisms, but also for risk stratification and therapy decisions[3]. Among the most intensively studied molecular aberrations are *DNMT3A*[4], *NPM1,* and *FLT3*-ITD mutations[5-7]. Co-occurrence of these three mutations is more frequent than can be explained by chance[8] and is associated with typical clinical features such as significantly higher white blood cell counts and prevalence in young women[9]. In addition, recent studies revealed genetic interaction of these three mutations in mice[10] and humans[11] further suggesting that triple-mutated AML represents a distinct entity with very poor outcome. At the same time, little is known about the molecular pathways driving leukemia and chemoresistance in these patients, e.g. it is unknown whether the three mutations have only additive effects or whether novel and specific pathways are induced in a synergistic way. Here we set out to determine genetic, phenotypic, and transcriptomic characteristics of triple-mutated AML and identified *HLF* as a specific target gene in this high-risk AML group.

Materials and Methods

Patient and cord blood samples

Peripheral blood and bone marrow specimens were collected from adult AML patients after obtaining written informed consent in accordance with the Declaration of Helsinki. Approval of the project was obtained from the Research Ethics Boards of the Medical Faculties of Martin-Luther University Halle (Saale), Heidelberg University, of Maisonneuve-Rosemont Hospital, CHU de Québec and University of Montreal, and of the Klinikum der Ludwig-Maximilians-Universität, Munich. Cord blood (CB) units of healthy infants were collected after obtaining written informed consent at the Department of Obstetrics at University Hospitals Halle (Saale) and Heidelberg following procedures that were approved by the Research Ethics Board of the Medical Faculty of Martin Luther University Halle-Wittenberg and the Medical Faculty of Heidelberg University. MN1/ND13 engineered cord blood CD34+ cells were generated as described[12]. Patient-derived xenografted (PDX) AML-491 cells were generated by serial transplantation of primary patient leukemia cells in NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice as described previously[13].

CRISPR/Cas9

Guide RNAs against *HLF*, *CD45*, and green fluorescent protein (*GFP*) were designed and purchased from Synthego. Sequences for guides are as follows: HLF#sg1: UUUGCUGGCAACAGCUGACC, HLF#sg2: CAAUGGGACUUGGUGUAUUG, sgCD45: GGUGCUGGUGUUGGGCGCAC, sgGFP: GGGCGAGGAGCUGUUCACCG.

Xenotransplantation

NOD.Rag1-; γcnull-SGM3 (NRGS) mice, which produce the three human cytokines SCF, GM-CSF, and IL3, were purchased from Jackson Laboratories. NOD.Cg-*Kit*$^{W\text{-}41J}$ *Prkdc*^{scid} *Il2rg*^{tm1Wjl}/WaskJ (NSGW41) mice carrying a homozygous *Kit* mutation were generated as described[14]. Mice were bred in specific pathogen free animal facilities at Martin-Luther University Halle (Saale) and German Cancer Research Center (DKFZ), Heidelberg. Animal experiments were approved and performed in accordance with all regulatory guidelines of the official committees (Landesverwaltungsamt Sachsen-Anhalt and Regierungspräsidium Karlsruhe).

Publicly available datasets

Sample and patient characteristics from The Cancer Genome Atlas AML patient cohort (http://cancergenome.nih.gov) were obtained from published work[8]. Microarray and clinical data for Verhaak[15] and Metzeler datasets were available through www.leukemia-gene-atlas.org[16]. Survival data in **Figure S1B** were adopted from[17]. Whole genome bisulfite sequencing (WGBS) data of *DNMT3A* mutated and wild-type patients of the TCGA cohort were downloaded from[18]. *HLF* ChIP data were downloaded from GEO accession GSE69817[19].

Additional methods

Additional methods are provided in Supplemental Information.

Results

Triple-mutated AML samples are characterized by an aberrant GPR56^{high}CD34^{low} immunophenotype and high LSC frequency

To identify unique characteristics of triple-mutated AML we analyzed the mutational landscape of 65 AML samples at diagnosis, which had been subjected to RNA-sequencing (RNA-Seq) as part of the Leucegene Project[20-25]. We grouped specimens based on their mutational status of the three genes *FLT3*-ITD *(F)*, *NPM1 (N)*, and *DNMT3A (D)* into “single”, “double”, and “triple” mutated samples, and those not mutated in these three genes hereafter called “triple wild-type” (WT) (**Figure 1A, Supplemental Table 1**). The group of triple-mutated AML was further subdivided based on whether the mutation in *DNMT3A* was located at amino acid position 882 (R882H, R882C) or elsewhere including missense, point-non-sense, and frameshift mutations (**Supplemental Table 1**). To assess whether the LSC marker GPR56[26] could help identify triple-mutated AML patients, we immunophenotyped all 65 AML specimens with regards to CD45, CD34, and GPR56, and correlated the three distinct, previously defined[26] flow cytometry profiles with genetic groups. We found that the great majority of triple-mutated AML, and in particular all *DNMT3A* R882-mutated triple AML samples, displayed a distinct immunophenotype characterized by a predominant GPR56 positive (GPR56^{pos}) population of which a minor subpopulation was also CD34^{pos} (profile 2, **Figure 1A and 1B**). This GPR56^{high}CD34^{low} phenotype was not shared by normal freshly isolated or cultured hematopoietic stem and progenitor cells (HSPCs) (**Figure S1A**), which indicated that the GPR56^{high}CD34^{low} profile (profile 2) was aberrant and leukemia-

specific. Of note, co-occurrence of mutations in the *FLT3* tyrosine kinase domain (TKD) together with *NPM1* and *DNMT3A* mutation was not associated with profile 2. Overall, detection of the aberrant profile 2 implied an 84-fold increased probability that a patient was triple-mutated at diagnosis suggesting that addition of GPR56 to flow cytometry panels might help identify this genetic subgroup (**Figure 1C,** Chi-square test, $P < 0.0001$).

In line with flow cytometry data, *GPR56* mRNA levels were also highest in triple-mutated AML compared to the other groups. (**Figure 1D**, see **Supplemental Table 2** for *P*-values). As allelic burden of *FLT3*-ITD (ITD-load) has been shown to play a more important role in prognosis than its presence alone[27], we determined *FLT3*-ITD mutant to wild-type ratios and found that even within the *FLT3*-ITD mutated samples *GPR56* expression levels increased with mutant allele frequency (**Figure 1E**). In parallel, we observed that the ITD-load was significantly higher when *NPM1* and *DNMT3A* were co-mutated (**Figure 1F,** $P = 0.019$, unpaired t-test, **Supplemental Table 3**). These data suggested that co-mutations in *NPM1* and *DNMT3A* better support expansion of the *FLT3*-ITD clone and potentially facilitate loss of heterozygosity (LOH) compared to other mutations that co-occur with *FLT3*-ITD.

To analyze the association between triple-mutated AML and LSC content, we interrogated our database containing LSC frequencies of 56 AML samples[26] and found significantly higher numbers of LSC^{high} samples in triple-mutated AML compared to other samples. This suggested that the three mutations synergize to induce specific self-renewal programs that enhance LSC activity (**Figure 1G, Supplemental Table 4**).

Further evidence for such synergy came from survival analyses of CN-AML patients within the German AMLCG cohort[17], which revealed significant prognostic value of *FLT3*-ITD only when both, *NPM1* and *DNMT3A* were mutated ($P = 0.0086$). *FLT3*-ITD on its own or in combination with mutations in either *NPM1* or *DNMT3A* alone showed a trend towards poor prognosis, but had no significant prognostic value confirming previous reports[11] in an independent patient cohort (**Figure S1B**).

Overall, these observations suggest that mutations in *NPM1*, *DNMT3A*, and *FLT3*-ITD interact with each other to drive a specific, aggressive subtype of AML characterized by high LSC frequency, high expression of the LSC-associated marker GPR56, and high *FLT3*-ITD allelic burden.

Triple-mutated leukemic subclones can be isolated based on GPR56 protein expression

Given the strong association between *FLT3*-ITD load and *GPR56* expression, we hypothesized that GPR56 might distinguish the triple-mutated from the *DNMT3A, NPM1* double mutated clone even within triple-mutated samples. To test this, we randomly selected ten triple-mutated specimens, sorted GPR56 and CD34 positive and negative fractions, and performed RNA-Seq to obtain information on gene expression and genetic alterations (outlined in **Figure 2A**). Indeed, we observed divergent *FLT3*-ITD mutant allele frequencies in the GPR56 positive and negative fractions in five of the ten samples (**Figure 2B**) and validated these results by semi-quantitative PCR using genomic DNA for two samples (**Figure 2C**). The co-presence of heterozygous, *bona fide* AML causing mutations in the genes *NPM1, DNMT3A*, and either *IDH1* (R132H) or *IDH2* (R140Q) in all three sorted fractions confirmed that all fractions contained a similarly high proportion of leukemic cells with no major contamination by normal cells (**Figure 2D, Supplemental Table 5**). In three samples (07H042, 09H002, 09H083) we found subclonal *PTPN11* and *NRAS* mutations, which were exclusively present in the GPR56 negative fractions further demonstrating that GPR56 positive and negative fractions had undergone individual clonal evolution. The most striking example was sample 09H083, in which copy neutral LOH occurred on (parts of) chromosome 13 after acquisition of an *FLT3*-ITD mutation (**Figure 2E, Figure S2**). It lost the *PTPN11* mutated minor subclone at relapse, which was accompanied by the loss of the GPR56 negative population (**Figure 2F**). In conclusion, GPR56 expression characterized not only triple-mutated patients, but also the triple-mutated leukemic subclones with high ITD load within individual samples.

Transcriptome analysis of triple-mutated AML samples reveals specifically high expression of *HLF*

To investigate the molecular mechanisms underlying the pathology of triple-mutated AML, we analyzed RNA-Seq data of 137 CN-AML samples contained in the Leucegene cohort. A set of genes potentially driven by all three mutations was pre-selected by requiring their expression to be most extreme in either the WT or the triple-mutated (DNF) group (n = 4025). These genes were subsequently grouped into ten clusters based on their expression profiles using an unsupervised clustering approach (partitioning around medoids; **Figures 3A-B**, **Supplemental Table 6**, see

Supplemental Information for details). Given that GPR56 positive and negative fractions distinguished the triple from the double mutated clone in several samples (**Figure 2**), we also performed differential gene expression analysis on RNA-Seq data of the sorted fractions. Indeed, we found great overlap of differentially expressed genes that distinguished triple from non-triple-mutated AML and GPR56 positive from negative fractions (**Figure 3A, 3B**).

To gain insight into potential pathways that might specifically be associated with triple-mutated AML, we performed Gene Ontology (GO) term enrichment analyses for the ten defined clusters (**Figures 3C, 3D,** full list of enriched GO terms provided in **Supplemental Table 7**). Cluster (CL)1 showed the most distinct triple-mutated AML defining expression profile for upregulated genes and, as expected, *GPR56* was part of it (**Figure 3A, 3B**). Moreover, it contained genes associated with cilium assembly and microtubule-based transport including several intraflagellar transport and Bardet-Biedl syndrome genes (**Supplemental Table 6,** see **Figure S3A-G** and **Supplemental Text** for additional information on transcriptome analysis).

As transcription factors (TFs) have been well described in orchestrating normal hematopoietic development[28], we next focused on the expression profiles of TFs in triple AML in comparison to the other genetic groups and to healthy blood cells. Among the six TFs significantly enriched in CL1 (**Figure 3E**) Hepatic leukemia factor (*HLF)* reached high expression levels almost exclusively in triple AML, whereas the other five TFs showed already slightly elevated expression levels in the “single” mutated groups (**Figure 3E, 3F, Figure S3C**). When analyzing co-expression patterns of these TFs in normal HSPCs we found that high concomitant *HLF* and *GLI2* expression was restricted to and therefore aberrant in AML, as it was neither observed in normal human CD34+ cord blood (CB), nor in CD34+CD33+ bone marrow myeloid progenitor cells (**Figure 3G**). In addition, high *HLF* expression and low expression of the CL10 genes *KLF12* (**Figure 3G**) and *HMGA2* (**Figure S3F**) was also an aberrant leukemia associated pattern, as normal HSPCs co-expressed these genes at high levels. In seven of eight triple-mutated AML samples, for which matched relapse samples were available, we observed higher *HLF* expression at relapse than at diagnosis (P=0.038, **Figure 3H**). The three samples that gained most at relapse though were *NPM1/DNMT3A* double mutated at diagnosis and had acquired an *FLT3*-ITD mutation at relapse (**Figure 3H**). In conclusion, our differential gene expression studies revealed *HLF* as one of the most triple-mutated AML defining genes.

Methylation levels at the *HLF* transcription start site correlate with *HLF* expression and *DNMT3A* mutational status

Missense mutations at position R882 in *DNMT3A* lead to reduced *de novo* DNA methyltransferase activity[29]. To investigate whether upregulation of *HLF,* e.g. by *NPM1c* and *FLT3*-ITD co-mutations, might be facilitated by loss of DNA methylation at the transcription start site, we compared DNA methylation levels at CpG sites along the *HLF* locus in AML samples with high and low *HLF* expression in the publicly available TCGA AML datasets[8]. In line with our hypothesis, we found that *HLF* mRNA expression anti-correlated with methylation levels at CpG sites close to the *HLF* transcription start site determined by 450K methylation arrays (**Figure 4A, Supplemental Table 8**). Moreover, methylation levels of CpGs close to the *HLF* transcription start site were significantly lower in *DNMT3A* R882 mutated versus *DNMT3A* wild type AML patients (**Figure 4B, Supplemental Table 9**). In line with 450K array data we also found decreased DNA methylation levels in *DNMT3A* mutated patients around the *HLF* transcription start site in publicly available WGBS data[18] (**Figure S4)**. Together, these data point towards a potential epigenetic regulation of *HLF* expression.

Loss of *HLF* reduces the CD34+GPR56+ LSC compartment *in vivo*

Before performing experiments with triple-mutated AML, we validated knockdown levels and impact of two small hairpin RNAs (shRNAs) targeting human *HLF* in CB CD34+ cells (**Figure 5A**). There was no significant difference in cell expansion or colony formation potential of lentivirally transduced fluorescent positive cells between knockdown (KD) and respective control cells using optimized culture conditions[30] (**Figures S5A, S5B**). While absolute cell counts were not affected, differentiation marker analysis during four-week *in vitro* culture revealed more rapid loss of the HSC-enriched CD34+CD45RA- population and accelerated myelo-monocytic differentiation upon *HLF* KD (**Figure 5B, Figure S5C**). Finally, *in vivo* experiments in NRGS mice showed significantly lower engraftment levels of *HLF* KD cells at all time points analyzed (**Figure 5C**). Of note, viability was not affected by *HLF* KD (**Figure S5D**).

To investigate the role of *HLF* in AML, we selected a triple-mutated sample with very high LSC frequency determined in previous studies (04H112)[26] and pursued two parallel approaches to reduce HLF expression (**Figure 5D**): one based on a plasmid-free CRISPR/Cas9 system and another based on lentiviral transduction of shRNAs. For both

approaches we used an optimized NSGW41 xenotransplantation model allowing engraftment of very low stem cell numbers without prior irradiation[14]. As primary AML cells rapidly change LSC activity *in vitro* despite optimized culture conditions[31], we injected triple-mutated AML cells directly after transfection with the most efficient sgRNA against HLF and Cas9 using sgRNA against GFP as negative control (**Figure S6A**). When harvesting cells 14 weeks after transplantation we determined cleavage efficiency on genomic DNA (**Figure S6B**) and confirmed loss of HLF protein by Western Blot in three out of five mice (**Figure 5E**).

Overall engraftment levels of human CD45+ cells were close to 100% (saturation) in all mice. In contrast, the CD34+GPR56+ compartment, which we previously showed to be highly enriched for LSCs[26], was significantly lower in mice, in which HLF protein was lost (hereafter called HLF knockout (KO) cells) compared to controls (**Figures 5F, 5G, Figures S6C, S6D**). At the same time, the double negative compartment, which we showed before to contain no or minimal LSC activity[26], was significantly increased in HLF KO cells compared to controls. When reintroduced in culture, HLF KO cells were not able to further expand in contrast to control cells (**Figure 5H**). After reinjecting high and low doses of HLF KO and control cells in secondary recipients, we detected human leukemic engraftment in all mice, but HLF KO cells reached high engraftment levels more rapidly (**Figure 5I**). While the CD34+GPR56+ fraction in control cells did not differ between primary and secondary transplantations, it was even more strongly reduced in secondary compared to primary mice transplanted with HLF KO cells (**Figures 5I, 5J, Figure S6E**). When re-introduced in culture, we observed again that HLF KO cells had significantly reduced proliferative capacity (**Figure 5K, Figure S6F**).

We noticed that some non-normal karyotype AML samples mimic the $GPR56^{high}CD34^{low}$ profile 2 of triple-mutated AML (**Figure S6G**). To test whether such samples respond in a similar way to HLF knockdown, we transduced AML-491 with shRNAs against HLF or shLuc control. We observed similar effects compared to those observed with triple-mutated AML 04H112 using the CRISPR/Cas9 system: HLF KD significantly reduced the CD34+GPR56+ compartment in primary and more severely in secondary recipients, and accelerated engraftment in secondary recipients (**Figure S6H-J**). Together, these results showed that loss of HLF impaired the function of $GPR56^{high}CD34^{low}$ AML cells.

Loss of HLF expression accelerates cell cycle progression and reduces the expression levels of *HES1* and *CDKN1C/p57*.

Given the distinct engraftment dynamics, we sought to further characterize cell cycle and proliferative properties of HLF KO and control cells. CFSE labelling of cells harvested from secondary recipients showed that sgHLF cells had divided more often when analyzed four days after experiment start (**Figure 6A, Figures S7A, S7B**). To determine, which phase of the cell cycle was most affected by loss of HLF, we performed EdU (5-ethynyl-2'-deoxyuridine) pulse-chase experiments (see also **Supplemental Methods**). Directly after the 90-minutes EdU pulse, we found a significantly higher fraction of EdU positive sgHLF cells compared to controls in the group injected with 10^5 (100K) cells (**Figure 6B, 6C**). This was accompanied by a significantly lower fraction of sgHLF cells in G1. There were only slight differences between sgHLF and control cells from mice injected with $2x10^6$ (2M) cells when assessed directly after the 90-minutes pulse. Importantly, despite these initial differences, we observed similar phenotypes when comparing sgHLF to sgGFP cells 14 hours later in both, the 100K and 2M groups: a significantly higher fraction of cells had become EdU positive accompanied by a reduction of the G1 fraction (**Figures 6B, 6C)**. Moreover, the fractions of G2/M cells identified by DAPI staining were significantly higher in sgHLF compared to control cells at t14h. A significant increase in S phase was also observed for cells that had not incorporated EdU ruling out an impact of EdU incorporation on the results (**Figure S7C**). We did not detect any significant difference in the fractions of quiescent, pyronin negative, G0 cells (**Figure S7D**).

To test whether loss of HLF sensitized the cells towards drug treatment, we exposed them to high concentrations of the anti-leukemia drugs AraC (50nM), Daunorubicin (50nM), and Etoposide (2µM) and observed significantly higher sensitivity of sgHLF versus control cells (**Figure 6D**). To corroborate our findings, we performed the same experiments again in the $GPR56^{high}CD34^{low}$ non-CN AML-491 and observed similar results in EdU, CFSE labelling, and drug treatment experiments (**Figure S7E-H**).

To gain insight into the potential mechanism, we had simultaneously performed RNA-Seq with the same triple-mutated AML sample 04H112 (**Figure 5D, Figure S7I**). RNA-Seq revealed downregulation of *HES1* and *CDKN1C* in HLF KD cells (**Figure 6E, Supplemental Table 10**). To validate these findings, we performed qPCR with CRISPR/Cas9 transfected cells from primary and secondary transplantations confirming significantly decreased expression of *HES1* and *CDKN1C* in HLF KO cells (**Figure 6F**).

To further support these data, we performed qPCR with a model leukemia (MN1/ND13) generated through overexpression of the two oncogenes MN1 and NUP98-HOXD13 in cord blood CD34+ cells[12] and observed a similar decrease in *HES1* and *CDKN1C* expression upon HLF KD indicating that the effect of HLF KD on *HES1* and *CDKN1C* expression was not specific to the triple-mutated AML sample 04H112 (**Figure 6G**). Similar effects were observed with AML-491, in which Pearson correlation analysis revealed a highly significant correlation between *HLF* and *CDKN1C* expression (**Figure S7J**). *HES1* expression was also decreased in healthy CD34+ cells five days after transduction with shRNAs against HLF (**Figure 6G**). In line with these findings, publicly available ChIP-Seq data obtained from murine HSPCs suggested direct binding of Hlf to the *Hes1* transcription start site (**Figure S7K**).

Having confirmed a functional role for *HLF* in high-risk triple-mutated AML, we questioned whether *HLF* expression itself was associated with patient outcome. We found that high *HLF* expression was significantly associated with poor overall survival (OS) and relapse-free survival (RFS) in the Leucegene dataset from time of diagnosis and after allogeneic stem cell transplantation (**Figure 6H, Figure S8A**). Similar results were found with the Verhaak and AMLCG datasets (www.leukemia-gene-atlas.org, **Figure S8B**). High *HLF* expression remained a prognostic factor of poor OS in multivariate analyses including age, white blood cell (WBC) counts and cytogenetic risk (**Supplemental Table 11, 12**). In summary, we propose that HLF plays a crucial role for triple-mutated AML by maintaining the CD34+GPR56+ compartment and slowing down cell cycle progression, and identified *HES1* and *CDKN1C* as novel *HLF* downstream targets possibly mediating these effects (**Figure 6I**).

Discussion

Here we determined transcriptomic and immunophenotypic characteristics of CN-AML triple-mutated for *NPM1*, *DNMT3A*, and *FLT3*-ITD. We found that presence of all three mutations was highly associated with an aberrant immunophenotype defined by high expression of the LSC marker *GPR56*[26], which itself correlated with *FLT3*-ITD allelic burden and allowed separation of triple and double mutated subclones within the same sample. Moreover, we performed functional *in vitro* and *in vivo* experiments and identified a key role for *HLF* in triple-mutated AML.

Among all TFs that were overexpressed in triple-mutated AML, *HLF* was clearly different from the others, as it was the only TF for which we observed no expression in “single”

DNMT3A, *NPM1*, or *FLT3*-ITD mutated AML samples pointing towards synergistic interaction of the three mutations. Furthermore, we previously found that *Hlf* was important for proliferation in an *Mn1* induced mouse leukemia model[32] supporting the rationale for studying *HLF* in AML. Recapitulation of the ancestral disease including maintenance of the stem cell compartment is a hallmark of self-renewing LSCs[33,34]. Here we found that HLF KO cells were not able to maintain the CD34+GPR56+ compartment, which we showed before to be highly enriched for LSCs[26] indicating that HLF was required for propagating the ancestral disease in serial transplantations.

Hlf was identified as a key regulator of stem cell activity in murine HSCs by maintaining quiescent HSCs in G0 and preventing them from entering cell cylce[19,35]. While we did not observe a difference in the G0 fraction between HLF KO versus control AML cells when harvested from mice at late time points, we cannot rule out that a difference was present in the early engrafting LSCs. Major alterations in cell cycle dynamics were however also detectable when harvesting cells at saturation. EdU pulse-chase and CFSE labelling experiments revealed that loss of HLF strongly enhanced cycling of triple-mutated AML cells. Interestingly, these differences were only visible with cells freshly harvested from mice, as HLF KO cells rapidly lost their expansion potential upon *in vitro* culture suggesting that the proliferative capacity of HLF KO cells was dependent on supporting niche factors.

RNA-Seq studies revealed the transcription factor *HES1* and the cyclin dependent kinase (CDK) inhibitor *CDKN1C/p57* as HLF target genes. Both, HES1 through upregulation of p21[36], and CDKN1C/p57 together with p27[37] have been shown to induce quiescence of normal CD34+ HSCs. *CDKN1C* also inhibited proliferation of different cancer cell types[38,39]. Besides its role in inducing cell cycle exit[37], *CDKN1C* seems to also affect other cell cycle phases, e.g. induction of *CDKN1C* in Jurkat cells reduced cyclin E and cyclin A activities, as well as the fraction of cells in S phase[40], which is in line with our EdU experiments, in which loss of HLF accelerated S phase entry.

While the known function of these target genes is strongly suggestive for a functional role in the HLF mediated phenotypes, more studies are needed to dissect the mechanisms downstream of HLF.

HLF had been identified in paediatric acute lymphoblastic leukemia (ALL) as a fusion partner of *E2A* causing a treatment resistant disease, while ALL induced by fusion of *E2A* with *PBX1* did not[41], suggesting that *HLF* is the key driver of chemoresistance in *E2A-HLF* mutated ALL. In line, we observed that HLF KO cells were more sensitive to *in*

vitro treatment with anti-leukemic drugs and found that high *HLF* expression itself correlated with poor survival of AML patients.

Together, our data establish *HLF* as a crucial transcription factor in triple-mutated AML, which modulates cell cycle dynamics and maintains the CD34+GPR56+ LSC-enriched compartment in this high-risk genetic subgroup.

Accession codes

Sequencing data are available through accession numbers GSE49642, GSE52656, GSE62190, GSE66917, GSE67039, GSE48843, GSE48846, GSE51984, GSE68623, GSE129094, and GSE128848. HLF ChIP data was downloaded from GSE69817

Acknowledgments

We acknowledge D. Gagné and V. Eckstein for technical support with cell sorting. Clinical specimens were collected and analyzed by the Banque de Cellules Leucémiques du Québec (BCLQ), supported by the Cancer Research Network of the Fonds de Recherche du Québec en Santé (FRQS). We thank B. Meier for performing *FLT3*-ITD ratio analyses. We thank T. Herold and K. Metzeler for providing AMLCG cohort data for survival analyses. This work was supported by the Government of Canada through Genome Canada and the Ministère de l'économie, de l'innovation et de l'exportation du Québec through Génome Québec. G.S. holds a Canada Research Chair in Molecular Genetics of Stem Cells, and J.H. holds a Research Chair in Leukemia supported by Industrielle-Alliance (Université de Montréal). A.R.P has been recipient of a postdoctoral fellowship granted by Fundación Ramón Areces and the Research Program "Atracción de Talento de la Comunidad de Madrid" (2017-T2/BMD-5532). This work was further supported by Deutsche Forschungsgemeinschaft (DFG PA2815/1-1), by a Max-Eder-Grant of the German Cancer Aid (70111531), and by the Leukemia and Lymphoma Society of Canada.

Declaration of Interests

The authors declare no competing financial interests.

Authorship

S.G. performed experiments, analyzed data, generated figures, and wrote the manuscript. A.R.P. performed computational analyses, generated figures, and wrote the manuscript. S.G. and A.R.P. contributed equally to this work. L.H. contributed to *in vivo*

experiments and edited the manuscript, A.B. performed sorting experiments, V.P.L., S.L., and P.G. performed RNA-Sequencing and mutational analyses of the Leucegene cohort samples, P.J. and J.X. helped perform *in vivo* experiments and lentiviral transductions, C.R. and C.M.T. performed and analyzed 450K arrays on AML samples, D.B.L. analyzed 450K arrays, S.I. and R.K.H. generated model leukemia cells, C.W. provided NSGW41 mice and supported *in vivo* experiments, B.V. and I.J. generated and characterized PDX AML cells, G.R.C. performed the survival analyses in the Leucegene cohort, J.H. and G.S. provided AML samples, RNA-Seq and clinical data of the Leucegene sample cohort, J.Z. performed bioinformatics analyses, revised the manuscript and co-supervised the project, F.B. and C.P. co-directed the project and wrote the manuscript.

Correspondence

Correspondence and material requests should be addressed to Caroline.Pabst@med.uni-heidelberg.de

Figure Legends

Figure 1. Triple-mutated AML samples are characterized by an aberrant GPR56^{high}CD34^{low} immunophenotype and high LSC frequency.

A) Mutational landscape of 65 AML samples at first diagnosis. Blue box indicates mutated, light grey box indicates non-mutated. Color coded bar on top shows grouping of samples according to their mutational status for *FLT3*-ITD, *NPM1*, and *DNMT3A* into “single” (n=16, orange group), “double” (n=14, green group), and “triple” mutated samples (n=22) split into those with a typical *DNMT3A* mutation at position R882 (n=13, black), and those with other *DNMT3A* mutations (n=9, dark grey). Samples not mutated for the three genes are comprised in the blue group (n=13). Color coded bar at the bottom indicates the immunophenotypic profile with regards to CD34 and GPR56 percentage determined by flow cytometry: profile 1 defined as CD34 % > GPR56 %, profile 2 as GPR56 % > CD34 %, and profile 3 comprises all CD34 negative samples (CD34 < 1%).

B) Representative examples of the three different immunophenotypic profiles as defined in A.

C) Numbers of triple mutations in patients with profile 2 versus other groups. Detection of the aberrant profile 2 at diagnosis implies an 84-fold greater chance to simultaneously harbor mutations in *FLT3*-ITD, *NPM1*, and *DNMT3A* (Fisher´s exact test, $P < 0.0001$).

D) GPR56 mRNA expression in AML samples at diagnosis in different genetic groups. Box and whiskers plot (Tukey) showing **R**eads **p**er **k**ilobase per **m**illion mapped reads (RPKM, transformed as log10(RPKM+0.001)) values for *GPR56* mRNA based on RNA-Sequencing data in genetic groups (n = 388). **WT**: not mutated for NPM1 (N) or DNMT3A (D) and no FLT3-ITD (F), **D**: D mutated, *NPM1* wt, no *FLT3*-ITD, **N**: *NPM1* mutated, no F nor D mutation, **ND**: N and D mutation, no *FLT3*-ITD, **DF**: D and F mutation, *NPM1* wt, **F**: *FLT3*-ITD, D and N not mutated, **NF**: N and F mutation, D wt, **DNF**: triple-mutated. Median RPKM for *GPR56*: 3.1 (WT), 6.6 (F), 12.47 (NF), 31.19 (DNF). *P*-values provided in **Supplemental** Table 2.

E) *Left*: Box and whiskers plot (Tukey) showing *GPR56* mRNA expression levels (transformed as log10(RPKM+0.001) in *FLT3*-ITD mutated patients with a mutant allele frequency > 0.5 (n = 34) and < 0.5 (n = 87). Medians were 9.25 versus 33.79, $P = 0.0005$, Mann Whitney U. *Right*: *FLT3*-ITD mutant allele frequencies in samples with *GPR56* mRNA expression above (n = 60) or below or equal to (n = 61) the median (RPKM 16). Allelic frequencies were 0.38 versus 0.55, $P < 0.0001$, unpaired t-test.

F) *FLT3*-ITD allelic frequencies in AML samples at diagnosis. Symbols represent individual samples, bars show average mutant allele frequencies in *FLT3*-ITD mutated patients with no mutation in *NPM1* or *DNMT3A* (F, n = 34), with *DNMT3A* mutation and *NPM1* wt (DF, n = 7), with *NPM1*, but no *DNMT3A* mutation (NF, n = 26), with *NPM1* and *DNMT3A* mutation (DNF, n = 42). Average frequencies were 0.36 (F), 0.42 (DF), 0.41 (NF), and 0.48 (DNF), $P = 0.018$ for F versus DNF, unpaired t-test. Of note, 24% of triple-mutated samples had *FLT3*-ITD allele frequencies above 0.5, suggesting that loss of heterozygosity (LOH) had occurred, versus only 8% in *FLT3*-ITD with non-mutated *NPM1* and *DNMT3A.*

G) Numbers of samples with high LSC frequency in triple-mutated AML versus other groups. 5 of 11 triple-mutated AML samples were categorized as “LSChigh” defined by an LSC frequency greater than 1:30,000 cells compared to only 2 out of 27 in the remaining samples (adapted from[26], $P = 0.0077$, Odds ratio 6.67, Chi-Square test).

Figure 2. Triple-mutated leukemic subclones can be distinguished from double mutated clones based on GPR56 expression.

A) Schematic overview of the sorting strategy of 10 triple-mutated AML samples. +/+ $CD34^{+}GPR56^{+}$, -/+ $CD34^{-}GPR56^{+}$, -/- $CD34^{-}GPR56^{-}$.

B) *FLT3*-ITD mutant allele frequencies determined by kmer approach in RNA-Seq data obtained from CD34+GPR56+ and CD34-GPR56- fractions. Divergent *FLT3*-ITD frequencies were found in the sorted fractions of samples marked in red, while no difference was found in samples marked in blue. Of note, in the five samples, in which *FLT3*-ITD ratios were not divergent in the sorted fractions, *FLT3*-ITD was close to 50% allele frequency i.e. was not subclonal in 12H007, 10H166, 09H043, while in two samples (14H007, 10H101) *FLT3*-ITD was below 50% in all fractions.

C) Confirmation of divergent *FLT3* mutant/wild type ratios in samples 07H042 and 07H062 by PCR using genomic DNA isolated from CD34+GPR56+ (+/+), CD34-GPR56+ (-/+), and CD34-GPR56- (-/-) sorted fractions (*upper panel*: bars indicating *FLT3* mutant/wild type ratios, *lower panel*: agarose gel showing FLT3 wild type band at 325 bp, +93 bp ITD in sample 07H062, and +54 bp ITD in 07H042).

D) Variant allele frequencies (VAF) of mutations with known leukaemogenic potential in GPR56 positive and negative fractions shown for five AML samples with divergent *FLT3*-ITD load. Mutations with VAF close to the diagonal line indicate no difference between the sorted fractions. See **Supplemental Table 5** for information on CD34-GPR56+ fractions.

E) Detailed VAF analysis in GPR56 positive and negative sorted fractions for sample 09H083.

F) *Left*: Comparison of VAF in unsorted sample 09H083 compared to the corresponding unsorted relapse sample 10H068. *Right*: FACS plots showing CD34 and GPR56 protein expression in sample 09H083 at diagnosis and in the corresponding relapse sample. The loss of the *PTPN11* clone at relapse is accompanied by the loss of the GPR56 negative population.

Figure 3. Differential gene expression in triple-mutated AML.

A) Heatmap of average normalized gene expression (Z-score) for each genetic group in the AML dataset. Splits separate the different-gene clusters and genes are sorted from high (top) to low (bottom) gene expression in triple-mutated AML. Only genes with an

average expression of library-normalized raw-read counts > 30 are shown. Letter code as defined in Figure 1D.
B) Heatmap of average Z-score normalized in GPR56 sorted fractions. Genes are sorted as in A. PP: GPR56+CD34+, MP: GPR56-CD34+, MM: GPR56-CD34-.
C) Enrichment analysis of biological processes for AML clusters. The background is defined using genes that are expressed and annotated to any ontology term. *P*-values and odd-ratios were calculated using Fisher's exact test, multiple-test correction using Benjamini and Hochberg method was applied to nominal *P*-values.
D) Box and whiskers plots of Z-score normalized gene expression for genes in clusters showing a gradual synergistic pattern compared to triple-mutated AML samples (i.e. CL 1-3 and CL 8-10) for each of the eight genetic groups.
E) Heatmaps of the expression profile for transcription factors (rows) in the selected clusters shown in D in each genetic group (columns).
F) Normalized read counts for *HLF*, *GLI2,* and *KLF12* in the defined, eight genetic groups (left) and in the sorted fractions (right). Letter code defined in Figures 1E and 3B.
G) Combinatorial scatter plot showing gene expression (log(RPKM+0.001)) of *HLF* and *GLI2* (left) and of *HLF* in combination with *KLF12* (right) in AML and normal CD34+ populations. Symbols represent individual samples.
H) *HLF* mRNA expression (RPKM) in eight matched diagnosis-relapse samples. Five pairs were triple-mutated at diagnosis and relapse (DNF), while three samples were *NPM1* and *DNMT3A* double mutated and gained an *FLT3*-ITD mutation at relapse (DN/DNF). Numbers indicate matched sample IDs.

Figure 4. Methylation levels at the transcription start site correlate with *HLF* expression and *DNMT3A* mutational status.
A) *Upper panel:* Schematic overview of the localisation of CpGs assessed by the Illumina 450K methylation array in the *HLF* gene (upper row shows entire *HLF* gene with exons and introns. Lower row shows enlarged regions assessed for CpG methylation; CpG positions are indicated by triangles, numbered black bars indicate gaps not shown in the enlarged regions). B) *Lower panel:* Methylation levels at indicated CpGs according to *HLF* mRNA expression levels in the TCGA patient cohort. Patients were grouped according to *HLF* expression (RPKM) into quartiles (n=156, box plots according to Tukey method, see Supplemental Table 8 for details). Only *P*-values < 0.005 are shown. ** *P* <

0.005, *** $P < 0.0005$. Correl r: Pearson correlation coefficients for direct correlation of methylation and *HLF* mRNA expression levels.

B) Methylation levels in *DNMT3A* R882 mutated (red, n = 27) and wild type patients (blue, n = 145) from the TCGA patient cohort. *DNMT3A* mutations other than R882 were excluded from the analysis given their undetermined impact on DNA methylation levels. Positions marked in red indicate CpGs with significant differences in methylation level between the two groups. *P*-values were determined by Mann Whitney U test and were Benjamini-Hochberg corrected. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$

Figure 5. Loss of HLF reduces the CD34+GPR56+ LSC compartment *in vivo*.

A) Knockdown level of *HLF* mRNA in CD34+ cord blood cells with two different shRNAs compared to shRNA against luciferase (shLuc) determined by q-PCR. *HLF* expression as percent of GAPDH expression was normalized to shLuc controls. Values indicate means (shHLF.3441, mean fraction (range) of shLuc expression 0.4 (0.2-0.67), shHLF.630 mean (range) fraction of shLuc expression 0.06 (0.03-0.12)). Bars and error bars represent means and standard deviation of three individual CD34+ cord blood infections.

B) *Left:* FACS plots showing differentiation of cord blood CD34+ cells 5 and 27 days after infection with shRNA.3441 against *HLF* or luciferase. Shown is one out of four replicates derived from two independent experiments. Protein expression of CD34 and CD45RA were tracked during a period of 27 days. Values indicate percentages. *Right*: Box plots showing fractions of CD34+CD45RA- cells on day 5 (upper panel, median percentage 30.85% versus 25.65%, $P = 0.02$) and CD34+CD45RA+ cells on day 27 (lower panel, 20.80% versus 7.2%, $P = 0.02$) after infection with shHLF.3441 or shLuc (four replicates of two independent infections, Mann Whitney U test, * $P < 0.05$).

C) Engraftment of cord blood CD34+ cells in NRGS mice after infection with shRNAs against *HLF* or luciferase using tagRFP as fluorescent marker. Horizontal lines indicate means, symbols represent individual mice. Shown are the percentages of human CD45+tagRFP+ cells in mouse bone marrow. At week 4 and 12, bone marrow was collected by aspiration from one femur, while bones from tibia, femur, pelvis, and spine were analyzed after sacrificing the mice in week 17. (Mean engraftment levels at week 4: 24.8% vs. 59.5%, n=6 per group, $P=0.01$; week 12: 0.6% vs. 10.1%, n=4 (shHLF group, no aspiration material for 2 mice), n=5 (shLuc group, no aspiration material for 1 mouse),

627 *P* = 0.046; week 17: 2.9% vs. 11%, n=5 (shHLF, one mouse died before week 17), n=6
628 (shLuc), *P* = 0.029).
629 D) Cartoon illustrating experimental setup of *in vitro* and *in vivo* experiments. See text
630 and methods for details. AM: Ametrine, + positive, BM: bone marrow, i.v.: intravenous,
631 M: million, K: thousand
632 E) Western Blot showing human HLF protein expression in triple-mutated primary
633 human AML cells (AML#04H112) harvested from mice 14 weeks after injection of cells
634 electroporated with either sgRNA against HLF (sgHLF) or GFP (sgGFP, negative
635 control) and Cas9 recombinant protein. GAPDH was used as loading control. Numbers
636 below Western Blot indicate the knockout (KO) efficiency (%) for HLF determined by
637 Sanger sequencing on the genomic DNA from the same cells (see Figure S6B for gDNA
638 results). Mouse IDs in bold indicate mice used for secondary transplantations.
639 F) Engraftment levels of total human CD45+ cells (left, mean percentages from left to
640 right 99%, 98%, 99%), and fractions of CD34+GPR56+ (middle, mean percentages from
641 left to right 15.80%, 38.5%, 34.9%), and CD34-GPR56- (right, mean percentages from
642 left to right 36%, 14%, 12%) cells of human CD45+ cells in primary recipient mice.
643 Shown are individual mice and means, * *P*-value < 0.05, ** *P*-value < 0.05, unpaired t-
644 test.
645 G) Representative FACS plots showing CD34 and GPR56 expression in sgGFP cells
646 (left, HLF protein not lost), sgHLF cells with confirmed loss of HLF protein (middle), and
647 those with sgHLF, but no loss of HLF protein. (See Figure S6 for complete data). Mice
648 #247 and #198 were used for secondary transplantations. Values indicate percentages.
649 # numbers indicate mouse IDs.
650 H) Proliferation curves for five sgGFP samples (blue) and three sgHLF samples with
651 confirmed loss of HLF protein (red). Shown is fold-increase in absolute cell counts per
652 well normalized to the start date of the culture. From each bone marrow sample six
653 replicate cultures were started. The average cell counts of the six cultures of each
654 sample were used to compare the five sgGFP versus the three sgHLF samples (mean
655 fold-change on day 3 was 2.3 vs. 1 and on day 5 was 3.2 vs 0.8). Cells were counted by
656 HTS-FACS. * *P* < 0.05, ** *P* < 0.005, unpaired t-test.
657 I) Overall human engraftment levels and CD34 and GPR56 surface expression in
658 secondary recipients injected with $2x10^6$ sgGFP (#246) or sgHLF (#198) cells. Shown
659 are means and individual values for week 4 and 7 bone marrow aspirates and final bone
660 marrow analysis *post mortem* in week 8 for total human CD45+ levels (left, mean

percentages week 4: 4.8% vs. 29%, week 7: 87% vs. 98%, final: 96% vs. 99%), CD34+GPR56+ fractions of human CD45+ cells (middle, mean percentages week 4: 13% vs. 8%, week 7: 26% vs. 0.7%, final: 37% vs. 1.6%), and CD34-GPR56- fractions of human CD45+ cells (right, mean percentages week 4: 9% vs. 18%, week 7: 6% vs. 45%, final: 5.5% vs. 34%). * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$, unpaired t-test. The third sgGFP mouse died prior to final bone marrow analysis so that data for this mouse are only available from week 4 and 7.

J) Representative FACS profiles showing CD34 and GPR56 expression of human sgGFP and sgHLF cells engrafted in secondary recipients eight weeks after injection of $2x10^6$ cells. Numbers in quadrants indicate percentages.

K) Proliferation curves of sgHLF and sgGFP (control) cells *in vitro* after harvest from secondary recipients eight weeks after injection of $2x10^6$ (2M) cells. Six cultures were started in 96- well plate formats from each mouse. Average cell counts of the six cultures per sample were used to compare the groups and were normalized to the start date of the culture. Cells were counted by HTS-FACS. * $P < 0.05$, ** $P < 0.005$, unpaired t-test.

Figure 6. Loss of HLF expression accelerates G1/S transition and reduces the expression levels of *HES1* and *CDKN1C*.

A) CFSE experiment performed with sgGFP and sgHLF cells harvested from secondary recipients. *Upper panel*: fractions of cells in generations 3-6 four days after incubation with CFSE. Shown are results of mice injected with $2x10^6$ sgGFP (n=2 mice, 3 cultures per sample) or $2x10^6$ sgHLF cells (n=3 mice, 3 cultures per sample). * $P < 0.05$, Gen: generation. *Lower panel*: Representative histograms showing distribution of generations four days after experiment start. Blue bar indicates starting CFSE intensity, green bar indicates background fluorescence intensity. Numbers indicate mouse IDs. Data from recipients of 10^5 cells available in Figures S7A, S7B.

B) Representative FACS plots of EdU experiment performed with sgGFP and sgHLF cells harvested from secondary recipients injected with 10^5 cells. Shown is distribution of cells in different cell cycle phases after 90 minutes pulse with EdU (t0) and 14 hours later (t14h). Numbers indicate percentages of total. DAPI was used to determine DNA content, EdU was detected in FITC channel.

C) EdU experiment performed with cells harvested from secondary recipients injected with either $2x10^6$ cells (sgGFP n=2, sgHLF n=3) or 10^5 cells (sgGFP n=3, sgHLF n=3). Shown are mean percentages (and standard deviations) of cells in the indicated cell

cycle phases after 90 minutes EdU pulse (above) and after 14 hours (below) out of all viable cells. M: million, K: thousand, * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$.

D) Compound sensitivity testing. Cells harvested from secondary recipients injected with either $2x10^6$ cells (sgGFP n=4 cultures from 2 different mice, sgHLF n=6 cultures derived from 3 different mice) or 10^5 cells (n=6 cultures derived from 3 different mice per group) were exposed to 50nM of AraC, 50nM of daunorubicin, or 2µM of etoposide, compared to DMSO (vehicle) only for 5 days. Viable cell counts on day 5 were normalized to cell counts in DMSO. * $P < 0.05$, *** $P < 0.0005$

E) *Left*: volcano plot showing log2-fold changes in mRNA expression (x-axis) and transformed *P* values (y-axis) for RNA-Seq data performed on triple-mutated cells after one round of *in vivo* expansion, followed by infection with shRNAs against HLF (shHLF.630, n=3) or shLuc control (n=3). Data points highlighted by colors represent genes with log2-fold change >1 (blue) or < -1 (red) and FDR<10%. Due to space constraints, not all gene symbols are displayed. See also Supplemental Table 10 for gene names. *Right*: Mean RPKM values for *HES1* (above) and *CDKN1C* (below), n=3 per group, * $P < 0.05$.

F) Expression levels in percent of *GAPDH* expression determined by qPCR for *HES1* and *CDKN1C* in sgHLF versus sgGFP cells from primary recipients (above, sgGFP: #247, #246 versus sgHLF: #197, #198, #199) and secondary recipients (below, sgGFP: #8013, #8014 versus sgHLF: #8248, #8249), * $P < 0.05$, ** $P < 0.005$.

G) *Above: HLF, CDKN1C* and *HES1* expression in percent of *GAPDH* expression determined by qPCR in the model leukemia MN1/ND13 upon lentiviral transduction with shRNA against *HLF* (sh.630) versus shLuc control after one round of *in vivo* expansion (n=3 recipients). *Below: HLF* and *HES1* expression in healthy cord blood CD34+ cells upon lentiviral transduction with two different shRNAs against *HLF* or shLuc control 5 days after infection normalized to shLuc. Numbers indicate mean fractions of shLuc.

H) Kaplan-Meier survival curves showing overall survival from time of diagnosis in patients from the Leucegene prognostic cohort with very high (4^{th} quartile, red line) versus low *HLF* expression (quartiles 1-3, blue line). Log-rank test.

I) Model indicating the proposed functional role of HLF in triple-mutated AML. HLF induces upregulation of *HES1* and *CDKN1C*, decelerates cell cyle progression, and maintains the CD34+GPR56+ compartment *in vivo*.

References

1. Robin M, Guardiola P, Dombret H, et al. Allogeneic bone marrow transplantation for acute myeloblastic leukaemia in remission: risk factors for long-term morbidity and mortality. *Bone Marrow Transplantation*. 2003;31(10):877–887.
2. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116(3):354–365.
3. Grossmann V, Schnittger S, Kohlmann A, et al. A novel hierarchical prognostic model of AML solely based on molecular mutations. *Blood*. 2012;120(15):2963–2972.
4. Im AP, Sehgal AR, Carroll MP, et al. DNMT3A and IDH mutations in acute myeloid leukemia and other myeloid malignancies: associations with prognosis and potential treatment strategies (vol 28, pg 1774, 2014). *Leukemia*. 2015;29(2):516–516.
5. Port M, Böttcher M, Thol F, et al. Prognostic significance of FLT3 internal tandem duplication, nucleophosmin 1, and CEBPA gene mutations for acute myeloid leukemia patients with normal karyotype and younger than 60 years: a systematic review and meta-analysis. *Ann Hematol*. 2014;93(8):1279–1286.
6. for the Acute Leukemia French Association (ALFA) and the Leucémies Aiguës Myéloblastiques de l'Enfant (LAME) Cooperative Groups, Boissel N, Leroy H, et al. Incidence and prognostic impact of c-Kit, FLT3 and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). *Leukemia*. 2006;20(6):965–970.
7. Wagner K, Damm F, Thol F, et al. FLT3-internal tandem duplication and age are the major prognostic factors in patients with relapsed acute myeloid leukemia with normal karyotype. *Haematologica*. 2011;96(5):681–686.
8. The Cancer Genome Atlas Research Network. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N Engl J Med*. 2013;368(22):2059–2074.
9. Loghavi S, Zuo Z, Ravandi F, et al. Clinical features of de novo acute myeloid leukemia with concurrent DNMT3A, FLT3 and NPM1 mutations. *Journal of Hematology & Oncology*. 2014;7(1):74.
10. Guryanova OA, Shank K, Spitzer B, et al. DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nature Medicine*. 2016;22(12):1488–1495.
11. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med*. 2016;374(23):2209–2221.
12. Imren S. Modeling de novo leukemogenesis from human cord blood with MN1 and NUP98HOXD13. *Blood*. 2014;124(24):3608–3612.
13. Vick B, Rothenberg M, Sandhöfer N, et al. An Advanced Preclinical Mouse Model for Acute Myeloid Leukemia Using Patients' Cells of Various Genetic Subgroups and In Vivo Bioluminescence Imaging. *PLoS ONE*. 2015;10(3):e0120925–.
14. Cosgun KN, Rahmig S, Mende N, et al. Kit regulates HSC engraftment across the human-mouse species barrier. *Cell Stem Cell*. 2014;15(2):227–238.
15. Verhaak RGW, Wouters BJ, Erpelinck CAJ, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*. 2009;94(1):131–134.

16. Hebestreit K, Gröttrup S, Emden D, et al. Leukemia Gene Atlas - A Public Platform for Integrative Exploration of Genome-Wide Molecular Data. *PLoS ONE*. 2012;7(6):e39148–.
17. Metzeler KH, Herold T, Rothenberg-Thurley M, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. *Blood*. 2016;128(5):686–698.
18. Spencer DH, Russler-Germain DA, Ketkar S, et al. CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell*. 2017;168(5):801–816.e13.
19. Wahlestedt M, Ladopoulos V, Hidalgo I, et al. Critical Modulation of Hematopoietic Lineage Fate by Hepatic Leukemia Factor. *CellReports*. 2017;21(8):2251–2263.
20. Lavallee VP, Gendron P, Lemieux S, et al. EVI1-rearranged acute myeloid leukemias are characterized by distinct molecular alterations. *Blood*. 2015;125(1):140–143.
21. Lemieux S, Sargeant T, Laperrière D, et al. MiSTIC, an integrated platform for the analysis of heterogeneity in large tumour transcriptome datasets. *Nucleic Acids Research*. 2017;45(13):e122–e122.
22. Lavallée V-P, Lemieux S, Boucher G, et al. RNA-sequencing analysis of core binding factor AML identifies recurrent ZBTB7A mutations and defines RUNX1-CBFA2T3 fusion signature. *Blood*. 2016;127(20):2498–2501.
23. Lavallée V-P, Baccelli I, Krosl J, et al. The transcriptomic landscape and directed chemical interrogation of MLL-rearranged acute myeloid leukemias. *Nat Genet*. 2015;47(9):1030–1037.
24. Baccelli I, Krosl J, Boucher G, et al. A novel approach for the identification of efficient combination therapies in primary human acute myeloid leukemia specimens. *Blood Cancer Journal*. 2017;7(2):e529–e529.
25. Lavallee VP, Lemieux S, Boucher G, et al. Identification of MYC mutations in acute myeloid leukemias with NUP98-NSD1 translocations. *Leukemia*. 2016;30(7):1621–1624.
26. Pabst C, Bergeron A, Lavallée V-P, et al. GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. *Blood*. 2016;127(16):2018–2027.
27. Schlenk RF, Kayser S, Bullinger L, et al. Differential impact of allelic ratio and insertion site in FLT3-ITD-positive AML with respect to allogeneic transplantation. *Blood*. 2014;124(23):3441–3449.
28. Wilson NK, Foster SD, Wang X, et al. Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Stem Cell*. 2010;7(4):532–544.
29. Russler-Germain DA, Spencer DH, Young MA, et al. The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. *Cancer Cell*. 2014;25(4):442–454.
30. Fares I, Chagraoui J, Gareau Y, et al. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science*. 2014;345(6203):1509–1512.
31. Pabst C, Krosl J, Fares I, et al. Identification of small molecules that support human leukemia stem cell activity ex vivo. *Nat Methods*. 2014;11(4):436–442.
32. Lai CK, Norddahl GL, Maetzig T, et al. Meis2 as a critical player in MN1-induced leukemia. *Blood Cancer Journal*. 2017;7(9):e613.
33. Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. 1994;367(6464):645–648.
34. Dick JE, Lapidot T. Biology of Normal and Acute Myeloid Leukemia Stem Cells.

Int J Hematol. 2005;82(5):389–396.

35. Komorowska K, Doyle A, Wahlestedt M, et al. Hepatic Leukemia Factor Maintains Quiescence of Hematopoietic Stem Cells and Protects the Stem Cell Pool during Regeneration. *CellReports*. 2017;21(12):3514–3523.
36. Yu X, Alder JK, Chun JH, et al. HES1 inhibits cycling of hematopoietic progenitor cells via DNA binding. *Stem Cells*. 2006;24(4):876–888.
37. Zou P, Yoshihara H, Hosokawa K, et al. p57Kip2 and p27Kip1 Cooperate to Maintain Hematopoietic Stem Cell Quiescence through Interactions with Hsc70. *Stem Cell*. 2011;9(3):247–261.
38. Georgia S, Soliz R, Li M, Zhang P, Bhushan A. p57 and Hes1 coordinate cell cycle exit with self-renewal of pancreatic progenitors. *Developmental Biology*. 2006;298(1):22–31.
39. Giovannini C, Gramantieri L, Minguzzi M, et al. CDKN1C/P57 Is Regulated by the Notch Target Gene Hes1 and Induces Senescence in Human Hepatocellular Carcinoma. *The American Journal of Pathology*. 2012;181(2):413–422.
40. Li G, Domenico J, Lucas JJ, Gelfand EW. Identification of multiple cell cycle regulatory functions of p57Kip2 in human T lymphocytes. *The Journal of Immunology*. 2004;173(4):2383–2391.
41. Kachroo P, Szymczak S, Heinsen F-A, et al. NGS-based methylation profiling differentiates TCF3-HLF and TCF3-PBX1 positive B-cell acute lymphoblastic leukemia. *Epigenomics*. 2018;10(2):133–147.

Figure 1

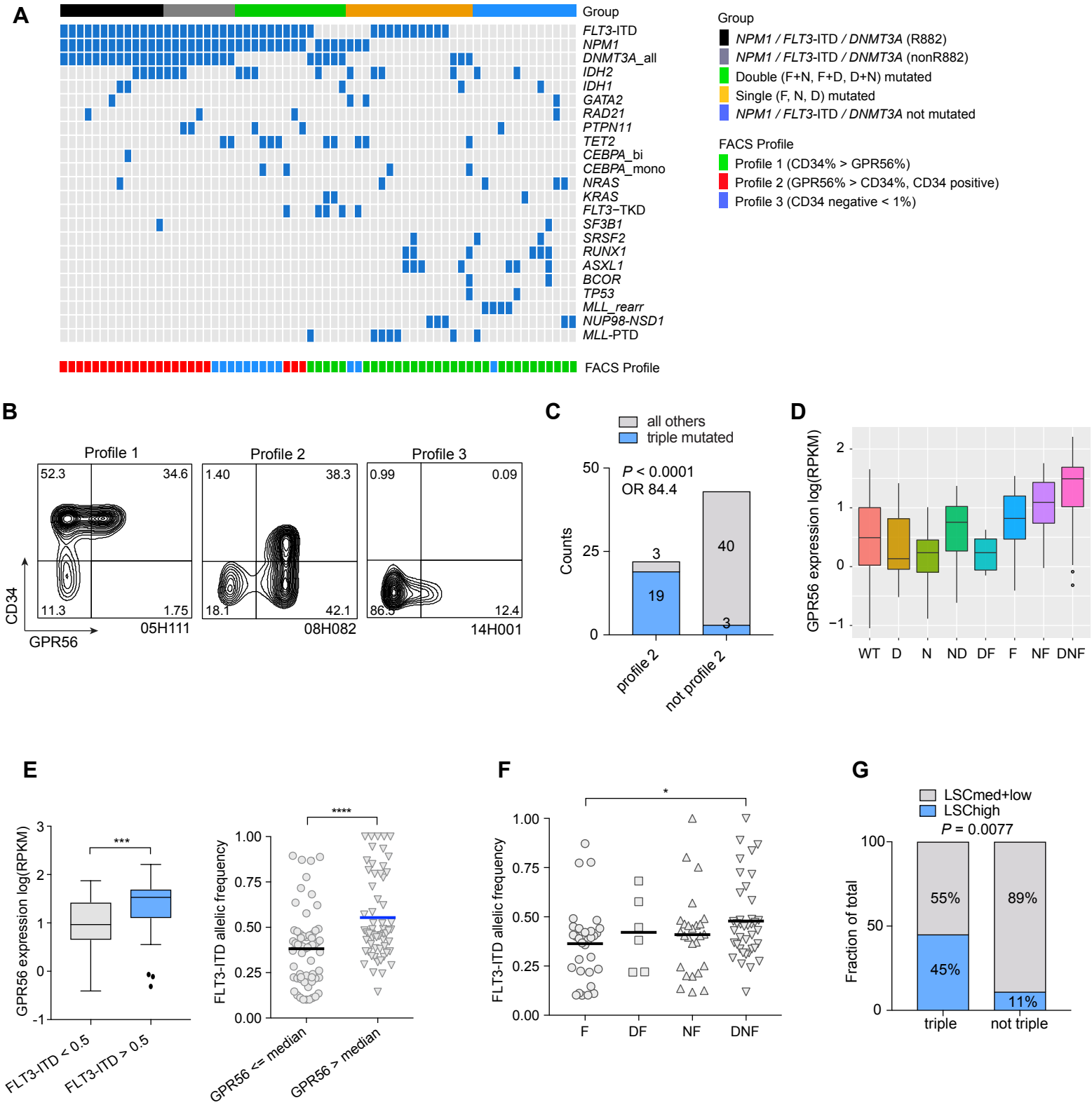


Figure 2

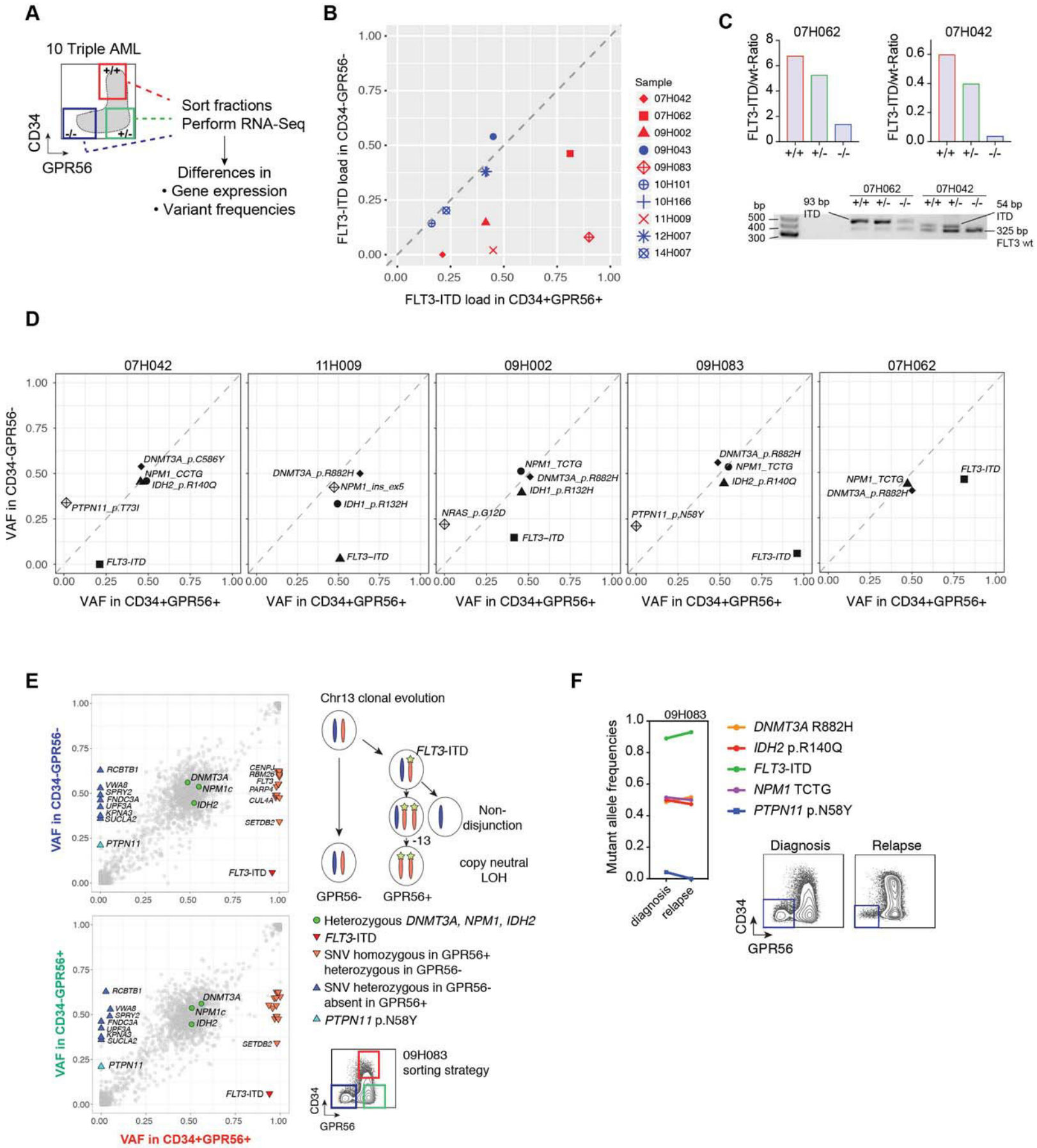
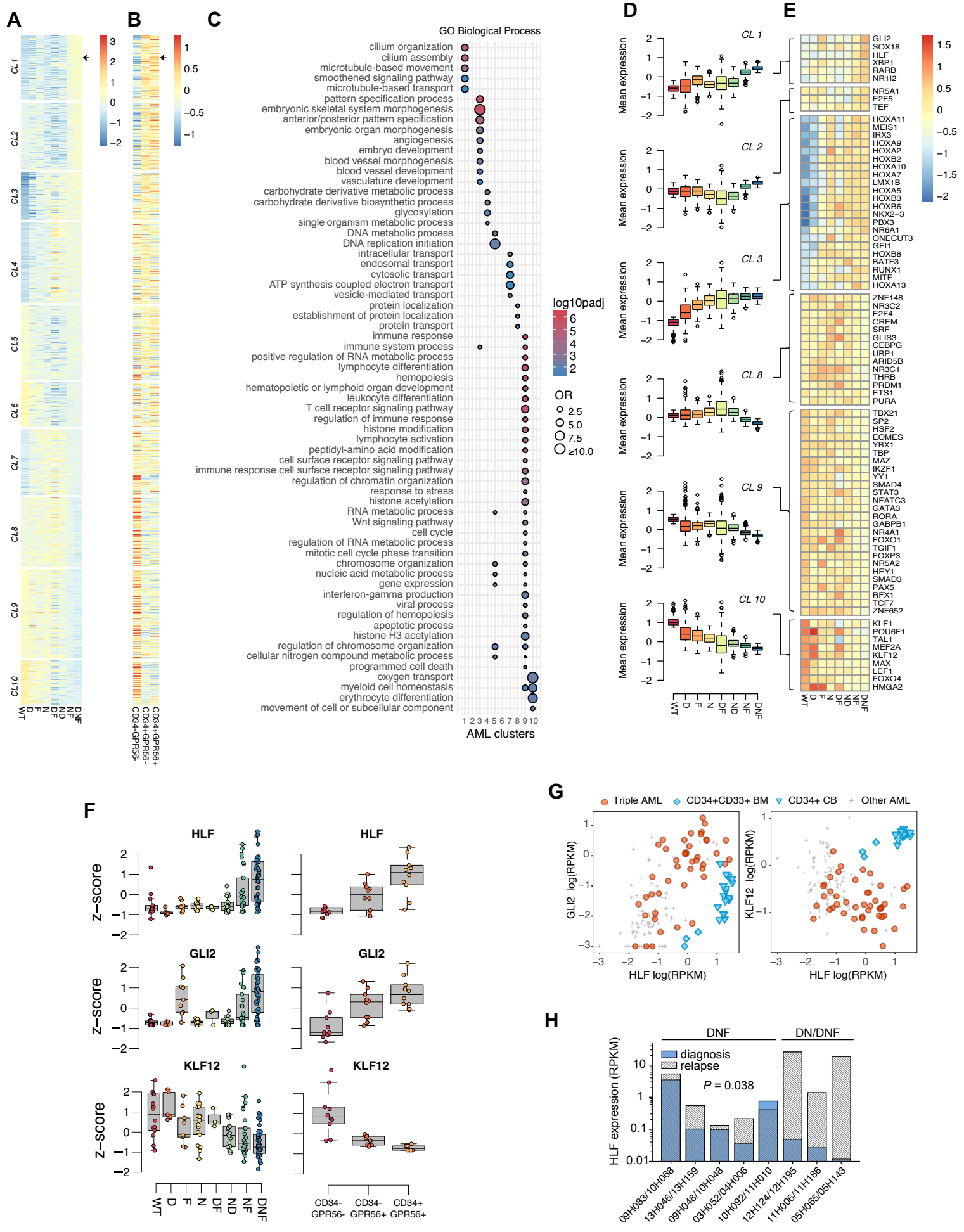
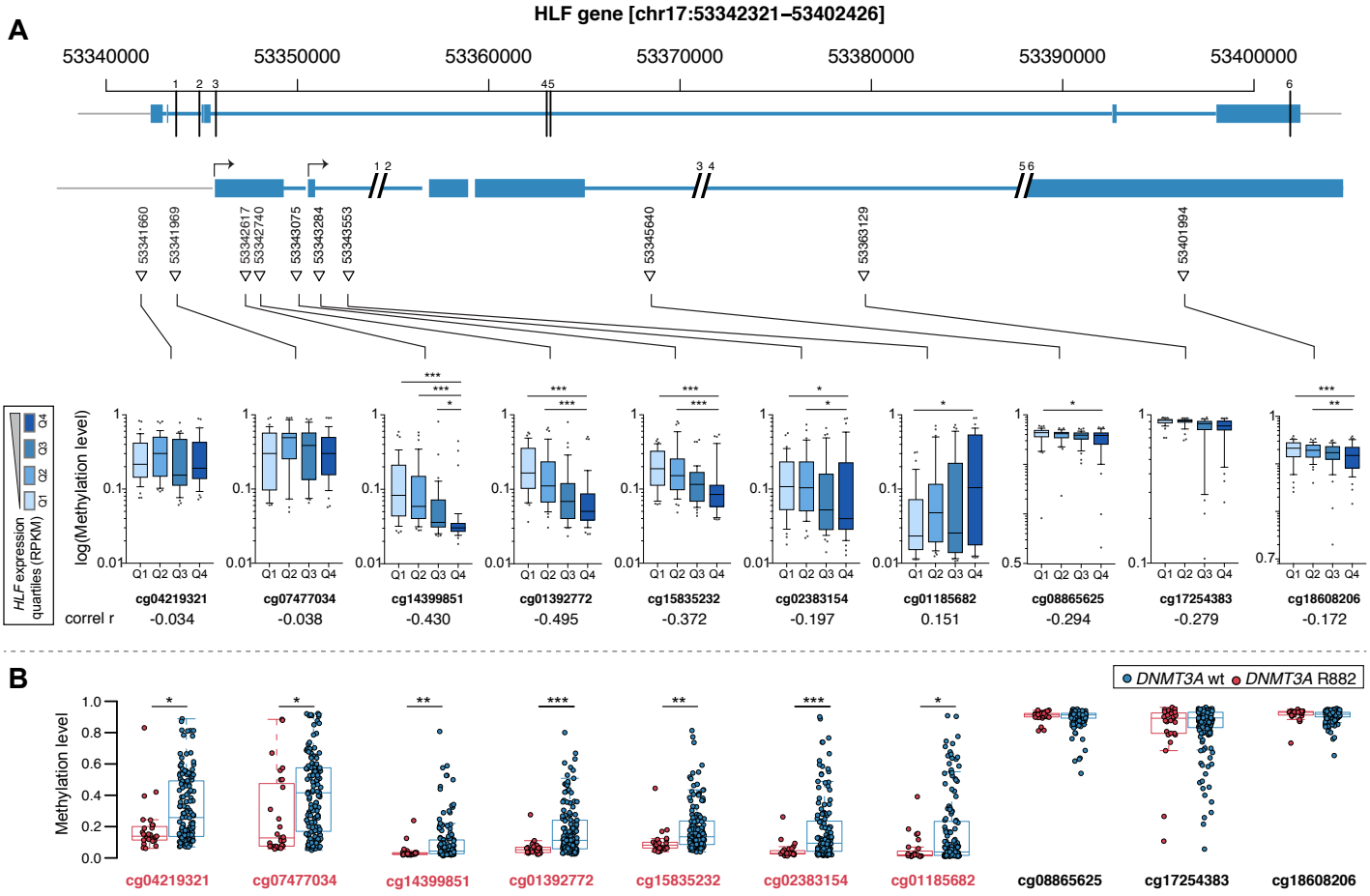


Figure 3





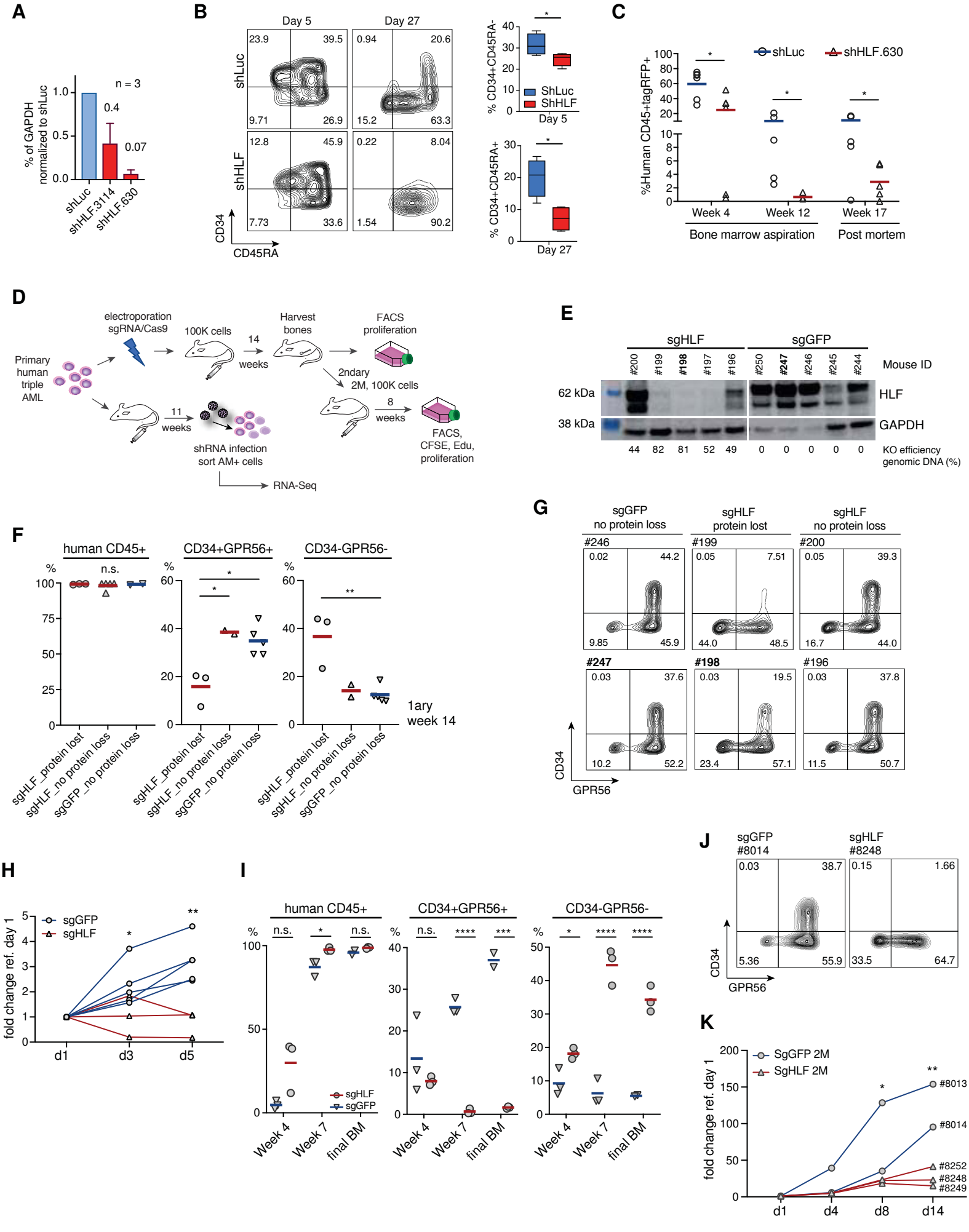
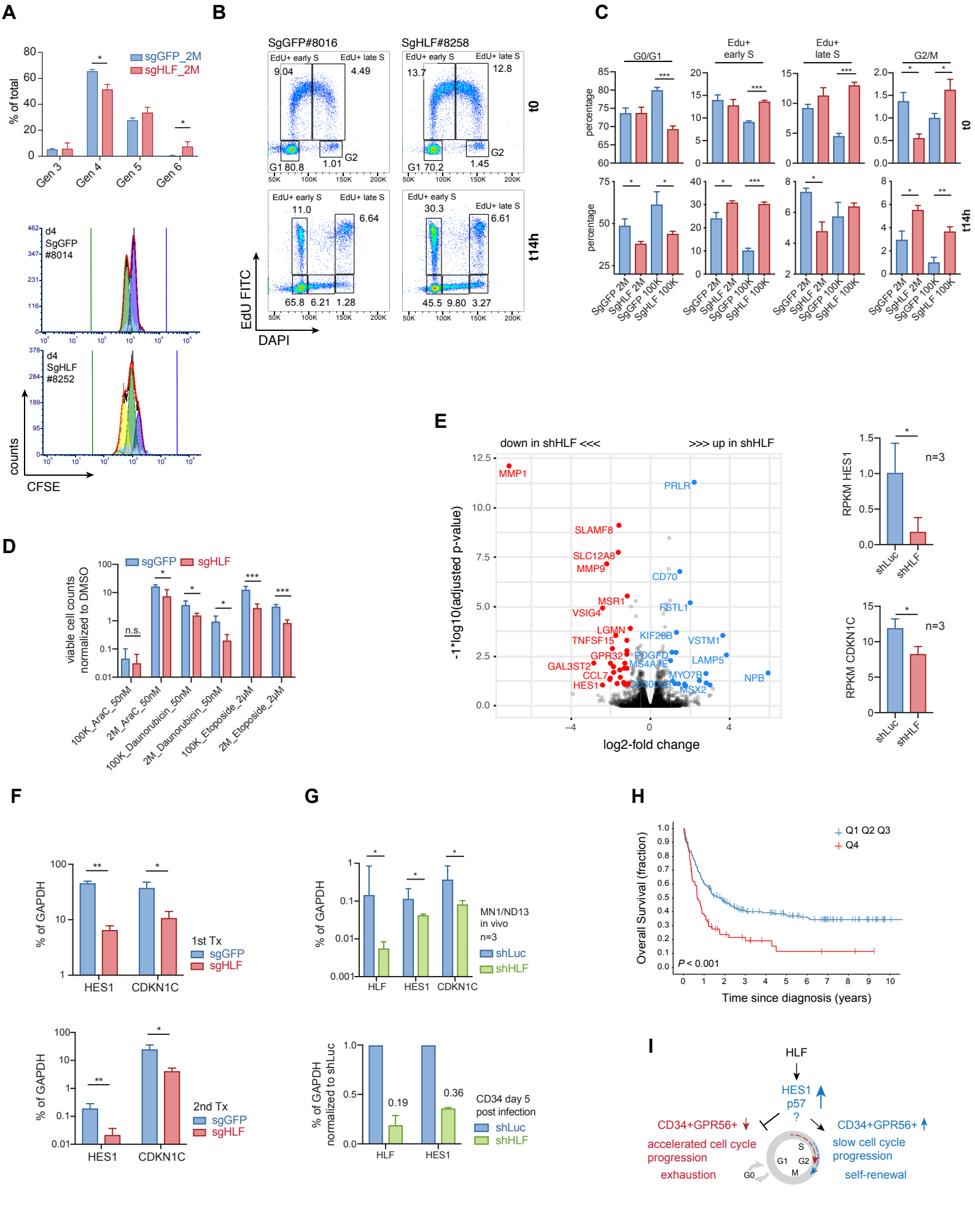


Figure 6



Prepublished online May 10, 2019;
doi:10.1182/blood.2018862383

Hepatic leukemia factor is a novel leukemic stem cell regulator in DNMT3A, NPM1, and FLT3-ITD triple-mutated AML.

Swati Garg, Armando Reyes-Palomares, Lixiazi He, Anne Bergeron, Vincent-Philippe Lavallée, Sébastien Lemieux, Patrick Gendron, Christian Rohde, Jianglong Xia, Prarabdha Jagdhane, Carsten Müller-Tidow, Daniel B. Lipka, Suzan Imren, R. Keith Humphries, Claudia Waskow, Binje Vick, Irmela Jeremias, Guillaume Richard-Carpentier, Josée Hébert, Guy Sauvageau, Judith Zaugg, Frédéric Barabé and Caroline Pabst

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.