Université de Montréal

Identifying regulators of Hepatic Leukemia Factor (*HLF*) expression in hematopoietic cells using a fluorescent reporter transgene.

Par

Hilary Johnson

Programme de Biologie Moléculaire,

Faculté de Médicine

Mémoire présenté à la Faculté de Médicine

en vue de l'obtention du grade de Maître es Science (M.Sc.)

en Biologie Moléculaire, option générale

January 10th 2022

Université de Montréal

Academic unit: Department of Medicine/Institute for Research in Immunology and Cancer, Faculty of Medicine/University of Montreal

---

This thesis entitled

Identifying regulators of Hepatic Leukemia (*HLF*) expression in hematopoietic cells using a fluorescent reporter transgene

Presented by

Hilary Johnson

Has been evaluated by a jury composed of the following people:

Dr. Trang Hoang
Chairman-rapporteur/Président

Dr. Martin Sauvageau
Member of the Jury

Dr. Guy Sauvageau
Research Director

Resumé

Le facteur de leucémie hépatique (HLF) est un facteur de transcription et l'un des gènes les plus sélectivement exprimés par les cellules souches hématopoïétiques (CSHs) humaines et de souris. L'expression de *HLF* identifie l'activité fonctionnelle des CSHs et est un gène clé du caractère souche maintenant la quiescence et l'auto-renouvellement.

Une surexpression de *HLF* a été observée dans les sous-groupes de leucémie myéloïde aiguë (LMA) comportant des mutations des gènes *NPM1* (*NPM1c*) et *FLT3* (*FLT3-ITD*), et il a été suggéré que ceci protège contre la mort cellulaire et les traitements chimio-toxiques. Cependant, la voie de signalisation responsable de la surexpression de *HLF* dans les leucémies avec mutation *NPM1c* reste inconnue. Puisque l'expression de *HLF* est associée à l'activité CSH et diminue au moment de la différentiation, *nous avons émis l'hypothèse que la fusion du gène HLF avec un gène rapporteur fluorescent (HLF-ZsG) dans une lignée de cellules leucémiques (IMS-M2) pourrait être utilisée pour identifier les gènes qui régulent l'expression de HLF dans les CSHs normales*.

La protéine NPM1 réside dans le noyau et régule la réparation de l'ADN, la croissance cellulaire et la prolifération. Des mutations qui entraînent sa localisation cytoplasmique (*NPM1c*) ont été observées dans la LMA, entraînant une prolifération et une différenciation aberrante des cellules. La relocalisation nucléaire de la protéine NPM1c entraîne la régulation à la baisse des gènes homéotiques dans les cellules IMS-M2.

Nous avons validé l'expression de la fusion HLF-ZsG dans les cellules IMS-M2 et confirmé la dépendance de l'expression de *HLF* à NPM1c. De plus, nous avons identifié *MEIS1*, *HOXA10* et *NKX2-3* comme étant des régulateurs potentiels de *HLF*. Nous avons également identifié un réseau régulant l'expression de *HLF* en aval de NPM1c aidant à comprendre les mécanismes de régulation de *HLF* dans les CSHs normales.

Mots-clés : *HLF*, LMA, CSH, réseaux génétiques, CRISPR, gènes rapporteurs fluorescents.

Abstract

Hepatic leukemia factor (HLF) is a transcription factor and one of the most selectively expressed genes in human and mouse hematopoietic stem cells (HSCs). HLF identifies functional HSC activity and is a key “stemness” gene that maintains quiescence and self-renewal.

Upregulation of *HLF* has been observed in NPM1c/*FLT3*-ITD-mutated acute myeloid leukemia (AML) and is thought to protect cells from death and chemotoxic insults. However, the pathways which cause *HLF* expression in NPM1c-mutated leukemia are unknown. Since *HLF* expression is associated with HSC activity and decreases upon differentiation, *we hypothesized that an HLF-ZsG fluorescent reporter in a leukemia cell line (IMS-M2) could be used as a readout to identify genes that regulate HLF in normal HSCs*.

The NPM1 protein resides in the nucleus and regulates DNA repair, cell growth and proliferation. Mutations which result in its cytoplasmic localization (NPM1c) have been observed in AML and lead to aberrant cell proliferation and differentiation. Nuclear relocalization of NPM1c results in the downregulation of homeobox genes in IMS-M2 cells.

We validated the HLF-ZsG reporter in IMS-M2 cells and confirmed the dependency of *HLF* expression on NPM1c. We then identified *MEIS1, HOXA10* and *NKX2-3* as potential regulators of *HLF*. We have identified a preliminary network of *HLF* expression downstream of NPM1c, which may also help us understand *HLF* regulation in normal HSCs.

Keywords : HLF, AML, HSCs, genetic networks, CRISPR, fluorescent gene reporter.

Contents

List of Tables

Appendices

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendices

List of acronyms and abbreviations

5-FU: 5-fluorouracil

bZIP: basic leucine zipper

BM: bone marrow

ChIP: chromatin immunoprecipitation

CRISPR: clustered regularly interspaced short palindromic repeats

ddPCR: droplet digital polymerase chain reaction

DSB: double-stranded break

EDTA: ethylenediaminetetraacetic acid

EMCV: encephalomyocarditis virus

EMP: erythro-myeloid progenitor

ETP: early T-cell progenitor

FACS: fluorescent activated cell sorting

FBS: Fetal bovine serum

FLT3: fms-like tyrosine kinase 3

GM-CSF: granulocyte-macrophage colony-stimulating factor

GMP: granulocyte/macrophage progenitor

HI FBS: heat-inactivated fetal bovine serum

HLF: Hepatic leukemia factor

HR: homologous recombination

HSCs: hematopoietic stem cells

HSPCs: hematopoietic stem and progenitor cells

INDEL: insertion/deletion

IMDM: Iscove Modified Dulbecco Media

IRES: internal ribosome entry site

KO: knocked out

LSC: leukemic stem cell

LT-HSCs: long-term HSCs

MEP: megakaryocyte/erythroid progenitor

MkP: megakaryocyte restricted progenitor

MPPs: multipotent progenitors

NHEJ: non-homologous end joining

NPM1: nucleophosmin1

ORF: open reading frame

PAR: proline and acidic amino-acid-rich

PB: peripheral blood

PBS: phosphate buffered saline

qPCR: quantitative polymerase chain reaction

rAAV6: recombinant adeno-associated virus pseudotype 6

RNP: ribonucleoprotein

RPMI: Roswell Park Memorial Institute

ROSE2: Rank Ordering of Super-Enhancers 2

RFP: red fluorescent protein

SCF: stem cell factor

sgRNA: single-guide RNA

TKD: tyrosine kinase domain

TPM: transcripts per million

UCB: umbilical cord blood

WT: wild-type

ZsG: ZsGreen

To my parents,

who ignited my passion for science,

yet always encouraged me to pursue what makes me happy.

Remerciements

I would like to thank all who shaped my time as a master's student, as it was an incredibly positive experience. This pursuit has taught me invaluable lessons which I will carry on for the rest of my life and has set the building blocks for my journey to become a patient, persistent and resilient scientist.

To begin, I would like to thank my research director, Dr. Guy Sauvageau, for welcoming me to his team and putting trust in me from the beginning of this journey. Not only am I grateful for his kindness, patience, understanding and support, but his remarkable work in science and medicine continues to inspire me every single day.

I would also like to say a very special thank you to Bernhard Lehnertz for teaching me experimental techniques from the most basic protocols to exciting and new technologies. Dr. Lehnertz continues to remind me of the compelling, exploratory and artistic nature of science, which is a passion I will never lose.

I would also like to thank Unain Ansari, Tara MacRae, Jalila Chagraoui, Nadine Mayotte, Isabelle Boivin, Simon Girard, Azadeh Hajmirza, Ossama Moujaber and all other lab members for their experimental support, for welcoming me to the team, and for always providing a lively and fun atmosphere in the lab. I would like to thank Dr. Marie-Eve Bordeleau for her very warm and consistent support and advice. Without this team, I believe that trying to complete a master's during the current global situation would have been a very different and difficult endeavour.

In addition, I would like to thank the collaborative work of all scientists and administrative personal at the IRIC and University of Montreal for helping to positively shape young scientists and provide such grand and exciting educational opportunities. A special thank you to Annie Gosselin and Angélique Bellemare-Pelletier for your help with flow cytometry experiments, members of the genomics platform for help with sequencing experiments and Dr. Julie Mantovani for her prompt and consistent administrative support. Thank you to Dr. Trang Hoang and Dr. Martin Sauvageau for their membership in my master's jury committee. A final thank you to the organizations who helped sponsor my studies (UdeM, SAÉ, IRIC and Mitacs). In conclusion, the time I have spent in this lab and at the IRIC

has allowed me to grow as a student and scientist in ways I didn't realize were possible and I am excited to see the paths and opportunities that this experience will allow me to take on in the future.

Please see page 52 for a contributions list for each figure and corresponding experiments.

CHAPTER 1 - INTRODUCTION

1.1 Hematopoiesis and HSCs

The hematopoietic system is a highly regenerative and regulated system that is responsible for producing a continuous supply of blood and immune cells and appropriately recovering after injury or insult.[1,2,3] The cells of the hematopoietic system are organized into a hierarchy in which hematopoietic stem cells (HSCs) lie at the apex.[4] HSCs are found in the peripheral blood (PB), bone marrow (BM), umbilical cord blood (UCB) and fetal liver and are defined by their ability to execute two functions: to self-renew and replenish their population and to maintain potential to differentiate into multiple lineage-committed progenitor cells which give rise mature blood cells.[5] Organization of this system is in part regulated by a network of transcription factors that influence HSC fate and function.[6]

HSCs can be distinguished into long-term repopulating HSCs (LT-HSCs) and short-term HSCs (ST-HSCs) based on their functional properties.[6] LT-HSCs reside in the BM and sustain long-term hematopoiesis through; asymmetric cell division, producing two daughter cells with different cellular fates and; symmetric divisions, producing two daughter cells with identical fates, sustaining self-renewal.[7,8] LT-HSCs can asymmetrically divide into cells that can either self-renew to sustain the hematopoietic stem cell population or into ST-HSCs or lineage-restricted progenitors which can differentiate into mature blood and immune cells of the hematopoietic system. LT-HSCs can also divide symmetrically to sustain self-renewal.[6,8,9] While LT-HCS sustain hematopoiesis for the entire life of the organism, ST-HSCs only provide hematopoiesis in the short term.[7] In addition to short-term hematopoiesis, ST-HSCs are responsible for providing early and temporary hematopoietic recovery after transplantation, as well as dividing and eventually differentiating into downstream progenitors of the hematopoietic system.[6,7,10]

A complex signalling network decides the fate of HSC differentiation. Based on canonical pathways, newer models suggest that different cues will initiate differentiation of HSCs into multipotent progenitors (MPPs) which can then differentiate into multilymphoid progenitors (MLP), megakaryocyte/erythroid progenitors (MEP), common myeloid

progenitors (CMPs) and granulocyte/macrophage progenitors (GMPs). MLPs give rise to lymphoid lineages, MEPs differentiate into megakaryocytes or erythrocytes (mature blood cells) and CMPs give rise to GMPs which differentiate into granulocytes (neutrophils, eosinophils, basophils or mast cells) as well as monocytes and subsequently macrophages.[6,11] Several MPP populations have been identified with different developmental biases. Classical models of hematopoietic have organized cells into a stepwise model of progression from HSCs to differentiated cells using evaluation of surface markers assessed by flow cytometry. However, recent advancements in single-cell sequencing have suggested a continuous progression of cell stages from primitive to differentiated cells, accomplished through gradual transitions of cells between stages of differentiation **(Figure 1).**[11,12]

Due to their relevance in hematopoiesis, much effort has gone into identifying the phenotypic markers that distinguish HSCs from their downstream progenitors, using both cell surface markers and single-cell sequencing. Until recently, a single molecular marker that can selectively identify HSCs had not been discovered, therefore a panel of markers has been used to distinguish this population from other cells of the blood system.[13] This identification system is based on the presence or absence of the expression of CD34, CD38, CD45RA, ITGA3, EPCR, CD90 and CD49f surface markers.[10] LT-HSC and ST-HSC populations can also specifically be discriminated from one another based on surface marker expression. $CD34^{+}CD45RA^{-}ITGA3^{+}$(CD49c)$EPCR^{+}$(CD201)$CD90^{+}$ expression has been found to identify LT-HSCs and $CD34^{+}CD45RA^{-}ITGA3(CD49c)^{-}EPCR^{+}CD90^{+}$ expression has been found to identify ST-HSCs when cord blood samples are cultured with the pyrimidoindole derivative molecule UM171 for expansion purposes.[10]

The distinction of the LT-HSC from ST-HSC populations, in addition to the distinction of HSC populations from downstream progenitors and differentiated cells, is important to allow scientists to study their mechanisms of self-renewal maintenance and differentiation. An understanding of LT-HSC and ST-HSC surface marker profile expression, in addition

to their molecular networks within the cell, also helps reveal how they may become deregulated to lead to blood malignancies.

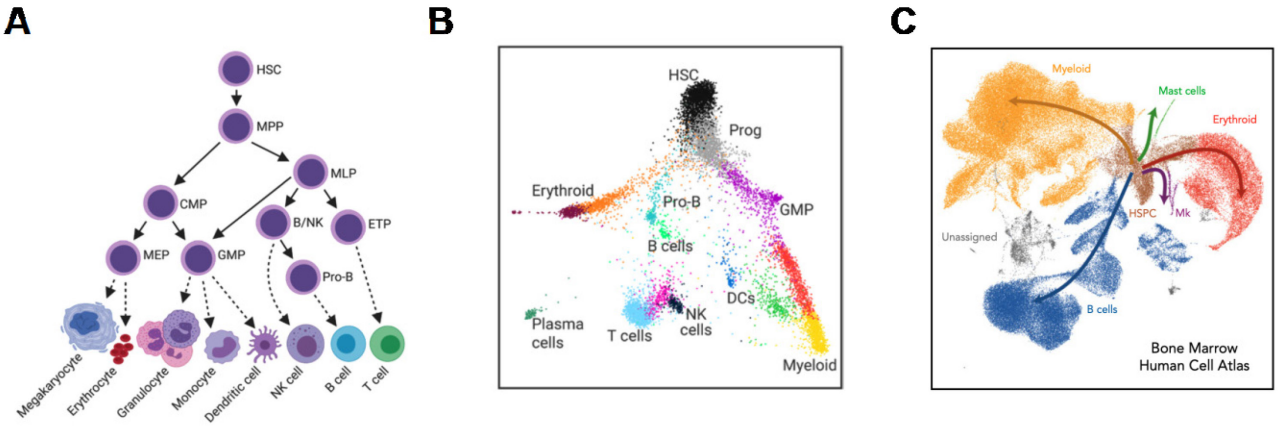


Figure 1 Different ways to model the hierarchy of the hematopoietic system. (A) Figure adapted from Acosta et al 2021.[11] Separation and organization of cells of the hematopoietic system using surface markers quantified by flow cytometry. (B) Figure adapted from Acosta et al 2021.[11] Single-cell sequencing profiles to study hematopoietic trajectories suggest a more gradual and continuous progression of differentiation. (C) Figure adapted from Acosta et al 2021.[12] (UMAP modelling of single-cell transcriptomes from the bone marrow mononuclear cell compartment also demonstrates a continuous progression of differentiation. Data downloaded from the Human Cell Atlas portal and processed by I.K. from Watcham et al., 2019.[12]

1.2 *HLF* Expression in Hematopoiesis

An intersection of the transcriptome profiles of LT-HSC versus depleted LT-HSC populations from UCB revealed *HLF* to be specifically expressed in the LT-HSC population.[13] Single-cell transcriptomic profiles of un-expanded and *ex vivo* expanded $CD34^+$ cord blood cells confirmed its specific expression within the hematopoietic stem and progenitor cell (HSPC) population.[13] *HLF* expression has also been detected in adult BM HSCs; gene-enrichment analysis of adult BM revealed *HLF* to be the most highly enriched BM HSPC gene.[14] Furthermore, *HLF* mRNA levels from BM HSPCs revealed specific expression in HSPC populations and a lack of expression in more differentiated populations.[15] *HLF* also has been demonstrated to play a role in embryonic

hematopoiesis.[14] Its expression was found to increase with HSC development in the embryo and to discriminate HSC-producing cells from erythro-myeloid progenitor (EMP)-producing cells in the embryo.[14] Studies of *HLF* expression during embryogenesis revealed *HLF* to be discriminatorily expressed in fetal liver HSCs as well.[14] During development, *HLF* has also been found to be highly expressed in the liver, lung and adult nervous system.[16] Finally, transcriptomic data has revealed that *HLF* expression declines as cells differentiate from HSCs towards lineage-committed progenitors, with no expression detected in mature peripheral cells.[13]

Due to its specific expression in HSPC populations and lack of expression in mature hematopoietic cells, it can be hypothesized that *HLF* carries out an essential role in HSC function. In fact, differentiation marker analysis after knockout of *HLF* in HSCs suggests that *HLF* keeps HSCs in a primitive state.[17] shRNA-mediated knockdown of *HLF* in $CD34^+$ cord blood cells also resulted in accelerated myelomonocytic differentiation compared to control cells.[17] In addition, *Hlf* deficient ($Hlf^{-/-}$) mice revealed reduced platelet numbers in the PB.[18] An accompanied reduction in phenotypic megakaryocyte restricted progenitor (MkP) cells suggests that an *Hlf* deficiency reduces differentiation propensity towards platelets specifically.[18] However, a normal frequency and distribution of mature blood cells in PB and BM suggest that mice can sustain normal hematopoiesis even when lacking *Hlf*-expression. In addition to platelet differentiation, *HLF* seems to direct myeloid and lymphoid commitment.[15] Overexpression of *HLF* in granulocyte/monocyte/lymphocyte progenitors resulted in a decreased fraction of B-cell based lymphoid progenitors and an increase in granulocyte-monocyte progenitors, as well as a lack of differentiation of B-cells, T-cells and NK-cells and an increased generation of myeloid cells.[15] Therefore, it can be concluded that the impact of *HLF* expression on directing hematopoietic differentiation starts early in differentiation progression and affects multiple progenitors and lineages.

In addition to its role in preventing proper and balanced differentiation of HSCs, *HLF* has been found to identify HSCs with reconstitution activity *in vivo* and has been demonstrated to impart self-renewal properties onto HSCs and their downstream progenitors.[2,13] Transplantation of bone marrow from *Hlf* knockout ($Hlf^{-/-}$) mice showed significantly

reduced donor chimerism in PB and BM compared to transplantation of control *Hlf*$^{+/+}$ cells in recipient mice.[18] Furthermore, serial transplantation of BM cells from primary *Hlf*$^{-/-}$ recipients to secondary and tertiary recipients revealed an even stronger reduction in reconstitution capacity in secondary recipients and negligible reconstitution in tertiary recipients, highlighting the role of *HLF* expression in maintaining the regenerative potential of the HSC population.[18]

Furthermore, enforced expression of *Hlf* was found to enhance self-renewal properties of HSCs.[2] Mice transplanted with transgenic HSCs overexpressing *Hlf* under doxycycline-inducible control revealed a dramatic and prolonged increase in BM reconstitution, HSC numbers and repopulating capacity.[4] Additionally, ectopic expression of *Hlf* in multipotent progenitor (MPP), CMP, GMP and MEP populations resulted in long-term sustained mixed myeloid colony-forming potential, compared to control transduced HSCs which showed limited colony-forming potential and an inability to maintain production of mixed myeloid lineages.[2] Inducing *Hlf* expression for 30 days in HSCs resulted in cultures that gave rise to megakaryocyte, macrophage, granulocyte and progenitor cells, compared to control cultures where *Hlf* was not induced and which only gave rise to macrophages.[2] In addition, *Hlf* transduced MPPs and myeloid progenitors generated a significantly greater number of colonies over time upon serial plating in methylcellulose, compared to control transduced cells which gradually lost the ability to generate colonies over time. These results demonstrate that continued *Hlf* expression is necessary to provide long-term self-renewal activity and myeloid lineage formation in HSC cultures.[2] Finally, and perhaps most interestingly, *HLF* has been identified as a member of a cocktail of transcription factors (including *LMO*2, *PBX1*, *ZFP37*, *PRDM5* and *RUNX1T1*), in which transient expression was found to be sufficient to provide long-term, multilineage transplantation potential onto committed myeloid and lymphoid progenitors and essentially reprogram them back to a “stem cell-like state.”[19]

HLF is thought to be able to maintain self-renewal properties of HSCs in part by regulating aspects of their cell cycle such as quiescence.[18] Quiescence is an important and temporary cell cycle state, often exhibited by HSCs, where cells reside mainly in the G0/G1 phase and remain non-proliferative.[7,20] An important feature of quiescent cells is

that they can re-enter the proliferative cycle during appropriate conditions, such as in response to growth signals, stress or injury. Therefore, by remaining highly quiescent, HSCs can mobilize in the face of injury to replenish damaged tissues and continuously replenish their population.[2,4,21]

HLF was identified as a regulator of quiescence in HSCs through functional HLF knockout and overexpression studies.[18] Genome-wide transcriptome analysis of purified *Hlf*$^{-/-}$ HSCs revealed a set of differentially expressed genes in *Hlf*$^{-/-}$ knockout compared to *Hlf*$^{+/+}$ HSCs.[18] *Hlf* overexpressing HSCs revealed that HLF binds to genes that regulate HSC activity. Overlapping transcriptomic data from downregulated genes in *Hlf*$^{-/-}$ knockout HSCs with chromatin immunoprecipitation (ChIP)-sequencing results of *Hlf* overexpressing HSCs revealed direct targets of HLF.[18] Specifically, among the genes which were downregulated in *Hlf*$^{-/-}$ HSCs, *Gfi1* and *Irf2*, two important regulators of HSC quiescence, were identified as direct targets of HLF.[18] The role of *Hlf* in HSC quiescence was further demonstrated when *Hlf*$^{-/-}$ knockout HSCs revealed significantly fewer *Hlf*$^{-/-}$ HSCs in the G0 cell cycle phase and significantly more in the G1 phase.[18] This change in cell-cycle status was only detected in the HSC population, compared to immediate downstream progenitors which were unaffected.[18] Finally, cell cycle labelling analysis revealed an increase in actively cycling *Hlf*$^{-/-}$ cells compared to controls.[17] In summary, changes in cell cycle status and targets of *HLF* strongly indicate this gene as a regulator of HSC quiescence.

In addition to maintaining quiescence, *HLF* also seems to play a role in protecting cells from apoptosis. Annexin-V staining to assess the apoptotic status of *Hlf* overexpressing cells revealed a significantly smaller proportion of apoptotic cells compared to controls.[4] Additionally, q-RT-PCR revealed that overexpression of *HLF* in primitive human hematopoietic cells induces the expression of *BCL2* expression, a regulator of apoptosis.[4] This interaction suggests that *HLF* may support anti-apoptotic mechanisms and protect premature cell death in HSCs in part by regulation of *BCL2*.[4] Finally, *Hlf*-transduced murine epidermal cells showed upregulation of anti-apoptotic genes, downregulation of regulators of cell death and higher viability after induction of TNFα stress compared to

controls.[22] These results demonstrate that high levels of *HLF* may help protect cells against apoptosis-mediated cell death.[22]

HLF seems to provide protection against chemotherapeutic stress in HSCs as well. Treatment of *Hlf*$^{-/-}$ mice with the myeloablative chemotherapeutic agent 5-fluorouracil (5-FU), which specifically targets and eliminates actively cycling cells, resulted in mice that were unable to recover or survive this treatment, compared to control mice which all survived.[18] Finally, sublethal irradiation of *Hlf*$^{-/-}$ mice showed significantly lower survival and a pronounced reduction in HSPCs compared to control mice.[18] The demonstrated role that *HLF* holds in protecting HSCs from apoptosis and maintaining their quiescent state may provide an explanation as to why *Hlf* expressing mice were able to recover from chemotherapeutic-induced stress or irradiation. By knocking out *HLF* expression in these studies, the protective mechanism that HLF held on otherwise healthy HSCs was released, causing them to die in the face of toxic stress and without allowing adequate recovery that would usually be provided from the reservoir of quiescent HSCs. [2,4,7,20,21]

In conclusion, *HLF* has been identified in many studies as a key player in the human HSC-specific gene signature. It has been defined as an essential “stemness gene” which influences control over crucial HSC properties such as differentiation, self-renewal and cell cycle regulation, including maintenance of quiescence and survival.

1.3 HLF Expression in Certain Leukemias

In addition to a healthy blood system, *HLF* has been shown to provide leukemic cells with drug resistance and has been shown to be specifically expressed in certain leukemias. In addition to the hematopoietic system, *HLF* has also been considered a “stemness” gene in leukemia.[18] Leukemias are blood cancers that arise from mutations in healthy BM HSCs that render these cells no longer able to perform their normal function such as differentiation into blood cells.[23,24] An inability of HSCs to differentiate in leukemia instead causes them remain as immature progenitor cells, also known as blast cells.[25] These cells then subsequently accumulate in the blood and bone marrow and interfere with normal hematopoesis.[25] Due to its relevance in stem cell maintenance and differentiation, it is hypothesized that *HLF* and its aberrant expression could contribute therefore to this malignancy.

In fact, HLF was first identified as part of the t(17:19) (q22;p13) chromosomal translocation in which E2A-HLF chimeric transcript was generated.[18,26] The E2A-HLF fusion is thought to be a driving factor of leukemogenesis in t(17:19)-ALL in part by preventing cancerous lymphoid precursor cells from undergoing apoptosis.[16] Finally, the t(17:19) ALL subtype expressing the E2A-HLF fusion is resistant to chemotherapy and is associated with high mortality rates and poor prognosis.[18]

High expression of *HLF* has also been detected in the triple-mutated acute myeloid leukemia (AML) subgroup, defined by the presence of *NPM1c*, *DNMT3A* and *FLT3*-ITD mutations.[17] In fact, the *NPM1/DNMT3A/FLT3*-ITD mutated AML subgroup is commonly seen in AML (in 6-8% of all AML cases) and presents a poor prognosis, overall survival and poor relapse-free status for patients who specifically express high *HLF* levels.[17,27] It was previously hypothesized that *HLF* expression within this AML subgroup is highest specifically when all three mutations were present, rather than only one or two.[17] In addition, as revealed by data from triple-mutated AML relapse samples, 7 out of 8 cases revealed higher expression of *HLF* at relapse than at diagnosis of leukemia, suggesting its role in contributing to the development of this cancer.[17] Particularly, samples that had gained the most expression of *HLF* were *NPM1c/DNMT3A* mutated at diagnoses and acquired the *FLT3-ITD* mutation at relapse, suggesting that not only is this combination malignant but is associated with high expression of *HLF*.[17]

When investigating its role in leukemia, *HLF* has been shown to contribute to the leukemic stem cell (LSC) phenotype. *In vivo* experiments revealed that CRISPR/Cas9 knockout (KO) of *HLF* in triple-mutated samples, followed by transplantation of these samples into sub-lethally irradiated mice, resulted in a significantly lower $CD34^{+}GPR56^{+}$ compartment of cells taken from the bone marrow of engrafted mice, indicating a loss of LSCs.[17] A significant reduction of this compartment was observed in primary recipients and was even more pronounced in secondary recipients.[17] Furthermore, reintroduction of *Hlf* KO triple-mutated cells from engrafted mice back into culture resulted in cells that were more actively cycling compared to controls, but which quickly displayed a significant reduction in proliferative capacity (reduced cellular fold change), compared to controls.[17] Finally,

knockdown of *HLF* in cord blood significantly reduced engraftment levels in sublethally irradiated mice.[17]

HLF expression was also found to protect cells from drug-induced stress.[17] Equal numbers of *HLF* knockout leukemic cells were treated with the leukemia chemotherapy medications Cytarabine (AraC), Daunorubicin, and Etoposide, resulting in significantly higher sensitivity as well as fewer viable cell counts of *HLF* knockout cells compared to control cells.[17] Furthermore, single-cell transcriptome data from leukemic bone marrow revealed clustering of *HLF*$^{+}$ cells into a quiescent, immature, HSC-like cluster.[28] These results, in combination with the understanding of the role of *HLF* in HSC quiescence,[18] suggests that *HLF* also likely protects leukemic cells from chemotherapeutic and toxic insults by preserving their quiescent state.[18]

Overall, it was concluded that the loss of *HLF* impairs the proliferation of triple-mutant leukemic cells, the engraftment potential in leukemic mice models and that *HLF* strongly contributes to maintaining the leukemic phenotype of triple-mutated leukemia.[16,17,18]

1.4 The NPM1c-Mutated AML Genetic Network

The nucleophosmin (NPM1) protein, a key player in this triple-mutated AML subgroup, is a multifunctional nucleolar phosphoprotein that plays a role in cell growth, proliferation, DNA repair, transformation and genome stability by regulating critical tumour suppressors such as *p53* and *ARF*.[29] NPM1 also has shuttling properties that allow it to migrate between different cellular locations using localization signals, interact with proteins and nucleic acids, and alter the location of many proteins to regulate many processes within the cell.[24,30,31,32] The role of NPM1 in leukemia was in part originally detected by increased amounts of its protein levels in highly proliferating and cancerous cells.[33] Subsequently, a mutation that prevents its proper localization, resulting in its stalled and aberrant localization in the cytoplasm, was identified in leukemic blasts.[33] Mutant NPM1 resulting in its constant export to the cytoplasm was therefore appropriately named the NPM1c mutation.[34] In fact, *NPM1c* mutations have been found in about 35% of patients with AML, making *NPM1* one of the most frequently mutated genes in this leukemia.[33] *NPM1* mutations are now considered driver mutations of AML and represent a distinct category in the WHO classification of AMLs.[34,35] Despite *NPM1c*-mutated AMLs presenting a

relatively favourable prognosis, the presence of a coinciding *FLT3-ITD* mutation provides this leukemia with a less favourable prognosis.[29] Investigation into the order of these mutations revealed that *NPM1* mutations precede *FLT3-ITD* mutations, which might be explained by the interaction of NPM1 with signalling proteins and its involvement in regulating cell proliferation.[29,36] This idea is further supported by the finding that NPM1 is able to alter the cellular localization of the mutated *FLT3* tyrosine kinase domain (TKD), leading to aberrant cellular signalling that contributes to leukemia.[36]

As suggested by the extensive cellular processes that WT-NPM1 regulates, as well as the prevalence of mutated *NPM1c* in AML, it was thought that removal of mutant NPM1c could reduce the leukemic phenotype in cells. In fact, nuclear relocalization or induced degradation of NPM1c using CRISPR/Cas9 in two NPM1c-mutated cell lines (OCI-AML3 and IMS-M2) significantly affected the differentiation and growth of AML cells.[30] Removal of NPM1c resulted in significantly reduced cell counts of both cell lines and the appearance of mature monocytic morphology and granulocytic differentiation for OCI-AML3 and IMS-M2 cells, respectively.[30]

Due to the vast cellular processes that NPM1 regulates, the transcriptome of OCI-AML3 and IMS-M2 cells was also assessed three days after nuclear relocalization of NPM1c.[30] Focusing on genes that displayed at least a 2-fold downregulation, 16 genes were identified which met this requirement in both cell lines.[30] These genes belonged to the HOXA cluster, HOXB cluster, or *MEIS1* (referred to in summary as HOX/MEIS1).[30] Not only are the HOXA and HOXB cluster genes highly expressed in HSCs, they have been implicated in HSC self-renewal.[30,39,51,52] Overexpression of HOX genes has also been shown to expand the hematopoietic stem and progenitor cell (HSPC) population.[37,38,39] In addition to their role in HSCs, deregulation of the HOX/MEIS1 transcriptional network is strongly associated with leukemic phenotype.[40] Forced overexpression of both *Hoxa9* and *Meis1* in murine primary bone marrow cells resulted in the leukemic transformation of recipient mice.[37] Finally, an observed downregulation in HOX/MEIS genes after nuclear relocalization of NPM1c further highlights their role in the leukemic phenotype of NPM1c-mutated AML.[30] In conclusion, the NPM1 protein regulates the transcriptional network of

healthy hematopoietic cells and contributes to leukemic transformation,[30] making it a desirable target for therapies.

1.5 Positioning HLF within the NPM1c-Mutated AML Network

Interestingly, nuclear relocalization of NPM1c in IMS-M2 cells also resulted in a 2-fold downregulation of *HLF* in addition to HOX/MEIS genes.[30] Downregulation of *HLF* was not mentioned in this study because it was not downregulated in OCI-AML3 cells and because this study only focused on genes which were downregulated in both IMS-M2 and OCI-AML3 cell lines.[30] However, considering that *HLF* expression has been shown to be upregulated in NPM1c-mutant leukemia when the *FLT3-ITD* mutation is present, and that OCI-AML3 cells are *FLT3*-WT, the lack of this mutation may explain the lack of *HLF* downregulation in OCI-AML3 cells after NPM1c nuclear relocalization.[30,41]

Despite the specific nature of *HLF* upregulation, as well as its contribution to the leukemic state in NPM1c-mutated leukemia, little is known about the genetic network that regulates *HLF* expression. Specifically, it is currently still not understood why *HLF* is specifically upregulated in *NPM1c/FLT3-ITD* mutated leukemia and not in other AML subgroups. Considering the relevance of *HLF* in normal HSCs in addition to NPM1c-mutated leukemia, a better understanding of the differences between the molecular networks which regulate its expression in both systems may help better inform the ways in which this leukemia arises from a healthy hematopoietic system.

Knowing that *HLF* is upregulated specifically in *NPM1c/FLT3-ITD* mutated leukemia, we can hypothesize that upregulation of *HLF* expression is downstream of the *NPM1c* and *FLT3-ITD* mutations. Additionally, recognizing that *NPM1c* regulates HOX/MEIS1 gene expression, it is possible that HOX/MEIS1 genes may regulate *HLF* expression downstream of *NPM1c* (**Figure 2**). In fact, gene interaction studies in MLL-fusion leukemias have revealed *HLF* as a direct target of *MEIS1* regulation, and NUP98-Hox fusion AMLs have revealed *HOXA7* and *HOXA9* expression to be involved in *HLF* regulation.[42,43] Recognizing that this interaction is more complex than a unidirectional interaction between a few genes, we sought out to disentangle and define a preliminary network of genes we think could regulate *HLF* expression. Considering and including other homeobox genes in our hypothesized network, such as *NKX2-3* and PBX genes

which are involved in the NPM1-specific TF network, we strived to identify and test genes that could regulate *HLF*.[27,44]

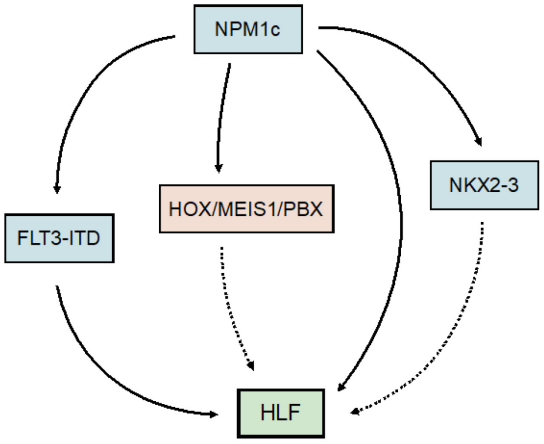


Figure 2. Hypothesized network of genes which regulate *HLF* expression in NPM1c-mutated AML. Solid arrows represent interactions that have already been demonstrated and dotted arrows represent our hypothesized interactions based on published literature.

1.6 CRISPR Engineering of a Genomic Reporter

Discovery of the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 gene editing system has allowed major advancements in experimental medicine in the context of gene investigations. Cas9-mediated double-stranded breaks (DSBs) at defined locations within the genome can create point mutations or large gene deletions through stimulation of error-prone non-homology end joining (NHEJ) repair mechanisms, or can insert entire genes through stimulation of homology directed repair (HDR) by providing a repair template containing the gene of interest.[3,45]

Taking advantage of CRISPR/Cas9-mediated gene insertion, in combination with recombinant adeno-associated viral vectors of serotype 6 (rAAV6)-mediated delivery of a repair template, the introduction of fluorescent reporter genes downstream of a gene of interest has been accomplished and demonstrated as a useful method for tracking gene expression.[3,13] After introducing the reporter cassette into cells and selecting for those that express it, other genes can be manipulated and the outcome on the expression of the gene of interest can be visually assessed using Fluorescent Activated Cell Sorting (FACS).[3,13] This CRISPR engineering technique avoids the cost and time associated with

gene expression quantification methods such as qPCR and has not shown to have any impact on cell viability or function when performed properly.[3,13] Furthermore, this technique is particularly useful in the context of studies that strive to identify the regulators of a specific gene, knowing that a full exploration of the genome, particularly by gene knockout, is a thorough approach to begin investigating the answer to this question. By using a fluorescent readout of gene expression, we can use FACS to quickly assess the potential regulators of a gene of interest.

When considering challenges associated with identifying the regulators of *HLF* expression in this study, such as the hypothesized involvement of multicomplex and master transcription factor HOX genes, the use of a fluorescent reporter and a CRISPR knockout approach as described above efficiently allows the identification of *HLF* regulators.

Engineering of an HLF reporter had recently been accomplished and validated in our laboratory in cord blood $CD34^+$ HSCs (**Figure 3**).[13] A CRISPR/Cas9 induced DSB in exon 4 of the *HLF* stop codon, followed by homologous recombination (HR) with a single-stranded repair template provided by rAAV6 viral transduction allowed the knock-in of a ZsGreen (ZsG) reporter cassette connected to the *HLF* open reading frame through an EMCV internal ribosome entry site (IRES).[13] P2A linkage of ZsG to either a puromycin resistance gene or a truncated EGF receptor allowed either antibiotic or antibody selection of reporter-expressing cells.[13] The final result is the co-expression of the HLF and ZsG to allow for visual tracking of *HLF* expression. This reporter was validated in *HLF* expressing (HepG2) and non-expressing (HEK293) cell lines. Droplet digital (ddPCR) was used to confirm homologous recombination and integration of the reporter cassette. Western analysis was used to confirm HLF expression in HLF expressing cell lines and no difference in protein expression between targeted and untargeted cells.[13] After validation, the HLF-ZsG reporter was introduced into *HLF*-expressing IMS-M2 leukemic cells to begin investigating the regulators of *HLF*.

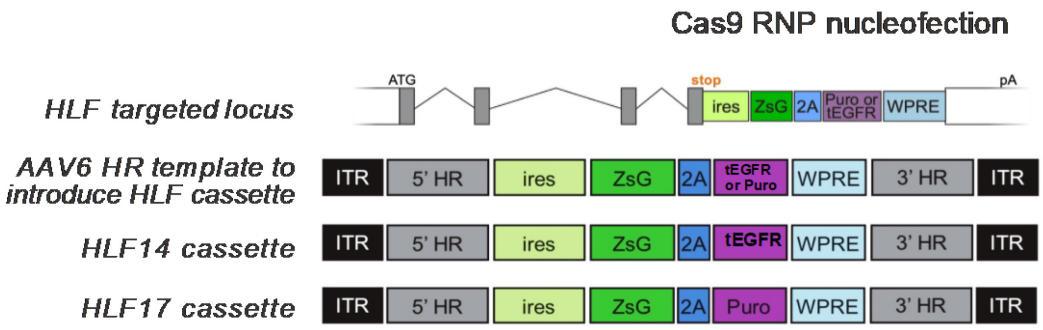


Figure 3. HLF-ZsG reporter. Adapted from Lehnertz et al. (2021).

General Rationale, Hypothesis and Objectives

The relevance of *HLF* expression in a hematopoietic and leukemic context makes it a fascinating point of study to identify the differences in the genetic networks of regulation between these two systems. In addition to identifying the primitive HSC population, *HLF*-expression is associated with functional HSC activity and declines as cells differentiate.[13] Curiously, *HLF* is also specifically expressed in *NPM1c/FLT3-ITD*-mutated leukemia.[17] The development of a ZsG reporter cassette, validated in HSCs, provides a reliable readout of *HLF* expression and a method to investigate the genes which regulate it.[13] Based on the specific and relevant expression of *HLF* in HSCs, I hypothesize that *HLF* expression in a leukemia cell line (IMS-M2), assessed in real time using a ZsG reporter, can be used as a readout to identify the genes or pathways that also regulate *HLF* in normal HSCs. We chose the *NPM1c*-mutated IMS-M2 cell line based on previous studies which looked at nuclear relocalization of the NPM1c protein and the subsequent transcriptomic changes of HOX/MEIS1 genes.[30] This cell line is an easily modifiable and inexpensive resource which will prove to be invaluable in the context of a small CRISPR screen to identify the regulators of *HLF*.

Objectives:

1. Validate the HLF-ZsG reporter in an AML cell line (IMS-M2).
2. Identify the putative regulators of *HLF* in the IMS-M2 reporter cell line.
3. Compare the regulators of *HLF* in a leukemic cell line to those in normal HSCs.

CHAPTER 2- RESULTS

2.1 Leucegene Analysis of *HLF* Expression in AML

Given *HLF*'s specific expression in HSCs, we hypothesized that it might also be specifically expressed in certain AMLs, as was initially demonstrated by Garg et al. (2019).[17] To address this hypothesis, we used transcriptome data from the Leucegene project (www.leucegene.ca) to determine its expression in various genetic subgroups of AML.[3,46,47,48] Based on previous claims in published literature on the dependency of *HLF* expression on *DNMT3A* and *FLT3-ITD* mutations,[17] we analyzed how its expression varied among these subgroups when one, both, or none of these mutations were present. Analysis of data from the Leucegene project revealed the highest expression of *HLF* within the normal karyotype (NK) *NPM1c* subgroup harbouring the *FLT3-ITD* mutation specifically **(Figure 4)**.

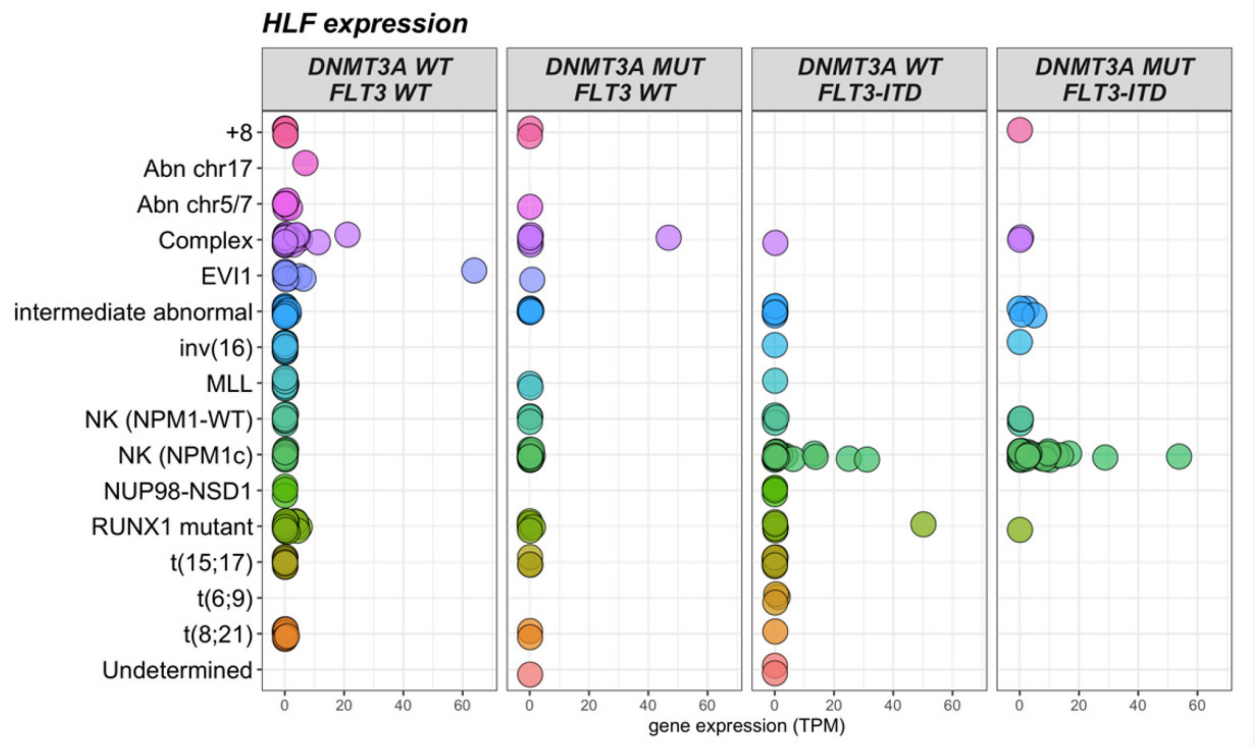


Figure 4. Analysis of *HLF* expression across different subgroups of AML harbouring either *DNMT3A* or *FLT3-ITD* mutations, none, or both using leucegene

cohort data. (n = 8 for +8, n = 1 for Abn chr17, n = 9 for Abd chr5/7, n = 62 for Complex, n= 11 for EVI1, n = 48 for intermediate abnormal, n = 28 for inv(16), n = 35 for MLL, n = 20 for NK (NPM1-WT, n = 97 for NK (NPM1c), n = 7 for NUP98-NSD1, n = 48 for RUNX1 mutant, n = 30 for t(15;17), n = 3 for t(6;9), n = 20 for t(8;21), n = 3 for Undetermined)

2.2 Transcriptomic Analysis of NPM1c Knockout IMS-M2 Cells

Given *HLF's* expression in *NPM1c/FLT3-ITD* AML,[17] as well as the role that NPM1c holds in driving leukemogenesis,[30] we hypothesized that *HLF* expression is directly dependent on oncogenic *NPM1c* in this context. Interestingly, Brunetti et al. (2018) published an approach where *NPM1c* expression could be specifically disrupted using CRISPR/Cas9 to study *NPM1c* dependent gene expression.[30] To address our hypothesis of the dependency of *HLF* on *NPM1c*, transcriptomic data of IMS-M2 cells published by Brunetti et al. (2018)[30] was downloaded and further analyzed to determine the genes that were highly downregulated after CRISPR/Cas9 knockout of mutant *NPM1c* in this cell line **(Figure 5)**. Normalized gene expression (quantified in transcripts per million; TPM) analysis revealed a significant decrease in *HLF* expression after knockout of NPM1c compared to knockout of CD19 and CD45. The selection and analysis of other genes from transcriptomic data published by Brunetti et al. (2018)[30] allowed us to assess their downregulation in the IMS-M2 cell line specifically. Knowing that *HLF* is not expressed in OCI-AML3 cells, their transcriptome profile was not of interest to us.[30] Furthermore, knowing that *HLF* is expressed in the IMS-M2 NPM1c-mutated leukemic cell line, an examination of the genes which are downregulated after NPM1c knockout in IMS-M2 cells specifically allows us to begin addressing the potential regulators of *HLF*.

As previously stated, mutant NPM1c has been shown to displace other proteins to the cytoplasm when it is also relocated there.[32] In particular, NPM1c has specifically been shown to interact with, and displace, the master transcription factor PU.1, encoded by the *SPI1* gene, causing large amounts of PU.1 to accumulate in the cytoplasm.[48] The expression *SPI1*, was therefore also included in our assessment of genes that were downregulated after NPM1c knockout in IMS-M2 cells. *FLT3* was selected in this analysis due to the relevance of *HLF* expression in *NPM1c/FLT3-ITD* mutated leukemia.[17] Furthermore, knowing that internal tandem duplication (ITD) mutations of the *FLT3* gene

(*FLT3-ITD*) have been shown to activate aberrant expression of *STAT5*, and that aberrant activation of this pathway has been recognized in leukemogenesis, *STAT5A* and *STAT5B* genes were included in this analysis as well.[49,50]

We also examined the expression of several homeobox genes (*HOXA9*, *HOXB8* and *MEIS1*) after *NPM1c* knockout because they presented more than 40% downregulation after loss of *NPM1c* in IMS-M2 cells, shown in published data.[30] In addition, previously published literature has identified *HLF* to be a direct target of *MEIS1* in MLL-fusion leukemias and a target of *HOXA7*, *HOXA9* and *HOXA10* mediated regulation in NUP98-Hox fusion AMLs. [42,43] *NKX2-3* was included for analysis because of its overexpression in NPM1c-mutated AML.[27] Furthermore, binding sites for HOXA10 have been identified in *HLF*, suggesting that it could also be a potential candidate regulator of *HLF*. [51] Homeobox genes *PBX1* and *PBX3* were included as well due to their ability to complex with HOX proteins, including MEIS1, to increase their DNA binding specificity.[53,54] Specifically, it has been discovered that MEIS1 requires cooperation with PBX1/PBX3 and HOXA9 in order to bind to DNA target sites, deregulate cell proliferation and hematopoietic differentiation and contribute to leukemogenesis.[54,55,56]

The CD34 and CD38 markers were included in our analysis due to the discovery of the NPM1c mutant protein in the cytoplasm of $CD34^+$ and $CD38^+$ leukemic cells, and the ability of $CD34^+$ NPM1c-expressing cells to initiate leukemia in immunocompromised mice.[57] We also included the *ADGRG1* gene, also known as the GPR56 marker, due to the specific detection of its expression in human LSCs in most AMLs.[46] Finally, we included the DNA replication protein RPA1 due to the reported role of NPM1c in DNA replication repair and the involvement of this protein in DNA replication.[58]

Transcriptomic data revealed significant downregulation of *HLF*, *HOXB8,* and *MEIS1* genes after knockout of NPM1c compared to control knockouts of sgCD19 and sgCD45 in IMS-M2 cells **(Figure 5)**. This finding allowed us to define a list of genes that are regulated by NPM1c and subsequently could be potential regulators of *HLF*. This data also confirms that *HLF* expression is regulated by NPM1c expression in IMS-M2 cells, as displayed by a decrease in HLF transcripts after sgNPM1c compared to sgCD19 and sgCD45.

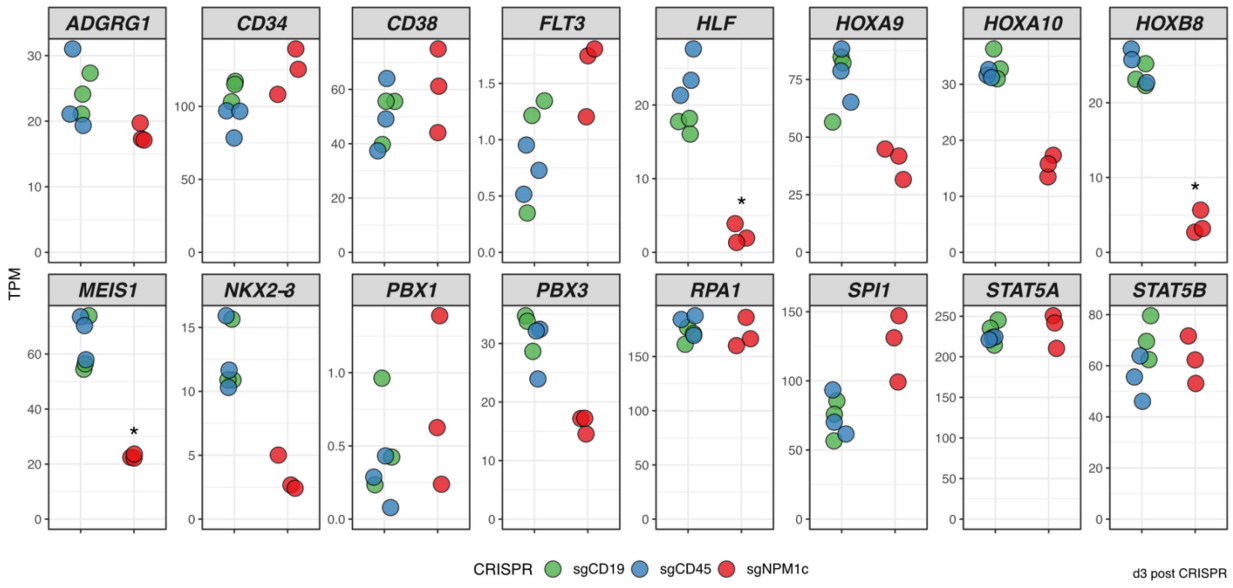


Figure 5. Transcriptomic analysis of IMS-M2 cells reveals expression of candidate regulators of *HLF* expression after sgNPM1c knockout compared to sgCD19 and sgCD45. N=3 biological replicates for sgCD19, sgCD45, and sgNPM1c. TPM = transcripts per kilobase million. * represents significant downregulation (q value ≤ 10^{-5} & log2(fold-change) > 1) for controls (sgCD19 & sgCD45) vs. sgNPM1c based on Kallisto/Sleuth analysis.[59]

2.3 Cells Targeted to Express a ZsGreen (ZsG) Fluorescent Reporter Downstream of HLF Allow Real Time Tracking of HLF Expression

Since our group has developed an HLF reporter in HSCs, we hypothesized that this same reporter could be introduced in a leukemic cell line to facilitate the identification of upstream regulators of *HLF* in leukemic cells. To accomplish this, CRISPR/Cas9 editing of the *HLF* locus in IMS-M2 leukemic cells was used to facilitate the introduction of the ZsG protein downstream of the *HLF* open reading frame (ORF). A double-stranded (DS) induced break at the end of the *HLF* ORF allowed homologous recombination (HR) with a repair template including a reporter cassette expressing either a truncated EGFR

(tEGFR) or puromycin resistance gene for selection of targeted cells, similar to our strategy used in HSCs.[13] The reporter cassette with a tEGFR is also noted as the HLF14 reporter and the cassette that includes a puromycin resistance gene is also noted as the HLF17 reporter to distinguish the two cassettes. These reporter cassettes are identical excluding their method of selection **(Figure 6A)**. Selection of HLF17-targeted cells using puromycin resulted in a 97% population of *HLF* expressing cells **(Figure 6B)**. Release of puromycin selection resulted in a reduction in HLF-ZsG expression to 80% after two weeks and 50% after 6 weeks.

HLF14 and HLF17 expressing cells were sorted into *HLF* expressing and non-expressing populations using FACS sorting for ZsG expression. Cells were sorted on a BD FACSAria II cell sorter (BD Bioscience) and both *HLF* expressing (ZsG+) and *HLF* non-expressing (ZsG-) were collected after the sort and put back in culture. After FACS sorting, antibody enrichment was used to select for cells that had integrated the HLF14 reporter cassette and media supplementation with puromycin was used to select for cells that had integrated the HLF17 reporter cassette. Selection was based on the principle that if cells were expressing the HLF14 cassette, they would also be expressing a truncated EGFR (tEGFR) receptor, which could be enriched for using antibodies. Similarly, if cells were expressing the HLF17 cassette, they would also be expressing the puromycin resistance gene and could be enriched for using media supplementation.[13] For HLF17-targeted cells, puromycin was removed from the media and *HLF* expression was recorded over time **(Figure 6C)**. It was found that HLF-ZsG expression in the *HLF* expressing populations was gradually lost over time for cells that had been targeted with either the HLF14 or HLF17 reporter. In addition, as expected, cells that had been targeted for the HLF14 or HL17 reporter and sorted into non-expressing populations did not display an acquisition in HLF-ZsG expression over time **(Figure 6C)**.

Together, this data suggests that, in addition to targeting of HSCs, introduction of an HLF-ZsG reporter could be accomplished in a leukemic cell line. Introduction of this reporter strategy into a leukemic cell line provides a cost and time effective method to identify the regulators of *HLF* expression.

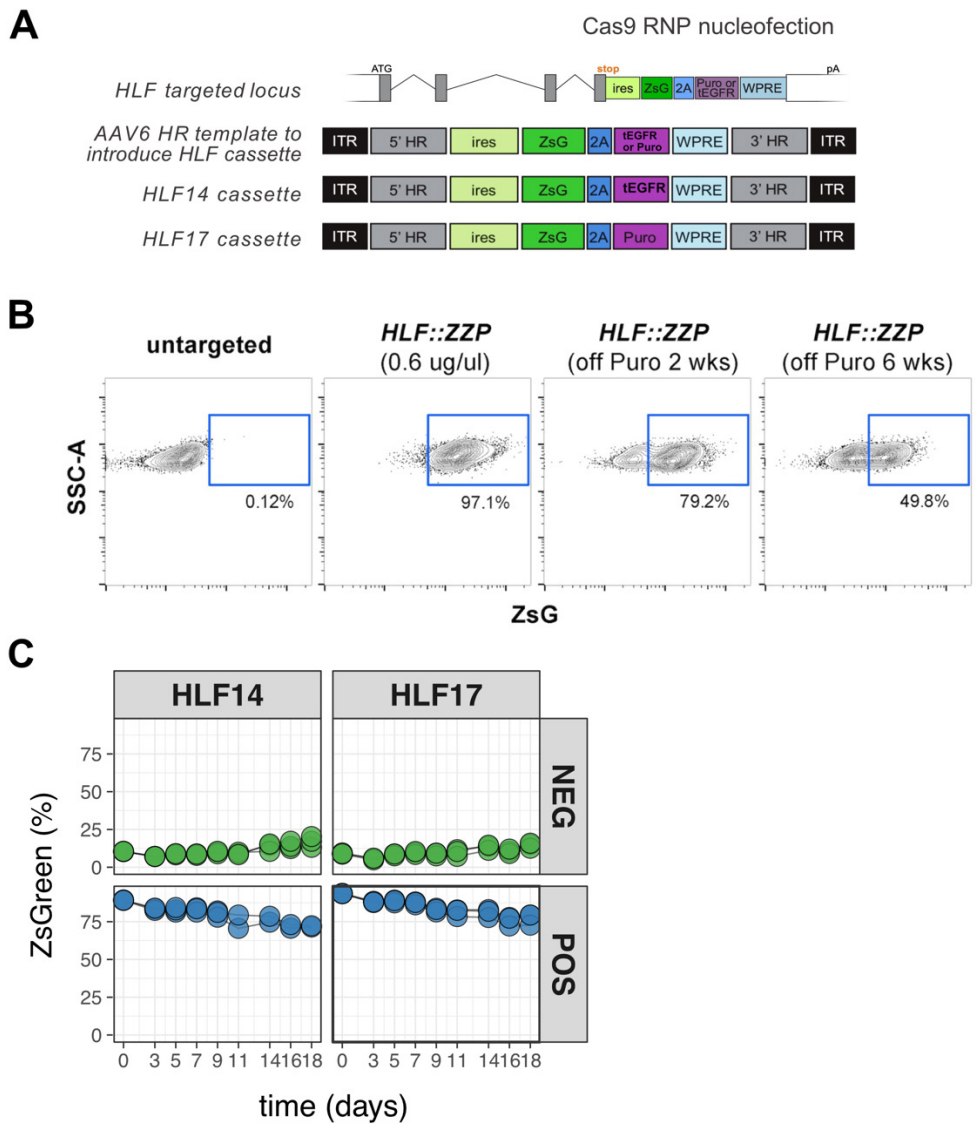


Figure 6. IMS-M2 cells targeted to express a ZsG fluorescent reporter downstream of *HLF* allowed real-time tracking of *HLF* expression. (A) Figure adapted from Lehnertz et al. 2021.[13] A DSB at the stop codon in exon 4 of the ORF of *HLF* allowed HR with a repair template including antibody selection using tEGFR (HLF14) or antibiotic selection using puromycin-supplemented media (HLF17).[13] Introduction of the HR repair template was provided using rAAV6-mediated transduction.[13] (B) Targeting of cells with the HLF17 reporter and selection using 0.6ug/ul of puromycin allowed assessment of ZsG expression over time using flow cytometry. (C) Cells were targeted with the HLF14 or HLF17 cassette and sorted for ZsG expression. ZsG expression was recorded every other day for 18 days.

2.4 Neither Integration of the HLF-ZsG Reporter nor Expression of HLF Affects Proliferation of IMS-M2 Cells

Since *HLF* expression in reporter targeted cells was heterogeneous, and because *HLF* has been implicated in AML cell survival and proliferation,[4,7,18,20,22,26] we hypothesized that *HLF* expressing and non-expressing cells might exhibit different proliferation behaviours. To address this hypothesis, IMS-M2 cells were targeted with either the HLF14 or HLF17 reporter and sorted into *HLF* expressing (positive) and *HLF* non-expressing (negative) populations using ZsG expression. These cells were then expanded and frozen into aliquots for all future experiments. HLF17 expressing cells were then selected for by supplementing media with puromycin. After targeting and sorting cells, we calculated cumulative population doublings for each population to determine the impact of the HLF-ZsG reporter or of *HLF* expression on cell proliferation. Cumulative population doublings of HLF14 and HLF17 positive and negative cells revealed very comparable profiles for these four populations **(Figure 7A)**. In addition to IMS-M2 cells that had been targeted with the HLF-reporter and sorted based on ZsG expression, untargeted cells also displayed identical cumulative population doubling dynamics **(Figure 7B)**. This data led us to conclude that neither genomic integration of the HLF-ZsG reporter, nor *HLF* expression, affects proliferation of IMS-M2 cells. This finding parallels the fact that OCI-AML3 cells do not express *HLF* and can still proliferate, suggesting that HLF is not entirely essential for this cell function.[30]

The fact that integration of the reporter did not affect proliferation of IMS-M2 cells allowed us to confidently move forward with experiments using this reporter; knowing that any changes in proliferation dynamics would be due to further manipulation of cells, and not from the introduction of the reporter. Furthermore, knowing that *HLF* expression does not affect proliferation of IMS-M2 cells meant that if we removed the expression of a potential *HLF* regulator and IMS-M2 cells exhibited a decrease in proliferation, then this was because expression of the potential *HLF* regulator was removed, and not because HLF expression was decreased.

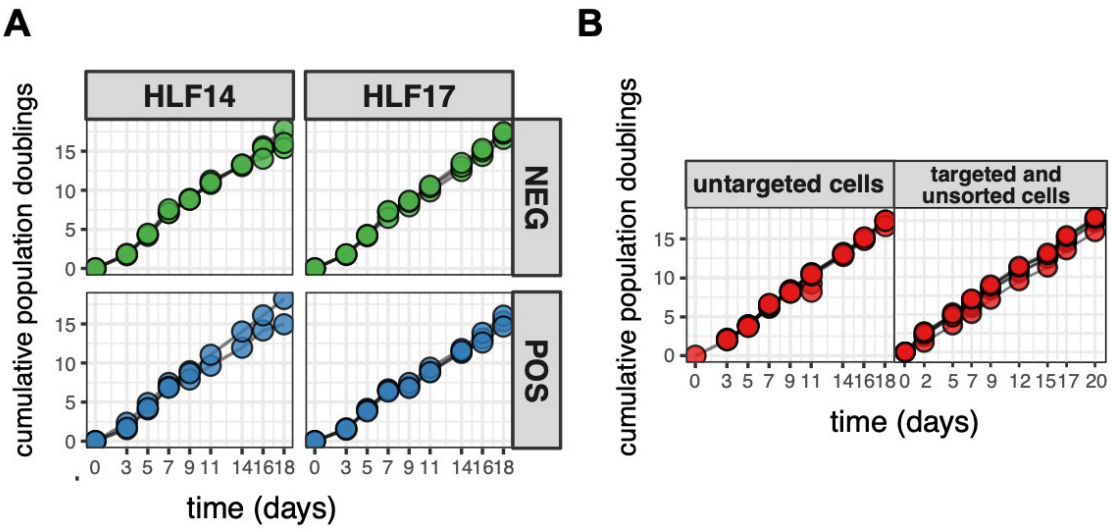


Figure 7. Neither the HLF-ZsG reporter nor the expression of HLF affects cell proliferation of IMS-M2 cells. (A) IMS-M2 cells were targeted with either the HLF14 or HLF17 reporter and sorted based on ZsG expression to determine cumulative population doubling patterns for *HLF* reporter expressing and non-expressing cells. (B) Untargeted cells, as well as targeted and unsorted IMS-M2 cells, were also assessed for cumulative population doublings over time.

2.5 Role of HLF as a Stem Cell Gene

Although *HLF* expressing and non-expressing cells demonstrated comparable proliferation kinetics *in vitro*, we hypothesized that differences in AML initiation and progression could be detected between these two populations. To test this, we transplanted HLF14 and HLF17 expressing (positive), and non-expressing (negative) populations into NSG mice and assessed disease latency (days), engraftment (%) and ZsG positivity (%). Results revealed no difference in transplantability or disease latency between positive and negative populations **(Figure 8A)**, however there was a slight difference in bone marrow engraftment revealed at experiment endpoint (week 4) with *HLF*-expressing cells displaying a slightly lower engraftment percentage **(Figure 8B)**. Lastly, ZsG positivity (%) was comparable to *in vitro* data in terms of a partial loss of ZsG percentage observed in the positive transplanted sub-fraction, with percentages starting at 80-90% and ending between 21.1-49.4% at endpoint (excluding an outlier of 7.52) **(Figure 8C)**. From this data, we can conclude that despite its role as a stem cell gene in

HSCs and some reported instances of triple mutated AML, *HLF* does not seem to hold the role of a stem cell gene in IMS-M2 cells. However, repeating the sorting and transplantation of *HLF*-expressing and non-expressing cells may be necessary to substantiate this finding.

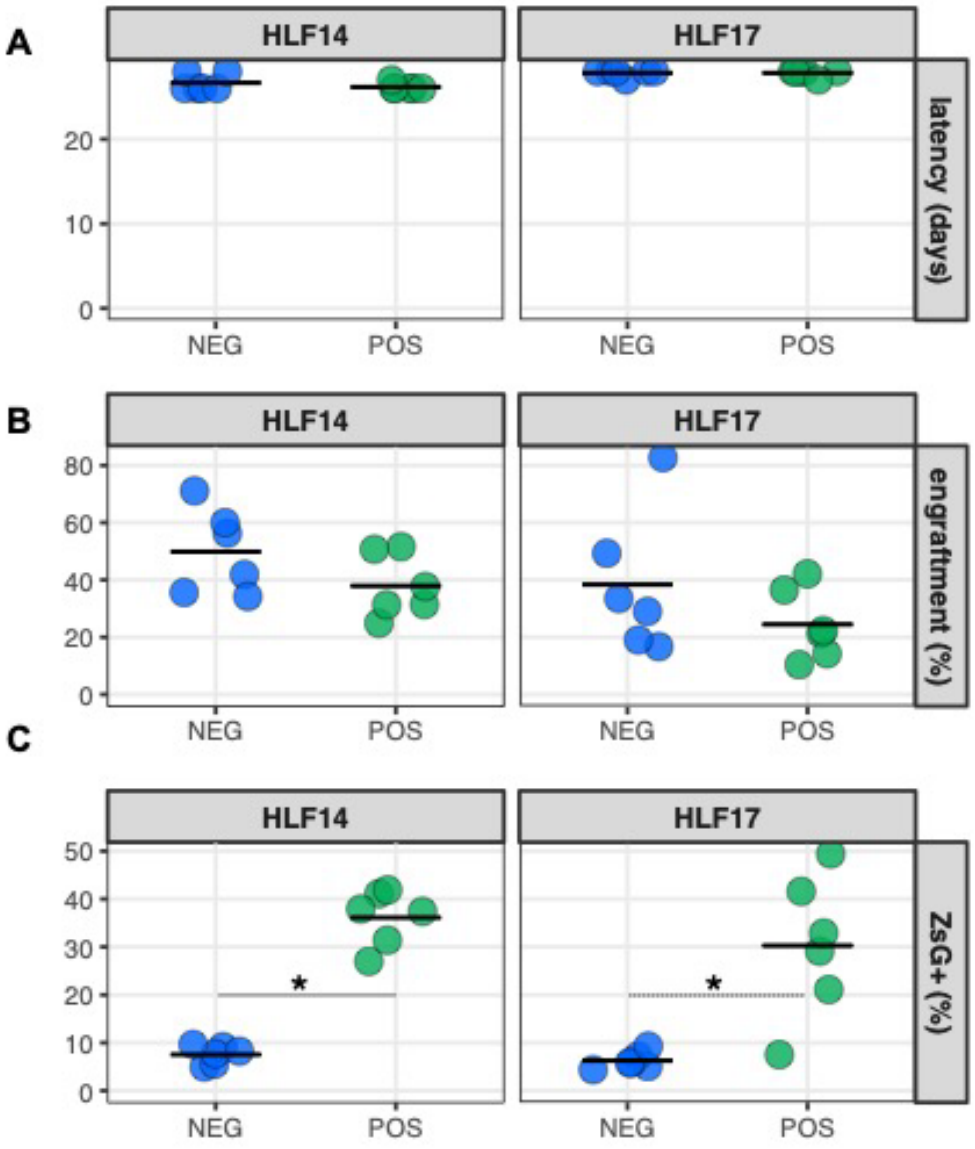


Figure 8. HLF expression does not correlate with transplantability or stem cell properties in IMS-M2 cells. (A) Disease latency (days) between *HLF* positive (pos) and negative (neg) populations. (B) Bone marrow engraftment (%) between *HLF* positive and negative populations at endpoint. (C) Percentage of ZsG expression within the positive-expressing sub-fraction of IMS-M2 cells at experiment endpoint. Each dot represents a

biological replicate. Asterisks represent a statistically significant difference between *HLF* positive (pos) and negative (neg) populations. Mean is denoted by a solid cross-bar. *p-value < 0.05 (Mann-Whitney U test, two-tailed).

2.6 NPM1c Knockout Induces Loss of *HLF* Expression, Reduced Proliferation and Differentiation in IMS-M2 Cells

After successfully introducing an HLF-ZsG reporter into IMS-M2 cells and establishing that this reporter does not affect proliferation of IMS-M2 cells, we hypothesized that we could knock out the expression of a known regulator of *HLF* and visualize a decrease in *HLF* expression through a reduction in ZsG fluorescence. Based on findings of upregulated *HLF* expression in triple-mutated AML,[17] and after confirming the dependency of *HLF* expression on the NPM1c mutation in IMS-M2 cells using transcriptomic data, we decided to first test the effect of knocking out the NPM1c mutation on *HLF* expression.

To accomplish this, mutant NPM1c was knocked out in IMS-M2 cells using a single guide RNA (sgRNA) designed by *Brunetti et al*. (2018)[30] which takes advantage of the four base pair insertion present within this mutation to target and knockout cells expressing the mutation. CRISPR/Cas9 knockout within the “safe-harbor” adeno-associated virus integration site 1 (AAVS1) locus in the same cell line was included as a control. The AAVS1 locus is found in intron 1 of the PPP1R12C (protein phosphatase 1 regulatory subunit 12C) gene. AAVS1 is considered a “safe-harbor” locus because genetic interference within this locus does not disrupt the molecular makeup of the cell.[60] Nucleofection of a Cas9/sgRNA complex including either sgAAVS1 or sgNPM1c was conducted using HLF17 targeted cells. After the release of puromycin selection, cellular fold expansion **(Figure 9A)**, cumulative population doubling **(Figure 9B)** and ZsG expression **(Figure 9B)** were tracked every other day for 11 days.

While sgAAVS1 cells modelled an exponential increase in fold expansion (600-fold increase by day 11), the fold expansion of sgNPM1c cells was greatly stunted and only reached a 200-fold expansion by day 11 **(Figure 9A)**. Similarly, despite a gradual increase in the cumulative population doublings of sgAAVS1 cells, sgNPM1c exhibited a gradual decline in cumulative population doublings after day five post nucleofection **(Figure 9B)**.

After the release of puromycin selection at day zero, ZsG expression gradually decreased in sgAAVS1 cells as expected, reaching around 75% by day 11 **(Figure 9B)**. On the contrary, ZsG expression in NPM1c knockout cells was drastically reduced at day one post-nucleofection until almost complete depletion by day seven. Violin plots of HLF-ZsG expression displaying a side-by-side comparison revealed a drastic loss of HLF-ZsG expression in NPM1c knockout cells (red) compared to AAVS1 knockout cells (blue) starting at day three post-nucleofection **(Figure 9D)**.

Finally, analysis of the morphological characteristics of NPM1c knockout cells revealed the presence of myeloid differentiation day nine post-nucleofection compared to AAVS1 knockout cells which did not display these properties **(Figure 9E)**.

In conclusion, the drastic reduction in ZsG expression after knocking out NPM1c, a known regulator of *HLF* expression, provided us with the confidence that the HLF-ZsG reporter could accurately report on a decrease in *HLF* expression when one of its regulators was removed. The substantial depletion in fold expansion exhibited by NPM1c knockout cells compared to control knockout cells can be explained by the roles that NPM1c holds in leukemic cell growth and proliferation.[30] Furthermore, the appearance of differentiation morphology after NPM1c knockout in IMS-M2 cells is consistent with published literature in which NPM1c affects differentiation of leukemic cells.[30] The expected phenotype of reduced cell proliferation and differentiation characteristics further provided us with confidence that this technique effectively produced a knockout. Finally, by establishing successful nucleofection parameters in IMS-M2 cells through NPM1c knockout, we could use these parameters to assess the outcome on *HLF* expression after knocking out other genes as well. The surface marker CD45 was knocked out using these exact same parameters to further test the efficacy of this technique. CD45 expression was stained seven days after nucleofection with and without Cas9. See **Appendix A** for flow cytometry profiles.

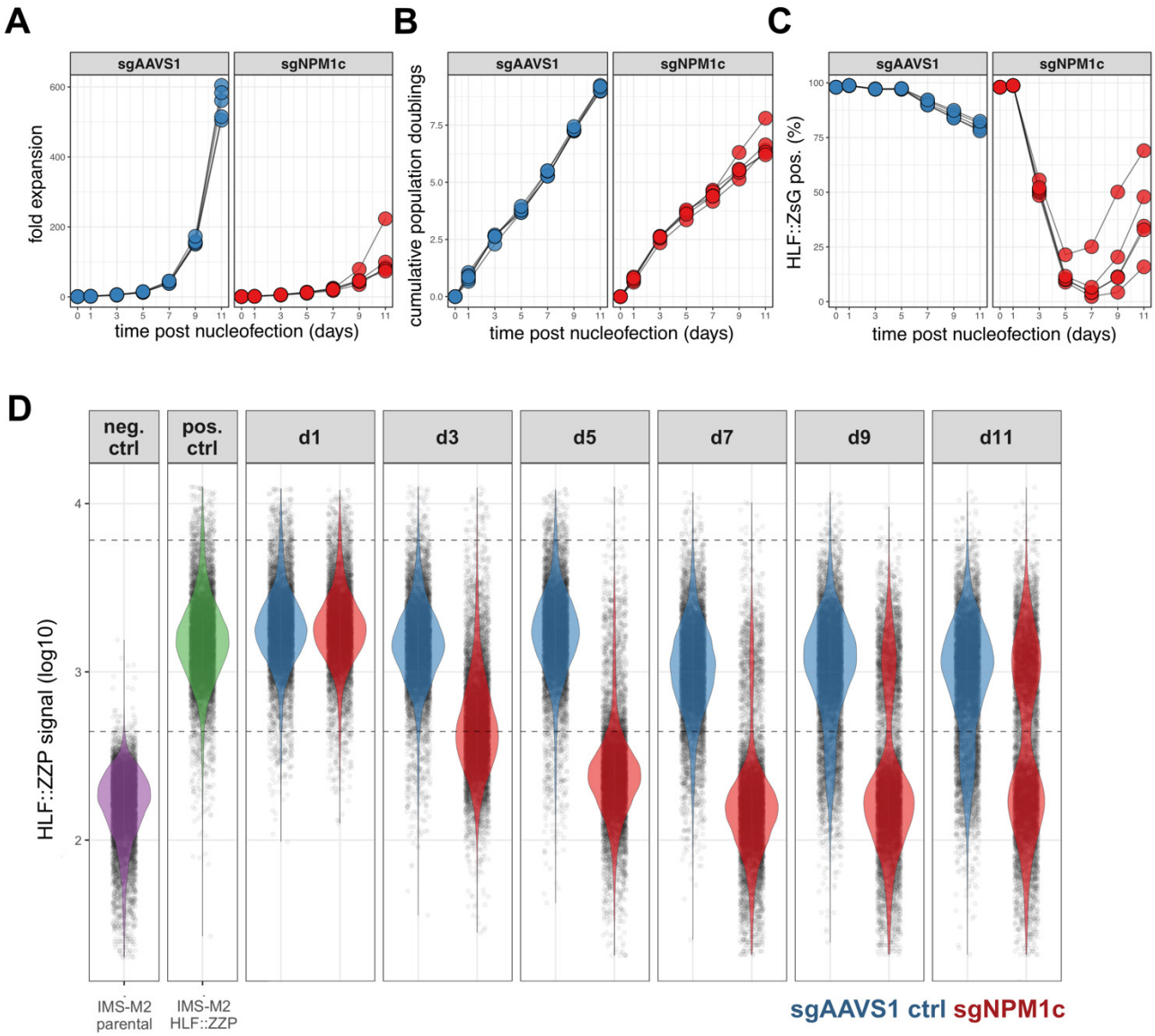

A
B
C
sgAAVS1
sgNPM1c
fold expansion
cumulative population doublings
HLF::ZsG pos. (%)
time post nucleofection (days)
D
neg. ctrl
pos. ctrl
d1
d3
d5
d7
d9
d11
HLF::ZZP signal (log10)
IMS-M2 parental
IMS-M2 HLF::ZZP
sgAAVS1 ctrl sgNPM1c
E
sgAAVS1
sgNPM1c

Figure 9. NPM1c Knockout Induces Loss of *HLF* Expression, Reduced Proliferation and Differentiation of IMS-M2 Cells. (A) Fold-expansion profiles of sgAAVS1 compared to sgNPM1c knockout cells. (n=5 for sgAAVS1 and sgNPM1c) (B) Cumulative population doublings between sgAAVS1 and sgNPM1c knockout cells. (C) ZsG expression in sgAAVS1 cells compared to NPM1c knockout cells. (D) Violin plots allow a side-by-side comparison of NPM1c and AAVS1 knockout cells at each timepoint. (E) Wright staining of NPM1c and AAVS1 knockout cells reveals the appearance of azurophilic granules, examples indicated by black arrows in NPM1c knockout cells compared to AAVS1 knockout cells.

2.7 shRNA-Mediated Knockdown Suggests MEIS1 as a Regulator of HLF

Based on the reduction in *HLF* expression after removing one of its known regulators, we hypothesized that we could use our established HLF-ZsG reporter cell line to identify other factors regulating *HLF* by knocking down their expression and observing changes in ZsG fluorescence.

To test this, shRNA vectors designed against Luciferase (Luc), *HLF, NKX2-3, HOXA9* and *MEIS1* were transduced into HLF17-expressing cells that had been selected using puromycin. These candidates were selected based on analysis of previously published literature which demonstrated a reduction in their expression after knockout of the mutant NPM1c protein, as described previously. shRNA hairpins were cloned into a vector that contain the fluorescent marker Ametrine to allow for the assessment of shRNA-expressing cells by flow cytometry **(Figure 10A**). Three different shRNAs per gene were designed, cloned, and transduced into cells (excluding Luc in which only one shRNA was included). Each shRNA targeted a different region of the transcript. Before infection with shRNA viral vectors, ZsG levels were recorded on the day of infection using a BD Biosciences Canto II flow cytometer to confirm the majority of cells were expressing *HLF* and its reporter.

Upon infection, puromycin selection was released to allow the observance of any potential changes in *HLF* expression. Day three post-infection, ZsG and Ametrine expression levels, as well as cell counts, were recorded to determine the number of cells per mL.

These parameters were recorded again every other day for 17 days for seven time points in total. The percentage of Ametrine-expressing cells was used as an indicator of cell proliferation **(Figure 10B)** whereas the percentage of ZsG-expressing cells within the Ametrine fraction was used as an indicator of *HLF* expression **(Figure 10C)**.

shLuc revealed consistent levels of Ametrine expression over time and a slight and gradual but expected loss of *HLF-ZsG* expression consistent with the release of puromycin selection. The three shRNAs targeting *HLF* did not impact proliferation or the percentage of Ametrine-expressing cells, however they did result in a drastic loss of *ZsG* expression compared to shLuc, as expected. All three knockdowns of NKX2-3 resulted in a small decrease in Ametrine expression, and one resulted in the loss of ZsG similar to the shHLF conditions. shRNA candidates against *HOXA9* and *MEIS1* resulted in consistent levels of Ametrine expression. The three shRNAs against *HOXA9* did not lead to a decrease in HLF-ZsG expression above what was expected due to the release of puromycin selection, whereas two of the three candidates against *MEIS1* resulted in a marked decrease in HLF-ZsG expression over time, producing a phenotype very comparable to the knockdown of *HLF*.

The difference in ZsG expression within the Ametrine positive and negative fractions for each knockdown of candidate HLF regulators at day ten was compared to these same profiles for shLuc at day three to determine the extent to which each shRNA resulted in a reduction of ZsG expression **(Figure 10D)**. This calculation revealed that shHLF and shMEIS1 result in a marked decrease in ZsG expression in the Ametrine positive population compared to the Ametrine negative fraction.

In conclusion, primary evidence suggests that MEIS1 regulates *HLF* expression without impacting cell proliferation. In addition, unchanged proliferation levels after knockdown of *HLF* confirms that *HLF* expression and cell proliferation in this cell line are not necessarily linked. Since the length of this screen is very long (17 days), we can be confident that it accurately assessed proliferation dynamics. Finally, a reduction in ZsG levels after knockdown of *HLF* also confirms that our reporter can relay *HLF* expression. Recognizing however that shRNA-mediated gene knockdown does not completely remove the expression of a gene,[61] we next hoped to confirm the phenotypes realized from our shRNA

approach and introduce more potential *HLF* regulators in a stepwise approach through conducting a small CRISPR/Cas9 knockout screen.

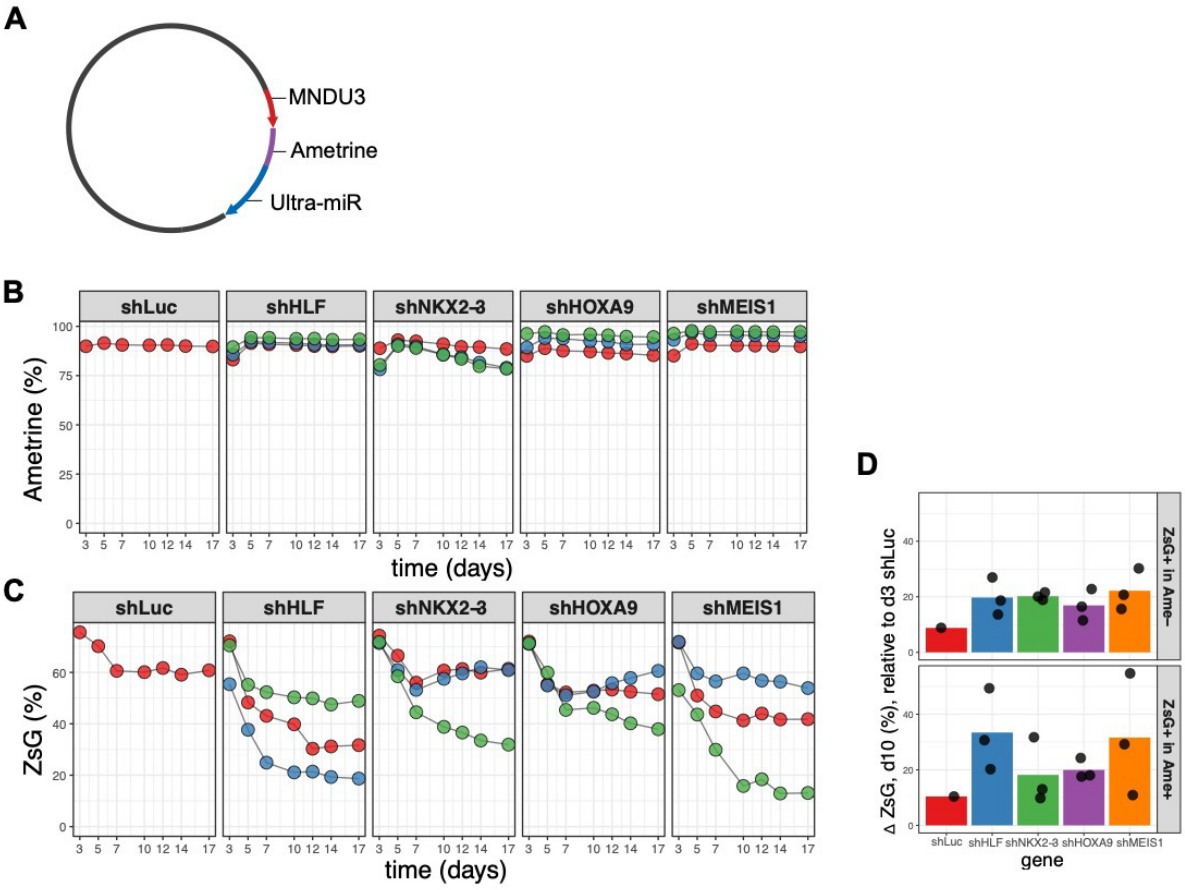


Figure 10. shRNA-Mediated Knockdown Suggests MEIS1 as a Regulator of *HLF*. (A) Schematic of the vector in which shRNA hairpin designs were cloned into, including Ametrine as a fluorescent marker. (B) Ametrine-expression dynamics over time allowed the assessment of proliferation of shRNA-expressing cells. (C) ZsG expression dynamics over time allowed assessment of *HLF* expression after knockdown of putative regulators and controls. (D) Percentage of ZsG within Ametrine-expressing and non-expressing cells at day 10 of knockdown for each putative HLF regulator was compared to shLuc cells at day 3. N=3 different shRNAs per gene, N=1 shLuc. One biological replicate per shRNA was performed and is shown in the figure. Statistical analyses could not be performed as data from only one biological replicate for 1 shRNA targeting Luciferase is available.

2.8 CRISPR Knockout Screen Reveals Regulators of HLF

To expand the list of potential *HLF* regulators, we then hypothesized that a small arrayed CRISPR knockout screen would allow us to confirm previously established HLF regulators, and identify new ones. In comparison to shRNA-mediated knockdown, which partially reduces expression of the gene through degradation of mRNA levels and is known to be subject to off-target effects,[61] a CRISPR/Cas9 knockout screen was selected for this experiment due to its ability to completely eliminate gene expression at the DNA level.[62] In this technique, single guide RNAs (sgRNAs) are used to direct a Cas9-induced DSB at a desired gene.[62] The cell must then rely on error-prone NHEJ repair mechanisms to repair this break, which often results in inserted or deleted bases and frameshift mutations that alter the coding frame of the gene and produce a non-functioning gene (gene knockout).[62] While shRNA-knockdown of a few potential regulators of *HLF* allowed us to test their ability to regulate *HLF*, a conclusive a CRISPR knockout screen provided us with more confidence to identify *HLF* regulators by completely knocking out their expression, rather than silencing it.[62] Recognizing that a genome-wide CRISPR knockout screen would quickly allow us to assess all potential regulators of *HLF*, attempts to generate a Cas9-inducible IMS-M2 cell line for this screen were made. However, an inability of IMS-M2 cells to take up the Cas9 protein rendered this experiment unsuccessful so far. Nevertheless, methylcellulose plating of IMS-M2 cells with human cytokines resulted in the production of IMS-M2 clones, suggesting that the production of a Cas9-inducible IMS-M2 cell line could be possible, and should be attempted again in the future (**Appendix D**).

We hypothesized that we could conduct a small, arrayed CRISPR knockout screen to identify genes that regulate HLF expression. CRISPR/Cas9-mediated knockouts of *AAVS1, ADGRG1, FLT3, HLF, HOXA10, HOXB8, MEIS1, NKX2-3, NPM1c*, *PBX1, PBX3, RPA1, SPI3, STAT5A, STAT5B* and *ZsG* were then conducted in an arrayed format to continue to identify the regulators of *HLF* **(Figure 11A)**. Three sgRNAs targeting different regions of each gene were designed using Synthego to avoid off-target effects. These guide RNAs were then cloned into a lentiviral vector which included a red fluorescent protein (RFP) fluorescent marker to allow for the visualization of sgRNA-expressing cells by flow cytometry **(Figure 11B)**. The RFP657 (Addgene #57824) marker was specifically

used (see **Appendix G** for cloning protocol). The two parameters that were tracked in this screen were proliferation of sgRNA expressing cells, assessed using RFP567 fluorescence, and *HLF* expression, assessed using ZsG fluorescence. The genes in this screen were selected based on their downregulation in IMS-M2 cells after nuclear relocalization of the mutant NPM1c protein, as described earlier. We also included control knockouts; *AAVS1,* a negative control where we did not expect to see significant changes in ZsG or RFP657 expression; *HLF*, *NPM1c* and *ZsG*, positive controls where we expected to see a significant decrease in ZsG expression and another positive control of *RPA1*, a gene that encodes an important DNA replication protein, where we expected to see a decrease in RFP657 expression since DNA replication and cell proliferation would be affected by this knockout.

After the Cas9 protein was nucleofected into sgRNA-expressing cells in an arrayed format at day zero, *HLF* expression (ZsG fluorescence) and cell proliferation (RFP567 fluorescence) were tracked and recorded every other day for two weeks (seven time points in total) **(Figure 11C)**. For the analysis of the screen, RFP657 and ZsG levels for each sgRNA at each day were compared (normalized) back to their levels at day zero by dividing the percentage of fluorescence at each time point by the percentage of fluorescence at day zero. Using viral vectors to introduce sgRNA constructs into cells allowed us to stably label these cells and conclude that any potential and significant changes in the RFP657 fluorescent marker were due to the changing biology of a cell, such as cell death, rather than a natural fluctuation of the expression of our viral vector. Puromycin selection was released on the day of Cas9 nucleofection (day zero) to allow the observance of any potential changes in *HLF-ZsG* expression without selecting for cells which express *HLF*.

Knockout of *AAVS1* was included as a negative control as its knockout is not known to produce any effect in cells.[60] Cell proliferation (RFP657 fluorescence), and *HLF* expression (ZsG fluorescence) for each sgRNA was compared to sg*AAVS1.* All three sgRNAs targeted against the negative control locus, *AAVS1*, did not result in any significant change in RFP657 expression over time, as expected, whereas ZsG

expression decreased to the extent that was expected due to the release of puromycin selection **(Figure 11D)**.

Evaluation of the positive controls within the screen (sgHLF, sgNPM1c, sgRPA1, sgZsG) allowed us to proceed with analysis of results and identify the regulators of HLF **(Figure 11D)**. One out of three sgRNAs against NPM1c led to a reduction in ZsG expression followed by a slightly delayed but prominent reduction in RFP657. This knockout corresponded to the sgRNA designed by *Brunetti et al*. (2018) to target the four base-pairs insertion present within the mutation.[30] The expected loss of HLF-ZsG expression after sgNPM1c provided us with the reliability that our construct accurately relayed the loss of *HLF* expression after disrupting the expression of a known HLF regulator. This assurance allowed us to move forward with analysis of the screen. Two of three sgRNAs against of *RPA1* led to a reduction in cell proliferation, as shown by a loss of the fraction of cells expressing RFP657. On the contrary, none of the sgRNAs targeting *RPA1* lead to a decrease in ZsG above what was expected with the release of puromycin selection. Two out of three sgRNAs targeting ZsG led to a reduction in ZsG expression, as expected, without impacting cell proliferation or RFP657 expression. Finally, HLF knockout cells displayed no significant reduction in RFP657 expression as well as an expected loss of ZsG expression following the release of puromycin selection.

A meta-analysis of the changes in ZsG and RFP657 expression from each knockout over time compared to the changes in these parameters from control *AAVS1* knockout cells allowed us to proceed with the analysis of knockouts which produced statistically significant changes in ZsG or RFP657 expression **(Figure 12A)**. Using a cut-off p-value of 0.05, we could see that knockout of *HOXA10, MEIS1* and *NKX2-3* led to a statistically significant change in ZsG expression. The magnitude of significance calculated using p-values showed that *NKX2-3* knockout cells displayed the greatest decrease in ZsG expression, followed by *MEIS1*, then *HOXA10.* The decrease in ZsG expression was displayed as early as day five post-nucleofection for *MEIS1* and *NKX2-3* knockout cells and day seven for *HOXA1*0 knockout cells. In addition to *HOXA10/MEIS1/NKX2-3, SPI1* and *STAT5A* knockout cells showed a more delayed but statistically significant difference in ZsG expression as well by day 14 and day nine for *SPI* and *STAT5A,* respectively.

Finally, knockout of *ADGRG1* resulted in a statistically significant change in ZsG expression at day five only, with *HLF-ZsG* expression recovering from day seven to the end of the screen (**Appendix D**). None of the sgRNAs produced any statistically significant differences in RFP657 expression. Analysis of individual sgRNA profiles showed that each statistically significant change in ZsG expression was a result of a significant decrease **(Figure 12B)**. See **Appendix D** for RFP657 and ZsG profile dynamics of all other sgRNA which did not produce statistically significant results.

In conclusion, results from our CRISPR screen reveal *HOXA10, MEIS1* and *NKX2-3* to be regulators of *HLF* expression while remaining dispensable for cell proliferation in IMS-M2 cells. The finding that loss of these factors did not reduce cell proliferation also assures us that the reduction in *HLF* expression was a result of cells downregulating *HLF* and not a result of cell numbers decreasing. Similar to results from shRNA-mediated knockdown of *HLF*, CRISPR/Cas9-mediated knockout of *HLF* confirmed our finding that *HLF* expression does not impact cell proliferation. While only one out of the three shRNAs targeting NKX2-3 seemed to reduce ZsG expression, our CRISPR screen showed that sgNKX2-3 resulted in a statistically significant decrease in *HLF* expression in IMS-M2 cells.

This screen also confirmed the impact of NPM1c on HLF regulation, demonstrated by a loss of ZsG expression in the knockout corresponding to the target of the 4-base pairs insertion in the NPM1c mutant.[30] Interestingly, the finding that *SPI1* and *STAT5A* resulted in a delayed loss of *HLF-ZsG* expression towards the end of the screen suggests that these genes may have a delayed impact on *HLF* regulation and that there may be other regulators that lie between *SPI1/STAT5A* and *HLF*. In summary, for the first time, *MEIS1, HOXA10* and *NKX2-3* have been demonstrated as regulators of *HLF* in NPM1c-mutated AML.

A

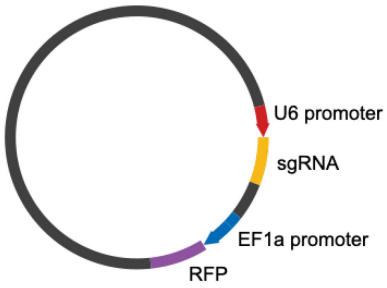
Legend:

Negative control

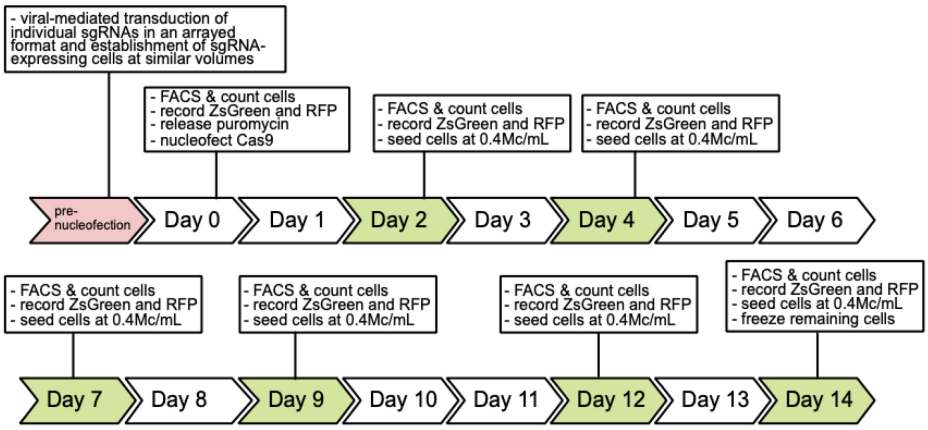
Positive control

AAVS1_sg1	NPM1c_sg1	HOXA10_sg1	SPI1_sg1
AAVS1_sg2	NPM1c_sg2	HOXA10_sg2	SPI1_sg2
AAVS1_sg3	NPM1c_sg3	HOXA10_sg3	SPI1_sg3
ADGRG1_sg1	PBX1_sg1	HOXB8_sg1	STAT5A_sg1
ADGRG1_sg2	PBX1_sg2	HOXB8_sg2	STAT5A_sg2
ADGRG1_sg3	PBX1_sg3	HOXB8_sg3	STAT5A_sg3
FLT3_sg1	PBX3_sg1	MEIS1_sg1	STAT5B_sg1
FLT3_sg2	PBX3_sg2	MEIS1_sg2	STAT5B_sg2
FLT3_sg3	PBX3_sg3	MEIS1_sg3	STAT5B_sg3
HLF_sg1	RPA1_sg1	NKX2-3_sg1	ZsG_sg1
HLF_sg2	RPA1_sg2	NKX2-3_sg2	ZsG_sg2
HLF_sg3	RPA1_sg3	NKX2-3_sg3	ZsG_sg3

B



C



D

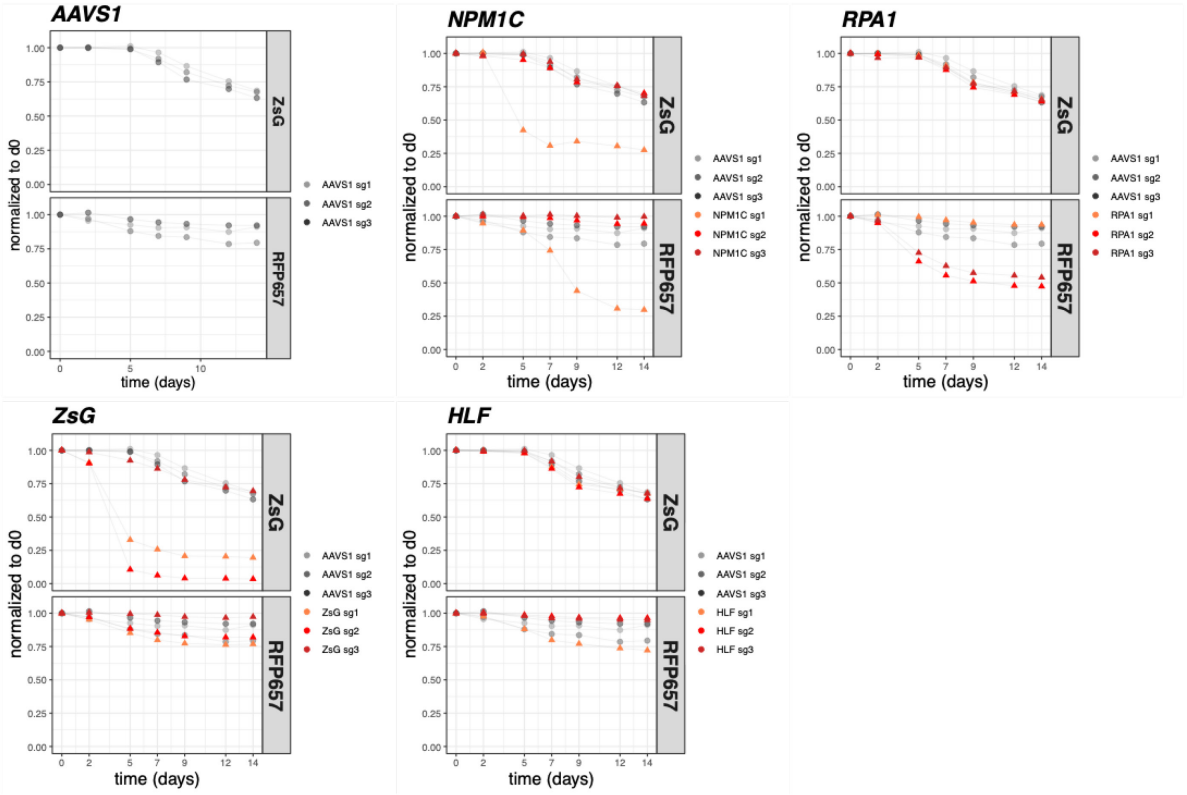


Figure 11. CRISPR arrayed knockout screen setup allows the identification of HLF regulators. (A) List of candidates in the screen including positive and negative controls. (B) Schematic of the vector into which sgRNA candidates were cloned. (C) Experimental schematic of CRISPR screen. (D) ZsG and RFP657 expression dynamics of control sgRNAs (sgAAVS1, sgNPM1c, sgRPA1, sgZsG and sgHLF). ZsG fluorescence serves as a proxy for *HLF* expression and RFP657 serves as a proxy of cell proliferation of sgRNA-expressing cells. N=3 different sgRNAs per gene. One biological replicate per sgRNA was performed and is shown in the figure. Two-tailed T-tests were performed to compare the average of the 3 sgRNAs for each gene with that of the 3 sgAAVS1 at each time point, however no significant differences were observed.

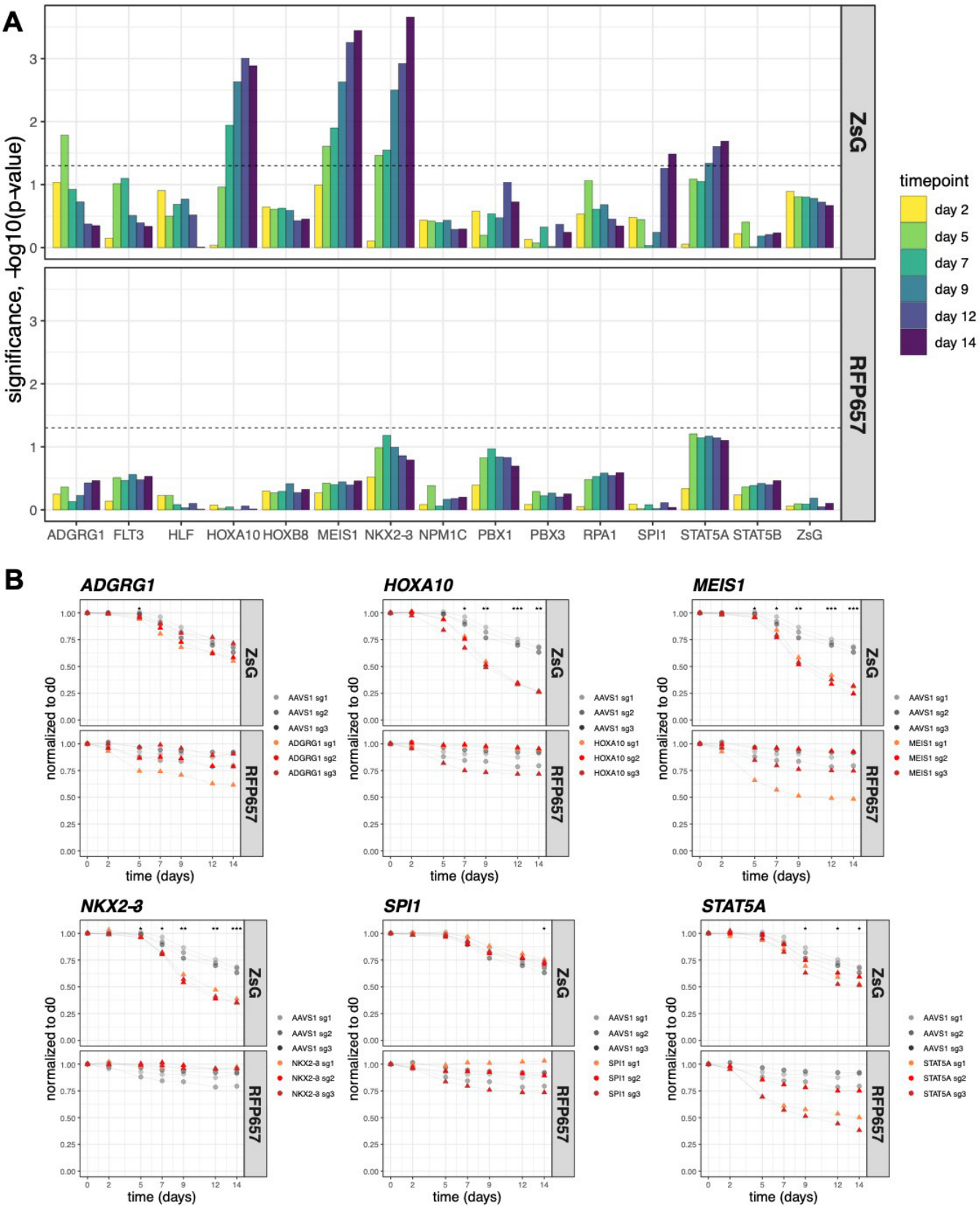

A
significance, −log10(p-value)
ZsG
RFP657
timepoint
day 2
day 5
day 7
day 9
day 12
day 14
ADGRG1 FLT3 HLF HOXA10 HOXB8 MEIS1 NKX2-3 NPM1C PBX1 PBX3 RPA1 SPI1 STAT5A STAT5B ZsG
B
ADGRG1
HOXA10
MEIS1
NKX2-3
SPI1
STAT5A
normalized to d0
time (days)
AAVS1 sg1
AAVS1 sg2
AAVS1 sg3
ADGRG1 sg1
ADGRG1 sg2
ADGRG1 sg3
HOXA10 sg1
HOXA10 sg2
HOXA10 sg3
MEIS1 sg1
MEIS1 sg2
MEIS1 sg3
NKX2-3 sg1
NKX2-3 sg2
NKX2-3 sg3
SPI1 sg1
SPI1 sg2
SPI1 sg3
STAT5A sg1
STAT5A sg2
STAT5A sg3

Figure 12. CRISPR arrayed knockout screen reveals HOXA10, MEIS1 and NKX2-3 to be regulators of HLF. (A) Meta-analysis of statistically significant changes in *ZsG* and RFP657 expression from candidate gene knockouts compared to *AAVS1* knockout cells. (B) Individual profiles of *HOXA10*, *MEIS1*, *NKX2-3*, *SPI1* and *STAT5A* knockout cells, which all produced statistically significant changes in ZsG expression. n=3 for each knockout. Asterisks of significance are assigned according to p-values as according: p-value ≤ 0.001 = ***, p-value ≤ 0.01 = **, p-value ≤ 0.05 = * based on two-tailed T-test.

CHAPTER 3 – DISCUSSION

3.1 Monitoring HLF-Expressing Dynamics Using a ZsG Reporter

Using CRISPR/Cas9 technology, we have validated an HLF reporter in a leukemic cell line that relays *HLF* expression using ZsG fluorescence as an output. After targeting IMS-M2 cells with the HLF-ZsG reporter and selecting cells using puromycin treatment, we were able to achieve a population in which the majority expressed the HLF-ZsG reporter. The subsequent observed and gradual loss of ZsG fluorescence in the *HLF* expressing population after releasing puromycin selection most likely occurred as cells began differentiating and subsequently downregulated *HLF* expression, similar to *HLF* expression dynamics observed in a healthy hematopoietic setting.[12-15,63] On the contrary, the slight gain of HLF-ZsG expression in HLF-ZsG non-expressing cells was most likely attributable to an incomplete FACS sorting which allowed the leakage of a very small proportion of *HLF* expressing cells into the HLF negative population (**Figure 6**).

Similar population doubling kinetics observed for HLF14 and HLF17 expressing and non-expressing cells as well as untargeted and unsorted cells reveals that neither integration of the reporter nor expression of *HLF* influences cell proliferation in IMS-M2 cells. This finding suggests that if we were to reduce the expression of an *HLF* regulator which resulted in decreased *HLF* expression and decreased cell proliferation, we could presume

that the decrease in cell proliferation was a result of the removal of an *HLF* regulator and not due to a decrease in *HLF* expression. Although the role of *HLF* expression in influencing cell cycle dynamics has already been assessed in other AML cell lines and in $CD34^+$ cord blood HSCs, future experiments in the context of this study could look at specific proliferation markers or cell cycle analysis of *HLF* expressing and non-expressing IMS-M2 cells to further assess if *HLF* has any specific role in regulating cell cycle progression in this leukemic cell line, albeit potentially small.

Finally, to assess that this reporter could accurately relay changes in *HLF* expression, we confirmed that knocking out a known regulator of *HLF*, such as NPM1, specifically led to a reduction in ZsG expression compared to control knockouts.

In addition to the reliability this reporter provides in identifying the regulators of our gene of interest, it also demonstrates paramount flexibility beyond the context of this study. By modifying the location of the Cas9-mediated DSB and homology arms of this repair template, the backbone of this reporter cassette can be used and designed to report on a wide variety of other genes of interest.[13,60,62] Similarly, the establishment of nucleofection parameters in this leukemic cell line to introduce this reporter and knock out genes of interest can be taken into consideration to help other scientists use this technique when introducing a fluorescent genomic reporter into similar AML cell lines.

3.2 Involvement of HLF in Stem Cell Properties

Despite its role as a stem cell gene in both healthy HSCs and LSCs, transplantation experiments using IMS-M2 cells did not indicate *HLF* as a stem cell gene in this cell line. Despite previous xenotransplantation assays which have demonstrated the role of HLF in maintaining the LSC compartment in NPM1c-mutated leukemia,[17] the differences in genetic landscapes between the NPM1c-mutated cell lines used in these experiments and in IMS-M2 cells might explain the differences in the stem cell properties that *HLF* imparts. The lack of expression of *HLF* in OCI-AML3 further highlights the heterogeneity of *HLF* expression in AML and suggests its potential non-essentiality as a stem cell gene in leukemia.[30] In addition, the decrease in proliferation after NPM1c relocalization in OCI-AML3 cells (HLF non-expressing) but not in IMS-M2 cells (HLF expressing)[30] suggests that perhaps HLF protects the survival of IMS-M2 cells rather than acting as a stem cell

gene in this leukemia. Finally, the lack of stem cell display in IMS-M2 cells may be explained by the fact that transplantation experiments were conducted using cells that had been sorted for *HLF* expression based on ZsG expression as determined by flow cytometry. Therefore, it is possible that a leak of the *HLF* positive population to the *HLF* negative population resulted in no significant changes in between engraftment percentages.

However, despite the previously defined role of HLF in maintaining the LSC compartment in NPM1c-mutated leukemia, serial transplantation of AML triple mutated samples that had *HLF* expression knocked out resulted in high engraftment levels more rapidly in secondary recipients than in control samples.[17] This suggests that, in opposition to its role in HSCs, *HLF* may not be relevant in providing engraftment potential to LSCs and parallels our findings that *HLF* did not impact engraftment potential in transplantation experiments using IMS-M2 cells.

Repeating transplantation experiments using *HLF* expressing and non-expressing IMS-M2 cells might help to address the role of *HLF* expression in LSCs. In addition to re-sorting IMS-M2 cells into *HLF* expressing and non-expressing populations and transplanting higher cell numbers than used before, knocking out *HLF* expression and transplanting knockout and control cells may provide clearer results regarding the role of *HLF* in providing engraftment potential to IMS-M2 cells.

In addition to transplantation experiments, and understanding the specific expression of *HLF* in HSPC populations, differentiation marker staining of *HLF* knockout IMS-M2 cells may help understand the stem cell role that *HLF* holds in this leukemic setting. Ultimately however, a lack of understanding of the genes and pathways which regulate *HLF* expression in a healthy hematopoietic system and a leukemic system may help uncover the roles in these two settings. To address this question, we then proposed to investigate the regulators of *HLF* in a leukemic cell line and use these results as a template to identify the genes which regulate *HLF* in a healthy hematopoietic system.

3.3 *HLF* Expression Depends on Mutant NPM1c

Analysis of leucegene data first allowed us to re-think *HLF* regulators that had been previously published. Despite previously published literature that argued the dependency of *HLF* expression on *NPM1c*, DNMT3A and *FLT3-ITD* mutations inclusively,[17] Leucegene cohort data revealed that the mutational status of *DNMT3A* does not affect *HLF* expression at all **(Figure 4)**. In fact, analysis of this data revealed that *HLF* expression more heavily depends on *NPM1c* and *FLT3-ITD* mutations in leukemia. Furthermore, in previously published literature of *HLF* expression in triple mutated AML, authors recognize that *HLF almost* reaches exclusive expression in triple-mutated AML and present data which shows very comparable expression of *HLF* in *NPM1c/FLT3-ITD* and *NPM1c/FLT3-ITD/DNMT3A* mutated subgroups, further confirming the potential lack of relevance of *DNMT3A* mutations in regulating *HLF* expression.[17]

Using our HLF-ZsG reporter, we were able to confirm the dependency of *HLF* expression on NPM1c after knocking out the *NPM1c* mutation in IMS-M2 cells and observing a significant reduction in ZsG expression compared to control knockouts of *AAVS1* **(Figure 9D and 9E)**. On top of determining this relationship in IMS-M2 cells, this proof-of-principle experiment validated the accuracy of our HLF-ZsG reporter by producing an expected phenotype (loss of *HLF* expression) when removing the expression of a known *HLF* regulator. Furthermore, this knockout allowed us to establish appropriate nucleofection parameters for IMS-M2 cells and use these parameters to knockout additional potential regulators of *HLF*. The gain of *HLF* expression seven days after knockout of *NPM1c* can be attributed to the fact that the presence of Cas9 after electroporation in cells is transient[64] and also suggests an ability of a small number of unedited cells to outcompete *NPM1c* knockout cells, allowing the population to gain *HLF* expression.

Interestingly, *FLT3* knockout cells did not lead to a significant reduction in *HLF* expression (**Appendix C**). However, the mutational status of *FLT3* in IMS-M2 cells is not known. Therefore, we propose that future experiments should sequence the *FLT3* locus to determine if the *FLT3-ITD* mutation is present and if so, target it with inhibitors or knock out its expression using CRISPR. This experiment would allow us to specifically determine the effect of the *FLT3-ITD* mutation on *HLF* expression using our HLF-ZsG IMS-M2

reporter cell line. Overall, the revealed dependency of *HLF* expression on the NPM1c mutation not only demonstrated the reliability of our HLF-ZsG reporter in tracking the regulators of *HLF*, but allowed us to begin exploring the regulators of *HLF* which lie downstream of NPM1c.

3.4 Defining the Network of HLF Regulators Downstream of NPM1c

Analysis of transcriptome data in IMS-M2 cells after knockout of mutant NPM1c further confirmed the dependency of *HLF* expression on NPM1c and helped us define a preliminary list of genes which could be potential regulators of *HLF*. In addition to *HLF*, transcriptome data revealed that nuclear relocalization of NPM1c resulted in downregulation of *HOXA9, HOXA10, HOXB8, MEIS1, NXK2-3* and *PBX3*. According to the change in expression of these genes upon NPM1c knockout specifically, we could begin to test if they in turn regulate the expression of *HLF* as well.

shRNA-mediated knockdown against three of these genes (*NKX2-3, HOXA9* and *MEIS1*) revealed *MEIS1* to be a strong candidate regulator of *HLF* without impacting cell proliferation (**Figure 10B** and **10C**). As expected, knockdown of *HLF* also resulted in a decrease in ZsG expression above what was expected with the release of puromycin upon infection of shRNA vectors. Similar to our findings that *HLF* non-expressing and expressing IMS-M2 cells had similar proliferation dynamics, shHLF did not reduce the proliferation of cells. One shRNA designed against *NKX2-3* seemed to produce a reduction in ZsG expression that was comparable to knockdown of *HLF*, suggesting that *NKX2-3* could also be a potential regulator of *HLF* expression in addition to *MEIS1*. Future experiments will need to address knockdown efficiencies of each shRNA through quantitative PCR (qPCR) to measure transcript levels of genes that were knocked down.

Interestingly, despite the upregulation of *HOXA9*, *MEIS1* and *NKX2-3* in *NPM1c*-mutated AML, and the dependency of *HLF* expression on the *NPM1c* mutation, *HOXA9* in turn does not seem to be a regulator of *HLF*. This finding in fact parallels published literature that HOXA9 is not required for leukemic transformation in E2A-HLF fusion leukemias.[65] Despite our initial expectations, shRNA experiments do not suggest that HOXA9 regulates

HLF expression. However, the reliance of HOX gene expression on the *NPM1c* mutation,[30] provided us with the grounds to further explore the potential regulation of other HOX genes on *HLF* expression.

The use of a small CRISPR arrayed knockout screen allowed us to investigate the potential of homeobox genes and other candidate genes to regulate *HLF* expression. First, evaluation of the positive controls (sgHLF, sgNPM1c, sgRPA1, sgZsG) allowed us to draw conclusions regarding the sgRNAs which produced a significant reduction in *HLF* expression. sgNPM1c resulted in a loss of HLF-ZsG expression for the sgRNA that targeted the 4-base pairs insertion found within the *NPM1c* mutant, as expected.[30] sgRPA1 also resulted in decreased RFP657 expression as a result of decreased cell proliferation, as expected due to its role in DNA replication.[58] This allowed us to confirm that our screen could report on knockouts which reduced cell proliferation in addition to HLF-ZsG expression. Similar to shRNA-mediated knockdown of *HLF*, CRISPR/Cas9-induced *HLF* knockout did not affect cell proliferation, confirming our previous findings regarding the dispensable nature of *HLF* on cell proliferation in IMS-M2 cells. However, unlike the loss of HLF-ZsG expression after shHLF knockdown, knockout of *HLF* did not significantly reduce ZsG expression. This can be explained because the guide sequences used in CRISPR/Cas9 gene knockouts are designed to induce a DSB in a particular gene, which the cell repairs partially using non-homology end joining (NHEJ) pathways.[66] NHEJ can result in insertions-deletions (INDELs) of unpredictable length, resulting in frameshift mutations.[66] These mutations could result in: the introduction of a premature stop codon, and the subsequent degradation of nascent mRNA through nonsense-mediated decay; or the removal of an entire amino acid that could result in the translation of a non-functional protein. [60,67,68] In the case of the *HLF* knockout condition, while HLF mRNA and protein were not functionally produced, the IRES element located downstream of HLF and just upstream of the ZsG coding sequence allowed transcription and translation of the ZsG mRNA and subsequent ZsG protein fluorescence **(Figure 6A)**.[13,69]

Results from our CRISPR screen revealed that *HOXA10, MEIS1* and *NKX2-3* regulate *HLF* while remaining dispensable for cell proliferation. Not only did this result introduce specific regulators into our hypothesized network of *HLF* regulation, it also confirmed the

phenotype of reduced HLF-ZsG expression after knockdown of *NKX2-3* with one of our shRNAs. While we have identified these regulators of *HLF*, the exact order of this network has not yet been defined. Considering that *HLF* expression disappears when NPM1c is knocked out or relocalized, we propose that future experiments include a rescue experiment involving knockout of NPM1c followed by overexpression of *HOXA10, MEIS1* and *NKX2-3* to see if we can recover *HLF* expression. We also propose that quantitative PCR (qPCR) be used to assess the expression of the other *HLF* regulators after one of them is targeted and that the gene be sequenced to confirm that a mutation was introduced. Not only will these experiments further validate the role of these potential *HLF* regulators, but they will also help us determine the hierarchy and order of genes which regulate *HLF* and will further help us solidify our understanding of the pathways of *HLF* regulation in *NPM1c*-mutated leukemia.

Targeting of *SPI1* and *STAT5A* also reduced ZsG expression, albeit to a lesser extent and later in the time course of the experiment than *HOXA10/MEIS1/NKX2-3.* sgSPI1 and sgSTAT5A reduced ZsG levels towards the end of the screen (12 and 14 days post-nucleofection of Cas9, respectively), suggesting that their effect on *HLF* expression may have been delayed because they rely on other genes in a pathway to regulate *HLF* first. For example, SPI1 is a protein known to be exported out of the nucleus and to the cytoplasm along with mutant NPM1c.[70] Therefore, while SPI1 may regulate *HLF* expression, it most likely works downstream of NPM1c and the disruption in protein localization that this mutant induces. Additionally, despite the encoding of *STAT5A/B* by paralog genes and the similar roles they hold in proliferation and cell cycle regulation in hematopoiesis and leukemogenesis,[50] the finding that targeting of *STAT5A* alone led to a significant reduction in *HLF* expression, but not *STAT5B*, was interesting. The demonstrated role of *STAT5A* in HSC self-renewal and differentiation and of *STAT5B* in immune cell signalling may help explain how only STAT5A regulated *HLF* expression.[50] Furthermore, the ability of *FLT3-ITD* to activate aberrant expression of STAT5 transcription factors, a pathway which has been recognized in leukemic transformation, suggests that *STAT5* genes may rely more heavily on upstream *FLT3* regulation to in turn regulate *HLF*.[49,50]

In conclusion, using a genomic fluorescent reporter of *HLF* expression in combination with functional CRISPR-based gene targeting experiments, we have defined a preliminary network of genes which regulate *HLF* in the NPM1c-mutated leukemic cell line IMS-M2 **(Figure 13).** While we have identified a few regulators of *HLF*, there is no doubt that a complex network of genes and pathways regulating *HLF* expression still lie downstream of NPM1c. Especially when considering the involvement of homeobox genes in *HLF* regulation, their roles as master transcription factors and their ability to regulate many genes,[30] a more extensive investigation of the genes and pathways which regulate *HLF* is still needed to fully unveil this network. We suggest addressing this through a genome-wide knockout screen using the HLF-ZsG reporter to uncover all potential regulators of HLF.

Considering the relevance of *HLF* expression in self-renewal, differentiation and cell cycle regulation of HSCs,[2-4,7,13-16,18,20,27] as well as its specific expression in NPM1c-mutated leukemia,[17] we can hypothesize that this gene serves as a relevant point of entry for uncovering the molecular differences in networks that regulate these two systems. Now that a few regulators of *HLF* have been identified, future experiments can compare their expression in HSCs to assess how and if this same network holds in normal HSCs as well.

For example, *MEIS1* and *NKX2-3* are known to be specifically expressed in HSCs and play a role in differentiation, with their expression declining as cells differentiate.[70-73] In addition, *MEIS1* is thought to be relevant in the stress response for this population.[74] Finally, binding sites for HOXA10 have even been identified in *HLF* in HSCs, and *HOXA10* has been shown to regulate HSC proliferation as well as erythroid and megakaryocyte development.[51] Therefore, the fact that the expression of *HLF* regulators identified in this study in a leukemic system also play relevant roles in HSCs differentiation and proliferation suggests that they may also regulate *HLF* in a healthy hematopoietic system.[51]

Future experiments may prioritize a genome-wide search to fill in the missing pieces in the network of genes and pathways which regulate *HLF*. After a list of these genes has been defined, a comparison of their regulation of *HLF* in HSCs may contribute to our

understanding of the ways in which these networks differ in their regulation of *HLF*. Distinguishing between the genetic landscape of healthy hematopoietic systems and leukemic ones is important in designing drugs that specifically target leukemic cells while leaving healthy ones intact. Knowing this, *HLF* may serve as an important point of reference for uncovering these differences. The specific and aberrant expression of *HLF* in NPM1c-mutated AML[17] may even allow it to serve as a molecular marker for this AML.

Unfortunately, it has been recognized that a majority of patients coping with AML fail to respond positively to chemotherapies in the long term.[17] Therefore, a better understanding of the genetic landscapes of AML subgroups and the ways in which diseases operate on a molecular level may help us design better treatments for them.

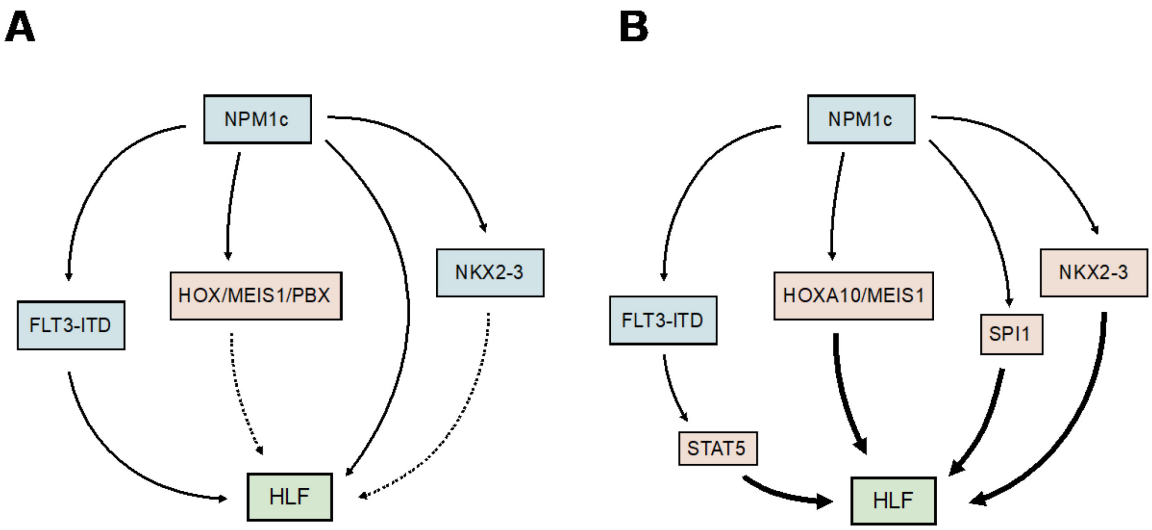


Figure 13. Hypothesized and defined network of *HLF* regulators in NPM1c-mutated AML. (A) Hypothesized network of *HLF* regulation in IMS-M2 cells. Solid arrows and blue boxes represent interactions which have already been demonstrated and dotted arrows represent our hypothesized interactions based on published literature. (B) Network of *HLF* regulation as uncovered in our study. Thin arrows and blue boxes represent interactions which have already been demonstrated and thick solid arrows represent interactions which have been uncovered in our study.

Concluding remarks

HLF is a relevant gene in defining HSC differentiation and self-renewal. Its expression has also been identified in NPM1c-mutated AML. Despite a previous lack of knowledge on the mechanism of regulation of *HLF* in these two systems, our work has revealed *HOXA10*, *MEIS1* and *NKX2-3* to be strong candidate regulators of *HLF* in the leukemic cell line IMS-M2. A comparison of *HLF* regulation in normal HSCs and leukemia may be exploited to better understand the ways in which a healthy hematopoietic system transforms to leukemia as well as inform decisions regarding the design and application of NPM1c-mutated chemotherapies.

CHAPTER 4 – METHODS

4.1 Leucegene cohort analysis

The Leucegene project is an initiative approved by the Research Ethics Boards of the Université de Montréal and Maisonneuve-Rosemont Hospital. As part of this project, RNA sequencing of 415 primary AML specimens from various cytogenetic groups was performed, including 110 samples that were also characterized by exome sequencing, as previously described.[73] All leukemia samples and paired normal DNA specimens were collected and characterized by the Quebec Leukemia Cell Bank (BCLQ). Normal bone marrow (BM) samples were obtained from the BCLQ and Lonza, and cord blood from Héma-Quebec.[47]

4.2 Cell culturing

IMS-M2 cells were cultured in RPMI 1640 media (ThermoFisher 11875093) supplemented with L-glutamine, 25mM HEPES and 10% heat-inactivated (HI) fetal bovine serum (FBS). Once IMS-M2 cells were targeted for the HLF17-reporter, 0.6ug/ul of puromycin was added to the media to select for targeted cells. For all experiments involving culturing of

IMS-M2 cells, cells were split every other day at $4x10^5$ cells/ml and maintained in T25 suspension flasks (Sarstedt #83.3910.500) at 37°C and 5% CO_2.

4.3 CRISPR engineering of a genomic *HLF* reporter

CRISPR/Cas9 was used to introduce the coding sequence for the fluorescent ZsGreen (ZsG) protein downstream of the *HLF* ORF so that when *HLF* is expressed, ZsG is as well and can be tracked in real time. A single guide ribonucleic acid (sgRNA) containing an internal ribosome entry site (IRES) element followed by the expression cassette for the coding sequence of ZsG and a P2A-linked Puromycin resistance gene or truncated EPCR receptor (tEPCR) was designed to bind downstream of the *HLF* ORF.[13] The expression cassette containing the puromycin resistant gene is also noted as the HLF17 reporter, and the one containing tEPCR is also noted as the HLF14 reporter. Delivery of a Cas9/sgRNA ribonucleoprotein (RNP) complex was used to direct a DSB at the 3'-end of the *HLF* open reading frame and delivery of a single-stranded recombinant adeno-associated viral vector (rAAV6) provided a repair template to cells for repair of the DSB through homologous recombination (HR).[13] See supplementary *Lehnertz et al. (2021)* file for a full detailed protocol of reporter targeting.

This HR event resulted in a transgenic locus that co-expresses the *HLF* ORF linked to the ZsGreen (ZsG) expression cassette through an encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES).[13] HLF17 cells were sorted for ZsG+ and maintained in media supplemented with puromycin for selection of targeted cells.

Integration of the HLF reporter was validated using droplet digital PCR (ddPCR) to detect the integrated reporter and Western blots to confirm the HLF protein in *HLF*-expressing cell lines solely.[13] In addition, transduction of the reporter in HEPG2 cells (which express *HLF*) and HEK293 (which do not express *HLF*) and detection of the reporter using ddPCR was further used to confirm integration of the HLF reporter in HEPG2 cells and not in HEK293 cells.[13] Nucleofection of cells with 3ug of Cas9 and 8ug of sgRNA was used to introduce the DSB in 10e6 cells per 100ul using the Lonza 4D-Nucleofector Core (Lonza 4d, DZ100 program #AAF-1002B). Cells were washed in PBS and taken up in 1M nucleofection buffer (Amaxa) containing the pre-assembled Cas9 sgRNA RNP complex (11ug total). After nucleofection, cells were plated at $4x10^5$ cells/ml in RMPI media

containing the reporter encoding rAAV6 at a multiplicity of infection (MOI) of 50,000. After the HLF17 reporter was introduced, cells were maintained in 10% HI FBS RPMI with 0.6ug/ul of puromycin and assessed using Fluorescent-Activated Cell Sorting (FACS) and the Canto II (BD Bioscience).

HLF-expressing and non-expressing cells were sorted into two different populations based on ZsGreen expression using a BD FACSAria II cell sorter (BD Bioscience). After sorting cells and establishing sufficient cell numbers, cells were analyzed through flow cytometry using a BD Bioscience Canto II flow cytometer.

4.4 Cloning

shRNA vectors were cloned into a lentiviral construct containing an Ultra-miR scaffold for optimized and consistent knockdown (**Figure 5.3A**). An ametrine fluorescent marker was included in these vectors for selection of shRNA-expressing cells (original MNDU vector from Dr. Humphries lab using ultra-miR sequence from Dr. Knott Addgene #81071). This vector was constructed under the expression of an MNDU promoter for high expression and high lentiviral vectors titers. shRNA vectors were cloned using Gibson assembly (Ultra-miR cloning by Gibson Assembly protocol, **Appendix F**).

sgRNA vectors were cloned into a pLKO5 backbone vector (Addgene #57824) into a gRNA scaffold under a U6 promoter for high expression (**Figure 5.3B**). This vector also included a tRFP657 fluorescent marker under an EF1α promoter to drive strong gene expression. sgRNA designs were cloned into the pLKO5 vector using Addgene's Genome-scale CRISPR Knock-Out (GeCKO) Target Guide Sequence Cloning Protocol (**Appendix F**).

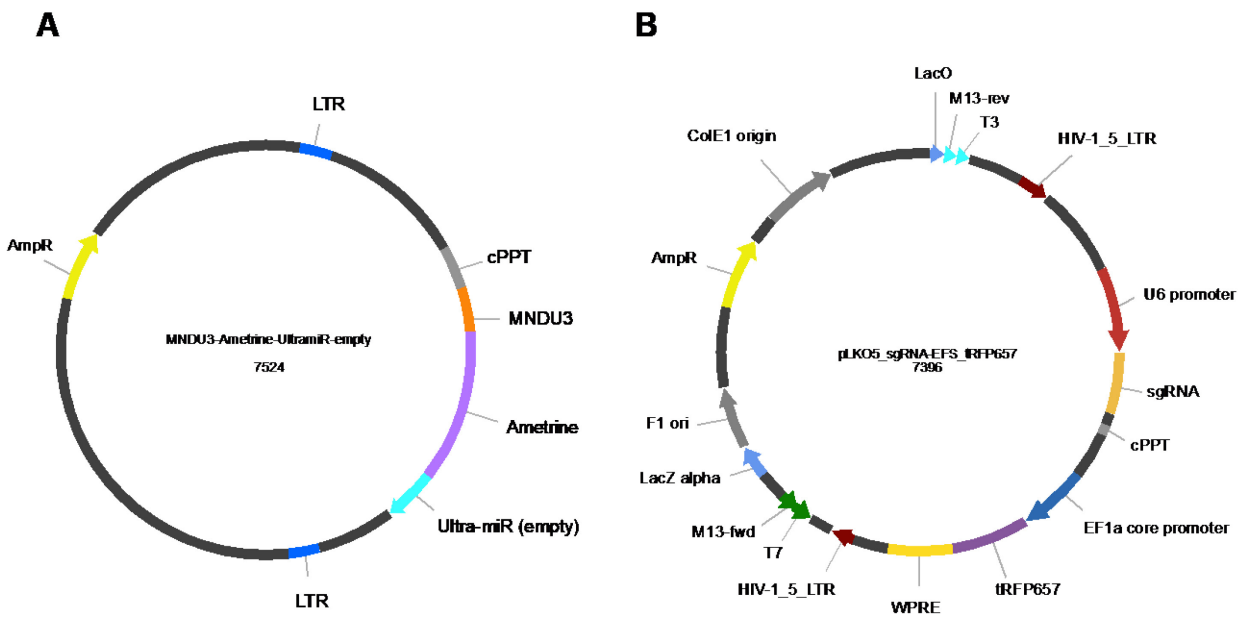


Figure 14. Schematic of lentiviral vectors used in cloning. (A) Empty shRNA vector for knockdown of gene candidates. Components include LTR, long terminal repeat; cPPT, central polypurine tract for increased integration of the viral genome into host cell nucleus upon infection[74]; MNDU3 promoter, a modified the U3 region of the myeloproliferative sarcoma virus LTR with the noncoding region removed[76]; Ametrine, fluorescent marker; Ultra-miR, a variation of the traditional miR30 scaffold for efficient shRNA processing[77] and AmpR, ampicillin resistance gene to allow for selection of vector-expressing cells in ampicillin present environments.[78] (B) Empty sgRNA vector for directing Cas9-induced gene knockouts at desired locations within the genome.[64] Components include lac operon (LacO and LacZ alpha) for efficient transduction into bacteria to allow the production of the vector DNA in large quantities[79,80]; M13-for and -rev to provide primer sites for sequencing of vector-expressing cells[81]; T3 and T7 promoters for transcription by T3 and T7 RNA polymerase respectively[79]; HIV-1 '5 LTR for successful integration into host cells,[74] U6 promoter recognized by RNA polymerase III for driving strong activity of small sgRNAs and reduced toxicity after viral transduction[64,82]; EF1α core promoter for strong constitutive gene expression[83,84]; tRFP657, tag red constitutive fluorescent reporter protein (Addgene #57824); WPRE, Woodchuck Hepatitis Virus Post-transcriptional Response Element to enhance gene expression after viral vector delivery[85] and F1 ori and ColE1 as the site for origins of replication in bacterial and mammalian host cells.[86-88]

4.5 Virus production and transduction

Lentivirus-mediated transduction of shRNA and sgRNA vectors was used to introduce them into IMS-M2 cells. Lentiviral particles were produced by HEK293TG cells. HEK293TG cells were seeded at $0.5x10^5$ cells/mL the afternoon before transfection in a 6-well tissue-culture treated plate (Sarstedt #83.3920.300) and maintained in 2mL of DMEM (ThermoFisher Gibco #1230052) supplemented with 10% HI FBS. Per each 9.6cm2 well in a 6-well plate, 1ug of lentiviral shRNA or sgRNA vector was transduced along with 0.67ug of PAX2 packaging plasmid and 0.2ug VSVG envelope glycoprotein plasmid using the jetPRIME transfection kid (Polyplus, jetPRIME #101000027). HEK293TG cells were maintained at 37°C and 5% CO_2. The media was changed 18 hours post-transfection with 2mL of DMEM containing 10% HI FBS. Viral supernatant was collected and filtered 48 hours post-transfection using a 3mL syringe (BD Biosciences Plastipak Sluer Slip Tip #309656) and a 33mm filter (Millex-GV Syringe Filter Unit, 0.2um, Millipore Sigma, #SLGV033RS). For infection of HLF-targeted IMS-M2 cells with viral vectors, $2x10^5$ cells were plated in 2mL of filtered viral HEK293TG supernatant followed by the addition of 1/1000 dilution of lentiboost and 6ul/ml of polybrene. Cells and filtered viral particles were centrifuged together at 1000g for 1h at 32°C to achieve infection of IMS-M2 cells with viral vectors. Cells were washed with PBS after infection and plated back at $4x10^5$ cells/mL. Transduction efficiency (ranging 20-95%) was assessed 48 hours post-infection by quantification of the fluorescent marker present in the viral vector using a BD Biosciences Canto II.

4.6 FACS analysis

FACS Canto II was used to track ZsG expression from the HLF-reporter cassette, Ametrine expression from shRNA-expressing cells, RFP657 expression from sgRNA-expressing cells and cell counts. To count cells, 1000 FACS counting beads (Invitrogen by ThermoFisher, CountBright™ Absolute Counting beads #C36950) were loaded with 100ul of cells. The number of beads and single cells present in the sample was recorded. 0.1ul of beads corresponds to one bead, so when 500 beads were counted, the corresponding volume was calculated and extrapolated using Rstudio to calculate how many cells were in the volume required to reach 500 beads. See **Appendix B** for the

script that was used to calculate cells/mL, fold expansion and population doubling numbers. Cell counts, HLF-ZsG expression and the fraction of shRNA and sgRNA-expressing cells were assessed and recorded every other day for two to three weeks. After cells were counted using counting beads and flow cytometry, they were seeded back at $4x10^5$ cells/mL.

4.7 Colony forming assays

The ability of IMS-M2 cells to form colonies was tested by plating various amounts of cells in methylcellulose-based media with and without human cytokines. Cell numbers of 100, 500, 5000, 10000 and 15000 were plated in 1ml of methylcellulose-based media. This media consisted of 1.04% methylcellulose in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 20% HI FBS, 1% deionized bovine serum albumin (BSA), 2mM glutamine, 200ug/mL holo-transferrin and 10^{-2} M 2-mercaptoethanol. Each condition of varying cell numbers was plated with and without cytokines. The media that included cytokines consisted of 100ng/mL human stem cell factor (hu-SCF), 10ng/mL hu-IL-3, 10ng/mL granulocyte-macrophage colony-stimulating factor (GM-CSF), 3 U/mL erythropoietin, 10ng/mL hu-IL-6 and 50ng/mL thrombopoietin). Methylcellulose plates were cultured at 37°C and 5% CO_2 and after 14 days, colonies were counted. See **Appendix E** for reagents used in methylcellulose plating).

4.8 Transplantation experiments

All experimental protocols followed the guidelines of the Animal Care Committee of the University of Montreal. To determine the potential of the *HLF* gene as a stem cell gene, HLF17 expressing IMS-M2 cells were sorted into *HLF* expressing and non-expressing populations based on ZsG expression using a BD FACSAria II. Six NSG mice were irradiated (200 Gy) and transplanted with 50,000 cells for each condition, for a total of 24 transplantations (six mice transplanted with HLF14 positive cells, six with HLF14 negative cells, six with HLF17 positive cells and six with HLF17 negative cells). Bone marrow, peripheral blood and spleen were harvested for flow cytometry analysis. Engraftment of human cells in mouse bone marrow was monitored based on CD45 expression using flow cytometry at 4 weeks post-transplantation. Cells were treated with red blood cell lysis

buffer, washed with FACS buffer (PBS, 2% FBS, 2mM EDTA) and stained. Cells were analyzed on the BD Bioscience Canto II. See **Appendix H** for supplementary analysis of xenotransplantation.

4.9 Arrayed CRISPR knockout screen

The genes selected for the CRISPR arrayed knockout screen included *ADGRG1, FLT3, HOXA10, HOXB8, MEIS1, NKX2-3, PBX1, PBX3, SPI1, STAT5A, STAT5B*. A negative control of *AAVS1* and positive controls of *HLF, NPM1c, RPA1* and *ZsG* were also included. Three sgRNA's per gene were designed and cloned into a vector containing a fluorescent marker of a red fluorescent protein RFP657. Cell counts, HLF-ZsG expression and RFP657 expression levels were quantified using a BD Biosciences Canto II.

HLF targeted IMS-M2 cells were individually transduced with one of 48 sgRNAs as described in section 4.7 and maintained in RPMI with 10% HI FBS and 0.6ul/ul of puromycin. On day zero, cells were nucleofected with Cas9 (Invitrogen by ThermoFisher, TrueCutTM Cas9 Protein v2 #AC36497) using the Lonza 4D-Nucleofecto Core (DZ100 program). 0.2ug of Cas9 per sgRNA was suspended in 1M nucleofection buffer and added to the nucleofector strip. $2x10^5$ cells were collected, resuspended in more 1M nucleofection buffer and added to the 1M nucleofection buffer containing Cas9 in the nucleofection strip. After nucleofection, cells were washed and plated back at $4x10^5$ cells/mL. Day zero marked the beginning of the screen and the release of puromycin selection. Cell counts, HLF-ZsG expression and RFP657-expression were assessed and recorded every other day for two weeks using a BD Biosciences Canto II.

As the knockout of AAVS1 was not expected to result in a change in *HLF* expression or cell proliferation, a comparison of each knockout was made to the knockout of AAVS1 to determine if *HLF* expression or cell proliferation significantly changed compared to that displayed by AAVS1 knockout cells.

ZsGreen expression was used as an indicator of *HLF* expression. HLF-ZsG expression within the RFP657-expressing population was specifically compared to HLF-ZsG expression within non-RFP657-expressing cells to determine how the impact of knocking out a certain gene affected the expression of *HLF*. The overall expression of RFP657 was

used as an indicator of the proliferation of sgRNA expressing cells to determine if a Cas9-mediated knockout of one of the genes caused a change in cell proliferation.

4.10 Bioinformatic analysis of CRISPR screen

Assessment in changes in HLF-ZsG expression as well as cell proliferation kinetics was assessed and recorded using a BD Biosciences Canto II and analyzed in Rstudio. RFP657 expression, as well as ZsG expression within the RFP657+ fraction, was recorded every other day for two weeks (seven total time points) and normalized back to expression levels at day zero to determine how changes in RFP657 and ZsG occurred over time. Significance was calculated using a two-sided alternative t-test. Levels of significance were calculated and annotated based on p-values associated with the t-test (p-value <= 0.001 ~ '****', p-value <= 0.001 ~ '***', p-value <= 0.01 ~ '**', p-val <= 0.05 ~ '*'). See Appendix B and C for RScript used to count cells and determine statistically significant changes in ZsG and/or RFP657 expression.

Contributions

Hilary Johnson (H.J.); Bernhard Lehnertz (B.L.); Tara MacRae (T.M.); Nadine Mayotte (N.M.); Simon Girard (S.G.).

Figure 5: Analysis performed by B.L. and H.J.

Figure 6: Analysis performed by B.L. and H.J.

Figure 7: Experiments and analysis performed by B.L. and H.J.

Figure 7A: Reporter designed by B.L.

Figure 7B: Experiment and analysis performed by B.L.

Figure 7C: Experiment and analysis performed by H.J. and B.L.

Figure 8: Experiments and analysis performed by H.J.

Figure 9: Experiments and analysis performed by B.L., T.M., and N.M.

Figure 10: Experiments and analysis performed by B.L.

Figure 11: Experiments performed by H.J. Analysis performed by H.J. and B.L. T.M. and S.G. assisted with cloning.

Figure 12: Experiments performed by H.J. Analysis performed by H.J. and B.L. T.M. and S.G. assisted with cloning.

Figure 12: Experiments performed by H.J. Analysis performed by H.J. and B.L.

Figure 15: Experiments and analysis performed by H.J.

Figure 16: Experiments performed by H.J. Analysis performed by H.J. and B.L.

Figure 17: Experiments performed by H.J. and N.M. Analysis performed by H.J.

All figures not listed above were constructed by H.J.

Bibliographical References

1. Eliasson P, Jönsson J-I. The hematopoietic stem cell niche: Low in oxygen but a nice place to be. *Journal of Cellular Physiology*. 2010;222(1):17-22. doi:10.1002/jcp.21908
2. Gazit R, Garrison BS, Rao TN, et al. Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. *Stem Cell Reports*. 2013;1(3):266-280. doi:10.1016/j.stemcr.2013.07.004
3. Bak RO, Dever DP, Porteus MH. CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nature Protocols*. 2018;13(2):358-376. doi:10.1038/nprot.2017.143
4. Shojaei F, Trowbridge J, Gallacher L, et al. Hierarchical and ontogenic positions serve to define the molecular basis of human hematopoietic stem cell behavior. *Developmental Cell*. 2005;8(5):651-663. doi:10.1016/j.devcel.2005.03.004
5. Lee JY, Hong S-H. Hematopoietic stem cells and their roles in tissue regeneration. *International Journal of Stem Cells*. 2020;13(1):1-12. doi:10.15283/ijsc19127
6. Cheng H, Zheng Z, Cheng T. New paradigms on hematopoietic stem cell differentiation. *Protein & Cell*. 2019;11(1):34-44. doi:10.1007/s13238-019-0633-0
7. Challen GA, Boles N, Lin K-YK, Goodell MA. Mouse hematopoietic stem cell identification and analysis. *Cytometry Part A*. 2009;75A(1):14-24. doi:10.1002/cyto.a.20674
8. Ting SB, Deneault E, Hope K, et al. Asymmetric segregation and self-renewal of hematopoietic stem and progenitor cells with endocytic AP2A2. *Blood*. 2012;119(11):2510-2522. doi:10.1182/blood-2011-11-393272
9. Höfer T, Rodewald H-R. Differentiation-based model of hematopoietic stem cell functions and lineage pathways. *Blood*. 2018;132(11):1106-1113. doi:10.1182/blood-2018-03-791517
10. Tomellini E, Fares I, Lehnertz B, et al. Integrin-α3 is a functional marker of ex vivo expanded human long-term hematopoietic stem cells. *Cell Reports*. 2019;28(4). doi:10.1016/j.celrep.2019.06.084

11. Acosta J, Ssozi D, van Galen P. Single-cell RNA sequencing to disentangle the blood system. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2021;41(3):1012-1018. doi:10.1161/atvbaha.120.314654.
License Number : 5281470391351. Requested from Wolters Kluwer Health, Inc. And Copyright Clearance Center.
12. Watcham S, Kucinski I, Gottgens B. New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. *Blood*. 2019;133(13):1415-1426. doi:10.1182/blood-2018-08-835355.
License Number: 5282771353238. Requested from Elsevier and Copyright Clearance Center.
13. Lehnertz B, Chagraoui J, MacRae T, et al. HLF expression defines the human hematopoietic stem cell state. *Blood*. 2021;138(25):2642-2654. doi:10.1182/blood.2021010745
14. Yokomizo T, Watanabe N, Umemoto T, et al. HLF marks the developmental pathway for hematopoietic stem cells but not for Erythro-myeloid progenitors. *Journal of Experimental Medicine*. 2019;216(7):1599-1614. doi:10.1084/jem.20181399
15. Wahlestedt M, Ladopoulos V, Hidalgo I, et al. Critical modulation of hematopoietic lineage fate by hepatic leukemia factor. *Cell Reports*. 2017;21(8):2251-2263. doi:10.1016/j.celrep.2017.10.112
16. Hitzler JK, Soares HD, Drolet DW, et al. Expression patterns of the hepatic leukemia factor gene in the nervous system of developing and adult mice. *Brain Research*. 1999;820(1-2):1-11. doi:10.1016/s0006-8993(98)00999-8
17. Garg S, Reyes-Palomares A, He L, et al. Hepatic leukemia factor is a novel Leukemic Stem Cell Regulator in DNMT3A, NPM1, and flt3-ITD triple-mutated AML. *Blood*. 2019;134(3):263-276. doi:10.1182/blood.2018862383
18. 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. *Cell Reports*. 2017;21(12):3514-3523. doi:10.1016/j.celrep.2017.11.084

19. Riddell J, Gazit R, Garrison BS, et al. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell*. 2014;158(1):226. doi:10.1016/j.cell.2014.06.014
20. Urbán N, Cheung TH. Stem cell quiescence: The challenging path to activation. *Development*. 2021;148(3). doi:10.1242/dev.165084
21. Tümpel S, Rudolph KL. Quiescence: Good and bad of stem cell aging. *Trends in Cell Biology*. 2019;29(8):672-685. doi:10.1016/j.tcb.2019.05.002
22. Waters KM, Sontag RL, Weber TJ. Hepatic leukemia factor promotes resistance to cell death: Implications for therapeutics and chronotherapy. *Toxicology and Applied Pharmacology*. 2013;268(2):141-148. doi:10.1016/j.taap.2013.01.031
23. Szade K, Gulati GS, Chan CKF, et al. Where hematopoietic stem cells live: The bone marrow niche. *Antioxidants & Redox Signaling*. 2018;29(2):191-204. doi:10.1089/ars.2017.7419
24. Zarka J, Short NJ, Kanagal-Shamanna R, Issa GC. Nucleophosmin 1 mutations in acute myeloid leukemia. *Genes*. 2020;11(6):649. doi:10.3390/genes11060649
25. Weiskopf K, Schnorr PJ, Pang WW, et al. Myeloid cell origins, differentiation, and clinical implications. *Microbiology Spectrum*. 2016;4(5).
26. Inaba T, Roberts WM, Shapiro LH, et al. Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia. *Science*. 1992;257(5069):531-534. doi:10.1126/science.1386162
27. Dovey OM, Cooper JL, Mupo A, et al. Molecular synergy underlies the co-occurrence patterns and phenotype of NPM1-Mutant Acute Myeloid leukemia. *Blood*. 2017;130(17):1911-1922. doi:10.1182/blood-2017-01-760595
28. Velten L, Story BA, Hernández-Malmierca P, et al. Identification of leukemic and pre-leukemic Stem Cells by clonal tracking from single-cell transcriptomics. 2018. doi:10.1101/500108
29. Balusu R, Fiskus WC, Rao R, et al. Targeting levels, aberrant localization or oligomerization of mutant nucleophosmin induces differentiation and loss of survival of human AML cells with mutant NPM1. *Blood*. 2011;114(22):2656-2656. doi:10.1182/blood.v114.22.2656.2656

30. Brunetti L, Gundry MC, Sorcini D, et al. Mutant NPM1 maintains the Leukemic State through Hox expression. *Cancer Cell*. 2018;34(3). doi:10.1016/j.ccell.2018.08.005
31. Brunetti L, Gundry MC, Goodell MA. New insights into the biology of acute myeloid leukemia with mutated NPM1. *International Journal of Hematology*. 2019;110(2):150-160. doi:10.1007/s12185-018-02578-7
32. Falini B, Bolli N, Liso A, et al. Altered nucleophosmin transport in acute myeloid leukaemia with mutated NPM1: Molecular basis and clinical implications. *Leukemia*. 2009;23(10):1731-1743. doi:10.1038/leu.2009.124
33. Ito K, Sportoletti P, Clohessy JG, Silvia G, Pandolfi PP. The role of nucleophosmin in hematopoietic stem cells and the pathogenesis of myelodysplastic syndrome. *Blood*. 2010;116(21):95-95. doi:10.1182/blood.v116.21.95.95
34. Falini B, Mecucci C, Tiacci E, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *New England Journal of Medicine*. 2005;352(3):254-266. doi:10.1056/nejmoa041974
35. Arnone M, Konantz M, Hanns P, et al. Acute myeloid leukemia stem cells: The challenges of phenotypic heterogeneity. *Cancers*. 2020;12(12):3742. doi:10.3390/cancers12123742
36. Grimwade D, Ivey A, Huntly BJ. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood*. 2016;127(1):29-41. doi:10.1182/blood-2015-07-604496
37. Kroon, E., Krosl, J., Thorsteinsdottir, U., Baban, S., Buchberg, A. M., & Sauvageau, G. (1998). HOXA9 transforms primary bone marrow cells through specific collaboration with meis1a but not pbx1b. *The EMBO Journal*, *17*(13), 3714–3725. https://doi.org/10.1093/emboj/17.13.3714
38. Alharbi, R. A., Pettengell, R., Pandha, H. S., & Morgan, R. (2012). The role of Hox genes in normal hematopoiesis and acute leukemia. *Leukemia*, *27*(5), 1000–1008. https://doi.org/10.1038/leu.2012.356
39. Doulatov, S., Vo, L. T., Chou, S. S., Kim, P. G., Arora, N., Li, H., Hadland, B. K., Bernstein, I. D., Collins, J. J., Zon, L. I., & Daley, G. Q. (2013). Induction of

multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell*, *13*(4), 459–470. https://doi.org/10.1016/j.stem.2013.09.002

40. Collins CT, Hess JL. Deregulation of the HOXA9/meis1 axis in acute leukemia. *Current Opinion in Hematology*. 2016;23(4):354-361. doi:10.1097/moh.0000000000000245
41. Quentmeier H, Reinhardt J, Zaborski M, Drexler HG. FLT3 mutations in acute myeloid leukemia cell lines. *Leukemia*. 2003;17(1):120-124. doi:10.1038/sj.leu.2402740
42. Roychoudhury J, Clark JP, Gracia-Maldonado G, et al. Meis1 regulates an HLF–oxidative stress axis in MLL-fusion gene leukemia. *Blood*. 2015;125(16):2544-2552. doi:10.1182/blood-2014-09-599258
43. Palmqvist L, Pineault N, Wasslavik C, Humphries RK. Candidate genes for expansion and transformation of hematopoietic stem cells by NUP98-hox fusion genes. *PLoS ONE*. 2007;2(8). doi:10.1371/journal.pone.0000768
44. Assi SA, Imperato MR, Coleman DJ, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nature Genetics*. 2019;51(1):151-162. doi:10.1038/s41588-018-0270-1
45. Wang H, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/CAS-mediated genome engineering. *Cell*. 2013;153(4):910-918. doi:10.1016/j.cell.2013.04.025
46. 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. doi:10.1182/blood-2015-11-683649
47. Lavallée V-P, Krosl J, Lemieux S, et al. Chemo-genomic interrogation of CEBPA mutated AML reveals recurrent CSF3R mutations and subgroup sensitivity to JAK inhibitors. *Blood*. 2016;127(24):3054-3061. doi:10.1182/blood-2016-03-705053
48. Simon C, Chagraoui J, Krosl J, et al. A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes & Development*. 2012;26(7):651-656. doi:10.1101/gad.186411.111

49. Choudhary C, Brandts C, Schwable J, et al. Activation mechanisms of STAT5 by oncogenic FLT3-ITD. *Blood*. 2007;110(1):370-374. doi:10.1182/blood-2006-05-024018
50. Maurer B, Kollmann S, Pickem J, Hoelbl-Kovacic A, Sexl V. STAT5A and STAT5B—twins with different personalities in hematopoiesis and leukemia. *Cancers*. 2019;11(11):1726. doi:10.3390/cancers11111726
51. Magnusson M, Brun AC, Miyake N, et al. HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development. *Blood*. 2007;109(9):3687-3696. doi:10.1182/blood-2006-10-054676
52. Argiropoulos, B., & Humphries, R. K. (2007). Hox genes in hematopoiesis and leukemogenesis. *Oncogene*, *26*(47), 6766–6776. https://doi.org/10.1038/sj.onc.1210760
53. Collins CT, Hess JL. Role of HOXA9 in leukemia: Dysregulation, cofactors and essential targets. *Oncogene*. 2015;35(9):1090-1098. doi:10.1038/onc.2015.174
54. Collins CT, Hess JL. Deregulation of the HOXA9/meis1 axis in acute leukemia. *Current Opinion in Hematology*. 2016;23(4):354-361. doi:10.1097/moh.0000000000000245
55. Blasi F, Bruckmann C. Meis1 in hematopoiesis and cancer. how MEIS1-PBX interaction can be used in therapy. *Journal of Developmental Biology*. 2021;9(4):44. doi:10.3390/jdb9040044
56. Shen W-F, Rozenfeld S, Kwong A, Kömüves Laszlo G., Lawrence HJ, Largman C. Hoxa9 forms triple complexes with PBX2 and Meis1 in myeloid cells. *Molecular and Cellular Biology*. 1999;19(4):3051-3061. doi:10.1128/mcb.19.4.3051
57. Martelli MP, Pettirossi V, Thiede C, et al. CD34+ cells from AML with mutated NPM1 harbor cytoplasmic mutated nucleophosmin and generate leukemia in immunocompromised mice. *Blood*. 2010;116(19):3907-3922. doi:10.1182/blood-2009-08-238899
58. Scott DD, Oeffinger M. Nucleolin and nucleophosmin: Nucleolar proteins with multiple functions in DNA repair. *Biochemistry and Cell Biology*. 2016;94(5):419-432. doi:10.1139/bcb-2016-0068

59. Pimentel, H., Bray, N. L., Puente, S., Melsted, P., & Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*, *14*(7), 687–690. https://doi.org/10.1038/nmeth.4324
60. Hayashi H, Kubo Y, Izumida M, Matsuyama T. Efficient viral delivery of Cas9 into human safe harbor. *Scientific Reports*. 2020;10(1). doi:10.1038/s41598-020-78450-8
61. Taxman DJ, Moore CB, Guthrie EH, Huang MT-H. Short hairpin RNA (shrna): Design, delivery, and assessment of Gene Knockdown. *Methods in Molecular Biology*. 2010;629:141-158. doi:10.1007/978-1-60761-657-3_10
62. Ishibashi A, Saga K, Hisatomi Y, Li Y, Kaneda Y, Nimura K. A simple method using CRISPR-Cas9 to knock-out genes in murine cancerous cell lines. *Scientific Reports*. 2020;10(1). doi:10.1038/s41598-020-79303-0
63. 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. doi:10.1126/science.1256337
64. DeWitt MA, Corn JE, Carroll D. Genome editing via delivery of Cas9 ribonucleoprotein. *Methods*. 2017;121-122:9-15. doi:10.1016/j.ymeth.2017.04.003
65. Ayton PM, Cleary ML. Transformation of myeloid progenitors by MLL Oncoproteins is dependent on Hoxa7 and HOXA9. *Genes & Development*. 2003;17(18):2298-2307. doi:10.1101/gad.1111603
66. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143
67. Tuladhar, R., Yeu, Y., Tyler Piazza, J., Tan, Z., Rene Clemenceau, J., Wu, X., Barrett, Q., Herbert, J., Mathews, D. H., Kim, J., Hyun Hwang, T., & Lum, L. (2019). CRISPR-Cas9-based mutagenesis frequently provokes on-target mrna misregulation. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-12028-5

68. Lalonde, S., Stone, O. A., Lessard, S., Lavertu, A., Desjardins, J., Beaudoin, M., Rivas, M., Stainier, D. Y., & Lettre, G. (2017). Frameshift indels introduced by genome editing can lead to in-frame exon skipping. *PLOS ONE*, *12*(6). https://doi.org/10.1371/journal.pone.0178700
69. Martinez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, Embarek AM. Insights into structural and mechanistic features of viral IRES elements. *Frontiers in Microbiology*. 2018;8. doi:10.3389/fmicb.2017.02629
70. Gu X, Ebrahem Q, Mahfouz RZ, et al. Leukemogenic nucleophosmin mutation disrupts the transcription factor hub that regulates granulomonocytic fates. *Journal of Clinical Investigation*. 2018;128(10):4260-4279. doi:10.1172/jci97117
71. Assi, S. A., Imperato, M. R., Coleman, D. J., Pickin, A., Potluri, S., Ptasinska, A., Chin, P. S., Blair, H., Cauchy, P., James, S. R., Zacarias-Cabeza, J., Gilding, L. N., Beggs, A., Clokie, S., Loke, J. C., Jenkin, P., Uddin, A., Delwel, R., Richards, S. J., … Bonifer, C. (2019). Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nature Genetics*, *51*(1), 151–162. https://doi.org/10.1038/s41588-018-0270-1
72. Xiang P, Wei W, Hofs N, et al. A knock-in mouse strain facilitates dynamic tracking and enrichment of MEIS1. *Blood Advances*. 2017;1(24):2225-2235. doi:10.1182/bloodadvances.2017010355
73. Zhang Y-H, Hu Y, Zhang Y, Hu L-D, Kong X. Distinguishing three subtypes of hematopoietic cells based on gene expression profiles using a support vector machine. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 2018;1864(6):2255-2265. doi:10.1016/j.bbadis.2017.12.003
74. Unnisa Z, Clark JP, Roychoudhury J, et al. Meis1 preserves hematopoietic stem cells in mice by limiting oxidative stress. *Blood*. 2012;120(25):4973-4981. doi:10.1182/blood-2012-06-435800
75. Lavallée V-P, Baccelli I, Krosl J, et al. The transcriptomic landscape and directed chemical interrogation of MLL-rearranged acute myeloid leukemias. *Nature Genetics*. 2015;47(9):1030-1037. doi:10.1038/ng.3371

76. Van Maele Bénédicte, De Rijck J, De Clercq E, Debyser Z. Impact of the central polypurine tract on the kinetics of human immunodeficiency virus type 1 vector transduction. *Journal of Virology*. 2003;77(8):4685-4694.
77. Haas DL, Lutzko C, Logan AC, et al. The Moloney murine leukemia virus repressor binding site represses expression in murine and human hematopoietic stem cells. *Journal of Virology*. 2003;77(17):9439-9450. doi:10.1128/jvi.77.17.9439-9450.2003
78. Knott SRV, Maceli AR, Erard N, et al. A computational algorithm to Predict Shrna potency. *Molecular Cell*. 2014;56(6):796-807. doi:10.1016/j.molcel.2014.10.025
79. Manna S, Harman A, Accari J, Barth C. Altering the selection capabilities of common cloning vectors via restriction enzyme mediated gene disruption. *BMC Research Notes*. 2013;6(1). doi:10.1186/1756-0500-6-85
80. Giordano TJ, Deuschle U, Bujard H, McAllister WT. Regulation of coliphage T3 and T7 RNA polymerases by the lac represser-operator system. *Gene*. 1989;84(2):209-219. doi:10.1016/0378-1119(89)90494-0
81. Wheatley RW, Lo S, Jancewicz LJ, Dugdale ML, Huber RE. Structural explanation for allolactose (lac operon inducer) synthesis by lacz β-galactosidase and the evolutionary relationship between Allolactose synthesis and the lac repressor. *Journal of Biological Chemistry*. 2013;288(18):12993-13005. doi:10.1074/jbc.m113.455436
82. Yanisch-Perron C, Vieira J, Messing J. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the m13mpl8 and puc19 vectors. *Gene*. 1985;33(1):103-119. doi:10.1016/0378-1119(85)90120-9
83. Xie J, Tai PWL, Brown A, et al. Effective and accurate gene silencing by a recombinant Aav-compatible MicroRNA scaffold. *Molecular Therapy*. 2020;28(2):422-430. doi:10.1016/j.ymthe.2019.11.018
84. Qin JY, Zhang L, Clift KL, et al. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS ONE*. 2010;5(5). doi:10.1371/journal.pone.0010611

85. Wang X, Xu Z, Tian Z, et al. The EF-1α promoter maintains high-level transgene expression from episomal vectors in transfected Cho-K1 cells. *Journal of Cellular and Molecular Medicine*. 2017;21(11):3044-3054. doi:10.1111/jcmm.13216
86. Zufferey R, Donello JE, Trono D, Hope TJ. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *Journal of Virology*. 1999;73(4):2886-2892. doi:10.1128/jvi.73.4.2886-2892.1999
87. Camps M. Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Patents on DNA & Gene Sequences*. 2010;4(1):58-73. doi:10.2174/187221510790410822
88. Johnson SA, Ormsby MJ, McIntosh A, Tait SWG, Blyth K, Wall DM. Increasing the bactofection capacity of a mammalian expression vector by removal of the F1 Ori. *Cancer Gene Therapy*. 2019;26(7-8):183-194. doi:10.1038/s41417-018-0039-9

Appendices

Appendix A: CD45 knockout and staining

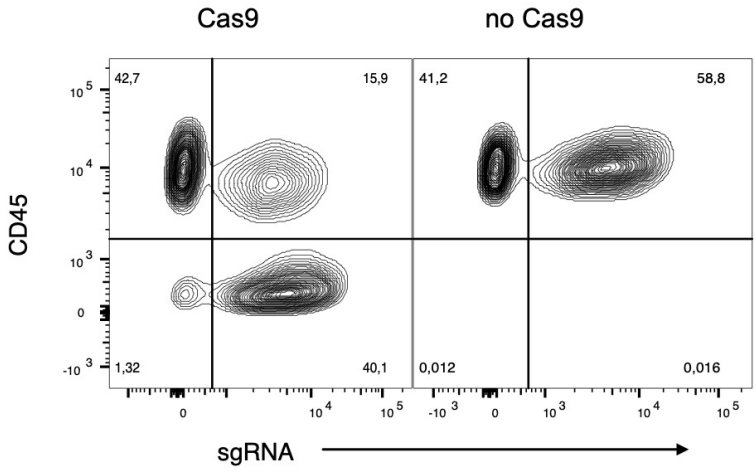


Figure 15. Knockout of CD45. Nucleofection of sgRNA expressing cells was conducted with and without Cas9 and stained for CD45 expression seven days post-nucleofection.

Appendix B: Sample of RScript Used for Counting Cells

Analysis using Rstudio was accomplished thanks to the Rstudio team for their development.

RStudio Team (2020). Rstudio: Integrated Development for R. Rstudio, PBC, Boston, MA URL http://www.rstudio.com/.

```
library(tidyverse)
library(ggplot2)
x <- read_delim('input/numbers.txt', delim = '\t', col_names = c('sample', 'beads', 'cells'), skip = 1) %>%
  separate(col = sample, into = c('tp', 'timepoint', 'sgRNA', 'rep'), sep = '_') %>%
  mutate(rep = gsub('.fcs', '', rep)) %>%
  mutate(rep = gsub('00', '', rep)) %>%
```

```
  select(-rep) %>%
  mutate(tp = gsub('d', '', tp)) %>%
  mutate(tp = as.numeric(tp)) %>%
  group_by(tp) %>%
  arrange(tp)

x1 <- x %>%
  mutate(ul_read = 100*beads/1000) %>%
  mutate(factor = 1000/ul_read) %>%
  mutate(cells_per_1000ul = factor*cells) %>%
  mutate(ul_cells_400k = as.integer(1000*400000/cells_per_1000ul)) %>%
  mutate(fill_up_to_1000ul = 1000-ul_cells_400k) %>%
  mutate(fold_expansion = cells_per_1000ul/400000) %>%
  mutate(pop_doubl = log2(fold_expansion))

x2 <- x1 %>%
  mutate(fold_expansion = case_when(day == 0 ~ 1, T ~ fold_expansion)) %>%
  mutate(pop_doubl = case_when(day == 0 ~ 0, T ~ pop_doubl)) %>%
  group_by(cond, pop, rep) %>%
  mutate(day = as.numeric(day)) %>%
  arrange(day) %>%
  mutate(cum_pd = cumsum(pop_doubl)) %>%
  mutate(sample = paste(cond, pop, rep, sep = '__')) %>%
  ungroup() %>%
  mutate(cond = factor(cond, levels = c('parental', 'HLF14', 'HLF17')))
write_delim(x = x1, path = 'output/table.xls', delim = '\t')
```

Appendix C: Sample of RScript Used to Track and Normalize RFP657 and ZsG Expression

```
library(tidyverse)

library(ggplot2)

# read input file to calculate change in RFP:

x <- read_delim('input/RFP_relative_change.txt', delim = '\t', col_names =
c('sample','RFPneg', 'RFPpos'), skip = 1) %>%

  separate(col = sample, into = c('tp', 'cond', 'carousel','sgRNA', 'rep'), sep = '_') %>%

  mutate(rep = gsub('.fcs', '', rep)) %>%

  mutate(rep = gsub('00', '', rep)) %>%

  select(-rep) %>%

  mutate(tp = gsub('d', '', tp)) %>%

  mutate(tp = as.numeric(tp)) %>%

  group_by(tp) %>%

  arrange(tp,sgRNA) %>%

  select(-carousel) %>%

  select(-cond,-RFPneg)


# annotate the samples

t <- c("AAVS1", "ADGRG1","FLT3","HLF","HOXA10","HOXB8","MEIS1","NKX2-
3","NPM1C","PBX1","PBX3","RPA1","SPI1","STAT5A","STAT5B","ZsG")

t1 <- sapply(t, function(x) rep(x,3)) %>%

  c(.)


s <- c("sg1","sg2","sg3")

s1 <- rep(s,16)


x1 <- x %>%

  mutate(gene = t1) %>%

  mutate(rep = s1)
```

```
# calculate change in RFP relative to day 0
x2 <- x1 %>%
  filter(tp == 0) %>%
  ungroup() %>%
  transmute(sgRNA,RFPref=RFPpos) %>%
  right_join(x1) %>%
  mutate(normalized_RFP = RFPpos/RFPref)

# read input file to calculate change in ZsG:
a <- read_delim('input/ZsG_relative_change.txt', delim = '\t', col_names =
c('sample','RFPnegZsGpos', 'RFPposZsGpos'), skip = 1) %>%
  separate(col = sample, into = c('tp', 'cond', 'carousel','sgRNA', 'rep'), sep = '_') %>%
  mutate(rep = gsub('.fcs', '', rep)) %>%
  mutate(rep = gsub('00', '', rep)) %>%
  select(-rep) %>%
  mutate(tp = gsub('d', '', tp)) %>%
  mutate(tp = as.numeric(tp)) %>%
  group_by(tp) %>%
  arrange(tp,sgRNA) %>%
  select(-carousel,-cond,-RFPnegZsGpos)

# annotate the samples
a1 <- a %>%
  mutate(gene = t1) %>%
  mutate(rep = s1)

# calculate change in ZsG relative to day 0
a2 <- a1 %>%
  filter(tp == 0) %>%
```

```
  ungroup() %>%
  transmute(sgRNA,ZsGref=RFPposZsGpos) %>%
  right_join(a1) %>%
  mutate(normalized_ZsG = RFPposZsGpos/ZsGref)

# combine x2 (change in RFP) and a2 (change in ZsG) data frames
d <- x2 %>%
  right_join(a2)

# rename columns to make more simple for graphing
d1 <- gather(d,value="value", key= "depletion", c(normalized_RFP,normalized_ZsG))
%>%
  mutate(depletion = factor(depletion, levels = c("normalized_ZsG", "normalized_RFP"),
         labels = c("ZsG","RFP657")))

# loop and graph individual genes compared to control
genes <- d1 %>%
  pull(gene) %>%
  unique(.) %>%
  .[-1]
Genes

# before the loop, an empty object is generated that we can deposit the statistics into,
and that will be saved at the end:

stats <- c()

for (g in genes) {
  message(g)
  d2 <- d1 %>%
    filter(gene == "AAVS1" | gene == g) %>%
```

```
  mutate(tp = as.numeric(tp)) %>%
  mutate(value = as.numeric(value)) %>%
  mutate(sample=paste(gene,rep))

# calculate p-values for ZsG:

pz <- d2 %>% filter(depletion == 'ZsG', tp > 0) %>%
  select(gene, tp, value, depletion) %>%
  group_by(tp) %>%
  nest(data = c(gene,value)) %>%                    # nest data for each timepoint
  mutate(pval = map_dbl(data, ~t.test(exact = F, value ~ gene, data = ., alternative =
        'two.sided')$p.value)) %>% # get p value for t-test
  mutate(sc = format(pval, scientific = T, digits = 3)) %>% # turns p-value into scientific
  mutate(significance = case_when(pval <= 0.0001 ~ '****',
                                  pval <= 0.001 ~ '***',
                                  pval <= 0.01 ~ '**',
                                  pval <= 0.05 ~ '*')) # assign asterisks

# calculate p-values for RFP657:
pr <- d2 %>% filter(depletion == 'RFP657', tp > 0) %>%
  select(gene, tp, value, depletion) %>%
  group_by(tp) %>%
  nest(data = c(gene,value)) %>%                    # nest data for each timepoint
  mutate(pval = map_dbl(data, ~t.test(exact = F, value ~ gene, data = ., alternative =
        'two.sided')$p.value)) %>% # get p value for t-test
  mutate(sc = format(pval, scientific = T, digits = 3)) %>% # turns p-value into scientific
  mutate(significance = case_when(pval <= 0.0001 ~ '****',
                                  pval <= 0.001 ~ '***',
                                  pval <= 0.01 ~ '**',
                                  pval <= 0.05 ~ '*')) # assign asterisks
```

```
  # combine p-values:
  pp <- bind_rows(pz, pr) %>%
    mutate(value = 1.03) %>%
    ungroup() %>%
    mutate(gene = g)

  # rbind into stats:
  stats <- bind_rows(stats, pp)

  # plot
  q <- ggplot(d2, aes(x = tp, y = value))
  q <- q + geom_line(aes(x = tp, y = value, group = sample),alpha = 0.6, lwd = 0.04)
  q <- q + facet_grid(facets = depletion ~ .)
  q <- q + geom_point(aes(colour = sample, shape = gene, alpha = gene), size = 2)
  q <- q + geom_text(data = pp, aes(label = significance))
  q <- q + scale_alpha_manual(values=c(0.5, 1.0))
  q <- q + scale_shape_manual(values=c(19, 17))
  q<-q+scale_colour_manual('',
values=c("grey60","grey40","grey20","sienna1","red","firebrick3"))
  q <- q + scale_y_continuous(limits = c(0,1.07), breaks = c(0, 0.25, 0.5, 0.75, 1))
  q <- q + scale_x_continuous(breaks = unique(d2$tp))
  q <- q + theme_bw() + theme(legend.position = 'right',
                              strip.text = element_text(face = 'bold', size = 14),
                              axis.title = element_text(size = 14),
                              plot.title = element_text(face = 'bold.italic', size = 14))
  q <- q + ylab("normalized to d0") + xlab('time (days)') + ggtitle(g) + guides(alpha =
        FALSE, shape = FALSE)
  q
```

```
  ggsave(q, filename = paste0('output/ZsG_vs_RFP_change/',g,'.pdf'), width = 5, height
= 5)
  ggsave(q, filename = paste0('output/ZsG_vs_RFP_change/',g,'.png'), width = 5, height
= 5, dpi = 300)  }


# save the stats that were compiled during each loop:


stats <- stats %>% select(gene, timepoint = tp, depletion, data, pval, sc, significance)
save(stats, file = 'output/statistics_summary.rdata')
stats1 <- stats %>% select(-data)
write_delim(x = stats1, file = 'output/statistics_summary.xls')
stats1 <- stats %>% select(-data)
write_delim(x = stats1, file = 'output/statistics_summary.xls', delim = '\t'
```

Appendix D: CRISPR sgRNAs Which did Not Produce Significant Results

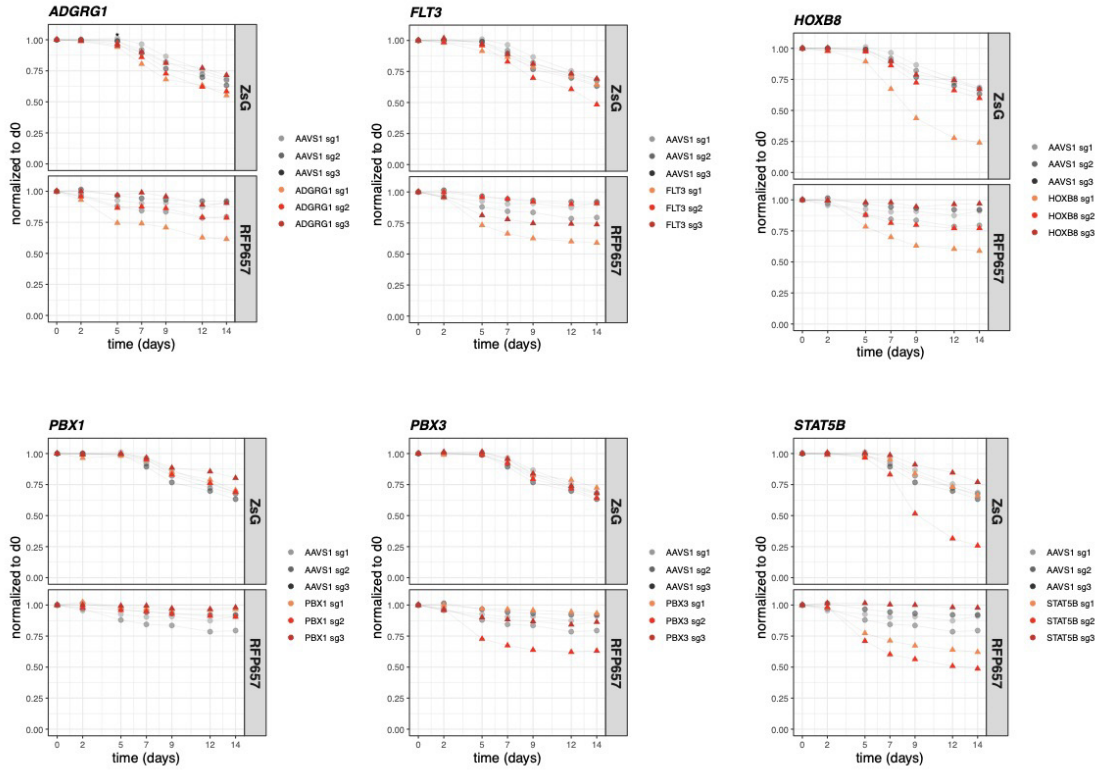


Figure 16. Non-Significant Results from CRISPR/Cas9 Knockouts. CRISPR/Cas9 knockouts which did not produce a statistically significant change in ZsG or RFP657 expression. From left to right then top to bottom: ADGRG1 (GPR56), FLT3, HOXB8, PBX1, PBX3, STAT5A.

Appendix E: Testing the Clonogenicity of IMS-M2 Cells in Methylcellulose

Number of cells plated	Cytokines added	Number of colonies
100	Human	0
100	None	0
500	Human	7
500	None	0
5000	Human	22
5000	None	0
10,000	Human	75
10,000	None	0
15,000	Human	~ 200

Table 1. Clonogenicity of IMS-M2 cells in methylcellulose. Colonies produced from unmodified IMS-M2 cells upon plating of different cell numbers in methylcellulose, with and without the addition of human cytokines.

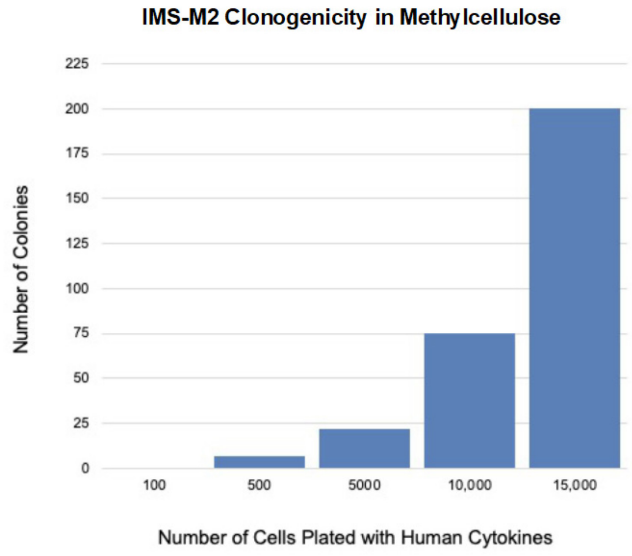


Figure 17. Number of colonies produced from IMS-M2 cells after plating different cell number with human cytokines.

Methylcellulose Cultures of Human Myeloid Clonogenic Progenitors

	Final concentration	For 30 mL, add	for 40 ml	par tubes	2 x 40ml	3 x 40 ml	4 x40 ml	5X 40 ml
Methylcellulose, 2% stock	1.04%	17 mL	20 ml	20 ml				
IMDM	4.5ml	4.5 mL	8.5 ml **	9	18	27	36	45
HI FBS	20%	6 ml	8 ml	8				
Deionized BSA, 30% stock	1%	1 mL	1.33 ml	1.3	2.6	3.9	5.2	6.5
Glutamine, 200 mM	2 mM	300 ul	400 ul	0.4	0.8	1.2	1.6	2
Holo-transferrin, 30 mg/mL	200 µg/mL	200 µL	266 ul	0.266	0.532	0.798	1.064	1.33
β-mercaptoethanol, 10^{-2}M	10^{-4}M	300 ul	400 ul	0.4	0.8	1.2	1.6	2
hu-SCF* (60ng/µL)	100 ng/mL	60 µL	67 ul	0.067	0.134	0.201	0.268	0.335
hu-IL-3*(80 ng/µL)	10 ng/mL	4 µL	5 ul	0.005	0.01	0.015	0.02	0.025
GM-CSF (100 ug/mL)	10 ng/mL	3 ul	4 ul	0.004	0.008	0.012	0.016	0.02
Epo* (220 U/mL)	3 U/mL	409 µL	545 ul	0.545	1.09	1.635	2.18	2.725
hu-IL-6* (100 ug/mL)	10 ng/mL	3 µL	4 ul	0.004	0.008	0.012	0.016	0.02
Tpo* (4 ng/µL)	50 ng/mL	375 µL	500 ul	0.5	1	1.5	2	2.5
				20.491	24.982			
		6.854 ml/30 ml			12.491	par tube		

** ajouté 500 ul car filtré

Table 2. Reagents in methylcellulose used for cloning human cell lines

Appendix F: Ultra-miR Cloning by Gibson Assembly Protocol

Ultra-miR cloning

Date:

PCR amplify UmiRs	ul/ reaction	X____rxns
Water	31	
5X GC buffer	10	
DMSO	1.5	
dNTPs (10mM ea)	1	
HairpinHpaF (10uM)	2.5	
HairpinHpaR (10uM)	2.5	
97mer template (1uM)	1	-
Phusion polymerase	0.5	
Total volume	50ul	49ul ea + template

Digest 1 - 5ug UmiR vector	
vector # concentration	
volume vector DNA	
volume water	
10X FD	3ul
HpaI	0.5ul
FastAP	0.5ul
total volume	30ul

98° 30sec / (98° 10s / 72° 10s) X 18 cycles / 72° 2min / 4° hold
check 5ul on gel (210bp product)

Purify both cut vector and PCR product with PCR purification kit
Invitrogen PureLink PCR purification kit -> elute in 50ul

thaw 1 tube/rxn of 5ul 2X **GA** mix on ice

keeping **on ice**, add:	
HpaI cut vector (~50ng)	1ul
PCR *	1ul
water	3ul

*include neg ctrl of vector only
mix, and place into thermocycler that is at **room temperature with lid pre-heated**
Start GA program on thermocycler:
50 degrees for 1hr -> hold at 4

Transform 2ul GA mix + 50ul Stbl4 Mix&Go
(or 5ul GA + 100ul bugs if inefficient)

Miniprep 2-4 colonies/sh

Screen clones	/rxn	X__rxn
Plasmid DNA	10ul	---
10X FD green	2ul	
*MluI	0.25ul	
*SalI	0.25ul	
water	7.5ul	
total	20ul	10ul ea

= 330bp with sh insert vs. 213bp empty
* MluI + AgeI for #3092

sequence clone#:
primer:
seq request #:

Seq results:

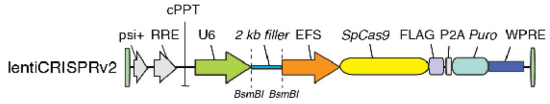
Appendix G: Genome-Scale CRISPR Knock-Out (GeCKO) Target Guide Sequence Cloning Protocol

GeCKO Genome-scale CRISPR Knock-Out | LENTIVIRAL CRISPR TOOLBOX | MIT ZhangLab

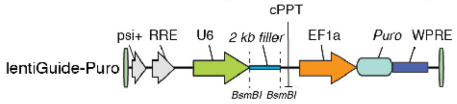
LentiCRISPRv2 and lentiGuide-Puro: lentiviral CRISPR/Cas9 and single guide RNA

CRISPR (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats) is a microbial nuclease system involved in defense against invading phages and plasmids. CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage. Lentiviral CRISPR/Cas can infect a broad variety of mammalian cells by co-expressing a mammalian codon-optimized Cas9 nuclease along with a single guide RNA (sgRNA) to facilitate genome editing (Shalem*, Sanjana*, *et al.*, Science 2014). Protocols for cloning into the lentiviral transfer plasmid and general considerations for producing lentivirus are described below. **Separate protocols are available for amplifying the genome-scale CRISPR knock-out (GeCKO) libraries. This protocol is for creating individual lentiviral CRISPR plasmids targeting a single genomic locus.**

lentiCRISPRv2 (one vector system): This plasmid contains two expression cassettes, hSpCas9 and the chimeric guide RNA. The vector can be digested using *BsmB*I, and a pair of annealed oligos can be cloned into the single guide RNA scaffold. The oligos are designed based on the target site sequence (20bp) and needs to be flanked on the 3' end by a 3bp NGG PAM sequence, as shown on the next page.



lentiGuide-Puro (two vector system): This plasmid expressed only the chimeric guide RNA. It does *not* contain Cas9. Please use lentiCas9-Blast (a separate lentiviral construct that delivers hSpCas9 and blasticidin resistance) to first integrate Cas9 into your cell line. The lentiGuide-Puro vector can be digested using *BsmB*I, and a pair of annealed oligos can be cloned into the single guide RNA scaffold. The oligos are designed based on the target site sequence (20bp) and needs to be flanked on the 3' end by a 3bp NGG PAM sequence, as shown on the next page.



Which vector to use: lentiCRISPRv2 is identical to the original lentiCRISPRv1 but produces nearly 10X higher titer virus. lentiGuide-Puro produces >100X higher titer virus over lentiCRISPRv1 and should be used in cell lines where Cas9 has already been integrated in (e.g. using the separate lentiCas9-Blast lentivirus). For applications where Cas9 cannot first be introduced (e.g. primary cells), lentiCRISPRv2 is recommended. After transduction, use puromycin to select for cells with lentiCRISPRv2 or lentiGuide-Puro.

Lentiviral production: Before starting any lentiviral work, please ensure compliance with your Environmental Health and Safety office and government/organization/university. Briefly, to make lentivirus, a transfer plasmid (e.g. lentiCRISPRv2 or lentiGuide-Puro) must be co-transfected into HEK293(F)T cells with the packaging plasmids pVSVg (AddGene 8454) and psPAX2 (AddGene 12260). As a positive control for viral production, we often use a CMV-EGFP lentiviral transfer plasmid (eg. AddGene 19319).

Target design notes and online resources: For application of Cas9 for site-specific genome editing in eukaryotic cells and organisms, we have computationally identified suitable target sites for the *S. pyogenes* Cas9 and calculated most likely off-targets within the genome. Please visit *http://www.genome-engineering.org* to access these Cas9 target design tools. Complete plasmid sequences, protocols, a discussion forum and additional information can be found at the Zhang Lab GeCKO website: *http://www.genome-engineering.org/gecko/* .

Citation: Please reference the following publications for the use of this material.

Improved lentiviral vectors and genome-wide libraries for CRISPR screening. Sanjana NE*, Shalem O*, Zhang F. *Nature Methods* (2014).

Genome-scale CRISPR-Cas9 knockout screening in human cells. Shalem O*, Sanjana NE*, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014). *Science*, 343, 83-7. DOI: 10.1126/science.1247005

Target Guide Sequence Cloning Protocol

In order to clone the target sequence into the lentiCRISPRv2 or lentiGuide-Puro backbone, synthesize two oligos of the following form. All plasmids have the same overhangs after *BsmBI* digestion and the same oligos can be used for cloning into lentiCRISPRv2, lentiGuide-Puro or lentiCRISPRv1.

***Example oligo design*: Note that the NGG PAM is not included in the designed oligos.**

***Oligonucleotide ordering tips*: Standard de-salted oligos (usually the most inexpensive synthesis) are sufficient for cloning. If not already resuspended, dilute each oligo to 100 μM in sterile water or TE.**

* * * * *

Lentiviral vector digestion, oligo annealing and cloning into digested vector:

1. Digest and dephosphorylate 5ug of the lentiviral CRISPR plasmid with *BsmBI* for 30 min at 37C:

5 ug	lentiCRISPRv2 or lentiGuide-Puro
3 ul	FastDigest *BsmBI* (Fermentas)
3 ul	FastAP (Fermentas)
6 ul	10X FastDigest Buffer
0.6 ul	100 mM DTT (freshly prepared)
X ul	ddH_2O
60 ul	total

2. Gel purify digested plasmid using QIAquick Gel Extraction Kit and elute in EB.

If *BsmBI* digested, a ~2kb filler piece should be present on the gel. Only gel purify the larger band. Leave the 2kb band.

3. Phosphorylate and anneal each pair of oligos:

1 ul	Oligo 1 (100 μM)
1 ul	Oligo 2 (100 μM)
1 ul	10X T4 Ligation Buffer (NEB)
6.5 ul	ddH_2O
0.5 ul	T4 PNK (NEB M0201S)
10 ul	total

Please use the T4 Ligation Buffer since the buffer supplied with the T4 PNK enzyme does not include ATP (or supplement to 1mM ATP).

Put the phosphorylation/annealing reaction in a thermocycler using the following parameters:

37°C	30 min
95°C	5 min and then ramp down to 25°C at 5°C/min

4. Dilute annealed oligos from Step 3 at a 1:200 dilution into sterile water or EB.

5. Set up ligation reaction and incubate at room temperature for 10 min:

X ul	*BsmBI* digested plasmid from Step 2 (50ng)
1 ul	diluted oligo duplex from Step 4
5 ul	2X Quick Ligase Buffer (NEB)
X ul	ddH_2O
10 ul	subtotal
1 ul	Quick Ligase (NEB M2200S)
11 ul	total

Also perform a negative control ligation (vector-only with water in place of oligos) and transformation.

6. Transformation into Stbl3 bacteria.
Lentiviral transfer plasmids contain Long-Terminal Repeats (LTRs) and must be transformed into recombination-deficient bacteria. We use homemade Stbl3 (propagated from Invitrogen C7373-03) and get excellent plasmid yields. Although other RecA- strains may work, we have found the most consistent transformations and yields using Stbl3.

Appendix H: Transplantation Analysis

Ab	# Molecula	Fluorochrome	Dilution
mCD45	613	APC Cy7	100
huCD45	612	PB	50
CD34	419	APC	50
CD38	608	PE Cy7	50
GPR56	751-FST	PE	50
Blocking 10x			

Table 3: Antibodies used for cell staining to determine engraftment parameters.

mouse_id	group	%Human	%ZsG+		%Human	%ZsG+		%Human	%ZsG+		
25955	HLF 14 POS wk4__BM_25955_011.fcs	31.4	27	wk4__PB_25955_001.fcs	32.4	24	wk4__SPL_25955_006.fcs	17.4	30.8	sacrifice 17 dec	
25956	HLF 14 POS wk4__BM_25956_012.fcs	24.9	41.1	wk4__PB_25956_002.fcs	28.7	32.9	wk4__SPL_25956_007.fcs	34.5	33.8	sacrifice 17 dec	
25957	HLF 14 POS wk4__BM_25957_013.fcs	31.3	41.9	wk4__PB_25957_003.fcs	9	30.7	wk4__SPL_25957_008.fcs	18.1	25.2	sacrifice 17 dec	
25958	HLF 14 POS wk4__BM_25958_014.fcs	51.6	37.9	wk4__PB_25958_004.fcs	13.6	36.4	wk4__SPL_25958_009.fcs	30.7	32.2	sacrifice 17 dec	
25959	HLF 14 POS wk4__BM_25959_015.fcs	37.7	31.4	wk4__PB_25959_005.fcs	12.1	26.3	wk4__SPL_25959_010.fcs	21.8	23.4	sacrifice 17 dec	
25960	HLF 14 POS wk4__BM_25960_017.fcs	50.7	37.4	wk4__PB_25960_001.fcs	16.4	28.4	wk4__SPL_25960_009.fcs	22.2	30.1	sacrifice 18 dec	
25961	HLF 14 NEG wk4__BM_25961_018.fcs	41.9	4.94	wk4__PB_25961_002.fcs	11.4	8.11	wk4__SPL_25961_010.fcs	6.35	9.04	sacrifice 18 dec	
25962	HLF 14 NEG wk4__BM_25962_019.fcs	34.2	5.54	wk4__PB_25962_003.fcs	7.04	5.1	wk4__SPL_25962_011.fcs	4.65	4.27	sacrifice 18 dec	
25963	HLF 14 NEG wk4__BM_25963_001.fcs	56.3	9.19							sacrifice 19 dec	higher CD34 and CD38
25964	HLF 14 NEG wk4__BM_25964_020.fcs	60	7.7	wk4__PB_25964_004.fcs	62.2	12.4	wk4__SPL_25964_012.fcs	13.1	6.55	sacrifice 18 dec	
25965	HLF 14 NEG wk4__BM_25965_021.fcs	35.6	9.66	wk4__PB_25965_005.fcs	12.5	10.5	wk4__SPL_25965_013.fcs	6.02	8.61	sacrifice 18 dec	
25966	HLF 14 NEG wk4__BM_25966_002.fcs	71.1	8.24							sacrifice 19 dec	higher CD34 and CD38
25967	HLF 17 POS wk4__BM_25967_003.fcs	36.5	29.1							sacrifice 19 dec	
25968	HLF 17 POS wk4__BM_25968_004.fcs	21.3	21.1							sacrifice 19 dec	
25969	HLF 17 POS wk4__BM_25969_005.fcs	10.4	7.52							sacrifice 19 dec	
25970	HLF 17 POS wk4__BM_25970_006.fcs	14.2	32.9							sacrifice 19 dec	
25971	HLF 17 POS wk4__BM_25971_007.fcs	42.1	41.6							sacrifice 19 dec	
25972	HLF 17 POS wk4__BM_25972_022.fcs	22.3	49.4	wk4__PB_25972_006.fcs	20.1	41.3	wk4__SPL_25972_014.fcs	3.3	22.1	sacrifice 18 dec	
25973	HLF 17 NEG wk4__BM_25973_008.fcs	82.7	7.23							sacrifice 19 dec	
25974	HLF 17 NEG wk4__BM_25974_023.fcs	33.5	5.04	wk4__PB_25974_007.fcs	5.1	3.67	wk4__SPL_25974_015.fcs	2.98	4.44	sacrifice 18 dec	
25975	HLF 17 NEG wk4__BM_25975_024.fcs	29	5.82	wk4__PB_25975_008.fcs	23.1	2.61	wk4__SPL_25975_016.fcs	8.32	3.45	sacrifice 18 dec	
25976	HLF 17 NEG wk4__BM_25976_009.fcs	16.7	5.82							sacrifice 19 dec	
25977	HLF 17 NEG wk4__BM_25977_010.fcs	49.3	9.32							sacrifice 19 dec	
25978	HLF 17 NEG wk4__BM_25978_011.fcs	19	4.39							sacrifice 19 dec	
	Mean	23.8	20.5								

Table 4: Human engraftment (%) and ZsG+(%) of each mouse used in xenotransplantation experiments.