PhD degree in Molecular Medicine

Curriculum in Molecular Oncology

European School of Molecular Medicine (SEMM)

University of Milan and University of Naples "Federico II"

Faculty of Medicine (MED/4)

in vivo shRNA screening to identify quiescence-related genes required for AML growth

Giulia De Conti

European Institute of Oncology (IEO), Milan

Student ID: R10310

Supervisor: Prof. Pier Giuseppe Pelicci

Added Supervisor: Dr. Emanuela Colombo

Academic year 2016-2017

Table of Content

1. List of Abbreviations

5-FU	5-fluorouracil
AML	Acute Myeloid Leukemia
APL	Acute Promyelocytic Leukemia
BC	Barcode
BM	Bone Marrow
BM MNC	Bone Marrow Mono-Nucleated Cell
bp	Base pair
CMP	Common Myeloid Progenitor
CSC	Cancer Stem Cell
ELDA	Extreme Limiting Dilution Assay
FAB	French-American-British
FACS	Fluorescence-Activated Cell Sorter
FBS	Fetal Bovine Serum
FCS	Fetal Calf Serum
FDA	Food and Drug Administration
CML	Chronic Myeloid Leukemia
GFP	Green Fluorescent Protein
GMP	Granulocyte-Macrophage Progenitor
GSEA	Gene Set Enrichment Analysis
HSC	Hematopoietic Stem Cell
KI	Knock-in
KO	Knock-out
LDTA	Limiting Dilution Transplantation Assay
LIC	Leukemia Initiating Cell
Lin	Lineage markers
LT-HSC	Long Term-Hematopoietic Stem Cell
LSC	Leukemia Stem Cell
MOI	Multiplicity Of Infection
MPP	Multi Potent Progenitor
NGS	Next Generation Sequencing
NK	Normal Karyotype
PB	Peripheral Blood
PRKI	PML-RARα knock-in
qPCR	Quantitative Polymerase Chain Reaction
RBC	Red Blood Cell
RFP	Red Fluorescent Protein
RT	Room Temperature
shRNA	Short hairpin RNA
SPL	Spleen
ST-HSC	Short Term-Hematopoietic Stem Cell
TU	Transducing Unit
WHO	World Health Organization
WT	Wild-Type

2. Figures Index

3. Tables index

4. Abstract

AML is hierarchically organized with at the apex Leukemia Stem Cells (LSCs), a rare cell population able to initiate and sustain the tumor growth. LSCs share many functional properties with normal Hematopoietic Stem Cells (HSCs) including self-renewal capacity and quiescence. Quiescent LSCs can survive to radiation and chemotherapy acting as a reservoir for leukemia relapse, the major cause of death for AML patients. Therefore, LSCs quiescence is critical for leukemia maintenance and few evidences suggest that quiescence regulation in pre-leukemic phase plays a pivotal role for leukemogenic process as well.

In this work, we demonstrated that the expression of NPMc+ or PML-RARα in HSCs is sufficient to enforce a quiescent stem cell gene expression profile. Therefore, we hypothesized that enhancement of the quiescent phenotype in HSCs could be a shared mechanism for leukemia development and maintenance. As an approach to examine the contribution of representative quiescence related genes in AML, we exploited RNA interference technology to perform *in vivo* screening. Among the target genes we found depleted during the screening, silencing of Stat1 or Sytl4 in AML blasts was sufficient to significantly decrease *in vitro* self-renewal and delay leukemia growth *in vivo*.

5. Introduction

5.1 Hematopoiesis is hierarchically organized with HSCs at the apex

The human body produces millions of blood cells every second for the entire lifetime[1]. Since most of the various mature blood cells are short-lived, Hematopoietic Stem Cells (HSCs) have the role of continuously replenish multipotent progenitors and committed precursors in the process called hematopoiesis. HSCs are the rare cell population at the top of the well characterized hematopoietic hierarchy, which is depicted as a series of branches that become progressively restricted towards each hematopoietic lineage (Figure 1). As other stem cells, HSCs are able to self-renew and to differentiate, producing all the mature cells of the blood, both myeloid and lymphoid. Self-renewal is defined as the property of every stem cell to, upon division, produce one (asymmetric division) or two (symmetric division) daughter cells which retain the capacity for self-renewal, ensuring a long-term maintenance or expansion of the stem cell compartment.

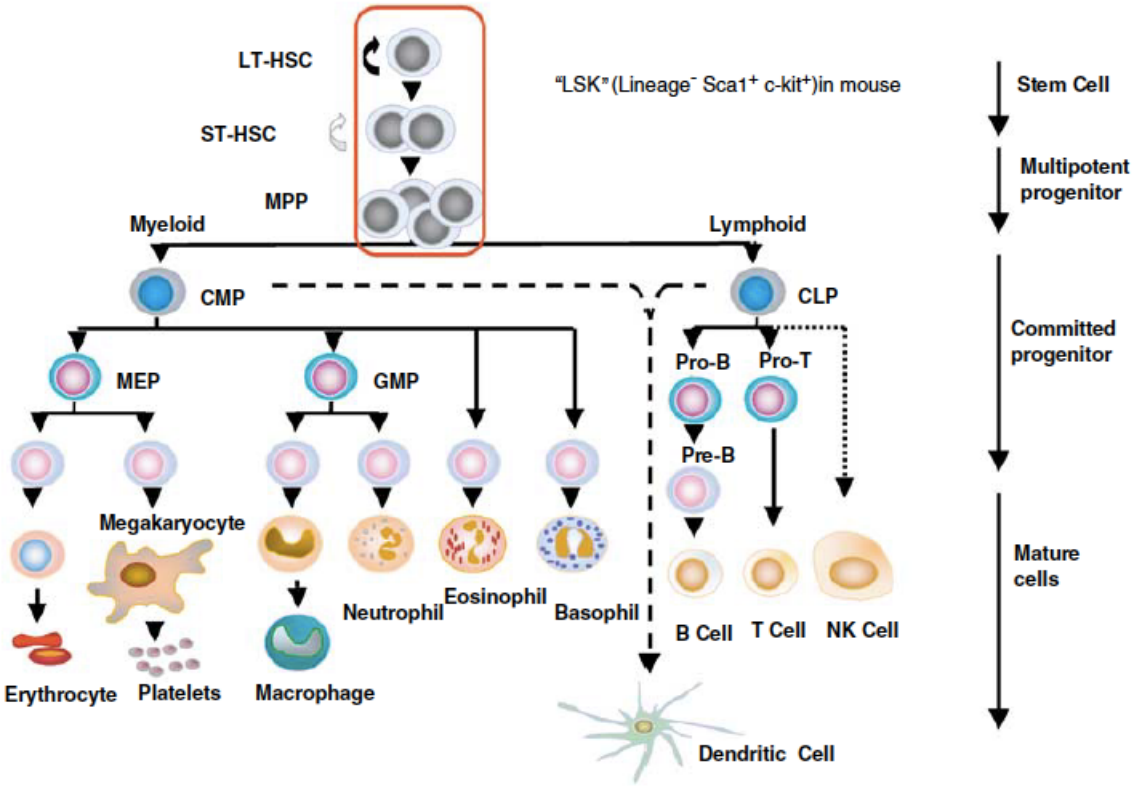

Figure 1. The hierarchical organization of hematopoietic cells. In mice, the Lin- Sca1+ cKit+ compartment is enriched in stem and multipotent progenitor cells that gradually differentiates towards the hematopoietic lineages (adapted from Larsson and Karlsson, Oncogene 2005).

In mammals, the first "primitive" hematopoiesis wave arises from the yolk sac, followed by the aorta-gonad-mesonephros (AGM) region of the embryo, the fetal liver and, in the adult, HSCs reside in the bone marrow (BM). HSCs are highly dependent on their niche for the regulation of self-renewal and differentiation potential, and their properties may vary depending on the different environment. For instance, while adult HSCs are mostly quiescent, when in the fetal liver they are mostly proliferating, to ensure the stem cell expansion needed for embryogenesis. In the adult bone marrow, HSCs are found to reside in two distinct niches, the endosteal and the vascular region[2-5] (Figure 2). Bone marrow cells synthetize important factors for HSCs such as thrombopoietin (TPO)[6], stem cell factor (SCF)[7] and CXC-chemokine ligand 12 (CXCL12)[8]. The niche microenvironment is therefore involved in HSCs quiescence and self-renewal control.

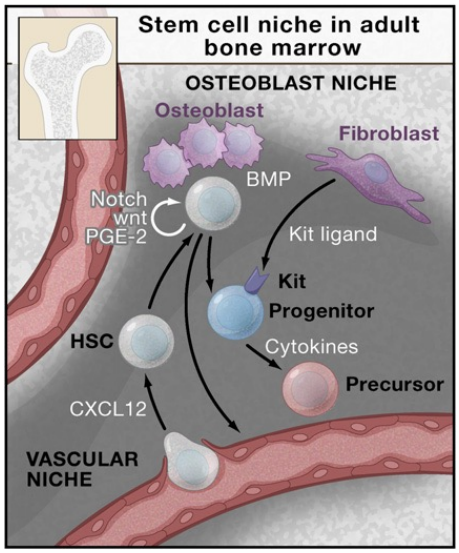

Figure 2. Adult HSCs reside in specialized niches in the bone marrow. HSCs are found both adjacent to osteoblasts (the osteobast niche) and to blood vessels (the vascular niche). Bone marrow also contains stromal cells that support hematopoiesis producing cytokines such as c-Kit ligand, interleukins, thrombopoietin (TPO), and erythropoietin (EPO) necessary for HSCs self-renewal and differentiation (adapted from Orkin and Zon, Cell 2008).

In homeostatic conditions, HSCs are mostly quiescent and only a small pool actively contributes to blood production[9]. Quiescence is defined as the reversible absence of cycling, also named G0 phase of the cell cycle. The ability of HSCs to remain in the quiescent phase of the cell cycle is thought to be fundamental for stemness maintenance, even though a rapid cell cycle entry is required for the efficient response to conditions of hematopoietic stress such as bone marrow injury. In 1999, Weissman and colleagues analyzed the kinetics of the murine long-term HSCs (LT-HSCs) cell cycle demonstrating that, in homeostatic conditions, about 75% are in G0 phase and are recruited into cell cycle every 57 days, on average[10]. More recent works confirmed these findings and computational models suggested that adult quiescent HSCs divide approximately five times during the mouse lifetime[9].

Stem cell quiescence is actively regulated by intrinsic mechanisms and microenvironmental signals, both converging towards characteristic transcriptional, epigenetic and metabolic profiles[11,12]. Quiescence has been linked to HSCs long-term potential, necessary for the maintenance of blood compartment throughout the whole lifetime of an individual.

Functional HSCs maintain the capacity to rapidly enter cell cycle in response to physiological stresses such as blood loss. Loss of HSCs quiescence, as following hyper-proliferative signals, leads to depletion of HSCs pool, accumulation of DNA damage and exhaustion of self-renewal potential[13,14]. Therefore, genes involved in quiescence regulation are critical for maintain self-renewal capacity in stem cells.

5.2 Use of surface markers to identify HSCs

In the last 30 years, many groups focused on the prospective isolation of HSCs. Florescence-activated cell sorting (FACS) technologies have improved the accuracy of single cell isolation according to surface markers expression or functional properties.

Thought the murine hematopoietic system has been more extensively characterized, it shares many critical features with the human, starting from the hierarchical organization. However, differences in the surface markers include the expression of the CD34 molecule that is exclusively present on human HSCs. Indeed, murine LT-HSCs are typically isolated using a combination of surface markers including Lin^-, $Sca1^+$, $cKit^+$, $CD34^-$ and $Flk2^-$ (Weissman approach) or Lin^-, $Sca1^+$, $cKit^+$, $CD150^+$ and $CD48^-$ (SLAM compartment)[4,15]. More recently, highly quiescent LT-HSCs have been isolated as $CD34^-$ $Flk2^-$ cells within the SLAM compartment[9]. Human HSCs, instead, have been historically described as enriched in Lin^-, $Thy1^+$, $CD34^+$, $CD38^{neg/low}$ cells[16]. More recently, Dick and colleagues described a combination of surface markers which is highly enriched in human functionally validated HSCs[17] and includes Lin^-, $Thy1^+$, $CD34^+$, $CD38^-$, $CD45RA^-$ and $CD49f^+$. Moreover, some functional properties can be used to enriched for HSCs, for instance the major efflux of Hoechst 3334 in the side-population[18] or their enhanced label retaining properties, upon BrdU or H2B-GFP labeling[9].

Recent studies have challenged the traditional hierarchical organization of hematopoietic system[19,20]. Dick and colleagues, combining single cell-sorting strategy and functional analysis, demonstrate that hematopoietic hierarchy might not follow the classical model. In

their model, oligopotent progenitors are the main component in the fetal liver hematopoiesis, while, in the adult BM, unilineage progenitors with myeloid or erythroid potential are prevalent. Human definitive hematopoiesis can be therefore subdivided in two main compartments with HSCs and MPPs at the top and committed unipotent progenitors at the bottom (Figure 3)[19].

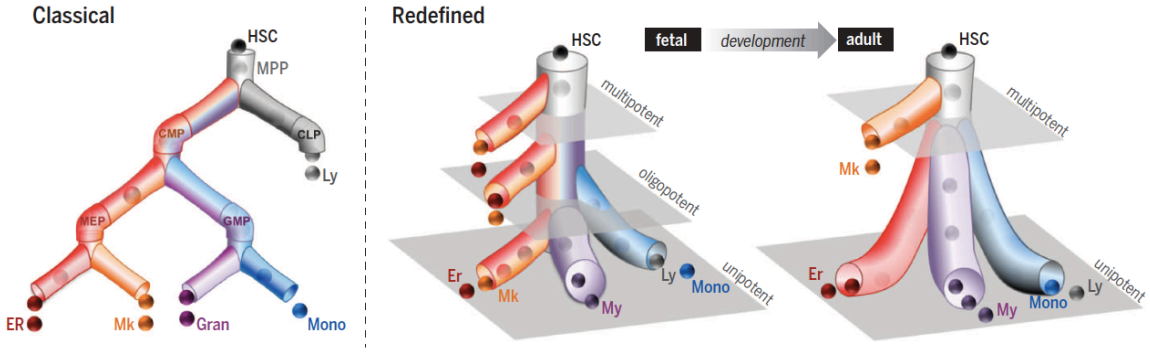

Figure 3. Redefined model for human hematopoietic system organization. Recent studies demonstrated that adult hematopoiesis is mainly organized two compartment: a multipotent one formed by HSCs and MPPs and a unipotent compartment with lineages committed progenitors (adapted from Notta et al., Science 2016).

5.3 Quiescence regulates HSCs self-renewal

Every HSCs fate decision, such cell division, differentiation, migration, cell death or self-renewal, must be highly regulated to avoid HSCs exhaustion and bone marrow failure, on one side, and uncontrolled proliferation and leukemic transformation, on the other.

Cell cycle is mainly controlled by cyclins, cyclin-depenent kinases (CDKs) and their inhibitors (CKIs, CDK inhibitors). Several CKIs of either the INK4 or CIP/KIP gene families have been described to play a critical role in regulation of HSCs quiescence. p21 (Cdkn1a), is a CKI belonging to the CIP/KIP family and it is one of the major G1 phase checkpoint regulator. p21 knock-out (KO) mice show increased HSCs number and proliferation in homeostatic conditions. Upon myelo-ablative stress, such as 5-fluorouracil (5-FU) treatment, p21 KO mice die due to HSCs depletion. Furthermore, serial bone marrow transplantation (BMT) highlighted an impairment in self-renewal potential of p21 KO

HSCs[21]. Another member of the CIP/KIP family, p57 (Cdkn1c), is highly expressed in quiescent HSCs and progressively decreases during differentiation toward multipotent progenitors (MPPs)[22]. p57 deficient HSCs are less quiescent and are defective in self-renewal potential. Moreover, abnormalities in p57 KO HSCs can be corrected by knocking p27 (Cdkn1b), another CKI, in the p57 locus, suggesting a functional overlap between the two genes in control HSCs homeostasis[23].

Cdk6 promotes cell-cycle progression and has the opposite role of the above described CKIs. Cdk6 ablation does not affect HSCs in homeostatic conditions, yet Cdk6 KO HSCs can not be activated upon proliferation signals and do not efficiently repopulate mice in competitive BMT[24]. Consistently, Cdk6 is non expressed in LT-HSCs, while it is present at high levels in short term-HSCs (ST-HSCs), where it may mediate rapid cell cycle entry upon stimulation. Thus, both LT- and ST-HSCs are quiescent, but the higher level of Cdk6 expression suggests that the ST-HSCs population is primed to enter proliferation. Enforced Cck6 expression in LT-HSCs is sufficient to shorten quiescence exit without impact on functionality. Therefore Cdk6 absence preserve LT-HSC pool from exhaustion[25].

Cell cycle entry and quiescence of HSCs are also regulated by many transcription factors, including Gata2, Gfi1 and Cited2. Among them, Gata2 is highly expressed in quiescent hematopoietic cells and Gata2 KO mice die during embryonic development for defects in HSCs[26]. Enforced GATA2 expression in human cord blood $CD34^+$ cells increases quiescence but reduces proliferation and performance in long-term culture-initiating cell assay (LTC-IC). Indeed, human HSCs expressing high level of exogenous GATA2 fail to contribute to hematopoiesis in NOD-SCID mice while $GATA2^{low}$ cells delay hematopoiesis contribution, remaining quiescent[27]. Therefore, Gata2 promotes quiescence both *in vitro* and *in vivo* and its expression must be fine tuned to maintain correct HSCs functionality.

Another factor involved in quiescence regulation is Gfi1, a zinc finger transcriptional repressor. Gfi1 KO mice show increased number of LT-HSCs in their bone marrow due to increased proliferation. Gfi1 KO HSCs, however, have decreased long-term reconstitution

capacity in competitive BMT, thus establishing an inverse correlation between HSCs proliferation and self-renewal potential[28].

Lastly, overexpression of the transcriptional coactivator CITED2 in human CD34+ cells resulted in enhanced HSCs colony-forming efficiency in vitro and a higher engraftment capacity in NSG mice. CITED2 enhances quiescence of $CD34^{+}CD38^{-}$ HSCs, in part by increasing p21 expression[29].

Soluble factors and cytokines also play a role in HSCs quiescence regulations, pointing out the active and fundamental role of the niche microenvironment. Thrombopoietin (TPO) signaling, mediated by the interaction between TPO and its receptor MPL, is critical for maintaining quiescence and self-renewal in adult HSCs[30]. HSCs that reside in the endosteal niche next to osteoblasts (which produce TPO) also express MPL. Inhibition of TPO/MPL signaling results in loss of quiescent HSCs while its activation results in p57 up-regulation[6].

Another example of environmental control of quiescence is the transforming growth factor-beta (TGFβ), a secreted factor present at high levels in the bone marrow and a well described negative regulators of hematopoiesis. However, conditional deletion of TGFβ receptor I (Tgfbr1) does not affect normal hematopoiesis and bone marrow cells are not defective in their ability to long-term repopulate recipient mice in serial BMT[31]. Accordingly, in homeostatic conditions, TGFβ signaling inhibition does not induce HSCs proliferation *in vivo*. On the contrary, under stress conditions, such as treatment with chemotherapeutic agents, TGFβ blockage accelerates hematopoietic reconstitution and delays HSCs re-enter in the G0 phase. p57 has been identified as one of the key downstream mediator of TGFβ[32].

Furthermore, a recent work showed that TGFβ is highly expressed by megacaryocites (MK) in the bone marrow niche, and MK ablation results in HSCs exit from quiescence and increased proliferation. Indeed, TGFb1 injection into mice depleted of MK is able to restore HSC quiescence while conditional deletion of Tgfb1 in MK phenocopies their complete ablation in the niche[33].

Lastly, CXCR4, the receptor of CXCL12 chemokine, is present on HSCs and its deletion in adult mice results in HSCs reduction in number, increased sensitivity to myelotoxic stress such as 5-FU and disadvantage in competitive BMT[8]. Moreover, deletion of Cxcl12 from mesenchymal cells in the niche is associated with marked loss of HSCs quiescence and long-term reconstitution ability[34].

Additionally, HSCs maintain the quiescence state by regulating specific metabolic pathways. Quiescent HSCs are characterized by a low mitochondrial activity and rely on glycolysis[35]. HIF1α, the hypoxia inducible factor 1α, regulates HSCs in hypoxic conditions and promotes glycolysis[36,37]. Indeed, HIF1α deficient HSCs lose their quiescence. More recently, dietary metabolites have been described to influence HSCs behavior. In particular, vitamin A, and its metabolite retinoic acid, have been linked to HSCs quiescence maintenance, while vitamin C signaling regulates HSCs self-renewal capacity[38,39].

5.4 Acute Myeloid Leukemia

Acute Myeloid Leukemia (AML) is a heterogeneous group of diseases characterized by immature and hyper-proliferating leukemic cells (i.e. leukemic blasts) infiltrating the bone marrow, blood and other tissues, eventually leading to hematopoiesis failure. Infiltration of the bone marrow and low number of normal terminally differentiated blood cells (anemia, neutropenia/lymphocytopenia and thrombocytopenia) are at the basis of the clinical manifestations of the disease, including fatigue, infections and hemorrhages, and, if not treated, AML has a rapid fatal outcome. AML is the most common type of a hematological malignancy in adults, with incidence increasing with age. In the United States, AML incidence ranges between 3-5 cases per 100,000 population. Only in 2016, about 20,000 new cases have been diagnosed and more than 10,000 patients died[40]. Despite advances in therapies for some subtypes of AML and, in particular, for younger patients, prognosis for the majority of cases remain poor as only less than 50% of patients are definitely cured[41]. The standard therapeutic approach, which has not significantly changed in the last 30 years,

consists in 7 days of cytarabine infusion and 3 days of anthracycline (e.g. doxorubicine), followed by allogeneic bone marrow transplantation for eligible patients. The induction therapy allows to rapidly reduce the number of blasts but the majority of patients undergoes relapse after initial remission[42]. This year, for the first time since 2000, a new drug, Midostaurin (Novartis), has been approved by FDA for the treatment of AML harboring mutations in FLT3 gene[43].

The different AML types can be categorized using the French-American-British classification (FAB), established in 1976. It defines 8 classes of AML from M0 to M7 according to the morphological evaluation of the blasts (Table 1).

Type	Name	Cytogenetics
M0	Acute myeloblastic leukemia, undifferentiated	
M1	Acute myeloblastic leukemia, without maturation	
M2	Acute myeloblastic leukemia, with granulocytic maturation	t(8;21) RUNX1-RUNX1T1, t(6;9)
M3	Acute promyelocytic leukemia (APL)	t(15;17) PML-RARα, t(11;17)
M4	Acute myelomonoytic leukemia	inv(16), del(16q), t(16;16)
M5	Acute monocytic leukemia	t(9;11) MLLT3-KMT2A, t(11;19)
M6	Acute erythroid leukemia	
M7	Acute megakaryoblastic leukemia	t(1;22)

Table 1. FAB classification of AML.

More recently, the World Health Organization (WHO) introduced a new classification, which incorporated additional disease features (genetic, morphological, immunophenotypic and clinical) and defined 6 AML subtypes[44] (Table 2).

Types	Genetic abnormalities
AML with recurrent genetic abnormalities	AML with t(8:21)(q22;q22); RUNX1-RUNX1T1
	AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11
	APL with PML-RARα
	AML with t(9;11)(p21.3;q23.3); MLLT3-KMT2A
	ML with t(6;9)(p23;q34.1); DEK-NUP214
	AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM
	AML (megakaryoblastic) with t(1;22)(p13.3;q13.3); RBM15-MKL1
	AML with BCR-ABL1 (provisional entity)
	AML with mutated NPM1
	AML with biallelic mutations of CEBPA
	AML with mutated RUNX1 (provisional entity)
AML with myelodysplasia-related changes	
Therapy-related myeloid neoplasms	
AML (not otherwise specified)	AML with minimal differentiation
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukemia
	Acute monoblastic/monocytic leukemia
	Acute erythroid leukemia
	Pure erythroid leukemia
	Acute megakaryoblastic leukemia
	Acute basophilic leukemia
	Acute panmyelosis with myelofibrosis
Myeloid sarcoma	
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis
	ML associated with Down syndrome

Table 2. WHO classification of AML and related neoplasms.

Accurate classification is fundamental for the proper diagnosis and treatment of the disease. Indeed, in combination with other prognostic factors such as the age, different AML subtypes are associated with different clinical outcome. The broad spectrum of chromosomal aberrations and mutations described in different AML allows the stratification into favorable, intermediate and adverse risk groups. In particular, translocations t(8;21) and t(15;17), which cause the expression of the oncogenes AML1-ETO and PML-RARα respectively, are indicative of favorable prognosis. Complex karyotype and monosomy are instead associated with poor prognosis due to high treatment failure risk. Finally, AML

without cytogenetic abnormalities (NK-AML, normal-karyotype), accounting for about 50% of all AML cases, often falls into the intermediate risk group[45]. Among NK-AML, gene mutations are fundamental to refine risk stratification. Mutations in CEBPA or NPM1 without concomitant mutation in FLT3 are associated with good prognosis, while mutations in KMT2A, DNMT3A or FLT3 itself are associated with worse prognosis[46].

One of the major obstacles in leukemia eradication is given by the extensive genomic and biological inter- and intra-tumoral heterogeneity. Extensive whole-exome sequencing experiments showed that each AML presents about 10 mutations per exome, yet more than a hundred of different genes were found to be recurrently mutated in AML patients[14,47]. AML intra-tumoral complexity is a well described phenomenon characterized by the presence of multiple genetic subclones within the leukemic bulk[48,49], which contributes to therapy failure and relapse, the major cause of death in leukemia patients.

Mutations in the NPM1 gene are found in 55% of NK-AML (which account for 45% of all AML cases)[50]. In the WHO classification, NPM1-mutated AML are recognized as a distinct category while, in the FAB classification, NPM1 mutations are associated with most subtypes except the M3. NPM1 gene encodes for nucleophosmin, a nucleus-to-cytoplasm shuttling protein. Mutations in NPM1, mainly found in exon 12, cause a frame shift in the C-terminus, which generate a new nuclear export signal and is associated with protein delocalization in the cytoplasm (NPMc+). The most common mutation consists in a 4 base pair insertion (type A mutation) and is found in about 80% of cases[50]. Despite NPM1 mutations were discovered more than 10 years ago, the associated molecular mechanisms are still unclear. Available data suggest that NPMc+ posses both loss- and gain-of-functions[51]. In murine models, expression of NPMc+ alone is not sufficient to induce leukemia, suggesting that cooperative mutations are needed[52]. Indeed, internal tandem duplication (ITD) of the FLT3 gene, which encode for a receptor tyrosine kinase, is frequently associated with NPMc+ in AML, suggesting genetic cooperation between the two mutations[50]. Notably, the variant allele frequency (VAF) of NPMc+ in AML is around 0.5,

while that of FLT3-ITD is much lower, indicating that only NPMc+ is present in heterozigosity in all cells of the samples, thus suggesting that the NPMc+ occurs prior to FLT3-ITD during leukemogenesis[53]. NPMc+ is also found in association with DNMT3A, IDH1, IDH2 and TET2, while it is mutually exclusive with chromosomal translocations (including PML-RARα and MLL-AF9) and mutations in MLL1, RUNX1, CEBPA and TP53[54,55]. Comparative analysis of the gene expression profile of NPMc+ and NPM1-WT AML allowed the identification of a NPMc+ associated signature characterized by down-regulation of CD34 and up-regulation of HOX, MEIS1 and PBX3 genes, which are known master regulators of HSCs self-renewal[56,57].

Acute promyelocytic leukemia (APL) is a distinct subtype of AML, accounting for 5-15% of all adult AML and characterized by a block of myeloid differentiation at the promyelocyte stage. Almost all patients harbor a translocation between chromosome 15 and 17 that leads to the formation of the oncogenic fusion protein PML-RARα[58]. PML-RARα transgenic mice develop APL after a long latency (6-12 months) and at variable penetrance, suggesting that PML-RARα needs secondary events to induce a full-blown leukemia[59-61]. Indeed, several and different additional genetic alterations (e.g. FLT3, WT1, NRAS and KRAS) have been described both in human and in murine APL[47,62-65]. Beyond genetic complexity, combinatorial treatment with all-trans retinoic acid (ATRA) and arsenic trioxide (ATO) is one of the best example of targeted therapy in cancer.

Treatment with ATRA and ATO induces disease remission in the vast majority of APL patients, suggesting that PML-RARα expression is indispensable for the maintenance of the disease. Indeed, both agents induce PML-RARα degradation and restoration of the normal PML and RARα signaling, resulting in blasts maturation[66]. Notably, treatment with ATRA or ATO alone is invariably associated with disease relapse, frequently due to somatic mutations in PML-RARα, PML or RARα, while ATRA-ATO combination induces disease cure, suggesting that the two agents target PML-RARα in different cellular compartment of the leukemia[67].

Rearrangements of MLL gene (KMT2A or MLL1) occur in about 10% of human AML. The MLL gene is located on chromosome 11 and it encodes for a H3K4 methyl-transferase mainly acting as a transcriptional activator. More than 50 translocation partner genes have been described in AML patients and among them, MLL-AF4, MLL-AF9, MLL-ENL, MLL-AF10 and MLL-AF6 are the most frequent. MLL-fusion proteins retain the functional N-terminal of MLL, fused to the C-terminal of the translocation partner. All the MLL-fusion genes are in frame and result in the expression of chimeric transcription factors capable of efficient transformation of HSCs and progenitor cells[68,69]. Critical transcriptional targets of the MLL-fusion proteins are the HOX cluster, MEIS1 and PBX3 regulators of HSCs self-renewal[70,71], CDK6, JMJD1C, MEF2C, EYA1, MYB and MECOM[72-77].

5.5 Leukemia Stem Cells

Only a fraction of leukemic cells is able to sustain the tumor growth. These cells share morphological and biological features with normal HSCs, and are termed Leukemia Stem Cells (LSCs)[78]. LSCs are experimentally defined by their ability to: (1) generate a xenograft resembling the original tumor, (2) self-renew, assessed by serial passages *in vivo*, and (3) give rise to daughter cells, which can proliferate but are unable to establish or maintain leukemia *in vivo* (i.e. non-LSCs)[79,80].

Seminal discoveries in AML in the 90s allowed the more general conception of the Cancer Stem Cell (CSC) model, which is today used to explain relevant features of the intra-tumor biological heterogeneity of both leukemia and epithelial tumors[81]. According to the CSC model, tumors are described as aberrant tissues harboring their own hierarchical organization, similar to that of their normal counterparts, with CSCs placed at the apex of the hierarchy. In this view, CSCs are considered the only cell type which possesses self-renewal properties and the intrinsic ability to perform asymmetric cell divisions, thus being responsible oh both tumor growth and generation of intra-tumoral cell

heterogeneity[82,83](Figure 4, right panel). Accordingly, CSCs are regarded as responsible for tumor initiation, maintenance, propagation and disease relapse after initial remission.

Although the CSC model is widely accepted within the scientific community, most currently used therapeutic approaches target highly proliferating cells and find their rationale on a different model of tumorigenesis, the so-called "stochastic model" of tumor growth, which implies that all cancer cells have equal self-renewal ability and tumorigenic potential (Figure 4, left panel). However, in both models, intra-tumoral heterogeneity is generated by the acquisition of novel somatic mutations that establish subclonal populations within the tumor. Hierarchically-organized tumors show an additional degree of intra-tumoral heterogeneity due to phenotypic transition between CSCs and non-CSCs. This plasticity has been well described in solid tumors and is highly dependent on microenvironmental stimuli[84].

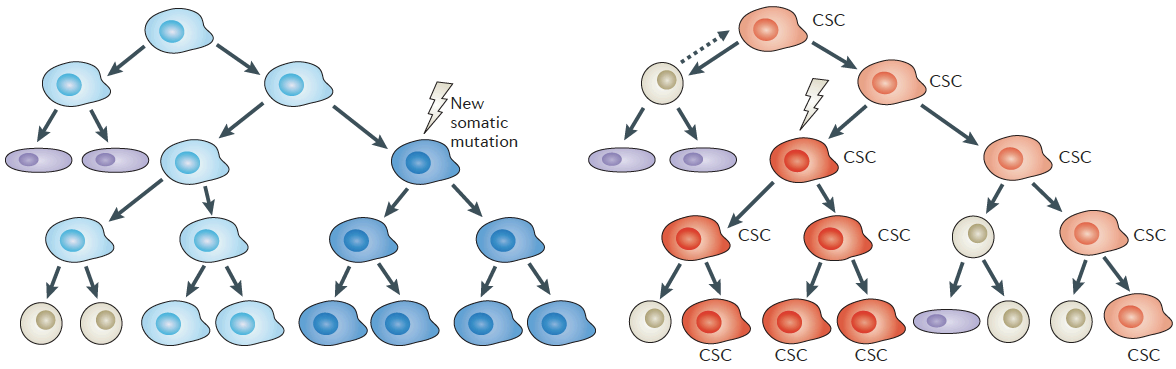

Figure 4. The two major models to describe tumor growth: stochastic vs hierarchical model. In the left panel is depicted the stochastic model of tumor organization in which each tumor cell has the same capability to growth. On the right, the hierarchical cancer organization is depicted: in this model, CSCs are the only cells able to self-renew and sustain the tumor growth. Nevertheless, a certain grade of plasticity can occur (dashed arrow). In both models, the acquisition of novel somatic mutations account for the genetic heterogeneity within the tumor (adapted from Beck and Blanpain, Nat Rev Cancer 2013).

Traditional chemo- and radiotherapies, which are able to target only actively cycling cells, have limited or no effects on LSCs. This intrinsic resistance can be due to the prevailing quiescent state of LSCs, and other mechanisms, such as the expression of drug efflux pumps, anti-apoptotic mediators and peculiar DNA damage repair response[85]. Thus, development of

novel therapeutic strategies aimed at targeting quiescent LSCs may impact on leukemia eradication.

Pioneering studies in John Dick's lab demonstrated the existence of LSCs in human AML by the prospective isolation and xenotransplantation into immunocompromised mice of different leukemic populations based on CD34 and CD38 antigen expression. The first studies showed that only the $CD34^{+}CD38^{-}$ compartment retains tumorigenic ability *in vivo*[78,86]. Nevertheless, there is great heterogeneity in the expression level of the two surface markers among different AML samples. Indeed, further studies identified LSCs at varying frequencies in different cell populations and, for instance, in NPM-mutated AML, LSC activity is commonly found in the $CD34^{-}$ compartment[87]. More recently, Dick and colleagues, combining deep sequencing and xenotransplantation, demonstrated that LSCs can either express HSCs/progenitors markers or be enriched in immunophenotypically committed cells[88]. Notably, regardless their immunophenotype, functionally validated LSCs are characterized by the same transcriptional program strongly related to stemness maintenance.

5.6 The AML cell of origin

Cancer development was initially described as the accumulation of sequential somatic mutations in a cell clone. If these mutations result in a growth advantage, the clone becomes the predominant cell population. This simplistic model has been intensively revised over years[82]. For many cancer types, including AML, the target cell of the transformation events is still not clear. The findings that AML blasts can co-expresses both myeloid and lymphoid markers[89] and the observation that the only cells capable of transplanting leukemia in xenograft models have a $CD34^{+}CD38^{-}$ phenotype[78,86], similar to that of normal human HSCs, lead to hypothesize that the cell of origin resides in this compartment. However, LSCs do not necessarily originate from the malignant transformation of normal HSCs[90]. Indeed, xenotransplantation experiments in later studies led to the detection of LSCs in AML

populations with variable phenotypes downstream of HSCs. In the case of APL, the PML-RARα fusion gene is absent in HSCs compartment while is found in more committed progenitors (i.e. promyelocytes) which are able to engraft in humanized ossicle xenograft mice[91,92]. Moreover, studies on murine models of spontaneous leukemogenesis have shown that oncogenic mutations can also confer self-renewal properties to cells with limited regenerative potential[93-95].

A recent study showed that the AML chromatin landscape pattern is able to reflect the transformed cell of origin, offering a new prognostic tool for patients[96]. Indeed, it is worth thinking that the biological properties of LSCs, then, may vary with regard to the immunophenotype, transcriptome, genetics, cell cycle status and multilineage potential of the cell of origin, as shown for other tumors, such as lymphoma[97,98].

5.7 Clinical implication of LSCs

Beside the cellular origin, the existence of LSCs has important implications for treatment, since these cells seem to be intrinsically resistant to traditional therapies, thus preventing disease eradication. Moreover, especially in AML, high LSCs frequency and better engraftment ability in immunodeficient mice are associated with more aggressive disease[99,100]. As well, the degree of overlap between LSCs and HSCs gene expression profiles correlates with poor clinical outcome[101]. In the last years, many groups tried to identify expression signatures able to classify AML according to their pathogenesis and prognosis[102]. In particular, Ng and colleagues have recently published a biomarker list of 17 genes related to stemness (the LSC17 score), differentially expressed between LSCs and non-LSCs, across 78 AML patients. The LSC17 score proved to be able to identify patients with poor outcomes in five independent patient cohorts belonging to different AML subtypes[103].

The ability of LSCs to survive chemotherapy and radiotherapy has been associated with functional properties and molecular mechanisms shared with normal HSCs, such as

quiescence[104,105], apoptosis evasion[106], enhanced DNA damage response[107] and lower concentration of reactive oxygen species (ROS)[108]. Quiescence, in particular has been linked to therapy resistance since the traditional therapies are usually specific for actively proliferating cells. Notably, cell cycle-restriction appears to be fundamental for leukemia development[107]. High-throughput analysis of neutral somatic mutations (i.e. microsatellite instability) at single cell level allowed the reconstruction of cell lineage trees in two AML patients, from diagnosis to relapse. These cell lineage trees showed that mutations present in minor slowly-cycling sub-clones were maintained after treatment and, suggesting that AML relapse can derive from rarely-dividing cells capable of surviving chemotherapy. Interestingly, this phenomenon was not observed in other hematological malignancies such as Acute Lymphoid Leukemia (ALL)[109]. Quiescent cells are characterized by low oxidative phosphorylation and LSCs have been shown to have lower ROS levels and oxygen consumption rates. At the same time, LSCs are highly dependent on oxidative phosphorylation, becoming more sensitive to Bcl-2 inhibitors[106,108]. Bcl-2 inhibitors are currently in clinical trials for hematological malignancies including AML.

Despite the many common features between LSCs and HSCs, it may be possible to target gene networks that are specific for the survival of LSCs. This is the case of miR-126, which was found to promote quiescence and increased self-renewal in primary LSCs, with an opposite function in HSCs[110]. The design of novel therapeutic strategies for AML treatment should therefore include the specific targeting of LSCs, while sparing normal HSCs.

Since markers commonly used to study LSCs in human AML are based on the immunophenotype of normal hematopoietic cells, a lot of effort has been invested into identifying specific molecules that discriminate LSCs from both normal HSCs and the bulk of leukemic blasts. Many surface markers have been proposed to be specific for LSCs in different AML subtypes but they often allow only the partial identification of the LSCs pool. The CD33 antigen, for instance, is highly expressed in AML blasts and in normal human myeloid committed progenitors, while it is not expressed on normal HSCs[111]. Therefore,

CD33 is a potential target for new therapies and specific antibodies are currently in clinical trials. In particular, GO (Gemtuzumab Ozogamicin), an anti-CD33 monoclonal antibody conjugated to the toxin calicheamicin, gave significantly increased remission rates in advanced and relapsed APL[112,113]. Since GO is highly active against APL, as compared to other AMLs, APL stem cells may reside in the $CD33^+$ compartment and posses a more mature phenotype[114]. However, in 2010 Pfizer voluntarily withdrew GO due to safety concerns and the drug is currently available only for palliative use. SGN-CD33A, a novel anti-CD33 antibody conjugated to two molecules of a pyrrolobenzodiazepine dimer drug, is currently in clinical trials for AML and, up to now, it has demonstrated anti-leukemic activity with 47% of complete remission and rapid blast clearance with relatively modest toxicity[115].

5.8 Targeting quiescence in LSCs

Inducing LSCs to proliferate might appear as a counterintuitive anti-leukemic strategy, because of the risk to accelerate disease progression or new mutations accumulation. However, experimental treatments with agents that induce cell cycle entry, in combination with standard chemotherapy, did not show evidences of accelerated disease progression[116,117]. Moreover, since quiescence regulation is similar in normal HSCs, this awaking strategy could affect normal hematopoiesis. Yet, clinical studies in hematological malignancies have not reported excessive toxicity, suggesting that a therapeutic window exists to distinguish normal and leukemic stem cells, at least for some neoplasms[118]. Thus, elucidation of the molecular mechanisms underlying maintenance of quiescence in LSCs might be critical for the identification of novel drug targets. Many approaches directed to target quiescence in LSCs have been described in Chronic Myeloid Leukemia (CML) in which LSCs phenotype is well characterized as $CD34^+CD38^-$.

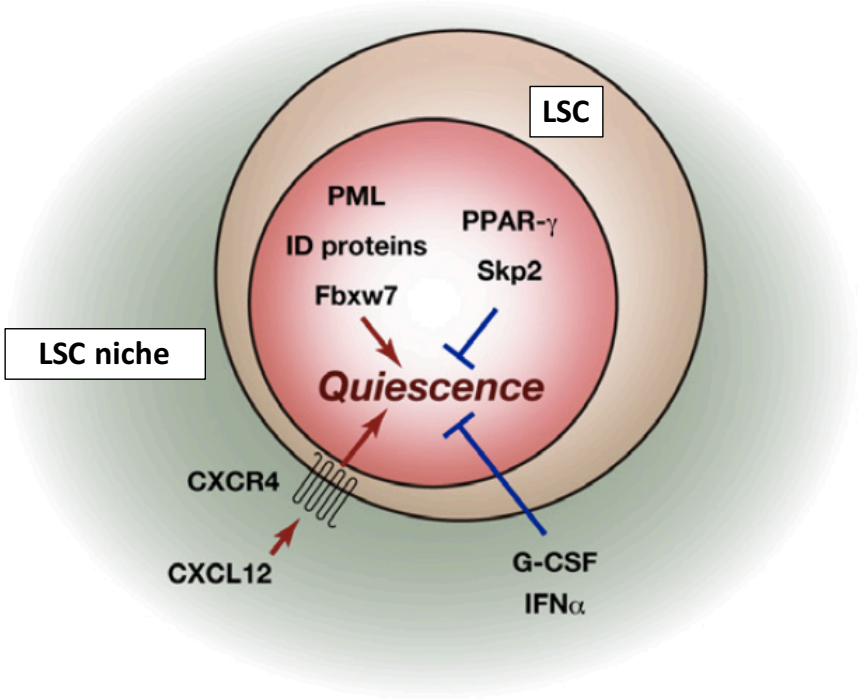

Figure 5. LSCs quiescence is regulated by intrinsic and extrinsic factors (adapted from Takeishi and Nakayama, Cancer Science 2016).

As already mentioned, self-renewal and quiescence of LSCs and HSCs is regulated by cell-autonomous mechanisms and microenvironment signals (Figure 5). Among the known cell autonomous regulators of LSCs quiescence is the promyelocytic leukemia protein (PML), first identified as a component of the PML–RARα fusion protein in patients with APL (reviewed in Mazza and Pelicci, 2013)[119]. It has been demonstrated that ATO triggers proteasome-dependent PML-RARα degradation and reactivation of PML signaling, leading to LSCs eradication[66,120]. ATO is also effective in the treatment of CML, where like PML inactivation, it reduces the proportion of quiescent LSCs[121]. The mechanisms by which PML regulates quiescence may involve metabolic processes, in particular fatty acid oxidation[121,122], and cell cycle control via regulation of p21, cyclins, and c-Myc[119].

As above-mentioned, the CKI p27 is necessary for HSCs quiescence maintenance. At the same time, p27 deletion in MLL-AF9 blasts reduces LSCs frequency and, inducing proliferation, it causes loss of drug resistance[123]. Another transcriptional regulator involved in quiescence regulation is the peroxisome proliferator-activated receptor-γ (PPARγ), a

nuclear receptor that governs fatty acid storage and glucose metabolism. The PPARγ agonist pioglitazone induces cell cycle entry of LSCs isolated from CML patients, an effect that was associated with down-regulation of STAT5 and CITED2[124]. The combination of pioglitazone and Imatinib (Novartis) reduces LSCs viability *in vitro* and a phase II clinical trial in CML is currently ongoing[125].

Among enzymatic functions, the F-box protein FBXW7, the substrate recognition subunit of the SKP1-CUL1-FBXW7 ubiquitin-ligase complex, has been described to regulate LSCs quiescence. The SKP1-CUL1-FBXW7 ubiquitin-ligase complex is responsible for the ubiquitination and subsequent degradation of many cellular substrates including c-Myc. In a mouse model of CML, Fbxw7 genetic ablation induces LSCs to enter cell cycle due to increased expression of c-Myc[126]. Moreover, Fbxw7-deficient LSCs become sensitive to Ara-C and Imatinib treatment leading to CML eradication. Another pathway tightly regulated in LSCs biology is mTOR. mTORC1 inhibition, as induced by rapamycin treatment, is sufficient to decrease LSCs number in a Pten-depleted AML model. Moreover, rapamycin treatment is able to restore HSCs function in the same context[127]. Though the mTORC2 complex is not extensively characterized, it has been recently demonstrated that depletion of one of its component, namely Rictor, leads to hyperactivation of mTORC1 and MLL-AF9 LSCs exhaustion after serial BMT, due to quiescence exit[128].

Regarding the extrinsic factors regulating LSCs quiescence, treatment with the granulocyte colony-stimulating factor (G-CSF), a cytokine that promotes maturation, differentiation and proliferation of myeloid cells, has been shown to induce LSCs cell cycle entry. G-CSF treatment in combination with Ara-C is able to reduce LSCs frequency improving survival in AML xenograft models[105]. Chemokine ligand 12 (CXCL12) is a chemoattractive cytokine with the function of retaining HSCs in the BM and maintain their quiescence, interacting with its receptor CXCR4 present on HSCs surface[8]. Targeting this axis with CXCR4 inhibitors is sufficient to sensitize AML and CML cells to chemotherapy[129-131]. Other microenvironment signals important for LSCs are the interferons (IFN), cytokines mainly

produced by cells of the immune system in response to infections and also tumor cells. In particular, IFNα treatment in CML patients is able to potentiates Imatinib therapy, though the underlying molecular mechanisms are not clear[132]. The effect of INFα treatment is instead well describe in normal HSCs where it induces cell cycle entry associated with down-regulation of p27, p57, Foxo3a and TGFβ, all genes that support HSCs quiescence[133-135]. Lastly, angiopoietin-1 (Ang1) is a cytokine known to promote HSCs quiescence via the interaction with its tyrosine-kinase receptor Tie2[136]. Ang1 signaling pathway has been recently described as a mechanism of quiescence maintenance in Evi1 highly expressing AML, protecting leukemic cells from anti-cancer drugs[137].

Growing evidences showed that also noncoding RNAs play important role in stem cell maintenance. microRNAs are small non-coding RNA molecules regulating gene expression at post-transcriptional level. For instance, miR-99 family consists of a group of microRNA highly expressed in HSCs and in acute myeloid LSCs. miR-99 promotes in vitro clonogenic capacity both in HSCs and LSCs, while its down-regulation reduces HSCs long-term reconstitution ability inducing cell cycle entry, and it reduces LSCs frequency in a MLL-AF9 driven AML mouse model by inducing blasts differentiation[138].

5.9 Pre-leukemic HSCs

With the advent of massive sequencing technologies, lots of information have been produced on the mutational status of both normal and malignant hematopoiesis. High-throughput sequencing of AML samples showed that mutations such as DNMT3A, IDH2 or TET2, found at diagnosis are often present in patients at complete morphological remission[139-142]. Moreover, investigation of residual normal HSCs in AML samples revealed that many patients harbor a population of HSCs bearing some of the mutations present in the corresponding leukemic cells. This so-called pre-leukemic stem cells are present both at diagnosis and relapse and their proportion in the stem cell compartment correlates with poor prognosis[139,140]. This finding is clinically relevant because relapses could arise both through

selection of pre-existing and resistant leukemic clones or through the evolution of pre-leukemic HSCs through the acquisition of additional mutations. In the last scenario, pre-leukemic HSCs will act as a reservoir for leukemic progression and they should be carefully evaluated during minimal residual disease monitoring.

Pre-leukemic mutations have been found in healthy individuals as well. On average, clonal hematopoiesis, as identified by a somatic mutation, is observed in 10% of healthy individuals older than 65 years old[143]. A more recent study, able to detect very low frequent mutations (0.0003 VAF) in the peripheral blood of 50-60 years-old healthy individuals, indicates that the 95% of them is characterized by clonal hematopoiesis, frequently harboring mutations in DNMT3A and TET2[144]. Moreover, mutations are present in multiple hematopoietic lineages, suggesting a long lived progenitor or a HSC as the cell of origin. Clonal hematopoiesis is associated with higher risk of hematological cancer development[143,145]. In contrast to leukemic cells, pre-leukemic cells are not characterized by aberrant self-renewal or differentiation potential and they require additional mutation to establish a frank leukemia.

5.10 Assays to study HSCs and LSCs

5.10.1 Transplantation into immunodeficient mice

Traditionally, HSCs functionality is assayed by transplantation of defined cell populations into irradiated recipient mice. In this setting, it is possible to dissect all the defining functions of stem cells, namely engraftment, self-renewal, bone marrow reconstitution and multi-lineage differentiation. Transplantation experiments, coupled with cell purification based on surface marker expression, have been instrumental for the characterization of HSCs and progenitors. As for normal HSCs, the study of LSCs depends and is influenced by the available experimental assays. The gold standard for human LSCs identification, in fact, remains transplantation into immunocompromised mice. However, xenotransplantation assay can introduce bias due to interspecies difference in cytokines, microenvironment and

immune system interactions[146]. Various xenograft models are suitable for the engraftment of self-renewing leukemia cells. Cells that are able to propagate leukemia in this context are functionally defined as Leukemia Initiating Cells (LICs), and they correspond to the experimental definition of LSCs[147]. Although, however, *in vivo* transplantation provided the first evidence for the existence of LSCs, this assay has some technical limitations. In particular, leukemogenic potential can not be assessed at single cell level and LSCs potential can be just segregated to a previously defined cell compartment. In addition, cells able to initiate the disease in mice could not be the same cells responsible for leukemia growth and maintenance in the patients. This problem is highlighted by the fact that the same leukemic population shows variable engraftment potential in different mouse strains: the more immunodeficient the mice are, the more they are permissive to human leukemic cells growth. Indeed, the transition from athymic mice, which lacked T-lymphocytes, to the SCID (severely combined immunodeficient) strains, which also lacke B-lymphocytes, allowed the first AML engraftment, and the introduction of more immunodeficient models (e.g. NOD or NSG mice, which also have defective innate immune system), as well as strains engineered to express human cytokines, led to significantly higher sensitivity in xenotransplantation assays[148,149].

5.10.2 Clonal tracking

It has been recently shown that unperturbed native hematopoiesis differs from post-transplantation hematopoiesis[150,151]. Single HSC reconstitution capacity, including self-renewal and differentiation, has been studied by lineage-tracing and clonal tracking experiments in mice. *In vivo* genetic lineage-tracing experiments are based on the expression of an inducible recombinase under the control of a stem cell-restricted gene (e.g. TIE2), which activates a stable reporter gene (e.g. YFP) in HSCs selectively, thus allowing their identification and tracing[150]. Lineage tracking experiments have been also conducted using inducible transposons, which allow *in situ* HSCs tagging without manipulation (such as

isolation, *ex vivo* transduction and transplantation in irradiated mice)[151]. Results obtained with these more physiological settings are consistent with model in which adult hematopoiesis is maintained by thousands of clones derived from long-lived multipotent progenitors, rather than from HSCs.

On the other hand, clonal tracking experiments combine viral cellular barcoding and NGS allowing an accurate investigation of HSCs clonal expansion in vivo[152,153]. These experiments, based on barcoded HSCs transplantation in irradiated recipient mice, showed that HSCs contribute at different extent to hematopoiesis and more than one HSCs population exists with distinct lineage bias[153]. HSCs heterogeneity is well characterized in terms of lineage bias and proliferative kinetics, especially during aging when the HSCs pool increases in size and accumulates preferentially myeloid-biased stem cells with reduced self-renewal potential[154].

As well as for normal hematopoiesis, clonal tracking strategy can be used to map leukemic cell kinetics *in vivo*. Lentiviral barcoding of leukemic blasts and transplantation in recipient mice allow the quantification of LIC frequency and the evaluation of leukemia clonal composition[155]. Beside LIC frequency estimation, clonal tracking experiments allow to study the heterogeneous proliferative capacity of single LICs as well as drug resistance mechanisms, as already described in solid tumors[156]. Moreover, it has an increased sensitivity because it allows single cell identification by detection of individual barcodes.

5.11 RNAi screenings

RNA interference (RNAi) was originally discovered in *C. elegans* in 1998[157] and it rapidly became a broadly used tool for loss-of-function studies, both *in vitro* and *in vivo*[158,159]. RNAi screens exploit a physiological mechanism that represses gene expression, primarily by causing the degradation of mRNA transcripts, allowing to interrogate the function of several genes simultaneously[160]. For a successful screen several factors have to be carefully evaluated such as library size, coverage and target identification and validation approaches.

RNAi library can be organized either in array format, in which each RNAi molecule is individually tested, or in a pool of shRNA (short hairpin RNA) molecules simultaneously assayed. The combination of this approach with retro- or lentiviral vectors allowed the possibility to perform screening in primary mammary cells, including stem cells. Indeed, several groups have performed RNAi screen in normal HSCs leading to new insights on self-renewal regulation[161-164]. Other groups have employed this tool to identify hematological disease-relevant genes, especially tumor-suppressors as in the case of Rad17 in Eμ-Myc lymphoma model[165]. Finally, several RNAi screens have been performed to identify potential therapeutic target in hematological malignancies. One of the best example is the work of Zuber and colleagues, in which they used a murine AML model, characterized by the expression of MLL-AF9 and $Nras^{G12D}$ oncogenes, to perform an *in vitro* RNAi screen targeting around 250 chromatin regulators[166]. The study lead to the identification of Brd4, a bromodomain protein which binds to acetylated histones regulating transcription, as necessary for leukemia maintenance, sustaining Myc expression. They also demonstrated that the Brd4 small molecule inhibitor JQ1 has an anti-leukemic effect *in vitro* and *in vivo*, establishing the basis for its clinical usage. The MLL-AF9 leukemia mouse model has been successfully used to identify other potential therapeutic targets via *in vivo* screens. In particular, Ebert's group identified Itgb3 and Csnk1a1 as conceptually ideal targets because dispensable for normal HSCs but required for AML *in vivo*[167,168]. Beyond studies on normal and malignant hematopoietic systems, the use of similar screening approaches has been extensively used in various solid tumors, leading to the identification of genes involved in the tumorigenic process[169,170].

6. Materials and Methods

6.1 AML mouse models

Mice were housed in a pathogen-free animal facility at European Institute of Oncology. The procedures related to animal use have been communicated and have been approved by the Italian Ministry of Health.

NPMc+ transgenic mouse model express the cDNA of the most frequent NPM mutation (mutA, found in about 80% of the patients)[50] into the Hprt locus and under the control of the pCAG promoter[171]. These mice express the oncogene upon CRE-mediated excision of the STOP cassette cloned between the pCAG promoter and the NPMc+ cDNA. Conditional NPMc+ mice were crossed with conditional Rosa26-EYFP mice[172] to obtain NPMc+$^{fl/-}$/YFP$^{fl/-}$ mice. In order to express the oncogene, bone marrow mononucleated cells (BM MNCs), isolated from NPMc+$^{fl/-}$/YFP$^{fl/-}$ or YFP$^{fl/-}$ mice, were treated *ex vivo* with TAT-CRE, a recombinant version of CRE recombinase fused with the HIV protein TAT. In details, total BM cells were isolated by grinding bones from posterior limbs and sternum. Total BM cells were layered onto density gradient (Histopaque® 1083, Sigma-Aldrich) and centrifuge at 1500 rpm, 45 min at 4°C. During centrifugation, BM-MNCs remain at the PBS-Histopaque® interface and can be collected. BM-MNCs were re-suspended 5x10^6/ml in serum-free media (Hyclone, USA) and incubated for 45 min at 37°C with 100 μg/ml of TAT-CRE. Transduction was stopped diluting samples with 10 volumes of BM-MNCs medium (IMDM (Gibco/Invitrogen, Carlsbad, CA), 12.5% heat inactivated fetal bovine serum (FBS), 12.5% Horse serum, 1% L-glutamine, 100 ng/ml SCF, 20 ng/ml IL3, and 20 ng/ml IL6 (PeproTech), 0.1% β-mercaptoethanol, and Hydrocortisone 10 ng/ml), cells were spun down, re-suspended in BM-MNCs medium and cultured at a density of 2x10^6 cells/ml for 48 hours. YFP+ cells were FACS sorted and intravenously injected into lethally irradiated (7.5 Gy) C57 BL/6 Ly5.1 recipient mice. Mice transplanted with NPMc+ expressing BM-

MNCs develop leukemia after a long latency (median 564 days) and with low penetrance (33%)[171].

For the PML-RARα leukemia model, we backcrossed the mCG-PR knock-in (KI) mice, provided by Tim Ley[61], in the C57 BL/6J strain. These mice expressed the PML-RARα oncogene in stem and early myeloid cells under the control of the murine cathepsin G promoter. Mice develop spontaneously a disease resembling the human APL, with 70% penetrance at 6-16 months of age (median 10 months). Murine APL can be serially injected in C57 BL/6 Ly5.1 recipient mice.

MLL-AF9 leukemia was generously provided by Dr Chi Wai So. Leukemia was obtained by MSCV-MLL-AF9-puro retroviral transduction and transformation assay (RTTA) of cKit+ cells isolated from BM-MNCs, as previously described[173]. In brief, cKit+ BM cells were FACS isolated from wild-type C57 BL/6 Ly5.1 mice, transduced by spinoculation with the retroviral vector expressing the MLL-AF9 oncogene, and serially re-plated in methylcellulose medium prior to injection into sub-lethally irradiated (5 Gy) C57 BL/6J recipient mice. Mice die of leukemia approximately one month after injection.

Murine AML were characterized based on blasts immunophenotype and oncogene expression. Mice were sacrificed at the first signs of pain and leukemic blasts were isolated from spleen and bone marrow after centrifugation through a Histopaque® gradient, stained with fluorochrome-conjugated antibodies against myeloid and lymphoid markers and analyzed by FACS. NPMc+ expression was assessed by western blot and immunofluorescence analysis as described in Mallardo et al., 2013. Quantitative polymerase chain reactions (qPCR), from blasts isolated from leukemic spleen, were performed according to standard techniques using primers specific for PML-RARα cDNA (Table 3). Blood smears were stained with May-Grünwald-Giemsa while bone marrow/spleen paraffin embedded samples were stained with haematoxylin-eosin, according to standard protocols, and used for AML diagnosis.

Primer name	Sequence 5'→3'
PR forward	AGGGACCCTATTGACGTTGAC
PR reverse	ACAGACAAAGCAAGGCTTGTAG
mTBP forward	TAATCCCAAGCGATTTGCTG
mTBP reverse	CAGTTGTCCGTGGCTCTCTT

Table 3. Primers used to check oncogene expression via qPCR.

6.2 Characterization of pre-leukemic transcriptional program

6.2.1 LT-HSCs RNA purification

NPMc+ pre-leukemic and wild-type LT-HSCs were purified from mice injected with NPMc+$^{fl/-}$/YFP$^{fl/-}$ or YFP$^{fl/-}$ recombined BM-MNCs, 4 months after transplantation. PML-RARα pre-leukemic and wild-type LT-HSCs were purified from the bone marrow of 10-12 weeks old mice. BM-MNCs were isolated as previously described and stained with fluorochrome-conjugated antibodies against Sca1, c-Kit, Flk-2, CD34 and the lineage markers antibodies cocktail (Lin: B220, Ter-119, CD3, Mac1, Gr1, CD4 and CD8), all provided by eBioscience. LT-HSCs were then sorted via FACS as Lin-, Sca1+, cKit+, CD34- and Flk-2- and RNA was immediately extracted with the PicoPure™ RNA Isolation Kit (ThermoFisher), according to manufacturer protocol. The integrity of RNA was analyzed by Bioanalyzer (Agilent).

6.2.2 Microarray analysis

RNA purified from WT and pre-leukemic LT-HSCs was analyzed on Affymetrix GeneChip® Mouse Gene 2.0 ST array. Double stranded cDNA synthesis was performed with Nugen® Pico WTA Systems V2 (NuGEN Technologies, Inc.). Microarray raw data were normalized by RMA using Partek Genomic Suite 6.6. A set of differentially expressed genes between pre-leukemic and WT samples was identified by applying a threshold of FC>|1.5| and FDR<0.1. Gene set enrichment analysis (GSEA, v2.14 software Broad Institute)[174] was used to investigate whether a gene set was significantly over-represented in

the transcriptome of the pre-leukemic cells. The curated gene sets collection (c2.all.v5.1.symbols.gmt) was downloaded from MSigDB. Moreover, three custom gene sets related to quiescence have been tested for enrichment. A gene set was identified as significantly enriched when associated with FDR q-val<0.1. A detailed description of the GSEA methodology and interpretation is provided at http://www.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html.

For microarrays validations qPCR reactions, 10 ng of total RNA purified from WT or pre-leukemic LT-HSCs was reverse-transcribed using Super Script IV (Invitrogen). cDNA was pre-amplified with TaqMan® PreAmp Master Mix Kit and then amplified using pre-designed TaqMan Gene expression assays (Thermo Fisher Scientific). All experiments were run in triplicate and results normalized to Gapdh and Gusb mRNA expression level.

6.3 shRNA screening and clonal tracking

6.3.1 Lentiviral libraries

The custom shRNA library used for the in vivo screenings was purchased from Cellecta Inc. In the pRSI16 lentiviral vector, shRNAs are cloned under the control of the U6 promoter and univocally associated with a 22-nucleotides barcode (BC). The vector also encodes for the puromycine resistance (PuroR) and the Red Fluorescence Protein (TagRFP) reporter gene, under the control of the UbiC promoter (Figure 6). The shRNA library (M1) contains 1,000 different shRNAs targeting 96 genes (about 10 shRNAs/gene), including 4 controls (Table 4). The control library is composed of 1,200 different 22-nucleotides BCs cloned in the same vector.

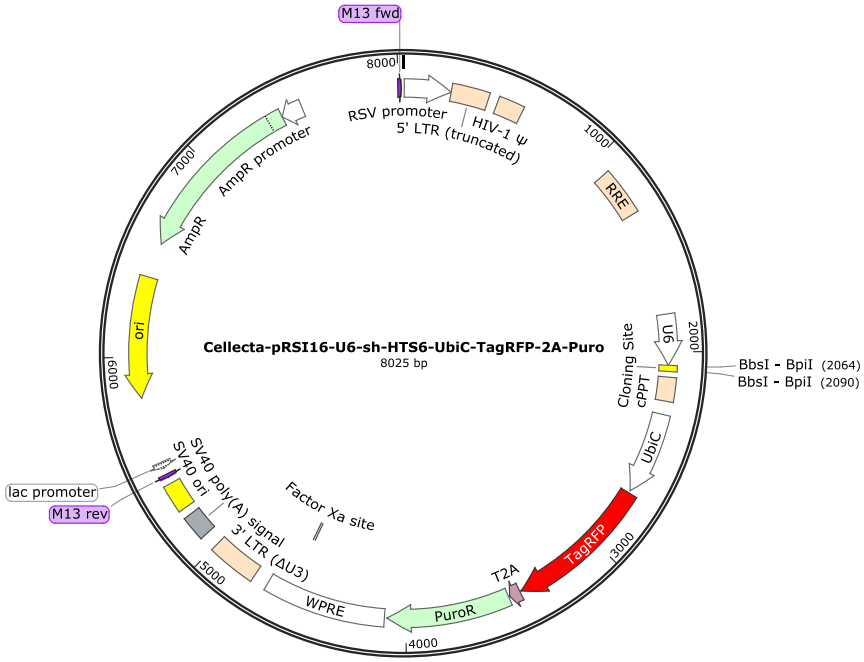

Figure 6. Map of Cellecta pRSI16 vector used for shRNA screenings.

Genes targeted in M1 shRNA library			
Abcb1b	Egr1	Jun	Rnf166
Acox1	Esr1	Kit	Rrad
Adcy9	Etnk1	Klf6	Slc44a2
Adgrg1	Fgl2	Kmt2a	Smad3
Adipor2	Flii	Map2k3	Smarca2
Adssl1	Fos	Mcfd2	Socs2
Angpt1	Fosb	Mecom	Stat1
Anxa4	Gabarapl1	Mpl	Stat3
Atxn1l	Gata2	Mprip	Stat5a
Bach1	Gbp2	Muc13	Stx3
Bgn	Gfi1	Mycbp2	Sytl4
Brd4	Hdac5	Naga	Tgfb1
Cables1	H2-Q10	Ndfip1	Tgfbr2
Cbfa2t3	Hoxa5	Ndrg1	Tie1
Cdkn1a	Ier3	Nfatc1	Tmbim1
Chrnb1	Igf1r	Npc1	Txnrd1
Cited2	Igtp	Nrip1	Ube2e1
Creg1	Il18r1	Pbx1	Vamp2
Csgalnact1	Irf1	Pdzk1ip1	Zbtb20
Ctla2a	Irf3	Pim1	Zfp68
Cyp4v2	Irf6	Pkd1	Psma1
Dcaf11	Itih5	Prdm16	Rpl30
Dock9	Itm2a	Rest	Polr2b
Dusp6	Jmjd1c	Rhob	LUC

Table 4. Genes targeted in M1 shRNA library.

For the clonal tracking experiments, we purchased a lentiviral library containing two consecutive 18-nucleotides barcodes for a total complexity of about $30x10^6$ different barcodes combinations (CellTracker™ Lentiviral Barcode Library, Cellecta Inc). The lentiviral backbone encodes for the PuroR and the TagRFP genes under the control of the UbiC promoter (Figure 7).

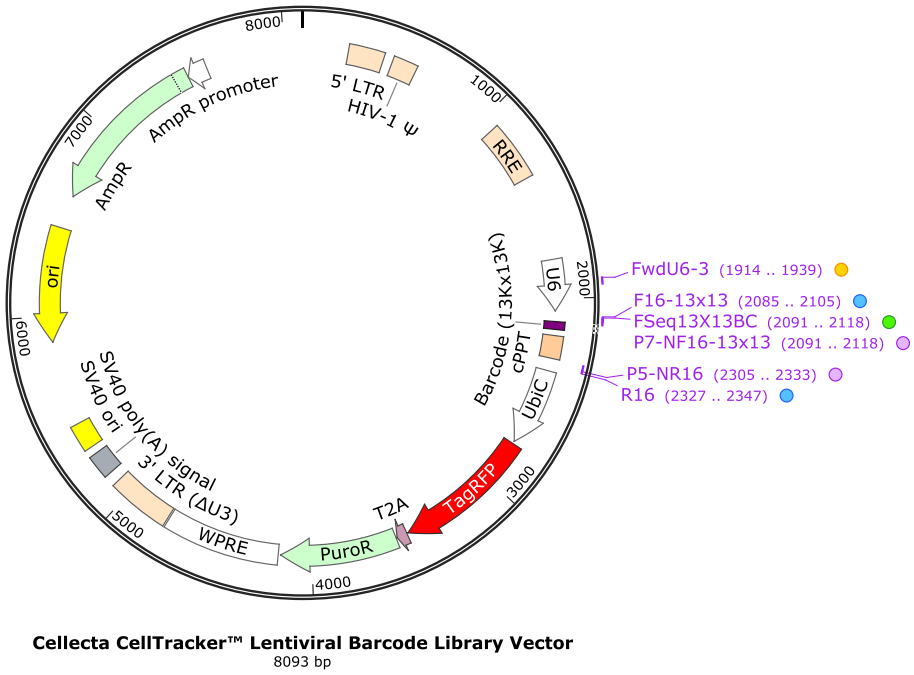

Figure 7. Map of Cellecta CellTracker™ Lentiviral Barcode vector used for clonal tracking experiments.

6.3.2 Viral production and titration

For lentiviral production, the 2nd generation packaging vectors pMD2.G and pCMVdR8.2 were used. 293T cells were transfected with Lipofectamine® and PLUS® reagents (ThermoFisher), according to manufacturer protocol. In brief, 293T were plated in 15 cm dishes in DMEM (Lonza), 2 mM L-glutamine, 100 U/ml penicillin/streptomycin, 10% Fetal Bovine Serum (FBS) and transfected when 70% confluence was reached. We mixed the library DNA and the packaging vectors DNA with the PLUS® reagent in 1.2 ml of serum-free Opti-MEM™: 6 µg library DNA, 7 µg pMD2.G, 24 µg pCMVdR8.2 and 60 µl PLUS®,

per dish. In an other tube we mixed 90 µl Lipofectamine® with 1.2 ml of serum-free OptiMEM™, per dish. Both solutions were incubated separately for 15 min at room temperature (RT) and then mixed together. After another 15 min incubation at RT, the DNA-Lipofectamine® mixture was added to the 293T dish. We usually transfect 10 293T 15 cm dishes, as suggested by Cellecta guidelines, in order to cover library complexity and obtain a high viral titer. Medium containing lentiviral particles was collected 24, 48 and 72 hours after transfection, replacing with fresh medium. Lentiviral supernatant was filtered through a 0.2 µm Nalgene™ Rapid-Flow™ Sterile Disposable Filter Unit (ThermoFisher), concentrated by ultracentrifugation at 24000 rpm, 2h at 4°C with a Optima L-90K ultracentrifuge (Beckman Coulter) and stored at -80°C. Lentiviral titer was evaluated by transdcucing $0.1x10^6$ 293T/well in 6 wells plate with serial dilution of the viral stock. 72h after infection, 293T were harvest and analyzed by FACS for the percentage of TagRFP+ cells. Viral titer, measured as transducing units per ml (TU/ml), was calculated with the following formula:

$$titer\left(\frac{TU}{ml}\right) = \frac{number\ of\ target\ cells\ (0.1x10^6) * (\%\ \text{TagRFP positive cells})}{volume\ of\ virus\ added\ (ml)}$$

6.3.3 Infection of AML blasts

Leukemic blasts are maintained in Iscove's Modified Dulbecco's Medium (IMDM, Lonza), 2 mM L-glutamine, 15% Fetal Calf Serum (FCS), 15% 5637-conditioned medium and 25% WEHI-3B-conditioned medium. Blasts are infected with lentiviral particles by spinoculation with MOI=3 at 2300 rpm, for 90 min at RT. For shRNA screenings, $24\text{-}50x10^6$ blasts/well were plated in 6 wells RetroNectin®-coated plates in presence of 4 mg/ml of polybrene. For clonal tracking experiments, $1x10^6$ blasts/well were plated in 24 wells RetroNectin® (Takara)-coated plates in presence of 4 mg/ml of polybrene. In order to dilute polybrene, 2 hours after spinoculation, one volume of fresh medium was added to each well. 24 hours

after transduction, half of the blasts was collected and froze as reference starting point (t0) and the remaining cells were intravenously injected in recipient mice. Transduction efficiency was measured by FACS analysis of TagRFP positive cells, 72 hours after transduction.

6.3.4 Genomic DNA extraction and samples preparation for NGS

Both for screening and clonal tracking experiments, leukemic blasts were harvested from spleen and bone marrow of moribund mice. Genomic DNA was isolated by phenol/chloroform extraction as follow: cells were resuspended in 5-10 ml P1 buffer with RNAse A (Qiagen) and 0.5% SDS and incubated 5 min at RT. The resuspension volume depends on the amount of cells. For clonal tracking experiments, NB4 spike-in cells were added prior to phenol/chloroform extraction. One volume of phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma Aldrich) was added to cell suspension and the mix was vigorously vortexed and centrifuged at 4200 rpm for 2h at RT. The upper phase was collected and washed with an additional volume of chloroform (VWR). DNA precipitation was performed adding 1 volume of isopropanol (Panreac AppliChem) and 0.125 volumes of 3M sodium acetate to the upper phase. After over night incubation, the mixture was spun at 4200 rpm for 1h at RT. DNA pellet was then washed with 10 ml ethanol 70%, spun at 4200 rpm 10 min at RT and air-dried. DNA was finally dissolved in nuclease-free water and quantified by NanoDrop ND-100 (Thermo Fisher).

According to Cellecta guidelines, libraries for NGS were prepared using all the genomic DNA extract from each sample, in order to ensure the complete representation of the barcodes. The strategy foresees two nested PCR with primers designed to include the sequencing adapters needed for Illumina HiSeq 2000 flow cell clustering. In the first PCR, DNA barcodes are amplified from the total genomic DNA while, in the second PCR, a sampling of the 1st reaction is used to remove non-specific PCR products and excess genomic DNA. For all reactions, Titanium® Taq DNA polymerase (Takara) was used because very

efficient in amplify high amount of DNA. The 1st PCR reaction was prepared as following: 25 µg genomic DNA, 0.3 µM of each primer, dNTPs 200 µM, 1x Titanium® Taq Buffer, 0.5 U Titanium® Taq DNA polymerase, in a final volume of 100 µl per tube. The 1st PCR program is:

Temperature	Time	Cycles
94°C	3 min	1
94°C	30 sec	16
60°C	10 sec	
72°C	20 sec	
68°C	2 min	1

After the 1st PCR round, all reactions from the same sample were combined together and 5 µl of the first amplification were used for the 2nd PCR with primers containing P5 and P7 adapter sequences. The 2nd PCR reaction was prepared as following: 5 µl 1st PCR, 0.5 µM of each primer, dNTPs 200 µM, 1x Titanium® Taq Buffer, 0.5 U Titanium® Taq DNA polymerase, in a final volume of 100 µl. The 2nd PCR program is:

Temperature	Time	Cycles
94°C	3 min	1
94°C	30 sec	16
65°C	10 sec	
72°C	20 sec	
68°C	2 min	1

Due to differences in the shRNA and clonal tracking vectors, specific primers are used during barcodes amplification (Table 5). 10 µl of the 2nd PCR was run on a 3% agarose gel to confirm the correct size of the amplicons: 251 bp for M1 and 1.2k libraries and 267 bp for clonal tracking library. PCR products were purified with QIAquick PCR purification kit (Qiagen) following the manufacturer's protocol and quantified by Qubit™ dsDNA HS Assay Kit (Thermo Fisher). DNA libraries were sequenced on Illumina HiSeq 2000 sequencing platform, with 51 bp single-end reads, using the GexSeq sequencing primer. 20-40 million reads were generated for each sample. The original Cellecta strategy does not

allow to insert any index during library preparation, therefore it was not possible to sequence more samples in the same lane of the flow cell. To overcome this important limitation, we modified primers used in the second PCR in order to insert different 6 bp indexes. With this optimization and a specific bioinformatics pipeline we were able to sequence up to 5 samples in the same lane.

Primer name	Sequence 5'→3'	Use
F2	TCGGATTCGCACCAGCACGCTA	shRNA 1st PCR
R2	AGTAGCGTGAAGAGCAGAGAA	shRNA 1st PCR
Gex1_NF2	CAAGCAGAAGACGGCATACGATCGCACCAGCACGCTACGCA	shRNA 2nd PCR
Gex2_NR2	AATGATACGGCGACCACCGAGAGCACCGACAACAACGCAGA	shRNA 2nd PCR
Gex1_2	CAAGCAGAAGACGGCATACGAT ACATCG CGCACCAGCACGCTACGCA	shRNA 2nd PCR (index2)
Gex1_4	CAAGCAGAAGACGGCATACGATTGGTCACGCACCAGCACGCTACGCA	shRNA 2nd PCR (index4)
Gex1_6	CAAGCAGAAGACGGCATACGATATTGGCCGCACCAGCACGCTACGCA	shRNA 2nd PCR (index6)
Gex1_7	CAAGCAGAAGACGGCATACGATGATCTGCGCACCAGCACGCTACGCA	shRNA 2nd PCR (index7)
Gex1_12	CAAGCAGAAGACGGCATACGATTACAAGCGCACCAGCACGCTACGCA	shRNA 2nd PCR (index12)
FwdHTS3	TCGGATTCAAGCAAAAGACGGCATA	clonal tracking 1st PCR
R2	AGTAGCGTGAAGAGCAGAGAA	clonal tracking 1st PCR
Gex1_Bpi	TCAAGCAGAAGACGGCATACGAAGACA	clonal tracking 2nd PCR
P5_NR2	AATGATACGGCGACCACCGAGACGAGCACCGACAACAACGCAGA	clonal tracking 2nd PCR
GexSeq	AGAGGTTCAGAGTTCTACAGTCCGAA	NGS primer

Table 5. Primers used for NGS library preparation.

6.3.5 Preparation of NB4 spike-in control

We selected independent clones from a DNA library containing 13,000 different barcodes (Cellecta, Inc.). This library differs from the CellTracker™ library for the barcode structure, indeed it contains just a single barcode of 18 bp. Since the two libraries are cloned in the same lentiviral backbone, barcodes can be amplified with the same PCR strategy, described in section 6.3.4. Single barcodes were selected by bacterial transformation and isolation of single colonies. Plasmidic DNA isolated from bacteria was analyzed by Sanger sequencing to confirm the presence of a single barcode and used to produce lentiviral particles, as

described in section 6.3.2. NB4 cells were transduced at low MOI with each barcode to ensure the presence of single integrant per cells. Infected cells were selected with 1 μg/ml puromycin for 48 hours and defined numbers of NB4 cells clones were combined to prepare the spike-in control for clonal tracking experiments.

6.3.6 NGS reads alignment

FASTQ files were prepared from the sequencing runs to count the number of reads per barcode. For each sample, reads were aligned to the respective reference sequences (1.2k or M1 library) using the Bowtie aligner[175] and by considering only those reads having, at most, three mismatches for each alignment. Barcodes frequency was calculated by normalizing each barcode read count on the total number of aligned reads. For subsequent analysis, fold change was calculated as the ratio of barcodes reads between the spleen and the reference samples, either for 1.2k and M1 screening samples.

For clonal tracking experiments, barcodes reads were perfect match aligned to the reference library using the Bowtie aligner. Spike-in reads were aligned with the same procedure to the known barcode sequences. All read counts were normalized based on the total number of reads per sample and frequency distribution was calculated.

6.3.7 shRNA screening bioinformatics analysis

In order to identify depleted genes during the screening we developed two distinct bioinformatics approaches. ECDF analysis: we summed the 1.2k FC distributions to obtain a unique empirical cumulative distribution function (ECDF). The empirical distribution function is an estimate of the cumulative distribution function that generated the points in the sample; specifically, the cumulative distribution function (CDF) of a real-valued random variable X, or just distribution function of X, evaluated at x, is the probability that X will take a value less than or equal to x. Therefore, by mean of resulting ECDF from 1.2k samples, we calculated an empirical pValue for each shRNA present in M1 samples. We

then aggregate these shRNA pValues among sample replicates with Fisher's Method in order to obtain a FC, a pValue and a qValue for each shRNA. With the purpose of define enriched and depleted genes, we evaluated how many depleted shRNAs we would find by recursively 1000 random picking 10 shRNAs, regardless the gene they target. One gene can be considered depleted if at least 5 out of 10 hairpins are significantly depleted, with 95% confidence. The threshold for significant enrichment is instead of at least 4 out of 10 shRNAs.

EdgeR analysis: EdgeR package was developed for RNAseq analysis but it can be used as well for shRNA screening statistical analysis[176]. Prior to proceed, samples were adjusted to account for differences in library size. In particular, TMM normalization was applied on counts and then a multidimensional scaling (MDS) plot was generated to assess the consistency between replicate samples. The barcodes variability was estimated under the assumption that common dispersion across barcodes is flat and the more the barcodes are present, the lower dispersion we have. Samples were analyzed by fitting a generalized linear model (GLM). We first analyzed the changes in barcodes abundance during the 1.2k screening to set a threshold for enrichment/depletion analysis in M1 screening. Statistical testing for changes in barcodes abundance between spleens and reference samples was carried out using exact test that allow results to be ranked by significance. We then applied the significance thresholds calculated on 1.2k control samples, to the shRNAs of M1 samples. Finally, since 10 hairpins per gene are present, we used Roast, a gene set analysis tool, to obtain a gene-by-gene ranking, rather than a shRNA specific one.

6.4 Limiting dilution transplantation assay

For the limiting dilution transplantation assay, MLL-AF9 cells, freshly isolated from moribund mice, were intravenously injected (from $5x10^5$ to 100 cells per mice) into sub-lethally irradiated (5 Gy) C57 BL/6 J recipient mice, with $2x10^5$ splenocytes as carrier. LIC frequency was measured used ELDA software[177].

6.5 Validation experiments

6.5.1 Virus production

Single shRNA targeting candidate genes were cloned in the same lentiviral vector used for the screening (pRSI16, see section 6.3.1). We used empty vector and shRNA directed against Luciferase as controls. Lentiviral particles were produced transfecting 293T cells with shRNA plasmid and 2nd generation packaging vectors, as reported in section 6.3.2.

6.5.2 Blasts infection and sorting

MA9 blasts were maintained in Iscove's Modified Dulbecco's Medium (IMDM, Lonza), 2 mM L-glutamine, 15% Fetal Calf Serum (FCS), 15% 5637-conditioned medium and 25% WEHI-3B-conditioned medium. 20x10^6 blasts/well were plated in 6-wells RetroNectin®-coated plates and infected with single shRNA lentiviral particles by spinoculation at 2300 rpm, for 90 min at RT, in presence of 4 mg/ml of polybrene. 2 hours after spinoculation, one volume of fresh medium was added to each well. Transduction efficiency was measured by FACS analysis 72 hours after transduction and infected blasts were sorted as TagRFP+ cells using MoFlo® Astrios™ cell sorter (Beckman Coulter).

6.5.3 RNA reverse-transcription and qPCR

RNA was extracted from TagRFP+ sorted cells using PicoPure™ RNA Isolation Kit (ThermoFisher), according to manufacturer protocol. 0.1-1 µg of total RNA was reverse-transcribed using ImProm-II™ Reverse Transcriptase kit (Promega). RNA was first incubated with random primers (0.5 µg/reaction) at 70°C for 15 min and then immediately on ice. The following mix was then added:

ImProm-II™ 5X reaction buffer	10 µl
$MgCl_2$ 25 mM	5 µl
dNTPs (10 mM each)	2.5 µl
recombinant RNAse inhibitor	1.5 µl
ImProm-II™ Reverse Transcriptase	1 µl
Nuclease-free water	to 50 µl

Reaction was incubated at 42°C for 70 min and then at 70°C for 15 min. cDNA was stored at -20°C.

For gene expression analysis, qPCR was performed using 10 ng of cDNA, 0.2 µM of both primers and 10 µl of FAST SYBR™ Green Master Mix, AmpliTaq® Fast DNA Polymerase (ThermoFisher) in a final volume of 20 µl per reaction in 96-well plate. Fluorescence accumulation during qPCR reaction was detected on CFX96 Touch™ Real-Time PCR Detection System (Biorad). Relative mRNA quantification was performed by the comparative ΔΔCt method using Tbp for normalization. Primers used are listed in Table 6.

Primer name	Sequence 5'→3'
Brd4 forward	AAAACTCCAACCCCGATGAG
Brd4 reverse	GAACCAGCAATCACGTCAAC
Stat1 forward	GCCGAGAACATACCAGAGAATC
Stat1 reverse	GATGTATCCAGTTCGCTTAGGG
Syt4 forward	AATGGTGTGAGGCTGGAAG
Syt4 reverse	ACCACTTCGCCATTACTGATC
Socs2 forward	TGAAGCATGAGCCTTTCCTC
Socs2 reverse	GCAGACACTGTCACCCAC
Gfi1 forward	TGGAGCAACACAAGGCAG
Gfi1 reverse	AGTACTGACAGGGATAGGGC
Tie1 forward	CCAGTGCCAGTGTCAAAATG
Tie1 reverse	CCTATGTTGAACTCCACCTCTG
Hoxa5 forward	CAAGCTGCACATTAGTCACG
Hoxa5 reverse	GGTAGCGGTTGAAGTGGAAT
Tbp forward	TAATCCCAAGCGATTTGCTG
Tbp reverse	CAGTTGTCCGTGGCTCTCTT

Table 6. qPCR primers used to check gene silencing.

6.5.4 Serial colony-forming-unit assay

500 TagRFP+ sorted MA9 blasts were plated in MethoCult M3434 (StemCell Technologies) in 35 mm dish. 7 days after plating, colony formation was quantified. For serial passages, 500 cells of each replicate were used for subsequent plating in MethoCult M3434.

6.5.5 Validation in vivo

$2x10^5$ TagRFP+ sorted MA9 blasts were intravenously injected in sub-lethally irradiated C57 BL/6 J 8-12 weeks old mice (3-4 mice/group). Latency in leukemia development was monitored and mice were sacrificed when moribund, according to animal facility guidelines. Leukemic infiltration was assessed in peripheral blood, spleen and bone marrow by measuring bt FACS the percentage of TagRFP+ cells over the total CD45.1+ population.

7. Aim of the project

Chemotherapy-resistant AML cells are thought to be enriched in quiescent LSCs and several publications highlighted the importance of quiescence for AML maintenance. However, additional studies are required to define the role of quiescence and underlying mechanisms during LSCs selection and leukemia development.

Our lab had previously demonstrated that cell cycle restriction is critical in preventing excess DNA-damage accumulation and functional exhaustion of LSCs in PML-RARα and AML1-ETO driven leukemia[107]. Moreover, NPMc+ expression in normal bone marrow is able to expand the LT-HSCs number without depleting the pool of quiescent LT-HSCs, therefore preserving their functionality (unpublished results).

We decided to investigate if induction of HSCs quiescence is a distinctive property of AML-associated oncogenes, the role of HSCs quiescence during the leukemogenic process, and if HSCs quiescence is required for LSCs maintenance of leukemia outgrowth. Therefore, the general aim of my thesis project is to investigate whether oncogene-induced pathways controlling HSCs quiescence, as identified in pre-leukemic cells, are also critical for the retention of LSCs self-renewal capacity and ultimately for leukemia maintenance. As experimental approach, we decided to take advantage of RNA interference technology to perform *in vivo* screenings to identify genes that could be required for the maintenance of AML. In particular, the specific aims of the project are:

1. To obtain gene expression profiles of pre-leukemic HSCs expressing different AML initiating oncogenes, namely NPMc+ and PML-RARα.
2. To identify common pathways enriched in the pre-leukemic transcriptional profiles, with a particular focus on stemness- and quiescence-related genes.
3. To select ~100 quiescence-related genes induced by the leukemic oncogenes in pre-leukemic HSCs to design a low complexity shRNA library.

4. To perform *in vivo* shRNA screening on NPMc+ and PML-RARα driven murine AML samples.
5. To develop a bioinformatics pipeline to identify genes significantly depleted or enriched during the screening in order to proceed with validation experiments *in vitro* and *in vivo*.

Employing this strategy on a number of leukemia characterized by different genetic alterations, we aimed at identifying not only genes fundamental for a single subtype but to discover key genes common to several AML, which may offer new therapeutic strategies in a wide group of patients.

8. Results

8.1 NPMc+ and PML-RARα AML models are characterized by a prolonged pre-leukemic phase

As model system, we started our investigation with two AML mouse models characterized by the expression of the cytoplasmic mutant of NPM1 (NPMc+)[171] or the fusion proteins PML-RARα[61]. In both model systems, leukemia recapitulates the main features of the corresponding human diseases and develops after a prolonged pre-leukemic phase.

NPMc+ transgenic mice harbor the cDNA of the most frequent NPM1 mutation (mutA, found in about 80% of the patients)[50] which is expressed upon CRE-mediated excision of the floxed STOP cassette present between the promoter and the NPMc+ cDNA[171]. Oncogene expression can be tracked *in situ* thanks to the co-expression of NPMc+ and the yellow florescent protein (YFP) (see Materials and Methods, section 6.1). Bone marrow mononucleated cells (BM-MNCs) isolated from NPM1c+$^{fl/-}$/YFP$^{fl/-}$ (NPMc+/YFP) or YFP$^{fl/-}$ control mice (YFP) were treated *ex vivo* with the recombinant TAT-CRE protein and FACS-sorted in YFP+ and YFP- subpopulations. Notably, NPMc+ is expressed only in the YFP+ population, with a recombination efficiency of ~85%, assessed by immunofluorescence (unpublished, not shown). BM MNCs expressing NPMc+ are transplanted in irradiated syngeneic recipient mice which develop leukemia with low penetrance (33.3%) and long latency (median 564 days)[171]. PML-RARα transgenic mice harbor the cDNA of the fusion protein targeted into the murine cathepsin G locus. These mice express constitutively low level of PML-RARα in HSCs and progenitors, and develop leukemia with high penetrance (90%) and 10 months median latency[61].

Thus, both NPMc+ and PML-RARα oncogenes are sufficient to initiate the leukemogenic process but require other genetic lesions for the full expression of a leukemia phenotype. Indeed, the two models are characterized by a prolonged phase whereby oncogenes are

expressed in an otherwise normal bone marrow. During the pre-leukemic phase, the organization of the hematopoietic tissue is near identical to that of normal mice, thus allowing direct comparison of identical cell types, such as the HSCs, and unambiguous identification of de-regulated gene expression. Given the long latency we observed prior of leukemia development, we were able to evaluate early perturbation occurring upon oncogene expression in normal hematopoietic cells, in order to define critical mechanisms during the leukemogenic process.

8.2 Gene expression profiles of pre-leukemic LT-HSCs showed an enforcement of quiescent stem cell transcriptional program

To investigate the transcriptional changes induced by oncogene expression in HSCs, we performed global gene expression analysis of LT-HSCs, purified as Lin-, Sca1+, cKit+, CD34- and Flk2- cells, from NPMc+ and PML-RARα pre-leukemic mouse models. Sorting strategy is depicted in Figure 8. LT-HSCs NPMc+/YFP+ or YFP+ have been isolated 4 months after BMT while LT-HSCs PML-RARα or WT have been isolated from 10-12 weeks old mice.

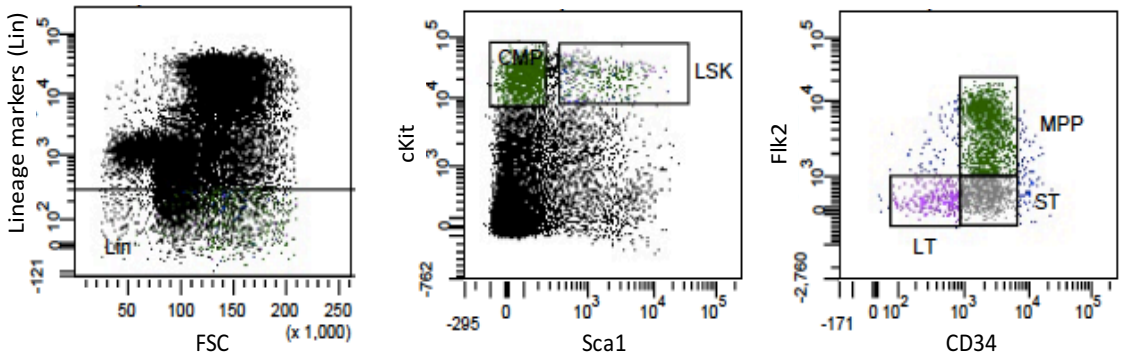

Figure 8. Pre-leukemic and WT LT-HSCs are sorted form from BM-MNCs. Representative FACS plot of sorting strategy used to isolate both WT and pre-leukemic LT-HSCs. LSK (Lin^-, $Sca1^+$, $cKit^+$) population was gated within the lineage negative cells. LT-HSCs were gated as CD34-, Flk2- cells within the LSK population.

RNA was extracted just after sorting and hybridized on Mouse Gene ST 2.0 Arrays. NPMc+ expression induced a total of 562 deregulations (322 up-regulations, FC>1.5; 240 down-regulations, FC<-1.5, FDR<0.1) while PML-RARα induced a total of 195 deregulations (117 up-regulations, FC>1.5; 78 down-regulations, FC<-1.5, FDR<0.1).

We ran a Gene Set Enrichment Analyses (GSEA)[174] on the two pre-leukemic expression profiles. GSEA revealed that NPMc+ expression in LT-HSCs: (1) induces the expression of genes associated with HSCs and reduces the expression of genes associated with mature hematopoietic cells[178,179] (Figure 9 A), (2) up-regulates genes involved in quiescence-related pathways such as TGFβ and TPO (Figure 9 B) and, (3) shows a marked correlation with the expression profiles from acute myeloid LSCs and human AML with mutated NPM1[101,180] (Figure 9 C).

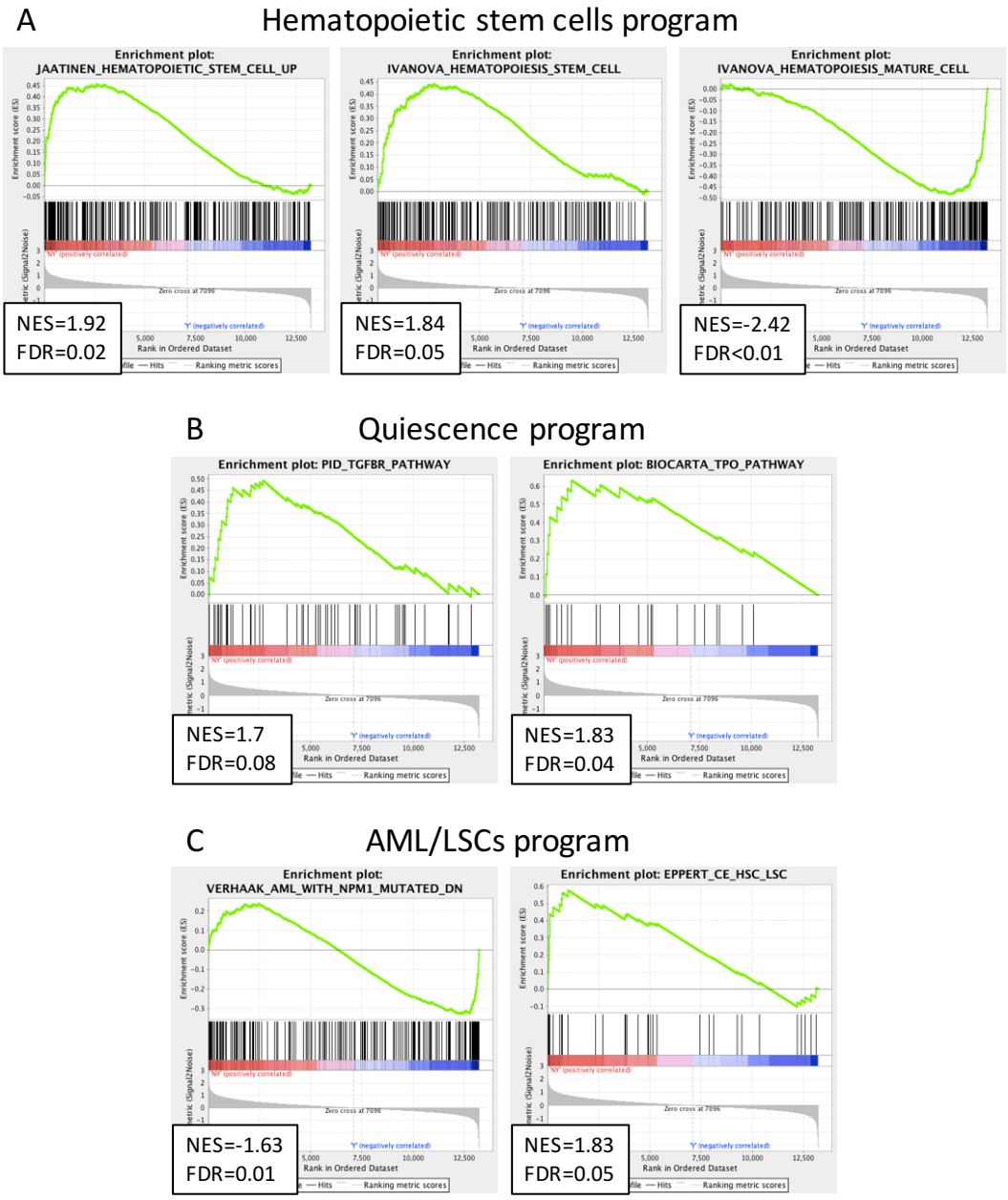

Figure 9. NPMc+ expression in LT-HSCs enforces a stem cell transcriptional program promoting quiescence. GSEA plots demonstrating enrichment levels of indicated gene sets in NPMc+ LT-HSCs compared to WT LT-HSCs. **A.** GSEA plot correlating genes up-regulated in NPMc+ LT-HSCs with HSCs transcriptional program and genes down-regulated in NPMc+ LT-HSCs with mature hematopoietic cells expression profile. **B**. GSEA plot correlating genes up-regulated in NPMc+ LT-HSCs with TGFβ and TPO pathways. **C**. GSEA plot correlating genes down-regulated in NPMc+ LT-HSCs with genes down-regulated in NPM1 mutated AML, and genes up-regulated in NPMc+ LT-HSCs with genes enriched in HSCs and LSCs. Normalized enrichment score (NES) and false discovery rate (FDR) are indicated.

As expected, among the up-regulated genes, we also found the HoxA cluster and Meis1, genes involved in self-renewal maintenance and often deregulated in AML, including NPMc+ AML[56]. Thus, GSEA data suggest that NPMc+ expression in LT-HSCs enforces a quiescent HSCs transcriptional program, which also characterizes LSCs. To further investigate molecular mechanisms related to quiescence regulation, we manually curated a list of genes directly involved in the maintenance of HSCs or LSCs quiescence, using data available in literature. Notably, we found many quiescence genes up-regulated by NPMc+ in LT-HSCs in our microarray data (Table 7).

Gene symbol	FC	Reference
Gata2	2.1	[27]
Angpt1	2	[136]
Smad3	1.9	[181]
Meis1	1.8	[182]
Cdkn2c	1.7	[183]
CD81	1.6	[184]
Cbfa2t3	1.6	[185]
Cdkn1a	1.6	[21]
Tgfbr2	1.6	[181]
Egr1	1.6	[186]
Cited2	1.4	[29]
Tgfb1	1.4	[181]
Mpl	1.4	[6]
Pml	1.4	[121]
Pbx1	1.4	[187]
Terc	1.4	[188]
Kit	1.3	[189]
Gfi1	1.3	[28]
Chd4	1.3	[190]
Cebpa	1.3	[191]

Table 7. NPMc+ expressing LT-HSCs up-regulated a set of quiescence positive regulators. For each quiescence gene is reported the FC of expression in NPMc+ LT-HSCs compared to WT LT-HSCs, based on microarray data, and the reference describing the role in quiescence regulation.

We validated by qPCR 15 out of 17 tested genes, including self-renewal genes (HoxA cluster and Meis1) and quiescence genes such as Gata2 and Gfi1 (Figure 10).

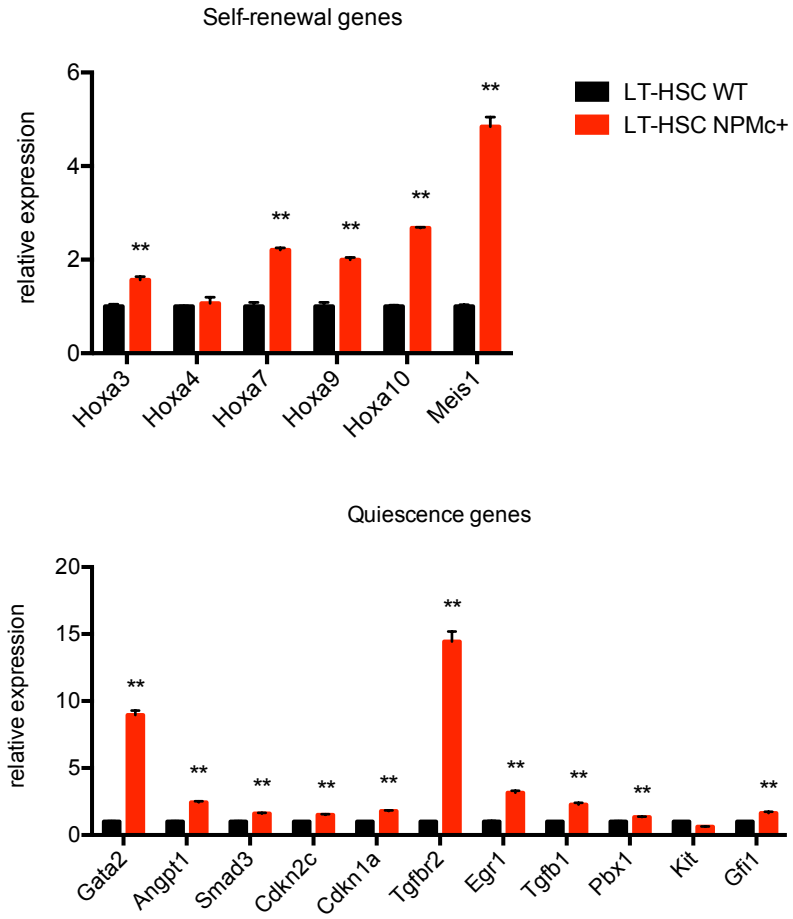

Figure 10. NPMc+ expression in LT-HSCs induces the up-regulation of self-renewal and quiescence genes. qPCR analysis of WT and pre-leukemic LT-HSCs RNA for self-renewal (upper panel) and quiescence genes (lower panel) up-regulated by NPMc+. Results were normalized to the expression levels in WT LT-HSCs (** $p<0.01$).

In the other model, GSEA revealed that PML-RARα expressing LT-HSCs are characterized by: (1) induced expression of genes associated with HSCs and reduced expression of genes associated with mature hematopoietic cells transcriptional profiles[179] (Fig. 11 A), (2) down-regulation of genes associated with proliferating HSCs[192,193] (Fig. 11 B), (3) up-regulation of genes involved in RXR/RAR pathways or bound by PML-RARα[194] (Fig. 11 C) and, (4) reduced expression of genes down-regulated in LSCs[195] (Fig. 11 D).

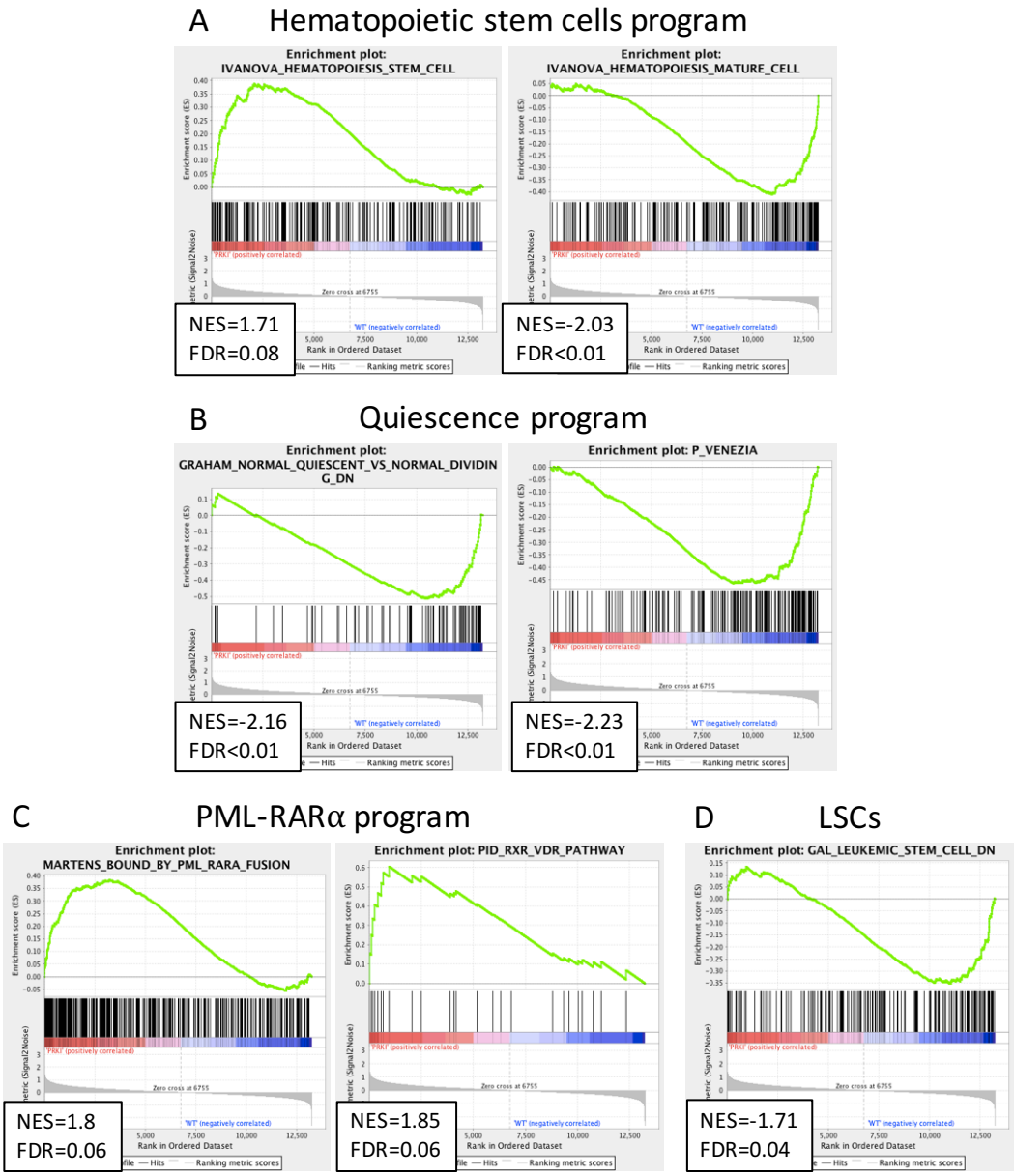

Figure 11. PML-RARα expressing LT-HSCs are enriched in stem cells and quiescence signatures. GSEA plots demonstrating enrichment levels of indicated gene sets in PML-RARα LT-HSCs compared to WT LT-HSCs. **A.** GSEA plot correlating genes up-regulated in PML-RARα LT-HSCs with genes up-regulated in HSCs and genes down-regulated in PML-RARα LT-HSCs with genes expressed in mature hematopoietic cells. **B**. GSEA plot correlating genes down-regulated in PML-RARα LT-HSCs with genes down-regulated in quiescent HSCs and proliferating genes. **C**. GSEA plot correlating genes up-regulated in PML-RARα LT-HSCs with genes bound by PML-RARα and with the retinoic acid/vitamin D nuclear receptors pathway. **D**. GSEA plot correlating genes down-regulated in PML-RARα LT-HSCs with genes down-regulated in LSCs. Normalized enrichment score (NES) and false discovery rate (FDR) are indicated.

Moreover, PML-RARα up-regulated the expression of a different set of quiescence positive regulators, described in Table 8. Interestingly, although PML-RARα and NPMc+ are very different leukemic oncogenes, we found 6 quiescence genes commonly up-regulated by both oncogenes: Gata2, Cbfa2t3, Cdkn1a, Cdkn2c, Smad3 and Tgfbr2. We validated by qPCR the PML-RARα dependent up-regulation in LT-HSCs of all the quiescence genes we tested (Figure 12).

Gene symbol	FC	Ref
Cdkn1a	2.1	[21]
Egr3	2.1	[196]
Hes1	1.7	[197]
Satb1	1.5	[198]
Tgfbr2	1.5	[181]
Junb	1.5	[199]
Fzd8	1.4	[200]
Cbfa2t3	1.4	[185]
Smad3	1.4	[181]
Atg7	1.4	[201]
G0s2	1.4	[202]
Cdkn2c	1.3	[183]
Gata2	1.3	[27]

Table 8. PML-RARα LT-HSCs up-regulate a set of quiescence positive regulators. For each gene is reported the FC of expression in PML-RARα LT-HSCs compared to WT LT-HSCs, based on microarray data, and the reference describing the role in quiescence regulation. Genes up-regulated in common with NPMC+ LT-HSCs are depicted in red.

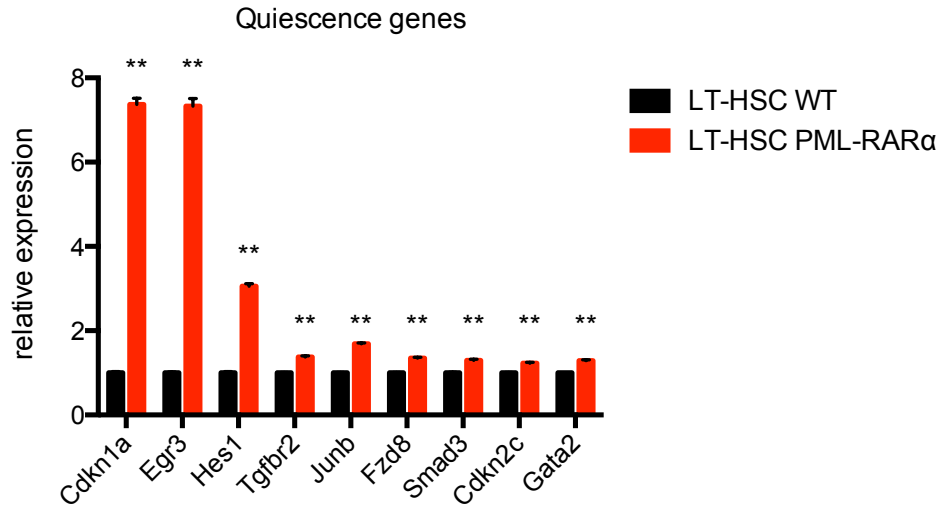

Figure 12. PML- RARα expression in LT-HSCs up-regulates quiescence genes. qPCR analysis of WT and pre-leukemic LT-HSCs RNA for a set of quiescence genes up-regulated by PML-RARα. Results were normalized to the expression levels in WT LT-HSCs (** $p<0.01$).

8.3 NPMc+ and PML-RARα up-regulate a common set of quiescence genes in LT-HSCs

Strikingly, both NPMc+ and PML-RARα LT-HSCs gene expression profiles are significantly enriched in a set of genes preferentially expressed in adult quiescent HSCs, compared to more proliferating HSCs (i.e. fetal liver-HSCs and HSCs mobilized with 5-FU treatment) or ST-HSCs[193] (Figure 13 A). Interestingly, among this signature, the genes accounting for the enrichment in both pre-leukemic transcriptional profiles showed a 30% overlap ($p<0.01$) (Figure 13 B).

Collectively, these data suggest that the expression of two unrelated oncogenes in LT-HSCs is sufficient to impose a transcriptional program that is highly specific for quiescent HSCs and it also correlates with LSCs/AML expression profiles. Thus, we tested the hypothesis that the enforcement of quiescence might be critical for the maintenance of the transformed clones during both the pre-leukemic and the leukemic phase by a reverse genetic approach. We selected, among genes up-regulated by NPMc+ and PML-RARα, ~100 genes involved

in stem cells quiescence regulation and we designed a shRNA library to perform *in vivo* genetic screening.

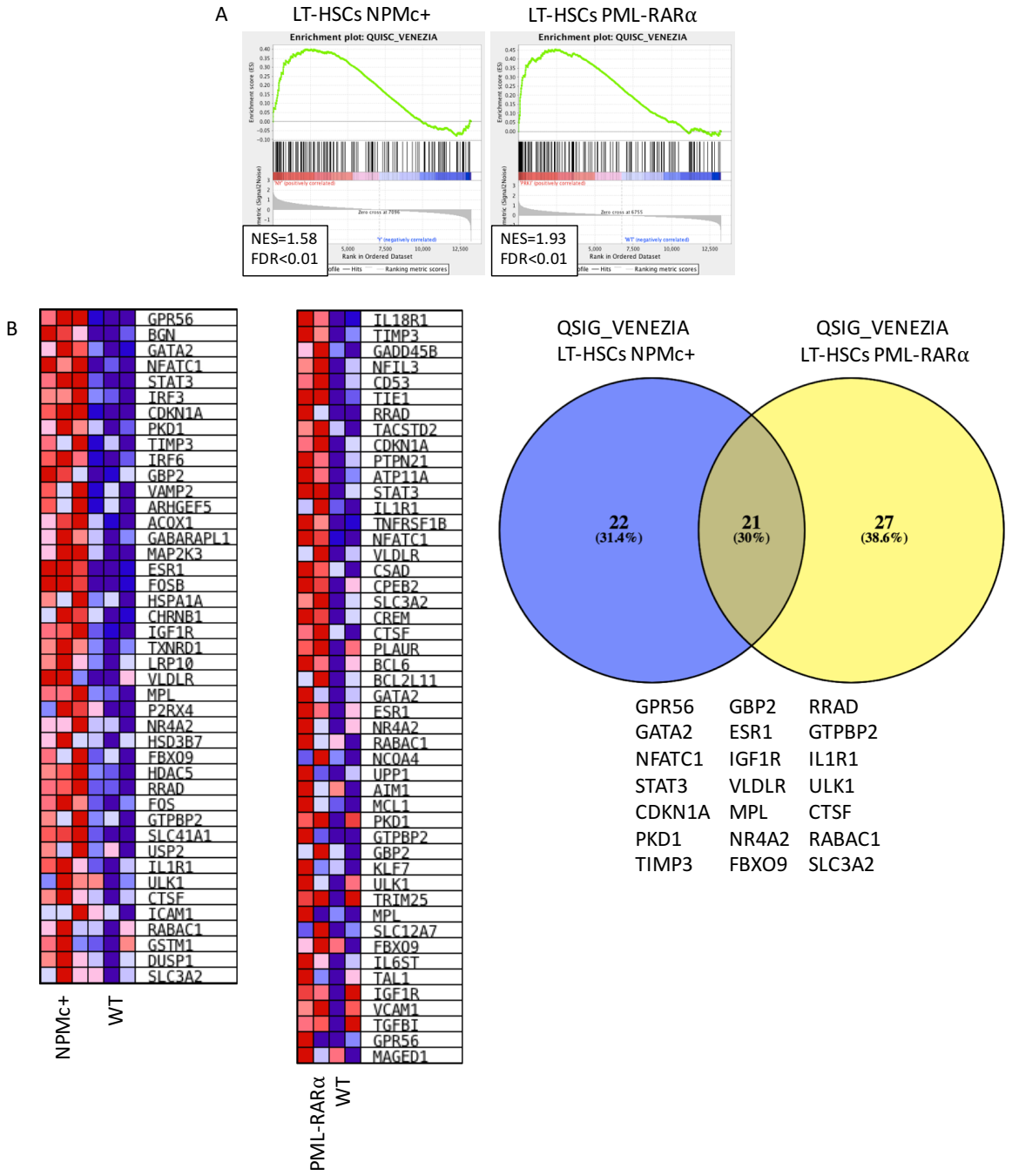

Figure 13. NPMc+ and PML-RARα LT-HSCs up-regulate the same set of genes characteristic of quiescent HSCs. A. GSEA plots showing the enrichment of a quiescent HSCs signature in NPMc+ and PML-RARα LT-HSCs compared to WT LT-HSCs. **B**. Heatmap showing the most enriched quiescent HSCs genes up-regulated by NPMc+ and PML-RARα, compared to WT LT-HSCs, and their overlap ($p<0.012$). Decreased gene expression is indicated by shades of blue; increased expression is indicated by shades of red.

8.4 The *in vivo* shRNA screening in NPMc+ and PML-RARα AML revealed clonality issues

We charged Cellecta Inc. with the design of a custom shRNA library cloned in the pRSI16 lentiviral vector. Lentiviral vectors allow transduction of both cycling and non-cycling cells, such as quiescent LSCs, and infected cells can be identified by the expression of the TagRFP reporter gene, also encoded by the pRSI16 backbone. The library used in the screen, from now on called M1 library, has a total complexity of 1,000 different shRNAs, targeting 92 genes and 4 controls, with 10 different hairpins for each gene. It is important to include control shRNAs (e.g. with an expected behavior with respect to the phenotype selected), such as shRNAs targeting essential genes, as controls of depletion, and shRNAs with no biological effect, as control of infection of a sufficient number of LSCs. In our library we inserted 20 hairpins targeting the essential genes Polr2b, Psma1 and Rpl30 and 20 shRNAs targeting the neutral gene Luciferase (LUC). Genes targeted in the screen are involved in diverse cellular process ranging from JAK/STAT pathway (Stat1, Stat3, Stat5a, Socs2), to TGFβ pathway (Tgfb1, Tgfbr2, Jun, Smad3, Cited2), some of them are involved in cytokines signaling (Il18r1, Irf1, Irf3, Irf6, Angpt1, Kit, Tie1, Mpl), in metabolic processes (Abcb1b, Acox1, Adcy9, Adssl1, Txnrd1) or in transcriptional and epigenetic regulation (Pbx1, Gata2, Jun, Fos, Hoxa5, Kmt2a, Hdac5, Brd4). Importantly, each shRNA is univocally associated with a unique 22 bp DNA barcode, allowing identification and quantification of the corresponding hairpins by NGS.

To study the behavior of barcodes during the screening *in vivo,* and better characterize our model system, we designed a control library composed by 1,200 different DNA barcodes but not associated to any shRNA (thereafter 1.2k library). Therefore, the *in vivo* distribution of the barcodes is not affected by gene silencing but it only reflects the intrinsic functional heterogeneity of the blasts. We performed experiments in parallel using the two libraries. The direct comparison between the shRNAs and the non-targeting barcodes allows

identification of the shRNAs that are enriched or depleted during leukemia growth, as compared to the neutral selection of the barcodes.

The experimental strategy for shRNA screening is depicted in Figure 14. Transduction of murine NPMc+ or PML-RARα leukemic blasts, obtained from our mouse models, was performed *in vitro* by spinoculation, using concentrated viral supernatant at low multiplicity of infection (MOI=3). 24 hours after transduction, half of the cells were harvested as reference point (t0), while the remaining were intravenously injected in two recipient mice. For the NPMc+ leukemia samples, recipient mice were sublehtally (5 Gy) irradiated 6-8 hours before transplantation to allow engraftment (that for this leukemia is dependent on prior irradiation).

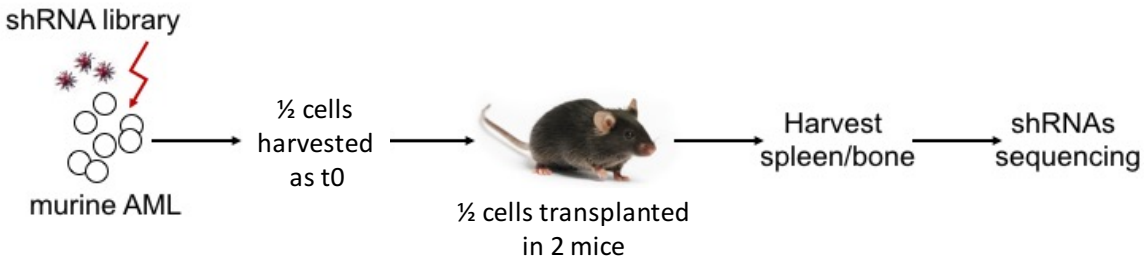

Figure 14. Strategy for shRNA screen *in vivo*.

Mice have been checked weekly for leukemia development by monitoring the peripheral blood (PB). When the percentage of blasts in the PB was >80% (measured as percentage of CD45.2+ cells in C57 BL/6 Ly5.1 recipient mice), mice were sacrificed and spleen (SPL) and bone marrow were collected. Genomic DNA was purified and libraries for NGS were prepared according to Cellecta protocol (see Materials and Methods section 6.3.4). NGS data have been processed as described in detail in the Materials and Methods section 6.3.6.

NGS data showed that in the reference samples, harvested 24 hours after transduction, all barcodes, of both the M1 and the 1.2k library were present, with a distribution range comprised within a 10-fold variance (Figure 15 A-B). This reflects the expected barcode

distribution, based on how libraries have been synthetized by the supplier. Therefore, there were no unbalanced representation of shRNAs/barcodes in the cells used for the *in vivo* screen.

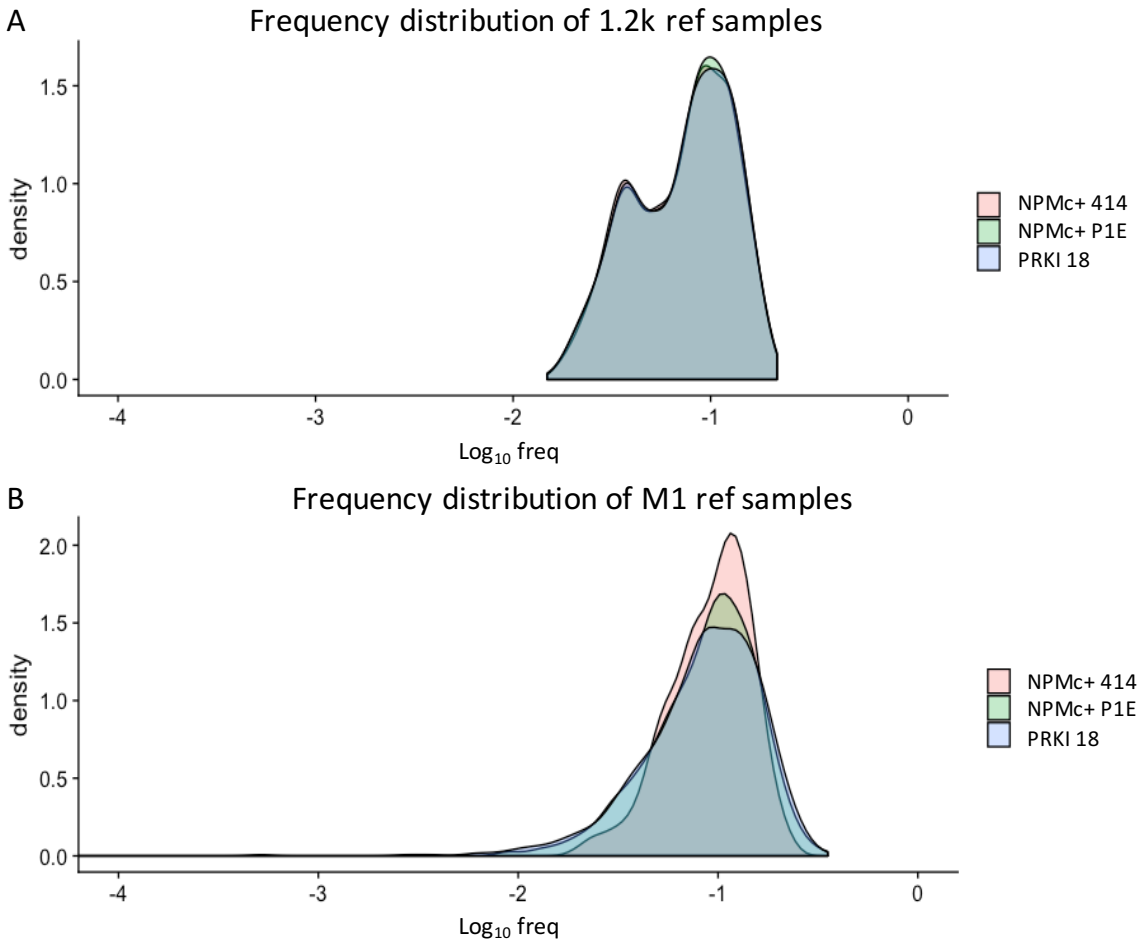

Figure 15. Barcode distribution is balanced in NPMc+ and PML-RARα reference samples (t0). A. Log_{10} frequency distribution of barcodes in reference samples infected with 1.2k library, 24 hours after transduction. **B**. Log_{10} frequency distribution of shRNAs in reference samples infected with M1 library, 24 hours after transduction. In NPMc+ P1E M1 ref sample is present one shRNA at very low frequency.

NGS analysis of the *in vivo* grown samples, instead, did not allow the identification of every barcode in the leukemic specimens. In experiments performed using the NPMc+ 414 AML, we were able to recover 86-96% and 85-92%, respectively, of shRNAs in the M1 screening and of the barcodes in the 1.2k screening. Moreover, their distribution showed a strong shift

towards the low frequencies due to a significant expansion of few barcodes, as shown in Figure 16 A-B.

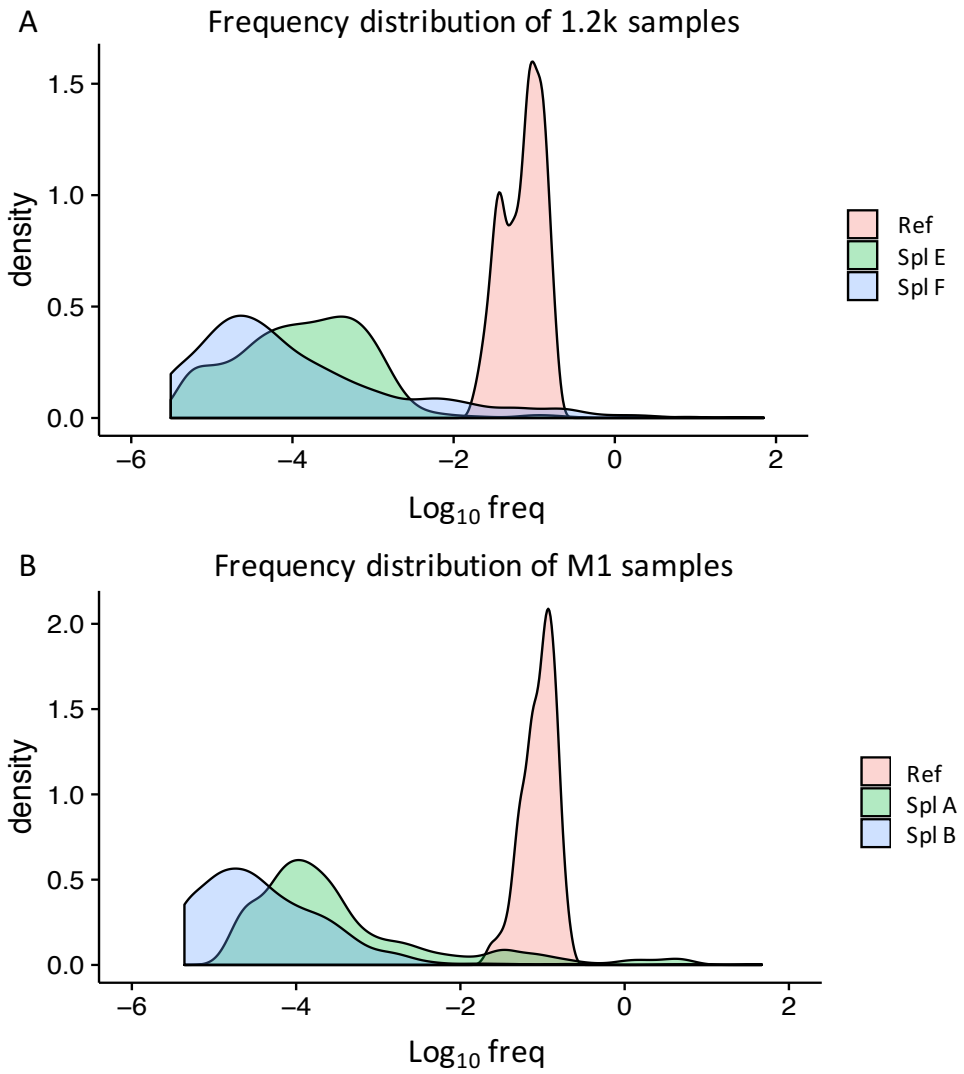

Figure 16. Barcode frequency distribution show a strong shift upon AML growth. **A**. Log_{10} frequency distribution of .2k library barcodes in reference sample (Ref) and in two spleens (Spl E and Spl F) after leukemia growth. Spl E and Spl F were collected from two mice transplanted with the same blasts transduced with 1.2k barcodes library (same infection of Ref sample). **B**. Log_{10} frequency distribution of shRNAs in reference sample (Ref) and in two spleens (Spl A and Spl B) after leukemia growth. Spl A and Spl B were collected from two mice transplanted with the same blasts transduced with M1 shRNA library (same infection of Ref sample).

In particular, in sample Spl A, the 20 most represented shRNAs accounted for ~80% of the sample reads, while in Spl B the 20 most represented shRNAs accounted for more than 99% of the sample, with the first 2 shRNAs accounting for >87% (Figure 17 B). The most represented hairpins in the SPL samples were not the most abundant in the reference sample and, among the top 20 hairpins, only one was found in both replicates, reflecting the poor correlation between the two samples (r=0.23) and suggesting that the expansion of selected shRNA was not biologically driven (Figure 17 D). Accordingly, the same behavior was found in in the control screening with the 1.2k library, where the 20 most represented BCs in the two replicates, Spl E and Spl F, accounted for >90% of the samples (Figure 17 A). The barcode-identity correlation between the two 1.2k replicates was lower than in M1 screening (r=0.16), however in this case it was expected since barcodes should distribute randomly *in vivo* (Figure 17 C). Unfortunately, we obtained similar results with another independent NPMc+ leukemia (NPMc+ P1E) and with one PML-RARα (PRKI 18) murine leukemia (not shown).

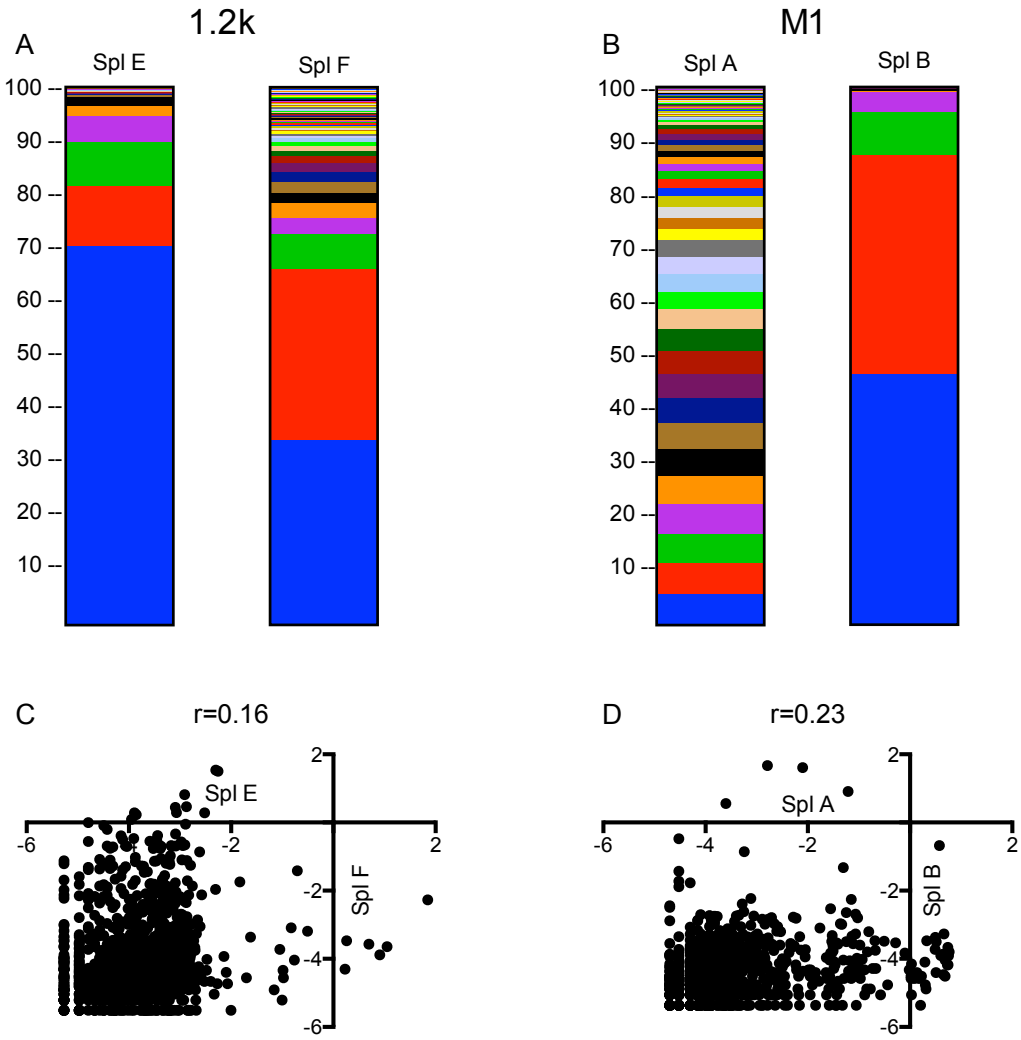

Figure 17. Few barcodes are highly expanded during *in vivo* screens with NPMc+ leukemia. **A**. Barcodes frequency distribution in 1.2k SPL samples. Spl E and Spl F were collected from two mice transplanted with the same blasts transduced with 1.2k barcodes library (same infection of Ref sample). **B**. shRNAs frequency distribution in M1 SPL samples. Spl A and Spl B were collected from two mice transplanted with the same blasts transduced with M1 shRNA library (same infection of Ref sample). *N.B. colors do not indicate barcodes identity. **C**. Pearson's correlation between barcodes representation in Spl E and Spl F (replicates of 1.2k screening). **D**. Pearson's correlation between shRNAs representation in Spl A and Spl B (replicates of M1 screening).

Together, these data suggest that individual LICs are extremely heterogeneous in their clonal expansion potential, with few LICs that posses a strong growth advantage over the others, and dramatically expanded *in vivo* after transplantation. From the perspective of our screening, the heterogeneous potential of LICs clonal expansion overcomes the selective pressure of shRNAs expression, thus preventing any meaningful bioinformatics analysis aimed at identifying biologically selected or counter-selected hairpins. Therefore, before

performing shRNA screening with AML samples in vivo, it is critical to evaluate the growth potential of individual LICs.

8.5 Clonal tracking experiments allow to study clonal growth *in vivo*

In order to further verify our hypothesis on clonal expansion and to set up a standard assay to evaluate in vivo clonal behavior of a given AML sample, we traced LICs *in vivo* using viral cellular insertion DNA barcoding and high-throughput sequencing. In particular, we performed clonal tracking experiments in which 1 million of leukemic blasts were transduced with a lentiviral library of DNA barcodes with a molecular complexity of 30 million, generated by the combination of 18+18 degenerated nucleotides sequences. *In vitro* transduction was performed at low multiplicity of infection (MOI=3) by spinoculation (see Materials and Methods section 6.3.3). The high library complexity and the low MOI used for the transduction assured integration of a unique DNA barcode into the cell genome (preliminary data on human AML, not shown) and identification of the cell progeny descending from each labelled cell. 24 hours after transduction, cells were intravenously injected into recipient mice and one aliquot has been frozen as reference (t0). Once mice developed leukemia, we harvested blasts from spleen and bone marrow to extract genomic DNA. Upon DNA library preparation, barcode sequences were identified by NGS (see Materials and Methods sections 6.3.4). For each sample, sequence reads were aligned to the reference library with no mismatch allowed and reads per barcode were normalized on total number of reads obtained. Reference samples were characterized by the presence of many barcodes at very low frequency (median 0.0003%, not shown), indicating that no barcode unbalance was present before transplantation.

8.6 Spike-in to set a threshold for clone identification

Due to the presence of several barcodes at low frequency, in order to establish a threshold for clone identification, we took advantage of a spike-in control. Prior to genomic DNA extraction from the *in vivo* grown samples, we added known quantities of NB4 cells (from 10 to 250,000 cells) each infected with a single different barcode of 18 nucleotides that can be amplified with the same PCR primers used for clonal tracking barcodes amplification. Therefore, during DNA library preparation for NGS, PCR product of samples containing the spike-in show two bands, one at the expected size of 267 bp and one smaller (249 bp) identifying barcodes in NB4 cells. After sample sequencing, we evaluated the relationship between the number of NB4 cells in the spike-in and the number of reads obtained for each barcode. As shown in Figure 18, we observed a linear relationship between the amount of cells harboring a specific barcode and the number of reads obtained for that barcode. Barcode associated to 10 cells did not retrieve any reads and barcode associated with 50 cells had a number of reads out of the linear range. However, considering only barcodes associated with higher amount of cells (>100 cells), we calculated a good linear correlation (R^2>0.98), indicating that PCR amplification does not saturate the relative frequency among different clones.

Spike-in barcodes	Nr. of NB4 cells	NPMc+ 414 spl spike-in reads	NPMc+ P1E spl spike-in reads
Clone 6 TGACGTGTGTGTTGTGGT	10	0	0
Clone 1 CAACACGTACTGGTGTCA	50	24	34380
Clone 4 GTTGGTGTACCAGTACCA	100	10002	40964
Clone 10 ACTGCATGACGTGTACTG	1000	21546	80382
Clone 9 GTGTTGCATGCACAACTG	10000	116498	736260
Clone 7 TGGTTGCACATGCACACA	50000	483138	3149386
Clone 5 ACACACGTGTCATGTGAC	250000	2426862	14642890

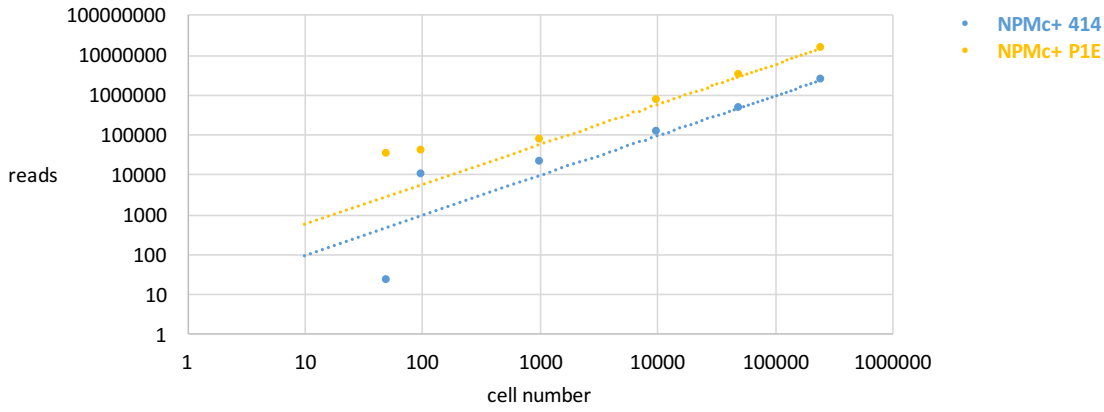

Figure 18. Spike-in control showed a linear relationship between the number of cells harboring a specific barcode and the NGS reads for the same barcode. In the table are reported the number of NB4 cells infected with a specific barcode and the number of reads retrieved by NGS in two independent NPMc+ AML samples. The graph in the lower panel show the linear relationship between number cells and NGS reads.

8.7 Clonal tracking *in vivo* confirms clonal expansion during leukemia growth

Clonal tracking experiments on two NPMc+ AML samples showed that very few clones are highly expanded *in vivo*. In particular, in both NPMc+ leukemia, only 2 barcodes account for >90% of total reads (Figure 19). Interestingly, identity of the predominant barcodes in the spleen and the bone marrow of the same mouse was conserved (Figure 19 A).

These findings confirmed our hypothesis that among leukemic cells, few LICs have a strong growth advantage and are clonally selected *in vivo*. This, in turn, is associated with the presence of minor clones at very low frequency, below the threshold of sensitivity set by the spike-in.

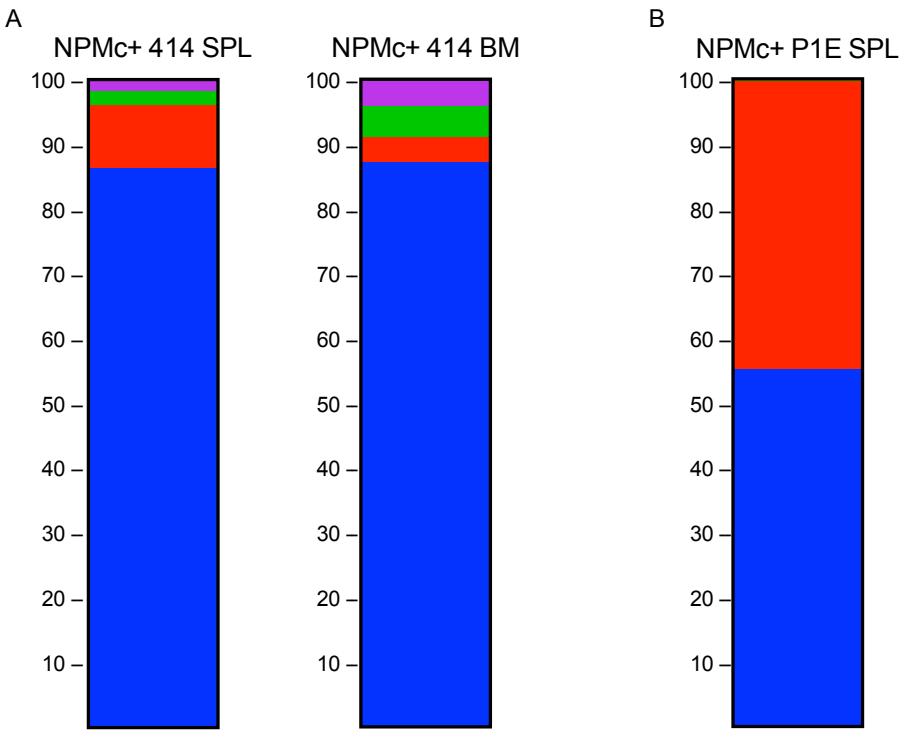

Figure 19. Clonal tracking confirmed a strong clonal expansion during NPMc+ leukemia growth *in vivo*. Clonal tracking barcodes frequency distribution in two independent NPMc+ leukemia samples. **A**. Barcodes distribution in NPMc+ 414 SPL and BM. **B**. Barcodes NPMc+ P1E SPL. All barcodes found in the samples are depicted in the graph. *N.B. colors do indicate barcodes identity only between SPL and BM in NPMc+ 414 samples.

Assuming that only LICs are able to grow upon transplantation in recipient mice, the number of clonal barcodes retrieved *in vivo* can be used to measure LIC frequency, with the following formula:

$$LIC\ freq = \frac{nr.\ transduced\ cells\ injected}{nr.of\ barcodes\ above\ the\ threshold}$$

However, in our experimental conditions, only a small proportion of clones were present at a frequency above the threshold set by the spike-in, due to the excessive clonal expansion of very few clones, therefore preventing a reliable evaluation of the LIC frequency in our AML samples (Table 9). Indeed, LICs with lower growth potential could not be scored in clonal representation because of the threshold set for clone identification.

	NPMc+ 414 SPL	NPMc+ P1E SPL
Total barcodes (BCs)	68	99
BCs considering spike-in threshold to 100 cells	5	2
Transduction efficiency	15%	14%
LIC frequency	1:30,000	1:70,000

Table 9. LIC frequency calculated by clonal tracking is affected by the low number of barcodes above the threshold.

In conclusion, two features of either NPMc+ or PML-RARa leukemias, namely relatively-low LIC frequency and high clonal heterogeneity of LICs, prevent the possibility of performing in vivo genetic screens, due to the low molecular complexity supported by both models.

8.8 The MLL-AF9 driven AML model

We then considered the possibility to use a different AML model. Notably, like human NPMc+ leukemia, human MLL-rearranged leukemia are characterized by the de-regulated expression of members of the HOX gene family[203], which are involved both in HSCs in self-renewal regulation and AML development. Both NPMc+ and MLL fusion proteins have been shown to cooperate with the BET chromatin regulator BRD4 in the regulation of gene expression[204,205]. In NPMc+ leukemia, inhibition of BRD4 by NPM1 is lost, while, in MLL-rearranged AML, the fusion partner is often part of the super elongation complex (SEC) that recruits BRD4 to promote transcription. Recently, the SEC complex, MLL1 and the H3K79 methyltransferase DOT1L have been shown to control MEIS1 and HOX genes expression in NPMc+ AML[206]. Moreover, in AML patients, NPMc+ and DNMT3A mutations often co-occur while both alterations are mutually exclusive with MLL-rearrangements[54,207], indicating that mutations in distinct epigenetic regulators share a convergent mechanism to induce self-renewal deregulation through MEIS1and HOX genes overexpression. Therefore, based on the similarity to NPMc+ AML, and the fact that it has been previously used to

perform shRNA screen both *in vitro* and *in vivo*[166,167,208], we decided to perform our genetic screens using the MLL-AF9 (MA9) murine AML.

8.9 The MLL-AF9 pre-leukemic LSK gene expression profile is enriched in quiescent stem cell signatures

To analyze MA9-dependent transcription in defined hematopoietic subpopulation, we generated pre-leukemic mice, by transplanting MA9 tranduced Lin- cells into lethally irradiated recipient mice. Mice were analyzed after 4 weeks from Lin- transplantation, to allow complete HSCs engraftment in the bone marrow niche. However, at this time point, the physiological organization of the bone marrow was totally lost (not shown), due to the ability of MA9 expression to induce accelerate leukemogenesis. Unfortunately, at 2 weeks after transplantation, control mice, transplanted with Lin- cells transduced with empty vector, had not yet completely reconstituted the bone marrow of the animals, while mice transplanted with MA9 Lin- cells showed a transformed-like phenotype (not shown). In conclusion, the MA9 behaves as a potent transforming oncogene in this model system, preventing appreciation of a pre-leukemic phase.

Thus, we decided to take advantage of previously reported transcriptional profiles of MA9 pre-leukemic LSK cells (Lin^-, $Sca1^+$ and Kit^+) derived from MA9 KI mice[209]. The LSK compartment is enriched in stem cells and it is composed by LT-HSCs, ST-HSCs and MPPs. MA9 KI mice harbor the fusion gene in the endogenous Mll locus and are characterized by a prolonged pre-leukemic phase (~6 months) with myeloid proliferation and expansion of LSK cells, as compared to WT mice. Interestingly, this phenotype is similar to what we observed in our NPMc+ model where, upon transplantation of NPMc+ expressing BM-MNCs, we scored an expansion of the donor LSK compartment (unpublished data). In addition, the authors described a MA9-associated gene signature composed by 192 genes up-regulated in four pre-leukemic subpopulations (LSK, CMP, GMP and CLP).

Interestingly, NPMc+ and PML-RARα LT-HSCs gene expression profiles are enriched in this MA9 signature (Figure 20).

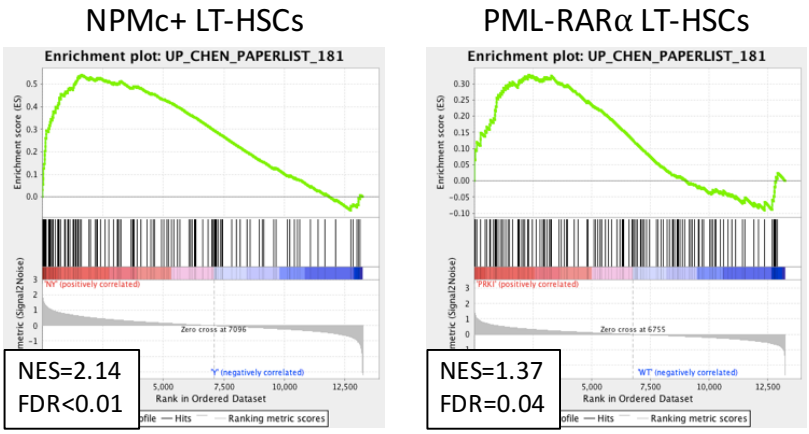

Figure 20. Genes up-regulated by MA9 are enriched in NPMc+ and PML-RARα LT-HSCs gene expression profiles. A. GSEA plot correlating genes up-regulated in NPMc+ LT-HSCs with genes up-regulated by MA9. **B**. GSEA plot correlating genes up-regulated in PML-RARα LT-HSCs with genes up-regulated by MA9.

Analysis of the MA9 LSK gene expression profile showed a strong up-regulation of genes directly involved in quiescence regulation such as Tie2 (FC=1.8) and its ligand Angpt1 (FC=1.4)[136], Gimap5 (FC=1.7)[210], Gata2 (FC=1.5)[27], Tal1 (FC=1.3)[211] and Pdk2 (FC=1.3)[37].

8.10 Clonal tracking and LIC frequency evaluation in MLL-AF9 leukemia

To evaluate LIC frequency and the extent of LIC clonal heterogeneity in the MA9 leukemia, we performed *in vivo* clonal tracking experiments: one million MA9 blasts were infected with the 30 million barcodes library and transplanted in recipient mice. Upon leukemia development, we prepared sequencing libraries from genomic DNA extracted form the collected leukemic spleens. Samples were submitted to NGS and reads aligned to the reference library with no mismatches allowed. We retrieve 3,636 barcodes from one reference sample (Ref) and 159 different barcodes from one spleen (SPL) (Figure 21). Notably, the most represented barcode account for 14.8% of total reads, as compared to the

NPMc+ leukemia, where the most represented barcodes accounted for 50-80% of the sample.

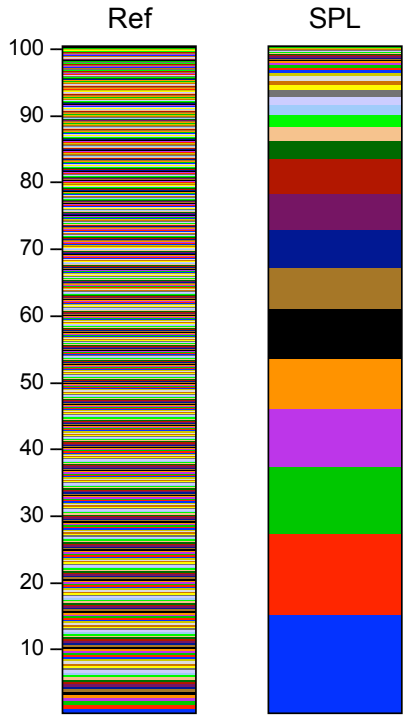

Figure 21. MA9 leukemia grows in vivo without major clonal expansion. Clonal tracking barcodes distribution in reference (Ref) and spleen (SPL) samples. All barcodes found in the samples are depicted in the graph. *N.B. colors do not indicate barcodes identity.

We identified 52 barcodes above the threshold, set by the spike-in (>100 cells). Considering the amount of injected MA9 blasts ($1x10^6$) and the transduction efficiency of 2%, measured by FACS analysis as percentage of TagRFP+ cells, we estimated a LIC frequency of 1 in 385 blasts. LIC frequency was also confirmed by Limiting Dilution Transplantation Assay (LDTA), the gold standard to evaluate LIC frequency in a given sample. Leukemic cells were transplanted into recipient mice at decreasing doses (from 500,000 to 100) and the proportion of mice that developed leukemia was used to calculate the number of self-renewing cells present in the sample. The results confirmed a high LIC frequency in MA9 leukemia model of 1:471 blasts (Table 10). Interestingly, in this AML model, LIC frequency estimation gave similar results with both the assay used, suggesting that all the LICs are able

to contribute to the growth of the leukemia *in vivo* and can be tracked under our experimental conditions.

Cell dose MA9	Mice	Dead
500000	2	2
100000	4	4
10000	4	4
1000	5	3
100	5	4

Confidence intervals for 1/(stem cell frequency)

Lower	Estimate	Upper
1253	471	177

Table 10. MA9 LIC frequency calculation by limiting dilution. LDTA was performed by injecting decreasing numbers of MA9 into recipient mice (range 100-500000). The number of mice died of leukemia is indicated for each cell dose. LIC frequency was calculated using ELDA software[177].

In conclusion, analysis of LIC frequency and clonal heterogeneity suggest that MA9 leukemia can support the molecular diversity of our libraries during *in vivo* screenings.

8.11 MLL-AF9 shRNA screening *in vivo*

In vivo screens were performed as described for NPMc+ and PML-RARα leukemia. Briefly, MA9 blasts were transduced *in vitro* with the M1 or 1.2k libraries at MOI=3 using the spinoculation protocol (see Materials and Methods section 6.3.3). Based on the calculated LIC frequency, we determined the number of MA9 blasts to be transduced in order to cover the complexity of our shRNA library. Indeed, especially in drop-out screens, every shRNA should be ideally present in 5-10 LICs to score its effect and avoid false negative results. To this end, estimating a transduction efficiency of ~20%, we independently transduced $24x10^6$ blasts for each replica. 24 hours after transduction, we collected half of the cells as reference points (t0) and used the other half for intravenously transplantation in recipient mice. A small

aliquot of cells was FACS analyzed 72 hours after transduction to evaluate infection efficiency, as percentage of TagRFP+ cells. Transduction efficiency was, on average, 17% and 22% for the M1 and 1.2k samples, respectively. We transplanted 6 mice for M1 and 4 mice for 1.2k screening. Three weeks post transplantation mice showed signs of leukemia development such as high percentage of donor cells and high WBC count in the peripheral blood, palpable splenomegaly and pale limbs. Therefore, we sacrificed all the animals and collected blasts from both the spleen and the bone marrow. All mice showed splenomegaly and high blasts infiltration in spleen and bone marrow, measured as percentage of CD45.1+ cells (MA9 blasts) on total cells retrieved from the same animals (C57 BL/6 J). We then proceed with genomic DNA extraction and library preparation for NGS (see Materials and Methods section 6.3.4). For each mouse we sequenced DNA from the spleen sample and the corresponding reference sample, for a total of 20 samples.

8.12 Identification of genes depleted during the *in vivo* screening

For each sample, barcodes sequencing reads were aligned to the corresponding reference library. Reference samples (Ref) had a uniform barcodes distribution comprised in around 10-fold variance for both libraries (Figure 22 A-C). Spleen samples (SPL), instead, showed a shift towards low frequencies for a number of barcodes, in both the M1 and in 1.2k library, much more evident in the M1 samples (Figure 22 B-D). However, this shift was not pronounced as the one observed in NPMc+ and PML-RARα screens, and we did not observe expansion of few barcodes, neither for M1 nor for 1.2k library. Indeed, the most abundant barcode accounted for <4% of total reads in each M1 sample, while it accounted for <0.5% of total reads in each 1.2k sample.

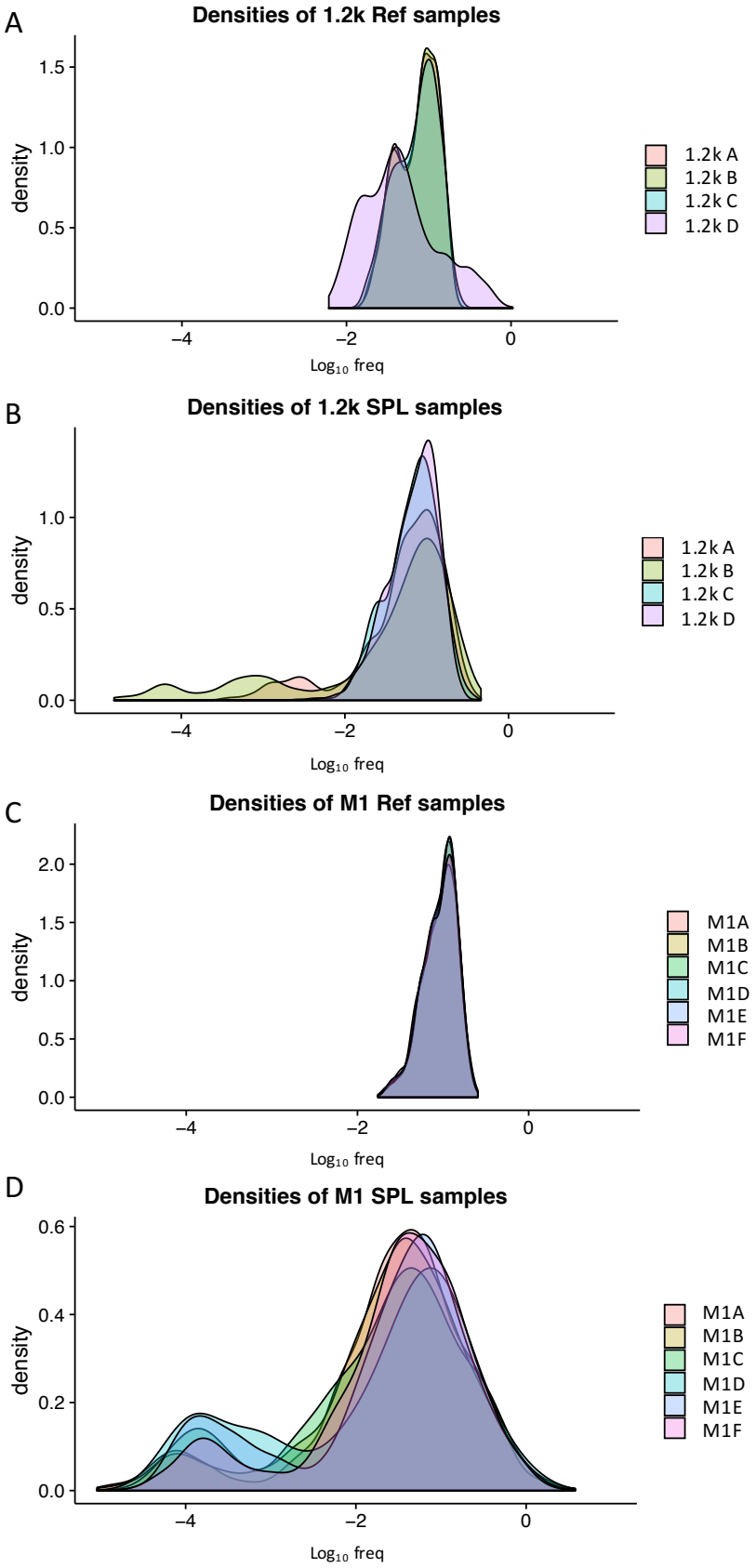

Figure 22. Barcode frequency distributions show a moderate shift during screening *in vivo*. **A**. Log_{10} barcodes distributions of 1.2k Ref samples. **B**. Log_{10} barcodes distributions of 1.2k SPL samples after *in vivo* screen. **C**. Log_{10} barcodes distributions of M1 Ref samples. **D**. Log_{10} barcodes distributions of M1 SPL samples after *in vivo* screen.

The correlation between M1 SPL samples was much higher than the one we obtained in the previous screens, with Pearson's correlation coefficients ranging from 0.44÷0.55 (Figure 23).

	M1A	M1B	M1C	M1D	M1E	M1F
M1A		0.55	0.53	0.48	0.51	0.52
M1B	0.55		0.53	0.47	0.52	0.54
M1C	0.53	0.53		0.44	0.49	0.52
M1D	0.48	0.47	0.44		0.47	0.48
M1E	0.51	0.52	0.49	0.47		0.49
M1F	0.52	0.54	0.52	0.48	0.49	

Figure 23. shRNAs have a similar distribution in the 6 screening replicates. Pearson's correlation of shRNAs abundance between each M1 SPL samples.

To increase the robustness of the screen, we analyzed 3 additional M1 samples (M1G, M1H and M1L) and one 1.2k, obtained from an independent *in vivo* screening, performed under identical experimental conditions. Samples were normalized and a multi-dimensional analysis was performed to visualized similarities among them. Interestingly, 2 out of the 3 additional M1 SPL samples, M1G and M1L, clustered together with the other M1 SPL samples, while all the reference samples clustered very close to each other (Figure 24 B). On the contrary, 1.2k SPL samples were less homogenously organized, with some of them clustering with the Ref samples (1.2k spl C, spl D, spl E and spl F), and other far away from each other (1.2k spl A and spl B) (Figure 24 A). Since the 1.2k barcodes should be devoided of any biological effect, they are expected to behave randomly *in vivo*.

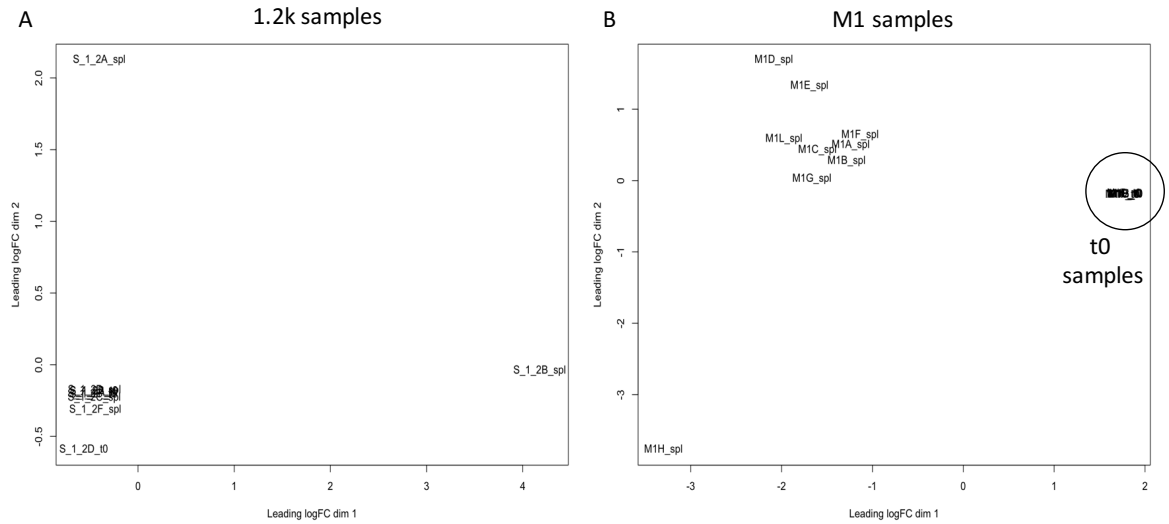

Figure 24. Multi-dimensional scaling (MDS) plot show relationship between screening samples. A. 4 out of 6 1.2k Spl samples cluster with 1.2k Ref samples while two 1.2k Spl samples cluster far away from each other. **B**. Most of M1 Spl samples cluster together far from Ref samples.

We then performed bioinformatics analyses of the 9 replicates for the shRNA screening and the 5 replicates for the control screening. We developed two different pipelines for the identification of genes depleted or enriched during screenings *in vivo*. In the first pipeline, we analyzed the barcodes distribution in the SPL and the respective Ref sample calculating the fold change for each barcode (FC=SPL/t0). Distribution of FC showed that M1 samples have a clear bimodal distribution of shRNAs that is less evident in 1.2k samples (Figure 25).

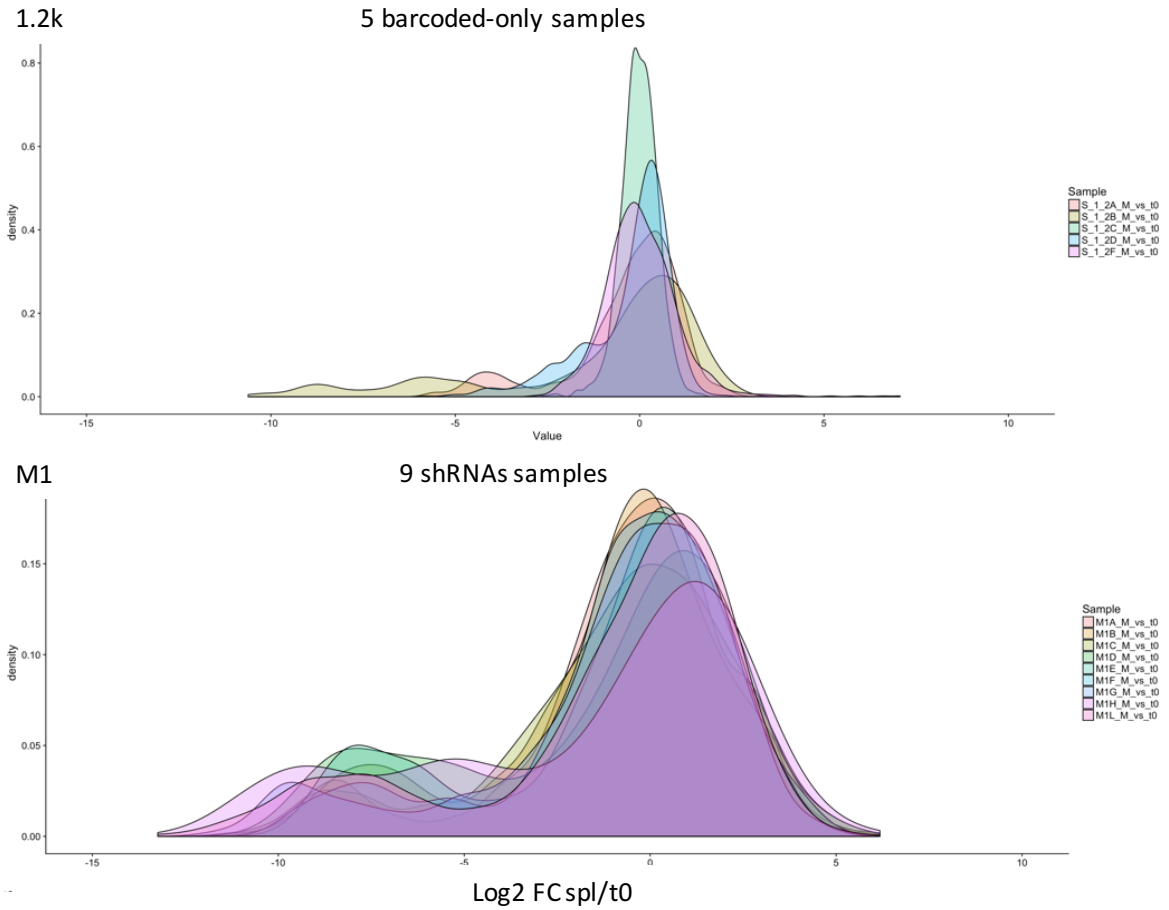

Figure 25. shRNAs fold change distribution shows a number of depleted hairpins *in vivo*. The upper panel shows the $\log_2$FC (SPL/t0) distributions of the barcodes in the 5 samples of 1.2k screening. The lower panel shows the $\log_2$FC (SPL/t0) distributions of the shRNAs in the 9 samples of M1 screening.

We then used the "random" distribution of the 1.2k barcodes *in vivo* to define a threshold for the identification of significantly depleted/enriched shRNAs in M1 samples. To calculate an empirical pValue for each shRNA of each M1 sample, we computed an empirical cumulative distribution function (ECDF) summing up all 1.2k control sample distributions. We then aggregated these pValues among replicates for each shRNA using the Fisher's Method, and obtained, for each shRNA, a FC, a pValue and a qValue. We identified 291 significantly depleted shRNAs and 223 significantly enriched shRNAs. Since for each gene were designed 10 hairpins, we established a threshold to define how many depleted/enriched shRNAs were sufficient to call a gene as depleted. To this end, we evaluated how many depleted shRNAs we would find by random picking 10 shRNAs among the others,

regardless the gene they target, for 1,000 times. This analysis revealed that if one gene has at least 5 of its hairpins depleted, we can define it as depleted with 95% confidence. We did the same for enrichment and we set the threshold to ≥4 shRNAs to define a gene as enriched with 95% confidence. With this analysis we identified 15 significantly depleted and 14 significantly enriched genes.

For the second bioinformatics workflow, we took advantage of a previously published tool based on the EdgeR pipeline[176]. Briefly, after sample normalization, we used the FC (SPL/t0) and pValue distributions of the 1.2k control samples to set a significance threshold to be applied to shRNA samples. We thus identified 167 significantly depleted shRNAs and 205 significantly enriched shRNAs. EdgeR package contains a gene set analysis-based tool, Roast, which allows to define if a gene is significantly depleted/enriched taking into consideration all the shRNAs targeting that gene. Roast analysis identified 19 significantly depleted and 18 significantly enriched genes. Among the depleted genes, in both analysis, we scored essential genes such as Rpl30 and Polr2b, and Brd4, already known to be critical for MA9 leukemia maintenance[166] (Figure 26).

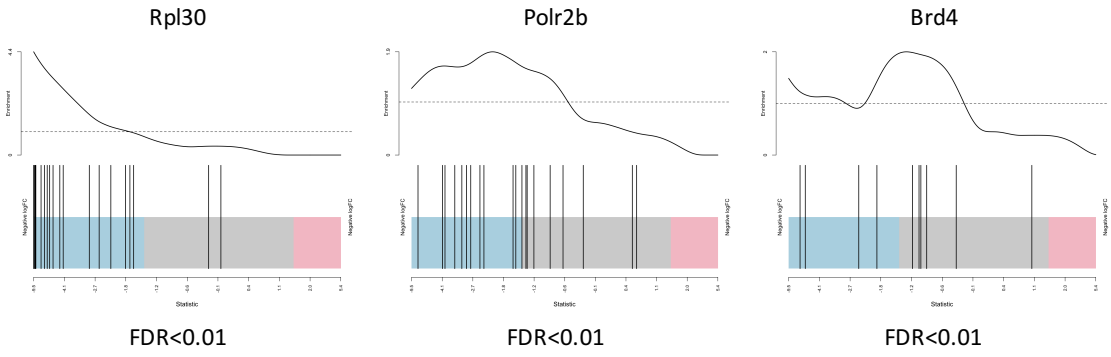

Figure 26. Essential genes are scored as depleted by Roast gene set analysis. Barcode plots highlighting the rank of shRNAs targeting a specific gene relative to all other hairpins present in the M1 library.

We decided to start validation experiments by targeting genes in common between the two analyses which overlap for 10 genes (Figure 27).

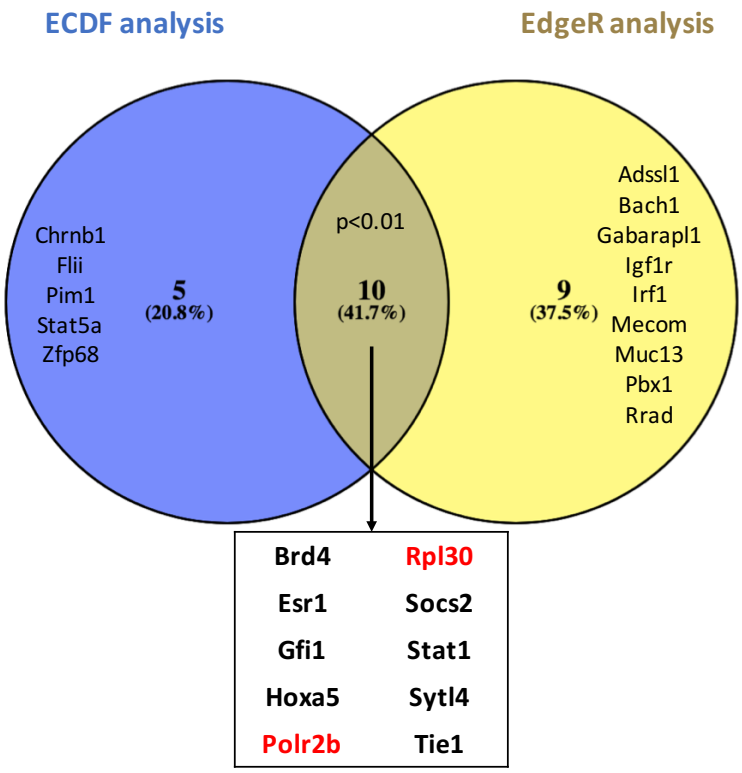

Figure 27. Depleted genes in common between the two analyses are selected for validation experiments. Overlap between the depleted genes identified with the first pipeline (ECDF, in blue) and depleted genes identified with EdgeR pipeline (in yellow).

8.13 Candidate genes validation

MA9 blasts were transduced with single shRNAs targeting each of the candidate gene, or with empty vector (EV) as control, and TagRFP+ cells were sorted 72 hours after transduction to obtained a pure population of infected blasts. RT-qPCR confirmed silencing of target genes at sorting time, compared to blasts infected with EV (Figure 28 A). To functionally test the effect of gene silencing *in vitro*, 500 sorted blasts were plated in methylcellulose (MC) cultures and colonies counted after one week. MA9 cells infected with shRNA targeting Brd4, Stat1 or Sytl4 were significantly less clonogenic, as compared to EV-transduced MA9 cells (Figure 28 B).

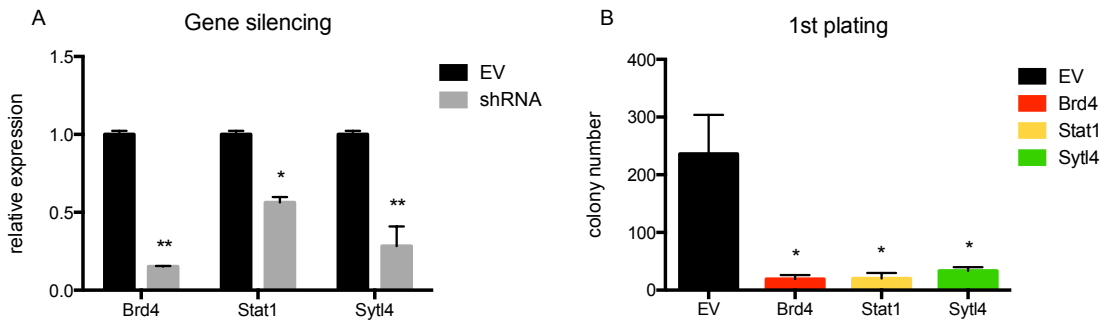

Figure 28. shRNA-mediated gene silencing reduces colony-forming efficiency *in vitro*. **A.** qPCR analysis of MA9 blasts infected with shRNAs targeting Brd4, Stat1 or Sytl4, compared to MA9 blasts infected with empty vector (EV), 72h post transduction. **B**. 500 MA9 TagRFP+ blasts infected with Brd4, Stat1 or Sytl4 shRNAs or EV were cultured in methylcellulose medium. The number of colonies scored after one week is shown (*p<0.05, **p<0.01).

Unexpectedly, upon two subsequent MC re-plating experiments, colonies quantification did not show the same differences scored in the first passage (not shown). However, when we counted only the TagRFP+ colonies (e.g. cells expressing the specific shRNAs) we scored a significant decrease in the Brd4, Stat1 and Sytl4 silenced samples, as compared to the control (Figure 29 A). Accordingly, upon serial passages, the percentage of TagRFP+ cells decreased in all shRNA samples, while it was maintained constant in the control blasts infected with EV (Figure 29 B). These data demonstrated that down-regulation of Brd4, Stat1 or Sytl4 was sufficient to decrease colony-forming efficiency of MA9 leukemia.

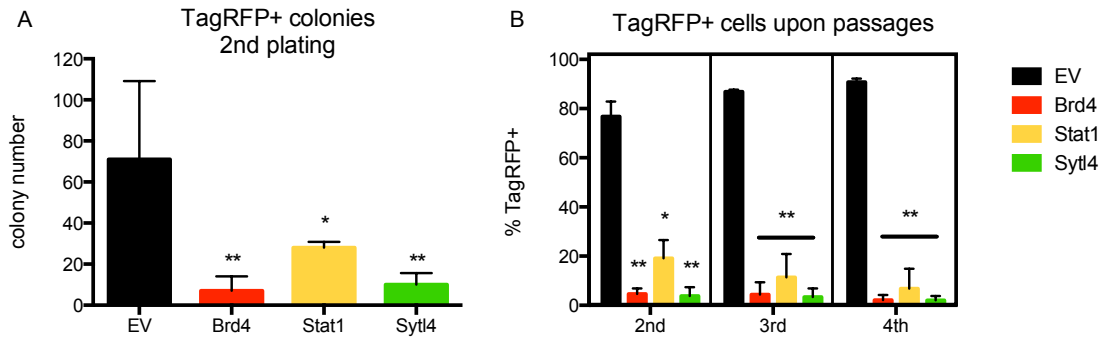

Figure 29. MA9 blasts infected with Brd4, Stat1 or Sytl4 shRNA are counter selected *in vitro*. **A**. 500 MA9 blasts from 1st plating in methylcellulose have been re-plated and TagRFP+ colonies only were counted one week later. **B**. At each re-plating in methylcellulose, the percentage of TagRFP+ cells was measured by FACS (*p<0.05, **p<0.01).

We next investigated the relevance of the candidate genes in leukemia maintenance *in vivo*. We intravenously injected $2x10^5$ TagRFP+ sorted blasts infected with EV, Brd4, Stat1, Sytl4 or LUC (Luciferase) shRNA in sub-lethally irradiated recipient mice and we monitored leukemia development. Silencing of candidate genes significantly delayed leukemia onset, as compared to controls (median survival: EV=33 days, LUC=34 days, Brd4=51 days, Stat1=57 days and Sytl4=79 days) (Figure 30). EV and LUC mice succumbed of leukemia with the latency observed in other experiments, confirming that lentiviral infection did not alter the growth potential of MA9 leukemia *in vivo*.

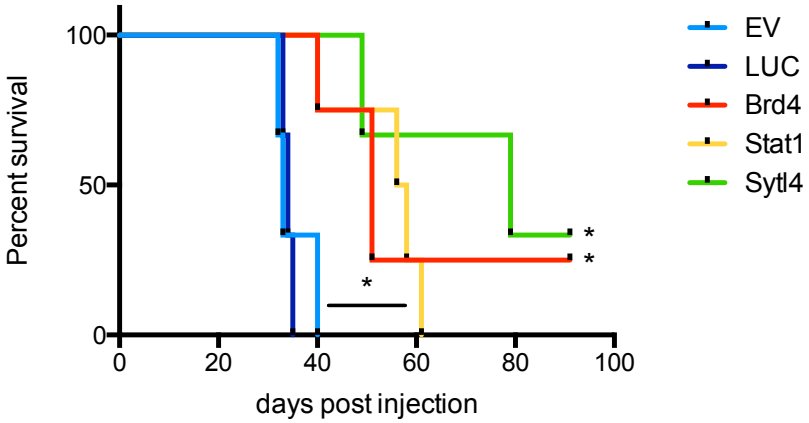

Figure 30. Brd4, Stat1 or Sytl4 down-regulation significantly delays MA9 leukemia burden. $2x10^5$ TagRFP+ sorted blasts were intravenously injected into sub-lethally irradiated mice and leukemia development was monitored (Mantel-Cox log-rank test *p<0.05).

Brd4 requirement for MA9 AML has already been extensively described and, in line with Zuber's work[166], we verified that blasts harboring Brd4-shRNA were depleted when leukemia appeared, by measuring the percentage of TagRFP+ blasts (CD45.1+) in PB, SPL and BM of leukemic mice (Figure 31 A). Interestingly, we obtained the same results for Stat1, suggesting that MA9 blasts expressing Stat1 shRNA were counter selected during leukemia growth *in vivo*. To test this hypothesis, we evaluated the mRNA level of Brd4 and Stat1 respectively, in leukemic cells collected when mice were sacrificed. As expected, outgrown leukemic blasts lost Brd4 or Stat1 silencing, with expression levels that were

comparable to controls (Figure 31 B-C). Collectively, these data demonstrated that Stat1, likewise Brd4, is critical for MA9 growth *in vivo*.

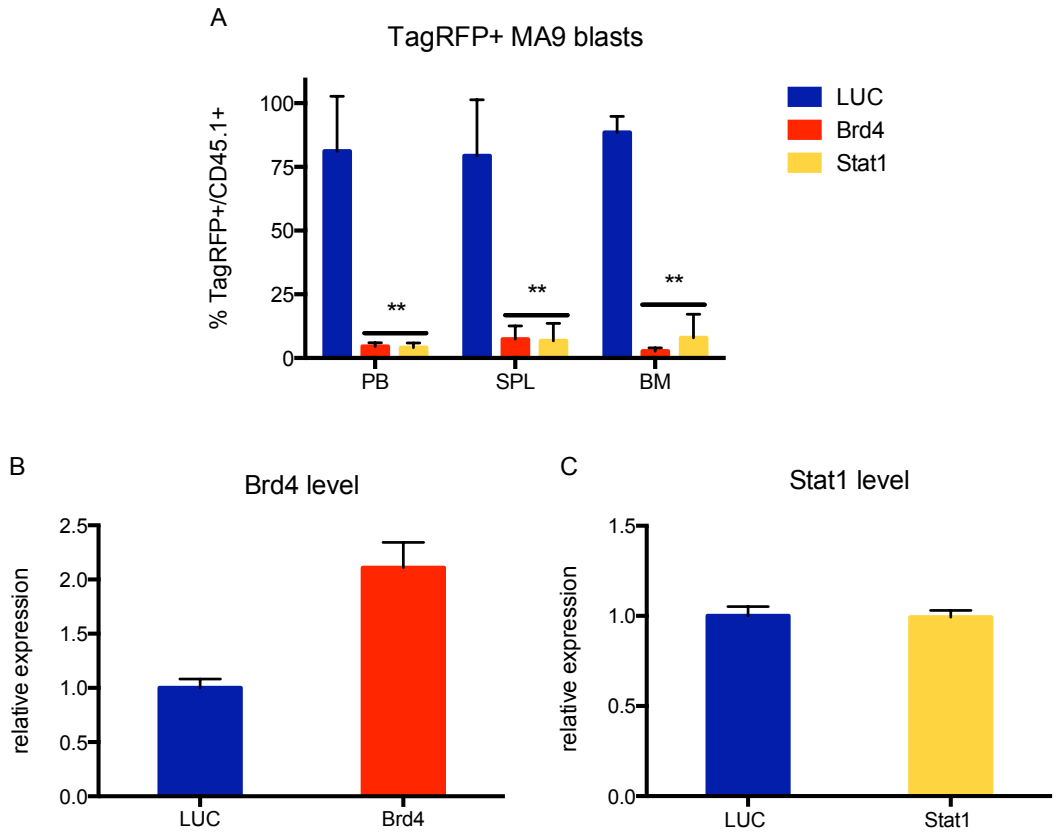

Figure 31. MA9 blasts harboring Brd4 or Stat1 shRNA are counter selected during leukemia growth *in vivo*. Mice injected with MA9 blasts infected with Brd4 or Stat1 shRNA eventually died of leukemia. **A**. Evaluation of TagRFP+ cells in leukemic peripheral blood (PB), spleen (SPL) and bone marrow (BM) of mice injected with blasts infected with LUC, Brd4 or Stat1 shRNAs. **B**. qPCR analysis of Brd4 mRNA level in AML deriving from blasts infected with Brd4 shRNA, compared to blasts infected with LUC shRNA. **C**. qPCR analysis of Stat1 mRNA level in AML deriving from blasts infected with Stat1 shRNA, compared to blasts infected with LUC shRNA (**$p<0.05$).

The most strikingly delay in leukemia latency was observed in mice transplanted with blasts harboring Sytl4-shRNA ($p=0.02$, Mantel-Cox log-rank test). The Sytl4 (or Slp4) gene encodes for the Synaptogamin like protein 4, also known as Granuphilin-A, and it interacts with Rab27a GTPase, and important regulator of exosome release[212]. Exosomes are small vesicles produced by cells in physiological conditions, which recently gained attention for their role in inter-tumoral trafficking of proteins, RNA and DNA. Therefore, we are

investigating this pathway by treating leukemic mice with GW4869, a sphingomyelinase (n-SMase) inhibitor which prevent exosome biogenesis (see Future plans).

We then started the validation of further 4 candidate genes identified as depleted during the screening *in vivo*: Socs2, Gfi1, Tie1 and Hoxa5. MA9 blasts were transduced and TagRFP+ cells sorted 72h after infection, as in the previous experiments. 500 sorted blasts were plated in MC cultures and colonies counted after one week. MA9 cells infected with shRNA targeting Gfi1, Tie1 or Hoxa5 were significantly less clonogenic, as compared to EV control, while blasts infected with Socs2 shRNA were not significantly different in clonogenic capacity (Figure 32 A). During subsequent passages in MC, blasts infected with Gfi1 shRNA were significantly counter selected, as shown in Figure 32 B by the percentage of TagRFP+ cells.

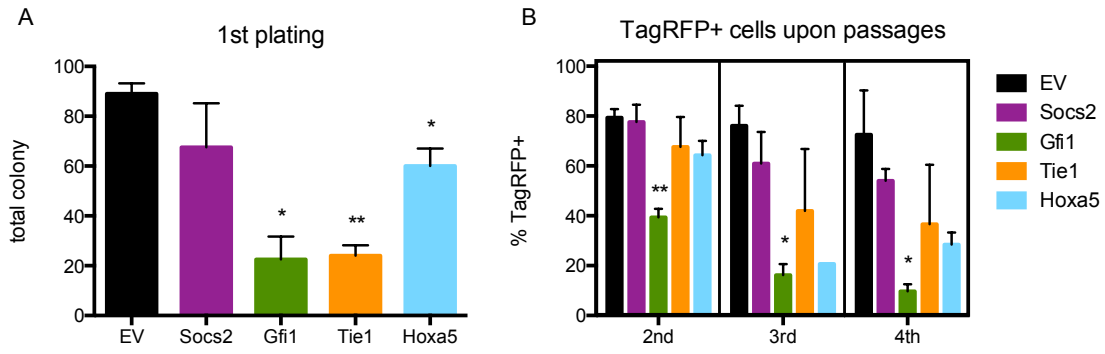

Figure 32. MA9 blasts infected with Gfi1, Tie1 or Hoxa5 shRNA have decreased clonogenic potential in MC. A. 500 MA9 TagRFP+ blasts infected with Socs2, Gfi1, Tie1 or Hoxa5 shRNAs or EV were cultured in methylcellulose medium. The number of colonies scored after one week is shown. **B**. 500 MA9 blasts from 1st plating in methylcellulose have been serially re-plated and percentage of TagRFP+ cells was scored at each passage by FACS (*p<0.05, **p<0.01).

However, RT-qPCR analysis at sorting time (i.e. 72h after transduction) revealed that only blasts infected with Socs2 shRNA showed a clear down-regulation of the target gene, while for Gfi1, Tie1 and Hoxa5, silencing was not effective at this time point (Figure 33).

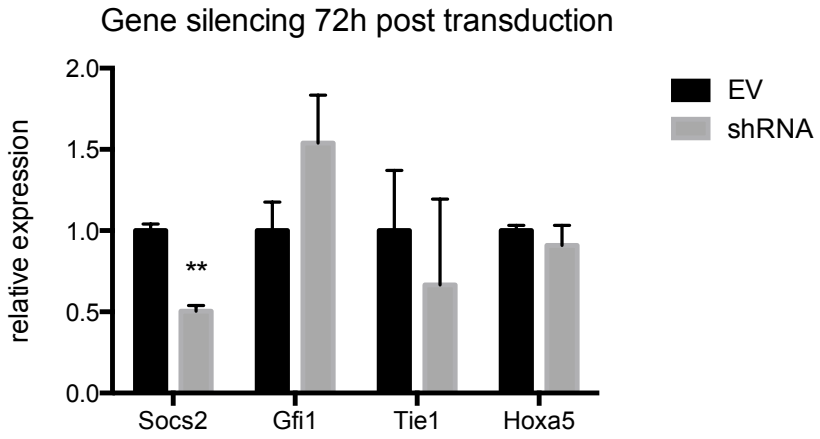

Figure 33. Socs2 was efficiently down-regulated 72 hours after transduction. qPCR analysis of MA9 blasts infected with shRNAs targeting Socs2, Gfi1, Tie1 or Hoxa5 compared to MA9 blasts infected with empty vector (EV), 72h post transduction (**p<0.05).

Given these not conclusive results, we are repeating the experiments with different hairpins. Since three days of culture might not be sufficient to obtain a shRNAs effect on target genes, we analyzed gene silencing 5 days after transduction in TagRFP+ sorted cells. As shown in Figure 34, we were able to down-regulate Gfi1 and Tie1 compared to control, therefore we will proceed in performing both in vitro and vivo experiments using these two shRNAs.

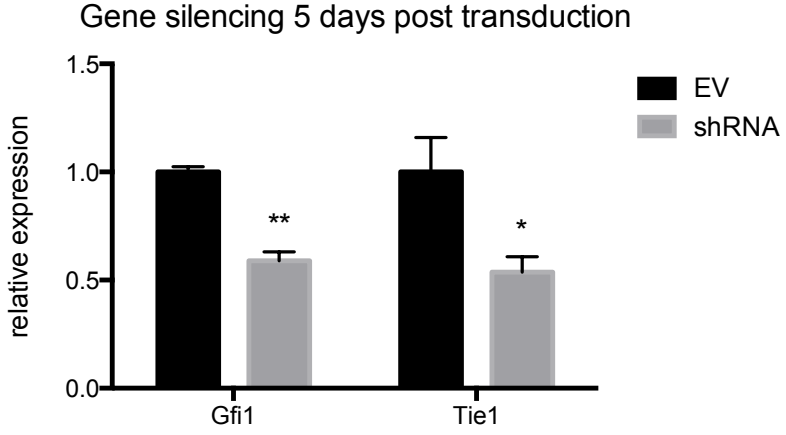

Figure 34. Gfi1 and Tie1 were down-regulated 5 days after transduction. qPCR analysis of MA9 blasts infected with shRNAs targeting Gfi1 or Tie1 compared to MA9 blasts infected with empty vector (EV) (*p<0.05, **p<0.01).

8.14 Future plans

We have shown that quiescence genes that are up-regulated by relevant AML-associated oncogene in pre-leukemic cells are critical for the maintenance of leukemia growth. Our next goal is to demonstrate that the anti-leukemic effect of the selected quiescence genes is indeed due to their ability to regulate LSCs quiescence. Except for Gfi1, the role in quiescence of HSCs for the other candidate genes has not yet established. Therefore, we will investigate the effect of their silencing in WT and pre-leukemic HSCs by performing cell cycle analysis on stem cell compartment isolated from recipient mice after long term transplantation. Since our hypothesis arose from the biological effect of NPMc+ expression in LT-HSCs, we will especially study the impact of gene silencing in this pre-leukemic context.

Moreover, we will address if down-regulating our targets we will be able to score a diminished quiescence in the LSCs compartment of MLL-AF9 leukemia, where their immunophenotype has been extensively characterized. Unfortunately, the identification, by surface markers, of LSCs in murine NPMc+ and PML-RARα leukemia is not well established. Finally, in order to obtain indications regarding a possible therapeutic strategy, we will treat leukemic mice with combinations of chemotherapeutic agents and specific drugs targeting the quiescence genes/pathways identified in the screening.

9. Discussion

One of the major obstacles in AML eradication is the genomic and biological heterogeneity of the tumor. AML genomic complexity is a well described phenomenon characterized by the presence of multiple subclones within the leukemic bulk population[48,49]. This intra-tumoral heterogeneity contributes to therapy failure and leukemia relapse, the major cause of death in AML patients. At biological level, AML are hierarchically organized and only LSCs, a minor population of the leukemic cells, are able to sustain tumor growth. Traditional chemo- and radiotherapies, which only target actively cycling cells, have limited effects on LSCs, mainly due to their quiescent state[85]. Quiescence is an intrinsic property of HSCs and it closely linked to long-term self-renewal capacity, necessary for the maintenance of blood compartment throughout the whole lifetime of an individual. Similarly, LSCs are enriched in quiescent cells with the ability to self-renew and propagate the disease. Therefore, the development of therapeutic strategies aimed to target quiescent LSCs may have a profound impact on leukemia eradication.

In this work, we demonstrated that two non-related leukemic oncogenes, namely NPMc+ and PML-RARα, are able to induce a common transcriptional program in HSCs. Specifically, a gene signature peculiar of quiescent stem cells is further enforced in the pre-leukemic HSCs expression profiles. Strikingly, a set of genes, already described to have a functional role in stem cells quiescence maintenance, was found to be up-regulated by NPMc+ and PML-RARα during the pre-leukemic phase. In the same way, preliminary data on AML1-ETO pre-leukemic HSCs expression profile showed an enrichment in the same quiescent HSCs signature (not shown), indicating that the perturbations induced by different leukemic oncogenes converge on similar pathways. Consistently, a series of data obtained expressing different oncogenes in HSCs suggested that a tight regulation of cell cycle is functionally relevant for leukemia development[107]. Moreover, $Nras^{G12D}$, a known leukemia initiating oncogene, has been described to exert a bimodal on HSCs[213]. $Nras^{G12D}$ expression is able to increase self-renewal potential in one subset of HSCs, maintaining their quiescent

state, while it increases the proliferation rate in another subset of HSCs. The author proposed a model where short lived but rapidly dividing $Nras^{G12D}$ HSCs presumably outcompete wild-type HSCs and are replenished over time by the quiescent $Nras^{G12D}$ HSCs, slowly recruited into cell cycle.

Our data suggested that the ability to enforce quiescence in HSCs is indeed a novel common feature of leukemia initiating oncogenes. Therefore, deeper characterization of the molecular mechanisms underlying oncogene-induced quiescence will shed light on functional relevant pathways in AML development, maintenance and response to therapies.

To assess if the quiescence-related genes induced in pre-leukemic phase are also relevant for leukemia maintenance, we investigated their role with *in vivo* shRNA screening. Unfortunately, we found an unexpectedly marked expansion of few clones during NPMc+ and PML-RARα leukemia growth, confirmed by lentiviral clonal tracking experiments. Clonal expansion reflects the intrinsic biological heterogeneity within the LSCs compartment in terms of growth potential. So far, few studies have used cellular barcodes to study AML clonal dynamics and, consequently, the relationship between LSCs functional heterogeneity and leukemia clonal composition remains not clear. In a leukemia mouse model, tracking of barcoded cells, from cancer initiation to leukemia development, showed that clonal composition fluctuates during leukemogenesis eventually resulting in oligoclonal leukemia[214]. Transplantation of human AML samples showed that xenografts are predominantly composed of few single genetically-defined subclones which are not the predominant in the primary sample[80]. On the contrary, cellular barcoding of human B-ALL samples showed high polyclonal composition upon xenotransplantation in NSG mice. Interestingly, clones are asymmetrically distributed in mouse niches highlighting the influence of the environmental selective pressure[155]. The major technical issue in clonal tracking experiments is cell transplantation in recipient mice which introduces by itself a strong selective proliferation pressure.

The strong clonal expansion we observed in NPMc+ and PML-RARα leukemia prevented any meaningful analysis of the screenings we performed. Therefore, we evaluated other AML models and we chose the MLL-AF9 driven leukemia, which showed a less prominent clonal growth *in vivo*. Interestingly, MLL-AF9 expression in normal stem cell compartment was able to up-regulate a set of quiescence genes, similarly to the other leukemia initiating oncogenes we examined. For these reasons, we performed the *in vivo* shRNA screening in this model, in parallel to a control screening with a non-targeting library composed by the same number of DNA barcodes. We implemented a novel strategy to analyze shRNA screening results based on the barcodes distribution of the non-targeting library *in vivo*. This library turned to be fundamental to account for the diverse proliferation capacity of LSCs upon *in vivo* transplantation. Two different bioinformatics pipelines allowed the identification of 10 significantly depleted genes, common to both the analyses. Among the targets we identified, we found Brd4, already described to play a critical role in MLL-AF9 leukemia maintenance[166]. Polr2b and Rpl30, two essential genes we introduced in the screening as positive controls, scored as depleted in both analysis. Other depleted genes included Esr1, Gfi1, Hoxa5, Socs2, Stat1, Sytl4 and Tie1. In particular, Stat1 or Sytl4 down-regulation in MLL-AF9 blasts was sufficient to significantly delay leukemia onset in mice.

Stat1, a transcription factor member of the STAT proteins family, is specifically activated by interferons and is involved in cell survival and pathogen response. Upon phosphorylation mediated by receptor associated kinases and Janus kinases (JAKs), STATs proteins form homo- or heterodimers and translocate to the nucleus where they regulate transcription of target genes. Stat1 is mainly described as a negative regulator of quiescence since INFα treatment is sufficient to trigger HSCs G0 exit, in a Stat1-dependent manner[133]. Moreover, Stat1 depletion is sufficient to rescue the HSCs hyperproliferative caused by KO of Irgm1, a well known HSCs quiescence positive regulator[215]. On the other side, Stat1 can induce cell growth arrest by inducing the expression of p21 and p27 and down-regulating cMyc in U937 cells[216,217]. Interestingly, in hematological malignancies models, Stat1 has been described to

play a role in LSCs expansion[218], in immunological surveillance escaping[219] and in therapy resistance[220]. Moreover, the constitutive activation of STAT proteins has been demonstrated in leukemic cell lines and AML blasts[221,222], laying the basis for several clinical trials with JAK/STAT inhibitors.

Furthermore, we showed that down-regulation of Sytl4 in MLL-AF9 blasts is sufficient to decrease colony-forming efficiency *in vitro* and significantly delay leukemia growth *in vivo*. Sytl4 interacts with Rab27a, a key regulator of exosomes release. Impaired exosome maturation has already been linked to HSCs exit from quiescence. Indeed, HSCs release extracellular vesicles containing cytokines such as Angptl2 and TPO which are necessary to maintain HSCs activity through an autocrine signaling effect[223]. AML patients are characterized by high plasma levels of blasts-derived exosomes which decrease after induction therapy, mirroring blasts reduction in the bone marrow. Moreover, patients achieving long-term remission typically have exosomes plasma levels comparable to healthy individuals, suggesting that this parameter might have prognostic significance[224]. Vesicles content partially depends on the parent cell type and, in particular, AML exosomes are important for the crosstalk between leukemic cells and the environment by modulating immune response, blast survival, angiogenesis and resistance to therapies[225]. Strategies to specifically block exosome release in AML need further mechanistic studies to avoid indiscriminate interfering with physiological processes.

As discuss in the future plans, our main goal is now to link the effect of the identified quiescence genes on leukemia growth to a direct role in quiescence maintenance. To this aim we will study both the effect of gene silencing in normal and pre-leukemic LT-HSCs in homeostatic conditions, and the extent of LSCs G0 exit, wherever they are immunophenotypically defined. To overcome this limitation, we will exploit functional assays for the identification of LSCs such as the H2B-GFP lentiviral labelling tool[226]. Leukemic blasts transduced with lentiviral Tet-Off H2B-GFP vector can be treated with

doxycycline to dilute the GFP signal upon cell divisions. With this system we can functionally isolate slow cycling LSCs in the fraction of GFP label-retaining cells.

The following step will be to investigate if the mechanism is conserved in human samples. To this purpose, we can take advantage of the large cohort of AML xenografts already available in the lab. Ultimately, we hope to discover new therapeutic targets able to induce LSCs to proliferate becoming sensitive to standard chemotherapy strategies. Indeed, the exclusive target of LSCs may be insufficient to obtain effective leukemia cure because of the hyperproliferation of the bulk population. On the contrary, a biological meaningful combination of therapies directed against both LSCs and proliferating blasts, would allow to reach a definitive leukemia eradication.

10. References

1. Rieger, M. A. & Schroeder, T. Hematopoiesis. *Cold Spring Harb Perspect Biol* **4,** (2012).
2. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425,** 841–846 (2003).
3. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425,** 836–841 (2003).
4. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121,** 1109–1121 (2005).
5. Sipkins, D. A. *et al.* In vivo imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435,** 969–973 (2005).
6. Yoshihara, H. *et al.* Thrombopoietin/MPL signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic niche. *Cell Stem Cell* **208,** 685–697 (2007).
7. Ogawa, M. *et al.* Expression and function of c-kit in hemopoietic progenitor cells. *J. Exp. Med.* **174,** 63–71 (1991).
8. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25,** 977–988 (2006).
9. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135,** 1118–1129 (2008).
10. Cheshier, S. H., Morrison, S. J., Liao, X. & Weissman, I. L. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 3120–3125 (1999).
11. Cheung, T. H. & Rando, T. A. Molecular regulation of stem cell quiescence. *Nat. Rev. Mol. Cell Biol.* **14,** 329–340 (2013).
12. Mendelson, A. & Frenette, P. S. Hematopoietic stem cell niche maintenance during homeostasis and regeneration. *Nat. Med.* **20,** 833–846 (2014).
13. Pietras, E. M., Warr, M. R. & Passegué, E. Cell cycle regulation in hematopoietic stem cells. *J. Cell Biol.* **195,** 709–720 (2011).
14. Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nature Reviews Genetics* **9,** 115–128 (2008).
15. Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 14541–14546 (2001).
16. Morrison, S. J., Uchida, N. & Weissman, I. L. The Biology of Hematopoietic Stem Cells. *Annu. Rev. Cell Dev. Biol.* **11,** 35–71 (1995).
17. Notta, F. *et al.* Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. *Science* **333,** 218–221 (2011).
18. Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. & Mulligan, R. C. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J. Exp. Med.* **183,** 1797–1806 (1996).
19. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351,** aab2116 (2016).
20. Paul, F. *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163,** 1663–1677 (2015).
21. Cheng, T. Hematopoietic Stem Cell Quiescence Maintained by p21cip1/waf1. *Science* **287,** 1804–1808 (2000).
22. Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the

regulation of hematopoietic stem and progenitor cell fates. *J. Exp. Med.* **202,** 1599–1611 (2005).
23. Matsumoto, A. *et al.* p57 Is Required for Quiescence and Maintenance of Adult Hematopoietic Stem Cells. *Stem Cell* **9,** 262–271 (2011).
24. Scheicher, R. *et al.* CDK6 as a key regulator of hematopoietic and leukemic stem cell activation. *Blood* **125,** 90–101 (2015).
25. Laurenti, E. *et al.* CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells. *Cell Stem Cell* (2015). doi:10.1016/j.stem.2015.01.017
26. Tsai, F. Y. *et al.* An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371,** 221–226 (1994).
27. Tipping, A. J. *et al.* High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood* **113,** 2661–2672 (2009).
28. Hock, H. *et al.* Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature* **431,** 1002–1007 (2004).
29. Korthuis, P. M. *et al.* CITED2-mediated human hematopoietic stem cell maintenance is critical for acute myeloid leukemia. *Leukemia* (2014). doi:10.1038/leu.2014.259
30. Qian, H. *et al.* Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell Stem Cell* **1,** 671–684 (2007).
31. Larsson, J. *et al.* TGF-beta signaling-deficient hematopoietic stem cells have normal self-renewal and regenerative ability in vivo despite increased proliferative capacity in vitro. *Blood* **102,** 3129–3135 (2003).
32. Brenet, F., Kermani, P., Spektor, R., Rafii, S. & Scandura, J. M. TGFβ restores hematopoietic homeostasis after myelosuppressive chemotherapy. *Journal of Experimental Medicine* **210,** 623–639 (2013).
33. Zhao, M. *et al.* Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nat. Med.* **20,** 1321–1326 (2014).
34. Greenbaum, A. *et al.* CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* **495,** 227–230 (2013).
35. Chandel, N. S., Jasper, H., Ho, T. T. & Passegué, E. Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. *Nat. Cell Biol.* **18,** 823–832 (2016).
36. Takubo, K. *et al.* Regulation of the HIF-1a Level Is Essential for Hematopoietic Stem Cells. *Stem Cell* **7,** 391–402 (2010).
37. Takubo, K. *et al.* Regulation of Glycolysis by Pdk Functions as a Metabolic Checkpoint for Cell Cycle Quiescence in Hematopoietic Stem Cells. *Stem Cell* **12,** 49–61 (2013).
38. Cabezas-Wallscheid, N. *et al.* Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. *Cell* (2017). doi:10.1016/j.cell.2017.04.018
39. Agathocleous, M. *et al.* Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature* **549,** 476–481 (2017).
40. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66,** 7–30 (2016).
41. Döhner, H., Weisdorf, D. J. & Bloomfield, C. D. Acute Myeloid Leukemia. *N Engl J Med* **373,** 1136–1152 (2015).
42. De Kouchkovsky, I. & Abdul-Hay, M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood Cancer J* **6,** e441 (2016).
43. Levis, M. Midostaurin approved for FLT3-mutated AML. *Blood* **129,** 3403–3406 (2017).

44. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127,** 2391–2405 (2016).
45. Estey, E. H. Acute myeloid leukemia: 2014 update on risk-stratification and management. *Am J Hematol* **89,** 1063–1081 (2014).
46. Patel, J. P. *et al.* Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *N Engl J Med* **366,** 1079–1089 (2012).
47. Riva, L. *et al.* Acute promyelocytic leukemias share cooperative mutations with other myeloid-leukemia subgroups. *Blood Cancer J* **3,** e147 (2013).
48. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150,** 264–278 (2012).
49. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–509 (2013).
50. Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med* **352,** 254–266 (2005).
51. Heath, E. M. *et al.* Biological and clinical consequences of NPM1 mutations in AML. *Leukemia* **31,** 798–807 (2017).
52. Sportoletti, P. *et al.* Mouse models of NPM1-mutated acute myeloid leukemia: biological and clinical implications. *Leukemia* (2014). doi:10.1038/leu.2014.257
53. Metzeler, K. H. *et al.* Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. *Blood* **128,** 686–698 (2016).
54. The Cancer Genome Atlas Research Network. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N Engl J Med* **368,** 2059–2074 (2013).
55. Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* **374,** 2209–2221 (2016).
56. Alcalay, M. *et al.* Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* **106,** 899–902 (2005).
57. Kohlmann, A. *et al.* Gene expression profiling in AML with normal karyotype can predict mutations for molecular markers and allows novel insights into perturbed biological pathways. *Leukemia* **24,** 1216–1220 (2010).
58. Longo, L. *et al.* Rearrangements and aberrant expression of the retinoic acid receptor alpha gene in acute promyelocytic leukemias. *J. Exp. Med.* **172,** 1571–1575 (1990).
59. Brown, D. *et al.* A PMLRARalpha transgene initiates murine acute promyelocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 2551–2556 (1997).
60. Grisolano, J. L., Wesselschmidt, R. L., Pelicci, P. G. & Ley, T. J. Altered myeloid development and acute leukemia in transgenic mice expressing PML-RAR alpha under control of cathepsin G regulatory sequences. *Blood* **89,** 376–387 (1997).
61. Westervelt, P. High-penetrance mouse model of acute promyelocytic leukemia with very low levels of PML-RAR expression. *Blood* **102,** 1857–1865 (2003).
62. Madan, V. *et al.* Comprehensive mutational analysis of primary and relapse acute promyelocytic leukemia. *Leukemia* **30,** 1672–1681 (2016).
63. Greif, P. A. *et al.* Somatic mutations in acute promyelocytic leukemia (APL) identified by exome sequencing. *Leukemia* **25,** 1519–1522 (2011).
64. Wartman, L. D. *et al.* Sequencing a mouse acute promyelocytic leukemia genome reveals genetic events relevant for disease progression. *J. Clin. Invest.* **121,** 1445–1455 (2011).
65. Ronchini, C. *et al.* PML-RARA-associated cooperating mutations belong toa

transcriptional network that is deregulated in myeloidleukemias. **31,** 1975–1986 (2017).
66. Nasr, R. *et al.* Eradication of acute promyelocytic leukemia-initiating cells through PML-RARA degradation. *Nat. Med.* **14,** 1333–1342 (2008).
67. Coombs, C. C., Tavakkoli, M. & Tallman, M. S. Acute promyelocytic leukemia: where did we start, where are we now, and the future. *Blood Cancer J* **5,** e304 (2015).
68. Krivtsov, A. V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442,** 818–822 (2006).
69. Somervaille, T. C. P. & Cleary, M. L. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell* **10,** 257–268 (2006).
70. Zeisig, B. B. *et al.* Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Molecular and Cellular Biology* **24,** 617–628 (2004).
71. Li, Z. *et al.* PBX3 is an important cofactor of HOXA9 in leukemogenesis. *Blood* **121,** 1422–1431 (2013).
72. Placke, T. *et al.* Requirement for CDK6 in MLL-rearranged acute myeloid leukemia. *Blood* **124,** 13–23 (2014).
73. Sroczynska, P. *et al.* shRNA screening identifies JMJD1C as being required for leukemia maintenance. *Blood* **123,** 1870–1882 (2014).
74. Schwieger, M. *et al.* Homing and invasiveness of MLL/ENL leukemic cells is regulated by MEF2C. *Blood* **114,** 2476–2488 (2009).
75. Wang, Q.-F. *et al.* MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. *Blood* **117,** 6895–6905 (2011).
76. Zuber, J. *et al.* An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. *Genes & Development* **25,** 1628–1640 (2011).
77. Arai, S. *et al.* Evi-1 is a transcriptional target of mixed-lineage leukemia oncoproteins in hematopoietic stem cells. *Blood* **117,** 6304–6314 (2011).
78. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367,** 645–648 (1994).
79. Hope, K. J., Jin, L. & Dick, J. E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat. Immunol.* **5,** 738–743 (2004).
80. Klco, J. M. *et al.* Functional Heterogeneity of Genetically Defined Subclones in Acute Myeloid Leukemia
. *Cancer Cell* (2014). doi:10.1016/j.ccr.2014.01.031
81. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14,** 275–291 (2014).
82. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414,** 105–111 (2001).
83. Beck, B. & Blanpain, C. Unravelling cancer stem cell potential. *Nat Rev Cancer* **13,** 727–738 (2013).
84. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23,** 1124–1134 (2017).
85. Wiseman, D. H., Greystoke, B. F. & Somervaille, T. C. P. The variety of leukemic stem cells in myeloid malignancy. *Oncogene* **33,** 3091–3098 (2014).
86. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3,** 730–737 (1997).
87. Taussig, D. C. *et al.* Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction.

Blood **115,** 1976–1984 (2010).
88. Shlush, L. I. *et al.* Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* (2017). doi:10.1038/nature22993
89. Smith, L. J. *et al.* Lineage infidelity in acute leukemia. *Blood* **61,** 1138–1145 (1983).
90. Visvader, J. E. Cells of origin in cancer. *Nature* **469,** 314–322 (2011).
91. Turhan, A. G. *et al.* Highly purified primitive hematopoietic stem cells are PML-RARA negative and generate nonclonal progenitors in acute promyelocytic leukemia. *Blood* **85,** 2154–2161 (1995).
92. Reinisch, A. *et al.* A humanized bone marrow ossicle xenotransplantation model enables improved engraftment of healthy and leukemic human hematopoietic cells. *Nat. Med.* (2016). doi:10.1038/nm.4103
93. Cozzio, A. *et al.* Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes & Development* **17,** 3029–3035 (2003).
94. Huntly, B. J. P. *et al.* MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **6,** 587–596 (2004).
95. Wojiski, S. *et al.* PML-RARalpha initiates leukemia by conferring properties of self-renewal to committed promyelocytic progenitors. *Leukemia* **23,** 1462–1471 (2009).
96. George, J. *et al.* Leukaemia cell of origin identified by chromatin landscape of bulk tumour cells. *Nat Commun* **7,** 12166 (2016).
97. Wright, G. *et al.* A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9991–9996 (2003).
98. Kridel, R. *et al.* Cell of origin of transformed follicular lymphoma. *Blood* **126,** 2118–2127 (2015).
99. van Rhenen, A. *et al.* High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. *Clin. Cancer Res.* **11,** 6520–6527 (2005).
100. Pearce, D. J., Taussig, D., Zibara, K., Smith, L. L. & Ridler, C. M. AML engraftment in the NOD/SCID assay reflects the outcome of AML: implications for our understanding of the heterogeneity of AML | Blood Journal. *Blood* (2006).
101. Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17,** 1086–1093 (2011).
102. Jung, N., Dai, B., Gentles, A. J., Majeti, R. & Feinberg, A. P. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nat Commun* **6,** 8489 (2015).
103. Ng, S. W. K. *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540,** 433–437 (2016).
104. Ishikawa, F. *et al.* Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nat Biotechnol* **25,** 1315–1321 (2007).
105. Saito, Y. *et al.* Induction of cell cycle entry eliminates human leukemia stem cells in a mouse model of AML. *Nat Biotechnol* **28,** 275–280 (2010).
106. Chan, S. M. *et al.* Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia. *Nat. Med.* (2015). doi:10.1038/nm.3788
107. Viale, A. *et al.* Cell-cycle restriction limits DNA damageand maintains self-renewal of leukaemiastem cells. *Nature* **457,** 51–56 (2008).
108. Lagadinou, E. D. *et al.* BCL-2 inhibition targets oxidative phosphorylation

and selectively eradicates quiescent human leukemia stem cells. *Cell Stem Cell* **12,** 329–341 (2013).
109. Shlush, L. I. *et al.* Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. *Blood* **120,** 603–612 (2012).
110. Lechman, E. R. *et al.* miR-126 Regulates Distinct Self-Renewal Outcomes in Normal and Malignant Hematopoietic Stem Cells. *Cancer Cell* (2016). doi:10.1016/j.ccell.2015.12.011
111. Peiper, S. C., Ashmun, R. A. & Look, A. T. Molecular cloning, expression, and chromosomal localization of a human gene encoding the CD33 myeloid differentiation antigen. *Blood* **72,** 314–321 (1988).
112. Breccia, M. *et al.* Sustained molecular remission after low dose gemtuzumab-ozogamicin in elderly patients with advanced acute promyelocytic leukemia. *Haematologica* **92,** 1273–1274 (2007).
113. Lo-Coco, F. *et al.* Gemtuzumab ozogamicin (Mylotarg) as a single agent for molecularly relapsed acute promyelocytic leukemia. *Blood* **104,** 1995–1999 (2004).
114. Walter, R. B., Appelbaum, F. R., Estey, E. H. & Bernstein, I. D. Acute myeloid leukemia stem cells and CD33-targeted immunotherapy. *Blood* **119,** 6198–6208 (2012).
115. Stein, E. M. *et al.* Interim Analysis of a Phase 1 Trial of SGN-CD33A in Patients with CD33-Positive Acute Myeloid Leukemia (AML). *Blood* **124,** 623–623 (2014).
116. Löwenberg, B. *et al.* Effect of priming with granulocyte colony-stimulating factor on the outcome of chemotherapy for acute myeloid leukemia. *N Engl J Med* **349,** 743–752 (2003).
117. Uy, G. L. *et al.* A phase 1/2 study of chemosensitization with the CXCR4 antagonist plerixafor in relapsed or refractory acute myeloid leukemia. *Blood* **119,** 3917–3924 (2012).
118. Takeishi, S. & Nakayama, K. I. To wake up cancer stem cells, or to let them sleep, that is the question. *Cancer Science* **107,** 875–881 (2016).
119. Mazza, M. & Pelicci, P. G. Is PML a Tumor Suppressor? *Front Oncol* **3,** 174 (2013).
120. Ablain, J., Nasr, R., Bazarbachi, A. & de The, H. The Drug-Induced Degradation of Oncoproteins: An Unexpected Achilles' Heel of Cancer Cells? *Cancer Discovery* **1,** 117–127 (2011).
121. Ito, K. *et al.* PML targeting eradicates quiescent leukaemia-initiating cells. *Nature* **453,** 1072–1078 (2008).
122. Ito, K. *et al.* A PML–PPAR-δ pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance. *Nat. Med.* **18,** 1350–1358 (2012).
123. Zhang, J. *et al.* p27kip1 maintains a subset of leukemia stem cells in the quiescent state in murine MLL-leukemia. *Mol Oncol* **7,** 1069–1082 (2013).
124. Prost, S. *et al.* Erosion of the chronic myeloid leukaemia stem cell pool by PPARγ agonists. *Nature* (2015). doi:10.1038/nature15248
125. Rousselot, P. *et al.* Pioglitazone together with imatinib in chronic myeloid leukemia: A proof of concept study. *Cancer* **123,** 1791–1799 (2017).
126. Takeishi, S. *et al.* Ablation of Fbxw7 eliminates leukemia-initiating cells by preventing quiescence. *Cancer Cell* **23,** 347–361 (2013).
127. Yilmaz, Ö. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441,** 475–482 (2006).
128. Fang, Y. *et al.* Rictor has a pivotal role in maintaining quiescence as well as stemness of leukemia stem cells in MLL-driven leukemia. *Leukemia* **31,** 414–422 (2017).

129. Zeng, Z. *et al.* Targeting the leukemia microenvironment by CXCR4 inhibition overcomes resistance to kinase inhibitors and chemotherapy in AML. *Blood* **113,** 6215–6224 (2009).
130. Nervi, B. *et al.* Chemosensitization of acute myeloid leukemia (AML) following mobilization by the CXCR4 antagonist AMD3100. *Blood* **113,** 6206–6214 (2009).
131. Weisberg, E. *et al.* Inhibition of CXCR4 in CML cells disrupts their interaction with the bone marrow microenvironment and sensitizes them to nilotinib. *Leukemia* **26,** 985–990 (2012).
132. Talpaz, M., Mercer, J. & Hehlmann, R. The interferon-alpha revival in CML. *Ann Hematol* **94 Suppl 2,** S195–207 (2015).
133. Essers, M. A. G. *et al.* IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature* **458,** 904–908 (2009).
134. Sato, T. *et al.* Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon-dependent exhaustion. *Nat. Med.* **15,** 696–700 (2009).
135. Pietras, E. M. *et al.* Re-entry into quiescence protects hematopoietic stem cells from the killing effect of chronic exposure to type I interferons. *Journal of Experimental Medicine* **211,** 245–262 (2014).
136. Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118,** 149–161 (2004).
137. Ichihara, E., Kaneda, K., Saito, Y., Yamakawa, N. & Morishita, K. Angiopoietin1 contributes to the maintenance of cell quiescence in EVI1(high) leukemia cells. *Biochem. Biophys. Res. Commun.* **416,** 239–245 (2011).
138. Khalaj, M. *et al.* miR-99 regulates normal and malignant hematopoietic stem cell self-renewal. *Journal of Experimental Medicine* (2017). doi:10.1084/jem.20161595
139. Corces-Zimmerman, M. R., Hong, W.-J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proceedings of the National Academy of Sciences* **111,** 2548–2553 (2014).
140. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 1–13 (2014). doi:10.1038/nature13038
141. Garg, M. *et al.* Profiling of somatic mutations in acute myeloid leukemia with FLT3-ITD at diagnosis and relapse. *Blood* **126,** 2491–2501 (2015).
142. Sun, Q.-Y. *et al.* Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). *Leukemia* **31,** 1–10 (2017).
143. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371,** 2477–2487 (2014).
144. Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat Commun* **7,** 12484 (2016).
145. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371,** 2488–2498 (2014).
146. Thomas, D. & Majeti, R. Biology and relevance of human acute myeloid leukemia stem cells. *Blood* (2017). doi:10.1182/blood-2016-10-696054
147. Goyama, S., Wunderlich, M. & Mulloy, J. C. Xenograft models for normal and malignant stem cells. *Blood* (2015). doi:10.1182/blood-2014-11-570218
148. Sarry, J.-E. *et al.* Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J. Clin. Invest.* **121,** 384–395 (2011).
149. Wunderlich, M. *et al.* AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and

IL-3. *Leukemia* **24,** 1785–1788 (2010).
150. Busch, K. *et al.* Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* **518,** 542–546 (2015).
151. Sun, J. *et al.* Clonal dynamics of native haematopoiesis. *Nature* **514,** 322–327 (2014).
152. Gerrits, A. *et al.* Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115,** 2610–2618 (2010).
153. Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat Biotechnol* **29,** 928–933 (2011).
154. Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M. & de Haan, G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *Journal of Experimental Medicine* **208,** 2691–2703 (2011).
155. Belderbos, M. E. *et al.* Clonal selection and asymmetric distribution of human leukemia in murine xenografts revealed by cellular barcoding. *Blood* **129,** 3210–3220 (2017).
156. Bhang, H.-E. C. *et al.* Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21,** 440–448 (2015).
157. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806–811 (1998).
158. Brummelkamp, T. R. A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science* **296,** 550–553 (2002).
159. Karlsson, C., Larsson, J. & Baudet, A. Forward RNAi screens in human stem cells. *Methods Mol. Biol.* **650,** 29–43 (2010).
160. Mohr, S. E., Smith, J. A., Shamu, C. E., Neumüller, R. A. & Perrimon, N. RNAi screening comes of age: improved techniques and complementary approaches. *Nat. Rev. Mol. Cell Biol.* **15,** 591–600 (2014).
161. Hope, K. J. *et al.* An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. *Cell Stem Cell* **7,** 101–113 (2010).
162. Wang, J. *et al.* A differentiation checkpoint limits hematopoietic stem cell self-renewal in response to DNA damage. *Cell* **148,** 1001–1014 (2012).
163. Baudet, A. *et al.* RNAi screen identifies MAPK14 as a druggable suppressor of human hematopoietic stem cell expansion. *Blood* **119,** 6255–6258 (2012).
164. Cellot, S. *et al.* RNAi screen identifies Jarid1b as a major regulator of mouse HSC activity. *Blood* **122,** 1545–1555 (2013).
165. Bric, A. *et al.* Functional identification of tumor-suppressor genes through an in vivo RNA interference screen in a mouse lymphoma model. *Cancer Cell* **16,** 324–335 (2009).
166. Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478,** 524–528 (2012).
167. Miller, P. G. *et al.* In Vivo RNAi Screening Identifies a Leukemia-Specific Dependence on Integrin Beta 3 Signaling. *Cancer Cell* 1–14 (2013). doi:10.1016/j.ccr.2013.05.004
168. Järås, M. *et al.* Csnk1a1 inhibition has p53-dependent therapeutic efficacy in acute myeloid leukemia. *jem.rupress.org*
169. Gargiulo, G. *et al.* In vivo RNAi screen for BMI1 targets identifies TGF-β/BMP-ER stress pathways as key regulators of neural- and malignant glioma-stem cell homeostasis. *Cancer Cell* **23,** 660–676 (2013).
170. Wuestefeld, T. *et al.* A Direct in vivo RNAi screen identifies MKK4 as a key regulator of liver regeneration. *Cell* **153,** 389–401 (2013).
171. Mallardo, M. *et al.* NPMc+ and FLT3_ITD mutations cooperate in inducing acute leukaemia in a novel mouse model. *Leukemia* **27,** 2248–2251 (2013).

172. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1,** 4 (2001).
173. Esposito, M. T. *et al.* Synthetic lethal targeting of oncogenic transcription factors in acute leukemia by PARP inhibitors. *Nat. Med.* **21,** 1481–1490 (2015).
174. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 15545–15550 (2005).
175. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10,** R25 (2009).
176. Dai, Z. *et al.* edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Res* **3,** 95 (2014).
177. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347,** 70–78 (2009).
178. Jaatinen, T. *et al.* Global gene expression profile of human cord blood-derived CD133+ cells. *STEM CELLS* **24,** 631–641 (2006).
179. Ivanova, N. B. *et al.* A stem cell molecular signature. *Science* **298,** 601–604 (2002).
180. Verhaak, R. G. W. Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* **106,** 3747–3754 (2005).
181. Yamazaki, S. *et al.* TGF-beta as a candidate bone marrow niche signal to induce hematopoietic stem cell hibernation. *Blood* **113,** 1250–1256 (2009).
182. Kocabas, F. *et al.* Meis1 regulates the metabolic phenotype and oxidant defense of hematopoietic stem cells. *Blood* **120,** 4963–4972 (2012).
183. Yuan, Y., Shen, H., Franklin, D. S., Scadden, D. T. & Cheng, T. In vivo self-renewing divisions of haematopoietic stem cells are increased in the absence of the early G1-phase inhibitor, p18INK4C. *Nat. Cell Biol.* **6,** 436–442 (2004).
184. Lin, K. K. *et al.* CD81 Is Essential for the Re-entry of Hematopoietic Stem Cells to Quiescence following Stress-Induced Proliferation Via Deactivation of the Akt Pathway. *PLoS Biol.* **9,** e1001148 (2011).
185. Fischer, M. A., Moreno-Miralles, I., Hunt, A., Chyla, B. J. & Hiebert, S. W. Myeloid translocation gene 16 is required for maintenance of haematopoietic stem cell quiescence. *EMBO J.* **31,** 1494–1505 (2012).
186. Min, I. M. *et al.* The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* **2,** 380–391 (2008).
187. Ficara, F., Murphy, M. J., Lin, M. & Cleary, M. L. Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell Stem Cell* **2,** 484–496 (2008).
188. Bruedigam, C. *et al.* Telomerase inhibition effectively targets mouse and human AML stem cells and delays relapse following chemotherapy. *Cell Stem Cell* **15,** 775–790 (2014).
189. Thorén, L. A. *et al.* Kit regulates maintenance of quiescent hematopoietic stem cells. *J. Immunol.* **180,** 2045–2053 (2008).
190. Yoshida, T. *et al.* The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes & Development* **22,** 1174–1189 (2008).
191. Ye, M. *et al.* C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence. *Nat. Cell Biol.* **15,** 385–394 (2013).

192. Graham, S. M., Vass, J. K., Holyoake, T. L. & Graham, G. J. Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal and chronic myeloid leukemia sources. *STEM CELLS* **25,** 3111–3120 (2007).
193. Venezia, T. A. *et al.* Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol.* **2,** e301 (2004).
194. Martens, J. H. A. *et al.* PML-RARα/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia. *Cancer Cell* **17,** 173–185 (2010).
195. Gal, H. *et al.* Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells. *Leukemia* **20,** 2147–2154 (2006).
196. Cheng, H. *et al.* Leukemic marrow infiltration reveals a novel role for Egr3 as a potent inhibitor of normal hematopoietic stem cell proliferation. *Blood* **126,** 1302–1313 (2015).
197. Yu, X. *et al.* HES1 inhibits cycling of hematopoietic progenitor cells via DNA binding. *STEM CELLS* **24,** 876–888 (2006).
198. Will, B. *et al.* Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. *Nat. Immunol.* **14,** 437–445 (2013).
199. Santaguida, M. *et al.* JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. *Cancer Cell* **15,** 341–352 (2009).
200. Sugimura, R. *et al.* Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell* **150,** 351–365 (2012).
201. Mortensen, M. *et al.* The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. *Journal of Experimental Medicine* **208,** 455–467 (2011).
202. Yamada, T., Park, C. S., Burns, A., Nakada, D. & Lacorazza, H. D. The cytosolic protein G0S2 maintains quiescence in hematopoietic stem cells. *PLoS ONE* **7,** e38280 (2012).
203. Armstrong, S. A. *et al.* MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30,** 41–47 (2002).
204. Dawson, M. A. *et al.* Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478,** 529–533 (2011).
205. Dawson, M. A. *et al.* Recurrent mutations, including NPM1c, activate a BRD4-dependent core transcriptional program in acute myeloid leukemia. *Leukemia* **28,** 311–320 (2014).
206. Kühn, M. W. M. *et al.* Targeting Chromatin Regulators Inhibits Leukemogenic Gene Expression in NPM1 Mutant Leukemia. *Cancer Discovery* (2016). doi:10.1158/2159-8290.CD-16-0237
207. Rose, D. *et al.* Subtype-specific patterns of molecular mutations in acute myeloid leukemia. *Leukemia* **31,** 11–17 (2017).
208. Zuber, J. *et al.* Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nat Biotechnol* **29,** 79–83 (2011).
209. Chen, W. *et al.* Malignant transformation initiated by Mll-AF9: gene dosage and critical target cells. *Cancer Cell* **13,** 432–440 (2008).
210. Chen, Y. *et al.* Critical role for Gimap5 in the survival of mouse hematopoietic stem and progenitor cells. *Journal of Experimental Medicine* **208,** 923–935 (2011).
211. Lacombe, J. *et al.* Scl regulates the quiescence and the long-term competence of hematopoietic stem cells. *Blood* **115,** 792–803 (2010).
212. Ostrowski, M. *et al.* Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat. Cell Biol.* **12,** 19–30– sup pp 1–13 (2010).
213. Li, Q. *et al.* Oncogenic Nras has bimodal effects on stem cells that sustainably

increase competitiveness. *Nature* 1–15 (2013). doi:10.1038/nature12830
214. Klauke, K. *et al.* Tracing dynamics and clonal heterogeneity of Cbx7-induced leukemic stem cells by cellular barcoding. *Stem Cell Reports* **4,** 74–89 (2015).
215. King, K. Y. *et al.* Irgm1 protects hematopoietic stem cells by negative regulation of IFN signaling. *Blood* **118,** 1525–1533 (2011).
216. Chin, Y. E. *et al.* Cell Growth Arrest and Induction of Cyclin-Dependent Kinase Inhibitor p21WAF1/CIP1 Mediated by STAT1. *Science* **272,** 719–722 (1996).
217. Dimberg, A., Karlberg, I., Nilsson, K. & Oberg, F. Ser727/Tyr701-phosphorylated Stat1 is required for the regulation of c-Myc, cyclins, and p27Kip1 associated with ATRA-induced G0/G1 arrest of U-937 cells. *Blood* **102,** 254–261 (2003).
218. Heuser, M. *et al.* Modeling the functional heterogeneity of leukemia stem cells: role of STAT5 in leukemia stem cell self-renewal. *Blood* **114,** 3983–3993 (2009).
219. Kovacic, B. *et al.* STAT1 acts as a tumor promoter for leukemia development. *Cancer Cell* **10,** 77–87 (2006).
220. Madapura, H. S. *et al.* Interferon γ is a STAT1-dependent direct inducer of BCL6 expression in imatinib-treated chronic myeloid leukemia cells. *Oncogene* **36,** 4619–4628 (2017).
221. Spiekermann, K., Biethahn, S., Wilde, S., Hiddemann, W. & Alves, F. Constitutive activation of STAT transcription factors in acute myelogenous leukemia. *Eur. J. Haematol.* **67,** 63–71 (2001).
222. Xia, Z., Baer, M. R., Block, A. W., Baumann, H. & Wetzler, M. Expression of signal transducers and activators of transcription proteins in acute myeloid leukemia blasts. *Cancer Research* **58,** 3173–3180 (1998).
223. Gu, H. *et al.* Sorting protein VPS33B regulates exosomal autocrine signaling to mediate hematopoiesis and leukemogenesis. *J. Clin. Invest.* **126,** 4537–4553 (2016).
224. Hong, C.-S., Muller, L., Whiteside, T. L. & Boyiadzis, M. Plasma exosomes as markers of therapeutic response in patients with acute myeloid leukemia. *Front Immunol* **5,** 160 (2014).
225. Zhou, J., Wang, S., Sun, K. & Chng, W.-J. The emerging roles of exosomes in leukemogeneis. *Oncotarget* **7,** 50698–50707 (2016).
226. Falkowska-Hansen, B. *et al.* An inducible Tet-Off-H2B-GFP lentiviral reporter vector for detection and in vivo isolation of label-retaining cells. *Exp. Cell Res.* **316,** 1885–1895 (2010).