PhD degree in Molecular Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: MED/04

Functional and genetic heterogeneity in acute myeloid leukaemia

Thaleia Vlachou

European Institute of Oncology (IEO), Milan

Matricola n. R09862

Supervisor: Prof. Pier Giuseppe Pelicci, MD, PhD

European Institute of Oncology (IEO), Milan

Added Supervisors: Dr. Chiara Ronchini, PhD

Italian Institute of Technology (IIT), Milan

Dr. Emanuela Colombo, PhD

European Institute of Oncology (IEO), Milan

Academic year 2016-2017

Acknowledgements

First and foremost, I would like to express my sincere and deepest gratitude to my mentor, Pier Giuseppe Pelicci, for giving me the opportunity to work on this project, for his support and guidance, for being an inspirational boss and teacher, for the liberty and challenges he gave me throughout these years. Working with him has been an invaluable experience that shaped my character and expanded my scientific horizons.

Next, I would like to thank the members of the group directly involved in this project: Chiara Ronchini, Emanuela Colombo and Andrea Cammarata. Chiara has been both a supervisor and a friend. I thank her for her endless support, kindness and patience, for always being available to help with experiments or discuss problems and ideas, and for her perseverance especially during the hard times. I would like to thank Emanuela for her guidance and helpful suggestions on key experiments, the insightful comments on data interpretation and the fruitful scientific discussions. I am grateful to Andrea for sharing his work and ideas with me, for teaching me practical skills and for his contribution in demanding experiments. A special thanks goes also to Angela Santoro, with whom I shared the project on Myc but also most of my anxiety and frustration in the last four years. I thank her for the long scientific discussions, for the motivation speeches and for her generosity.

This project would have never reached its present form without the extensive and thorough bioinformatics analysis, for which I am indebted to Laura Riva and her team: especially Jole Costanza for the long hours of work, Margherita Bodini and Giorgio Melloni. I would also like to thank Luisa Lanfrancone and Daniela Bossi for the exceptional collaboration on the melanoma studies, as well as Daniela Besozzi, Marco S. Nobile and Paolo Cazzaniga for their wonderful work on the mathematical modelling. Finally, I am grateful to Paola Dalton for her guidance on scientific writing, the constructive criticism and the kind advice.

Last but not least, I would have never even embarked on this journey without the moral support of my closest friends and, of course, my parents Abraham and Maria. I thank them for having faith in me and for encouraging me to go after my dreams, but most importantly I thank them for teaching me to never give up.

Table of Contents

List of figures

List of tables

List of abbreviations

5-FU	5-fluorouracil
AML	Acute myeloid leukaemia
BM	Bone marrow
BrdU	Bromodeoxyuridine
CN-AML	Cytogenetically normal AML
cPPT	Central polypurine tract
CR	Complete remission
CSC	Cancer stem cell
EF-1α	Elongation factor 1α
ELDA	Extreme limiting dilution analysis
FAB	French-American-British classification
FACS	Fluorescence activated cell sorting
FBS, NA	Foetal bovine serum, North-American
FBS, SA	Foetal bovine serum, South-American
GFP	Green fluorescent protein
H2B	Histone 2B
HSC	Haematopoietic stem cell
HSPCs	Haematopoietic stem and progenitor cells
IF	Intrafemoral
IP	Intraperitoneal
IT	Intratibial
IV	Intravenous
LIC	Leukaemia initiating cell
LRC	Label-retaining cell
LSC	Leukaemia stem cell
LT-HSC	Long-term repopulating HSC
LTR	Long terminal repeats
MAF	Minor allele frequency
MDS	Myelodysplastic syndrome
MFI	Mean fluorescence intensity
MOI	Multiplicity of infection
MPD	Myeloproliferative disease
NGS	Next generation sequencing
NSG	NOD SCID IL2RG null
PB	Peripheral blood
PDX	Patient derived xenograft
PuroR	Puromycin resistance marker
RFP	Red fluorescent protein
RRE	Rev response element
SIN	Self inactivating
SNP	Single nucleotide polymorphism

SNV	Single nucleotide variant
SPL	Spleen
T2A	*Thosea asigna* virus 2A translational cleavage site
TCGA	The Cancer Genome Atlas
TIC	Tumour initiating cell
tTA	Tet trans-activator
TU	Transducing units
UbiC	Ubiquitin C
VAF	Variant allele frequency
VSV-G	Vesicular stomatitis virus protein G
WES	Whole exome sequencing
WHO	World Health Organisation
WPRE	Woodchuck hepatitis virus post-transcriptional regulation element

Abstract

Acute myeloid leukaemia (AML) is the most frequent leukaemia in adults, and still represents a disease with an unmet medical need, with 50-60% of patients relapsing within 3 years after diagnosis. AMLs are characterised by a high degree of intra-tumour heterogeneity, both at the biological and the genetic level, which is critical for tumour maintenance and response to treatments. Biologically, AMLs are organised hierarchically, with rare stem-like cells (leukaemia stem cells, LSCs) endowed with the unique properties of self-renewal and differentiation. Genetically, AMLs harbour patient-specific combinations of different driver mutations, which are organised within individual cases in sub-clones with distinct growth properties. We hypothesized that tumour maintenance and relapse in AMLs are driven by the selective expansion of quiescent sub-clones within the LSC population, which serve as the genomic and functional reservoir of the tumour. The experimental strategy we employed to test this hypothesis is based on the xenotransplantation of human leukaemias, the implementation of an *in vivo* clonal tracking approach, the functional isolation of leukaemic subpopulations with diverse proliferation histories and whole-exome sequencing (WES) of bulk and isolated leukaemic subpopulations. Our aims were to assess the proliferative hierarchy of LSCs and to examine their intrinsic genetic heterogeneity. We identified two functional LSC classes, quiescent and cycling, that are in equilibrium in the tumour and largely share the same clonal architecture. We further observed that genetic leukaemic clones appear to consist of a high number of individual LSCs, the majority of which exhaust upon serial transplantation. Finally, by genetic analyses of isolated leukaemic subsets, we were able to detect a specific enrichment for rare mutations in the quiescent compartment of two patient xenografts. Our data indicate that tumour evolution is sustained by the quiescent LSC pool and suggest that their highly proliferating counterpart has a finite lifespan. We expect that

the results of our studies will provide new insights into the mechanisms of disease progression and treatment response in AML, and potentially reveal novel therapeutic approaches.

1 Introduction

1.1 Acute myeloid leukaemia (AML)

Acute myeloid leukaemia (AML) is a clonal disorder of the blood that is characterised by the aberrant proliferation of abnormal and, often, poorly differentiated myeloid progenitors (blasts), which primarily accumulate in the bone marrow (BM), invade the peripheral blood (PB) and occasionally infiltrate other extramedullary tissues. These cells interfere with normal haematopoiesis, resulting in signs and symptoms of anaemia, and thrombocytopenia and leukopenia. In the absence of treatment, patients succumb to the disease, usually within one year after diagnosis, due to fatal complications associated with the BM failure (haemorrhage and infection) (Estey and Dohner 2006).

AML is one of the most common haematological malignancies in adults, with an estimated incidence of 8.8 cases *per* 100,000 men and women in Europe for the year 2012 (Ferlay, Steliarova-Foucher et al. 2013). It is a highly heterogeneous disease, both biologically and clinically, with variable prognosis among patients and a high mortality rate. The French-American-British (FAB) classification system was established in 1976 as a first attempt to categorise AML sub-types according to morphological, immunophenotypic and cytochemical criteria (Bennett, Catovsky et al. 1976). The current World Health Organisation (WHO) classification, last updated in 2016, incorporated cytogenetics and molecular genetics to define four principal categories of AML with clinical significance: AML with recurrent genetic abnormalities, AML with myelodysplasia-related changes, therapy-related myeloid neoplasms and AML not otherwise specified (Table 1.1) (Vardiman, Thiele et al. 2009, Arber, Orazi et al. 2016). Based on aetiology alone, three classes of AML can be recognised: secondary AML (following a myelodysplastic syndrome, MDS, or a myeloproliferative disease, MPD), therapy-related AML (usually

from cytotoxic chemotherapy for a solid tumour), and *de novo* AML (Lindsley, Mar et al. 2015).

Table 1.1 World Health Organisation (WHO) classification of acute myeloid leukaemia (AML) and related neoplasms.

Adapted from Arber, Orazi et al., Blood, 2016.

Type	Cytogenetic, morphological and other characteristics
AML with recurrent genetic abnormalities	t(8;21)(q22;q22.1); *RUNX1-RUNX1T1* inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11* *PML-RARA* (acute promyelocytic leukaemia, APL) t(9;11)(p21.3;q23.3); *MLLT3-KMT2A* t(6;9)(p23;q34.1); *DEK-NUP214* inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM* t(1;22)(p13.3;q13.3);*RBM15-MKL1* (megakaryoblastic AML) *BCR-ABL1* (provisional entity) mutated *NPM1* biallelic mutations of *CEBPA* mutated *RUNX1* (provisional entity)
AML with myelodysplasia-related changes	Complex karyotype (3 or more abnormalities) Unbalanced abnormalities -7/del(7q) del(5q)/t(5q) i(17q)/t(17p) -13/del(13q) del(11q) del(12p)/t(12p) idic(X)(q13) Balanced abnormalities t(11;16)(q23.3;p13.3) t(3;21)(q26.2;q22.1) t(1;3)(p36.3;q21.2) t(2;11)(p21;q23.3) t(5;12)(q32;p13.2) t(5;7)(q32;q11.2) t(5;17)(q32;p13.2) t(5;10)(q32;q21.2) t(3;5)(q25.3;q35.1)
Therapy-related myeloid neoplasms	Therapy-related myelodysplastic syndrome (t-MDS) Therapy-related AML (t-AML)
AML, not otherwise specified	AML with minimal differentiation AML without maturation AML with maturation Acute myelomonocytic leukaemia Acute monoblastic/monocytic leukaemia Pure erythroid leukaemia Acute megakaryoblastic leukaemia Acute basophilic leukaemia Acute panmyelosis with myelofibrosis
Myeloid sarcoma	
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis (TAM) Myeloid leukaemia associated with Down syndrome

Despite the advances in the field, the backbone of therapeutic intervention for non-promyelocytic AML has remained essentially unaltered for the last 40 years. The standard-care treatment is provided in two phases, induction and consolidation. Typically, patients who are considered fit for intensive chemotherapy receive a "3+7" induction regimen, which consists in the intravenous administration of an anthracycline (daunorubicin, idarubicin or mitoxantrone) for the first 3 days of treatment and the continuous infusion of cytarabine from day 1 through to day 7. About 60-85% of patients below the age of 60 respond to the induction therapy and initially achieve complete remission (CR), which is defined by a blast count <5% in the BM and an increase in the neutrophil (>100) and platelet (>100,000) counts (Estey and Dohner 2006, Dohner, Weisdorf et al. 2015). Consolidation therapy is usually provided to prolong CR with 2-4 cycles of intermediate doses of cytarabine. Allogeneic haematopoietic stem cell (HSC) transplantation is considered as an alternative to consolidation chemotherapy, given the availability of a compatible donor and the patient's overall health status and risk assessment (Dohner, Weisdorf et al. 2015).

Nevertheless, 50-60% of patients will relapse within 3 years after diagnosis. Salvage therapy, in the form of aggressive chemotherapy or allogeneic HSC transplantation, may still be an option at this stage but the prognosis after the first recurrence becomes dismal, especially when relapse occurs after a brief CR period (< 6 months). New targeted therapies for AML treatment (including epigenetic modifiers, tyrosine kinase inhibitors, cell-cycle and signalling inhibitors and antibody-based therapies) have been recently developed and tested in clinical trials, but none of them has shown therapeutic promise as a single-agent so far (Dohner, Weisdorf et al. 2015). The molecular basis of chemoresistance in AMLs remains unknown and there are currently no available markers to safely predict treatment outcome. It is thought, however, that disease recurrence is caused by the persistence of residual chemoresistant leukaemic cells at CR, as a consequence of the intra-tumour biological and genetic heterogeneity in AMLs.

1.2 The genomic landscapes of AML

AML diagnosis, classification and patient risk stratification are routinely performed by analyses of BM aspirates and PB smears and consist of the examination of the morphology and relative representation of leukocytes or other nucleated cells, lineage determination using immunophenotyping by flow cytometry or cytochemistry, karyotype characterisation and mutation testing (Dohner, Estey et al. 2010).

Cytogenetic findings alone permit patients' risk to be categorised as favourable [t(15;17), t(8;21) and inv(16)/t(16;16)], intermediate [normal karyotype or t(9;11)] or adverse [del(7), del(5), inv(3)/t(3;3), t(6;9), t(v;11) or complex karyotype] – with very different cure rates (Grimwade, Walker et al. 1998, Dohner, Estey et al. 2010). Well-characterised recurrent chromosomal abnormalities are identified in ~52% of all adult patients diagnosed with primary AML. However, about 40-50% of all cases belong to the group of cytogenetically normal AML (CN-AML) which is generally associated with intermediate risk, but still remains very heterogeneous and the characterisation of the patient mutational profile is required to better define prognosis (Saultz and Garzon 2016).

Studies on animal models and hereditary leukaemias suggested that the co-operation of two types of genetic alterations is required for the development of AML. The "two-hit" model of leukaemogenesis, thus, foresees two complementary classes of mutations (Gilliland and Griffin 2002, Gilliland, Jordan et al. 2004). Class I mutations (*e.g. FLT3-ITD*, *FLT3-TKD*, *NRAS*, *KRAS*, *TP53, c-KIT*) confer a survival or proliferative advantage to haematopoietic progenitors, but are not sufficient by themselves to induce frank AML appearance. Class II mutations (e.g. *NPM1* and *CEBPA*) block myeloid differentiation, and like class I mutations they cannot cause leukaemia in the absence of additional co-operating genetic lesions (Gilliland, Jordan et al. 2004, De Kouchkovsky and Abdul-Hay 2016). A third class of mutations can be added, encompassing alterations in epigenetic regulators (e.g. *DNMT3A*, *TET2*, *IDH1* and *IDH2*), which affect both self-renewal and differentiation and

have gradually gained increasing attention in recent years (De Kouchkovsky and Abdul-Hay 2016).

Key genetic alterations that are currently known to affect prognosis and are predictive of the response to treatment had been already uncovered before the next generation sequencing (NGS) era. Constitutive activation of the fms-like tyrosine kinase 3 (*FLT3*) receptor, due to in-frame tandem duplications (*FLT3-ITD*) or point mutations in the tyrosine kinase domain (*FLT3-TKD*), is found in ~1/3 of AML cases. Point mutations (*FLT3-TKD*) have been associated with a favourable outcome, while the presence of *FLT3-ITD* is indicative of higher relapse risk and lower overall survival (Mead, Linch et al. 2007, Grimwade, Ivey et al. 2016). Mutations in the *RAS* family genes (*KRAS*, *NRAS*) are also commonly found in AML, but their presence alone is of low prognostic value. The co-occurrence with *cKIT* mutations or core-binding factor (*CBF*) chromosomal re-arrangements, however, is predictive of a poor outcome (Grimwade, Ivey et al. 2016).

Mutations in the nucleophosmin gene (*NPM1*) leading to the cytoplasmic mislocalization of the NPM1 chaperone protein are highly recurrent in AML, accounting for > 50% of the CN-AML cases. According to the current WHO classification (Table 1.1) (Arber, Orazi et al. 2016), *NPM1* mutated AML (or *NPM1c*) is a distinct disease entity, generally characterised by favourable prognosis. However, both studies in animal models and clinical data have shown that the co-occurrence with *FLT3-ITD* negatively impacts on the disease outcome (Vassiliou, Cooper et al. 2011, Kuhnl and Grimwade 2012, Mallardo, Caronno et al. 2013). Similarly, mutations in the gene encoding CCAAT/enhancer binding protein α (*CEBPA*) consist another class of genetic lesions recurrently found in CN-AML (~10% of cases) and associated with relevant clinical features. In the majority of the cases, both *CEBPA* alleles are affected, often due to inheritance of a germline mutation and subsequent acquisition of an additional somatic mutation. AMLs carrying biallelic *CEBPA* mutations are generally characterised by a favourable outcome and usually do not associate

with *FLT3-ITD* (Grimwade, Ivey et al. 2016). A provisional entity for *RUNX1* mutated AMLs has been also included in the last updated version of WHO classification (Table 1.1) (Arber, Orazi et al. 2016). The presence of *RUNX1* mutations is usually mutually exclusive with *NPM1* and *CEBPA* lesions, and is associated with secondary AML and generally predicts a poor outcome (Grimwade, Ivey et al. 2016).

Recent studies using massive parallel sequencing have highlighted the importance of epigenetic modifier mutations in cancer progression (Mardis, Ding et al. 2009, Ley, Ding et al. 2010, Yamashita, Yuan et al. 2010). Among those, DNA methyltransferase 3A (*DNMT3A*) has been found to be mutated in ~30% of CN-AML patients. *DNMT3A* mutations often co-occur with *NPM1c* and *FLT3-ITD*, suggesting a potential synergistic effect that promotes leukaemogenesis and disease progression leading altogether to an adverse outcome. Mutations affecting DNA methylation have also gained attention, with ten eleven translocation 2 (*TET2*) and isocitrate dehydrogenase enzymes (*IDH1* and *IDH2*) identified as additional crucial players (Grimwade, Ivey et al. 2016).

Genome-wide studies have shed light on our knowledge of the genomic landscape, transcriptional and methylation profiles of different tumour types. For AML, there are an estimated 13 mutations *per* case, 8 of which are considered to be passenger mutations (Cancer Genome Atlas Research 2013). Mutations have been classically designated as "drivers" when they confer a selective advantage to a tumour clone and as "passengers" when they are neutral (Stratton, Campbell et al. 2009). However, this classification is mainly based on the observed frequency of mutations in the patient and could be misleading, as different microenvironments or external pressures, such as chemotherapy, may lead to the selective expansion of minor tumour sub-clones carrying mutations initially considered as passengers (Ding, Ley et al. 2012). Furthermore, neutral mutations can persist during tumour evolution and contribute to the intrinsic genetic heterogeneity, given that they have no detrimental effects on tumour growth and, therefore, do not alter

clonal competition (Figure 1.1) (Williams, Werner et al. 2016). Under these conditions, there is the possibility of the appearance of additional mutations in the same clone, which may eventually synergistically confer a competitive advantage (Welch, Ley et al. 2012). Although AMLs have a relatively low mutation rate, as compared to solid tumours, they are nonetheless characterised by high patient-to-patient heterogeneity and despite the increasing number of cases available for meta-analyses, the discrimination of true cancer-associated genes and driver mutations from the non-relevant background remains cumbersome. (Lawrence, Stojanov et al. 2013).

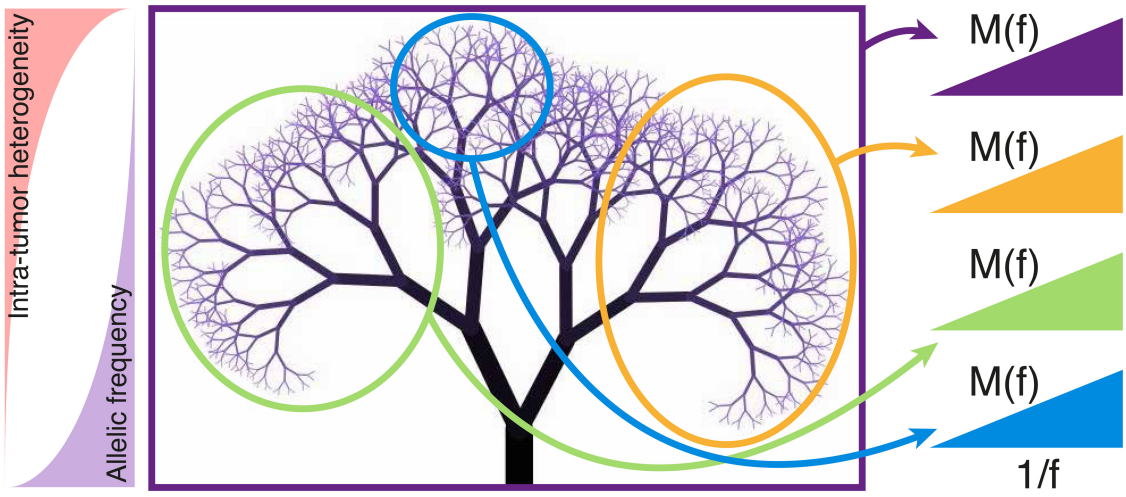

Figure 1.1 Neutral evolution contributes to intra-tumour genetic heterogeneity.

After the key oncogenic event that triggers tumorigeneis, a large number of mutations, M(f), is accumulated linearly which inversely correlates with their respective allelic frequencies, f, leading altogether to increased intrinsic genomic heterogeneity. Bulk sequencing approaches cannot faithfully depict the full spectrum of this heterogeneity due to the large number of cells sampled which typically leads to the identification of mutations belonging to the "trunk" of the phylogenetic tree. Reprinted by permission from Macmillan Publishers Ltd: Nature Genetics http://www.nature.com/ng/index.html (Williams, Werner et al. 2016) , copyright © 2016.

A very recent study by Papaemmanuil and colleagues included data from 1,500 patients and proposed a more comprehensive classification scheme for AML, solely on the basis of genomic features. In more detail, 5,234 driver mutations were identified across 76 genes

and genomic regions. Distinct patterns of mutation co-occurrence or mutual exclusivity led to the definition of 11 AML classes with relevant clinical features. In addition to the ones already recognised by the WHO classification (Table 1.1) (Arber, Orazi et al. 2016), three novel disease entities were introduced, encompassing mutations in genes encoding chromatin/RNA-splicing regulators or both (18% of cases), *TP53* mutations/chromosomal aneuploidies or both (13% of cases) and, provisionally, $IDH2^{R172}$ mutations (1% of cases). Using this classification scheme at least one driver mutation was scored in 96% of the patients analysed and two or more drivers were found in the 86% of the cohort (Papaemmanuil, Gerstung et al. 2016).

1.3 Patterns of clonal evolution in AML

De novo AML is a disease with an estimated long latency, also termed "pre-leukaemia". The genetic composition and clonal architecture at leukaemia presentation may allow the inference of the tumour evolution model from the pre-leukaemic state, but very few clinical or experimental data are available to confirm these hypotheses. It is thought that during the pre-leukaemic phase normal haematopoietic stem and progenitor cells (HSPCs) accumulate mutations that will gradually lead to clonal haematopoiesis, albeit still maintaining differentiation potential at this stage (Jan, Snyder et al. 2012, Corces-Zimmerman, Hong et al. 2014, Shlush, Zandi et al. 2014). On the other hand, studies in elderly healthy individuals with clonal haematopoiesis showed that the presence of leukaemia-associated mutations correlated with a higher risk of leukaemia development (Genovese, Kahler et al. 2014, Jaiswal, Fontanillas et al. 2014, Xie, Lu et al. 2014). Finally, the existence of multiple genetic sub-clones at diagnosis (Shlush, Chapal-Ilani et al. 2012, Klco, Spencer et al. 2014, Paguirigan, Smith et al. 2015) supports a more intricate branching evolution pattern, as compared to the linear model of successive mutation acquisition in a single clone.

In the case of AML, it has been recently demonstrated that relapse–specific DNA mutations can be either acquired or selected for during chemotherapy. More precisely, the genomes of primary and relapsed tumours from 8 AML patients were sequenced pair-wise and in 3 out of 8 instances the dominant clone at diagnosis persisted after the treatment and eventually acquired additional mutations, while in the rest of the cases a minor sub-clone expanded (Figure 1.2) (Ding, Ley et al. 2012). It should be noted however, that (due to technical limitations) this study may have underestimated the clonal heterogeneity of AMLs at diagnosis, given that only a minor subpopulation of quiescent or slow-cycling cells is expected to be able to escape treatment. This means that some DNA mutations found in relapsed AML tumours could be represented by very low frequencies and might, therefore, not be detectable in the primary tumours. Similarly, in a study regarding *NPM1c* AMLs, relapse was shown to derive from the dominant clone in the primary tumour only in half of the cases, while a minor sub-clone was selected to expand by the changing environment in the rest (Kronke, Bullinger et al. 2013).

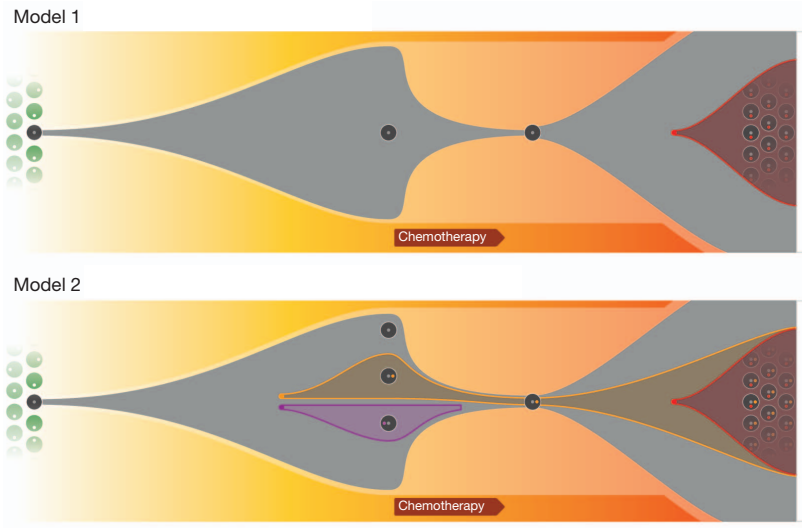

Figure 1.2 Models of tumour evolution in AML upon chemotherapy.

Whole genome sequencing of 8 pairs of primary and relapse AML patient samples revealed two patterns of tumour evolution upon chemotherapy. According to Model 1 (upper panel), the dominant clone of the primary leukaemia persists during the treatment and gives rise to the relapse clone, which subsequently acquires additional mutations. Model 2 (lower panel), depicts an alternative scenario in which multiple clones co-exist in the primary leukaemia and a minor sub-clone is selected to expand and evolve into the

relapse tumour under the environmental pressure of chemotherapy (Ding, Ley et al. 2012).

It is becoming increasingly evident that the intra-tumour heterogeneity is under-estimated. Technological advances with the development of NGS, including single-cell sequencing, as well as sophisticated mathematical modelling have allowed us to obtain a better understanding of the intrinsic complexity of cancer genomes. As far as solid tumours are concerned, multiple sampling from different tumour sites has enabled researchers to decipher information regarding the extent of intra-tumour heterogeneity (Gerlinger, Rowan et al. 2012). This is obviously not a possibility in the case of haematological malignancies, like AML. Sequencing of morphologically or immunophenotypically distinct subpopulations has provided an alternative for the evaluation of intra-tumour genetic heterogeneity at a higher resolution. In fact, this approach allowed the detection of relapse-specific mutations in specific subsets of the primary leukaemia (Ding, Ley et al. 2012, Klco, Spencer et al. 2014). Additional technical refinements, such as increasing the depth of sequencing and the use of multiple mutation calling algorithms, have been also shown to provide substantial improvement in elucidating the clonal architecture of the AML genome (Griffith, Miller et al. 2015).

However, both multi-region or immunophenotype-specific sequencing and sequencing approaches with increased coverage/depth can only measure the average signal of a bulk population, missing rare variants that could be present in only very minor tumour sub-clones. Furthermore, inferring tumour clonal architecture becomes more challenging when it comes to low frequency mutations. Single-cell sequencing provides a useful tool to elaborate on intra-tumour genetic and clonal heterogeneity, however technical limitations impede with the clear designation of biologically relevant diversity beyond technical errors. Multiple techniques can be used for the isolation of single-cells that are abundant in tumours (*e.g.* serial dilution, microfluidics, fluorescence activated cell sorting – FACS),

but none of them is as efficient for rare cancer cells (< 1% of the tumour) (Navin 2014). Therefore, to overcome this hurdle, the isolation of relevant tumour subpopulations should be envisioned prior to the application of single-cell sequencing approaches for the identification of relapse-specific mutations in primary tumours. It should be noted, however, that whole-genome amplification may have allowed the application of whole genome and whole exome sequencing approaches starting with a single-cell input, but it also came with the price of low and non-uniform coverage that is associated with a high rate of false positive and false negative errors (Navin 2014).

Nevertheless, informative studies at single cell level allowed delineation of the sequence of mutation acquisition in MDS to the development of a secondary AML (Hughes, Magrini et al. 2014) and most importantly revealed the persistence of HSCs with relevant mutations after chemotherapy (Jan, Snyder et al. 2012). Similar studies have also suggested polyclonality at the relapse emergence (Shlush, Chapal-Ilani et al. 2012), hinting to the potential co-existence of multiple mechanisms of chemoresistance. The scenario of a symbiotic relationship of tumour sub-clones, that could either compete for resources in the niche or synergise to promote tumour growth should be taken into account in the development of novel therapeutic strategies (McGranahan and Swanton 2015).

1.4 The Cancer Stem Cell (CSC) model

Tumours can be envisioned as aberrant organs endowed with a complex and hierarchical cellular organisation, similar to their normal tissue counterparts, with cancer stem cells (CSCs) placed at the apex of this hierarchy. *In vitro* clonogenic and *in vivo* transplantation assays paved the way for the identification of tumoral cell populations with variable tumorigenic potential and the designation of CSCs as the only cells able to sustain tumour initiation and progression *in vivo* (Figure 1.3). Pioneering studies in John Dick's lab demonstrated the existence of such a population in AML by the prospective isolation and xenotransplantation into immunocompromised murine hosts of leukaemic populations with different combinations of CD34 and CD38 antigen expression. These studies documented that only the $CD34^{+}CD38^{-}$ compartment retains tumorigenic ability *in vivo* (Lapidot, Sirard et al. 1994, Bonnet and Dick 1997). There is great heterogeneity in the expression profile of the CD34/CD38 surface markers among different patient samples and leukaemia stem cells (LSCs) have been detected with varying frequencies in more than one fraction in many AML samples (Taussig, Miraki-Moud et al. 2008, Taussig, Vargaftig et al. 2010). For instance, LSC activity is commonly found in the CD34- compartment in *NPM*-mutated AMLs, which are generally characterised by low CD34 expression (Taussig, Vargaftig et al. 2010). Nevertheless, in CD34+ leukaemias, the $CD34^{+}CD38^{-}$ subset has been associated with higher tumorigenic capacity through limiting dilution xenotransplantation assays (Eppert, Takenaka et al. 2011, Sarry, Murphy et al. 2011). Furthermore, the existence of CSCs, or otherwise stated, tumour initiating cell (TIC) populations has been since demonstrated also in numerous solid tumours, including breast cancer (Al-Hajj, Wicha et al. 2003), glioblastoma (Singh, Hawkins et al. 2004) and colorectal cancer (O'Brien, Pollett et al. 2007).

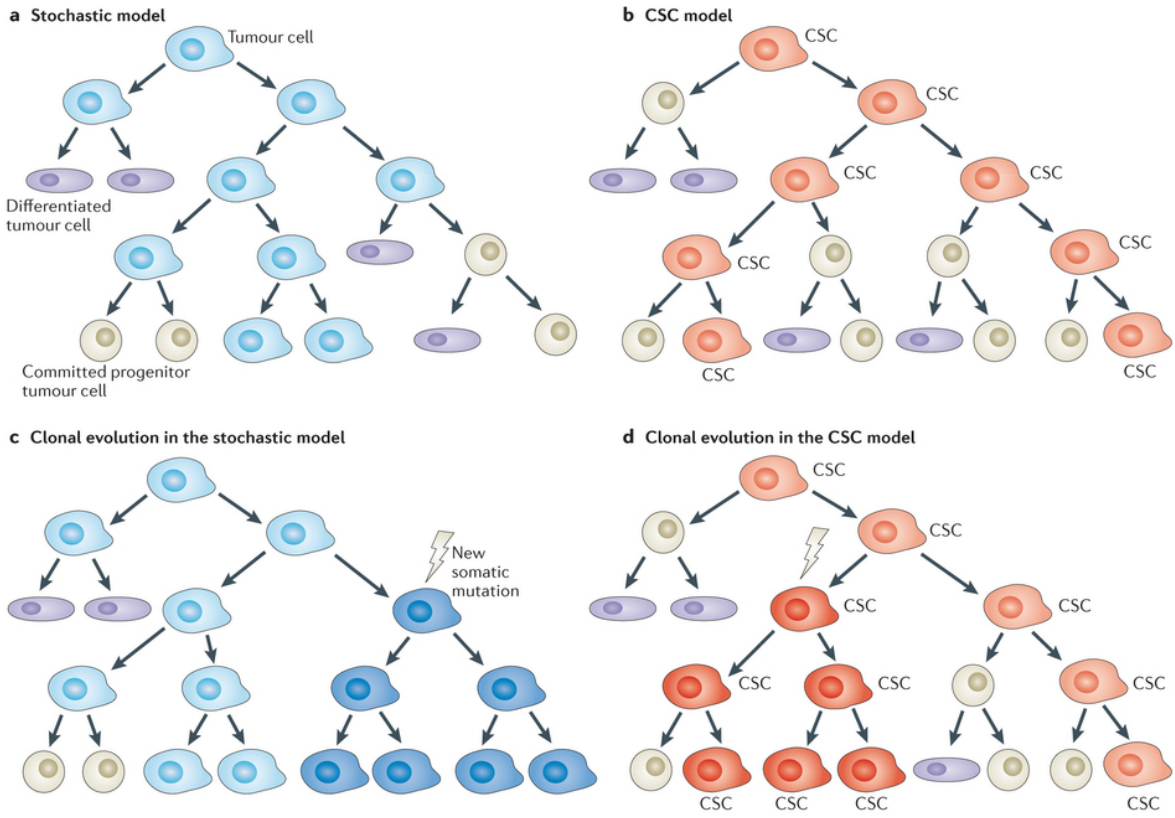

Figure 1.3 Stochastic *versus* cancer stem cell (CSC) model of tumour growth.

The stochastic model of tumour growth (a) foresees that all tumour cells have equal self-renewal potential and can randomly produce differentiated progeny, which contributes to the intra-tumoral heterogeneity. In the CSC model (b), instead, only a small fraction of the tumour cells retains the ability of long-term self-renewal and propagation of the disease. In both cases, the acquisition of new mutations leads to a higher level of intrinsic heterogeneity in the tumour. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer http://www.nature.com/nrc/index.html (Beck and Blanpain 2013), copyright © 2013.

According to the Cancer Stem Cell model, CSCs are the only cells within a tumour that retain the ability of tumour initiation, maintenance, and propagation and should therefore be considered the only ones responsible for disease relapse after initial CR. Indeed, in glioblastomas, cells positive for the CSC marker Prominin-1 (CD133+) were found enriched after ionizing radiation both *in vitro*, after short-term cultures, and *in vivo*, in the brains of immunocompromised mice, as well as in tumour samples obtained from patients. Importantly, the higher frequency of CD133+ cells after radiation resulted in the development of more aggressive tumours upon transplantation in secondary recipient mice,

suggesting that glioblastoma CSCs might survive radiotherapy and lead to the relapse of the disease (Bao, Wu et al. 2006). Similarly, breast CSCs identified by aldehyde dehydrogenase 1 expression (ALDH1) in patient samples that did not achieve pathological CR after treatment, have been associated with resistance to chemotherapy (Tanei, Morimoto et al. 2009). The clinical relevance of such a hierarchical organisation is further highlighted by the increased frequency of CSCs in the most aggressive tumours (van Rhenen, Feller et al. 2005, Pece, Tosoni et al. 2010), as well as the dominance of a CSC transcriptional signature (Ginestier, Hur et al. 2007, Eppert, Takenaka et al. 2011). Therefore, there is a need for efficient therapeutic interventions that target specifically this population for complete tumour eradication (Figure 1.4) (Reya, Morrison et al. 2001).

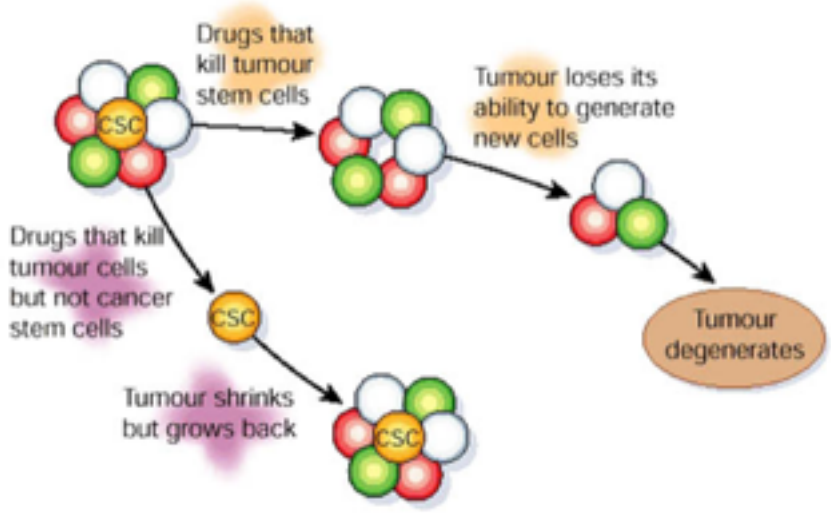

Figure 1.4 CSCs may evade standard therapeutic approaches.

CSCs could be intrinsically resistant to current chemotherapeutic approaches and therefore persist during treatment and lead to the emergence of the relapse. CSC-specific treatments, instead, should result in complete tumour eradication. Reprinted by permission from Macmillan Publishers Ltd: Nature http://www.nature.com/ (Reya, Morrison et al. 2001), copyright © 2001.

However, there are also tumours that do not seem to follow strictly the CSC model or at least suggest a higher plasticity within this intrinsic hierarchy. In melanoma, nearly all cells are considered to have virtually equal tumorigenic potential, as single cell

transplantations in immunocompromised murine hosts have proven their high *in vivo* tumour reconstitution capacity (Quintana, Shackleton et al. 2008). Also in this case, however, there is some intra-tumoral phenotypic diversity (Quintana, Shackleton et al. 2010) and functional heterogeneity (Roesch, Fukunaga-Kalabis et al. 2010).

It is not clear whether the persistence of chemoresistant cells is stochastic or whether pre-defined functional leukaemic sub-sets are specifically endowed with the capacity to evade cytotoxic treatments (Shlush and Mitchell 2015). Remission has been occasionally found to be clonal in AML (Fialkow, Janssen et al. 1991) and recent studies have demonstrated, more specifically, the persistence of pre-leukaemic HSPCs at remission (Corces-Zimmerman, Hong et al. 2014, Shlush, Zandi et al. 2014). Despite the controversies regarding the immunophenotypic characterisation of LSCs, relapse-specific mutations in a sub-group of *FLT3-ITD* AMLs could be traced back solely in the $CD34^{+}CD38^{-}$ leukaemic population of the primary tumour (in 70% of the cases, the mutations could not be detected in the rest of the primary tumours), providing a link between the genetics of LSC immunophenotype and relapse as opposed to the dominant clone at diagnosis (Shlush and Mitchell 2015).

1.5 Intra-tumour functional heterogeneity: beyond genetics

LSCs share many characteristics with their normal counterparts, including the ability to self-renew and to generate phenotypically diverse progeny, which are fundamental for the formation of a new tumour or organ, respectively (Reya, Morrison et al. 2001). HSPC populations with distinct self-renewal and lineage reconstitution capacities have been characterised in the murine BM and can be prospectively isolated on the basis of their immunophenotype, with long term repopulating HSCs (LT-HSCs) governing a well-established hierarchy. Even though human HSCs have not been as thoroughly studied due to experimental constrains, the differentiation hierarchy appears to be faithfully recapitulated in the mouse models (Rieger and Schroeder 2012).

In addition to the defined set of surface markers, different levels of functional heterogeneity have been described for normal HSCs, as they can be further classified as active, quiescent and/or dormant depending on their cell-cycle state (Weissman 2000, Wilson, Laurenti et al. 2008, Foudi, Hochedlinger et al. 2009, Takizawa, Regoes et al. 2011). LT-HSCs have been reported as predominantly quiescent and HSC dormancy appears to be essential for the maintenance of a long-term repopulating ability (Wilson, Laurenti et al. 2008, Qiu, Papatsenko et al. 2014, Bernitz, Kim et al. 2016). Furthermore, it has been shown that the clonal composition of the LT-HSC pool after transplantation is greatly skewed and not equally distributed among different skeletal locations (Verovskaya, Broekhuis et al. 2014). On the other hand, recent studies on native haematopoiesis have demonstrated that progenitors with extended lifespan can sustain a wide and highly dynamic clonal complexity with relatively minor contribution from the HSC compartment (Sun, Ramos et al. 2014, Busch, Klapproth et al. 2015, Sawen, Lang et al. 2016).

Drawing parallels in the cancer setting, CSC content may vary between different genetic subsets within the same tumour, as mutations conferring stem properties or significant proliferation advantage can lead to the dominance of a particular genetic clone (Kreso and

Dick, 2014). Moreover, diverse proliferation kinetics within tumour sub-clones and individual LSCs could provide an explanation for the main patterns of tumour evolution observed upon chemotherapy (Figure 1.2 and Figure 1.5). (Klco, Spencer et al. 2014)

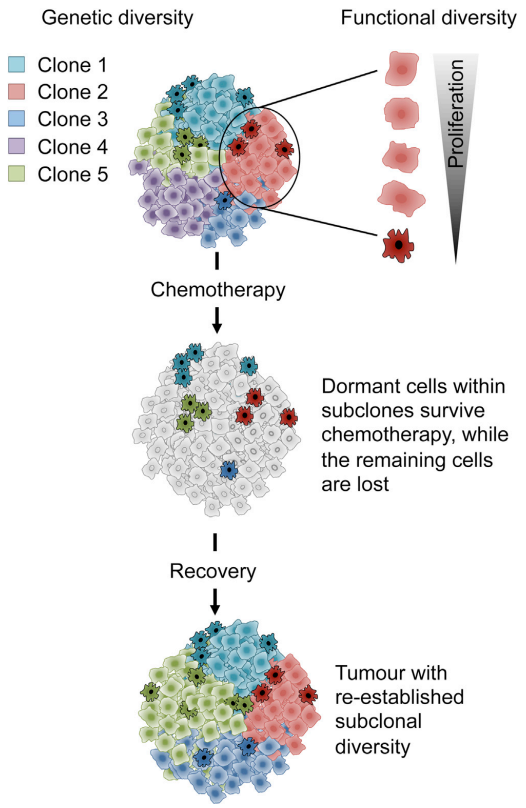

Figure 1.5 The impact of intra-tumour functional diversity on treatment response.

Each genetic clone is composed by a heterogeneous population of cells with diverse functional properties, such as proliferation. Non-proliferating or dormant cells within each clone may persist during chemotherapy and give rise to the relapse clones. Reprinted from Cell Stem Cell (Kreso and Dick 2014), copyright © 2014 , with permission from Elsevier.

Indeed, indirect evidence points to LSC quiescence as the source of resistance and eventual relapse. Quiescence has been linked to therapy resistance (Ishikawa, Yoshida et al. 2007, Saito, Uchida et al. 2010), while cell-cycle restricted LSCs are crucial for the development of leukaemia (Viale, De Franco et al. 2009). Longitudinal single cell analyses in two AML patients, from diagnosis to relapse, allowed lineage reconstruction depicting the genomic distance based on microsatellite mutations. In contrast to studies on driver mutations, this

approach enables the approximation of cell depth at relapse or, in other words, the number of cell replication events prior to the emergence of the relapse clone. Although lacking direct evidence, this study indicated that mutations carried in minor slowly expanding sub-clones were maintained after treatment, suggesting that cells with shorter replicative histories may be more apt to survive chemotherapy (Shlush, Chapal-Ilani et al. 2012).

2 Hypotheses and Aims of the project

Tumours, including leukaemias, are characterised by a high degree of intra-tumour heterogeneity, which is considered critical for tumour maintenance and growth, as well as for response to treatments. Two levels of heterogeneity are well recognised: biological and genetic. The first level refers to phenotypic cellular diversity. It has been previously shown that individual human AMLs are biologically heterogeneous and that only distinguishable cell subpopulations have reproducibly high clonogenic capacity (Bonnet and Dick 1997). Furthermore, a subpopulation of cell-cycle restricted LSCs appears to play a pivotal role in tumour initiation and maintenance (Viale, De Franco et al. 2009). This quiescent LSC subset could also evade current chemotherapeutic approaches that mainly target highly proliferative cells. The second level, referring to genetic heterogeneity, is also well documented in leukaemias. In a recent study from our group, a total of 200 AML samples were analysed and 1,400 genes were found mutated at an allelic frequency greater than 25%. Out of these, only roughly 10% were mutated in more than one patient per cytogenetic group and 3% were mutated at a statistically significant frequency (Riva, Ronchini et al. 2013).

Whole genome and whole exome sequencing studies have paved the way to a better understanding of clonal dynamics and dominance in several cancer types, indicating the presence of multiple genomes within a single tumour. It has been reported that leukaemias generally carry a lower number of mutations *per* case compared to adult solid tumours (Lawrence, Stojanov et al. 2013). However, studies of our group and others proposed that the genomic landscape of AMLs is potentially much more complex than previously thought due to the presence of a high number of low frequency mutations, which cannot be easily identified by the current sequencing and analysis strategies (Bodini, Ronchini et al. 2015, Griffith, Miller et al. 2015). These observations, together with the preliminary

evidence indicating that LSCs are distinguished by different cycling and self-renewal potential (Viale, De Franco et al. 2009), suggest that different LSC-derived clones might harbour different genomes.

These levels of genetic complexity might be responsible for the capacity of individual tumours to adapt and evolve under diverse environmental pressures such as chemotherapy. In this view, it is reasonable to think that cellular and genetic hierarchy within distinct tumour subpopulations converge to define the adaptability of a tumour as a whole. Since CSCs are thought to be the only cells responsible for tumour growth and maintenance *in vivo*, we postulated the following hypothesis: tumour growth, maintenance and relapse in AMLs are driven by the selective expansion of tumour sub-clones, presumably originating within the quiescent LSC population, which function as the genomic and functional reservoir of the leukaemia. Our experimental strategy to test this hypothesis consists in a) the generation of patient derived xenografts (PDX) of human leukaemias in immunocompromised mice, b) the analysis of the clonal evolution of the leukaemias *in vivo*, under steady-state conditions or during chemotherapy treatment, c) the implementation of a label-retaining assay for *in vivo* cell division tracking and purification of leukaemic cell populations with different proliferation potential, d) the identification of single nucleotide variants (SNVs) by NGS technologies.

The first aim of this study is to assess the functional heterogeneity and proliferative hierarchy of LSCs by *in vivo* clonal tracking, using a lentiviral barcode library. This approach should allow us to track the progeny of each cell initially labelled with a unique molecular identifier (barcode), which has been stably integrated in the host genome by means of lentiviral transduction. Since the assay is based on a transplantation procedure, it allows selective analysis of LSCs. The second aim is to separate LSCs with different proliferative histories (based on their label retention capacity) and assess their genetic complexity by whole-exome sequencing (WES). Our results provide new insights on

clonal evolution during disease progression in AML and set the grounds for the identification of novel therapeutic strategies.

3 Materials and Methods

3.1 Generation of AML patient derived xenografts

3.1.1 Patient Samples

Human leukaemic blasts were collected from the peripheral blood (PB) or other relevant tissues (*e.g.* BM) from patients through routine sample acquisition during healthcare delivery at the IEO hospital and other collaborating centres (*e.g.* "Policlinico" of Milan). Informed consent was obtained from all enrolled patients in compliance with the principles of the World Medical Association Declaration of Helsinki.

3.1.2 Xenotransplantation of human AMLs

It has been already documented that patient derived tumour xenografts at low passage (generally lower than 10) largely retain the phenotypic, histological and cytogenetic characteristics of the parental disease (Tentler, Tan et al. 2012). Furthermore, this setting foresees the expansion of the number of cancer cells initially obtained from the patient, enabling thus further manipulations and the implementation of more elaborate experimental procedures. For this purpose, we generated a bank of human AML xenografts by intravenous (IV) or intraperitoneal (IP) injections of ~10^7 CD3-depleted mononuclear BM or PB cells, freshly collected from AML patients, in 6-8 week old NOD SCID IL2RG null (NSG) immunocompromised mice. NSG mice were chosen as a host, based on their demonstrated ability to support greater engraftment of human haematopoietic stem cells (HSCs, CD34+), as compared to all other strains of immunocompromised mice (Agliano, Martin-Padura et al. 2008). Overall success of engraftment was around 70%, while the latency of the disease in the recipient mice was variable among different patient samples (from 4 weeks to 8 months).

3.2 Animal experimentation

All experiments including animals were conducted in accordance with the Italian Law (Legislative Decrees no. 116/92 and no. 26/2014), which enforces the EU Directives (86/609/EEC of 24 November 1986 and 2010/63/EU of 22 September 2010) on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. Mice were housed and taken care of accordingly to the guidelines set out in Commission Recommendation 2007/526/EC - June 18, 2007. At the end of the experiments, all mice were euthanized by inhalation of high concentrations of CO_2.

3.2.1 Limiting dilution and serial transplantation assays

In order to establish the LSC frequency and the tumorigenic properties of bulk leukaemias and isolated quiescent or cycling leukaemic populations for each PDX, we performed limiting dilution and serial transplantation assays. For this purpose, groups of 8-12 week old NSG mice were injected with leukaemic blasts at scalar cell doses (from 1,000,000 to 100 cells) IV or intra-BM directly in the femur (intrafemoral, IF) or in the tibia (intratibial, IT) of the hind legs. Direct delivery of the cells in the BM has been shown to improve the efficiency of engraftment (McKenzie, Gan et al. 2005). In the case of IF or IT injections, mice were anaesthetised with 2.5% avertin (delivered IP; 100% avertin: 10 g of 2,2,2-tribromoethanol 97%, Sigma-Aldrich, in 10 ml of 2-methyl-2-butanol ≥ 99%, Sigma-Aldrich) or by isoflurane inhalation. Successfully engrafted mice developed leukaemia with variable latencies, depending on the PDX and the number of cells injected (from 4 weeks to 8 months). All injected mice were monitored for a maximum period of one-year post-transplantation and euthanized by CO_2 inhalation when blast infiltration reached ~80% in the PB. At the end of the experiments BM and spleen (SPL) cells were collected and either re-transplanted immediately into secondary recipient mice, or stored in freezing medium (10% DMSO, Merck, in South-American foetal bovine serum [FBS, SA],

Microgem) in liquid nitrogen. Mice found dead without prior confirmation of leukaemia engraftment were excluded from all subsequent analyses. Finally, the LSC frequency was calculated using available computational tools (extreme limiting dilution analysis, ELDA; http://bioinf.wehi.edu.au/software/elda/) (Hu and Smyth 2009).

3.2.2 Engraftment analysis

PB (30-50 µl) was collected from the lateral tail-veins of injected and control (non-injected) mice at regular intervals, after the transplantation of the leukaemic blasts, for engraftment analysis. EDTA (0.5 M, pH 8.0) was used as anticoagulant and the red blood cells were lysed in a hypotonic salt solution (8.125 mg/ml NH_4Cl, 1 mg/ml $KHCO_3$, 0.13 mM EDTA in dH_2O). The level of engraftment of the human blasts in the transplanted animals was estimated by flow cytometry analyses of PB cells stained with an antibody specific for the human CD45 haematopoietic surface antigen (hCD45; see section 3.4.1).

3.2.3 Administration of 5-fluorouracil (5FU)

5-fluorouracil (5-FU) is an antimetabolite widely used as a chemotherapeutical agent in the treatment of breast, head and neck and, most commonly, colorectal cancer. Its mechanism of action primarily involves the inhibition of the enzyme thymidylate synthase, consequently blocking the thymidine formation required for DNA synthesis (Longley, Harkin et al. 2003). It has been previously shown that 5-FU selectively eliminates rapidly dividing haematopoietic progenitor and stem cells in the normal mouse BM, thus allowing subsequent cell-cycle entry of the highly quiescent LT-HSCs (Randall and Weissman 1997, Wilson, Laurenti et al. 2008). Therefore, 5-FU is a useful tool that can be employed to enforce an environmental pressure for the selective expansion of quiescent LSCs. For this purpose, a single dose of 150 µg 5-FU *per* g of body weight was administered IP to NSG mice soon after transplantation (3-10 days, in order to allow homing of the human blasts). As previously described, the level of engraftment was routinely checked in the PB

of the treated recipients and the BM and SPL cells were collected upon overt leukaemia manifestation.

3.2.4 *In vivo* label-retaining assay

Human leukaemic cells were transduced *in vitro* with an inducible Tet-Off lentiviral vector encoding the H2B-GFP fusion protein (Figure 3.1) (Falkowska-Hansen, Kollar et al. 2010), prior to their xenotransplantation into NSG mice and chasing with doxycycline *in vivo*. Doxycycline was administered to the experimental animals engrafted with H2B-GFP expressing blasts through specially modified feed (containing 625 mg doxycycline *per* kg of pellets). The conditional shut down of transgene expression by doxycycline led to gradual reduction of the GFP signal, allowing for the segregation of subpopulations within the tumour based on their respective proliferation rate.

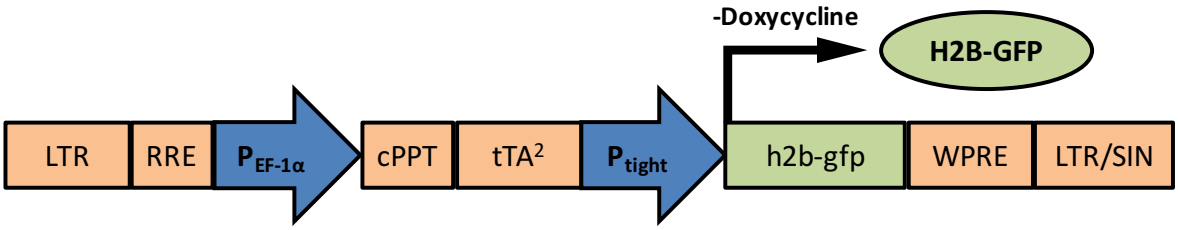

Figure 3.1 Schematic representation of the Tet-Off H2B-GFP lentiviral vector.

In the absence of doxycycline, EF-1α drives the expression of tTA^2 enabling thus the production of the H2B-GFP fusion protein. $P_{EF\text{-}1\alpha}$, elongation factor 1α promoter; tTA^2 and P_{tight}, Tet trans-activator and promoter, respectively; LTR/SIN, self inactivating retroviral long terminal repeat sequences; RRE, HIV-1 Rev response element; cPPT, central polypurine tract; WPRE, *Woodchuck* hepatitis virus post-transcriptional regulation element. Adapted from Falkowska-Hansen, Kollar et al., Exp Cell Res, 2010.

3.2.5 *In vivo* bromodeoxyuridine (BrdU) incorporation

Exposure to bromodeoxyuridine (BrdU; a thymidine analogue) *in vitro* or *in vivo* can be used to label newly synthesized DNA. Immunostaining of the incorporated BrdU allows identification of the fraction of actively dividing cells (Gratzner and Leif 1981). To study

in vivo cell-cycle kinetics of our H2B-GFP expressing PDX models, after a 10-day chasing period with doxycycline, we provided a short pulse of BrdU to both doxycycline-treated and control (untreated) leukaemic mice. The mice were injected twice IP, at a 6-hour interval, with 1 mg of BrdU (stock solution: 10 mg/ml in sterile PBS, BD) and sacrificed 6 hours after the second injection (12-hour pulse). BM and SPL mononuclear cells were collected from the animals and stained with antibodies specific for BrdU, as well as for other surface and intracellular antigens, for subsequent flow cytometry analysis (see section 3.4)

3.3 Mammalian cell culture

All mammalian cell cultures were maintained at 37°C, 5% CO_2 and 20% O_2 and handled according to the principles of aseptic technique.

3.3.1 Lentivirus production

3.3.1.1 Packaging cell line and plasmids

HEK293T cells, or simply 293Ts, are very commonly used for the production of lentiviruses due to their high transfection efficiency (Pear, Nolan et al. 1993). They carry a temperature sensitive mutant of SV-40 large T-antigen (tsA1609neo) and were derived from the human primary embryonic kidney 293 cell line, which was originally established by stable transfection with Ad5 sheared DNA (Graham, Smiley et al. 1977, DuBridge, Tang et al. 1987). The 293T cells were cultured in DMEM (Lonza) supplemented with 10% FBS, SA (Microgem), 2 mM L-Glutamine (Lonza), 100 U/ml penicillin and 100 U/ml streptomycin (Pen/Strep stock, Lonza).

For the production of 3rd generation lentiviral particles (Dull, Zufferey et al. 1998) we co-transfected 4 plasmids into 293T cells:

- **pMD2.G:** envelope plasmid (Addgene #12259) encoding the G protein of vesicular stomatitis virus (VSV-G), which allows targeting of a wide variety of mammalian cells.
- **pMDLg/pRRE:** 3rd generation packaging plasmid (Addgene #12251) containing the *gag* and *pol* coding sequences and the HIV-1 Rev responsive element (RRE).
- **pRSV-Rev:** 3rd generation packaging plasmid (Addgene #12253) encoding the HIV-1 Rev shuttle protein, which promotes the nuclear export of viral pre-mRNA by binding to the RRE.
- **pH2BGFP**: the transfer vector harbouring the Tet-Off H2B-GFP expression cassette, flanked by long terminal repeats (LTRs) that facilitate host genome integration (Figure 3.1). To improve safety, a deletion in the 3' LTR renders the virus "self-inactivating" (SIN) after integration (Zufferey, Dull et al. 1998).

3.3.1.2 Plasmid DNA isolation (maxiprep)

Plasmid DNA was purified from large-scale bacterial cultures with the use of the NucleoBond® Xtra Maxi kit (MACHEREY-NAGEL), according to the manufacturer's instructions. Cells were resuspended, lysed in alkaline conditions and subsequently neutralised with the addition of appropriate buffers (Birnboim and Doly 1979). Lysates were centrifuged at ≥ 5,000 x g for 10 minutes in order to precipitate cell debris, protein-SDS complexes, chromosomal DNA and high molecular weight RNA. Supernatants, containing the plasmid DNA, were then transferred to a NucleoBond® Xtra Column, included in the kit, and cleared by the specially designed column filter. Next, plasmid DNA was bound to the NucleoBond® Xtra Silica Resin and, after a washing step, eluted in a high-salt concentration buffer. Finally, the eluted DNA was precipitated with the addition of isopropanol and the resulting pellet re-dissolved in a low volume of dH_2O. The purified plasmids were quantified on the NanoDrop 1000 spectrophotometer (Thermo Scientific)

and appropriate diagnostic restriction enzyme digestions were performed, before any use in further applications.

3.3.1.3 Transient transfection of 293T cells

The H2B-GFP lentivirus was produced by transient transfection of 293T cells, using the calcium phosphate precipitation method (Barde, Salmon et al. 2010), adapted for 3rd generation vectors. In detail, low passage 293T cells were seeded in 10 cm plates to obtain sub-confluent (~70%) cultures. Then, for each 10 cm plate the following mix was prepared:

- pMD2.G: 3 µg
- pMDLg/pRRE: 5 µg
- pRSV-Rev: 2.5 µg
- pH2BGFP: 10 µg
- $CaCl_2$ (2M) 61 µl
- sterile dH_2O to 500 µl

The mix was added drop-wise to 500 µl of 2X HBS (HEPES buffered saline: 250 mM HEPES pH 7.0, 250 mM NaCl and 150 mM Na_2HPO_4), constantly bubbling with the pipette aid. After 15 minutes of incubation at room temperature, the precipitate was added drop-wise to the 10 cm plates and immediately mixed with the culture medium by gentle swirling. The cells were incubated at 37°C for 8-12 hours, after which the transfection medium was replaced with fresh 293T medium. The viral supernatant was collected twice, 24 and 48 hours later. Both harvests, pooled together, were filtered through a 0.22 µm filter unit and concentrated 1,000x by ultracentrifugation at 50,000 x g for 2 hours at 16°C. The resulting viral stock was immediately stored at -80°C.

3.3.1.4 Virus titration

The functional titre was determined for each batch of viral stock independently, by infection of 293T cells with serial dilutions of the frozen stock corresponding to 0.0001-1 µl of virus (Barde, Salmon et al. 2010). More precisely, 250,000 cells *per* well were plated in a 12-well plate in 0.5 ml of medium and transduced in duplicate with 1, 0.1, 0.01, 0.001 or 0.0001 µl of concentrated virus. The cells were collected 72 hours post-infection, washed and fixed in a 4% formaldehyde solution (diluted in PBS from 37% stock, VWR). The percentage of the GFP+ infected cells was estimated by flow cytometry analysis and the titre was, then, calculated with the aid of the equation below, considering only the dilutions of viral stock which yielded 1-20% GFP+ cells:

$$Titre\ (\frac{TU}{\text{ml}}) = \frac{\text{Number of target cells (count at day 1)} x [\frac{\%\ positive\ cells}{100}]}{volume\ of\ vector\ (ml)}$$

The titre was finally reported as transducing units *per* ml (TU/ml) and typically ranged between 10^8-10^{10} TU/ml.

3.3.2 Isolation of human AML blasts from patient derived xenografts

Human AML blasts were recovered from the BM and SPL of the engrafted mice and the resulting cell suspensions were filtered through a 100 µm cell strainer in order to remove aggregates. When required, mononuclear cell isolation was performed using Ficoll-Paque PLUS (GE Healthcare Life Sciences) separation media, which allows for segregation of different blood cell types in distinct layers due to differential cell migration during centrifugation (Boyum 1968). In brief, single cell suspensions of BM and SPL cells in PBS were layered carefully over an equal volume of Ficoll-Paque and centrifuged at 400 x g for 30 minutes at 16°C, without brake. Highly purified, viable mononuclear cells were

then harvested from the interphase between the Ficoll-Paque layer and the PBS. After two washing steps with PBS, followed by centrifugation at 400-500 x g for 10 minutes, in order to remove residual Ficoll-Paque and platelets, the cells were finally resuspended in the appropriate medium for subsequent culture, long-term storage or other applications.

3.3.3 Cell culture of human AML blasts

Human leukaemic blasts coming from different PDX models presented diverse culture requirements. They were generally cultured in RPMI medium supplemented with 20% North-American FBS (FBS, NA), variable concentrations of 5637 (bladder carcinoma cell line)-conditioned medium (5637 CM; 1-20%) (Myers, Katz et al. 1984, Quentmeier, Zaborski et al. 1997) and/or 1% insulin-transferrin-sodium selenite (ITS) liquid media supplement (Table 3.1).

Table 3.1 Composition of the cell culture media used for different AML xenografts.

Component	Manufacturer	AMLIEO20	AML1	AML5	AML9
RPMI / 1640	Lonza	yes	yes	yes	yes
FBS, NA	HyClone™	20%	20%	20%	20%
5637 CM	Produced in-house	1%	1%	10%	20%
ITS (100x)	Sigma-Aldrich	-	-	1%	1%
Gentamycin (50mg/ml)	Lonza	-	0.2%	-	-

3.3.4 Lentiviral transduction of human AML blasts

Lentiviral transduction of the human AML blasts with the H2B-GFP vector was performed by a single round of spin-infection at 750 x g, for 45 minutes at room temperature. For the clonal tracking experiments, the human blasts were infected with a lentiviral library consisting of $30x10^6$ different barcodes, cloned in the pRSI vector (provided by Cellecta). Pilot experiments with the empty pRSI vector led to the optimization of the spin-infection protocol for the barcoded library at 1,200 x g for 90 minutes at 37°C. Cells were typically

plated at a density of 0.5-1x10^6 cells *per* well in 24-well non-tissue culture treated plates (optionally pre-coated with RetroNectin®, Takara) in the presence of 8 µg/ml of Polybrene (Sigma-Aldrich). Transduction efficiency was measured by flow cytometry analysis of GFP positive cells for the H2B-GFP vector or TagRFP positive cells for the barcoded library.

3.4 Flow cytometry analysis and fluorescence activated cell sorting

Fluorescence activated cell sorting (FACS) was performed using FACSAria™ (BD) and MoFlo® Astrios™ (Beckman Coulter), in order to isolate either H2B-GFP+ cells after the transduction *in vitro* with the Tet-Off lentiviral vector or label-retaining cells (LRCs) after chasing *in vivo* with doxycycline. BD FACSCalibur™ or FACSCanto™ II were used instead for multi-parametric and cell-cycle flow cytometry data acquisition. Finally, MoFlo® Astrios™ was also used for the detection of TagRFP+ cells, infected with the barcoded library or empty pRSI vector. All flow cytometry data were analysed with the aid of the FlowJo™ 8.8.7 platform or later versions.

3.4.1 Engraftment and surface marker expression analysis

Human leukaemic cells derived from the PB, SPL and BM of mice transplanted with H2B-GFP expressing blasts were assessed for the expression of the fusion protein, TagRFP and haematopoietic cell surface markers. Cells were incubated at 4°C for at least 30 minutes in blocking buffer (10% BSA in PBS) and then stained with antibodies specific for human CD45 (hCD45; APC conjugated, clone 2D1, BD), human CD34 (hCD34; PE conjugated, clone 8G12, BD) and human CD38 (hCD38; PerCP-Cy5.5 conjugated, clone LS198-4-3, Beckman Coulter). For the staining, the cells were incubated at 4°C for one hour at a concentration of 10^6 cells *per* 50 µl of staining buffer (1% BSA in PBS) containing the

specific antibodies at a 1:100 dilution. After a washing step with PBS, the cells were fixed in 4% formaldehyde and analysed by flow cytometry.

3.4.2 Cell-cycle analysis

Cell-cycle analysis was performed by staining with a dye that binds to total DNA (Hoechst 33342, Sigma-Aldrich) coupled with immunofluorescent staining for the intracellular Ki-67 antigen and/or incorporated BrdU after *in vivo* labelling (see section 3.2.5).

3.4.2.1 DNA content analysis

Hoechst 33342 binds to AT-rich double-stranded DNA in both live and fixed cells (Arndt-Jovin and Jovin 1977). In our experimental setting, we labelled fixed cells overnight at 4°C in a 2 µg/ml solution in PBS. After data acquisition on FACSCanto^{TM}II, we performed cell-cycle analyses using the Watson Pragmatic (Watson, Chambers et al. 1987) univariate model on FlowJo. In brief, the G0/G1 and G2/M peaks were approximated by Gaussian distributions, setting the ratio of their respective means to 1:2. The S-phase was then fitted with a polynomial function to map exactly the shape of the experimental data. Once models had been fitted for all the samples, we created gates comprising the G0/G1, S-phase and G2/M cell populations accordingly in order to assess other parameters (*e.g.* GFP expression).

3.4.2.2 Ki-67 analysis

Ki-67 expression is strictly associated with cell proliferation. The antigen is found in the nucleus of cycling cells (G1, S, G2, M cell-cycle phases), but is not be detected during G0 (Gerdes, Lemke et al. 1984). To identify the G0 population in our samples, we stained cells, after fixation and permeabilization (with BD Cytofix-Cytoperm^{TM} buffer), using a Ki-67 antibody (PE-Cy7 conjugated, clone B56, BD) diluted 1:100 in 1X Perm/Wash buffer (BD Perm/Wash^{TM} buffer 10X diluted in dH_2O) and incubating 1 hour on ice. Finally, after a washing step with 1X Perm/Wash buffer, the cells were stained for total

DNA levels with Hoechst 33342, as previously described (see section 3.4.2.1), and kept at 4°C until FACS analysis.

3.4.2.3 BrdU incorporation analysis

Evaluation of the DNA synthetic activity, by BrdU incorporation analysis coupled with total DNA content staining, enhances the resolution of S-phase from G0/G1 and G2/M events. For this purpose, ~10^7 BrdU pulsed leukaemic blasts (see section 3.2.5) were optionally stained for cell surface antigens (see section 3.4.1) an then fixed and permeabilized with BD Cytofix/Cytoperm™ and Cytoperm™ Permeabilization Plus buffers. After an additional fixation step with BD Cytofix/Cytoperm™, cells were treated with DNase (diluted to 150 μg/ml in PBS) for 1 hour at 37°C to expose the incorporated BrdU. We stained the cells with an anti-BrdU antibody (APC BrdU Flow kit, BD; diluted 1:200 in 1X Perm/Wash buffer) and for other intracellular antigens (*i.e.* Ki-67, see section 3.4.2.2), incubating 1 hour on ice. Finally, the cells were stained for total DNA levels with Hoechst 33342, as previously described (see section 3.4.2.1), and kept at 4°C until FACS analysis.

3.4.3 Proliferation analysis

The conditional repression of H2B-GFP expression, upon doxycycline treatment *in vitro* or *in vivo*, resulted in a gradual loss of the intensity of the fluorescent signal in proliferating cells. In fact, in the presence of doxycycline, at each cell division, the H2B-GFP protein is expected to be equally distributed to the two daughter cells, allowing thus monitoring of the number of cell divisions that any given cell has undergone. Therefore, PB, BM and SPL samples from both doxycycline-treated and control mice were collected at regular time intervals during the chasing period and analysed by flow cytometry. The histogram of the fluorescence intensity of the GFP+ cells was then generated using FlowJo.

FlowJo's proliferation platform was used for the modelling of cell divisions, starting from the original population (defined by untreated controls) and searching for peaks with

diminishing fluorescence with an approximate ratio of 0.5 *per* generation and peak coefficient of variation (peak CV) around 4-7%. Each cell generation could be then analysed separately by creating the corresponding gates through the proliferation platform.

The overall goodness of fit of the model was finally assessed visually and by the mean root mean square (mRMS) value, which is a measure of the distance of the composite model line from the histogram of the data. The fitted data were further described by the following statistics:

- **Division Index:** the average number of cell divisions that a cell in the original population has undergone (including the undivided cells).
- **Proliferation Index:** the average number of divisions that a dividing cell in the original population has undergone.
- **% Divided**: the percentage of original cells that have divided.

3.5 *In vivo* clonal tracking

A lentiviral barcode library (CellTracker^{TM}, Cellecta) was used to track the clonal composition of growing leukaemias *in vivo*. Our experimental strategy was based on the stable integration (*via* lentiviral transduction) of a unique molecular marker (barcode) in the genome of single LSCs. Sequencing of barcodes in the DNA of leukaemias developed after transplantation of library-labelled samples allowed us to track the progeny of individual cells and estimate the relevant contribution of each clone to tumour formation. The pRSI vector, in which the libraries are cloned, contains a TagRFP transduction reporter and the PuroR selection marker (Figure 3.2). Furthermore, each library construct is unequivocally identified by a specific barcode, which consists of two consecutive 18-nucleotide degenerated sequences (Figure 3.2). Due to the presence of 13,000 x 13,000 different combinations of the two 18-nucleotide sequences, the complexity of the clonal tracking libraries reaches approximately the number of $30x10^6$ unique barcodes.

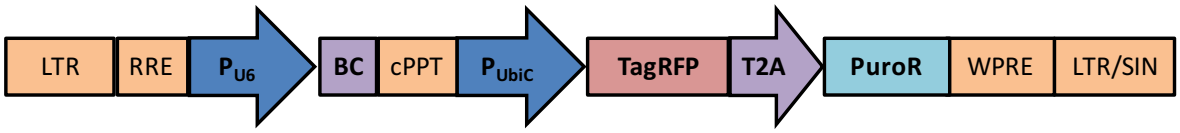

Figure 3.2 Schematic representation of the pRSI lentiviral vector (Cellecta).

P_{U6}, human U6 promoter; BC, barcode; P_{UbiC}, Ubiquitin C promoter drives expression of TagRFP and PuroR; TagRFP, monomeric red fluorescent protein (Evrogen) as transduction reporter; PuroR, puromycin-resistance marker for the selection of transduced cells; T2A, *Thosea asigna* virus 2A translational cleavage site containing 18 amino acid residues; LTR/SIN, self inactivating retroviral long terminal repeat sequences; RRE, HIV-1 Rev response element; cPPT, central polypurine tract; WPRE, *Woodchuck* hepatitis virus post-transcriptional regulation element.

3.5.1 Genomic DNA extraction for barcode amplification

Both uninfected and H2B-GFP infected human AML blasts were targeted *in vitro* with the lentiviral barcode library (see section 3.3.4) at a low multiplicity of infection (MOI) and transplanted into NSG mice after a 24-hour incubation at 37°C. Upon leukaemia presentation, BM and SPL cells were collected from the engrafted mice and the genomic DNA was extracted, either from the bulk tumours or from sorted quiescent and cycling populations. For this purpose a phenol/chloroform extraction protocol was used (Sambrook and Russell 2006) followed by isopropanol precipitation according to Cellecta's specifications.

In detail, tumour cells were resuspended in TES buffer (10 mM Tris-Cl pH 8.0, 5 mM EDTA pH 8.0, 0.5% SDS) containing RNase A (100 µg/ml) and incubated for 5 min at room temperature. One volume of phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma-Aldrich) was added to the samples, mixed thoroughly by vortex until the formation of an emulsion and centrifuged at maximum speed for 60 minutes at room temperature. After centrifugation, the upper aqueous phase was collected and the process repeated with

the addition of 1 volume of chloroform (VWR) to remove phenol residues. DNA was then precipitated from the upper phase, by adding 1 volume of isopropanol and 0.125 volumes of 3 M sodium acetate, washed with 70% ethanol and air-dried. Pellets were finally resuspended in sterile dH_2O and the DNA concentration was measured on NanoDrop or Qubit® 2.0 (Invitrogen).

3.5.2 Barcode amplification

Barcode amplification was performed by two rounds of PCR, using nested primers designed to include sequences complementary to the immobilized primers necessary for generating amplification clusters in Illumina HiSeq 2000 (Table 3.2, Figure 3.3).

Table 3.2 Primers used for barcode amplification and sequencing.

Application	Round	Name	Sequence
Barcode amplification	1st	FwdHTS3	TCGGATTCAAGCAAAAGACGGCATA
		R2	AGTAGCGTGAAGAGCAGAGAA
	2nd	Gex1-Bpi	TCAAGCAGAAGACGGCATACGAAGACA
		NR2	AATGATACGGCGACCACCGAGACGAGCAC CGACAACAACGCAGA
Sequencing		GexSeqS	AGAGGTTCAGAGTTCTACAGTCCGAA

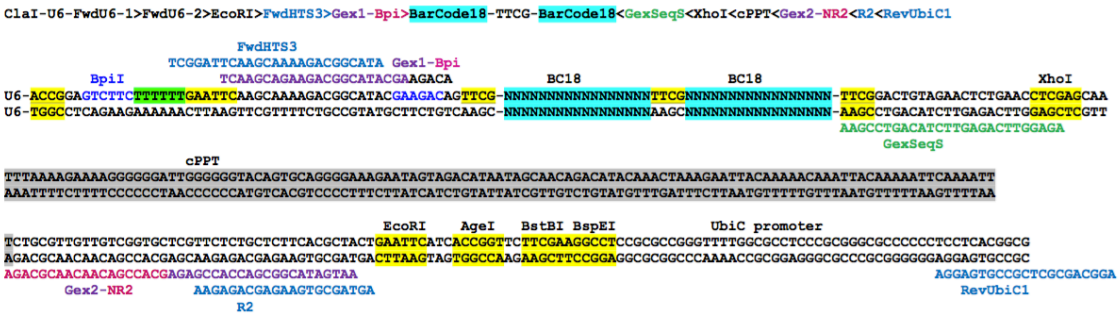

Figure 3.3 Nested PCR strategy for barcode amplification

In the first PCR round, the FwdHRS3 and R2 primers are employed to produce a 275bp amplicon, which is used as a template in the second PCR round to amplify a ready-to-sequence 267 bp fragment with the the Gex-Bpi and Gex2-NR2 primers.

For the first PCR round, we used as input the entire amount of genomic DNA obtained, to ensure full representation of the library of barcodes in our samples. For this purpose,

Titanium *Taq* DNA polymerase (Takara), which is very efficient in the amplification of high amounts of DNA, was used for the bulk tumour samples and Phusion® High-Fidelity DNA polymerase (New England Biolabs) was used for the sorted populations. For each sample, a master mix was prepared according to Table 3.3 and the PCR performed following the programme outlined in Table 3.4.

Table 3.3 Master-mix preparation for PCR round 1.

High DNA input (25µg/reaction)		**Low DNA input (250ng/reaction)**	
Component	**Final concentration**	**Component**	**Final concentration**
Titanium Taq buffer (10X)	1X	5X Phusion HF buffer	1X
dNTP mix (10 mM each)	200 µM each	dNTP mix (10 mM each)	200 µM each
FwdHTS3 primer (10 µM)	0.3 µM	FwdHTS3 primer (10 µM)	0.5 µM
R2 primer (10 µM)	0.3 µM	R2 primer (10 µM)	0.5 µM
dH_2O	(to 100 µl)	dH_2O	(to 50 µl)
Titanium Taq polymerase (50X)	5U/ 100 µl	Phusion DNA polymerase	1U/ 50 µl

Table 3.4 Thermocycling conditions for PCR round 1

Extension temperature was set to 72°C or 68°C for Titanium *Taq* and Phusion DNA polymerase amplifications accordingly.

Process	**Temperature**	**Time**	**Cycles**
Initial denaturation	94°C	3 min	1
Denaturation	94°C	30 sec	16
Annealing	60°C	10 sec	
Extension	72°C/68°C	20 sec	
Final extension	72°C/68°C	2 min	1

After the first round of PCR, all the reactions *per* sample were combined together and extensively mixed. A small sample of this mix (typically 1-10 µl) was used to perform the

second round nested-PCR, in order to separate the amplified barcodes from non-specific PCR products and the excess of genomic DNA. The primers used for this second amplification contain the P5 and P7 adapter sequences for the Illumina flow cell (Table 3.2, Figure 3.3). For each sample, a master mix was prepared according to Table 3.5 and the PCR performed following the programme outlined in Table 3.6.

Table 3.5 Master-mix preparation for PCR round 2.

High DNA input (25µg/reaction)		**Low DNA input (250ng/reaction)**	
Component	**Final concentration**	**Component**	**Final concentration**
Titanium Taq buffer (10X)	1X	5X Phusion HF buffer	1X
dNTP mix (10 mM each)	200 µM each	dNTP mix (10 mM each)	200 µM each
Gex1-Bpi primer (10 µM)	0.5 µM	Gex1-Bpi primer (10 µM)	0.5 µM
NR2 primer (10 µM)	0.5 µM	NR2 primer (10 µM)	0.5 µM
dH_2O	(to 100 µl)	dH_2O	(to 50 µl)
Titanium Taq polymerase (50X)	5U/ 100 µl	Phusion DNA polymerase	1U/ 50 µl

Table 3.6 Thermocycling conditions for PCR round 2.

Extension temperature was set to 72°C or 68°C for Titanium *Taq* and Phusion DNA polymerase amplifications accordingly.

Process	**Temperature**	**Time**	**Cycles**
Initial denaturation	94°C	3 min	1
Denaturation	94°C	30 sec	16
Annealing	66°C	10 sec	
Extension	72°C/68°C	10 sec	
Final extension	72°C/68°C	2 min	1

The PCR products (10µl *per* reaction) were analysed by gel-electrophoresis on a 3.5% agarose gel in 1X TAE (1X Tris-acetate-EDTA buffer) to confirm the correct size (267 bp)

and equal densities of the amplicon bands for all samples. When necessary, we repeated the second PCR, adjusting the amount of input DNA from the first round to obtain the same yields among all amplified samples. The products of two 2nd round reactions were then combined and purified using the QIAquick PCR purification kit (QIAGEN), according to the manufacturer's indications. When non-specific PCR products had been detected during the electrophoresis, we used the QIAquick Gel Extraction kit (QIAGEN) to isolate the 267 bp target DNA fragments after the PCR purification step. As described in the manufacturer's protocol, the purified PCR products were separated by electrophoresis on a 3.5% agarose gel in 1X TAE and the desired bands were excised using a disposable sterile scalpel. The gel slices were dissolved in the designated buffer included in the kit and after the addition of isopropanol, the DNA was bound to the silica membrane of the provided spin columns and washed with an ethanol containing buffer. Finally, the DNA was eluted in a small volume (30-50 μl) of dH_2O and quantified using Qubit.

3.5.3 Barcode sequencing and enumeration

Since the lentiviral barcode library and the PCR primers are complementary to the immobilized P5 and P7 primers (blue and red respectively in Figure 3.3) necessary for generating amplification clusters in Illumina's HiSeq 2000, the amplicons can be directly sequenced through the GexSeqS primer (Table 3.2). The barcodes were identified by aligning each sequencing read to the reference sequences of the library barcodes (provided by Cellecta) without mismatches using Bowtie (Langmead, Trapnell et al. 2009), enabling thus the conversion of raw NGS data into number of reads for each barcode.

3.6 Mutational analysis

3.6.1 Whole exome sequencing (WES)

High quality genomic DNA was obtained from bulk leukaemias and sorted cycling or quiescent populations with the use of the DNeasy® Blood and Tissue kit (QIAGEN). In detail, the cells were resuspended in the lysis buffer containing RNase A (100 µg/ml) and Proteinase K (75 µg/ml) and incubated at 56°C for 20 minutes. After the addition of ethanol (96-100%), the lysates were loaded on the microspin columns and the DNA was selectively bound to the silica-based membrane. Contaminants and enzyme inhibitors were removed by two washing steps and the DNA was finally eluted in the appropriate volume of dH_2O (50-100 µl).

Starting from 200 ng fractionated DNA *per* sample, the whole-exome of isolated populations and bulk tumour cells was captured with the use of the SureSelect^{XT} Human All Exon kit (Agilent Technologies). Paired-end sequencing at high coverage (threshold >1-2%) with 101 nt read length was performed using the Illumina HiSeq 2000 platform. Data were analysed using an appropriately developed bioinformatics pipeline (see section 3.6.2) based on MuTect (Cibulskis, Lawrence et al. 2013) for the identification of single nucleotide variations (SNVs).

3.6.2 Mutation calling

Although exome capture sequencing is highly species-specific, the presence of contaminating murine stromal cells in our samples could impinge on the sensitivity and specificity of NGS analysis (Rossello, Tothill et al. 2013). For this reason, we used Xenome (Conway, Wazny et al. 2012) to separate reads of the human graft (hg19) from the ones of the murine host (mm9). Graft-specific reads were then aligned to the hg19 reference genome using BWA (Li and Durbin 2009). Next, we performed the following NGS data pre-processing steps according to GATK best practices (McKenna, Hanna et al.

2010): local realignment, duplicate marking and base quality recalibration. SNV identification was performed by MuTect (Cibulskis, Lawrence et al. 2013), comparing each tumour sequence to the corresponding normal. All "REJECT" mutations were discarded and we considered only mutations in the target region. Finally, we used the D-ToxoG tool (Costello, Pugh et al. 2013) to remove artefacts introduced during library preparation. All positions with a minimum coverage of 10 reads (for both the tumour and normal samples of each confrontation pair) were annotated with ANNOVAR (Wang, Li et al. 2010). Finally, the frequency of each mutation was calculated dividing the number of reads supporting the variant to the total coverage at the variant site.

3.6.3 Clonal analysis

For each patient, a unique file was built containing all the positions identified as mutated in at least one of the sequenced samples. For each mutated position a Python script was used to re-count the reads (carrying or not the variant) directly from the BAM files. We excluded variants in non-coding regions, variants present in highly repetitive regions, variants annotated as single nucleotide polymorphisms (SNPs) [*i.e.* annotated in ExAC (Lek, Karczewski et al. 2016), ESP (http://evs.gs.washington.edu/EVS) or 1KG (http://www.internationalgenome.org/home) with Minor Allele Frequency (MAF) higher than 0.01 (MAF>0.01)] and transgene-specific variants (introduced by the H2B-GFP vector). SciClone (Miller, White et al. 2014) was employed to cluster exonic mutations by variant allele frequency (VAF) similarity, using Bayesian mixture modelling of beta distributions (minimum depth = 10, maximum number of clusters = 10). ClonEvol (Dang, White et al. under review) was then used to infer clonal evolution models (p-value ≤ 0.05) from the clusters of variants (sub-clones) previously identified by SciClone. Finally, the inferred models were visualized with the fishplot R package function (Miller, McMichael et al. 2016).

4 Results

4.1 LSCs have variable growth potential *in vivo*

AMLs are functionally heterogeneous and hierarchically organized, containing only a minor fraction of cells with leukaemogenic capacity (Bonnet and Dick 1997). These cells, commonly termed as leukaemia stem cells (LSCs) or leukaemia initiating cells (LICs), are thought to be responsible for tumour maintenance and progression. We first investigated whether the pool of LSCs is inherently heterogeneous in terms of growth potential *in vivo*. To address this question, we used a clonal tracking approach, envisioning the identification of each LSC by a unique molecular marker (barcode), which is stably integrated in the genome of the target cells by lentiviral transduction of a library of molecularly distinct barcodes (see section 3.5).

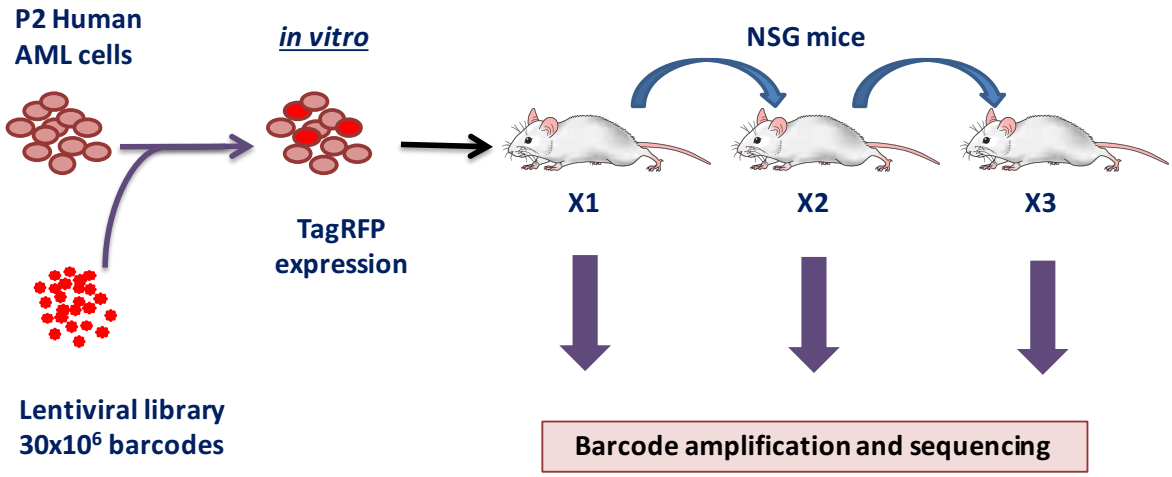

Figure 4.1 Experimental strategy for the *in vivo* clonal tracking.

Freshly isolated leukaemic cells, derived from xenotransplanted human samples, were infected at a low MOI with a lentiviral library containing $30x10^6$ unique molecular barcodes. The infected blasts were then serially transplanted into NSG mice. After leukaemia development, the recipients were sacrificed and the spleen and bone marrow cells of each passage (X1 to X3) were collected for barcode amplification and sequencing.

Limiting dilution and serial transplantation assays have been considered as the gold standard for the identification and enumeration of both normal and cancer stem cells (CSCs). However, the clonal tracking approach involves transplantation of a mixed population of CSCs, thus permitting monitoring of individual CSCs in a more physiological context. Sequencing of the barcodes in the tumours originating from transduced cells (X1) and cells from serially transplanted (secondary X2 and tertiary X3) recipient mice, should enable us to track the progeny of each individual CSC and estimate the relevant contribution of each leukaemic clone to tumour formation and evolution (Figure 4.1).

4.1.1 Infection of human AML blasts with the 30x10^6 barcode library

Our group has established a collection of patient derived xenografts (PDX) from 33 primary human AMLs (passage 1, P1), of which 16 were successfully passaged in secondary recipients (passage 2, P2). For the purposes of this study, P2 freshly isolated cells were used for the *in vitro* transduction with the 30x10^6 barcode lentiviral library, which was followed by xenotransplantation to monitor the contribution of individual LSCs in tumour growth *in vivo*.

To avoid multiple viral integrations *per* cell, we first defined the appropriate experimental conditions to ensure high infection efficiency of the human AML blasts at a low MOI. A single spin infection at 1,200 x g for 90 minutes at 37ºC was chosen as the most efficient transduction protocol. Importantly, the infected cells presented no apparent deficits in homing and engraftment upon transplantation into NSG mice and the presence of the TagRFP+ blasts could be easily monitored in the PB of the transplanted animals (data not shown).

4.1.2 *In vivo* clonal tracking upon serial passaging

We used a high complexity 30x10^6 barcode library in order to minimize the possibility of marking multiple cells with the same barcode. After the infection step, 5x10^5 cells were

transplanted in NSG recipients, while the remaining were kept for the evaluation of infection efficiency, by flow cytometry analysis.

The use of a library with such a high complexity rendered the discrimination of background sequencing noise extremely challenging. Based on binomial statistics, at least 5 observations (*i.e.* 5 reads *per* barcode) are required to claim a significant difference from 0 (p-value < 0.05). However, in order to set a biologically meaningful threshold, a spike-in experiment was performed (Cammarata A., unpublished), which showed a linear correlation between the numbers of cells marked by a specific barcode present in a sample, and the number of reads obtained for the same barcode. Linearity, however, was lost when a barcode was represented by as few as 10 cells, while the detection of the barcode altogether was not uniformly reliable for less than 10,000 cells. Setting as threshold the number of reads corresponding to the 10,000 cell input, we retrieved an average of ~6.5% of the barcodes in the various experiments, which accounted however for the ~99% of the reads *per* tumour in all cases. In other terms, the background threshold that we used allowed analyses of 99% of the cells of each tumour, with a sensitivity of clonal identification of 10,000 cells *per* clone.

The relative contribution of each LSC to the whole tumour population (X1) *in vivo* was highly variable, ranging from ~0.03% to ~19%, thus suggesting that the LSC compartment is highly heterogeneous in terms of growth potential *in vivo*. However, by comparing the clonal evolution of the same X1 in the different mice of passages X2 and X3 of the serial transplantation (Figure 4.1), we found a strikingly similar clonal composition *per* passage both in terms of clone numbers and clone identity. Moreover, the common clones among the recipients of the same donor always grew to occupy the same proportion of the tumour population, suggesting that growth potential, under a given environmental condition, is an inherent and “stable” property of each clone.

Notably, a strong clonal selection was observed upon serial transplantation under steady state conditions. Only a few clones survived and expanded in X2 and X3 (Figure 4.2). In fact, only a minor proportion of clones (~8%) was able to constantly expand throughout the passages, suggesting that the majority of LSCs are endowed with a limited proliferative potential and that the observed changes in clonal composition in the X2 and X3 tumours reflect the functional exhaustion of individual clones.

The strong clonal selection observed upon serial transplantation could be perturbed by external selective pressures such as 5-FU (in collaboration with Cammarata A., unpublished; Figure 4.2 and Figure 4.3), suggesting that different clones may adapt differently to the changing environment. Importantly, the clones selected by 5-FU maintained their growth potential when challenged by transplantation (in the absence of 5-FU treatment), suggesting that regardless of their adaptability to the changing environment, different clones have similar capacity to fit the selecting environment.

In conclusion, leukaemias, at steady state, are composed by multiple clones, each arising from a single LSC. Different LSCs have different self-renewal and growth potential, equally fit to the external selective pressure of the environment, and different capacities to adapt to a changing environment.

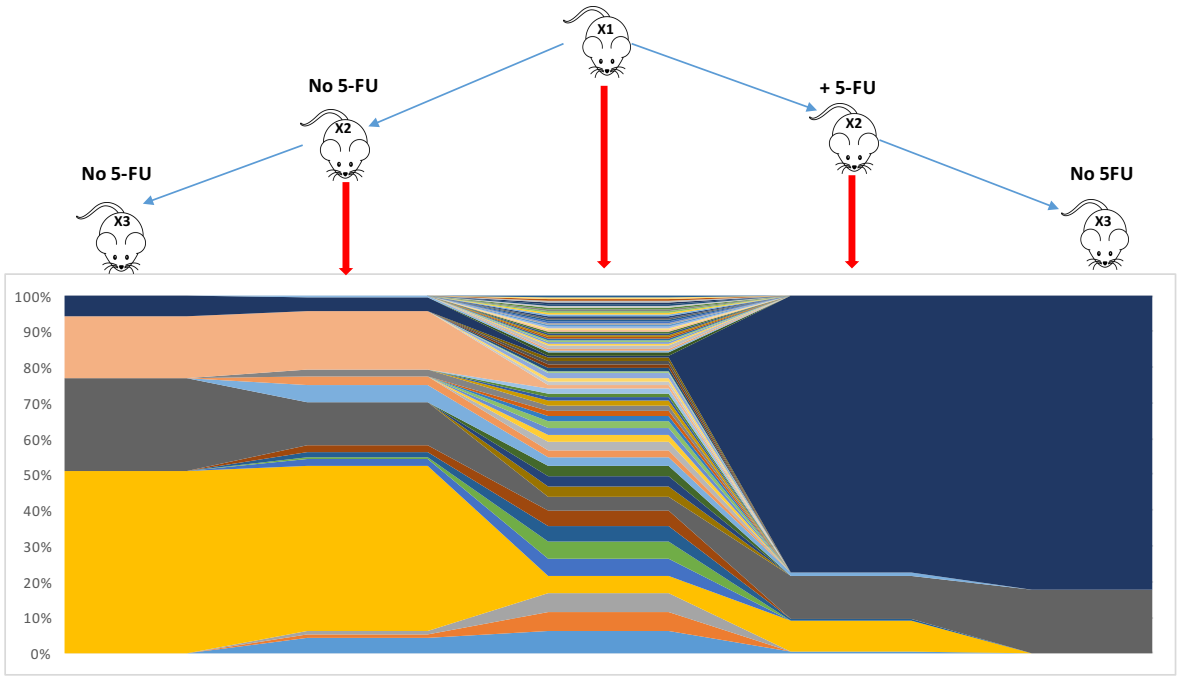

Figure 4.2 Clonal selection upon serial passaging and 5-FU treatment.

The upper scheme depicts the experimental outline and the graph below it shows a representative example of the relative contribution of each barcoded clone to tumour formation, for each passage (X1, X2 and X3) and condition (No 5-FU, + 5-FU). The inferred expansion or shrinkage of clones upon serial passaging is shown below the light blue arrows. All unique barcodes are depicted by a different colour and can be therefore univocally identified in all samples, allowing for the visual observation of their evolution under different environmental stresses (*i.e.* serial transplantation and 5-FU treatment). The clones are presented in an ascending order, based on their relative contribution to X1 (n=5, one representative example is shown). The clonal selection upon serial passaging (No 5-FU) is illustrated on the left (n=4 *per* passage, one representative example is shown for each passage) and the effect of 5-FU administration at X2 (+ 5-FU), which led to a diverse clonal selection, is shown on the right (n=4 *per* passage, one representative example is shown for each passage) (Cammarata A., unpublished).

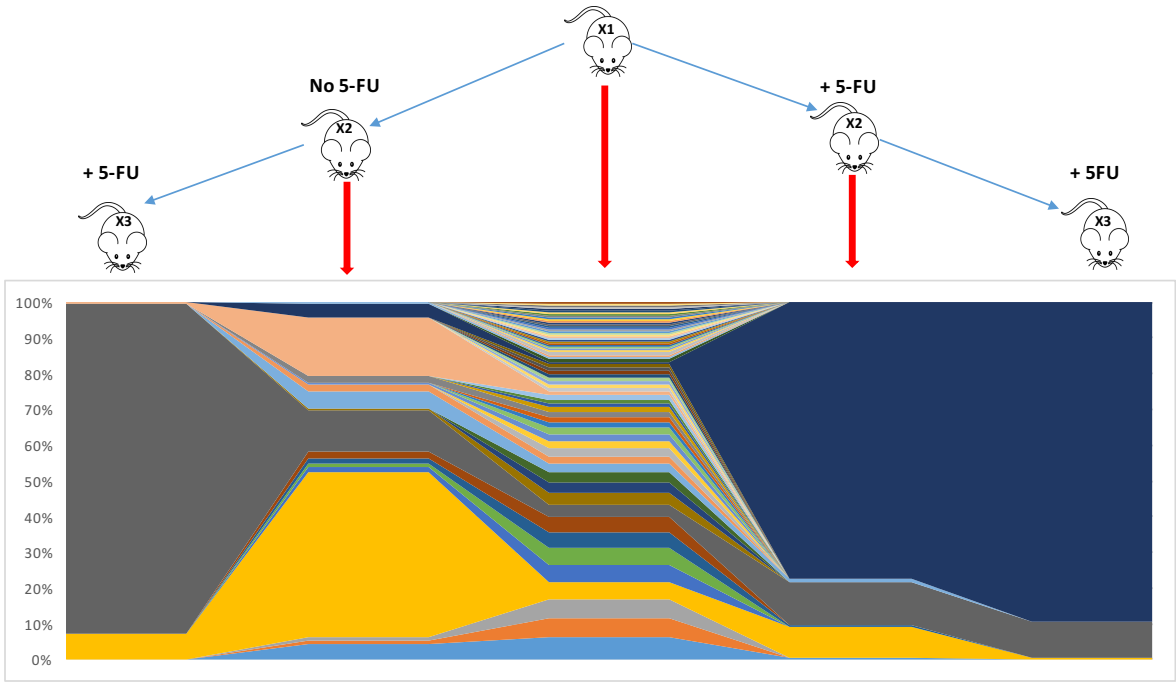

Figure 4.3 Minor clones are selected for expansion upon 5-FU treatment

As in Figure 4.2, the upper scheme depicts the experimental outline and the graph below it shows a representative example of the relative contribution of each barcoded clone to tumour formation, for each passage (X1, X2 and X3) and condition (No 5-FU, + 5-FU). The inferred expansion or shrinkage of clones upon serial passaging is shown below the light blue arrows. The graph shows the effect of 5-FU administration on the clonal selection process both at X2 (+5-FU X2 on the right; the same X1 and X2 cases that were previously shown in Figure 4.2) and X3 (+5-FU X3, one case representative of 4 *per* condition) passages of transplantation of the leukaemia. However, after the first administration of 5-FU, further treatments appeared to be ineffective in the clonal selection process (+5-FU X3 on the right) (Cammarata A., unpublished).

4.2 Functional isolation of leukaemic populations with different proliferation histories

Our *in vivo* clonal tracking data suggest the existence of a pool of functionally heterogeneous LSCs in AMLs, in terms of proliferation capacity (different clone-sizes at X1), self-renewal potential (clonal expansion or extinction upon serial transplantation) and capacity to adapt to a changing environment (different clonal expansion after 5-FU treatment). We next investigated proliferation and self-renewal kinetics of AML cells *in vivo*, using an *in vivo* label-retaining assay based on the inducible expression of the H2B-GFP labelling fusion protein (see section 3.2.4 and Figure 4.4).

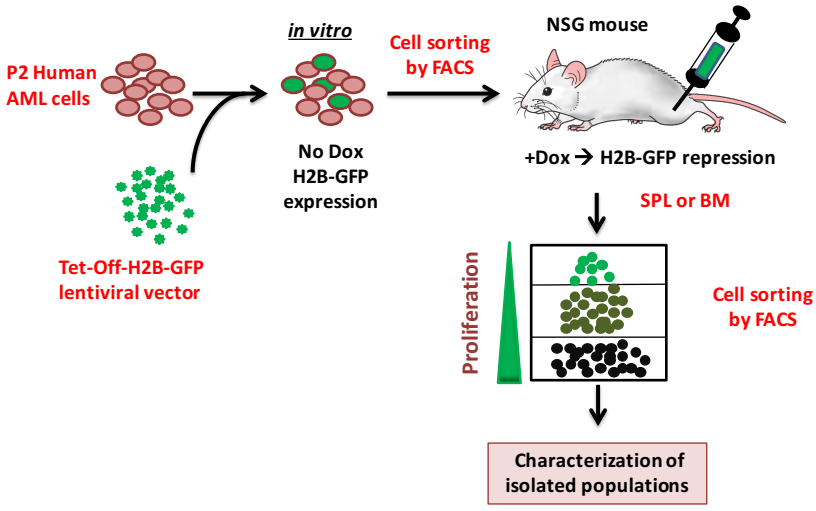

Figure 4.4 Experimental strategy for the *in vivo* H2B-GFP label-retaining assay

Human AML cells, derived from a passage 2 (P2) PDX, were transduced *in vitro* with the Tet-Off H2B-GFP lentiviral vector, in the absence of doxycycline (Dox) in order to allow for the expression of the H2B-GFP transgene. The H2B-GFP+ infected cells were isolated by FACS and transplanted into NSG mice. Upon leukaemia manifestation, the mice were treated with Dox *in vivo* to shut off the H2B-GFP expression (chasing period). At the end of the chasing period, the bone marrow (BM) and spleen (SPL) cells were collected from the mice and FACS-sorted in three subpopulations, based on their GFP fluorescence intensity, for further characterisation.

The experimental strategy outlined in Figure 4.4 is based on the xenotransplantation of human AMLs infected with the H2B-GFP lentivirus and stably expressing the fusion protein in the absence of doxycycline (Falkowska-Hansen, Kollar et al. 2010). The prolonged suppression of H2B-GFP expression by *in vivo* administration of doxycycline (chasing period), enables identification and isolation of leukaemic cell populations endowed with different label-retaining capacity which, in turn, correlates with their cycling properties.

4.2.1 Generation of H2B-GFP expressing human AML xenografts

The first step towards the set-up of the *in vivo* label-retaining assay was to generate PDX models of AML that stably express the H2B-GFP transgene in the absence of doxycycline. For this purpose, P2 leukaemic blasts were targeted *in vitro* by the Tet-Off H2B-GFP lentiviral vector in doxycline-free media, to allow production of the fusion protein. As expected, the GFP signal was restricted to the nucleus of the infected cells (Figure 4.5).

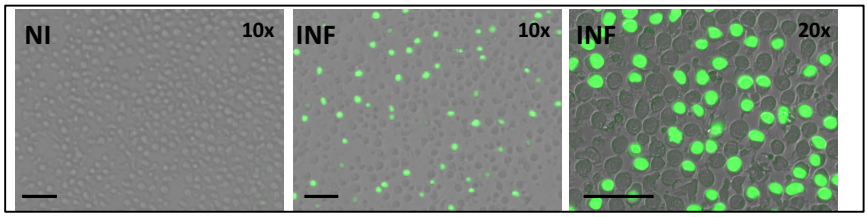

Figure 4.5 Fluorescence microscopy of H2B-GFP expression in human AML blasts.

Left panel: control (non-infected, NI) cells. Middle and right panel: cells infected (INF) with the H2B-GFP lentivirus. Magnification 10x or 20x as indicated. Scale bar represents 50 µm.

Human leukaemic blasts coming from different patient samples typically presented variable culture requirements *in vitro* (see section 3.3.3) and infection efficiencies with the H2B-GFP vector (Figure 4.6). Our aim was to obtain a relatively homogeneous population of leukaemic H2B-GFP expressing cells. We used a relatively low MOI (10 or 20), to avoid extensive cell death, and the percentage of GFP+ cells obtained under these conditions was generally low (< 25%; Figure 4.6).

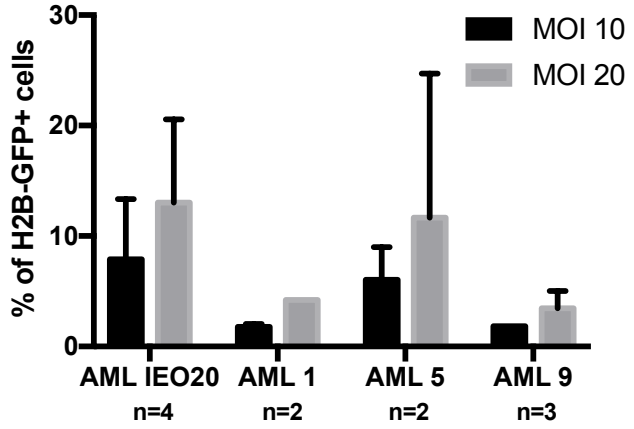

Figure 4.6 Transduction efficiency of human AML blasts with the H2B-GFP vector.

The transduction efficiency at MOI 10 and 20 for different PDXes is represented by the percentage of H2B-GFP+ cells detected by FACS 2-3 days post infection. Mean and standard deviation of biological replicas are shown.

However, we were able to isolate a highly pure H2B-GFP+ population by FACS from all AML samples tested, after a single spin infection, and to transplant from 30,000 to 50,000 sorted cells *per* NSG mouse. Finally, GFP+ leukaemias were obtained from two PDX models, AML IEO20 and AML 9 (Figure 4.7). In the other two samples tested (*i.e.* AML 1 and AML 5) the combination of low infection efficiency with low LSC content was restrictive. Importantly, in the absence of doxycycline, expression of the H2B-GFP transgene in AML IEO20 and AML 9 was stable *in vivo* across all haematopoietic tissues and upon serial transplantation (Figure 4.8).

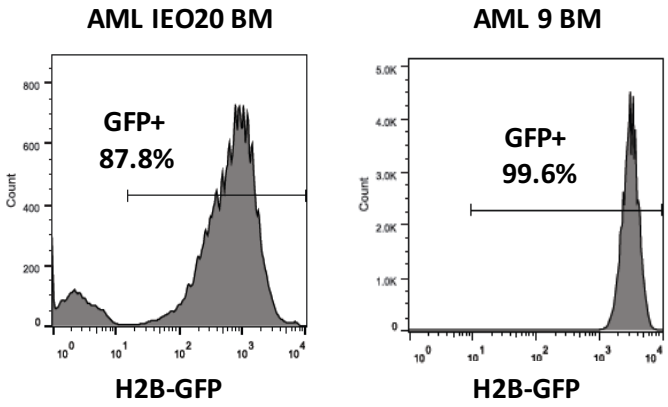

Figure 4.7 Flow cytometry histograms of H2B-GFP expression *in vivo*.

The distribution of the H2B-GFP fluorescence, in the absence of doxycycline, is shown for the human (hCD45+) population in the BM of mice engrafted with H2B-GFP+ expressing AML IEO20 (left panel) and AML9 (right panel) blasts.

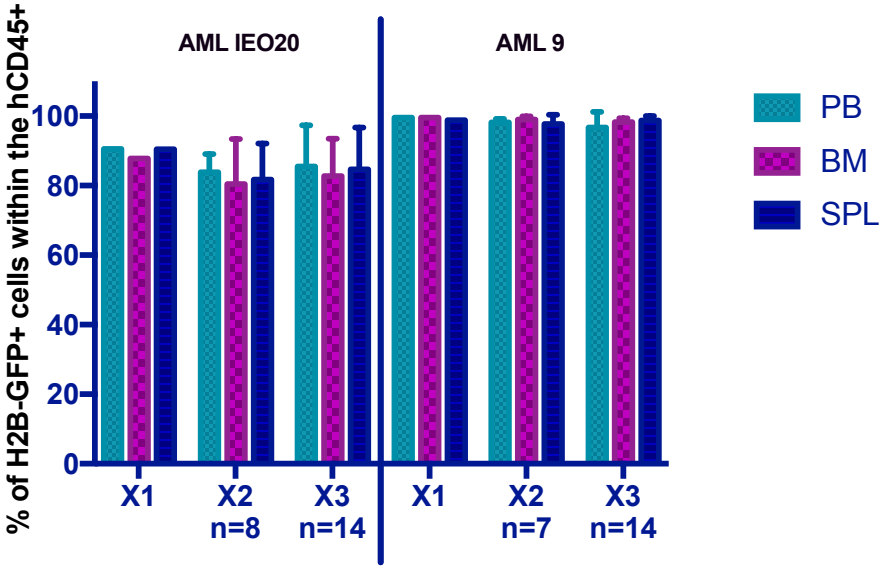

Figure 4.8 H2B-GFP expression *in vivo* in the absence of doxycycline.

Percentage of H2B-GFP+ cells within the human (hCD45+) population in the PB, BM and SPL of mice transplanted with H2B-GFP+ AML IEO20 and AML 9 blasts, as defined by flow cytometry analysis. Mean and standard deviation are shown for the first (X1), second (X2) and third (X3) passages of transplantation of the infected blasts.

4.2.2 H2B-GFP expression is tightly controlled by doxycycline

The design of the H2B-GFP vector offers the possibility to control extrinsically the regulation of transgene expression in the infected cells, by administration of doxycycline *in vitro* or *in vivo*. This feature is of key importance in view of the downstream experimental applications, as loss of the H2B-GFP signal will correlate with cell proliferation rate. As already discussed in the previous paragraph, the levels of H2B-GFP *in vivo* were stable in the absence of doxycycline (Figure 4.8), indicating that vector leakiness and/or transgene silencing did not interfere with our experimental conditions. Preliminary *in vitro* experiments further showed that treatment with doxycycline at a concentration of 10 ng/ml was sufficient to shut off the expression of H2B-GFP, resulting in rapid and time-dependent decrease of the Mean Fluorescence Intensity (MFI) of the GFP+ population (Figure 4.9).

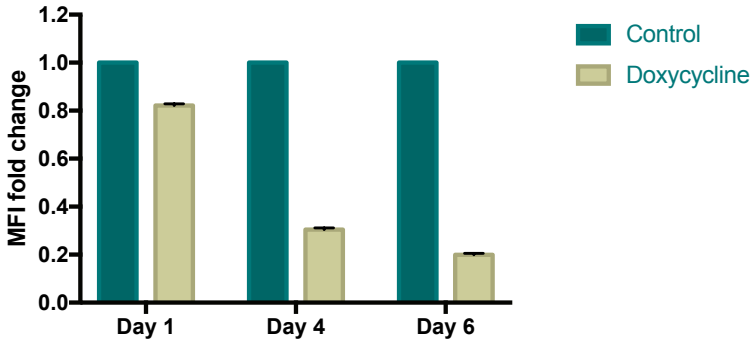

Figure 4.9 Regulation of H2B-GFP expression *in vitro*.

The mean fluorescence intensity (MFI) of H2B-GFP+ leukaemic blasts (AML IEO20) in the presence of 10 ng/ml of doxycycline was monitored by flow cytometry one, four and six days after plating. Control: cells cultured in doxycycline-free media. Doxycycline: cells cultured in the presence of 10 ng/ml doxycycline. All values were normalised to the MFI of the controls *per* day. Mean and standard deviation are shown.

In order to regulate H2B-GFP expression *in vivo*, doxycycline was administered to the experimental mice through the feed, while control animals were kept on a normal diet. The gradual loss of the fluorescent signal of the H2B-GFP tagged human population could be easily monitored in the PB of the treated mice (Figure 4.10).

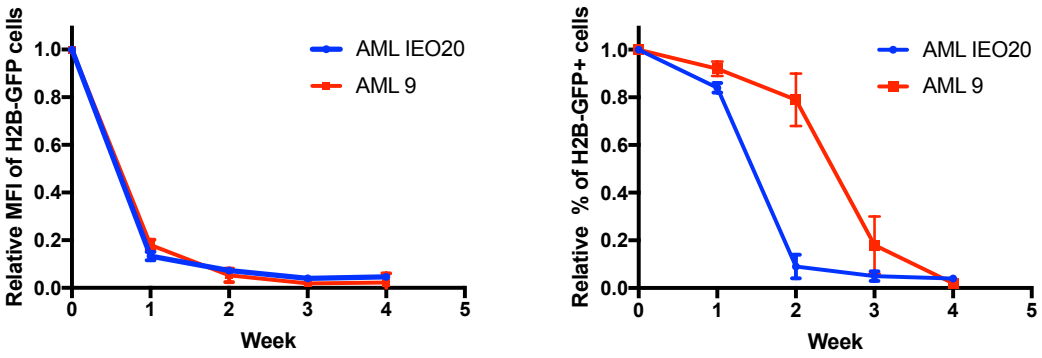

Figure 4.10 Regulation of H2B-GFP expression *in vivo*.

Relative decrease in the MFI (left panel) and in the percentage (right panel) of the H2B-GFP+ cells, as monitored by flow cytometry of the human (hCD45+) cells in the PB of mice (n=5) transplanted with the H2B-GFP+ AML IEO20 and AML 9. All values were normalised to the equivalent MFI and percentages of hCD45+ cells within the GFP+ gate of the control group (*per* PDX and time point). Mean and standard deviation are shown.

Notably, the rate by which the mean fluorescence intensity (MFI) of the H2B-GFP+ population decreased was indistinguishable in the two PDX models used (Figure 4.10, left panel), despite their difference in disease latency (one month for AML IEO20 and two months for AML 9). However, due to their respective initial H2B-GFP fluorescence distributions (Figure 4.7), the decrease in the percentage of cells that retained a detectable H2B-GFP+ signal was faster in AML IEO20 than AML 9 (Figure 4.10, right panel). Finally, GFP- populations, upon removal of doxycycline at the end of the chasing period, re-acquired their initial GFP+ signal both *in vitro* and *in vivo* (data not shown). Taken together, these data demonstrate the tight control of H2B-GFP expression by the Tet-Off trans-activator, while the kinetics of the dilution of the H2B-GFP signal *in vivo* suggest that this assay allows identification and selection of quiescent or slowly cycling cells.

4.2.3 The H2B-GFP label-retaining cells are quiescent

The accurate control of the system by doxycycline was essential for the implementation of the *in vivo* H2B-GFP label-retaining assay, yet its reliability needed to be further evaluated on the basis of the cell-cycle properties of the label-retaining cells (LRCs) obtained (LRC+). To this end, we performed an *in vivo* BrdU incorporation assay in combination with Ki-67 characterisation of the GFPhigh, GFPlow and GFPneg populations in the SPL and BM of mice subjected to a short chasing period (Figure 4.11).

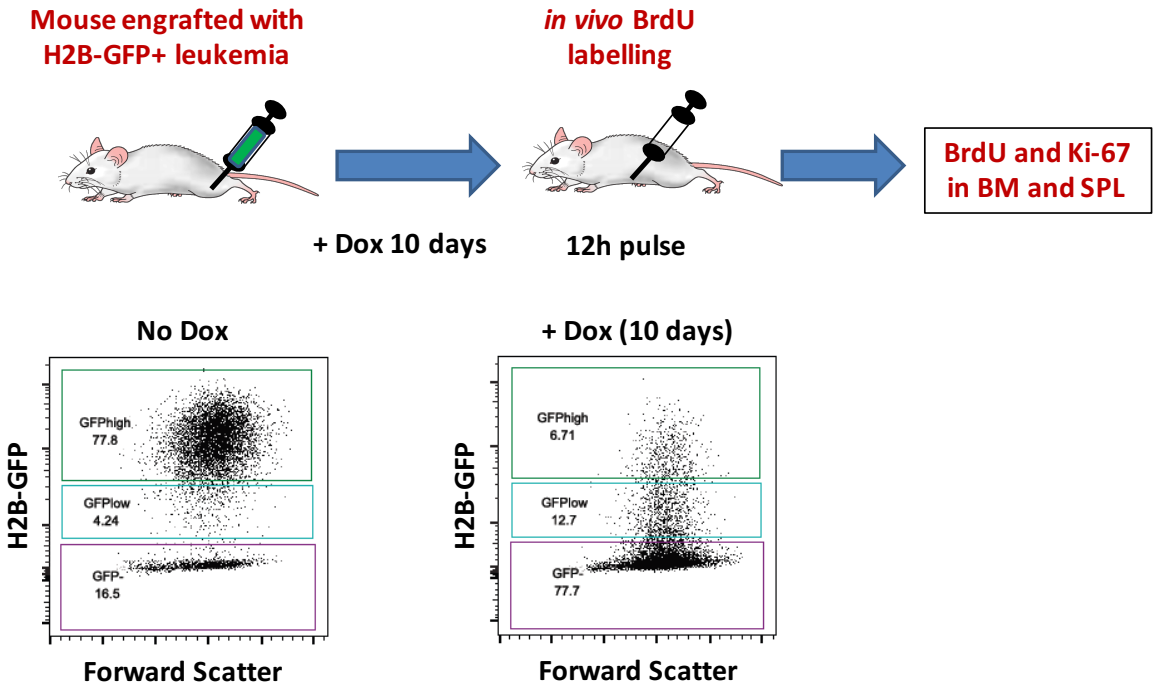

Figure 4.11 Experimental outline and gating strategy for BrdU and Ki-67 analyses.

A 12-hour pulse of BrdU was administered IP to mice engrafted with H2B-GFP+ blasts after a 10-day treatment with doxycycline (+ Dox) for 10 days. BM and SPL cells were then collected for evaluation of BrdU incorporation and Ki-67 status. The flow cytometry dot plots depict representative H2B-GFP expression profiles of AML IEO20 control (No Dox, no doxycycline treatment; left panel) and doxycycline-treated cells (+ Dox; right panel) after a chasing period of 10 days. For both AMLs the gates for the GFPhigh, GFPlow and GFPneg (GFP-) subpopulations were set on the dot plot of the control cells and applied accordingly to all doxycycline-treated samples for the flow cytometry analysis.

In particular, animals already engrafted with H2B-GFP+ blasts and treated with doxycycline for 10 days were given a 12-hour BrdU pulse by two IP injections (one injection every 6 hours). As expected, the percentages of GFPhigh, GFPlow and GFPneg cells varied greatly between the two AML models. Due to the different profiles of H2B-GFP distribution in the starting populations (Figure 4.7), more than 80% of the H2B-GFP+ blasts coming from AML IEO20 lost any detectable GFP signal upon the 10-day chasing period (Figure 4.12). Conversely, in the same timeframe, AML 9 H2B-GFP+ cells largely retained a positive GFP signal (Figure 4.12).

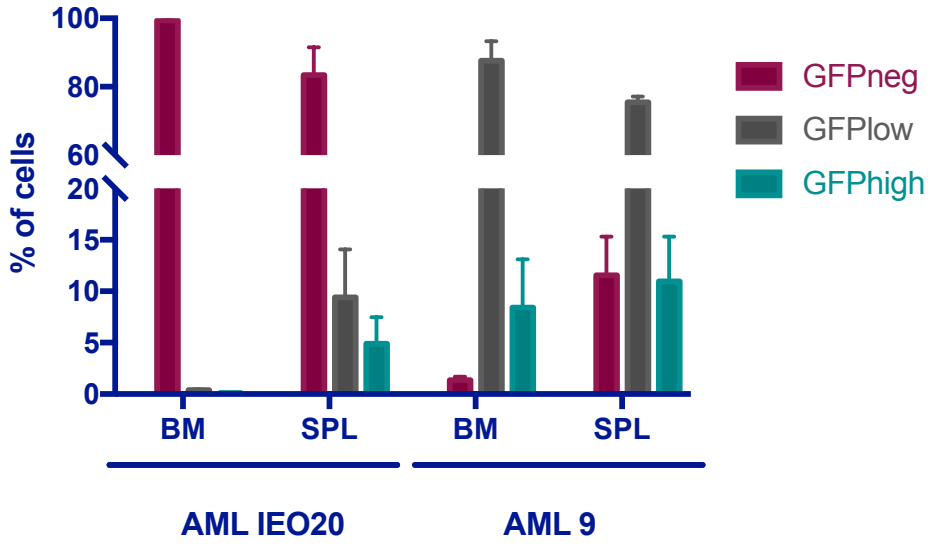

Figure 4.12 Percentages of H2B-GFP subsets in the BM and SPL after 10 days of chasing.

Mice (n=3) transplanted with H2B-GFP+ AML IEO20 or AML 9 cells were treated with doxycycline for 10 days. The percentages of GFPhigh, GFPlow and GFPneg in the BM and SPL are reported, following the gating strategy depicted in Figure 4.11. Mean and standard deviation are shown.

About 12-53% of the cells were found to be actively cycling in the BM and SPL of the two PDX models at 10 days after chasing. The fraction of BrdU incorporating cells was higher in the GFPneg cells, it decreased in the GFPlow compartment and declined further in the GFPhigh subset of doxycycline-treated mice (Figure 4.13). Furthermore, as revealed by

immunostaining, >70% of GFPhigh cells in the BM and >80% in the SPL were negative for Ki-67 at the end of the 10-day doxycycline treatment (Figure 4.14).

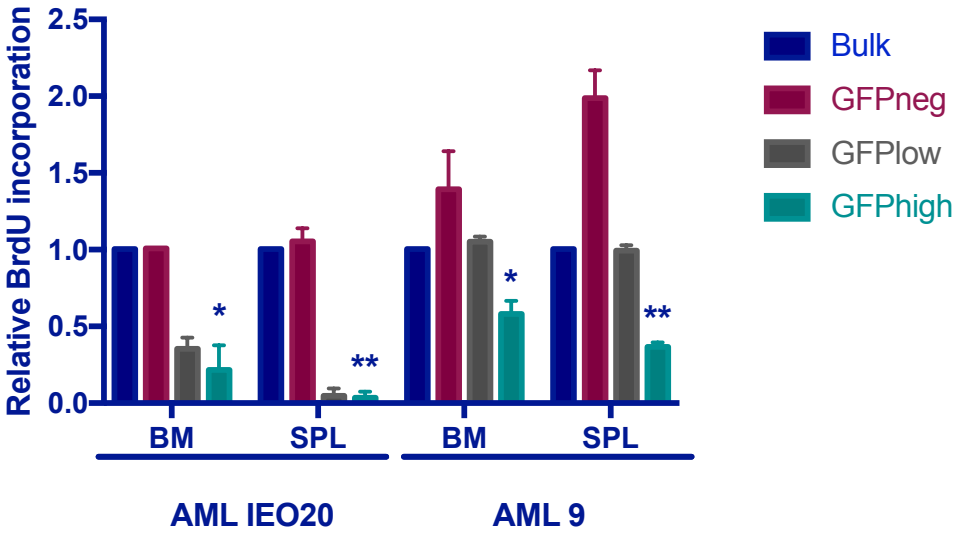

Figure 4.13 BrdU incorporation analysis in the H2B-GFP subsets.

The relative fraction of actively cycling cells (BrdU+) is shown for the bulk (entire leukaemic population) and the GFPhigh, GFPlow and GFPneg subsets (as defined in Figure 4.11) in the BM and SPL of mice (n=3) engrafted with H2B-GFP+ AML IEO20 or AML 9, after a 10-day chasing period with doxycycline. All values were normalised to the corresponding of the bulk population *per* tissue and PDX model. Mean and standard deviation are shown. Student's t-test was performed for GFPhigh *vs.* GFPneg. * p-value < 0.05; ** p-value < 0.01.

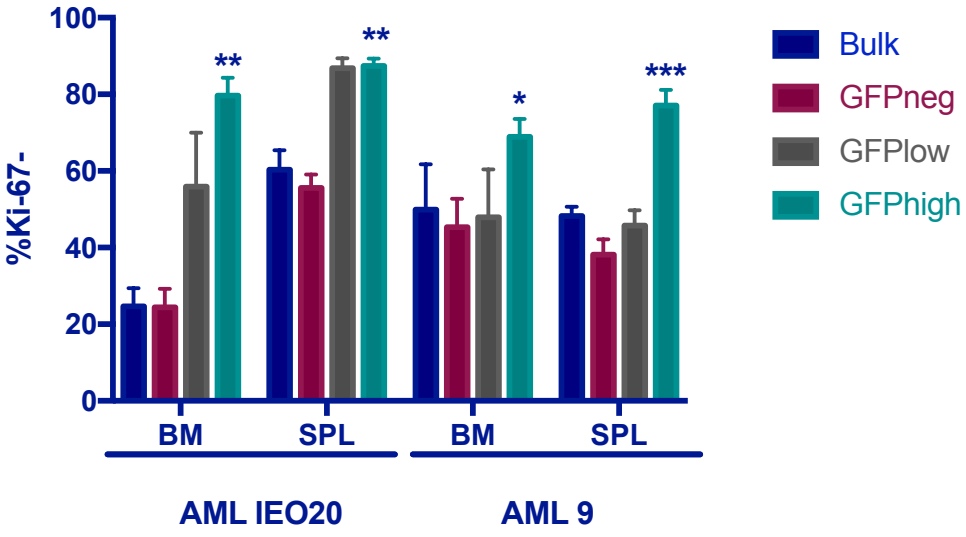

Figure 4.14 Ki-67 analysis in the H2B-GFP subsets.

The percentage of Ki-67- cells is shown for the bulk (entire population) and the GFPhigh, GFPlow and GFPneg subsets (as defined in Figure 4.11) in the BM and SPL of mice (n=3) engrafted with H2B-GFP+ AML IEO20 or AML 9, after a 10-day chasing period with doxycycline. Mean and standard deviation are shown. Student's t-test was performed for GFPhigh *vs.* GFPneg. * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

Thus, both analyses indicated that the loss of the H2B-GFP signal upon doxycycline treatment was linked to an active cell-cycle state. This was exemplified in the case of AML IEO20, where virtually all actively cycling cells were GFPneg (Figure 4.15).

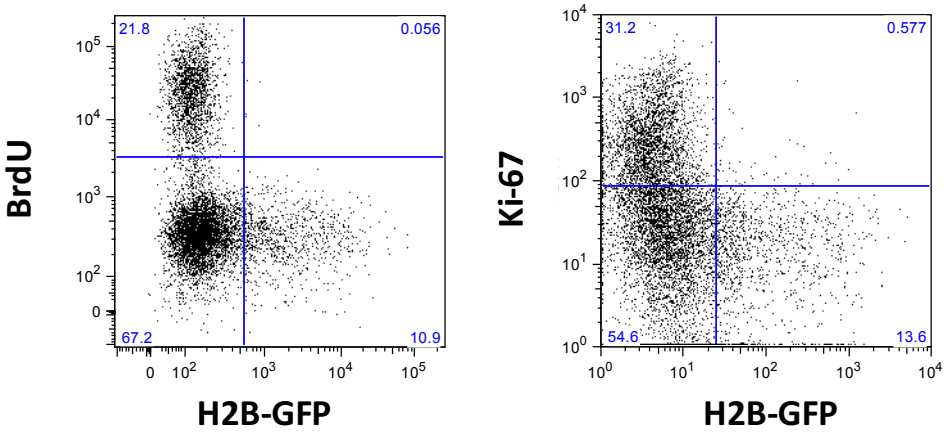

Figure 4.15 Flow cytometry dot plots of BrdU and Ki-67 *vs.* H2B-GFP expression.

Representative flow cytometry dot plots of AML IEO20 SPL blasts, obtained after a 10-day treatment with doxycycline *in vivo*. BrdU incorporation *vs.* H2B-GFP expression on the left panel; Ki-67 *vs.* H2B-GFP expression on the right panel. The percentage of each population is indicated in the corresponding quadrants.

To further characterise the cell-cycle status of the GFPhigh, GFPlow and GFPneg subsets, we combined BrdU incorporation and Ki-67 analyses with total DNA staining by Hoechst 33342 (see section 3.4.2). First, we created gates for the G0/G1, S and G2/M phases, by cell-cycle analyses of DNA content, using the Watson (pragmatic) model on FLowJo. Then, within the G0/G1 gate, the G0 phase was defined with higher stringency by BrdU and Ki-67 double negativity. Consistently with the above reported results, we found a

significant enrichment of cells in the G0 phase of the cell-cycle within the GFPhigh subpopulation (Figure 4.16).

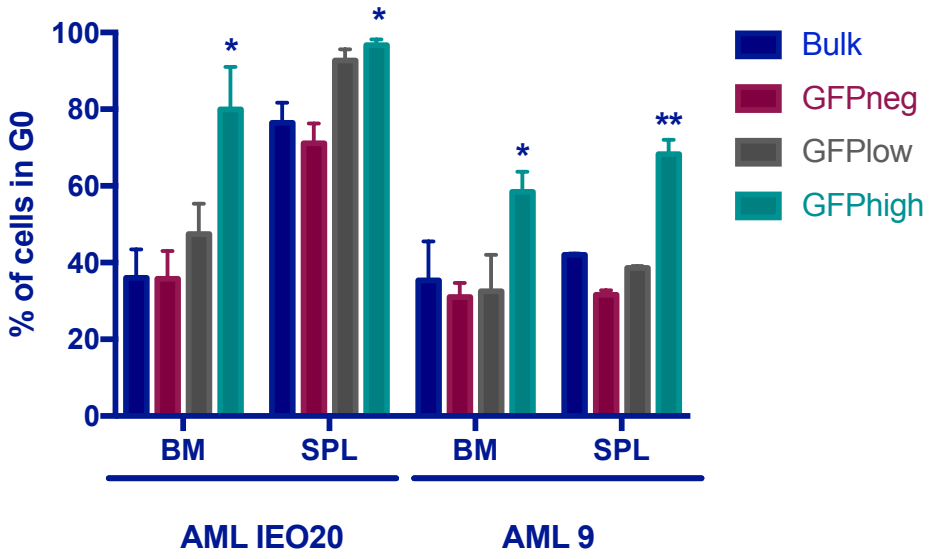

Figure 4.16 Percentage of cells in G0 within the H2B-GFP subsets.

The percentage of G0 (BrdU-Ki-67-) cells is shown for the bulk (entire population) and the GFPhigh, GFPlow and GFPneg subsets (as defined in Figure 4.11) in the BM and SPL of mice (n=3) engrafted with H2B-GFP+ AML IEO20 or AML 9, after a 10-day chasing period with doxycycline. Mean and standard deviation are shown. Student's t-test was performed for GFPhigh *vs.* GFPneg. * p-value < 0.05; ** p-value < 0.01.

In conclusion, these results demonstrate that the cells retaining a high GFP signal are predominantly quiescent, and, more in general, that the H2B-GFP label-retaining assay allows for the identification and isolation of leukaemic subsets based on their proliferation properties.

4.2.4 *In vivo* cell division tracking

We have shown that the H2B-GFP expression is tightly regulated by doxycycline in our two PDX model systems and that the LRC+ (GFPhigh) cells identified after 10 days of chasing are largely quiescent. Thus, we used the H2B-GFP label-retaining assay to study the *in vivo* proliferation kinetics of AML cells. Ideally, a homogeneously H2B-GFP

expressing starting population (before any doxycycline administration) that divides in a synchronous manner would allow the visual identification of each generation of cells as a distinct peak in the H2B-GFP fluorescence histogram. As expected, patterns of H2B-GFP expression were less homogeneous in our two AML models, thus challenging the inference of the number of divisions. To circumvent this limit, we used FlowJo's proliferation platform to analyse the H2B-GFP dilution data obtained after 10 days of chasing with doxycycline.

As shown above, more than 80% of the AML IEO20 cells were already GFPneg by the 10-day time point (Figure 4.12), due to the initial distribution of the H2B-GFP fluorescence (Figure 4.17, left panel). Performing the proliferation analysis with the bulk doxycycline-treated AML IEO20 cells, allowed us to monitor only up to four divisions *in vivo*. All cells that divided more than 4 times appeared to have completely lost any detectable H2B-GFP signal (Figure 4.17, right panel).

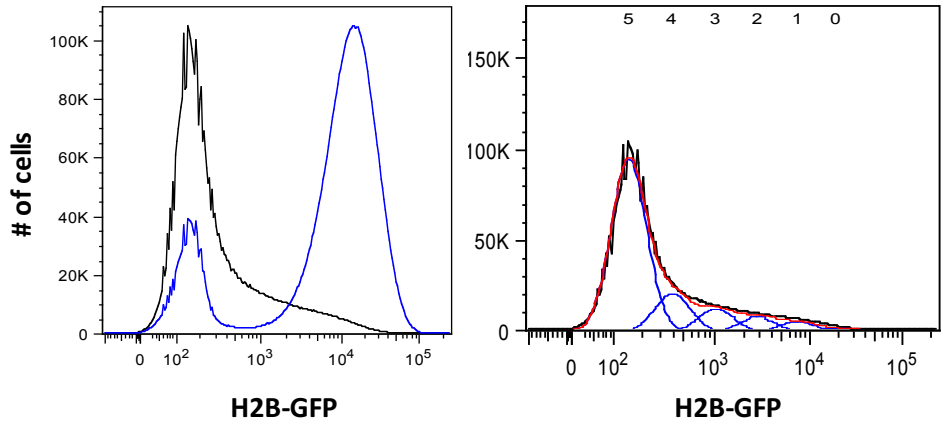

Figure 4.17 *In vivo* division tracking in AML IEO20 based on the H2B-GFP dilution.

A representative H2B-GFP fluorescence histogram is shown (black lines) for AML IEO20 SPL cells, after 10 days of chasing. Left panel: Overlay with the starting population (blue line). Right panel: division modelling using FLowJo's proliferation platform. Each modelled generation is depicted by a blue peak and the composite model by the red line. The number of generations is indicated on the top of the histogram.

The sharp peak of a bright H2B-GFP signal in the starting population of AML 9, instead, enabled us to track up to 8 rounds of cell division (Figure 4.18). Since no information regarding cell division can be any longer obtained from cells when they become GFPneg, we decided to pre-select the H2B-GFP+ population for the modelling of the SPL cells of AML 9, in order to exclude any contaminating mouse stromal cells. In the case of AML IEO20, this issue was not of great importance as we were confident that any such contamination would have had negligible effects on the analysis readout, given the high abundance of GFPneg human cells at the end of the 10-day chasing period.

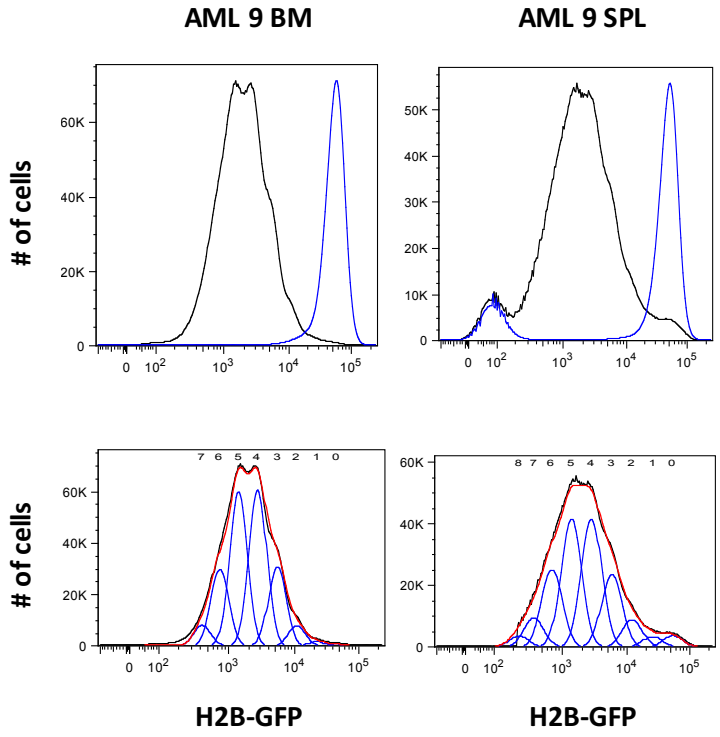

Figure 4.18 *In vivo* division tracking in AML 9 based on the H2B-GFP dilution.

Representative H2B-GFP fluorescence histograms are shown (black lines) for AML 9 BM (left panels) and SPL (right panels) cells, after 10 days of chasing. Upper panels: Overlay with the corresponding starting population (blue line). Lower panels: division modelling using FLowJo's proliferation platform. Each modelled generation is depicted by a blue peak and the composite model by the red line. The number of generations is indicated on the top of the histogram.

The models fitted to the H2B-GFP dilution data of AML 9 allowed us to infer more detailed information on the proliferation kinetics of the labelled cells (Table 4.1). According to the proliferation statistics calculated by FlowJo, all cycling cells underwent an average of 2.87-3.41 cell divisions in the samples analysed for the 10-day chasing period. Importantly, ~7.4% of the BM and ~23% of the SPL original population (at the start of the chasing) remained undivided within the same timeframe.

Table 4.1 Proliferation statistics for AML9.

Peak CV, peak coefficient of variation; Div. Index, division index; Prol. Index, proliferation index; % Divided, percentage of cells in the original population that have divided; mRMS, mean root mean square value (see section 3.4.3).

	# peaks	**Peak ratio**	**Peak CV**	**Div. Index**	**Prol. Index**	**% Divided**	**mRMS**
BM 1	8	0.472	6.12	2.58	2.88	89.5	149
BM 2	8	0.488	5.32	3.18	3.41	93.3	179
BM 3	8	0.49	5.55	3.2	3.38	94.9	296
SPL 1	9	0.471	5.8	2.19	2.87	76.3	790
SPL 2	9	0.468	6.15	2.3	3.08	74.6	382
SPL 3	9	0.464	6.21	2.7	3.35	80.4	1317

In accordance with the cell-cycle analysis outlined in the previous section (see 4.2.3), we observed that implementation of the GFPhigh gate (defined by the fluorescence distribution of control non-doxycycline treated mice, as in Figure 4.11) allows selection of cells that have undergone no cell division, or maximum one, during the chasing period (Figure 4.17 and Figure 4.18).

To summarise, the H2B-GFP dilution profile obtained after a 10-day chasing period allowed us to retrieve information on the *in vivo* proliferation kinetics from both PDX models, albeit with a lower resolution for AML IEO20 due to the initially heterogeneous

expression of the fusion protein prior to doxycycline administration. The lower MFI of the H2B-GFP signal in the starting population (Figure 4.17) of the AML IEO20 accounted for a faster extinction of the GFP signal altogether, although the kinetics of MFI loss were close to identical between the two leukaemias (Figure 4.10).

In conclusion, our data demonstrate that AMLs are composed of populations of cells with highly heterogeneous proliferative potential. In the analysed timeframe (10 days), ~85% of the cells had undergone 3.16 (+/- 0.25) divisions, with ~15% of the initial population remaining GFPhigh, compatible with a proliferative history of zero or maximum one cell division. We consider this latter population of GFPhigh cells as quiescent.

4.3 Biological characterization of isolated quiescent and cycling leukaemic cells

The *in vivo* H2B-GFP label-retaining assay established an experimental setting for the isolation of quiescent and cycling leukaemic populations from PDX models. To test the frequency of LSCs and their tumorigenic properties in both compartments, we transplanted FACS-sorted GFPhigh and GFPneg cells from X1 AMLs under limiting dilution (from 5,000 to 100 cells). Even though doxycycline treatment could be extended to a maximum of 4 weeks in AML IEO20 and 6 weeks in AML 9, based on their respective disease latencies, we opted for a 3-week chasing period to avoid unnecessary animal suffering and to ensure a good retrieval of cells by sorting. In general, at the end of the 3-week chasing, the GFPhigh cells represented <1% of the bulk leukaemia for both PDX models, while the GFPneg accounted for ~80% of the tumour (Figure 4.19). Engraftment of H2B-GFP+ blasts was scored for both GFPhigh and GFPneg populations (Table 4.2), indicating that two functional types of LSCs with diverse cell-cycle properties (quiescent *vs.* proliferating) exist and can be prospectively isolated through the H2B-GFP label-retaining assay.

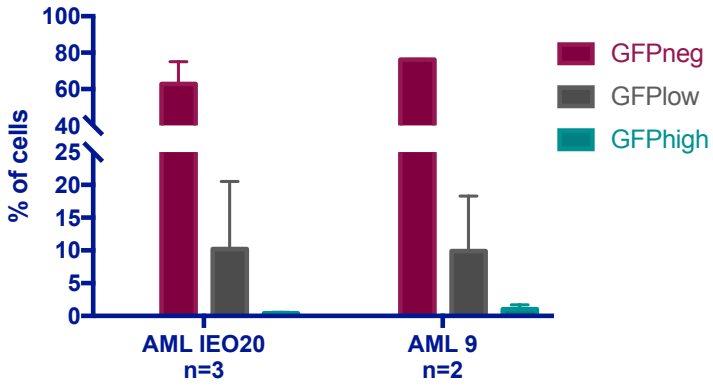

Figure 4.19 Percentages of H2B-GFP subsets in the SPL after 3 weeks of chasing.

Mice transplanted with H2B-GFP expressing AML IEO20 or AML 9 blasts were treated with doxycycline for 3 weeks. The percentages of GFPhigh, GFPlow and GFPneg cells in the SPL are reported, following the gating strategy depicted in Figure 4.11. Mean and standard deviation of independent experiments are shown.

Table 4.2 Limiting dilution transplantation of quiescent (GFPhigh) and cycling (GFPneg) leukaemic cells.

NSG mice were injected with GFPhigh or GFPneg leukaemic blasts at scalar cell doses (from 5,000 – 100 cells) and leukaemia engraftment was checked at regular intervals in the PB of the transplanted animals. All injected mice were monitored for a maximum period of one-year post-transplantation and were euthanized by CO_2 inhalation when blast infiltration reached ~80% in the PB. LSC frequency calculation was performed using ELDA (Hu and Smyth 2009).

	AML IEO20		**AML 9**	
Cell dose	**Quiescent**	**Cycling**	**Quiescent**	**Cycling**
5,000	-	-	2/3	2/3
1,000	3/5	2/6	1/3	0/3
500	0/6	1/8	0/3	0/3
250	0/3	1/6	-	-
100	0/3	-	-	-
LSC frequency	**1:2483 (1:7517-1:820)**	**1:2512 (1:6734-1:937)**	**1:4348 (1:14093-1:1341)**	**1:6953 (1:27035-1:1788)**

The LSC frequency was calculated for all transplanted populations using the Extreme Limiting Dilution Analysis, ELDA (Hu and Smyth 2009), web tool (Table 4.2). Taking into account the representation of each population in the tumour at the end of the 3-week chasing period (Figure 4.19), we estimated that the relative proportion of quiescent *vs.* proliferating LSCs should be roughly close to 1:174 for AML IEO20 and 1:47 for AML 9.

Notably, for both leukaemias, no significant difference was observed in the LSC frequencies of the corresponding GFPhigh (quiescent) and GFPneg (cycling) subsets. However, the quiescent AML 9 LSCs propagated more aggressive secondary leukaemias (X2), with a significantly lower median survival (Figure 4.20). To test whether the quiescent and cycling LSCs (X1) have similar long-term self-renewal potential, we performed a serial transplantation assay. To this end, equal numbers of blasts (bulk population) from secondary leukaemias (X2), that originated from isolated quiescent and

cycling X1 cells, were re-transplanted in parallel in tertiary recipient mice (X3). Contrary to X2 of AML 9, no significant difference was scored in the median survival of the transplanted mice at X3. In addition, all the engrafted animals reached high levels of blast infiltration in a very synchronous manner leading to a sharp downslope of the Kaplan-Meier curve (Figure 4.21). Taken together these data show that quiescent and proliferating LSCs have comparable self-renewal and tumorigenic potential.

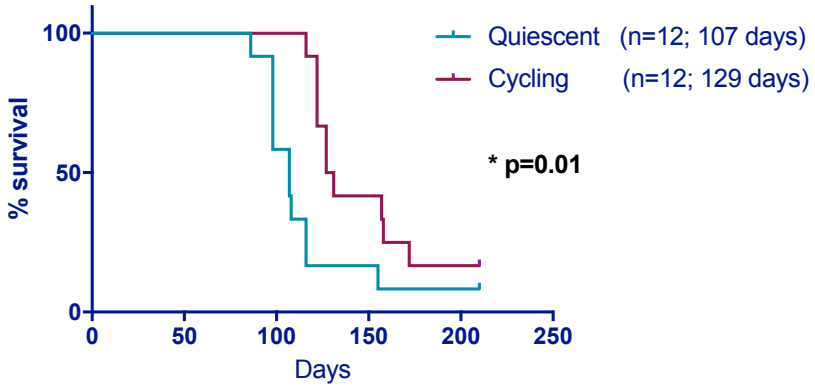

Figure 4.20 Kaplan-Meier curve of mice transplanted with isolated quiescent and cycling cells (X2, AML 9).

Quiescent and cycling leukaemic cells were isolated by FACS as GFPhigh and GFPneg cells, respectively, using the *in vivo* H2B-GFP label-retaining assay and a chasing period of 3 weeks. The sorted populations were transplanted into NSG mice (10,000 cells *per* mouse) and blast engraftment was monitored in the PB of the recipients for a total period of 7 months. Log-rank test was performed to test the differences in survival.

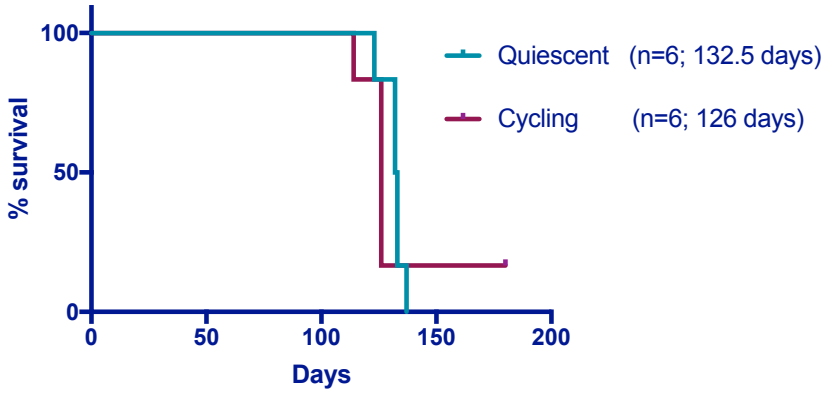

Figure 4.21 Kaplan-Meier curve of tertiary recipient mice transplanted with X2 leukaemias generated from isolated quiescent and cycling cells (AML 9).

Bulk leukaemic cells obtained from the X2 passage of isolated quiescent and cycling cells (X1) were transplanted into tertiary recipients (10,000 cells *per* mouse) and blast engraftment was monitored in the PB of the X3 mice for a total period of 6 months.

In conclusion, as normal HSCs, LSCs exist in two pools, quiescent and proliferating. In contrast to HSCs, the two pools have comparable self-renewal properties and proliferating LSCs are more prevalent.

4.4 Quiescence is a dynamic functional LSC state

4.4.1 Contribution to clonal variability

We next investigated whether the quiescent and proliferating LSCs give rise to the same tumour sub-clones, and whether their progeny has similar self-renewal potential. To this end, we analysed the relative clonal composition of proliferating and quiescent subpopulations, and how this architecture evolves upon serial transplantation. Experimentally, we utilised the $30x10^6$ barcode library in combination with the *in vivo* H2B-GFP label-retaining assay described earlier (Figure 4.22).

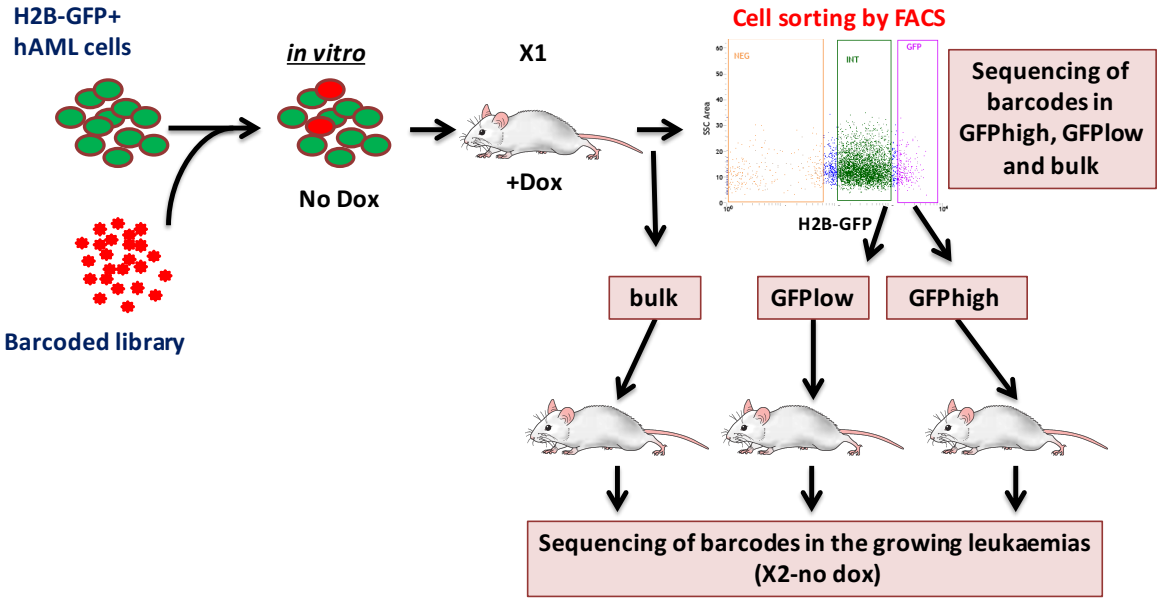

Figure 4.22 Experimental outline of clonal tracking in quiescent and cycling cells.

H2B-GFP+ leukaemic blasts were infected *in vitro*, at a low MOI, with the $30x10^6$ lentiviral barcode library and transplanted into immunocompromised recipient mice (X1). Upon leukaemia manifestation, the mice were treated with doxycycline in order to allow the segregation of quiescent and cycling barcoded leukaemic cells, on the basis of their H2B-GFP fluorescence signal. At the end of a 10-day chasing period, bulk and sorted GFPhigh and GFPlow cells were retrieved for serial transplantation and barcode sequencing (X1). The X2 leukaemias that developed either from the bulk X1 or the isolated X1 quiescent and cycling blasts were also collected for barcode amplification and sequencing.

AML 9 human leukaemic blasts stably transduced with the H2B-GFP vector were, thus, subjected to a second round of infection with the TagRFP+ library of 30x10^6 barcodes (Figure 4.23, left panel) and successfully transplanted into NSG mice. The percentage of TagRFP+GFP+ cells was generally maintained *in vivo* and could be easily monitored in the PB of the engrafted mice (Figure 4.23, right panel).

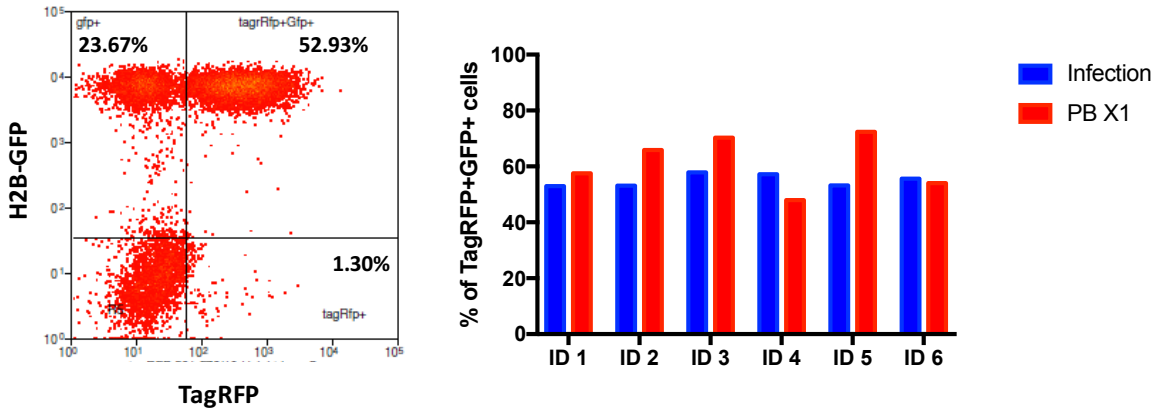

Figure 4.23 *In vivo* propagation of H2B-GFP+ AML 9 cells infected with the pRSI 30x10^6 barcode library.

Left panel: representative flow cytometry dot plot of H2B-GFP+ AML 9 blasts transduced with the 30x10^6 barcode library (at MOI 2) 5 days post infection. Right panel: the percentages of TagRFP+GFP+ AML 9 blasts measured by flow cytometry 5 days post-infection *in vitro* (Infection) and in the PB of the corresponding recipient mice (PB X1) are shown for 6 replicas (ID 1, ID 2, ID 3, ID 4, ID 5 and ID 6).

Animals transplanted with leukaemic cells infected with both vectors (X1) were put on a doxycycline diet for the isolation of a population highly enriched in quiescent LSCs. In this experimental setting, we chose a 10-day chasing period to allow higher cell retrieval from the GFPhigh population, in order to cover, as much as possible, the complexity of the barcode library in the sample. At the end of the chasing, SPL cells of treated mice were sorted based on levels of H2B-GFP expression. One part of the TagRFP+ bulk, quiescent (GFPhigh) and cycling (GFPlow) cells were re-transplanted in equal cell doses into secondary recipients (X2), while the rest was kept for barcode amplification and sequencing (Figure 4.22).

Barcode enumeration in the bulk, cycling and quiescent populations at X1 revealed a very similar clonal composition among the three samples, in terms of numbers of barcodes identified, molecular identity of the barcodes and clone size (percentage of total reads corresponding to each barcode; Figure 4.24). Furthermore, the clonal composition of the secondary leukaemias (X2) was also quite similar especially between the recipients of GFPhigh and GFPlow X1 cells, while there seemed to be a slightly stronger clonal selection in the X2 of the bulk (Figure 4.24).

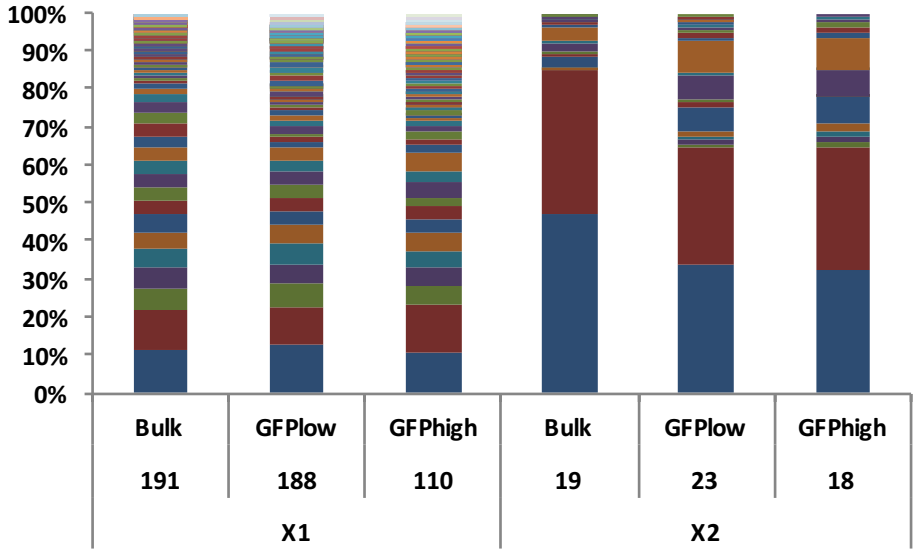

Figure 4.24 Clonal tracking in the bulk (Bulk), cycling (GFPlow) and quiescent (GFPhigh) cells of AML 9.

Distribution of barcode frequencies within each sample. Each barcode is represented by the same colour in all samples. The numbers of barcodes that cumulatively make up for 99% of the reads for each sample are indicated accordingly.

These data indicate that cycling and quiescent LSCs give rise to the same clonal distribution, suggesting that each tumour sub-clone is supported by cycling and quiescent LSCs. Strikingly, the clonal selection process at X2 was not affected by the functional state (cycling or quiescent) of the LSC compartment, suggesting that the two LSC compartments have similar growth potential.

4.4.2 The functional heterogeneity is maintained upon serial passaging of isolated quiescent and cycling LSCs

We next asked if there is a hierarchical link between quiescent and proliferating LSCs, analysing the capacity of each subpopulation to re-form the same pool of quiescent and proliferating cells. To this end, we sorted the GFPhigh and GFPneg leukaemic subsets from AML 9, transplanted in parallel equal numbers of cells from the isolated populations and, upon leukaemia engraftment, treated the secondary recipients (X2) with doxycycline for 3 weeks in order to isolate X2 quiescent and cycling leukaemic cells (Figure 4.25).

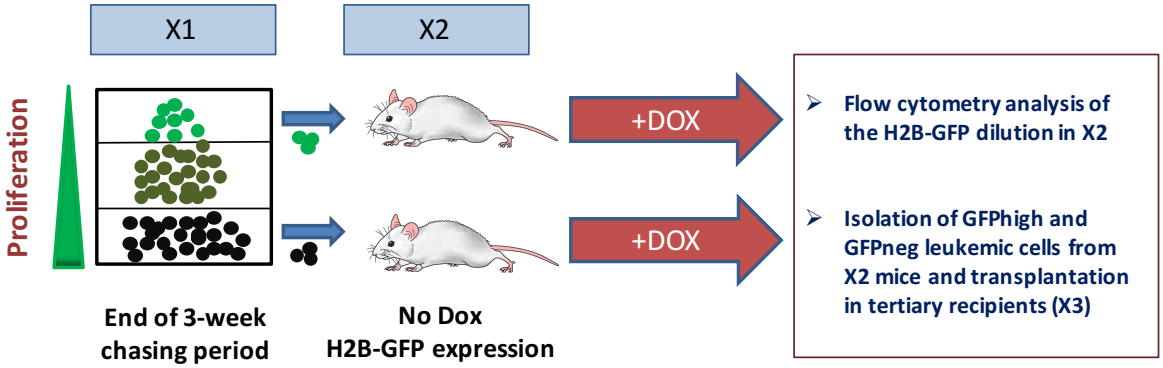

Figure 4.25 Experimental strategy for the characterisation of X3 leukaemias originating from isolated quiescent (GFPhigh) and cycling (GFPneg) cells after two rounds of chasing with doxycycline at X1 and X2.

Quiescent and cycling leukaemic cells were isolated by FACS as GFPhigh and GFPneg cells, respectively, using the *in vivo* H2B-GFP label-retaining assay and a chasing period of 3 weeks. The sorted populations were transplanted into NSG mice that were kept on a normal diet to allow for the re-expression of the H2B-GFP protein. When blast engraftment reached ~30% in the PB of the X2 recipients, we proceeded to another 3-week round of doxycycline administration. At the end of the second chasing period, flow cytometry analysis was performed and secondary (X2) quiescent and cycling cells were isolated by FACS and serially passaged to tertiary recipients (X3).

The profiles of the dilution of the H2B-GFP signal were comparable between the X2 leukaemias originating from X1 quiescent (GFPhigh) and cycling (GFPneg) cells (Figure 4.26), as well as to the profiles obtained for 3 weeks of chasing at X1 (Figure 4.19).

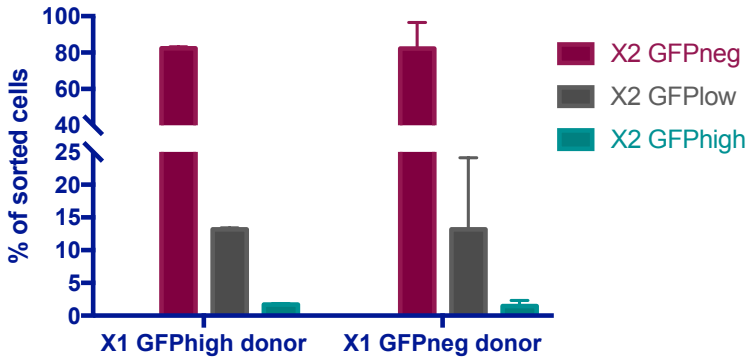

Figure 4.26 Percentages of H2B-GFP subsets in the SPL after 2 rounds of chasing.

Mice transplanted with X1 GFPhigh and GFPneg AML 9 blasts were treated with doxycycline for 3 weeks and GFPhigh, GFPlow and GFPneg cells were sorted from the X2 SPL, following the gating strategy depicted in Figure 4.11. Mean and standard deviation of two independent experiments are shown.

Importantly, as we already reported for the bulk X2 of AML 9 (Figure 4.21), also the isolated X2 quiescent and cycling cells successfully engrafted into tertiary recipients in a quite synchronous manner (Figure 4.27 and Figure 4.28).

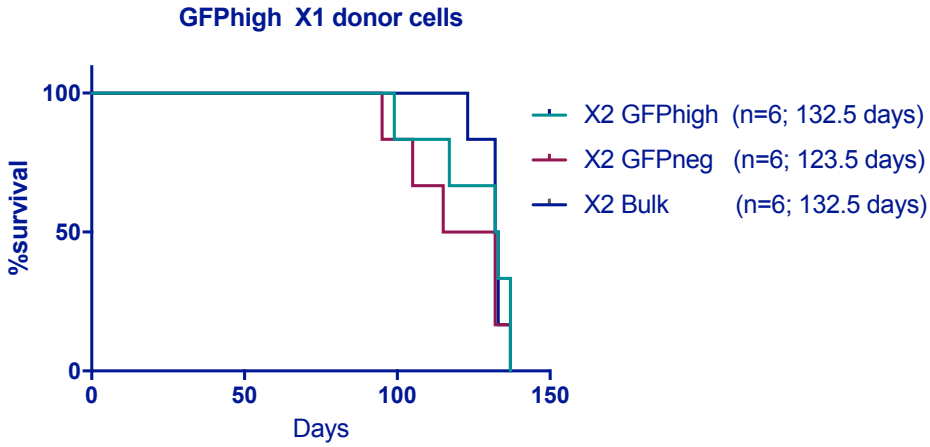

Figure 4.27 Kaplan-Meier curve of tertiary recipient mice transplanted with X2 leukaemic populations generated from isolated X1 quiescent cells (AML 9).

GFPhigh, GFPneg and Bulk leukaemic cells obtained from the X2 passage of isolated quiescent (GFPhigh X1donor) cells were transplanted into tertiary recipients (10,000 cells *per* mouse) and blast engraftment was monitored in the PB of the X3 mice for a total period of 6 months.

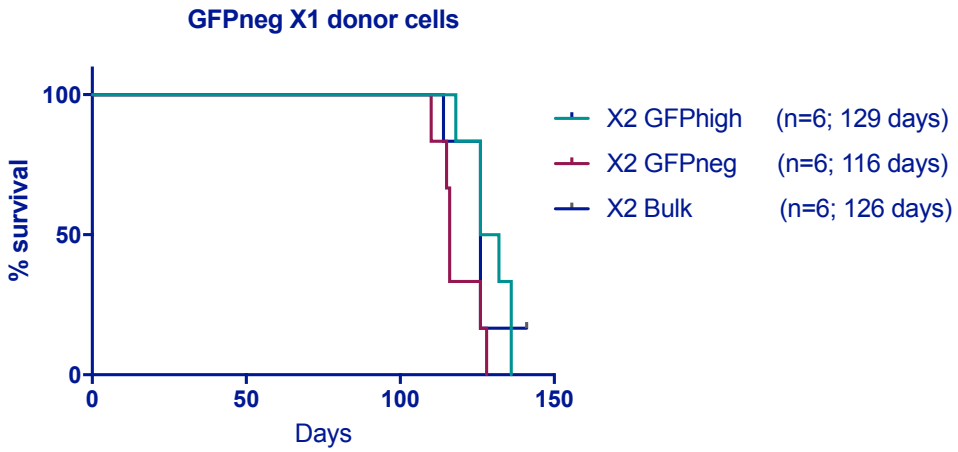

Figure 4.28 Kaplan-Meier curve of tertiary recipient mice transplanted with X2 leukaemic populations generated from isolated X1 cycling cells (AML 9).

GFPhigh, GFPneg and Bulk leukaemic cells obtained from the X2 passage of isolated cycling (GFPneg X1 donor) cells were transplanted into tertiary recipients (10,000 cells *per* mouse) and blast engraftment was monitored in the PB of the X3 mice for a total period of 6 months.

Taken together, these data show that the functional heterogeneity of LSCs is maintained after transplantation of isolated quiescent or cycling LSCs in secondary recipient mice, demonstrating that the two functional LSC states of proliferation and quiescence are not deterministic and that the two populations are maintained in equilibrium in the growing leukaemias. Quiescent LSCs can enter the cell-cycle when challenged by transplantation and cycling LSCs are endowed with the ability to generate a *de novo* pool of tumorigenic quiescent LSCs. Altogether, these observations highlight the importance of efficient targeting of both populations in the clinical setting in order to achieve complete tumour eradication.

4.5 Quiescence fuels tumour evolution

4.5.1 Sub-clone variance upon diverse environmental pressures

To summarise our data so far, we have set up an assay that allows segregation of LSCs, based on their cell-cycle properties, and identified a quiescent tumorigenic compartment which largely shares the same clonal composition with the bulk of the leukaemia. Upon serial passaging, we observed that individually barcoded LSCs exhaust and, strikingly, the pattern of clonal selection is almost identical in different recipient mice. We next examined how this LSC selection is reflected on the genetics of the tumour by performing WES of the bulk AML 9 leukaemias at X1 and X2. Whole exomes of both passages were captured by SureSelect targeted enrichment and sequenced by Illumina HiSeq. The sequences were then analysed with an appropriately developed bioinformatics pipeline (see section 3.6.2). More precisely, MuTect was used to call somatic mutations, single nucleotide variants (SNVs), in each sequenced tumour population against the corresponding normal DNA obtained from T-cells of the original AML patient. Finally, SciClone (Miller, White et al. 2014) was used to cluster the mutations based on variant allele frequency (VAF) similarity and clonal evolution models (p-value ≤ 0.05) were inferred by ClonEvol (Dang, White et al. under review).

In contrast to what we observed by *in vivo* clonal tracking, only a small number of genetic clones were identified by SciClone which was, however, maintained from X1 to X2 (Figure 4.29, left panel). Furthermore, the inferred clonal evolution model (Figure 4.29, right panel) predicted the selective expansion of one of the two major sub-clones from X1 to X2. Bringing together the results obtained through barcoding of individual LSCs and the mutational analysis in the bulk of the tumour, we can postulate that each genomic clone contains many LSCs, which have variable self-renewal potential and equal fitness in the selective environment.

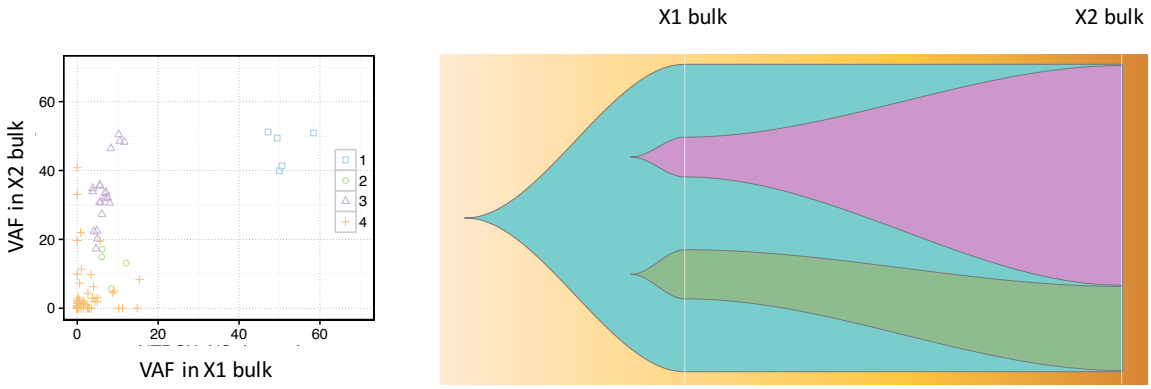

Figure 4.29 Clonal evolution upon serial transplantation (AML 9).

Mutations identified in the bulk of AML 9 X1 and X2 were clustered by SciClone according to VAF similarity into 4 genetic clones (left panel). ClonEvol was used to infer the clonal evolution model (right panel).

We next investigated the mutational landscape of quiescent *vs.* proliferating AML subpopulations in X1 and X2. To this end, we performed WES of FACS-sorted GFPhigh, GFPlow and GFPneg subpopulations from AML 9, as generated by the H2B-GFP label-retaining assay after a 3-week chasing period. As expected by their relative representation in the corresponding bulk tumour (Figure 4.19), the highly proliferating populations (GFPneg) were very similar to the bulk leukaemia of the corresponding passage of AML 9 (Figure 4.30, right panels). On the contrary, the GFPhigh populations were clearly distinct on the genetic level (Figure 4.30, left panels), while the GFPlow represented an intermediate situation (Figure 4.30, middle panels). More precisely, low frequency mutations (VAF < 0.2) identified in the bulk of both passages were found at higher frequencies in the GFPhigh and GFPlow compartments (Figure 4.30), indicating the selective accumulation of unique mutations in the quiescent or slowly cycling leukaemic populations of X1 and X2. On the other hand, a fraction of the high frequency mutations (VAF > 0.2) in the bulk X2 were underrepresented in the corresponding GFPlow and GFPhigh (Figure 4.30, lower panels), suggesting that different genetic clones may have variable proportions of quiescent and proliferating subpopulations.

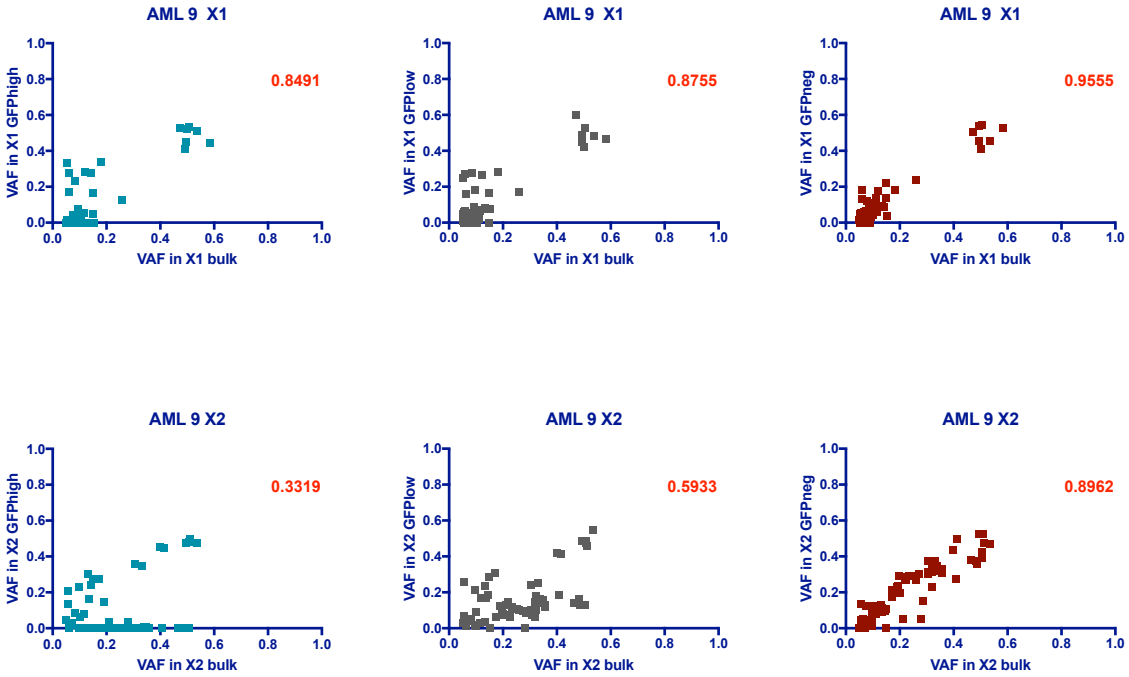

Figure 4.30 Correlations of the variant allele frequency (VAF) for the mutations identified in the bulk leukaemic populations of the X1 and X2 passages of AML 9.

Each point of each graph corresponds to a single SNV with VAF ≥ 0.05 in the bulk leukaemic population of X1 (upper panels) or X2 (lower panels). The VAF in the bulk was always plotted on the X-axis and the VAF in the corresponding GFPhigh (left panels), GFPlow (middle panels) or GFPneg (right panels) subpopulation was plotted on the Y-axis. The values of the Pearson correlation coefficient, r, are shown in red.

We then investigated the effect of the selective pressure of 5-FU on the genetic evolution of the tumour. For this purpose, we administered IP a single dose of 5-FU (150 μg *per* g of body weight) to NSG mice soon after transplantation at X1 and, upon leukaemia presentation, we treated the "relapsed" (+5-FU X1) mice with doxycycline for 3 weeks (Figure 4.31). Notably, heterogeneously cycling leukaemias were generated after the 5-FU treatment, which allowed the isolation of quiescent (GFPhigh) and cycling (GFPlow and GFPneg) cells. As previously demonstrated by the *in vivo* clonal tracking approach (see section 4.1.2), the impact of 5-FU on the clonal evolution of the bulk was also revealed by the SciClone analysis as one of the major tumour sub-clones was counter-selected by the changing environment (Figure 4.32).

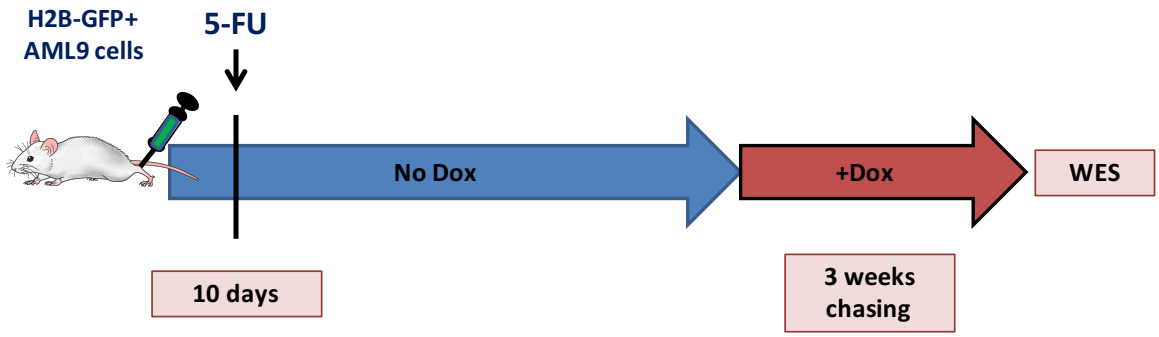

Figure 4.31 Experimental outline for the generation of a "relapsed" leukaemia in the H2B-GFP+ AML 9 PDX model by 5-FU administration at X1.

We transplanted 500,000 H2B-GFP expressing AML 9 blasts into NSG recipient mice and, 10 days later, administered IP a single dose of 5-FU (150 µg *per* g of mouse body weight). Upon leukaemia engraftment (~30% hCD45+ cells in the PB), the mice were put on doxycycline diet for a chasing period of 3 weeks to allow the isolation of +5-FU quiescent and cycling population for WES and subsequent mutational analyses.

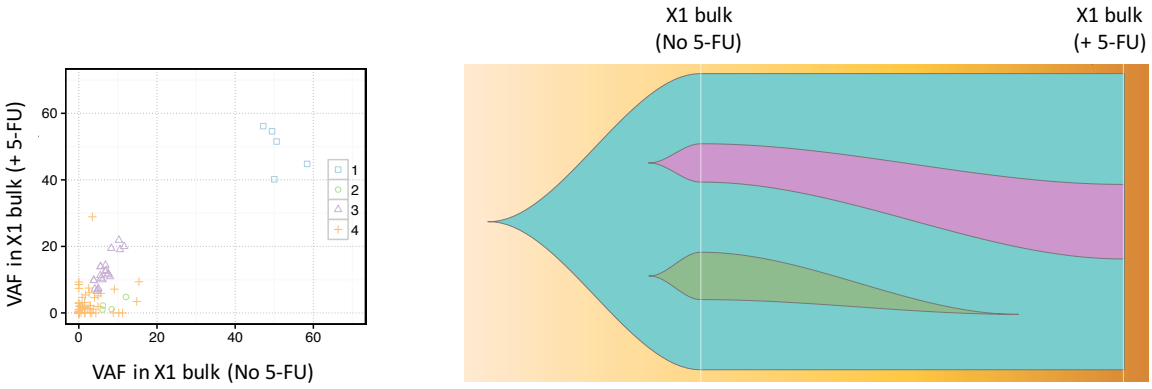

Figure 4.32 Clonal evolution upon 5-FU treatment (AML 9).

Mutations identified in the bulk of AML 9 X1 in the absence (No 5-FU) or presence (+5-FU) of 5-FU were clustered by SciClone according to VAF similarity into 4 genetic clones (left panel). ClonEvol was used to infer the clonal evolution model (right panel).

Furthermore, contrary to our observations in the absence of 5-FU at X1 and X2 (Figure 4.30), the correlation between the GFPhigh or the GFPlow and the bulk was very high and similar to the one observed for the GFPneg (GFPhigh *vs.* bulk, r = 0.9358; GFPlow *vs.* bulk, r = 0.9365; GFPneg *vs.* bulk, r = 0.9465) (Figure 4.33). The high genetic homogeneity between all samples suggests synchronous re-growth of a specific

subpopulation after 5-FU treatment, which is consistent with the cell-cycle specific cytotoxicity of 5-FU.

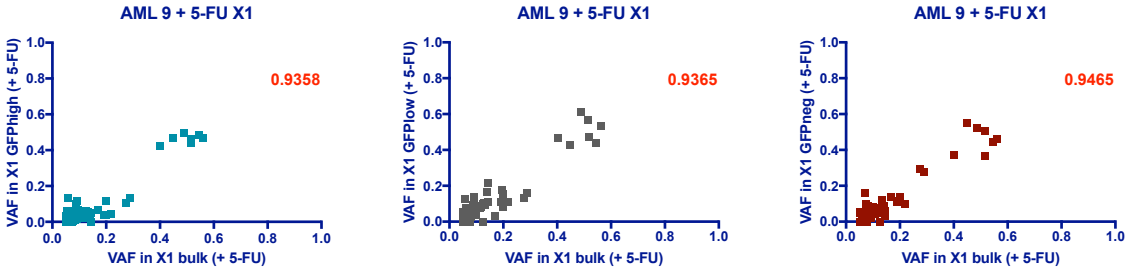

Figure 4.33 Correlations of the variant allele frequency (VAF) for the mutations identified in the bulk leukaemic population of the X1 of AML 9 after 5-FU administration (AML 9 + 5-FU X1).

Each point of each graph corresponds to a single SNV with VAF ≥ 0.05 in the bulk leukaemic population of AML 9 + 5-FU X1. The VAF in the bulk was always plotted on the X-axis and the VAF in the corresponding GFPhigh (left panel), GFPlow (middle panel) or GFPneg (right panel) subpopulation was plotted on the Y-axis. The values of the Pearson correlation coefficient, r, are shown in red.

Next, to examine which functional LSC subpopulation is responsible for the genetic selection by 5-FU, we transplanted isolated +5-FU X1 quiescent (Q, GFPhigh) and cycling (C, GFPneg) cells into secondary recipient mice to obtain +5-FU Q-X2 and C-X2 leukaemias, which were subjected to WES (Figure 4.34).

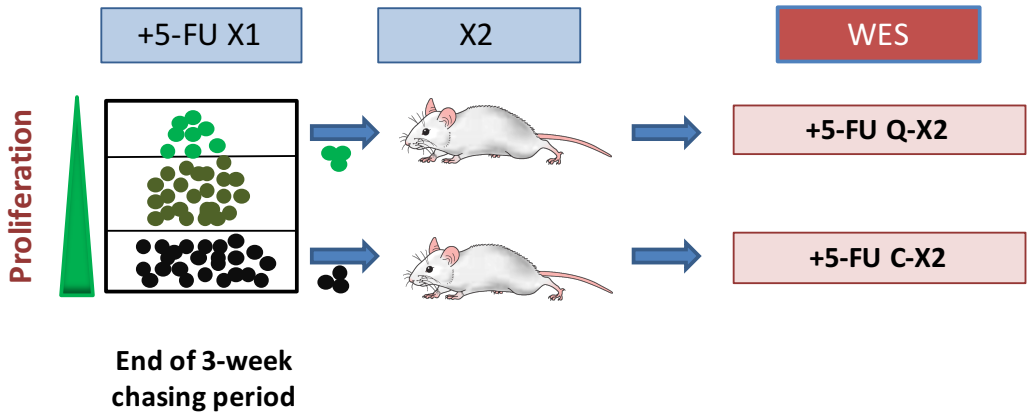

Figure 4.34 Experimental outline for the serial transplantation of isolated quiescent and cycling leukaemic populations after 5-FU treatment to obtain X2 leukaemias for WES (AML 9).

Quiescent (Q) and cycling (C) blasts were isolated from +5-FU X1 leukaemias by FACS as GFPhigh and GFPneg cells, respectively, after a chasing period of 3 weeks. The sorted populations were transplanted into NSG mice to obtain secondary (+5-FU Q-X2 and C-X2) leukaemias for WES and subsequent mutational analyses.

The clonal analysis performed with SciClone on the +5-FU Q-X2 and +5-FU C-X2 leukaemias revealed distinct patterns of tumour evolution. A minor sub-clone was selectively expanded upon transplantation of the quiescent LSCs obtained after 5-FU treatment, whereas transplantation of their proliferating counterpart led to the maintenance of the previously adapted clonal composition to the + 5-FU environment (Figure 4.35, right panels). In detail, the mutational analysis in +5-FU Q-X2 leukaemia indicated the selection of specific SNVs, which were present at low frequency (VAF < 0.1) in the X1 donor cells (Figure 4.35, left pannels). Notably, among the selected mutations we found SNVs in the exonic regions of *PIK3CB* (Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Beta), *MTMR11* (Myotubularin Related Protein 11) and *OGT* (O-Linked N-Acetylglucosamine (GlcNAc) Transferase).

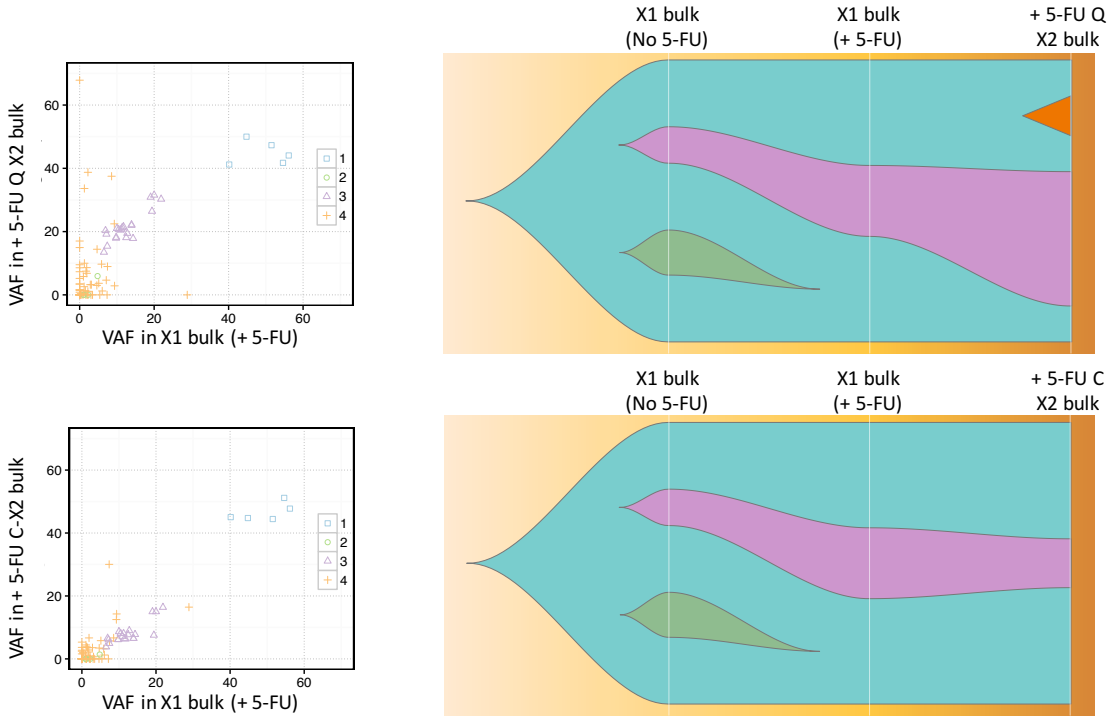

Figure 4.35 Clonal evolution upon transplantation of isolated quiescent (GFPhigh) and proliferating (GFPneg) leukaemic populations after 5-FU treatment (AML 9).

Mutations identified in the confronted samples were clustered by SciClone according to VAF similarity into 4 genetic clones (left panels). ClonEvol was used to infer the clonal evolution model (right panels).

Bringing together our data from the two clonal analyses performed (*i.e.* barcoding and SciClone), we can conclude that each genomic clone contains both quiescent and proliferating LSCs. However, the genomic pattern of quiescent cells is unique under steady state conditions and directs diverse clonal selection upon environmental pressure, suggesting that quiescent LSCs are heterogeneous with respect to their mutational landscape and cell-cycle re-entry probability.

4.5.2 Intra-tumour genetic heterogeneity: the hidden genome of the quiescent leukaemic subpopulation

To better understand the diverse clonal selection patterns driven by the expansion of quiescent and proliferating LSCs (Figure 4.35), we decided to study in greater depth the intra-tumour genetic heterogeneity under steady state conditions. For this purpose, we considered all mutations called by MuTect in the isolated quiescent (GFPhigh) and cycling (GFPlow and/ or GFPneg) leukaemic populations, including very low frequency mutations (VAF < 0.05) and compared them with the ones found in the corresponding bulk of both AML 9 and AML IEO20 at X1.

As already shown in Figure 4.7 (left panel), H2B-GFP expression in AML IEO20 is heterogeneous and, upon transplantation in recipient mice, we consistently detected, prior to any doxycycline administration, a minor population (~10% of the tumour) of non-labelled H2B-GFP- leukaemic cells (Figure 4.8). Therefore, we first checked whether this stable, initial H2B-GFP- population was genetically distinct from the ~90% H2B-GFP+ labelled population of the bulk leukaemia for this PDX. To this end, we isolated by FACS these two populations (H2B-GFP+ and H2B-GFP-) from the SPL cells of a control mouse engrafted with AML IEO20, continuously kept on a normal (doxycycline-free) diet, and

compared their genomes by WES. Indeed, as shown in Figure 4.36, the H2B-GFP- cells contained many high frequency mutations (VAF > 0.2) that were not found in the labelled H2B-GFP+ population. We reasoned that these SNVs would result in "contaminating" mutations in the GFPneg cycling population of AML IEO20 isolated after the 3-week chasing with doxycycline. By introducing false positives, these mutations would inevitably interfere with the interpretation of our results and we, therefore, decided to exclude this cycling population from the mutational analysis of AML IEO20, concentrating only on the GFPhigh and GFPlow subsets.

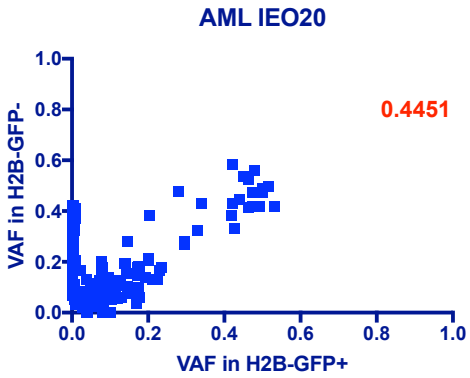

Figure 4.36 Correlation of the variant allele frequency (VAF) for the mutations identified in the H2B-GFP- and H2B-GFP+ subsets of the non-treated AML-IEO20 (prior to doxycycline administration).

Each point of the graph corresponds to a single SNV that was called by MuTect in at least one of the two samples sequenced (H2B-GFP- and H2B-GFP+ subsets prior to chasing). For each SNV, the VAF in the H2B-GFP+ leukaemic population was plotted on the X-axis and the VAF in the H2B-GFP- on the Y-axis. The value of the Pearson correlation coefficient, r, is shown in red.

As expected by their relative representation in the corresponding bulk tumour (Figure 4.19), the cycling populations (GFPlow for AML IEO20; GFPlow and GFPneg for AML 9) were very similar to their respective bulk leukaemias both in terms of number (Figure 4.37) and identity of the SNVs called (Figure 4.38). On the contrary, the GFPhigh was found to comprise a high number of very low frequency mutations that were not present in the bulk or the cycling subsets (GFPneg and/or GFPlow) of the tumour (Figure 4.37). In

more detail, mutations present at high frequency (VAF > 0.2) were generally shared among bulk and quiescent populations, while the vast majority of the SNVs identified specifically in the GFPhigh compartment were characterised by a VAF < 0.1 (Figure 4.37 and Figure 4.38).

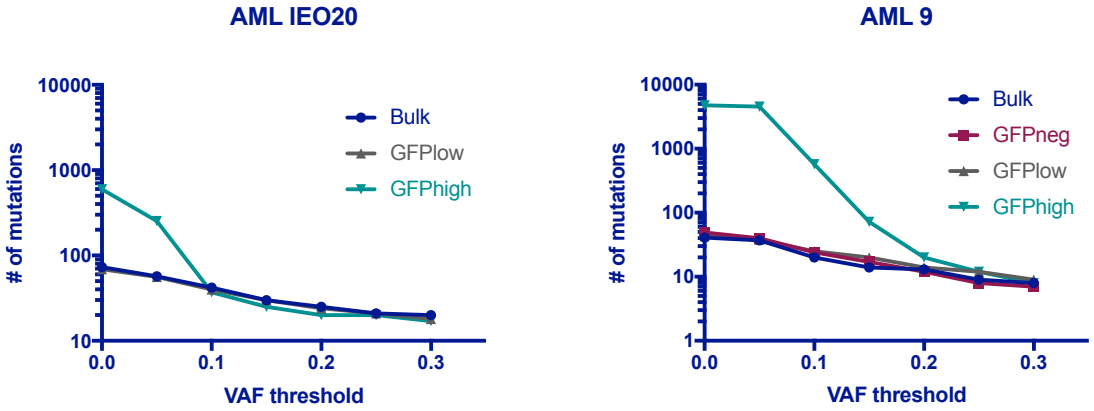

Figure 4.37 Quiescent cells carry a large number of low frequency mutations.

The graphs report the numbers of SNVs identified in the bulk (Bulk), cycling (GFPneg or GFPlow) and quiescent (GFPhigh) subpopulations in AML IEO20 (left panel) and AML 9 (right panel), using different VAF thresholds.

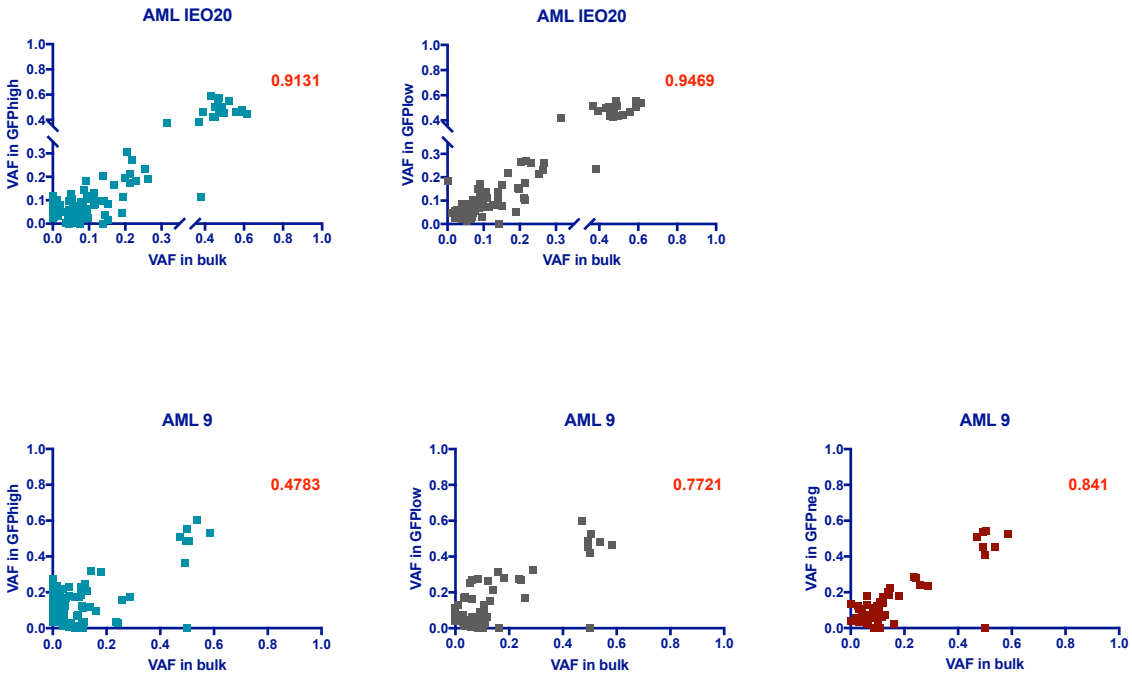

Figure 4.38 Correlations of the variant allele frequency (VAF) for the mutations identified in the GFP subsets and the bulk of each AML.

Each point of each graph corresponds to a single SNV that was called by MuTect in at least one of the two confronted sequences at a time (GFPhigh, GFPlow or GFPneg *vs.* bulk). The VAF in the bulk leukaemic population was always plotted on the X-axis and the VAF in the corresponding GFPhigh (left panels), GFPlow (middle panels) or GFPneg (right panel) subpopulation was plotted on the Y-axis. The values of the Pearson correlation coefficient, r, are shown in red. The plots for AML IEO20 are presented in the two upper panels and the plots for AML 9 in the three lower panels.

Due to the relatively low percentage (~1%) of the GFPhigh cells at the end of the chasing period (Figure 4.19), we often obtained as few as 300,000 cells by FACS. We wondered whether the observed enrichment of the GFPhigh population in very low frequency mutations could be linked to the disproportional number of cells retrieved after sorting compared to the bulk and cycling populations, which were always in abundance (usually in the range of millions of cells). To test this possibility, we randomly sampled aliquots of 300,000 and $3x10^6$ cells from the bulk of AML 9, proceeded to DNA isolation and library preparation for each sample independently, following the exact same protocol employed and already described for the GFP subsets, and compared their whole exome sequences. Notably, both the low cell input (300K) and the high cell input (3M) samples carried a low number of mutations (51 and 53, respectively), which was comparable to the one detected in the bulk and cycling populations of AML 9 (Figure 4.37, right panel). Moreover, a good correlation (r=0.8715) was found between the VAFs scored in the two sequences for all the SNVs identified in at least one of them (Figure 4.39). Therefore, we concluded that the number of cells from which we isolated the DNA for genomic analysis was not a variable that impacted on our experimental procedures nor affected our mutational analysis, indicating, importantly, that our reported observations for the GFPhigh populations were not biased by this parameter.

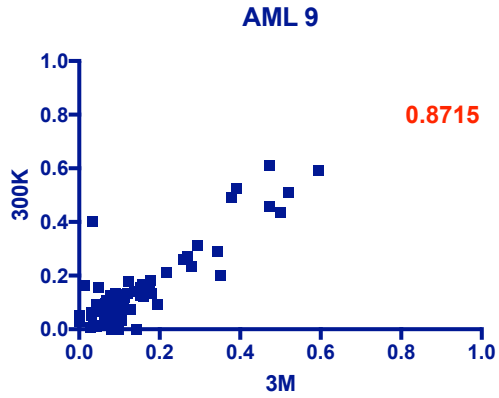

Figure 4.39 The starting number of cells used for DNA extraction does not impact on the number and VAF of the mutations scored by WES.

Each point of the graph corresponds to a single SNV that was called by MuTect in at least one of the two samples sequenced (300K cell input and 3M cell input, both sampled randomly from the bulk of AML 9). The VAF in the high-cell (3M) input sample was plotted on the X-axis and the VAF in the low-cell (300K) input on the Y-axis. The value of the Pearson correlation coefficient, r, is shown in red.

Our data show that the quiescent GFPhigh cells are characterised by the presence of a very high number of low frequency mutations that are unique among the leukaemic populations analysed. This observation strongly indicates that the quiescent leukaemic compartment could play a pivotal role in the progression of the disease as it appears to carry a hidden genome that could not be previously revealed by bulk sequencing approaches.

However, no clonal segregation was scored in the 99% of the tumour by our barcoding approach (Figure 4.24). This led us to hypothesize that the excess of mutations in the GFPhigh subset is a property of the quiescent state of the cells. To further investigate on this phenomenon, we repeated the same analysis in the presence of 5-FU. As previously described, we challenged the quiescent LSCs to enter the cell-cycle by a 5-FU insult soon after transplantation at X1 and, upon leukaemia presentation, we treated the "relapsed" (+5-FU X1) mice with doxycycline (Figure 4.31).

In contrast to steady state conditions, all GFP subsets presented an equal and low number of mutations overall (42 for +5-FU X1 GFPhigh and 45 for the GFPlow and GFPneg).

Notably, the GFPhigh subset was depleted of the mutational overload observed in X1 GFPhigh cells in the absence of 5-FU (Figure 4.40 and Figure 4.41), suggesting that the rare quiescent cells with excess mutations are counter-selected during clonal expansion.

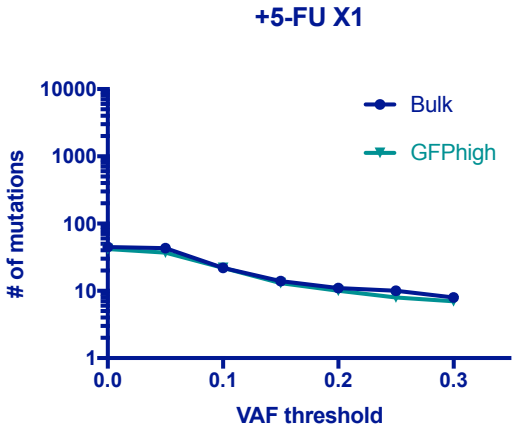

Figure 4.40 Quiescent cells isolated after 5-FU treatment carry equal numbers of mutations with the corresponding bulk (+5-FU X1 AML 9).

The graph reports the numbers of SNVs identified in the bulk (Bulk) and the quiescent (GFPhigh) subpopulation in the +5-FU X1 of AML 9, using different VAF thresholds.

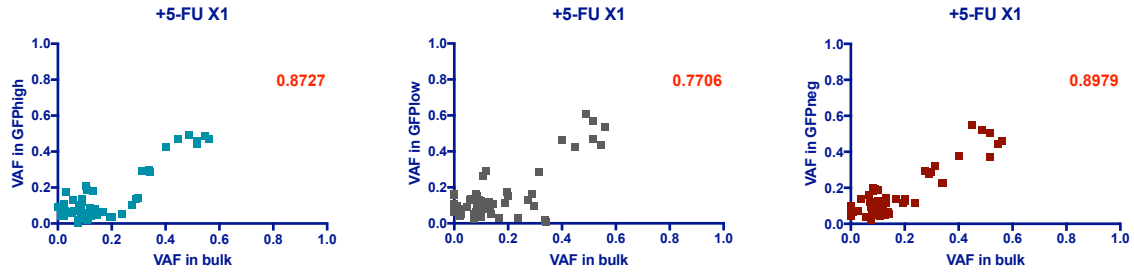

Figure 4.41 Correlations of the variant allele frequency (VAF) for the mutations identified in the GFP subsets and the bulk of the leukaemia generated after 5-FU administration (+5-FU X1).

Each point of each graph corresponds to a single SNV that was called by MuTect in at least one of the two confronted sequences at a time (GFPhigh, GFPlow or GFPneg *vs.* bulk). The VAF in the bulk leukaemic population was always plotted on the X-axis and the VAF in the corresponding GFPhigh (left panel), GFPlow (middle panel) or GFPneg (right panel) subpopulation was plotted on the Y-axis. The values of the Pearson correlation coefficient, r, are shown in red.

In summary, our data suggest that each genomic clone in a growing leukaemia has its own reservoir of quiescent LSCs, with similar fitness within the niche microenvironment. Once these quiescent LSCs enter the cell-cycle, they are destined to exhaust and, simultaneously, to replenish the quiescent pool. Importantly, during quiescence, LSCs accumulate new mutations, forming a reservoir of genomes that can be selected by diverse environmental pressures (*e.g.* 5-FU in our experimental setting). Upon mobilization of quiescent LSC populations the excess of mutations was indeed lost, suggesting that highly-mutated LSCs are counter-selected.

5 Discussion

The interplay between functional and genetic heterogeneity in cancer is constantly gaining attention in the field, as more studies reveal the complexity of the hierarchical CSC organization and clonal evolution patterns in tumours (Kreso and Dick 2014). Here we have reported the existence of two functional classes of LSCs in AML, which maintain a high tumorigenic capacity, but differ in terms of cell-cycle entry rate. Although quiescence has been previously associated with chemoresistance in AML (Ishikawa, Yoshida et al. 2007, Saito, Uchida et al. 2010), our H2B-GFP PDX leukaemia models offer the advantage of monitoring cell proliferation kinetics over time, granting access to a better perception of the dynamic regulation of quiescence *in vivo*. Both LSC types can switch functional classes, meaning quiescent LSCs may enter the cell-cycle and cycling LSCs can revert to a quiescent state in order to maintain an equilibrium between the two populations (Figure 5.1). Therefore, efficient treatment approaches should be designed to target both compartments in parallel in order to achieve complete tumour eradication.

We have further shown that human AMLs are composed of a few genetic clones and each of them is sustained by many individual LSCs, both quiescent and cycling, with variable long-term self-renewal potential. The vast majority of individual LSCs eventually exhaust upon serial passaging *in vivo*, but the genetic heterogeneity of the bulk tumour is maintained under steady state conditions. Importantly, we were able to assess the intrinsic LSC genetic heterogeneity and found that, even though under steady state conditions they largely share a common clonal architecture, a very high number of *de novo* low frequency mutations is concealed in the small pool of quiescent leukaemic cells (~1% of the leukaemia). Upon diverse environmental pressures, relevant phenotypes can be selected, but the excess of mutations is lost (Figure 5.1). The mechanisms by which these mutations

are instilled during quiescence and then selected to fuel tumour adaptation and evolution remains an open issue with important clinical implications.

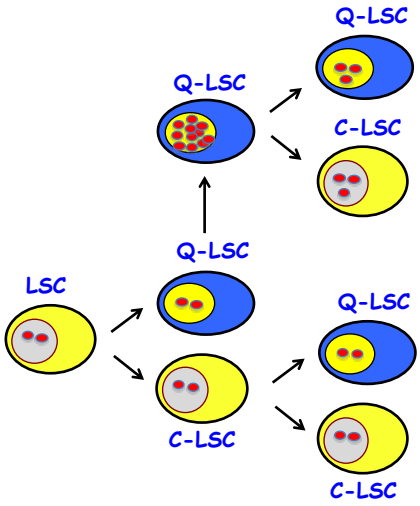

Figure 5.1 Quiescent LSCs fuel tumour heterogeneity.

LSCs include quiescent (Q-LSC) and cycling (C-LSC) cells. Q-LSCs accumulate high numbers of mutations, the majority of which are counter-selected upon cell-cycle re-entry. A few relevant mutations, which may confer a competitive advantage for tumour growth in a given microenvironment, are, however, selected from the functional and genetic reservoir of the Q-LSCs eventually leading to the adaptive evolution of the tumour.

We were able to measure proliferation properties, clonal distribution and mutational load of LSCs in two human AML xenografts, using the *in vivo* H2B-GFP label-retaining assay in combination with a clonal tracking approach and serial xenotransplantation in immunocompromised mice. In more detail, we defined the two leukaemic cell populations on the basis of their label-retention capacities: one rapidly dividing during the course of a 3-week chasing period, resulting in the complete loss of any detectable H2B-GFP signal, and a second that remained essentially undivided and retained a high intensity of H2B-GFP fluorescence. The tumorigenic properties of these two populations were tested by limiting dilution and serial transplantation assays, and by *in vivo* clonal tracking. Both compartments contained LSCs with long-term leukaemogenic potential and exhibited

similar behaviours of clonal selection. Although adult SCs are thought to be predominantly quiescent in normal tissues, accumulating data support a model of co-existing but spatially separated zones of active (proliferating) and resting (quiescent) SCs in the niche microenvironment (Li and Clevers 2010). Distinct functions can be ascribed to active and resting SCs and, most likely, different SC assays (e.g. serial transplantation and lineage tracing) provide biased measures for a specific SC-type at a time (Li and Clevers 2010). These results are in line with previous work of the group on the ErbB2 murine model of breast tumorigeneis, where non-LRC ErbB2-tumour cells (obtained from mammosphere cultures after *ex vivo* labelling with the PKH26 lipophilic dye) were shown to retain tumorigenic properties upon orthotopic transplantation in congenic recipients (Cicalese, Bonizzi et al. 2009). Notably, preliminary data on two H2B-GFP expressing melanoma xenografts further confirm our observations on the leukaemia PDX models, suggesting that the documented CSC-behaviour is not leukaemia-specific and may, instead, have relevant implications for the wider cancer research field (Bossi D., Vlachou T. and Lanfrancone L., unpublished).

To discern more information regarding the internal organisation of the functionally heterogeneous LSC compartment we performed proliferation analyses using FlowJo, which allowed the inference of the number of divisions we can monitor in our models, the percentage of divided and quiescent cells in the original leukaemic population and the division rate (average number of divisions in a given time). As a parallel approach and in order to accommodate the high complexity of an asynchronously dividing cell population monitored *in vivo*, we applied mathematical modelling on the flow cytometry data obtained from the BM at three time points during the chasing period (days 0, 10 and 21) to depict the following scenarios (in collaboration with the University of Bicocca; Nobile M.S., Cazzaniga P. and Besozzi D., unpublished):

- Model 1: one population of cells (dividing cells only);

- Model 2: two populations of cells (non-dividing cells and dividing cells);
- Model 3: three populations of cells (non-dividing cells, slowly dividing cells and fast dividing cells).

Particle Swarm Optimisation was used to estimate the relative contribution of each sub-population in the original pool of leukaemic cells, at the start of the chasing, as well as the average time between two successive divisions in each case. The difference between the expected and the simulated data was determined by both the Kullback-Leibler divergence and the Hellinger distance and the lowest values were provided by Model 3. The best fit corresponded to an average division interval of 53.41 hours for the fast cycling cells and 207.57 hours for the slowly diving cells, assuming an estimated initial ratio of 4 non-dividing : 47 slowly cycling : 49 fast cycling cells at the start of the chasing period. Testing of a fourth model envisioning two dividing populations (a fast and a slow one, the latter encompassing the quiescent cells) is underway.

Similar studies have been performed to decipher cell-cycle and proliferation kinetics in the normal HSC compartment, using H2B-GFP, CFSE or BrdU label-dilution flow cytometry data after long term *in vivo* chasing periods (Wilson, Laurenti et al. 2008, Foudi, Hochedlinger et al. 2009, Takizawa, Regoes et al. 2011, Bernitz, Kim et al. 2016). Quiescent HSCs were consistently found to retain long-term reconstitution ability, while some discrepancies noted between the various proposed models are very likely mirroring differences in the experimental setup (*e.g. in situ* labelling/chasing *vs.* transplantation of pre-labelled cells). Depending on the study, the active HSCs appear to divide once in about every 10-36 days, while the dormant ones are estimated to enter cell-cycle every 100-150 days (Wilson, Laurenti et al. 2008, Takizawa, Regoes et al. 2011). Nevertheless, in all cases, dividing LSCs seem to exhibit faster kinetics in our PDX models (~8.6 days for the slowly dividing and ~2.2 days for the fast dividing cells). It should be noted, however, that we could not reproducibly extend the chasing period beyond the 3-week timeframe due to

animal welfare reasons and therefore we cannot predict the behaviour of the LRC leukaemic population beyond that time point. Assuming a model of two dividing populations, with different cell-cycle entry rate, and using the label-dilution data we have up to the 3-week time point, we may be able to make a more direct comparison with the proposed models for normal HSCs.

We showed by serial transplantation and chasing experiments that quiescent and proliferating LSCs are linked as they can reciprocally produce one another. Consistently, no clonal segregation was observed between the two populations by *in vivo* barcoding. However, we also showed that repeated cell-cycle entry upon serial transplantation of barcoded leukaemias progressively leads to functional exhaustion of the majority of proliferating LSCs, similarly to normal HSCs (Qiu, Papatsenko et al. 2014, Walter, Lier et al. 2015, Bernitz, Kim et al. 2016). Taken together, these data suggest that the long-term survival and propagation of the LSC clone depends on its ability to continuously replenish a pool of quiescent "immortal" LSCs, presumably through asymmetric proliferative divisions (Figure 5.1).

Intriguingly, the quiescent compartment presented a high level of intrinsic genetic diversity compared to the bulk leukaemia. AML genomes are, in general, relatively "simple" and present mutations rates comparable to corresponding normal samples (Welch, Ley et al. 2012, Lawrence, Stojanov et al. 2013, Roberts and Gordenin 2014). Even though the excessive number of quiescence-specific mutations was an unexpected result, still awaiting formal validation, we reckon that the high rate of mutation occurrence might be a property of the functional state of quiescence.

Two main questions remain unanswered: (i) what is the molecular mechanism that leads to the generation of this immense mutational spectrum in a resting leukaemic population and (ii) why do these mutations get depleted upon cell-cycle re-entry.

The notion of mutation acquisition in highly proliferating tumour cells is generally accepted, but the understanding of high mutations rates at quiescence seems to be counter-intuitive at first. However, if one thinks that replication-related mutagenesis coincides with selection while mutations occurring in quiescent cells are not under imminent selective pressure, the higher mutation score in quiescent leukaemic populations or the fact that this excess of mutations is lost when the cells are forced to enter the cell-cycle should not come as a surprise. In the normal BM, it has already been shown that HSCs accumulate DNA damage due to aging-related processes or upon oncogene activation. Reportedly, when HSCs carrying persistent DNA damage are forced to enter the cell-cycle, they face rapid functional exhaustion (Rossi, Seita et al. 2007, Walter, Lier et al. 2015). Therefore, the quiescent state of normal HSCs appears to be prophylactic towards apoptosis and senescence in the presence of unrepaired genetic lesions. On the other hand, studies in the normal haematopoietic system have also shown that quiescent stem cells ensure their long-term survival potential by displaying enhanced abilities to respond to persistent DNA damage (Viale, De Franco et al. 2009, Mohrin, Bourke et al. 2010). Moreover, these cells reportedly employ the error-prone non-homologous end-joining repair (NHEJ)-mediated DNA damage repair mechanism, instead of the high-fidelity homologous recombination, which in turn leads to the accumulation of genetic abnormalities that may persist *in vivo* and potentially contribute to the development of haematopoietic malignancies (Mohrin, Bourke et al. 2010). Similarly, in the context of LSCs, we can envision that persistent DNA damage coupled with increased DNA-repair activity might impart a mutator phenotype that ultimately contributes to disease progression. Our data, indeed, suggest that minor tumour sub-clones within the LSC quiescent compartment can be selected under the environmental pressure of chemotherapy and expand in the relapse tumour.

Studies on the TCGA and other public datasets have revealed that mutations tend to be repaired at higher rates in genes that are actively transcribed and in early replicating regions, creating thus a mutational strand asymmetry in the cancer genome (Roberts and

Gordenin 2014, Haradhvala, Polak et al. 2016). Stretches of single-stranded DNA remain exposed during replication, transcription or double-strand break repair and are vulnerable to genotoxic attacks. In bulk sequencing approaches, a heterogeneous population consisting of asynchronously dividing cells is typically analysed, and, hence, only mutations that are fixed on both strands can be identified. Since the sorted GFPhigh cells have been isolated precisely on the premise of their quiescent state, we can envision the existence of *de novo* "single-strand" mutations, mainly residing on the non-transcribed DNA strand. Single strand lesions were indeed identified in our dataset, predominantly G>T transversions. The high incidence of this type of base substitutions is an indicator of high oxidative stress, which is a known mutagenic mechanism, albeit not typically associated with quiescence (Juntilla, Patil et al. 2010, Lagadinou, Sach et al. 2013).

True G>T "single-strand" mutations cannot be discriminated by artefacts introduced at the step of acoustic DNA fragmentation during library preparation (Costello, Pugh et al. 2013). In bulk sequencing any such lesion scored only on one of the parental DNA strands would be reasonably considered to be an artefact, since all cells are expected to have undergone multiple rounds of division between the moment of the oxidative stress insult in the living cells and the DNA isolation, ensuring the presence of a C>A transversion on the complementary strand. We believe that our approach should allow us to "catch" a snapshot of this type of mutations in their genesis and therefore at single-strand level: if the mutations really appear at quiescence they should be identified as single-strand lesions in our isolated quiescent populations. In fact, a massive rate of single-strand G>T transversions was scored in the GFPhigh samples under steady state conditions, reaching the overwhelming number of 30,000 SNVs for AML 9, which were excluded from all the analyses presented in the Results section. We are now testing an alternative method of enzymatic DNA fragmentation prior to library preparation for target enrichment in parallel in pairs of bulk and quiescent leukaemic samples. This approach should reveal the actual number of single-strand lesions captured in the quiescent state *in vivo* and enable us to

better appreciate the role of these mutations in tumour progression. If the hypothesis of *de novo* single-strand mutations is sustained by additional data, we can postulate that in great part these mutations get repaired upon cell-cycle re-entry.

Regardless of whether the quiescence-specific mutations are found only on one or on both strands of the DNA, the absence of selective pressure, precisely due to the quiescent state of the cells, is most likely the factor that renders these cells more permissive to sustain a high mutation load. However, when they are prompted to enter the cell-cycle, the selective processes of tumour evolution will take action by eliminating cells carrying disadvantageous mutations. The clinically relevant question is whether the few mutations that are not repaired or counter-selected upon cell-cycle entry can promote an adaptive response of the tumour. Our data show, indeed, that, upon selective expansion of quiescent LSCs, minor genetic sub-clones carrying mutations associated with tumour aggressiveness and chemotherapy-resistance can emerge, suggesting a mechanism for the development of refractory relapse tumours.

Appendix

During the course of my PhD studies I closely collaborated with a fellow student on a project aimed at studying the effects of de-regulated Myc on the self renewal properties of normal and cancer mammary stem cells. A manuscript reporting the results of this study has been recently submitted and we are waiting for the editorial decision. The title, list of authors and the abstract are provided in the following pages.

Loss of p53 in breast cancer leads to Myc activation and reprogramming of cancer stem cells

Angela Santoro[1]^, Thalia Vlachou[1]^, Lucilla Luzi[1], Cristina Elisabetta Pasi[1], Linsey Reavie[2], Giorgio Melloni[3], Paola Bonetti[3], Errico D'Elia[1], Lucia Casoli[3], Arianna Sabò[3], Francesco Nicassio[3], Maria Cristina Moroni[1], Gaetano Ivan Dellino[1, 4], Bruno Amati[1, 3] and Pier Giuseppe Pelicci[1, 4]*

[1]Department of Experimental Oncology, European Institute of Oncology, via Adamello 16, 20139 Milan, Italy.

[2]Friedrich Miescher Institute, Maulbeerstr 66, 4058 Basel, Switzerland

[3]Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, via Adamello 16, 20139 Milan, Italy

[4]Dipartimento di Oncologia ed Emato-Oncologia, Universita' degli Studi di Milano, via Santa Sofia 9, 20142 Milan, Italy

^ Equally contributing

*Correspondence to Pier Giuseppe Pelicci: piergiuseppe.pelicci@ieo.eu

Abstract

Loss of p53 function is invariably associated with cancer. Its role in tumor growth was recently linked to its effects on cancer stem cells (CSCs), though underlying molecular mechanisms remain unknown. Here we show that *c-myc* is a transcriptional target of p53 in mammary stem cells (MaSCs), and its activation in breast tumors is a consequence of p53-loss. Constitutive Myc expression leads to increased frequency of MaSC symmetric divisions and mammary progenitors reprogramming. In tumors, Myc activation is necessary and sufficient to maintain the CSCs phenotype and tumor growth. Our data indicate Myc as the effector of p53-loss in mammary CSCs, and suggest that continuous reprogramming of mammary progenitors contributes to the maintenance of the expanding pool of CSCs.

References

Agliano, A., I. Martin-Padura, P. Mancuso, P. Marighetti, C. Rabascio, G. Pruneri, L. D. Shultz and F. Bertolini (2008). "Human acute leukemia cells injected in NOD/LtSz-scid/IL-2Rgamma null mice generate a faster and more efficient disease compared to other NOD/scid-related strains." Int J Cancer **123**(9): 2222-2227.

Al-Hajj, M., M. S. Wicha, A. Benito-Hernandez, S. J. Morrison and M. F. Clarke (2003). "Prospective identification of tumorigenic breast cancer cells." Proc Natl Acad Sci U S A **100**(7): 3983-3988.

Arber, D. A., A. Orazi, R. Hasserjian, J. Thiele, M. J. Borowitz, M. M. Le Beau, C. D. Bloomfield, M. Cazzola and J. W. Vardiman (2016). "The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia." Blood **127**(20): 2391-2405.

Arndt-Jovin, D. J. and T. M. Jovin (1977). "Analysis and sorting of living cells according to deoxyribonucleic acid content." J Histochem Cytochem **25**(7): 585-589.

Bao, S., Q. Wu, R. E. McLendon, Y. Hao, Q. Shi, A. B. Hjelmeland, M. W. Dewhirst, D. D. Bigner and J. N. Rich (2006). "Glioma stem cells promote radioresistance by preferential activation of the DNA damage response." Nature **444**(7120): 756-760.

Barde, I., P. Salmon and D. Trono (2010). "Production and titration of lentiviral vectors." Curr Protoc Neurosci **Chapter 4**: Unit 4 21.

Beck, B. and C. Blanpain (2013). "Unravelling cancer stem cell potential." Nat Rev Cancer **13**(10): 727-738.

Bennett, J. M., D. Catovsky, M. T. Daniel, G. Flandrin, D. A. Galton, H. R. Gralnick and C. Sultan (1976). "Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group." Br J Haematol **33**(4): 451-458.

Bernitz, J. M., H. S. Kim, B. MacArthur, H. Sieburg and K. Moore (2016). "Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions." Cell **167**(5): 1296-1309 e1210.

Birnboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**(6): 1513-1523.

Bodini, M., C. Ronchini, L. Giaco, A. Russo, G. E. Melloni, L. Luzi, D. Sardella, S. Volorio, S. K. Hasan, T. Ottone, S. Lavorgna, F. Lo-Coco, A. Candoni, R. Fanin, E. Toffoletti, I. Iacobucci, G.

Martinelli, A. Cignetti, C. Tarella, L. Bernard, P. G. Pelicci and L. Riva (2015). "The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations." Blood **125**(4): 600-605.

Bonnet, D. and J. E. Dick (1997). "Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell." Nat Med **3**(7): 730-737.

Boyum, A. (1968). "Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g." Scand J Clin Lab Invest Suppl **97**: 77-89.

Busch, K., K. Klapproth, M. Barile, M. Flossdorf, T. Holland-Letz, S. M. Schlenner, M. Reth, T. Hofer and H. R. Rodewald (2015). "Fundamental properties of unperturbed haematopoiesis from stem cells in vivo." Nature **518**(7540): 542-546.

Cancer Genome Atlas Research, N. (2013). "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." N Engl J Med **368**(22): 2059-2074.

Cibulskis, K., M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander and G. Getz (2013). "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples." Nat Biotechnol **31**(3): 213-219.

Cicalese, A., G. Bonizzi, C. E. Pasi, M. Faretta, S. Ronzoni, B. Giulini, C. Brisken, S. Minucci, P. P. Di Fiore and P. G. Pelicci (2009). "The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells." Cell **138**(6): 1083-1095.

Conway, T., J. Wazny, A. Bromage, M. Tymms, D. Sooraj, E. D. Williams and B. Beresford-Smith (2012). "Xenome--a tool for classifying reads from xenograft samples." Bioinformatics **28**(12): i172-178.

Corces-Zimmerman, M. R., W. J. Hong, I. L. Weissman, B. C. Medeiros and R. Majeti (2014). "Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission." Proc Natl Acad Sci U S A **111**(7): 2548-2553.

Costello, M., T. J. Pugh, T. J. Fennell, C. Stewart, L. Lichtenstein, J. C. Meldrim, J. L. Fostel, D. C. Friedrich, D. Perrin, D. Dionne, S. Kim, S. B. Gabriel, E. S. Lander, S. Fisher and G. Getz (2013). "Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation." Nucleic Acids Res **41**(6): e67.

Dang, H. X., B. S. White, S. M. Foltz, C. A. Miller, J. Luo, R. C. Fields and C. A. Maher (under review). ClonEvol: inferring and visualizing clonal evolution in multi-sample cancer sequencing.

De Kouchkovsky, I. and M. Abdul-Hay (2016). "'Acute myeloid leukemia: a comprehensive review and 2016 update'." Blood Cancer J **6**(7): e441.

Ding, L., T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendl, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson and J. F. DiPersio (2012). "Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing." Nature **481**(7382): 506-510.

Dohner, H., E. H. Estey, S. Amadori, F. R. Appelbaum, T. Buchner, A. K. Burnett, H. Dombret, P. Fenaux, D. Grimwade, R. A. Larson, F. Lo-Coco, T. Naoe, D. Niederwieser, G. J. Ossenkoppele, M. A. Sanz, J. Sierra, M. S. Tallman, B. Lowenberg, C. D. Bloomfield and L. European (2010). "Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet." Blood **115**(3): 453-474.

Dohner, H., D. J. Weisdorf and C. D. Bloomfield (2015). "Acute Myeloid Leukemia." N Engl J Med **373**(12): 1136-1152.

DuBridge, R. B., P. Tang, H. C. Hsia, P. M. Leong, J. H. Miller and M. P. Calos (1987). "Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system." Mol Cell Biol **7**(1): 379-387.

Dull, T., R. Zufferey, M. Kelly, R. J. Mandel, M. Nguyen, D. Trono and L. Naldini (1998). "A third-generation lentivirus vector with a conditional packaging system." J Virol **72**(11): 8463-8471.

Eppert, K., K. Takenaka, E. R. Lechman, L. Waldron, B. Nilsson, P. van Galen, K. H. Metzeler, A. Poeppl, V. Ling, J. Beyene, A. J. Canty, J. S. Danska, S. K. Bohlander, C. Buske, M. D. Minden, T. R. Golub, I. Jurisica, B. L. Ebert and J. E. Dick (2011). "Stem cell gene expression programs influence clinical outcome in human leukemia." Nat Med **17**(9): 1086-1093.

Estey, E. and H. Dohner (2006). "Acute myeloid leukaemia." Lancet **368**(9550): 1894-1907.

Falkowska-Hansen, B., J. Kollar, B. M. Gruner, M. Schanz, P. Boukamp, J. Siveke, A. Rethwilm and M. Kirschner (2010). "An inducible Tet-Off-H2B-GFP lentiviral reporter vector for detection and in vivo isolation of label-retaining cells." Exp Cell Res **316**(11): 1885-1895.

Ferlay, J., E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. Coebergh, H. Comber, D. Forman and F. Bray (2013). "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012." Eur J Cancer **49**(6): 1374-1403.

Fialkow, P. J., J. W. Janssen and C. R. Bartram (1991). "Clonal remissions in acute nonlymphocytic leukemia: evidence for a multistep pathogenesis of the malignancy." Blood **77**(7): 1415-1417.

Foudi, A., K. Hochedlinger, D. Van Buren, J. W. Schindler, R. Jaenisch, V. Carey and H. Hock (2009). "Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells." Nat Biotechnol **27**(1): 84-90.

Genovese, G., A. K. Kahler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landen, M. Hoglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Gronberg, C. M. Hultman and S. A. McCarroll (2014). "Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence." N Engl J Med **371**(26): 2477-2487.

Gerdes, J., H. Lemke, H. Baisch, H. H. Wacker, U. Schwab and H. Stein (1984). "Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67." J Immunol **133**(4): 1710-1715.

Gerlinger, M., A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N. Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C. R. Santos, M. Nohadani, A. C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P. A. Futreal and C. Swanton (2012). "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing." N Engl J Med **366**(10): 883-892.

Gilliland, D. G. and J. D. Griffin (2002). "The roles of FLT3 in hematopoiesis and leukemia." Blood **100**(5): 1532-1542.

Gilliland, D. G., C. T. Jordan and C. A. Felix (2004). "The molecular basis of leukemia." Hematology Am Soc Hematol Educ Program: 80-97.

Ginestier, C., M. H. Hur, E. Charafe-Jauffret, F. Monville, J. Dutcher, M. Brown, J. Jacquemier, P. Viens, C. G. Kleer, S. Liu, A. Schott, D. Hayes, D. Birnbaum, M. S. Wicha and G. Dontu (2007). "ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome." Cell Stem Cell **1**(5): 555-567.

Graham, F. L., J. Smiley, W. C. Russell and R. Nairn (1977). "Characteristics of a human cell line transformed by DNA from human adenovirus type 5." J Gen Virol **36**(1): 59-74.

Gratzner, H. G. and R. C. Leif (1981). "An immunofluorescence method for monitoring DNA synthesis by flow cytometry." Cytometry **1**(6): 385-393.

Griffith, M., C. A. Miller, O. L. Griffith, K. Krysiak, Z. L. Skidmore, A. Ramu, J. R. Walker, H. X. Dang, L. Trani, D. E. Larson, R. T. Demeter, M. C. Wendl, J. F. McMichael, R. E. Austin, V. Magrini, S. D. McGrath, A. Ly, S. Kulkarni, M. G. Cordes, C. C. Fronick, R. S. Fulton, C. A. Maher, L. Ding, J. M. Klco, E. R. Mardis, T. J. Ley and R. K. Wilson (2015). "Optimizing cancer genome sequencing and analysis." Cell Syst **1**(3): 210-223.

Grimwade, D., A. Ivey and B. J. Huntly (2016). "Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance." Blood **127**(1): 29-41.

Grimwade, D., H. Walker, F. Oliver, K. Wheatley, C. Harrison, G. Harrison, J. Rees, I. Hann, R. Stevens, A. Burnett and A. Goldstone (1998). "The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties." Blood **92**(7): 2322-2333.

Haradhvala, N. J., P. Polak, P. Stojanov, K. R. Covington, E. Shinbrot, J. M. Hess, E. Rheinbay, J. Kim, Y. E. Maruvka, L. Z. Braunstein, A. Kamburov, P. C. Hanawalt, D. A. Wheeler, A. Koren, M. S. Lawrence and G. Getz (2016). "Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair." Cell **164**(3): 538-549.

Hu, Y. and G. K. Smyth (2009). "ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays." J Immunol Methods **347**(1-2): 70-78.

Hughes, A. E., V. Magrini, R. Demeter, C. A. Miller, R. Fulton, L. L. Fulton, W. C. Eades, K. Elliott, S. Heath, P. Westervelt, L. Ding, D. F. Conrad, B. S. White, J. Shao, D. C. Link, J. F. DiPersio, E. R. Mardis, R. K. Wilson, T. J. Ley, M. J. Walter and T. A. Graubert (2014). "Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing." PLoS Genet **10**(7): e1004462.

Ishikawa, F., S. Yoshida, Y. Saito, A. Hijikata, H. Kitamura, S. Tanaka, R. Nakamura, T. Tanaka, H. Tomiyama, N. Saito, M. Fukata, T. Miyamoto, B. Lyons, K. Ohshima, N. Uchida, S. Taniguchi, O. Ohara, K. Akashi, M. Harada and L. D. Shultz (2007). "Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region." Nat Biotechnol **25**(11): 1315-1321.

Jaiswal, S., P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler and B. L. Ebert (2014). "Age-related clonal hematopoiesis associated with adverse outcomes." N Engl J Med **371**(26): 2488-2498.

Jan, M., T. M. Snyder, M. R. Corces-Zimmerman, P. Vyas, I. L. Weissman, S. R. Quake and R. Majeti (2012). "Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia." Sci Transl Med **4**(149): 149ra118.

Juntilla, M. M., V. D. Patil, M. Calamito, R. P. Joshi, M. J. Birnbaum and G. A. Koretzky (2010). "AKT1 and AKT2 maintain hematopoietic stem cell function by regulating reactive oxygen species." Blood **115**(20): 4030-4038.

Klco, J. M., D. H. Spencer, C. A. Miller, M. Griffith, T. L. Lamprecht, M. O'Laughlin, C. Fronick, V. Magrini, R. T. Demeter, R. S. Fulton, W. C. Eades, D. C. Link, T. A. Graubert, M. J. Walter, E. R. Mardis, J. F. Dipersio, R. K. Wilson and T. J. Ley (2014). "Functional heterogeneity of genetically defined subclones in acute myeloid leukemia." Cancer Cell **25**(3): 379-392.

Kreso, A. and J. E. Dick (2014). "Evolution of the cancer stem cell model." Cell Stem Cell **14**(3): 275-291.

Kronke, J., L. Bullinger, V. Teleanu, F. Tschurtz, V. I. Gaidzik, M. W. Kuhn, F. G. Rucker, K. Holzmann, P. Paschka, S. Kapp-Schworer, D. Spath, T. Kindler, M. Schittenhelm, J. Krauter, A. Ganser, G. Gohring, B. Schlegelberger, R. F. Schlenk, H. Dohner and K. Dohner (2013). "Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia." Blood **122**(1): 100-108.

Kuhnl, A. and D. Grimwade (2012). "Molecular markers in acute myeloid leukaemia." Int J Hematol **96**(2): 153-163.

Lagadinou, E. D., A. Sach, K. Callahan, R. M. Rossi, S. J. Neering, M. Minhajuddin, J. M. Ashton, S. Pei, V. Grose, K. M. O'Dwyer, J. L. Liesveld, P. S. Brookes, M. W. Becker and C. T. Jordan (2013). "BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells." Cell Stem Cell **12**(3): 329-341.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Lapidot, T., C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M. A. Caligiuri and J. E. Dick (1994). "A cell initiating human acute myeloid leukaemia after transplantation into SCID mice." Nature **367**(6464): 645-648.

Lawrence, M. S., P. Stojanov, P. Polak, G. V. Kryukov, K. Cibulskis, A. Sivachenko, S. L. Carter, C. Stewart, C. H. Mermel, S. A. Roberts, A. Kiezun, P. S. Hammerman, A. McKenna, Y. Drier, L. Zou, A. H. Ramos, T. J. Pugh, N. Stransky, E. Helman, J. Kim, C. Sougnez, L. Ambrogio, E. Nickerson, E. Shefler, M. L. Cortes, D. Auclair, G. Saksena, D. Voet, M. Noble, D. DiCara, P. Lin, L. Lichtenstein, D. I. Heiman, T. Fennell, M. Imielinski, B. Hernandez, E. Hodis, S. Baca, A. M.

Dulak, J. Lohr, D. A. Landau, C. J. Wu, J. Melendez-Zajgla, A. Hidalgo-Miranda, A. Koren, S. A. McCarroll, J. Mora, R. S. Lee, B. Crompton, R. Onofrio, M. Parkin, W. Winckler, K. Ardlie, S. B. Gabriel, C. W. Roberts, J. A. Biegel, K. Stegmaier, A. J. Bass, L. A. Garraway, M. Meyerson, T. R. Golub, D. A. Gordenin, S. Sunyaev, E. S. Lander and G. Getz (2013). "Mutational heterogeneity in cancer and the search for new cancer-associated genes." Nature **499**(7457): 214-218.

Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H. H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur and C. Exome Aggregation (2016). "Analysis of protein-coding genetic variation in 60,706 humans." Nature **536**(7616): 285-291.

Ley, T. J., L. Ding, M. J. Walter, M. D. McLellan, T. Lamprecht, D. E. Larson, C. Kandoth, J. E. Payton, J. Baty, J. Welch, C. C. Harris, C. F. Lichti, R. R. Townsend, R. S. Fulton, D. J. Dooling, D. C. Koboldt, H. Schmidt, Q. Zhang, J. R. Osborne, L. Lin, M. O'Laughlin, J. F. McMichael, K. D. Delehaunty, S. D. McGrath, L. A. Fulton, V. J. Magrini, T. L. Vickery, J. Hundal, L. L. Cook, J. J. Conyers, G. W. Swift, J. P. Reed, P. A. Alldredge, T. Wylie, J. Walker, J. Kalicki, M. A. Watson, S. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, P. Westervelt, M. H. Tomasson, D. C. Link, T. A. Graubert, J. F. DiPersio, E. R. Mardis and R. K. Wilson (2010). "DNMT3A mutations in acute myeloid leukemia." N Engl J Med **363**(25): 2424-2433.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, L. and H. Clevers (2010). "Coexistence of quiescent and active adult stem cells in mammals." Science **327**(5965): 542-545.

Lindsley, R. C., B. G. Mar, E. Mazzola, P. V. Grauman, S. Shareef, S. L. Allen, A. Pigneux, M. Wetzler, R. K. Stuart, H. P. Erba, L. E. Damon, B. L. Powell, N. Lindeman, D. P. Steensma, M. Wadleigh, D. J. DeAngelo, D. Neuberg, R. M. Stone and B. L. Ebert (2015). "Acute myeloid leukemia ontogeny is defined by distinct somatic mutations." Blood **125**(9): 1367-1376.

Longley, D. B., D. P. Harkin and P. G. Johnston (2003). "5-fluorouracil: mechanisms of action and clinical strategies." Nat Rev Cancer **3**(5): 330-338.

Mallardo, M., A. Caronno, G. Pruneri, P. R. Raviele, A. Viale, P. G. Pelicci and E. Colombo (2013). "NPMc+ and FLT3_ITD mutations cooperate in inducing acute leukaemia in a novel mouse model." Leukemia **27**(11): 2248-2251.

Mardis, E. R., L. Ding, D. J. Dooling, D. E. Larson, M. D. McLellan, K. Chen, D. C. Koboldt, R. S. Fulton, K. D. Delehaunty, S. D. McGrath, L. A. Fulton, D. P. Locke, V. J. Magrini, R. M. Abbott, T. L. Vickery, J. S. Reed, J. S. Robinson, T. Wylie, S. M. Smith, L. Carmichael, J. M. Eldred, C. C. Harris, J. Walker, J. B. Peck, F. Du, A. F. Dukes, G. E. Sanderson, A. M. Brummett, E. Clark, J. F. McMichael, R. J. Meyer, J. K. Schindler, C. S. Pohl, J. W. Wallis, X. Shi, L. Lin, H. Schmidt, Y. Tang, C. Haipek, M. E. Wiechert, J. V. Ivy, J. Kalicki, G. Elliott, R. E. Ries, J. E. Payton, P. Westervelt, M. H. Tomasson, M. A. Watson, J. Baty, S. Heath, W. D. Shannon, R. Nagarajan, D. C. Link, M. J. Walter, T. A. Graubert, J. F. DiPersio, R. K. Wilson and T. J. Ley (2009). "Recurring mutations found by sequencing an acute myeloid leukemia genome." N Engl J Med **361**(11): 1058-1066.

McGranahan, N. and C. Swanton (2015). "Biological and therapeutic impact of intratumor heterogeneity in cancer evolution." Cancer Cell **27**(1): 15-26.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

McKenzie, J. L., O. I. Gan, M. Doedens and J. E. Dick (2005). "Human short-term repopulating stem cells are efficiently detected following intrafemoral transplantation into NOD/SCID recipients depleted of CD122+ cells." Blood **106**(4): 1259-1261.

Mead, A. J., D. C. Linch, R. K. Hills, K. Wheatley, A. K. Burnett and R. E. Gale (2007). "FLT3 tyrosine kinase domain mutations are biologically distinct from and have a significantly more favorable prognosis than FLT3 internal tandem duplications in patients with acute myeloid leukemia." Blood **110**(4): 1262-1270.

Miller, C. A., J. McMichael, H. X. Dang, C. A. Maher, L. Ding, T. J. Ley, E. R. Mardis and R. K. Wilson (2016). "Visualizing tumor evolution with the fishplot package for R." bioRxiv.

Miller, C. A., B. S. White, N. D. Dees, M. Griffith, J. S. Welch, O. L. Griffith, R. Vij, M. H. Tomasson, T. A. Graubert, M. J. Walter, M. J. Ellis, W. Schierding, J. F. DiPersio, T. J. Ley, E. R.

Mardis, R. K. Wilson and L. Ding (2014). "SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution." PLoS Comput Biol **10**(8): e1003665.

Mohrin, M., E. Bourke, D. Alexander, M. R. Warr, K. Barry-Holson, M. M. Le Beau, C. G. Morrison and E. Passegue (2010). "Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis." Cell Stem Cell **7**(2): 174-185.

Myers, C. D., F. E. Katz, G. Joshi and J. L. Millar (1984). "A cell line secreting stimulating factors for CFU-GEMM culture." Blood **64**(1): 152-155.

Navin, N. E. (2014). "Cancer genomics: one cell at a time." Genome Biol **15**(8): 452.

O'Brien, C. A., A. Pollett, S. Gallinger and J. E. Dick (2007). "A human colon cancer cell capable of initiating tumour growth in immunodeficient mice." Nature **445**(7123): 106-110.

Paguirigan, A. L., J. Smith, S. Meshinchi, M. Carroll, C. Maley and J. P. Radich (2015). "Single-cell genotyping demonstrates complex clonal diversity in acute myeloid leukemia." Sci Transl Med **7**(281): 281re282.

Papaemmanuil, E., M. Gerstung, L. Bullinger, V. I. Gaidzik, P. Paschka, N. D. Roberts, N. E. Potter, M. Heuser, F. Thol, N. Bolli, G. Gundem, P. Van Loo, I. Martincorena, P. Ganly, L. Mudie, S. McLaren, S. O'Meara, K. Raine, D. R. Jones, J. W. Teague, A. P. Butler, M. F. Greaves, A. Ganser, K. Dohner, R. F. Schlenk, H. Dohner and P. J. Campbell (2016). "Genomic Classification and Prognosis in Acute Myeloid Leukemia." N Engl J Med **374**(23): 2209-2221.

Pear, W. S., G. P. Nolan, M. L. Scott and D. Baltimore (1993). "Production of high-titer helper-free retroviruses by transient transfection." Proc Natl Acad Sci U S A **90**(18): 8392-8396.

Pece, S., D. Tosoni, S. Confalonieri, G. Mazzarol, M. Vecchi, S. Ronzoni, L. Bernard, G. Viale, P. G. Pelicci and P. P. Di Fiore (2010). "Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content." Cell **140**(1): 62-73.

Qiu, J., D. Papatsenko, X. Niu, C. Schaniel and K. Moore (2014). "Divisional history and hematopoietic stem cell function during homeostasis." Stem Cell Reports **2**(4): 473-490.

Quentmeier, H., M. Zaborski and H. G. Drexler (1997). "The human bladder carcinoma cell line 5637 constitutively secretes functional cytokines." Leuk Res **21**(4): 343-350.

Quintana, E., M. Shackleton, H. R. Foster, D. R. Fullen, M. S. Sabel, T. M. Johnson and S. J. Morrison (2010). "Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized." Cancer Cell **18**(5): 510-523.

Quintana, E., M. Shackleton, M. S. Sabel, D. R. Fullen, T. M. Johnson and S. J. Morrison (2008). "Efficient tumour formation by single human melanoma cells." Nature **456**(7222): 593-598.

Randall, T. D. and I. L. Weissman (1997). "Phenotypic and functional changes induced at the clonal level in hematopoietic stem cells after 5-fluorouracil treatment." Blood **89**(10): 3596-3606.

Reya, T., S. J. Morrison, M. F. Clarke and I. L. Weissman (2001). "Stem cells, cancer, and cancer stem cells." Nature **414**(6859): 105-111.

Rieger, M. A. and T. Schroeder (2012). "Hematopoiesis." Cold Spring Harb Perspect Biol **4**(12).

Riva, L., C. Ronchini, M. Bodini, F. Lo-Coco, S. Lavorgna, T. Ottone, G. Martinelli, I. Iacobucci, C. Tarella, A. Cignetti, S. Volorio, L. Bernard, A. Russo, G. E. Melloni, L. Luzi, M. Alcalay, G. I. Dellino and P. G. Pelicci (2013). "Acute promyelocytic leukemias share cooperative mutations with other myeloid-leukemia subgroups." Blood Cancer J **3**: e147.

Roberts, S. A. and D. A. Gordenin (2014). "Hypermutation in human cancer genomes: footprints and mechanisms." Nat Rev Cancer **14**(12): 786-800.

Roesch, A., M. Fukunaga-Kalabis, E. C. Schmidt, S. E. Zabierowski, P. A. Brafford, A. Vultur, D. Basu, P. Gimotty, T. Vogt and M. Herlyn (2010). "A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth." Cell **141**(4): 583-594.

Rossello, F. J., R. W. Tothill, K. Britt, K. D. Marini, J. Falzon, D. M. Thomas, C. D. Peacock, L. Marchionni, J. Li, S. Bennett, E. Tantoso, T. Brown, P. Chan, L. G. Martelotto and D. N. Watkins (2013). "Next-generation sequence analysis of cancer xenograft models." PLoS One **8**(9): e74432.

Rossi, D. J., J. Seita, A. Czechowicz, D. Bhattacharya, D. Bryder and I. L. Weissman (2007). "Hematopoietic stem cell quiescence attenuates DNA damage response and permits DNA damage accumulation during aging." Cell Cycle **6**(19): 2371-2376.

Saito, Y., N. Uchida, S. Tanaka, N. Suzuki, M. Tomizawa-Murasawa, A. Sone, Y. Najima, S. Takagi, Y. Aoki, A. Wake, S. Taniguchi, L. D. Shultz and F. Ishikawa (2010). "Induction of cell cycle entry eliminates human leukemia stem cells in a mouse model of AML." Nat Biotechnol **28**(3): 275-280.

Sambrook, J. and D. W. Russell (2006). "Purification of Nucleic Acids by Extraction with Phenol:Chloroform." Cold Spring Harbor Protocols **2006**(1): pdb.prot4455.

Sarry, J. E., K. Murphy, R. Perry, P. V. Sanchez, A. Secreto, C. Keefer, C. R. Swider, A. C. Strzelecki, C. Cavelier, C. Recher, V. Mansat-De Mas, E. Delabesse, G. Danet-Desnoyers and M.

Carroll (2011). "Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rgammac-deficient mice." J Clin Invest **121**(1): 384-395.

Saultz, J. N. and R. Garzon (2016). "Acute Myeloid Leukemia: A Concise Review." J Clin Med **5**(3).

Sawen, P., S. Lang, P. Mandal, D. J. Rossi, S. Soneji and D. Bryder (2016). "Mitotic History Reveals Distinct Stem Cell Populations and Their Contributions to Hematopoiesis." Cell Rep **14**(12): 2809-2818.

Shlush, L. I., N. Chapal-Ilani, R. Adar, N. Pery, Y. Maruvka, A. Spiro, R. Shouval, J. M. Rowe, M. Tzukerman, D. Bercovich, S. Izraeli, G. Marcucci, C. D. Bloomfield, T. Zuckerman, K. Skorecki and E. Shapiro (2012). "Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability." Blood **120**(3): 603-612.

Shlush, L. I. and A. Mitchell (2015). "AML evolution from preleukemia to leukemia and relapse." Best Pract Res Clin Haematol **28**(2-3): 81-89.

Shlush, L. I., S. Zandi, A. Mitchell, W. C. Chen, J. M. Brandwein, V. Gupta, J. A. Kennedy, A. D. Schimmer, A. C. Schuh, K. W. Yee, J. L. McLeod, M. Doedens, J. J. Medeiros, R. Marke, H. J. Kim, K. Lee, J. D. McPherson, T. J. Hudson, H. P.-L. G. P. Consortium, A. M. Brown, F. Yousif, Q. M. Trinh, L. D. Stein, M. D. Minden, J. C. Wang and J. E. Dick (2014). "Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia." Nature **506**(7488): 328-333.

Singh, S. K., C. Hawkins, I. D. Clarke, J. A. Squire, J. Bayani, T. Hide, R. M. Henkelman, M. D. Cusimano and P. B. Dirks (2004). "Identification of human brain tumour initiating cells." Nature **432**(7015): 396-401.

Stratton, M. R., P. J. Campbell and P. A. Futreal (2009). "The cancer genome." Nature **458**(7239): 719-724.

Sun, J., A. Ramos, B. Chapman, J. B. Johnnidis, L. Le, Y. J. Ho, A. Klein, O. Hofmann and F. D. Camargo (2014). "Clonal dynamics of native haematopoiesis." Nature **514**(7522): 322-327.

Takizawa, H., R. R. Regoes, C. S. Boddupalli, S. Bonhoeffer and M. G. Manz (2011). "Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation." J Exp Med **208**(2): 273-284.

Tanei, T., K. Morimoto, K. Shimazu, S. J. Kim, Y. Tanji, T. Taguchi, Y. Tamaki and S. Noguchi (2009). "Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers." Clin Cancer Res **15**(12): 4234-4241.

Taussig, D. C., F. Miraki-Moud, F. Anjos-Afonso, D. J. Pearce, K. Allen, C. Ridler, D. Lillington, H. Oakervee, J. Cavenagh, S. G. Agrawal, T. A. Lister, J. G. Gribben and D. Bonnet (2008). "Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells." Blood **112**(3): 568-575.

Taussig, D. C., J. Vargaftig, F. Miraki-Moud, E. Griessinger, K. Sharrock, T. Luke, D. Lillington, H. Oakervee, J. Cavenagh, S. G. Agrawal, T. A. Lister, J. G. Gribben and D. Bonnet (2010). "Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction." Blood **115**(10): 1976-1984.

Tentler, J. J., A. C. Tan, C. D. Weekes, A. Jimeno, S. Leong, T. M. Pitts, J. J. Arcaroli, W. A. Messersmith and S. G. Eckhardt (2012). "Patient-derived tumour xenografts as models for oncology drug development." Nat Rev Clin Oncol **9**(6): 338-350.

van Rhenen, A., N. Feller, A. Kelder, A. H. Westra, E. Rombouts, S. Zweegman, M. A. van der Pol, Q. Waisfisz, G. J. Ossenkoppele and G. J. Schuurhuis (2005). "High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival." Clin Cancer Res **11**(18): 6520-6527.

Vardiman, J. W., J. Thiele, D. A. Arber, R. D. Brunning, M. J. Borowitz, A. Porwit, N. L. Harris, M. M. Le Beau, E. Hellstrom-Lindberg, A. Tefferi and C. D. Bloomfield (2009). "The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes." Blood **114**(5): 937-951.

Vassiliou, G. S., J. L. Cooper, R. Rad, J. Li, S. Rice, A. Uren, L. Rad, P. Ellis, R. Andrews, R. Banerjee, C. Grove, W. Wang, P. Liu, P. Wright, M. Arends and A. Bradley (2011). "Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice." Nat Genet **43**(5): 470-475.

Verovskaya, E., M. J. Broekhuis, E. Zwart, E. Weersing, M. Ritsema, L. J. Bosman, T. van Poele, G. de Haan and L. V. Bystrykh (2014). "Asymmetry in skeletal distribution of mouse hematopoietic stem cell clones and their equilibration by mobilizing cytokines." J Exp Med **211**(3): 487-497.

Viale, A., F. De Franco, A. Orleth, V. Cambiaghi, V. Giuliani, D. Bossi, C. Ronchini, S. Ronzoni, I. Muradore, S. Monestiroli, A. Gobbi, M. Alcalay, S. Minucci and P. G. Pelicci (2009). "Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells." Nature **457**(7225): 51-56.

Walter, D., A. Lier, A. Geiselhart, F. B. Thalheimer, S. Huntscha, M. C. Sobotta, B. Moehrle, D. Brocks, I. Bayindir, P. Kaschutnig, K. Muedder, C. Klein, A. Jauch, T. Schroeder, H. Geiger, T. P.

Dick, T. Holland-Letz, P. Schmezer, S. W. Lane, M. A. Rieger, M. A. Essers, D. A. Williams, A. Trumpp and M. D. Milsom (2015). "Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells." Nature **520**(7548): 549-552.

Wang, K., M. Li and H. Hakonarson (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." Nucleic Acids Res **38**(16): e164.

Watson, J. V., S. H. Chambers and P. J. Smith (1987). "A pragmatic approach to the analysis of DNA histograms with a definable G1 peak." Cytometry **8**(1): 1-8.

Weissman, I. L. (2000). "Stem cells: units of development, units of regeneration, and units in evolution." Cell **100**(1): 157-168.

Welch, J. S., T. J. Ley, D. C. Link, C. A. Miller, D. E. Larson, D. C. Koboldt, L. D. Wartman, T. L. Lamprecht, F. Liu, J. Xia, C. Kandoth, R. S. Fulton, M. D. McLellan, D. J. Dooling, J. W. Wallis, K. Chen, C. C. Harris, H. K. Schmidt, J. M. Kalicki-Veizer, C. Lu, Q. Zhang, L. Lin, M. D. O'Laughlin, J. F. McMichael, K. D. Delehaunty, L. A. Fulton, V. J. Magrini, S. D. McGrath, R. T. Demeter, T. L. Vickery, J. Hundal, L. L. Cook, G. W. Swift, J. P. Reed, P. A. Alldredge, T. N. Wylie, J. R. Walker, M. A. Watson, S. E. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, J. E. Payton, J. D. Baty, S. Kulkarni, J. M. Klco, M. H. Tomasson, P. Westervelt, M. J. Walter, T. A. Graubert, J. F. DiPersio, L. Ding, E. R. Mardis and R. K. Wilson (2012). "The origin and evolution of mutations in acute myeloid leukemia." Cell **150**(2): 264-278.

Williams, M. J., B. Werner, C. P. Barnes, T. A. Graham and A. Sottoriva (2016). "Identification of neutral tumor evolution across cancer types." Nat Genet **48**(3): 238-244.

Wilson, A., E. Laurenti, G. Oser, R. C. van der Wath, W. Blanco-Bose, M. Jaworski, S. Offner, C. F. Dunant, L. Eshkind, E. Bockamp, P. Lio, H. R. Macdonald and A. Trumpp (2008). "Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair." Cell **135**(6): 1118-1129.

Xie, M., C. Lu, J. Wang, M. D. McLellan, K. J. Johnson, M. C. Wendl, J. F. McMichael, H. K. Schmidt, V. Yellapantula, C. A. Miller, B. A. Ozenberger, J. S. Welch, D. C. Link, M. J. Walter, E. R. Mardis, J. F. Dipersio, F. Chen, R. K. Wilson, T. J. Ley and L. Ding (2014). "Age-related mutations associated with clonal hematopoietic expansion and malignancies." Nat Med **20**(12): 1472-1478.

Yamashita, Y., J. Yuan, I. Suetake, H. Suzuki, Y. Ishikawa, Y. L. Choi, T. Ueno, M. Soda, T. Hamada, H. Haruta, S. Takada, Y. Miyazaki, H. Kiyoi, E. Ito, T. Naoe, M. Tomonaga, M. Toyota, S. Tajima, A. Iwama and H. Mano (2010). "Array-based genomic resequencing of human leukemia." Oncogene **29**(25): 3723-3731.

Zufferey, R., T. Dull, R. J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini and D. Trono (1998). "Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery." J Virol **72**(12): 9873-9880.

Ιθάκη

Σα βγεις στον πηγαιμό για την Ιθάκη,
να εύχεσαι νάναι μακρύς ο δρόμος,
γεμάτος περιπέτειες, γεμάτος γνώσεις.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον θυμωμένο Ποσειδώνα μη φοβάσαι,
τέτοια στον δρόμο σου ποτέ σου δεν θα βρεις,
αν μέν' η σκέψις σου υψηλή, αν εκλεκτή
συγκίνησις το πνεύμα και το σώμα σου αγγίζει.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον άγριο Ποσειδώνα δεν θα συναντήσεις,
αν δεν τους κουβανείς μες στην ψυχή σου,
αν η ψυχή σου δεν τους στήνει εμπρός σου.

Να εύχεσαι νάναι μακρύς ο δρόμος.
Πολλά τα καλοκαιρινά πρωιά να είναι
που με τι ευχαρίστησι, με τι χαρά
θα μπαίνεις σε λιμένας πρωτοειδωμένους·
να σταματήσεις σ' εμπορεία Φοινικικά,
και τες καλές πραγμάτειες ν' αποκτήσεις,
σεντέφια και κοράλλια, κεχριμπάρια κ' έβενους,
και ηδονικά μυρωδικά κάθε λογής,
όσο μπορείς πιο άφθονα ηδονικά μυρωδικά·
σε πόλεις Αιγυπτιακές πολλές να πας,
να μάθεις και να μάθεις απ' τους σπουδασμένους.

Πάντα στον νου σου νάχεις την Ιθάκη.
Το φθάσιμον εκεί είν' ο προορισμός σου.
Αλλά μη βιάζεις το ταξείδι διόλου.
Καλλίτερα χρόνια πολλά να διαρκέσει·
και γέρος πια ν' αράξεις στο νησί,
πλούσιος με όσα κέρδισες στον δρόμο,
μη προσδοκώντας πλούτη να σε δώσει η Ιθάκη.

Η Ιθάκη σ' έδωσε τ' ωραίο ταξείδι.
Χωρίς αυτήν δεν θάβγαινες στον δρόμο.
Άλλα δεν έχει να σε δώσει πια.

Κι αν πτωχική την βρεις, η Ιθάκη δεν σε γέλασε.
Έτσι σοφός που έγινες, με τόση πείρα,
ήδη θα το κατάλαβες η Ιθάκες τι σημαίνουν.

Κ. Π. Καβάφης