Unravelling the molecular role of the long non-coding RNA FAM30A in modulating acute myeloid leukaemia stem cells dynamics

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I - Biowissenschaften

der Martin-Luther-Universität Halle-Wittenberg,

vorgelegt

von Herrn Jaime Calvo Sánchez

[redacted]

1. Gutachter/in: Prof. Dr. Stefan Hüttelmaier
2. Gutachter/in: Prof. Dr. Daniel Zenklusen
3. Gutachter/in: PD Dr. med. Jessica Höll

Datum der öffentlichen Verteidigung: 12.12.2024

Table of contents

1. INTRODUCTION

1.1- Transition from normal to malignant haematopoiesis

1.1.1.- Normal haematopoiesis

The life-long production of blood cells is governed by a process known as haematopoiesis. Acting as a highly organized, yet highly responsive system, it ensures the steady replenishment of mature blood cells as these cells lack the ability to proliferate indefinitely and often have limited lifespans. In humans, this process undergoes a developmental shift in location: initial blood cell production occurs in the blood islands of the yolk sac, transitions temporarily to the liver during later embryogenesis and eventually settles in the bone marrow and thymus, establishing definitive haematopoiesis in adults. A key distinction between primitive and definitive haematopoiesis lies in the presence of long-term repopulating stem cells, which possess the ability to restore the haematopoietic system in an adult host subjected to lethal irradiation (Jagannathan-Bogdan & Zon, 2013). This process caters to various biological demands, including the production of red blood cells, the generation of platelets, and the formation of myeloid and lymphoid cells, among other processes. Disturbances in the homeostasis of this process can result in haematological pathologies such as leukaemias, lymphomas, anemias, thrombocytopenias and immunodeficiencies (Loughran et al., 2020; Olson et al., 2020; Orkin & Zon, 2008).

The haematopoiesis process results in the production of various types of mature immune cells which can be assigned to two main lineages: the lymphoid lineage including T-cells, B-cells and NK-cells as well as the myeloid lineage comprising erythrocytes (blood cells), megakaryocytes (platelets), monocytes/macrophages and granulocytes (neutrophils, eosinophils, basophils and mast cells) (Kondo et al., 1997; Weissman, 2000). The classical model of haematopoietic differentiation depicts a hierarchical tree structure, wherein a succession of branch fate decisions ultimately gives rise to distinct mature cell types. Sitting at the apex of the tree, the initiation of the haematopoiesis process begins fundamentally with a small and rare population of pluripotent hematopoietic stem cells (HSCs). The prevailing hypothesis suggests that all haematopoietic lineages originate from this common ancestor, the HSC, which is thought to be the only bone marrow (BM) cell population capable of long-term self-renewal and multilineage differentiation (Orkin, 2000; Shizuru et al., 2005). HSCs hold paramount importance in this intricate process, serving as the foundation for the entire haematopoietic cascade thus a lot of research has been focused on studying and functionally characterizing these particular cells.

1.1.2.- Haematopoetic stem cells - definition, niche and identification

HSCs are primitive cells defined on the basis of two key features: (1) the capacity to produce all immune cells (i.e. multipotency) as well as differentiation to accomplish the functional lineage-committed maturation, and (2) to proliferate while fully maintaining its own population numbers (i.e. self-renewal). The current gold standard to functionally test whether a certain HSC population possesses these characteristics is by transplanting a single HSC into a myeloablated host which should result in long-term reconstitution of both lymphoid and myeloid lineages in primary and serial transplantation (Matsuzaki et al., 2004; Smith et al., 1991). HSCs can be found in different organs such as peripheral blood (PB) and BM. They constitute less than 0.01-0.2% of the total cell population in the BM of humans (Shin et al., 2014; Upadhaya et al., 2018). In order to ensure their life-long functional integrity, the balance between HSCs self-renewal and differentiation must be tightly regulated. This is accomplished through a complex interplay of both cell-intrinsic factors (i.e. transcriptional and epigenetic regulators), metabolic pathways, cell-extrinsic factors or local cues from the BM microenvironment, commonly known as the "haematopoietic stem cell niche" or BM niche (Morrison & Scadden, 2014a; Tamma et al., 2017). In contrast to the high turnover of lineage-committed progenitors, most of these HSCs largely reside in a dormant or quiescent G_0 state which protects them against genotoxic insults as well as environmental stresses while the bulk of haematopoiesis is ensured by these lineage-restricted progenitors (Cheshier et al., 1999; Passegué et al., 2005; Sawai et al., 2016; J. Sun et al., 2014). Much of this quiescence state is controlled by a hypoxic microenvironment, self-renewal promoting cytokines and growth factors provided by mesenchymal stromal cells, endothelial cells, osteoblasts and nerve cells (Chandel et al., 2016; A. Wilson et al., 2008).

Identifying and collectively describing HSCs based on their immunophenotype is crucial for their characterization and multiple cellular surface proteins have been found. The adhesion molecule CD34 is commonly used as a convenient cell surface marker for HSCs and early lineage-committed progenitors. CD34+ BM cells have been shown to possess high colony-formation potential in long-term *in vitro* cultures ultimately allowing differentiation of blood lineages in immunocompromised mice (Andrews et al., 1989). HSCs have been functionally profiled through *in vivo* studies as the (Lin-) CD34+ CD38- CD90+ CD49f+ CD45RA- fraction of haematopoietic cells. However, the CD34+ CD38- immunophenotype is predominantly used as this subset is already enriched for HSCs with a high self-renewal capacity (Baum et al., 1992; Lansdorp et al., 1990; Strauss et al., 1986).

1.1.3.- The myeloid lineage and its intricate regulatory network

Depending on its progressively diminished self-renewal ability in response to the earliest differentiation signals, HSCs can be assigned to three categories: long-term (LT-HSC), short-term (ST-HSC) and multipotent progenitors (MPPs). LT-HSCs have a higher self-renewal capacity compared to ST-HSCs which are lineage-committed and thus cannot sustain their self-renewal capacity (Cheng et al., 2020). MPPs have limited or no long-term self-renewal capacity as observed in donour stem or progenitor cells where they were not able to restore haematopoiesis completely (K. Busch et al., 2015; Morrison & Weissman, 1994). In line with the current "classical haematopoiesis model", a strict bifurcation occurs when MPP give rise to lineage-committed progenitors, including common myeloid progenitors (CMPs) or common lymphoid progenitors (CLPs), each of which lacks the ability for self-renewal (Akashi et al., 2000; Kondo et al., 1997). At the second branching point in the myeloid lineage, CMPs subsequently develop into granulocyte-macrophage progenitors (GMPs) which generate granulocytes/monocytes and megakaryocyte-erythroid progenitors (MEPs) which differentiate to megakaryocytes and erythrocytes, respectively (Figure 1).

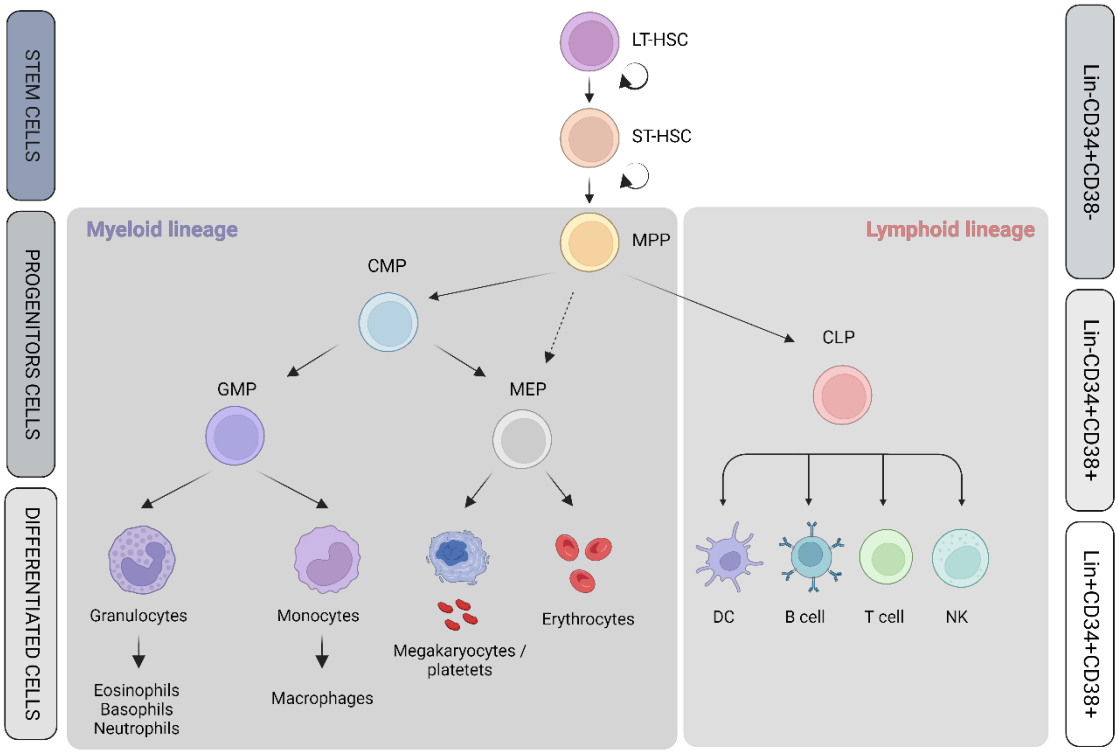


Figure 1. Classical model of the human haematopoiesis tree. Simplified representation of the haematopoiesis process following a hierarchical organization with the HSCs sitting at the top (with different degrees of self-renewal capacity). HSCs and certain types of MPPs lie within the CD34+CD38- fraction, downstream restricted progenitors in Lin-CD34+CD38+ and terminally differentiated cells in the Lin+ (lineage) fractions. Circular arrows indicate self-renewal capacities.
-Continued on next page-

LT-HSC: long-term haematopoietic stem cell, ST-HSC: short-term haematopoietic stem cell, MPP: multipotent progenitor; CMP: common myeloid progenitor, CLP: common lymphoid progenitor, GMP: granulocyte-monocyte progenitor; MEP: megakaryocyte-erythrocyte progenitor; DC: dendritic cell; NK: natural killer cell. Dotted line represents alternative pathways that have been proposed. Figure was adapted from (Doulatov et al., 2012).

Based on the hypothetical model, key intrinsic transcription factors (TFs) together with cytokines, epigenetic regulators and miRNAs collectively regulate the stepwise differentiation process into mature blood cells (Cheng et al., 2020).

The maturation of myeloid cells is orchestrated by a number of lineage-instructive TFs. For instance, at the stem-cell level, the runt-related transcription factor (RUNX1; also known as AML1) is part of the core-binding factor complex (CBF) and also JUNB, MYC, SCL, GFI1, ETV6 and GATA2, among others (Rosenbauer & Tenen, 2007). These TFs ultimately control specific gene expression programs which often induce transcription of growth factor receptors or growth factors important for myeloid differentiation. Among them, RUNX1 is an essential transcription factor at multiple stages of haematopoiesis and for establishing definitive HSCs (Kalev-Zylinska et al., 2002) as well as it plays a significant role in early development of embryonic HSCs (Boisset et al., 2010). RUNX1 targets several genes many of which are also pivotal transcription regulators involved in the formation of all haematopoietic lineages, especially in the myeloid compartment (Hoogenkamp et al., 2007).

RUNX1 regulates gene expression of various haematopoietic genes, including those coding for growth factors (e.g. GM-CSF, IL3, MPO), surface receptors (e.g. CD34, FLT3, M-CSF receptor), signalling molecules (e.g. CDKN1A, BCL2) or transcription factors (e.g. STAT3, MYB). However, conditional deletion of RUNX1 in adult haematopoietic cells did not cause complete loss of adult haematopoiesis. This indicates that RUNX1 is not required for maintenance of HSCs but significant downstream lineage-specific defects were seen (Ichikawa et al., 2004; Levantini et al., 2011; Okuda et al., 1996).

RUNX1 is found in close genetic proximity with TFs specific to the HSC/MPP compartment (N. K. Wilson et al., 2010), yet it is also involved in association with TFs complexes linked with MEPs (Goldfarb, 2009; Pimkin et al., 2014), GMPs (Rosenbauer & Tenen, 2007) as well as within the lymphoid compartment (Collins et al., 2009; Niebuhr et al., 2013). This implies a key role of RUNX1 in controlling self-renewal and differentiation decisions, especially in the myeloid-cell fate program. For example, one of the major RUNX1 target genes is PU.1 whose transcriptional predominance is one of the earliest events biasing HSCs to myeloid commitment by modulating expression levels of GM-CSF, IL-3 cytokines and M-CSFR cytokines receptors (Gerritsen et al.,

2019; Harada et al., 2013). Minimal amounts of PU.1 expression are required for HSCs maintenance, whereas elevated levels trigger myeloid differentiation (G. Huang et al., 2008; Imperato et al., 2015).

As Figure 2 shows, the transcription of the human RUNX1 gene is controlled by two distinct promoters that generate ≥12 different RUNX1 mRNA isoforms and three main protein isoforms. These differ in their 5´ and 3´untranslated region (UTR), coding sequence, RNA stability and translation efficiency that ultimately contribute to diverse effects on haematopoiesis (Levanon et al., 2001; Miyoshi et al., 1995). The distal promoter (P1 promoter) regulates the production of RUNX1C isoform and it is located 130 kb upstream of the proximal promoter (P2 promoter) that controls transcription of RUNX1B and RUNX1A isoforms (Ghozi et al., 1996; Tanaka et al., 1995). Interestingly, the P1 promoter contains binding motifs for several key haematopoietic transcription factors while the usage of the P2 promoter is more basic (Martinez et al., 2016). All three protein isoforms contain the Runt homology domain necessary for DNA binding and heterodimerization with CBFβ (De Bruijn & Dzierzak, 2017). Only RUNX1B and RUNX1C contain downstream transcriptional regulatory domains (transactivation or TAD and inhibitory domains) required for recruiting essential cofactors (Rozen et al., 2023).

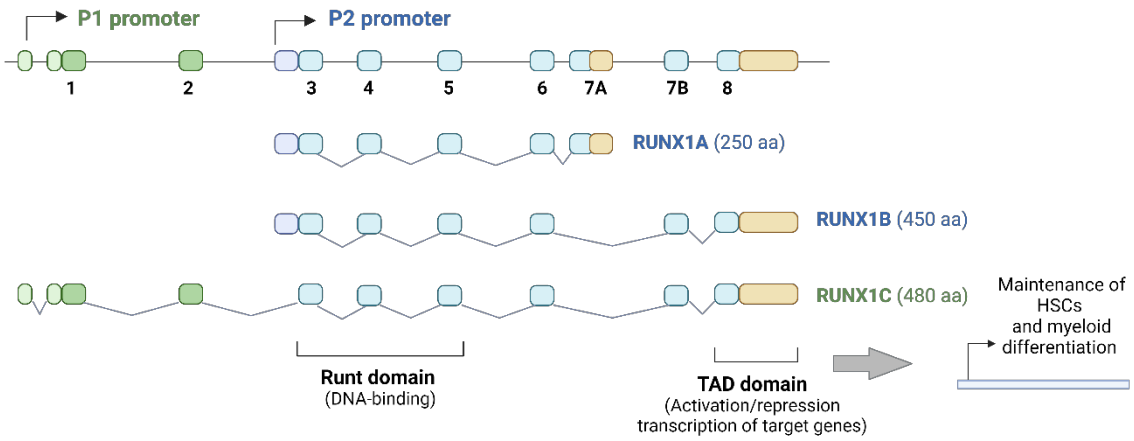


Figure 2. RUNX1 genomic locus and its main protein isoforms. The RUNX1 genomic locus on human chromosome 21 is shown with the location of the proximal and distal promoters. The three main transcriptional isoforms of RUNX1 (with P1-derived exons highlighted in green and P2-derived isoforms in blue) are shown with their protein sizes (amino acids, aa) and the main structural domains highlighted. TAD: transactivation domain. HSPCs: Haematopoetic stem and progenitor cells. Light green and blue exons represent the 5´UTR from P1 and P2-derived isoforms, respectively. Orange exons represent the 3´UTR from the RUNX1 isoforms.

RUNX1B and RUNX1C are the longer isoforms and they differ by 27 amino acids at their N-terminus which results also in different 5´UTR. The 5´UTR of RUNX1B spans more than 1.6 kb comprising GC-rich regions and putative internal ribosomal entry sites (IRES) elements that suggest differential control of translation compared to RUNX1C (Nieke et

al., 2017). RUNX1A lacks the transactivation and inhibitory domains at the C-terminus, while retaining the DNA-binding domain and has been long proposed as a natural dominant repressor (Levanon et al., 2001; Miyoshi et al., 1991) and contributes to leukaemogenesis (X. Liu et al., 2009). This structural divergence has been shown to enhance DNA binding and antagonises the transcriptional programs driven by RUNX1B/C (Gialesaki et al., 2023; Guo et al., 2012; Liu et al., 2009). Consistenly, expression of the different RUNX1 isoforms results in diverse effects in cell growth and differentiation in a myeloid progenitor cell line (Telfer & Rothenberg, 2001).

Several discrepancies exist concerning the temporal expression and the importance of each RUNX1 isoform in both embryonic and adult haematopoiesis. The P2 promoter usage is far more heterogenous and its promoter activity precedes that of P1 before emergence of HSCs. RUNX1B is expressed in a specific endothelial subpopulation prior to haematopoietic commitment and its expression is more limited and restricted to MEPs and lymphoid lineages (Draper et al., 2016; Menegatti et al., 2021). RUNX1A constitutes a small portion within the overall pool of RUNX1 isoforms and its enforced expression has been shown to influence self-renewal capacity and long-term expansion of HSPCs *ex vivo* (Ran et al., 2013). On the other hand, P1 is the dominant promoter in adult haematopoiesis thus RUNX1C is the most abundant protein isoform in definitive haematopoietic cells in adult BM (Challen & Goodell, 2010). Its expression correlates with early haematopoietic markers and it is restricted to a subpopulation of CD34+ cycling HSCs (Ferrell et al., 2015) as well as its function is required for maintenance of HSPCs with engraftment potential (Lis et al., 2017; Sugimura et al., 2017). Moreover, RUNX1C is the earliest isoform and its upregulation is required during haematopoietic specification of HSPCs (Navarro-Montero et al., 2017) and its genetic deletion results in numerous lineage-specific defects (Draper et al., 2016).

Due to RUNX1 significance throughout haematopoiesis, any dysregulation in its regulatory network directly contributes to malignant haematopoiesis and its disruption leads to a failure of the entire blood development (Okuda et al., 1996; Q. Wang et al., 1996). In fact, mutations in the RUNX1 gene stand out as a highly prevalent genetic event in haematological malignancies, especially in acute myeloid leukaemia (AML) (Bellissimo & Speck, 2017; Gaidzik et al., 2016). Interestingly, studies reported higher frequency of the CD34+ compartment in RUNX1-mutated AML blasts which might reflect the association with more immature AML phenotypes (Langabeer et al., 2002; Schnittger et al., 2011).

1.1.4.- Clonal disease evolution

Under normal conditions, the haematopoietic compartment is an inherently tumour-suppressive system continuously regulating the small size of the HSC pool. The mechanism that keeps the pool constant relies on an asymmetrical division of the HSC, assuring that always one of the daughter cells remains an HSC (Riether et al., 2015). Even though haematopoiesis remains a highly efficient process for decades during human life, it is inevitably affected by the erosive effects of ageing and the acquisition of somatic DNA mutations (De Haan & Lazare, 2018; Mohrin et al., 2010). In response to regenerative signals that lead to excessive HSC activation in the BM niche, a shift towards differentiation occurs at the expense of self-renewal resulting in the depletion of the HSC compartment. Conversely, under conditions where differentiation is affected and self-renewal prioritized, the HSC compartment is excessively proliferative, detrimentally affecting the effective maintenance of the production of blood cells. For instance, processes such as chronic inflammation, heritable genetic variation and genotoxic stress can favour the emergence of mutations that inherently augment HSC fitness and drive clonal expansion of mutant HSCs (Bick et al., 2020; Coombs et al., 2017; Gibson et al., 2017). This phenomenon is referred to as clonal haematopoiesis (CH), and it typically involves genetic alterations (i.e. genetic mutations, chromosomal changes) in a limited set of oncogenes that are also frequently mutated in leukaemia, such as DNMT3A, TET2, ASXL1, TP53, RUNX1, BCR-ABL1 and SF3B1 (Genovese et al., 2014; Jaiswal & Ebert, 2019; McKerrell et al., 2015; Xie et al., 2014). In this framework, these so-called driver mutations confer a clonal growth advantage and can be selected through the oncogenic process. Simultaneously, cells that accumulate neutral mutations (also called passenger mutations) which lack any fitness advantage are, therefore, exempt from natural selection (Bowman et al., 2018; Pich et al., 2022). It's noteworthy that CH is not exclusive to malignant conditions -only established as a precursor condition to haematologic malignancies- and can also be observed in healthy individuals (Fabre et al., 2022). In general terms, CH heightens the risk of developing myeloid neoplasms through diverse potential mechanisms such as alterations in the BM microenvironment, chronic inflammatory signals in clonal effector cells or abnormal immune responses, among others. When this is further disrupted due to the heightened self-renewal capacity of transformed HSCs with signalling pathways working synergistically, it can lead to leukaemia, myeloma and lymphoma (Mendez & Patnaik, 2023; Tuval & Shlush, 2019). Further classification of these three blood cancer types includes acute myeloid leukaemia (AML), B-cell acute lymphoblastic leukaemia (B-ALL), T-cell acute lymphoblastic leukaemia (T-ALL), myelodysplastic syndrome (MDS), myeloproliferative neoplasm (MPN), chronic lymphocytic leukaemia (CLL), and follicular lymphoma (FL)

(Taylor et al., 2017). Even though for most cancers the cancer-initiating cell harbouring the transforming mutations is unknown, there is considerable evidence that most of leukaemia types arise from mutations that accumulate in HSCs (Reya et al., 2001).

Longstanding observations have enabled the identification of pre-leukemic mutations in HSC-derived clones (also named pre-leukemic stem cells or pre-LSCs). These pre-LSCs give rise to healthy blood and immune cells and the occurrence of additional mutational events can hinder the differentiation process, leading to the malignant expansion of abnormal progenitor cells (Warner et al., 2004). Because leukaemogenesis is a multi-step process, additional “hits” are required in pre-LSCs for the formation of fully transformed leukemic stem cells (LSCs) (Pandolfi et al., 2013). Whereas pre-LSCs consisting of self-renewing HSPCs are able to initiate leukaemia *in vivo* with variable latency, LSCs are functionally defined to induce fully penetrant leukaemia with rapid onset in murine transplantation models (Shlush et al., 2017).

The current hypothesis is that cancer evolves through a hierarchical tree structure stemming and maintained by a small pool of cancer stem cells (CSCs) with self-renewal capacities, giving rise to differentiated non-cancerous cells via asymmetric division. Interestingly, the first cancer type for which this model was rigorously tested was AML (Bonnet & Dick, 1997; Lapidot et al., 1994). Since then, this CSC hypothesis has expanded to solid tumours (Dick, 2003; Kreso & Dick, 2014). In haematological malignancies that arise from, and are maintained by LSCs (i.e. CML, AML), significant efforts are being directed towards identifying these therapy-resistant LSCs in order to improve the clinical outcome of these patients. This subject will be explored in detail later within the scope of this thesis.

1.2.- Acute myeloid leukaemia

1.2.1.- Definition, epidemiology and pathophysiology

AML is a haematological malignancy characterized by the clonal expansion of immature myeloid precursors coupled with arrested differentiation that accumulate in the peripheral blood and bone marrow. Consequently, the production of healthy and functional immune cells is significantly reduced resulting in impaired haematopoiesis (Deschler & Lübbert, 2006; Döhner et al., 2015). This is reflected in a list of clinical manifestations which consists of a combination of leucocytosis and signs of bone marrow failure with a rapid onset of symptoms such as anaemia and thrombocytopenia. If left untreated, this disease may be fatal within weeks after diagnosis secondary to infection or bleeding (Versluis et al., 2022).

AML is the most commonly diagnosed acute leukaemia among the adult population, accounting for approx. 80% of all cases, with a median age at diagnosis of 68 years. The incidence of AML increases with age, exponentially increasing to more than 10 times in patients older than 65 years. The estimated 5-year overall survival (OS) is 30% and greatly differs between various groups depending on age and comorbidities, reaching 50% in younger (< 60 years old) but less than 10% in older patients (Döhner et al., 2015; National Cancer Institute, 2019; Sasaki et al., 2021). Although significant progress in the treatment of AML has led to major improvements in the outcome for younger patients, prognosis and survival in the elderly has only marginally improved. This is mainly because until now the current chemotherapy regime has remained the standard of care for AML patients for over 40 years and tolerability in this group of patients is limited (Meyers et al., 2013; A. Shah et al., 2013). The disease outcome in older patients who are unable to receive intense chemotherapy without unacceptable side effects remains unfavourable, with more than 70% of them dying within the first year after diagnosis (Döhner et al., 2022; Sasaki et al., 2021; Shimony et al., 2023).

As mentioned before, the pathogenesis of AML involves the clonal transformation of myeloid stem cells through the acquisition of chromosomal rearrangements and multiple gene mutations that confer proliferative advantage and impair normal haematopoietic differentiation. Studies on animal models contributed to the development of a two-hit model of leukaemogenesis which serves a conceptual framework to categorize the manifold mutations linked to AML (Gary Gilliland & Griffin, 2002). This model proposes that AML originates as a consequence of at least two oncogenic mutations which are classified in two categories: class I mutations that confer a proliferative and/or survival advantages but have no effect on differentiation (e.g. mutations in FLT3, KIT or RAS) and class II mutations that impair haematopoietic differentiation and subsequent apoptosis (i.e. PML-RARA, RUNX1-RUNX1T1 or MLL rearrangements) (Kelly & Gilliland, 2002; Takahashi, 2011). Depending on the clinical ontogeny, AML can be subdivided in three categories: the most common type is *de novo* AML (approx. 80% of all cases) that arises as a new condition, secondary AML or sAML can develop after an antecedent myeloid malignancy (myelodysplastic syndrome or other myeloid neoplasms) and therapy-related AML or t-AML that originates after leukaemogenic therapy (i.e. topoisomerases II, alkylating agents or radiations) (Patel et al., 2012; Pfirrmann et al., 2012; Sill et al., 2011).

1.2.2.- Cytogenetics and molecular abnormalities

AML stands out as one of the cancer types to be extensively studied by innovative high-throughput microarray and sequencing technologies (Shivarov & Bullinger, 2014). This

allowed the identification of multiple recurrent somatically acquired driver mutations coexisting in competing clones which makes AML a very complex and dynamic disease. The mutational landscape of AML is very heterogenous with mutations identified in more than 97% of the patients (Ding et al., 2012; Welch et al., 2012). Only a few mutations (FLT3, NPM1 DNMT3A) are present in more than 25% of AML patients (Döhner et al., 2015), emphasizing the idea that clinical diagnosis of AML encompasses a diverse spectrum of genetically distinct malignancies. From a mechanistic standpoint, these driver mutations generally lead to hyper-activation of signalling cascades that trigger uncontrolled cell proliferation, resistance to apoptosis and aberrantly affect the function of transcription factors vital for normal haematopoietic differentiation.

Targeted sequencing has revealed certain clinically relevant mutations with prognostic and diagnostic value (i.e. FLT3, RUNX1, IDH1, CEBPA) and progress is being made to understand the pathogenic principles governing AML (Papaemmanuil et al., 2016; Richard-Carpentier & DiNardo, 2019). In an effort to dissect the functional crosstalks and interactions between the oncogenic proteins that drive leukaemogenesis, The Cancer Genome Atlas (TCGA) project analyzed the genome of 200 AML patients and described seven functional categories of genes that are commonly mutated in *de novo* AML according to their biological function. This includes mutations in genes involved in signalling pathways (FLT3, KRAS, NRAS and KIT), in genes involved in chromatin modifications and DNA-methylation processes (DNTM3A, TET2, IDH1, IDH2, ASXL1, EZH2 and MLL fusions), mutations or gene fusions in myeloid transcription factors (RUNX1, CEBPA and GATA2 mutations, RUNX1-RUNX1T1, PML-RARA, MYH1-CBFB fusions), mutations in tumour suppressor genes (WT1, TP53, PHF6), mutations in spliceosomal-complex genes (SRSF2, SF3B1, U2AF1 and ZRSR2), mutations in cohesin-complex genes (RAD21, STAG1, STAG2, SMC1A and SMC3) and mutations in the nucleophosmin gene (NPM1) (summarized in Figure 3).

AML pathogenesis requires a limited number of genetic mutations along with one or two cooperative mutations to establish the founding clone (two-hit hypothesis). Interestingly, some of these transcription-factor fusions have been shown to be relevant for disease initiation (Corral et al., 1996; Higuchi et al., 2002; Kuo et al., 2006; Westervelt et al., 2003) but are also mutually exclusive to other common mutations found in AML genes, suggesting that one activating mutation is adequate for AML pathogenesis (Abdel-Wahab & Levine, 2013; Bullinger et al., 2017).

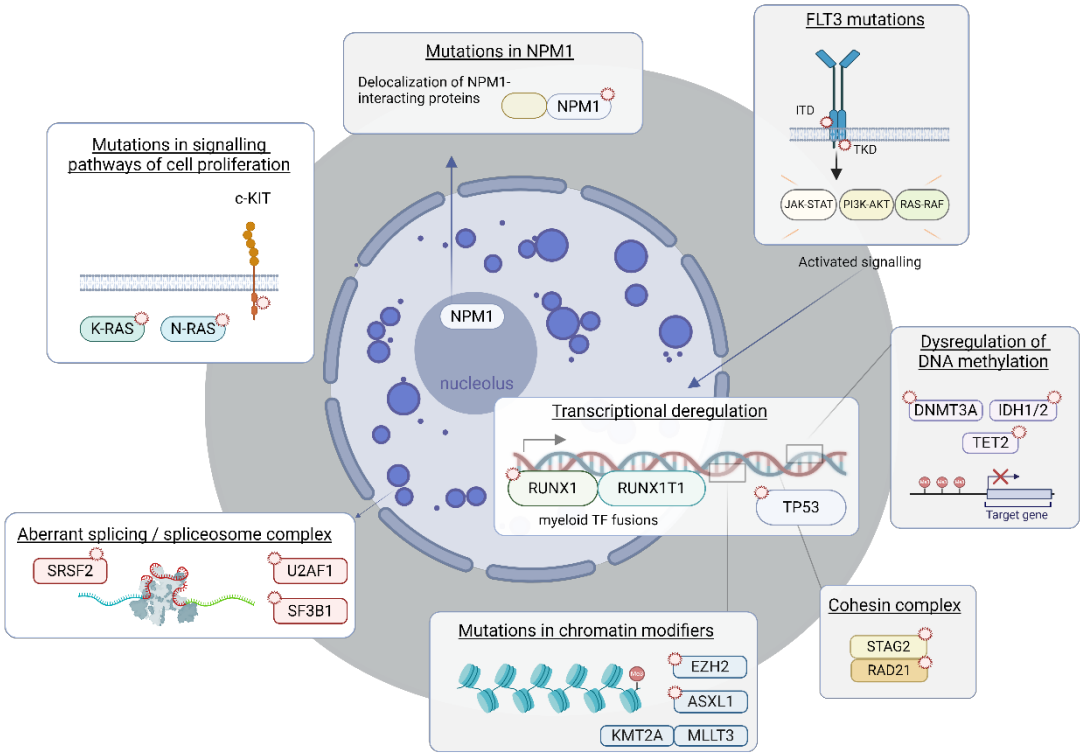


Figure 3. Overview of the mutational landscape in AML. General depiction of the most common mutations found in AML patients and organized based on the data from the functional categories described in The Cancer Genome Atlas (TCGA) network. TF: transcription factor. Red stars represent mutations in the assigned gene. Figure adapted from (Döhner et al., 2015).

Nevertheless, studies integrating mutational analysis with clinical outcome in different AML cohorts suggest that mutations of genes encoding for epigenetic modifiers (i.e. DNMT3A, TET2, ASXL1 or IDH1) are not sufficient for leukemogenesis on their own (Challen et al., 2011; Fisher et al., 2010; Moran-Crusio et al., 2011; Quivoron et al., 2011). Interestingly, these “passenger lesions” were identified in elderly individuals before the occurrence of a clinically apparent AML per se suggesting that these mutations influence the clonal outgrowth of pre-LSCs but they require subsequent additional mutations.

1.2.3.- Disease classification and risk stratification

The progression of AML classification, transitioning from morphology-based to cytogenetic/genetic-focused, underscores the acknowledgment of the critical role played by AML subtypes. This involves examining bone marrow specimens and blood smears morphologically, assessing surface or cytoplasmic markers through flow cytometry, identifying chromosomal abnormalities using conventional cytogenetic testing, and, in more recent approaches, screening for specific molecular genetic mutations (Short et al., 2018).

Based on the morphologic heterogeneity and differentiation status of cells, AML was first classified into eight French–American–British (FAB) subtypes: M0 or minimal myeloid

differentiation (3% of cases); M1 or poorly differentiated myeloblasts (15-20%), M3 or promyelocytic (5-10%), M4 myelo-monoblastic (20%), M5 monoblastic (2-9%), M6 or erythroblastic (3-5%) and M7 or megakaryoblastic (3-12%) (J. M. Bennett et al., 1976, 1985, 1991; Cheson et al., 1990). Although the FAB system lacks the molecular and genetic information that underpins the heterogenous mutational landscape of AML, this diagnostic criterion has not been abandoned but other more precise classifications have been developed.

The latest revised version of the World Health Organisation (WHO) classification system provides an overview of AML subtypes with an emphasis on the integration of clinical, cytomolecular and pathological parameters. This system is arranged mainly into two main groups: AML with defining genetic abnormalities and AML defined by differentiation (similar to the FAB system). The first type includes characteristic gene fusions (e.g. RUNX1::RUNX1T1 or (t(8;21)) and genetic rearrangements involving e.g. KMT2A (11q23), among others (Alaggio et al., 2022).

In 2022, a panel of experts published an updated version of the European LeukemiaNet (ELN) including recommendations for the diagnosis and management of AML which also includes guidelines for risk prediction. Here, depending on the presence of certain genetic alterations, AML patients are divided into favourable, intermediate and adverse risk groups and they also refined the prognostic value of certain genetic abnormalities (Döhner et al., 2022).

1.2.4.- Diagnostic procedures and prognostic factors

Typically, AML is defined by a percentage of immature leukaemic blasts of myeloid origin (at least 20%) in a bone marrow aspirate of peripheral blood smears. Mutations in NPM1 (25-35% of all cases), CEBPA (6-10%) and FLT3-ITD (approx. 20%) are commonly used in clinical practice as diagnostic and predictive markers. Currently, there are other additional mutations included in the prognostic panel (i.e. RUNX1, ASXL1 and TP53) which correlate with inferior clinical outcome. The risk of treatment-related mortality has significantly diminished, especially among older patients, due to considerable advancements in supportive care. However, despite the substantial decrease in treatment-related death rates in recent years, the issue of the disease becoming chemotherapy-resistant remains a significant concern (Liersch et al., 2014; Othus et al., 2014).

1.2.5.- Current and novel therapies in clinical development

The fundamental therapeutic approach for AML patients has remained largely unchanged for over four decades, probably due to the constellation of genetic

aberrations which made it difficult to adopt new personalized approaches to each patient´s karyotype (Döhner et al., 2015; Michaelis, 2018; Short et al., 2018). Initially, first-line treatment involves continuous infusion of Cytarabine (or Ara-C) in combination with the anthracycline Daunorubicin for 3 and 7 days (also known as the “7+3” regime), which remains the mainstay in order to achieve complete remission or response (CR). This means that less than 5% of leukaemic blasts in the bone marrow with no circulating blasts or evidence of extramedullary disease are present (X. Chen et al., 2015). The main goal of the cytotoxic chemotherapy is to impair DNA and RNA synthesis from proliferative leukaemic blasts. CR is normally achieved in 60-85% of adults who are 60 years or younger with a 5-year survival rates of 30-40%. However, in older patients the treatment results are inferior (40-50%) with a 5-year survival rates of less than 10-15% (Kantarjian et al., 2021).

Over the past decade new targeted therapies in combination with initial consolidation therapy have improved clinical outcome due to cytogenetic testing to facilitate appropriate therapy. Remission is a crucial clinical endpoint and it is a requirement for further post-remission treatment such as a high dose cytarabine-based consolidation therapy (salvage) or allogeneic haematopoietic stem cell transplantation (allo-HSCT), the latter one being the most effective treatment to reduce disease relapse in patients with intermediate or adverse risk AML (Koreth et al., 2009). Although the majority of AML patients respond to induction chemotherapy, in case there is an absence of CR after at least two courses of induction chemotherapy regime or allo-HSCT, current literature defines this concept as refractory AML. This group of patients does not generally respond well to chemotherapy treatment and a wide array of salvage regimens is used, making patients response to be broadly influenced by age, karyotype and previous regimens (Othus et al., 2015; Wattad et al., 2017). In patients which are non-eligible for initial intensive chemotherapy (i.e. patients above the age of 65 years or with refractory AML), non-intensive regimes have reshaped the therapeutic landscape. This involves the combination of low-dose cytarabine (LDAC), hypomethylating agents such as azacitidine (CC-486) and venetoclax (B-cell lymphoma [BCL]-2 inhibitor) (DiNardo et al., 2020).

1.2.6.- Relapsed and refractory AML

Despite the previously outlined progress in upfront therapy for AML patients, relapse is the most common cause of treatment failure (DeWolf & Tallman, 2020). At this stage, the primary issue is determining whether a patient is fit to undergo intensive salvage chemotherapy treatment. Approximately, 30-50% of the patients who initially achieve a CR and 20-30% of patients who receive HSCT subsequently will develop relapsed

disease (M. de Lima et al., 2021; Ossenkoppele & Schuurhuis, 2016). The prognosis of relapsed AML patients is broadly poor with a 5-year survival rate of less than 10%.

Mounting evidence traces back the origin of relapsed AML to therapy-resistant leukaemia cells which exhibit stemness-related properties, the so-called leukaemia stem cells (LSCs). The current hypothesis is that disease relapse correlates with founder clone recurrence, subclonal expansion and dynamic changes due to relapse-specific acquired somatic mutations (Rapaport et al., 2021; Vosberg & Greif, 2019). This has been demonstrated by deep-sequencing of purified subsets of AML populations and xenotransplantation assays from relapsed AML samples (Eppert et al., 2011; Shlush et al., 2017). Early identification of this subset of leukaemia cells in AML patients is critical, as it facilitates the accurate allocation of patients to salvage regimens or enrolment to targeted therapies prior to overt AML relapse.

1.3.- Acute myeloid leukaemia stem cells (AML LSCs)

1.3.1.- Discovery, concept and ontogeny of AML LSCs

AML set the precedent to be the first cancer type where the existence of cancer stem cells was experimentally proven. The earliest observations of slow-cycling LSCs in myeloid leukaemia patients can be traced back to more than 50 years go (Clarkson et al., 1967) and later demonstrated with empirical evidence showing the ability of AML LSC to initiate leukaemia in NOD/SCID mice in transplantation assays using a patient sample (Lapidot et al., 1994). This study set a standard method to functionally validate tumour-initiating cells and its self-renewal capacity through serial-dilution xenografts. Since then, substantial progress has been made in the identification of the wide variety of malignant stem cell progenitors responsible to originate and propagate leukaemia in AML patients. This was achieved using fluorescence-activated cell sorting (FACS) to isolate engrafting populations from AML samples (Bonnet & Dick, 1997; Kreso & Dick, 2014). Subsequent improvements of immunocompromised mice strains (i.e. expressing human cytokines) have increased the proportion of engrafting AML samples (Wunderlich et al., 2010). Moreover, further studies have broadly demonstrated that AML LSCs are refractory to standard-of-care chemotherapy regimens (Ishikawa et al., 2007) even though this view has been challenged by other groups that proved that LSCs that survive chemotherapy are molecular and metabolically different (Boyd et al., 2018; Farge et al., 2017). Unfortunately, innovative treatments designed to eradicate AML LSCs have not yet transitioned into clinical practice.

AML LSCs are functionally defined as a rare and heterogenous subpopulation of leukaemia cells with a few distinguishable properties: extensive self-renewal capacity,

mostly in a quiescent state with a distinct metabolism, arrested differentiation, resistant to drugs and localized in protective microenvironments, among other features (Ishikawa et al., 2007). Nevertheless, due to the vast heterogeneity of the driver mutations present among AML patients, the biological properties of AML LSCs greatly differ from patient to patient. As mentioned before, the clonal evolution of AML is a highly dynamic and multi-step process. Defining the cell of origin of LSCs is of great importance in the AML field as it gives us insights of how this disease initiates and evolves. Depending on their developmental stage, AML LSCs can be subdivided in two different types: pre-LSCs and LSCs (also called leukaemia-initiating cells). Seminal discoveries analysing mutational patterns have led to the identification of pre-leukemic clonal states in HSCs (named pre-LSC) which arise due to the sequential acquisition of "early" somatic DNA mutations. Consequently, these first hits (i.e. in epigenetic regulators) generally enhance self-renewal potential and impair the haematopoietic differentiation potential (Corces et al., 2016a; Lu et al., 2012; Shlush et al., 2014). These pre-LSCs are maintained in a "primed" state and have a propensity to transform into LSCs even though they still can contribute to normal haematopoiesis (Corces et al., 2016a; Corces-Zimmerman et al., 2014; Sato et al., 2016) (Figure 4).

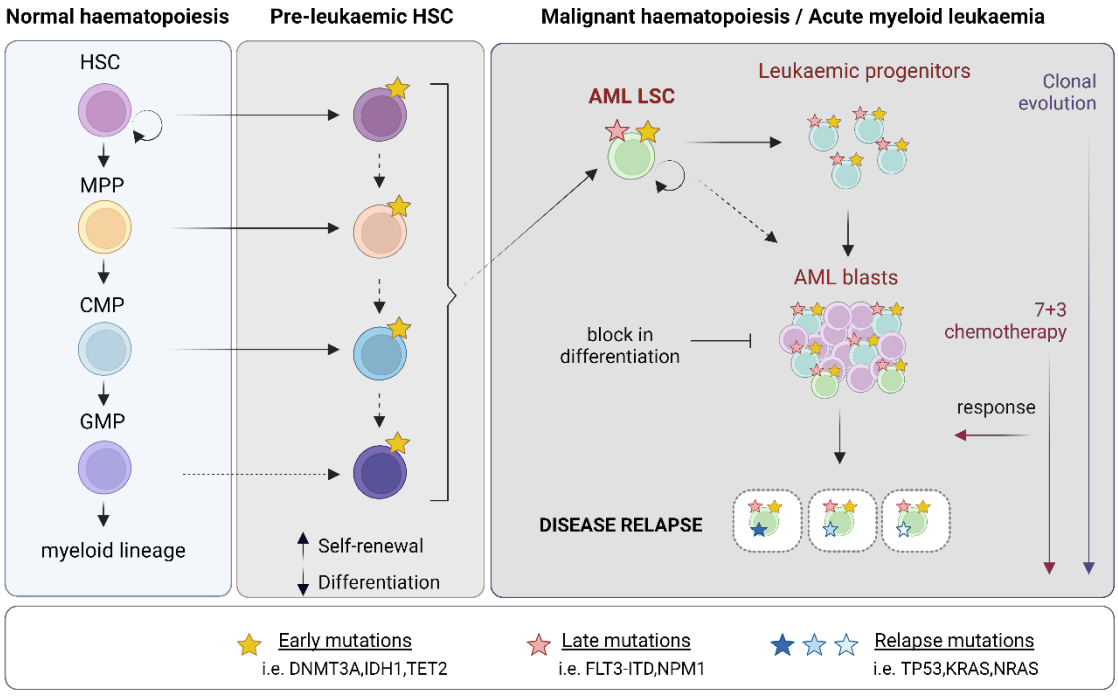


Figure 4. Model of how AML LSCs arise and evolve during disease progression. Depiction of how acute myeloid leukaemia originates from the transformation of normal quiescent or cycling haematopoietic stem cells (HSCs) with high self-renewal capacity, multipotent progenitors (MPPs) or myeloid-committed progenitors (CMPs, GMPs) that acquire primary/early mutations giving rise to pre-leukaemic stem cells (pre-LSCs). Over time, with the acquisition of secondary/late mutations, these further transform into acute myeloid leukaemia stem cells (AML LSCs).
-Continued on next page-

As a result, this sequence of clonal events ultimately leads to the onset of full-blown leukaemia. Following standard induction chemotherapy regime (7+3), a population of therapy-resistant LSCs harbouring additional mutations can survive treatment and hold the capacity to initiate leukaemia thus represent a clinically relevant disease reservoir. Figure adapted from (Long et al., 2022; Vetrie et al., 2020)

In parallel, “late” mutations are mainly found in genes affecting signalling pathways (i.e. FLT3) that give a proliferation advantage and promote an arrest in differentiation that ultimately result in the malignant expansion of aberrant progenitor cells (Gary Gilliland & Griffin, 2002; Zheng & Small, 2005). Only after pre-LSCs transition into LSCs do they differentiate and proliferate vigorously into leukaemic blasts triggering the onset of AML in the bone marrow. These second hits are crucial as the vigorous accumulation of the so-called “blasts” is fuelled by the presence of fully competent LSCs.

1.3.2.- Hallmarks of AML LSCs

As mentioned above, AML LSCs are defined functionally. First, they are highly competent at their unlimited self-renewal capacity and thus able to exhibit long-term survival in optimized *ex vivo* co-culture systems (Udomsakdi et al., 1992) and engraftment into immunocompromised mice (Bonnet & Dick, 1997; Eppert et al., 2011; Lapidot et al., 1994). Importantly, their ability to engraft and initiate human AML in immunocompromised mice is the gold-standard property defining an AML LSC and engraftment levels directly correlate with LSC frequencies and poor clinical outcome (Griessinger et al., 2018; Pearce et al., 2006; Ran et al., 2009). Across several studies, AML patient samples have shown a wide range of percentage of engraftment and also capable to be propagated for multiple serial transplants (Curtis et al., 1984; Delmer et al., 1989). More than 20 years ago, other groups alternatively developed less laborious assays to interrogate stemness (requiring less cell number from AML samples) and allowed the quantitative measure of the clonogenic self-renewal capacity of LSCs and predict AML patient survival. Examples of these techniques are co-culture of AML cells with bone marrow stroma in long-term initiating cell asays (LTC-IC) or colony-forming (CFU) assays in semi-solid medium (i.e. methylcellulose) (Blair & Sutherland, 2000).

AML LSCs are largely cell cycle quiescent (G_0 phase) in order to shield them from genotoxic insults and prevent their exhaustion to mitigate the effects of leukaemia proliferation. This biological characteristic also makes these slow cycling cells to evade cytotoxic effects of chemotherapeutic agents which target rapidly proliferating cells (Guan & Hogge, 2000). This relatively dormant condition is tightly controlled at the molecular level by growth factors, microRNAs and signalling pathways in the bone marrow endosteal region where they generally reside (Niu et al., 2022; Vetrie et al., 2020). Nevertheless, there are instances where the assumed requirement for a

quiescent state does not hold true, as certain LSCs exhibit certain features of actively cycling cells (Iwasaki et al., 2015). Previous studies have shown in AML primary samples that AML LSCs are initially quiescent but they spontaneously exit G_0 phase, thus demonstrating that AML colony-forming cells are more active in cell cycle than their healthy HSC counterparts (Guan & Hogge, 2000). Linked to this quiescent state, AML LSCs have been shown to be metabolically less active (with low oxidative state) compared their HSC counterparts (Cheng et al., 2020). Dysregulation of the reactive oxygen species (ROS) homeostasis (too high or low levels) directly affect LSCs as high levels of these highly reactive byproducts can promote differentiation or impair stemness properties by forcing cells out of quiescence (Adane et al., 2019; Ito & Ito, 2018). Targeting ROS homeostasis is another extensively exploited avenue to target chemotherapy-resistant LSCs. For instance, the recently developed small molecule BCL-2 inhibitors (i.e. venetoclax) have been shown to selectively target ROS-low producing AML LSCs (Lagadinou et al., 2013).

LSCs are considered to harbour inherent therapy resistance to anti-proliferative treatments. Classical chemotherapy regimens primarily target the actively cycling leukaemic blasts resulting in a relatively effective initial remission. Wealthy evidence suggests that AML LSCs preferentially resist to chemotherapy treatment and thus represent a cellular reservoir believed to be the basis for relapsed disease. For example, a study showed a substantial surge in LSC frequencies (from 10 to 90-fold-increase) from diagnosis to relapse in AML patients (Testa et al., 2002).

However, recent studies have shown that both quiescent and cycling AML LSCs are sensitive to chemotherapy (Shlush et al., 2017) and only pre-LSCs can survive chemotherapy (Corces-Zimmerman et al., 2014; Rothenberg-Thurley et al., 2018). During relapse, residual minor clones identified from initial diagnosis emerge and become dominant, incorporating new mutations induced by DNA damage from chemotherapeutic agents. Overall, the sequential accumulation of mutations accelerates LSC clonal diversification with new sub-clonal architectures post-chemotherapy which are more challenging to target, resulting in worse clinical outcome (Corces et al., 2016; Rothenberg-Thurley et al., 2018). AML LSCs also exhibit heightened activity of multi-drug resistance proteins that act as efflux transporters of biomolecules relevant for tumour resistance LSCs (de Grouw et al., 2006; Ho et al., 2008; Vasconcelos et al., 2021).

Several signalling pathways have been implicated in the maintenance, self-renewal and drug-resistance of AML LSCs, including the NOTCH pathway (Reya et al., 2001; Takebe et al., 2011), IL-2/STAT5 pathway (Wingelhofer et al., 2018), hypoxia pathway (Y. Wang et al., 2011), TGFβ pathway (Tabe et al., 2013), NF-κB pathway (Bosman et al., 2016; Gasparini et al., 2014), TNF-α pathway (Kagoya et al., 2014), WNT/β-catenin pathway (Pepe et al., 2022), and Hedgehog pathway (Jagani et al., 2010) (summarized in Figure 5). Aberrant activation of these pathways in AML LSCs occurs frequently through genetic and epigenetic mechanisms, correlating with leukaemia initiation, progression, and resistance to treatment. Other mechanisms of drug resistance include different anti-apoptotic mechanisms (i.e. BCL-2, MCL-1, mutated TP53), senescence resistance, aberrant RNA epigenetic modifications (i.e. m6A) or remodelling of the BM microenvironment, among others (Niu et al., 2022).

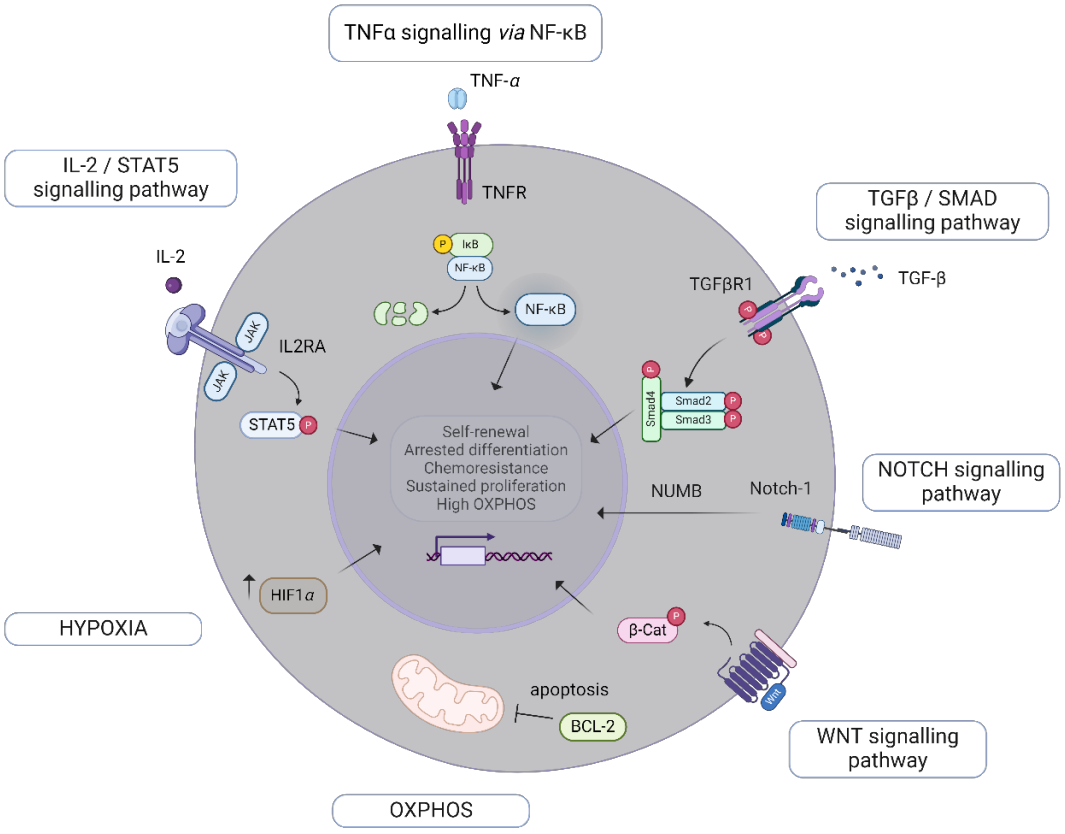


Figure 5. Pathways commonly associated with maintenance and survival of AML LSCs. The illustration summarizes some of the major upregulated pathways involved in the increased self-renewal capacity and cell proliferation, block in differentiation, intrinsic resistance to chemotherapeutic agents and reliance on high oxidative phosphorylation (OXPHOS) that have been reported so far in AML LSCs. TNF-α: tumour necrosis factor alpha, TNFR: tumour necrosis factor receptor, IκB: IkappaB, NF-κB: nuclear factor kappa B, TGFβ: transforming growth factor beta, TGFβR1: transforming growth factor beta receptor I, β-Cat: β-catenin, BCL-2: B-cell lymphoma 2, HIF1α: hypoxia-inducible factor 1-alpha, IL2RA: interleukin-2 receptor alpha. Figure adapted from (Bencomo-Alvarez et al., 2019; Rodrigues et al., 2021).

1.3.3.- AML LSC-specific surface antigens and its potential therapeutic targeting

Over the years, studies that couple immunophenotyping of AML samples and xenotransplantation assays have identified several proteins expressed in functionally defined LSC-enriched populations by comparing AML versus BM samples from healthy donours. The most commonly accepted immunophenotype of AML LSCs is the expression of CD34 and lack of CD38 (CD34+CD38−) as demonstrated by multiple studies via engraftment in immunodeficient mice (Eppert et al., 2011; Goardon et al., 2011; Ishikawa et al., 2007; Sarry et al., 2010) and *in vitro* LTC-IC assays (Blair & Sutherland, 2000). Consistently, this fraction was shown to exhibit higher self-renewal potential and a more immature phenotype. Detailed characterization of the surface immunophenotype showed that CD34+ AML LSCs are hierarchically organized and primarily detected in multipotent progenitors (Lin−CD34+CD38−CD90−CD45RA+) that give rise to granulocyte-macrophage progenitors committed or GMP-like LSCs (Lin⁻CD34+CD38+CD123+CD45RA+) (Goardon et al., 2011). Examination of transcriptional profiles and differentiation potential has elucidated that CD34+ AML LSCs originate from an HSC-like progenitor that acquired heightened self-renewal capacity and gained expression of surface markers from mature myeloid cells (Kirstetter et al., 2008; Somervaille et al., 2009).

However, LSCs can also be present in other subpopulations with less frequency like CD34+CD38+ and even rarer in the CD34- fraction (Kreso & Dick, 2014). Recent studies advocate that LSC might co-exist in all CD34 and CD38 defined subpopulations as their LSC activity was detected in all fractions and their leukaemia-initiating activity assessed via engraftment into mice (Ng et al., 2016). Notably, there were hardly leukaemia cases originating from the CD34- or CD34+/CD38+ fractions and only the CD34+CD38- population and their concomitant subpopulations contained the most important leukaemia-initiating cells. This hypothesis has been confirmed by other studies that observed that only the CD34+CD38- population correlates with engraftment, *in vivo* therapy resistance and minimal residual disease (Terwijn et al., 2014; Van Rhenen et al., 2007). The rare but engrafting CD34- AML LSCs fraction can also induce leukaemia (Quek et al., 2016) and may express other progenitor markers that in turn give rise to a non-LSC population with also a more mature myeloid immunophenotype (Sarry et al., 2010; Taussig et al., 2010).

Unfortunately, there is not yet a distinctive marker that is universally expressed across AML patients solely on CD34+CD38− LSCs but not on the leukaemic blast progeny or on normal HSCs. This complexity partially stems from to the vast heterogeneity of AML, not only among patients but also within the same sample. Regardless of these

challenges, there is a growing list of several cell surface markers upregulated in CD34+CD38− AML LSCs compared to normal CD34+CD38− HSC counterparts that have been identified and serve as biomarkers for LSC enrichment, disease progression or as drug targets. Some examples are: CD25 (Saito, Uchida, et al., 2010), CD33 (Jilani et al., 2002), CD123/IL3RA (Jin et al., 2009) and GPR56 (Pabst et al., 2016), among others. Some of these markers have been demonstrated to be enriched in leukaemia-initiating populations in certain AML patient-derived xenografts (PDX). For instance, CD25 has barely detectable expression levels in normal HSCs but highly expressed in AML LSCs and in bulk AML blasts which is correlated with poor prognosis thus allowing better risk stratification (Aref et al., 2020; Terwijn et al., 2009). These quiescent and multidrug-resistant CD25-positive cells (CD25+CD34+CD38-) from AML patients bear LSC-like molecular signatures (Gönen et al., 2012) and were also capable of leukaemic engraftment in serial transplantation assays in mice (Kageyama et al., 2018; Saito, Uchida, et al., 2010). The lack of universal surface markers for the identification and isolation of heterogeneous AML LSCs poises a significant obstacle, prompting recent investigations to focus on "stemness" gene metrics capable of integrating various AML LSC subtypes across extensive AML sample cohorts.

1.3.4.- AML LSC-specific gene signatures

Similar to their HSC counterparts, LSCs employ analogous molecular mechanisms that drive self-renewal capacity, cell cycle quiescence and differentiation into more committed leukaemic progenitor cells. Despite the biological parallels between AML LSCs and HSCs or MPPs, robust prognostic gene profiles displaying mRNA (Eppert et al., 2011; Ng et al., 2016), ncRNA (Bill et al., 2019) or epigenetic signatures (N. Jung et al., 2015) have been developed and used as drug discovery tools to identify compounds that target AML LSCs *in vitro* and *in vivo* (Laverdière et al., 2018). Assessment of these grouped gene signatures have provided a further refined prognostic model that can predict clinically important parameters tightly linked to poor survival and failure to standard therapy across all AML subtypes.

Initially, Gal et *a*l. were the first study to identify gene expression profiles specific for AML LSCs by comparing CD34+CD38- cells to CD34+CD38+ from AML patients. They identified a list of up- and downregulated genes between the two populations indicating that a big fraction was shared with normal HSCs (Gal et al., 2006). In addition, in order to identify mRNA expression profiles of protein coding genes associated with AML LSCs, Eppert et *al*. compared global gene expression profiles of functionally validated LSC-enriched fractions (i.e. via xenotransplantation assays) with fractions without LSCs and yielded an LSC-specific signature (LSC-R). Further comparison between LSC and

normal HSC-gene signatures led to the identification of a “core enriched” signature comprised of 44 genes altered in processes such as self-renewal, niche dependence and cell cycle quiescence (CE-HSC-LSC) (Eppert et al., 2011). Noteworthy, another study developed a 17-gene stemness signature (LSC17) with differentially expressed genes between 138 LSC+ and 89 LSC- cell fractions from AML patients validated by xenotransplantation assays. It has been demonstrated that AML patients with high LSC17 score do poorly with standard therapy and this score is currently being translated into the clinics with promising results enabling up-front, risk-adapted treatment decisions in patients treated with intensive chemotherapy regimens in paediatric and adult AML cohorts (Bill et al., 2020; Ng et al., 2022; Vasseur et al., 2023). Interestingly, this LSC17 signature includes several protein-coding genes with reported functions in AML LSCs and even therapeutically targetable (i.e. CD34, CDK6 (Z. Li et al., 2021); SOCS2 (Vitali et al., 2015), AKR1C3 (Verma et al., 2016)) but other genes, mostly non-coding, have so far, no reported function.

The LSC17 score is no longer predictive within established cytogenetic and cytomolecular risk groups across pediatric AML (pAML) patients, thus a more robust signature was developed, named LSC47 (pLSC47) signature (B. J. Huang et al., 2022). Moreover, based on the LSC17, the pediatric LSC6 signature (pLSC6) was also developed as it was specifically generated using clinical outcome data from pAML patients (Elsayed et al., 2020). In addition, another group developed a largely mutation-independent LSC DNA methylation signature (LSC71) that comprises 65 new genes that were not previously reported and are significantly hypomethylated in LSC compared to their blast progeny (N. Jung et al., 2015).

1.3.5.- Relevance of RNA-binding proteins function in AML LSCs

While transcriptional and epigenetic mechanisms that underly leukaemogenesis are being thoroughly investigated in AML (Y. Sun et al., 2018; Takei & Kobayashi, 2019), other multiple post-transcriptional events remain comparatively understudied. A myriad of processes take place between RNA synthesis and protein biogenesis which includes (alternative) splicing, mRNA 5´capping, polyadenylation, nuclear export, stabilization, decay and translation of the mRNA, among others. These aforementioned functions are finely orchestrated by dynamic interactions between RNA molecules (i.e. mRNAs, microRNAs and long non-coding RNAs) and RNA-binding proteins (RBPs) via their RNA-binding domains (RBD) or through unconventional RNA binding to ultimately form ribonucleoprotein (RNP) complexes (Corley et al., 2020; Lunde et al., 2007). These structurally well-defined RBDs (i.e. RNA-recognition motif (RRM) or KH domain) recognise consensus sequences in their targets to regulate RNA stability, metabolism

and function. Because the same target sequences can be shared by many different RNAs, the regulation of a cohort of transcripts can be controlled by a single RBP and consequently regulate a plethora of tissue or cell-specific molecular pathways (Müller-McNicoll & Neugebauer, 2013). Conversely, RNAs can also bind to the RBP to affect its localization, biological function and fate (Hentze et al., 2018).

Considering their relevance in shaping the transcriptome and proteome, there is a growing interest in investigating the role of RBPs in controlling myeloid differentiation. In addition, alterations in RBPs such as dysregulation or mutations have also been linked to AML initiation and maintenance (Kharas et al., 2010; Neelamraju et al., 2018; Palanichamy et al., 2016). Global expression analysis in normal and AML samples suggest that RBP expression profiles can not only segregate in myeloid cell types (HSCs, progenitor cells, monocytes) but also in leukaemic cells (LSCs and leukemic blasts) (Saha et al., 2019). In healthy HSCs, a number of RBPs have been reported to be relevant in maintaining HSCs stemness and quiescence (Copley et al., 2013; Rentas et al., 2016). Given that HSCs and LSCs share core transcriptional programs that regulate self-renewal and differentiation, it is to be expected that hijacking normal RBP function can result in malignant haematopoiesis, regardless of their necessity for normal HSCs. In fact, comparison of mRNA expression levels of RBPs between AML patients and human CD34+ HSCs uncovered a RBP network significantly upregulated and crucial for AML survival (E. Wang et al., 2019).

Notably, recent reports are uncovering the molecular dependencies of certain RBPs that are selectively enriched in the fraction of AML LSCs, in particular those involved in splicing, RNA modification, translation and mRNA export (Paris et al., 2019; Park et al., 2014; Vu, Pickering, et al., 2017; Vujovic et al., 2023; E. Wang et al., 2019; Weng et al., 2018). For instance, a study identified a set of 128 RBPs preferentially expressed in LSCs and that are important for self-renewal function, from which 32 were essential for *in vivo* leukaemic propagation (Vujovic et al., 2023). For instance, HNRNPA1 and KHDRBS1 were shown to impact post-transcriptional regulation of RUNX1 isoform generation that contribute to dysregulation in leukaemia and HSC function (Davis et al., 2021; Gialesaki et al., 2023). Moreover, loss-of-function approaches targeting ELAVL1 highlighted a significant inhibitory effect to LSCs through controlling LSC stemness, differentiation arrest and survival (Vujovic et al., 2023). Another LSC-specific RBP is SYNCRYP who was shown to indirectly co-regulate target mRNAs implicated in LSC stemness gene expression programs (Vu, Prieto, et al., 2017).

Mushashi-2 (MSI2) stands out as one of the most relevant AML LSC-specific RBP. Functional studies have demonstrated its enhanced expression in haematological malignancies as well as its well-established role in leukaemic cell fate and stem cell self-renewal. While high MSI2 expression is predominantly observed in haematological neoplasms, MSI1 expression is not as commonly reported. This discrepancy may stem from the recognized role of MSI2 as a regulator within the haematopoietic system (Kharas & Lengner, 2017; Park et al., 2014). MSI2 is highly expressed in more primitive cells like LT-HSCs, ST-HSCs and MPPs and its expression levels declines at later stages of myeloid differentiation. In this context, lack of MSI2 leads to reduced engraftment and depletion of HSCs *in vivo*, poorer engraftment and defects in lympho-myeloid primed progenitors (LMPPs) activity due to failure in maintaining quiescence (De Andrés-Aguayo et al., 2011; Kharas et al., 2010). Moreover, another study highlighted the relevance of MSI2 as an important regulator of the HSC translatome and showed that conditional ablation of MSI2 resulted in a failure of HSC maintenance and engraftment with increased commitment divisions. In contrast, forced expression of MSI2 in human cord blood cells increased long-term repopulating activity of LT-HSC and ST-HSC, enhanced long-term engraftment and impaired myeloid differentiation thus suggesting that MSI2 function is relevant for the maintenance and self-renewal of HSCs in mice (Park et al., 2014). Consistent with this, the relevance and requirement of MSI2 function in leukaemia has been demonstrated in mouse and human systems, especially in myeloid leukaemias. MSI2 has been thoroughly investigated as an independent adverse prognostic marker in AML patients and increased MSI2 expression drives the aggressiveness of leukaemia (Byers et al., 2011; Kharas et al., 2010). The requirement of MSI2 for AML progression was further observed in a panel of human AML cell lines as MSI2 depletion resulted in reduced proliferation, differentiation and increased apoptosis (Kharas et al., 2010).

Mammalian MSI2 has four isoforms and all of them include two highly conserved RRMs in the N-terminal region followed by a less-conserved C-terminal region that allows protein-protein interactions and its susceptible to different post-translation modifications (PTMs) (M. Li et al., 2020). In humans, RRM1 provides preferential RNA binding specificity to tandem UAG-containing target sequences whereas RRM2 has a complementary role by increasing RNA affinity (Ruth Zearfoss et al., 2014; Sakakibara et al., 2001). In general terms, MSI proteins are thought to bind to the 3´UTR of their target RNAs and interact with the polyA-binding protein thus competing for eIF4G leading to inhibition of translation initiation (Ito et al., 2010; Kawahara et al., 2008). However,

more recent studies suggest that MSI2 also promotes the translation of its target mRNAs (Vu, Prieto, et al., 2017; Yeh et al., 2023).

Integrative genome-wide approaches like individual nucleotide resolution cross-linking and immunoprecipitation (i.e. iCLIP, HITS-CLIP) coupled with polysome profiling and other complementary techniques like HyperTRIBE (identification of RBP targets by RNA editing) have revealed that even though MSI2 has a vast mRNA interactome, translational differences are restricted only to a small fraction of these transcripts (Karmakar et al., 2022; Nguyen et al., 2020). Of particular interest is the fact that MSI2 gain- or loss-of-function significantly alters several central pathways involved in RNA biogenesis, cell cycle and stem self-renewal and different strategies have characterized a handful of validated MSI2-targets relevant in the AML context. For instance, transcripts encoding for components of the TGFβ pathway (i.e. SMAD3, TGFBR1) were shown to be directly regulated by MSI2 (Park et al., 2014). Other relevant AML transcripts that have been found to be regulated by MSI2 at the translational level include FLT3 (frequently mutated and overexpressed in AML patients) (Hattori et al., 2017), BCL2 (pivotal anti-apoptotic effector in AML LSCs) (Kwon et al., 2015), TSPAN3 (required for maintaining the self-renewal AML LSCs and subsequent disease initiation) (Kwon et al., 2015), CDKN1A, NUMB and MYC (Vu, Prieto, et al., 2017), among others.

Other recent studies have provided more insights on the MSI2 oncogenic role and its interactome in AML LSCs. By combining MSI2 conditional knockout in an MLL-AF9 mouse model and subsequent target profiling, MSI2 was demonstrated to be required for LSC maintenance through controlling the efficient translation of the oncogenic self-renewal program, including pro-LSC factors such as IKZF2, MYC and HOXA9 (Park et al., 2015). Similarly, another study uncovered a functionally dysregulated RBP interactome controlled by MSI2 together with SYNCRIP that post-transcriptionally regulate the myeloid leukaemia self-renewal program (Vu, Prieto, et al., 2017). Noteworthy, it was discovered that there was a significant increase in MSI2 RNA-binding activity in LSCs compared to normal HSCs and progenitor cells that resulted in a selective control of MSI2-driven translation of its oncogenic targets. This might explain the differential requirement of MSI2 in leukaemia (i.e. AML LSCs) compared with normal haematopoiesis (i.e. HSCs) (Nguyen et al., 2020) (Figure 6). Another AML LSC feature that has been linked to MSI2 function is chemoresistance. Recently, a study demonstrated that genetic silencing of MSI2 resulted in an increased sensitivity to daunorubicin which suggests that selective MSI2 inhibition in combination with chemotherapy agents could be an attractive therapeutic strategy in AML treatment (Han et al., 2015).

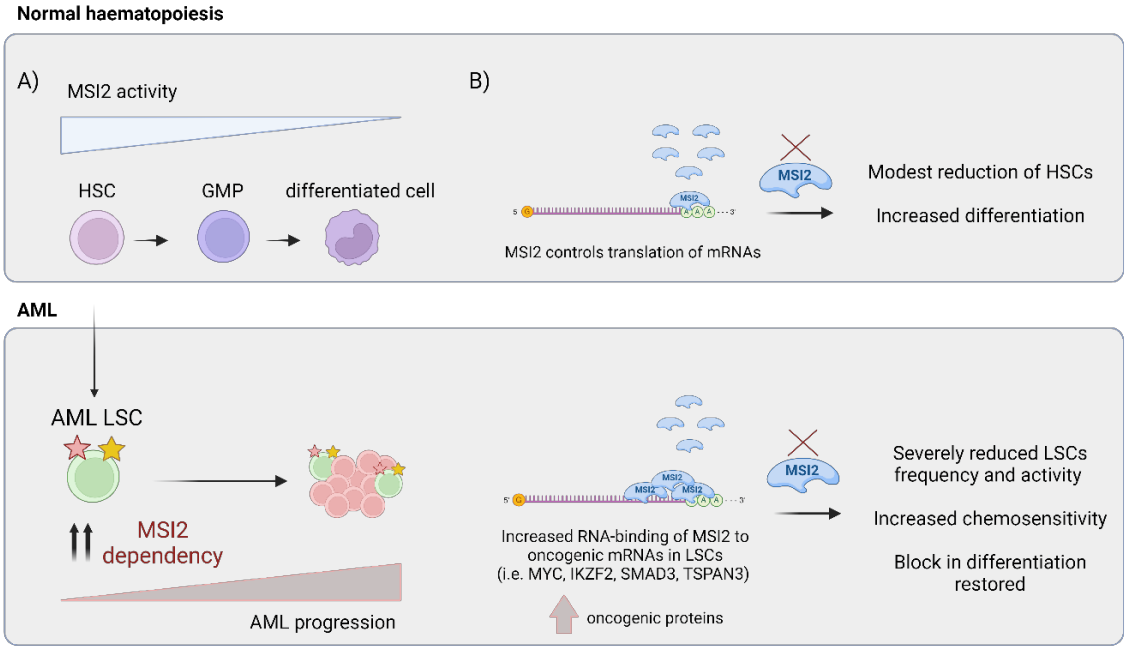


Figure 6. Differential requirement of MSI2 in AML LSCs vs normal HSCs. A) Depiction showing in the upper panel the progressively diminished MSI2 activity in stem (HSC) and progenitor cells (GMP) compared to myeloid differentiated cells. Lower panel shows the higher MSI2 activity in fully transformed acute myeloid leukemia stem cells (AML LSCs) compared to HSCs resulting in malignant transformation. B) Upper panel shows mechanism of action of how MSI2 binds to the mRNAs associated with self-renewal and maintenance of the stem cell compartment. The lower panel illustrates heightened RNA binding by MSI2, leading to decreased levels of target oncogenic proteins, thereby reinforcing a positive-feedback loop crucial for maintaining LSC function. Depleting MSI2 in AML LSCs notably impacts the protein levels of these targets compared to normal HSCs. IKFZF2: IKAROS Family Zinc Finger 2, SMAD3: Mothers against decapentaplegic homolog 3, TSPAN3: Tetraspanin 3. Figure adapted from (Nguyen et al., 2020; Park et al., 2015).

Notably, recent strategies have been directed towards the development and characterization of small molecule antagonists designed to target AML LSC-specific oncogenic RBPs, including MSI2. On this pursue, Ro 08-2750 (Ro) was demonstrated to bind selectively to the RRM1 of MSI2 and inhibit downstream MSI2 targets which resulted in a reduction of the disease burden *in vivo* (Minuesa et al., 2019).

The leukaemogenic role of MSI2 has also been investigated in other leukaemia types. Initially identified as an oncogenic fusion protein of HOXA9 in chronic myeloid leukaemia (CML) in accelerated phase (Barmpouti et al., 2003), MSI2 was shown to be essential for regulation of progression to blast crisis in CML patients (Kharas et al., 2010) and identified as an early indicator for poorer prognosis (Ito et al., 2010). Moreover, higher expression of MSI2 has been linked with MDS stem cells as well as with a more aggressive phenotype and poorer prognosis in high-risk MDS (Taggart et al., 2016). Overall, substantial evidence indicates that MSI2 is involved in the effective translation

of the oncogenic LSC self-renewal program, suggesting MSI2 as a potential therapeutic target for myeloid leukaemia stem cells.

1.4.- Long non-coding RNAs in malignant haematopoiesis

High-throughput sequencing technologies have demonstrated that the human genome undergoes pervasive transcription with nearly 90% of its DNA sequences transcribed into RNA but only a minor fraction of the genome is ultimately dedicated to actually encode proteins (Andrews et al., 1989; Eddy, 2001). Initially argued to be non-functional “junk” DNA, a huge family of non-coding transcripts referred to as non-coding RNA (ncRNA) is receiving attention in recent years, shifting the perspective that these RNA molecules act as a simple intermediary of protein synthesis towards a functional RNA molecule regulating gene expression and genome organization. Despite most of these ncRNAs were identified decades ago, we are recently witnessing an increasing number of biological and functional assignments. Overpowering proof has shown aberrant ncRNA expression profiles and subsequent affected downstream signalling processes which have been directly linked to cancer development and progression (Calvo Sánchez & Köhn, 2021; Nemeth et al., 2024; Slack & Chinnaiyan, 2019). This has prompted researchers to focus on this field of research in order to define their clinical relevance and exploit their potential use as future prognostic biomarkers or therapeutic targets (Adams et al., 2017; Deveson et al., 2017).

In general, ncRNAs can be broadly classified into two major classes based on an arbitrary transcript size cut-off of 200 nucleotides (nt) although many ncRNAs may fall into both sides of this threshold. Small ncRNAs (<200 nt) comprise different housekeeping ncRNAs such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) indispensable for protein synthesis, microRNAs (miRNAs) that target and silence complementary mRNA transcripts, small nucleolar RNA (snoRNAs) that perform post-transcriptional modifications or small nuclear RNAs (snRNAs) that represent the core part of the spliceosomal complex, among others. On the other side, long non-coding RNAs (lncRNAs; >200 nt) are gaining attention as they play an important regulatory role as modulators of gene expression at the epigenetic, transcriptional and post-transcriptional level during cell differentiation and development (Costa, 2005; Morris & Mattick, 2014). Thus, every lncRNA regulating cell fate (i.e. cell pluripotency, lineage commitment or apoptosis) during haematopoiesis, if deregulated, might contribute to leukaemia and other haematological malignancies. In fact, lncRNAs are rapidly gaining recognition in the AML field and the list of relevant lncRNAs, either abnormally expressed and used as prognostic markers or mutated genes, is growing exponentially even though the majority of them has not yet been functionally characterized.

1.4.1.- Biogenesis and mechanisms of lncRNAs

LncRNAs are functionally defined as transcripts over 200 nucleotides that appear to lack protein coding potential. Initially thought to be large by-products transcribed by RNA polymerase II (Pol II) with no biological relevance, with the advent of deep sequencing technologies, thousands of lncRNAs have been annotated and their role in tumorigenesis is gradually elucidated. Mirroring mRNAs in their biogenesis, structure and maturation, lncRNAs are mostly transcribed by Pol II and then may undergo mRNA-like processing steps such as canonical splicing, 5´-capping and 3´polyadenylation (Iyer et al., 2015; Statello et al., 2020). Additionally, protein-coding genes can yield transcript variants that lack coding capacity, thus expanding the extensive repertoire of annotated lncRNAs in human cells. The majority of lncRNAs are suggested to be expressed at lower levels and in a more tissue- and cell type-specific manner than most mRNAs (Derrien et al., 2012), they contain fewer exons and they are less evolutionary conserved across mammals (Hezroni et al., 2015). According to their genomic origins and transcription site by Pol II, lncRNAs can be classified as (1) sense lncRNAs transcribed on the same direction as coding exons; (2) antisense lncRNAs which are transcribed in the opposite strand of coding exons, (3) intergenic lncRNAs lying within two protein coding regions and (4) emerging from overlapping known non-coding regions (i.e. enhancer and promoter regions, introns of protein-coding genes or pseudogenes). Moreover, as mRNAs, their transcription is also tightly controlled by lineage-determining chromatin modifying complexes and transcription factors (Cabili et al., 2011; Washietl et al., 2014).

The realm of lncRNAs is characterized by its multifaceted nature concerning their mechanism of action which is directly tied to their subcellular localization. Unlike mRNAs that normally engage with the translational machinery in the cytoplasm, a larger proportion of lncRNAs are localized in various nuclear compartments (i.e. chromatin, nucleoplasm or subnuclear domains). LncRNAs can also be found in the cytoplasm, nuclear bodies, mitochondria or even released into exosomes (C. J. Guo et al., 2020; Tian & Manley, 2016). Based on the localization where these lncRNA carry out their regulatory roles, lncRNAs can be classified to act in *cis* (absolutely nuclear and accumulated at their sites of transcription where they affect the expression of nearby genes), nuclear acting in *trans* (localized to the nucleus but need to be relocated from their site of synthesis to affect gene regulation) and those which are cytoplasmic and function in *trans* affecting the expression of distal genes (X. Chen et al., 2015).

Nuclear lncRNAs are implicated in epigenetic regulations via either the recruitment of activator or repressor chromatin-modifying complexes to chromatin sites, transcriptional regulations through guiding or preventing the recruitment of transcription factors onto

target promoters or splicing regulations via sponging components of the spliceosome machinery. On the other side, cytoplasmic lncRNAs usually exert its function by blocking translation of mRNAs, affecting mRNA stability and modulating its degradation, serving as scaffolds with the ability to bind multiple effector molecules such as RBPs and promoting/inhibiting their function or acting as small regulatory RNA sponges for recruiting miRNAs (also called competitive endogenous RNAs or ceRNAs), among other mechanisms (Marchese et al., 2017; Salmena et al., 2011). Moreover, the adaptable and modular scaffold nature of lncRNAs allows them to bind several protein factors that might not interact or collaborate functionally solely through protein-protein interactions (Guttman & Rinn, 2012; Zappulla & Cech, 2004).

1.4.2.- Functional implications of lncRNAs in normal haematopoiesis

The relatively low expression levels and conservation across species of lncRNAs compared to mRNAs (Pang et al., 2006), along with their highly specific expression patterns across different cell types, lineages, and stages of development or disease, indicate that lncRNAs play integral roles particularly throughout the haematopoiesis process (Guttman & Rinn, 2012; Washietl et al., 2014) .

HSCs and multipotent progenitors require a tightly regulated gene-regulatory program to orchestrate cell fate transitions such as lineage commitment and differentiation. Given that all blood cell types originate from a single HSC, it is understandable that the majority of studies have concentrated on uncovering the functions of lncRNAs that regulate pathways governing HSC maintenance. Interestingly, expression of lncRNAs has been found to be elevated in HSCs and that is dynamically expressed during the differentiation process (M. Luo et al., 2015). These HSC-specific lncRNAs closely correlate with protein-coding genes with well-known functions in the regulation of haematopoiesis and cell fate decisions (Z. Wu et al., 2019). By deep RNA-sequencing, another study uncovered a subset of 159 unannotated lncRNAs uniquely expressed in HSCs but not in their differentiated progenies (M. Luo et al., 2015). This evidence was largely congruent with results from another study that revealed that hundreds of lncRNAs are co-expressed with lineage-specific transcription factors in HSCs compared to that in lineage-primed progenitors (Cabezas-Wallscheid et al., 2014). Consistently, another study performed systematic profiling of the non-coding transcriptome within the hematopoietic compartment defined fingerprint lncRNAs that regulate lineage HSC proliferation and differentiation (Schwarzer et al., 2017).

For instance, Spehd was shown to be required for haematopoiesis as its reduced levels led to decreased capacity of HSCs for multilineage differentiation (Delás et al., 2019).

Similarly, HOTTIP lncRNA was shown to regulate self-renewal and differentiation of HSCs. Notably, its aberrant expression led to an AML-like disease *in vivo* by reprogramming the haematopoietic chromatin and transcriptional program (H. Luo et al., 2019). Moreover, the lncRNA HOTAIR was reported to be upregulated in CD34+ AML LSCs compared to HSCs and required or proliferation and self-renewal of LSCs as well as for development and progression of AML (Gao et al., 2018). There is a growing list of lncRNAs involved in HSC self-renewal and quiescence; some examples are XIST (Yildirim et al., 2013), MALAT1 (X. Y. Ma et al., 2015), H19 (Venkatraman et al., 2013) and MEG3 (Sommerkamp et al., 2019), among others.

Besides their crucial function in HSCs maintenance, a plethora of studies have delved into the role of specific lncRNAs in controlling the differentiation of HSCs into myeloid precursors (Alvarez-Dominguez et al., 2014; Schwarzer et al., 2017). For example, the lncRNA HOTAIRM1 appears to be the most prominent upregulated lncRNA during granulocyte differentiation and myeloid cell maturation via stabilizing gene expression of the key transcriptional regulators HOXA1 and HOXA4 (Z. H. Chen et al., 2017; Langston & Gudas, 1992; X. Zhang et al., 2009). Other lncRNAs have been revealed in different myeloid maturation stages such as maturation of granulocytes, erythropoiesis and others (Qiu et al., 2021). With compelling evidence showcasing the diverse effects of lncRNAs involved in haematopoietic fate decisions, particularly within the HSC compartment, there is significant potential for the identification and utilization of lncRNAs as biomarkers in leukaemia.

1.4.3.- Prognostic relevance of lncRNAs in AML

Since lncRNAs are differentially expressed during normal and malignant haematopoiesis (Schwarzer et al., 2017), this class of ncRNAs is attracting much attention and mounting evidence show their key role in AML pathogenesis. Over the past years, genome-wide expression studies have highlighted the relevance of lncRNAs as independent prognostic markers in AML patients and correlated their distinct expression patterns with genetically defined AML subtypes and their clinical outcomes (De Clara et al., 2017; Mer et al., 2018; T. Sun et al., 2022; Zhu et al., 2023). For instance, a pioneering study established a prognostic score comprised of 48 lncRNAs in cytogenetically normal AML patients that associated distinctively with most recurrent AML mutations such as FLT3-ITD, NPM1, CEBPA or RUNX1 but not with DNMT3A and TET2 thus suggesting mutually exclusive oncogenic pathways activated by lncRNAs (Garzon et al., 2014). Using whole transcriptome profiling, Garzon et *al.* studied lncRNAs in more detail as independent prognostic markers and correlated their expression with distinct mRNA and miRNA signatures, providing meaningful information about which leukaemogenic pathways were

altered (Papaioannou et al., 2017). These scores were later incorporated in a clinical setting and an unfavourable lncRNA score correlated with poor treatment response and lower survival rates (Tsai et al., 2019). Since then, several prognostic lncRNA signatures focused in regulatory mechanisms that drive AML pathogenesis have been reported, with a focus in m6A regulatory genes (Zhong et al., 2022), autophagy (C. Zhao, Wang, et al., 2021) or mitochondrial cell-death (Zhu et al., 2023), among others.

Another group built a 11-gene mRNA expression scoring system (Chuang et al., 2015). Despite the variations in the statistical analysis and AML cohorts used in the study, only three of these genes (i.e. GPR56, KIAA0125, TM4SF1) were reported in another independent gene signature (Metzeler et al., 2008). Recent studies are focusing on the construction of lncRNA signatures that correlate with LSCs activity in order to identify and characterize AML LSC-specific lncRNAs with functional relevance and therapeutic potential. Originally identified as mRNAs and without confirmed non-coding potential, numerous ncRNAs reside within these signatures and await characterization. In pursuit of identifying prognostic biomarkers that can correlate with AML LSCs activity, Eppert et *al*. was the first group to identify mRNA expression profiles of protein-coding genes that correlate with functionally validated AML LSCs. A comparison of this gene signature with another derived from normal HSCs led to a “core enriched” (CE)-HSC-LSC signature comprising 44 stem cell-associated genes (including a subset of not yet annotated lncRNAs) highly expressed in LSCs (Eppert et al., 2011). High expression of this score was validated to have a negative prognostic impact in cytogenetically normal (CN)-AML patients (Metzeler et al., 2013). Subsequently, a study generated a CE-HSC-LSC-derived signature consisting of 111 lncRNAs that correlated strongly with LSCs activity (Bill et al., 2019).

1.4.4.- FAM30A is an LSC-specific prognostic marker in AML with no reported function

Noteworthy, FAM30A (also called KIAA0125) was the most consistently deregulated lncRNA in younger and older AML patients in the lncRNA-based signature derived from Bill M et *al.* FAM30A is also among other established prognostic adult and paediatric signatures commonly used in the clinics for the identification of AML LSCs (Eppert et al., 2011; Elsayed et al., 2020; B. J. Huang et al., 2022; N. Jung et al., 2015, S. W. K. Ng et al., 2016).

The biochemical and molecular features of FAM30A are poorly understood, as its non-coding nature has not been definitively confirmed through *in vitro* experiments. Its prognostic relevance in AML and other haematological malignancies has been primarily

assessed through computational prediction studies (Hung et al., 2021; Y. H. Wang et al., 2021). Even though FAM30A has been proposed to have context-specific oncogenic (Diniz et al., 2019; T. Zhang et al., 2023) or tumour-suppressor potential (Yang, Zhao, et al., 2019; Yao et al., 2017) in different tumour types, diseases or biological processes (D. S. de Lima et al., 2019), the underlying mechanism whereby this lncRNA exert its function still remains unexplored.

1.4.5.- Regulatory role of lncRNAs in AML pathogenesis

In AML, the growing characterization of relevant deregulated lncRNAs involved in leukemogenesis has revealed the variety of their exerted mechanisms (examples summarized in Figure 7). For example, HOTAIRM1 (HOXA Transcript Antisense RNA Myeloid-Specific 1) is one of the most well studied lncRNA and its myeloid-restricted function has been investigated in AML. This oncogenic lncRNA was demonstrated to enhance in *cis* the expression of their neighbouring HOXA cluster gene locus and myeloid differentiation markers like CD11B via recruitment of the UTX/MLL epigenetic complex to their promoters as well as acting in *trans* as a competing endogenous RNA as sequestering miRNAs (X. Chen et al., 2015; X. Q. D. Wang & Dostie, 2017; X. Zhang et al., 2009). Overexpression of the lncRNA HOXBLINC was identified in NPM1-mutated AML patients and critical for leukaemogenesis via recruitment of MLL1, chromatin domains and increasing promoter accessibility (Z. H. Chen et al., 2017). Another example of a *cis*-regulatory lncRNA is the unconventional RUNXOR which is transcribed from an upstream promoter of the RUNX1 gene. This lncRNA was found to be upregulated in AML patient samples. The proposed mechanism in KG-1 cells is that through its 3´end, this lncRNA interacts with RUNX1 promoter and enhancers and promotes interchromosomal contacts with distant gene loci of recurrent RUNX1 translocation partners (i.e. EVI1, ETO) as well as recruiting epigenetic modifiers such as EZH2 to specific promoters (H. Wang et al., 2014). Another lncRNA reported to have functions through RUNX1 is LOUP which recruits RUNX1 to the enhancer and promoter regions of the SPI gene to promote its transcription to ultimately induce myeloid differentiation. Interestingly, the AML-specific oncogenic protein fusion RUNX1-ETO or t(8;21) disrupts this interaction thus preventing LOUP-mediated molecular functions (Trinh et al., 2021).

Although less characterized in comparison with nuclear lncRNAs, cytoplasmic lncRNAs have also been studied in the AML context. Once exported to the cytosol, lncRNAs presumably undergo similar processing pathways compared to mRNAs and associate with diverse RBPs or miRNAs. Paradoxically, a large fraction of cytoplasmic RNAs is found in polysomal fractions and engaged by the translation machinery thus affecting

translation of mRNAs (Carlevaro-Fita et al., 2016). For example, PU.1-AS lncRNA associated to polysomes in the cytoplasm and decreases the translation of PU.1 mRNA, its antisense coding gene and vital transcription factor for normal haematopoiesis, via selective binding to the translation initiation factor eIF4A (Ebralidze et al., 2008). The lncRNA NEAT1 is downregulated in AML and has tumour suppressor activity. NEAT1 deficiency was shown to enhance AML LSC self-renewal and AML progression. When translocated to the cytoplasm, NEAT1 recruits and promotes the degradation of crucial components from the WNT signalling pathway and inactivates its downstream effectors ultimately affect AML LSC self-renewal capacity (Yan et al., 2021). Another example of an oncogenic cytoplasmic lncRNA is HOXA10-AS which is overexpressed in KMT2A-rearranged AML. Mechanistically, it was demonstrated that the stem-cell specific HOXA10-AS lncRNA acts *in trans* to induce NF-kB target genes and induce leukemogenesis by blocking monocytic differentiation (Al-Kershi et al., 2019). The cytoplasmic lncRNAs can also function as ceRNAs to recruit certain miRNAs and regulate target mRNA translation. For example, CRNDE elicits its oncogenic function via sponging miR-181 which in turn downregulates NOTCH2, part of an important signalling pathway involved in differentiation and proliferation of AML cells (X. Ma et al., 2020).

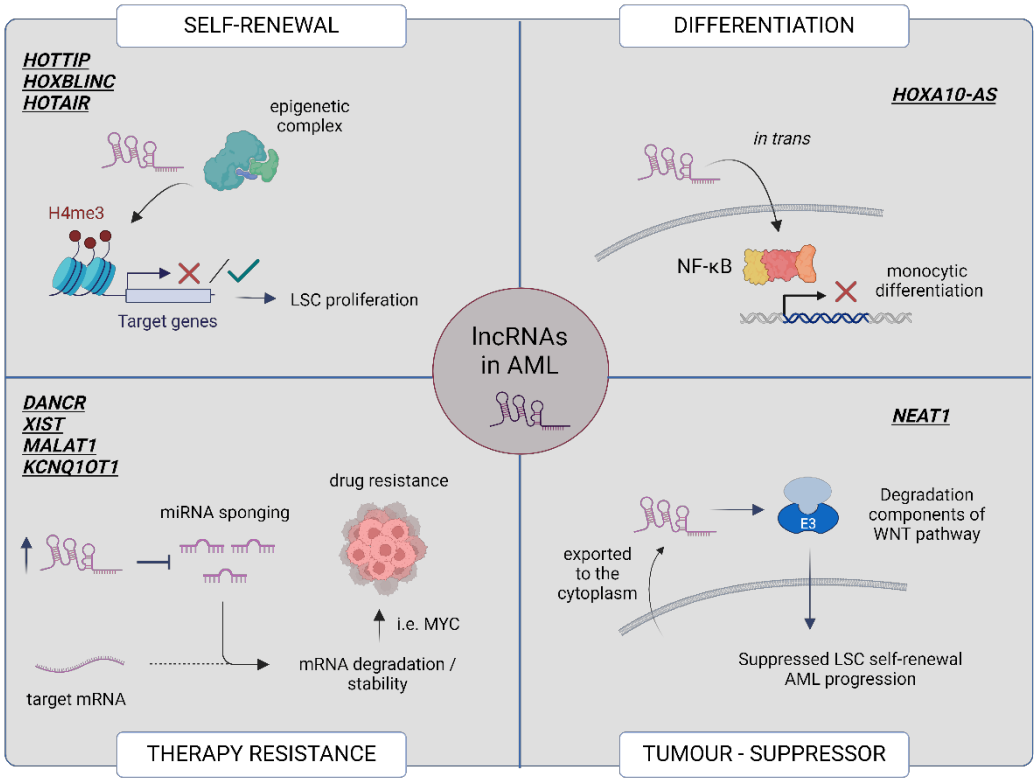


Figure 7. General overview of the mechanisms of lncRNAs involved in AML progression. Schematic displaying some examples of lncRNAs driving AML progression, especially in the LSC context.

-Continued on next page-

HOTTIP: HOXA transcript at the distal tip, HOXBLINC: HOXB locus-associated long non-coding RNA, HOTAIR: HOX antisense intergenic RNA, H4me3: tri-methylation at Histone 3, DANCR: Differentiation antagonizing non-protein coding RNA, XIST: X-inactive specific transcript, MALAT1: Metastasis-associated lung adenocarcinoma transcript 1, KCNQ1OT1: KCNQ1 Opposite strand/antisense transcript 1, NEAT1: Nuclear enriched abundant Transcript 1.

Recent research studies are also highlighting the relevance of lncRNAs in regulating chemotherapy resistance in AML. For example, the lncRNA KCNQ1OT1 was found to sponge the miRNA miR-193a-3p and induce expression of TSPN3, which is important of AML progression and chemoresistance (H. Sun et al., 2020). Another LSC-specific lncRNA, LINC00152, was shown to induce chemoresistance of AML cells via acting *in trans* on PARP1 gene expression (Cui et al., 2021). In the same line, knockdown of LINC00239 increased sensitivity to doxorubicin in AML cells via downregulating the PI3K/AKT/mTOR pathway and promote cellular apoptosis (Yang, Dai, et al., 2019) and the lncRNA SNHG5 regulates chemoresistance by targeting the miR-32/DNAJB9 pathway involved in autophagy (D. Wang et al., 2020). In addition, the well-known scaffolding lncRNA MALAT1 was found to be overexpressed in AML patients. Downregulation of MALAT1 led to increased sensitivity to Cytarabine which can be mechanistically explained via the decoy of miR-96, involved in AML proliferation (Hu et al., 2019). Notably, the lncRNA DANCR was recently identified as an important regulator of AML LSCs activity by regulating their self-renewal capacity and quiescence (Bill et al., 2019). Bill et *al.,* also unveiled a distinctive lncRNA signature specific to AML LSCs, some of which await functional characterization (i.e. FAM30A). These AML LSC-specific lncRNAs stand as promising targets for novel therapeutic approaches and significant ongoing progress is being made in this field.

1.4.6.- Therapeutic strategies to target lncRNAs in AML

The nascent but potential field of leveraging lncRNA for cancer therapy comprises several phases, from the identification of optimal oncogenic lncRNA targets with high therapeutic index to designing appropriate drugs for high-precision targeting. LncRNAs possess unique biochemical and molecular strengths that enhance their potential as therapeutic targets such as tissue- or cell-specific expression and their unique sequence specificity also renders them remarkable therapeutic targets by minimizing off-target effects. On the contrary, they also exhibit notable weaknesses including repetitive sequences, inadequate annotation and a high rate of sequence evolution (Winkle et al., 2021). There are four principal therapeutic strategies that aim to suppress the activity (loss of function) and/or expression of oncogenic lncRNAs: RNA interference (RNAi) approaches, antisense oligonucleotides (ASOs), CRISPR-Cas systems and small molecule approaches.

RNAi has been until recently the standard loss-of-function study to discover important cancer lncRNA drivers (Beermann et al., 2018; Gutschner et al., 2011; Delás et al., 2017; Tiessen et al., 2019). In essence, RNAi strategies generally utilize chemically engineered transcripts of 20-40 nt in length that, upon complementary binding to the target RNA, triggers the degradation of the formed RNA duplex via canonical cellular machinery (Chang et al., 2012). Subsequently, this can lead to transient (small interfering RNA, siRNA) or sustained (short hairpin RNA, shRNA) downregulation of the target RNAs. For instance, knockdown of the LSC-specific lncRNA DANCR resulted in decreased self-renewal capacity and quiescence and DANCR-targeting *in vivo* increased the survival of mice (Bill et al., 2019). Although RNAi approaches are relatively fast and easy to use, its effectiveness in targeting nuclear lncRNAs has long been debated as the RNAi machinery is predominantly located in the cytoplasm (Lennox & Behlke, 2016; Maamar et al., 2013). To overcome the constraints of RNAi, researchers have increasingly turned to the genome-editing tool known as clustered regularly interspaced short palindromic repeats-CRISPR-associated endonuclease (CRISPR-Cas9) in recent years. This shift has allowed the functional characterization of several lncRNAs in AML (Bester et al., 2018; Fernando et al., 2017; M. Ng et al., 2019). This system relies on the site-specific cleavage by endonuclease Cas9 at a target DNA sequence guided by previously designed single-guide RNA (sgRNA) in order to achieve long-term gene editing (Miller et al., 2007; Wiedenheft et al., 2012). Different strategies can be employed for targeting lncRNAs: removing the entire locus of the lncRNA gene (Bassett et al., 2014), deleting the major promoter region of the lncRNA gene (Aparicio-Prat et al., 2015), introducing transcription termination within the lncRNA gene signals guided by epigenetic machinery (Janga et al., 2018) or stronger suppression of lncRNA expression using CRISPR interference system (modified Cas9 fused with transcriptional repressors) (Perez-Pinera et al., 2013).

In particular, research studies have been focusing on ASOs, which have arguably become the forefront contender in lncRNA therapeutics due to their favourable pharmacological properties (i.e. tolerability, activity, stability) and customizable design as it allows chemical modifications of the structure. ASOs represent a group of chemically synthesized short fragments of single-stranded oligonucleotides that recognize their target RNAs via Watson-Crick base-pairing (Crooke et al., 2021; Rinaldi & Wood, 2018). Depending on their design and mode of targeting, there are two ASOs modalities. ASOs can be “Gapmers” that bind to their target lncRNAs through sequence complementarity and trigger their degradation (simple loss of function or knockdown) via RNaseH in the nucleus (Monia et al., 1993). Alternatively, ASOs can also be steric blocker (SB-ASOs)

that bind their target RNAs, without degrading it, and disrupt lncRNA interactions with other molecules, hence their downstream effect (Havens & Hastings, 2016). Comprehensive studies have demonstrated that, in comparison to RNAi, ASOs exhibit significantly greater efficacy in depleting both nuclear lncRNAs and, to a considerable extent, cytoplasmic lncRNAs (including nascent RNA transcripts) (Lennox & Behlke, 2016). ASOs-mediated downregulation of the stem cell-specific lncRNA HOXA10-AS led to decreased leukemic growth of KMT2A-rearranged AML blasts *in vivo* (Al-Kershi et al., 2019). Moreover, several examples of using ASOs to succesfully target lncRNAs in the AML field have been reported *in vitro* and *in vivo* (Bester et al., 2018; De Clara et al., 2017; Gourvest et al., 2021; Papaioannou et al., 2017). Notably, the liposome-incorporated ASOs called Prexigebersen (BP1001) has been demonstrated to exert potent antitumour activity in the treatment of patients with newly diagnosed and relapsed/refractory AML in combination with decitabine and venetoclax in a phase 2 clinical trial (ClinicalTrials.gov, NCT02781883).

An alternative method for drugging lncRNA is using small molecules. Akin to SB-ASOs, their mechanism typically involves inducing lncRNA loss of function by disrupting its native structure or hindering recognition of the binding partner(s) (Costales et al., 2020). They offer significant advantages including optimal cell permeability and pharmacokinetic profiles, yet their development is hindered by uncertainties regarding toxicities and off-target effects. Consequently, most small molecule inhibitors concentrate on disrupting lncRNA-protein interactions, an area where the understanding is slightly more advanced (Disney et al., 2016; R. Zhao et al., 2022). For example, compounds that target XIST, a well-known oncogenic lncRNA in AML, reduce its conformational space and displace binding of the epigenetic modifiers PRC2 and SPEN (Jiang et al., 2024; C. Wang et al., 2020).

There is unequivocal evidence regarding the contribution of lncRNAs in the development, progression, and drug resistance in AML. Understanding the functional role of lncRNAs is still at an early stage, given their complex involvement in gene regulation through dynamic interactions with diverse molecules or protein complexes. In elucidating the functional and clinical relevance of lncRNAs that lack prior characterization, perturbation experiments—particularly loss-of-function studies—play a crucial role. Among these lncRNAs, some have been identified as significant prognostic markers for high-risk AML and are part of established AML LSC signatures; nevertheless, yet they remain to be characterized. Unfortunately, the development of lncRNA-targeted therapeutics is constrained by this limited understanding, resulting in a lack of therapeutic examples targeting lncRNAs, for instance, in AML. However, as nucleic acid-targeting therapeutic

strategies progress rapidly with enhanced toxicity profiles and pharmacokinetics, the identification and evaluation of targeted lncRNAs in AML could potentially be translated into clinical applications to improve survival rates in AML patients.

1.5.- Aims of the study

Long non-coding RNAs (lncRNAs) are increasingly emerging as important prognostic biomarkers and potential therapeutic targets in cancer. It has been known for many years that high FAM30A expression significantly correlates with signatures of leukaemic stem cell (LSC) activity and poor prognosis in both adult and paediatric acute myeloid leukaemia (AML). However, the precise molecular role of FAM30A in AML LSCs is poorly understood. Thus, it is hypothesized that FAM30A could directly regulate AML LSC activity so the primary objectives of this thesis are as follows:

I. **Investigating the biological and molecular attributes of FAM30A and identification of potential protein binding partners.**

II. **Evaluation of how modulation of FAM30A expression phenotypically affects AML LSCs activity.**

III. **To propose a mechanistic model of a putative oncogenic axis involving FAM30A, its protein interactors and subsequent downstream effects, as these could provide a rationale for effective anti-LSCs targeting.**

IV. **The assessment of the leukaemogenic potential of FAM30A *in vivo.***

To achieve these goals, KG-1a, a monocytic AML cell line with known LSC activity, was primarily used as a mechanistic model to study the molecular function of FAM30A. Screening and further analysis of protein binding partners of FAM30A were determined using RNA-pulldowns alongside mass spectrometry analysis and immunoprecipitation (IP) assays. After establishing fluorescently-tagged stable cell lines for FAM30A gain- or loss-of-function, RNA-sequencing analysis, cell viability and apoptosis assays, cytotoxic treatment with chemotherapeutic agents, colony forming cell (CFC) assays, and lineage commitment were assessed for each cell line *in vitro*. To unravel the regulatory network between FAM30A and its protein binders as well as the effects on downstream targets, protein translation inhibition assays, public CLIP-sequencing data, IP and luciferase-reporter assays were used. Ultimately, in order to assess the oncogenic potential of FAM30A in promoting leukaemic engraftment, cell lines were transplanted into immunocompromised mice and engraftment in the bone marrow as well as analysis of the immunophenotype (via FACS) of the engrafting cell populations was subsequently analysed after 6 weeks.

2. MATERIALS AND METHODS

2.1.- MATERIALS

2.1.1.- Cell lines

Cell lines were either purchased from DSMZ or kindly acquired from:

- Group of Prof. Dr. Hüttelmaier group, Institute of Molecular Medicine, University of Halle.
- Group of Prof. med. Lutz P. Müller group, Innere Medizine Medizin IV (Hematology and Oncology), Universitätsklinikum Halle.

Table 1. Cell lines used for cell culture

Name	Origin / Cell type	Distributor	Reference
MV4-11	Acute monocytic leukemia	UKH Halle, DSMZ	(Drexler et al., 2004)
MOLM-13	Acute myeloid leukemia	DSMZ	(Matsuo et al., 1997)
ME-1	Acute myeloid leukemia	DSMZ	(Yanagisawa et al., 1991)
GDM-1	Acute myelomonocytic leukemia	DSMZ	(Ben-Bassat et al., 1982)
KG-1a	Acute myeloid leukemia	DSMZ	(Koeffler et al., 1980)
K562	Chronic myeloid leukemia	Hüttelmaier, DSMZ	(Lozzio & Lozzio, 1975)
HEK293T	Human embryonic kidney	Hüttelmaier, DSMZ	(DuBridge et al., 1987)

2.1.2.- Bacteria

The bacterial strain employed for transformation of cloned plasmids in this study was *Escherichia coli* TOP10 (genotype: F-*mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX*74 *recA*1 *deo*R *ara*D139 Δ(*ara leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G). Bacteria were cultivated in LB (lysogeny broth) liquid medium (Luria-Bertani recipe – 1 % (w/v) Tryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl), supplemented with antibiotics (30 µg/mL Kanamycin or 150 µg/mL Ampicillin) for bacterial selection. For the initial transformation and selection of recombinant clones, bacteria were plated on LB medium supplemented with agar (1.5% w/v) and 100 mg/mL Ampicillin or 50 mg/mL Kanamycin for colony growth, and subsequently analyzed by PCR for positive transformants.

2.1.3.- Animals

All mice were housed at the Beatson Research Unit (University of Glasgow, UK) under the supervision of the staff of Prof. David Vetrie (Wolfson Wohl Cancer Research Centre, University of Glasgow). The mouse strain used for this study was NRG-3GS [(NOD.Cg-Rag1tm1Mom Il2rgtm1Wjl Tg(CMV-IL3,CSF2,KITLG)1Eav/J].

2.1.4.- Chemicals, reagents and cell culture

The chemicals used in this thesis were purchased from Sigma-Aldrich, Roth, PeqLab, Thermo Fisher Scientific unless otherwise stated. Cell culture dishes and flasks were acquired from TPP (Techno Plastic Products) and serological pipettes and methylcellulose cell culture dishes from Corning. The cell culture growth medium (RPMI, DMEM) and associated supplements for cell culture such as GlutaMAX, Opti-MEM, Trypsin and PBS were obtained from Thermo Fisher Scientific. The FBS was purchased from Thermo Fisher Scientific and PAN-Biotech. Methylcellulose-based medium including cytokines was acquired from Stem Cell Technologies. Centrifuge tubes used for sucrose-based gradient fractionations were purchased from Beckman Coulter.

Dynabeads conjugated with streptavidin or protein A for protein, RNA pulldowns and immunoprecipitation studies were also purchased from Thermo Fisher Scientific. Enzymes and reaction buffers required for PCR reactions, cloning and *in vitro* transcription were purchased to New England Biolabs and Promega. The Master Mixes for qRT-PCR were manufactured by High Qu GmbH and Promega and the protein and nucleotide ladder used for agarose and SDS-polyacrylamide gels respectively were purchased from Thermo Fisher Scientific.

2.1.4.1.- Buffers and reagents

Table 2. Recipes for buffers and reagents.

Name	Receipt
PBS (phosphate buffered saline)	140 mM NaCl 2.6 mM KCl 8 mM Na_2HPO_4 1.5 mM KH_2PO_4
PBS-T (PBS – Tween)	1 x PBS 1% Tween-20
5x TBE (Tris/Borate/EDTA)	450 mM Tris 450 mM Boric acid 10 mM EDTA
TAE (Tris/Acetate/EDTA)	40 mM Tris 20 mM Glacial acetic acid 1 mM EDTA
NuPAGE (blotting buffer)	50 mM Tris 40 mM Glycerin

	0.04 % SDS 10 % Methanol
Ponceau-S	0.1 % Ponceau-S 5 % acetic acid
Total Lysis Buffer	50 mM Tris (pH 7,4) 50 mM NaCl 1 % SDS 2 mM $MgCl_2$
Gradient Buffer	10 mM HEPES-KOH (pH 7,2) 5 mM $MgCl_2$ 150 mM KCl
Gradient Lysis Buffer	1x Gradient Buffer 0,5% NP-40
Immunoprecipitation Buffer	1x Gradient Buffer 1% Triton X-100
Lysis Buffer for LC-MS/MS	1x Gradient Buffer 1% DDM
TRIzol	0.8 M Guanidine thiocyanate 0.4 M Ammonium thiocyanate 0.1 M Sodium acetate (pH 4,5) 5 % Glycerin 48 % Roti®Aqua Phenol
Cell fractionation Buffer	10 mM HEPES-KOH pH 7,2 5 mM $MgCl_2$ 150 mM KCl 100-200 µg/µL Digitonin
BW Buffer (Binding-Wash Buffer)	5 mm Tris-HCl pH 7,5 0.5 mm EDTA 1 m NaCl
Trypsin-EDTA	0.05 % Trypsin 0.4 mM EDTA sterile PBS

2.1.5.- Antibodies

Table 3. Primary antibodies used for Western Blot and immunofluorescence studies.

Primary Antibody	Species	Company	Catalog number
anti-MSI2	mouse	Biorbyt	orb69392
anti-MSI2	rabbit	Bethyl	A303-215A
anti-HNRNPA1	rabbit	Cell Signaling /NEB	#8443
anti-ACTB	mouse	Sigma-Aldrich	A5316
anti-LaminB1	rabbit	Cell Signaling /NEB	#13435 (D9V6H)
anti-HIF1α	rabbit	Cell Signaling /NEB	#14179 (D2U3T)
anti-PTBP1	mouse	Serum	(Chou et al., 2000)
anti-EEF2	mouse	Santa Cruz	sc-166409
anti-VCL	mouse	Sigma-Aldrich	V9131
anti-RUNX1	mouse	Santa Cruz	sc-365644 (A-2)
anti-NUMB	rabbit	Cell Signaling /NEB	#2756 (C29G11)

anti-SMAD3	rabbit	Cell Signaling /NEB	#9523 (C67H9)
anti-MYC (c-MYC)	rabbit	Cell Signaling /NEB	#5605 (D84C12)
anti-RPL7	rabbit	Bethyl	A300-741A
Anti-STAT5	mouse	Thermo-Fisher	ST5-8F7
Anti- pSTAT5 (Tyr694)	rabbit	Cell Signaling/NEB	#4322 (D47E7)

Table 4. Fluorescently-conjugated secondary antibodies used in western blot and immunofluorescence studies.

Antibody	Species reactivity	Company	Catalog number
IRDye680RD donkey	mouse	LI-COR	926-68072
IRDye680RD donkey	rabbit	LI-COR	926-68073
IRDye800CW donkey	mouse	LI-COR	926-32212
IRDye800CW donkey	rabbit	LI-COR	926-32213
Alexa Fluor® 488 AffiniPure F(ab')$_2$ Fragment Donkey IgG (H+L)*	rabbit	DiaNova	711-546-152

*Antibody used for immunofluorescence studies.

Table 5. Antibodies used in flow cytometry studies (FACS).

Antibody	Fluorophore	Company	Catalog / Order no.
anti-CD34 human	VioBlue	Miltenyi Biotec	130-113-744
anti-CD34 human	BV-510	BD Biosciences	745070
anti-CD25 human	VioBright-B515	Miltenyi Biotec	130-113-287
anti-CD38 human	FITC	Miltenyi Biotec	130-119
anti-CD11B human	VioBlue	Miltenyi Biotec	130-110-558
anti-HLA-DR human	APC-Cy7	BD Biosciences	335796
anti-CD33 human	FITC	Miltenyi Biotec	130-111-135
anti-CD45 human	FITC	BD Biosciences	561865
Annexin-V	FITC	BD Biosciences	560931

2.1.6.- Standard Kits and Assays

Table 6.- Commercial kits and systems.

Name	Company	Catalog / Order no.
OneTaq® 2X Master Mix with Standard Buffer	New England Biolabs	M0482
Q5® High-Fidelity 2X Master Mix	New England Biolabs	M0492

Zero Blunt™ TOPO™ PCR Cloning Kit	New England Biolabs	451245
Agarose Gel Extraction Kit - Column Kit	Jena Bioscience	PP-202L
Plasmid Mini-Prep Kit - Column Kit	Jena Bioscience	PP-204L
PureLink™ HiPure Plasmid-Midiprep-Kit	Thermo Fisher Scientific	K210004
PCR Cloning Kit	New England Biolabs	E1202
PureLink™ Genomic DNA-Minikit	Thermo Fisher Scientific	K182002
CellTiter-Glo® Assay System	Promega	G7571
Dual-GloTM Luciferase Assay System	Promega	E1960
DC Protein Assay	Bio-Rad	5000111

2.1.7.- Oligonucleotides and probes

All oligonucleotides were purchased from Eurofins Genomics GmbH.

Table 7. Sequences of oligonucleotides used for molecular cloning.

Name	Sequence (5´--> 3´)	Restriction site
FAM (FAM30A repeats) _s	GGTTTAAACTTGACCTCTTTCAG CCCTTGGTTACAGGA	PmeI
FAM (FAM30A repeats) _as	GGCTCGAGAACTGATCAAATTTG GTGTTACAAG	XhoI
RUNX1-BS (binding sites) _s	AATTCTGTGGTTAACCACATGTG GTTAACCACAA	EcoRI
RUNX1-BS (binding sites) _as	AGCTTTGTGGTTAACCACATGTG GTTAACCACAG	HindIII
RUNX1-P1 promoter _s	GGAGATCTGGTTTCATTAAAACA GTCCTTTTATTTCAA	BglII
RUNX1-P1 promoter _as	GGCTCGAGCGCTTCCTCCTGAA AATGCACCCTCTTCTG	XhoI
RUNX1-P2 promoter _s	GGAGATCTGCCGCTGACGCACG CAGCAAGTGAGACGCG	BglII
RUNX1-P2 promoter _as	GGCTCGAGCATCCAGCCAAGAA GCCAGTTAATTCAAAA	XhoI

Table 8. Sequences of oligonucleotides used for RT-qPCR.

Name	Sense (5´--> 3´)	Antisense (3´--> 5´)
FAM30A endg (endogenous)	CCCACAATATAGACAGGCATTGTCAC T	CCAGGCAGATTTCCCAAATCCGGC T
FAM30A ex (exogenous)	CCATCGTGGAACAGTACGAGCGCGC CG	CCAGGCAGATTTCCCAAATCCGGC T
FAM30A Exons 1-2	GACATCGGGAACGGACGG	GACCCAGGAGGTCTGGCATA
FAM30A Exons 1-3	CCTTTTCAGCAGACGCCCT	AGCATCTGGCGAGGACCTTC
FAM30A Exons 4-5i	TCTCACAATGCGTGTTGAGCTT	TCCATAGCAGCCTGCCTACA

FAM30A Exon 5i	GAATCTATGCCTGTCACGGGG	TCCATAGCAGCCTGCCTACA
FAM30A Exon 5e	AACGTGCCCTTGTAACACCA	AAGGTGTCCATTCTGACAGCC
GPR56	CTGCTTCTGCAACCACTTGAC	GAGCAGAGGTAGGCGGCAAT
DNMT3B	AGCCCATGCAACGATCTCTC	CCAGGAACCGTGAGATGTCC
CD52	CAGCCTCCTGGTTATGGTACAG	ACAGGGCTAAGGCTGAGACG
NGFRAP1	GAGAGGAAGACCGCCCTTTG	TCCTCAACTGCAGCTCCCTA
SMIM24	CCCTTCTGGTGCTGGAGTTTC	TTCTGAACGTGGTCTCCTCCT
MMRN1	CCCCACGGGAAACATACCTC	TCTGAGGACAGGATCCACCG
NYNRIN	ACACTCGCTCAAGATGCTTTCA	AAGAGATTCCCCGCAAAGGTG
CPXM1	CTTTGTGCTAAGTGCCAACCTC	CAGCAATCCCAAGCTCCGT
AKR1C3	AGAAGTAAAGCTTTGGAGGTCACA	GGACCAACTCTGGTCGATGAAAA
ZBTB46	GCGGCGCTCATGAGTAAGAA	ATGGCCGAGAAGCTGCAGTA
EMP1	TGGAGACCAGCAAGTATTGTCC	TTAGAGCCATTGGCAGCATCA
ARHGAP22	AAGCCTAGTGATGGGGGAGC	GTCACCTGTGTCCCTTGTAGAG
NEAT1	TGATGCCATCTCACAGGCAG	TGCTGCGTATGCAAGTCTGA
U6	GTGCTCGCTT CGGCAGCACA TATAC	AAAATATGGA ACGCTTCACG AATTTG
ACTB	TCCCTGGAGAAGAGCTACG	GTAGTTTCGTGGATGCCACA
EEF2	GGAGTCGGGAGAGCATATCA	GGGTCAGATTTCTTGATGGG
PPIA	GTCTCCTTTGAGCTGTTTGCAG	ATGGACAAGATGCCAGGACCC
BCL2	AAAAATACAACATCACAGAGGAAGT	CCTTGGCATGAGATGCAGGA
TGFBR1	GCAGAGCTGTGAAGCCTTGAGA	TGCCTTCCTGTTGACTGAGTTG
TSPAN3	CCTTCCGACCTCTATGCTGAGG	GAATAGCTGCAAATGCCAGTGCG
VRK1	AGACCCCAAAAGATGTCACG	CCAAGGAAGATGGCCAGTAA
MALAT1	GAGAAGTGTCAGCCTCACCTG	TGCTTTTAACAGGCTTCTGGA
RUNX1A	CTCCCTGAACCACTCCACTG	TTAACATCTCCAGGGTGCTGTGTC TTC
RUNX1B/C	CTCCCTGAACCACTCCACTG	CAGAGAGGGTTGTCATGCCG
RUNX1C	GTGCATTTTCAGGAGGAAGCG	CGTGGACGTCTCTAGAAGGATTC
RUNX1T	TCTGCAGAACTTTCCAGTCG	AAGGCGCCTGGATAGTGCAT
SMAD3	GCCTGTGCTGGAACATCATC	TTGCCCTCATGTGTGCTCTT
NUMB	CGTTCGCACCGGAAAATGTA	TTCCAGTTTTTCCAAAGAAGCCT
MSI2	TGTTGTGGAGAAAGTCTGTGAGATT	CATGGTGTAAGGCAGTCCCC

Table 9. Sequences of oligonucleotides used for cloning of shRNAs.

Name	Sequence (5´--> 3´)	Restriction site

shFAM30A _s	GATCCGCACAGAGGGTTTGGCTTTTACTAGTAAA GCCAAACCCTCTGTGCTTTTTG	BamHI
shFAM30A _as	AATTCAAAAAGCACAGAGGGTTTGGCTTTACTAG TAAAAGCCAAACCCTCTGTGCG	EcoRI
shMSI2 _s	GATCCGCACAGAGGGTTTGGCTTTTACTAGTAAA GCCAAACCCTCTGTGCTTTTTG	BamHI
shMSI2_as	AATTCAAAAAGCACAGAGGGTTTGGCTTTACTAG TAAAAGCCAAACCCTCTGTGCG	EcoRI
shHNRNPA1 _s	GATCCACAACAATCAGTCTTCAAATTTACTAGTAA TTTGAAGACTGATTGTTGTTTTTTG	BamHI
shHNRNPA1 _as	AATTCAAAAAACAACAATCAGTCTTCAAATTACTA GTAAATTTGAAGACTGATTGTTGTG	EcoRI

Table 10. Sequences of oligonucleotides used for in vitro transcription.

Name	Sequence (5´--> 3´)
CTRL (Control RNA) _s	GTAATACGACTCACTATAGGGCTTCTGCCTCCGGGTA AAATAAACTGCAAAAAGGATTCCCAGGAAAC
CTRL (Control RNA) _as	CTGTAACCAAGGGCTGAAAGAGGGAACGGTTTCCTG GGAATCCTTTTTGCA
FAM (FAM30A repeats) _s	GTAATACGACTCACTATAGGATGACATTCACATGAACG GGCACATACAGGAAAACACGGTAATGTAATTAGAATA
FAM (FAM30A repeats) _as	TTCTGGCTACTTTTCTCTGACTATTCTAATTACATTACC GTGTTTTCCTGT
MMut (FAM30A repeats with mutated MSI2 binding sites) _s	GTAATACGACTCACTATAGGATGACATTCACATGAACG GGCACATACAGGAAAACACGGTAATGTAATATCAAAT
MMut (FAM30A repeats with mutated MSI2 binding sites) _s	TTCTGGGATCTTTTCTCTGAGATTTGATATTACATTACC GTGTTTTCCTGT

2.1.8.- Plasmids

Table 11. Commercial vectors used for cloning.

Name	Antibiotic resistance	Company (cat. no.) & Reference
pCR™-Blunt II - TOPO™	Kanamycin	Thermo Fisher Scientific (450245)
pmirGLO	Ampicillin	(B. Busch et al., 2016)
pLVX_Crimson-shRNA3-Control (pLVX shC)	Ampicillin	Clontech (632164)
pLVX_RFP (pLVX EV)	Ampicillin	(B. Busch et al., 2016)
pmirGLO Promo+Stopp*HindIII	Ampicillin	Prof. Dr. Stefan Hüttelmaier,

Table 12. Plasmids used for eukaryotic expression.

Final plasmid	Vector	Restriction sites
pLVX_FAM	pLVX_RFP	PmeI, XhoI
pLVX_shFAM30A	pLVX_Crimson-shRNA3-Control	BamHI, EcoRI
pLVX_shMSI2 (shM1)	pLVX_Crimson-shRNA3-Control	BamHI, EcoRI
pLVX_shHNRNPA1 (shA1)	pLVX_Crimson-shRNA3-Control	BamHI, EcoRI
X2 RUNX1	pmirGLO Promo+Stopp*HindIII	EcoRI, HindIII
X3 RUNX1	pmirGLO Promo+Stopp*HindIII	EcoRI, HindIII
RUNX1 P1 promoter	pmirGLO	PmeI, XhoI
RUNX1 P2 promoter	pmirGLO	PmeI, XhoI

2.1.9.- Chemical compounds

Table 13. Chemotherapeutics, inhibitors and detergents

Name	Company (cat. no.) & Reference
CAS 285986-31-4 (STAT5 inhibitor)	573108-10MG. Sigma-Aldrich
BC-96 (anti-human CD25 Antibody)	302602. BioLegend
Cytosin β-D-Arabinofuranosid (Ara-Cytarabine)	147-94-4. Sigma-Aldrich
Daunorubicin -hydrochlorid (Daunorubicin)	23541-50-6. Sigma-Aldrich
NP-40 (IGEPAL®)	9002-93-1. Sigma-Aldrich
Triton-X100	3051.3. Roth.
Digitonin	D1414-100MG. Sigma-Aldrich
Emetine	E2375. Sigma-Aldrich
Cycloheximide	C4859-1ML. Sigma-Aldrich
Polybrene	SC-134220. Santa Cruz Biotechnology
Puromycin	#P9620. Sigma-Aldrich

Protease/Phosphatase Inhibitor cocktail (100X)	#5872S. Cell Signaling

2.1.10- Instruments

Table 14. List of used devices.

Application / Experiment	Name of the device (company)
Imaging	Nikon TE-2000E Nikon TS-100 Leica SP5X Leica SP8X
Infrared-Scanner	LiCOR Odyssey Infrared Scanner
SDS-PAGE	NuPAGE® Bis-Tris Electrophoresis System (Life Technologies)
Western Blot	XCell IITM Mini-Cell Blot Module (Thermo Fisher)
Magnetic Rack	DynaMag™-2 Magnet (Invitrogen)
Real-Time PCR	Light Cycler 480 II 384-well plates (Roche) iCycler iQ 96-well plates (Bio-Rad)
Spectroscopy	Tecan Spark® multimode microplate reader GloMax® 96 well Microplate Luminometer
LC-MS / MS	Orbitrap Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific)
Thermocycler	Mastercycler Nexus II 96well (Eppendorf)
Centrifuges	Heraeus Biofuge Stratos Eppendorf Centrifuge 5424 R Eppendorf MiniSpin
Ultracentrifugation	Optima™ L-90K Ultracentrifuge (Beckman Coulter)
Agarose gel electrophoresis	Peqlab PerfectBlue Mini L
Cell counting	T20 Automated Cell Counter (Bio-Rad) Neubauer (0.1 mm / 0.0025 mm^2)
Flow cytometry	MACSQuant Analyzer (Miltenyi Biotec)
FACS sorting	BD FACSMelody™ Cell Sorter (BD Biosciences)

2.2.- METHODS

2.2.1.- Experiments involving cell culture

2.2.1.1.-Cell cultivation

The leukaemia cells used in this thesis, both AML cell lines (MV4-11, MOLM-13, ME-1, GDM-1 and KG-1a) and CML cell lines (K562), were cultured in Roswell Park Memorial Institute (RPMI 1640, Thermo Fisher Scientific) supplemented with 20% fetal bovine serum (FBS), 1% GlutaMAX (L-Alanyl-L-glutamin) and 1% (v/v) penicillin-streptomycin. For subsequent experiments, at every passaging fast-growing cells (MV4-11, MOLM-13 ME-1, K562) were usually seeded at lower cell density (0.1-0.5 x 10^6 cells/mL) and slow-growing and more stem-like cells (GDM-1 and KG-1a) were maintained at higher cell density (0.5-1 x 10^6 cells/mL).

For the adherently growing HEK293T cell line, the cultivation conditions were in Dulbecco's modified Eagle's medium (DMEM) supplemented in 10% FBS and 1% GlutaMAX. Prior to passaging, the medium was discarded and cells were washed with 1x PBS. Following this, cells were detached via trypsinization (0.05% Trypsin and 0.4mM EDTA) and the process was halted by the addition of fresh DMEM supplemented with 10% FBS. Afterwards, cells were counted and seeded into cell culture dishes. All cell lines were kept in the incubator at 37ºC and 5% CO_2.

2.2.1.2.-Lipofection of DNA

Transfection of plasmid DNA into cells have been performed using the transfection reagents Lipofectamine™ 3000 (Thermo Fisher Scientific), DharmaFECT kb (Horizon Discovery Ltd.) or Viromer Red (formerly known as Lipocalyx) and following manufacturer´s instructions. DharmaFECT kb was utilized as a substitute for Viromer Red due to the discontinuation of the latter's production. Regardless the transfection method employed, a day prior to transfection HEK293T cells were seeded at a density of 3.5 x10^5 cells/mL in a 6-well plate in a total volume of 2 mL of growing medium. For Lipofectamine™ 3000, 2 µg of plasmid DNA and 4 µl of P3000 reagent were mixed in 100 µl of serum-free medium (Opti-MEM, Thermo Fisher Scientific) in a 1.5 ml reaction tube. In a separate reaction tube, 4 µl of Lipofectamine 3000 were combined with 100 µl of Opti-MEM. The contents of both tubes were thoroughly mixed, incubated for 10 minutes at room temperature, and then added drop-wise to the cell suspension. In the case of DharmaFECT, 4 µg of plasmid together with 7 µL of DharmaFECT kb reagent mixed in 200 µL of Opti-MEM was used per transfection. This transfection method was used in K562 cells as it yielded better transfection rates. Downstream analyses were performed 48-72 hours post transfection. Considering the typically low transfection rates

observed in leukemia cells with lipofection-based techniques, we ultimately resorted to lentiviral transduction to assess the long-term effects.

2.2.1.3.-Lentiviral transduction and generation of stable cell lines

To generate lentiviral particles, HEK293T cells were transfected using Lipofectamine 3000 or DharmaFECT kb reagents following the manufacturer's protocol (refer to chapter 2.2.1.2). Briefly, 3 x 10^6 HEK293T cells were seeded in 4 mL of growth medium in a 6 cm culture dish the day prior transfection. Two packaging plasmids, psPAX2 (Addgene) and pMD2.G (Addgene), along with transfer plasmids (pLVX EV-RFP, pLVX FAM-RFP; pLVX shC – Crimson, pLVX-shFAM30A, pLVX-shMSI2, pLVX-shHNRNPA1) were diluted in Opti-MEM and transfected into the cells. Growth media were replaced 8 hours post-transfection. Lentivirus was harvested at 24- and 48-hours post-transfection and filtered through a 0.45 µM filter. After an overnight incubation at 4°C, lentiviral particles were centrifuged at 2000 rpm at 4°C for 30 minutes. The resulting pellets were resuspended in 500 µl PBS and stored at -80°C if required in the long-term. Generally, the transduction of leukaemia cells (KG-1a, MV4-11 and K562 cells) was performed at the same day of the last harvest the media with the lentiviral particles including 10 µg/µL polybrene was carefully resuspended with 2 x 10^6 cells (previously washed with PBS) and the mix was centrifuged at 24°C for 1 hour at 1000 rpm and subsequently cells were grown for another 3 days incubated with 4 µg/mL for selection. Afterwards, cells were washed three times with PBS and grown in RPMI with 20% FBS. Cells expressing RFP and/or Crimson were sorted using a FACS Melody cell sorter. The viral titer was determined by fluorescence titering with a flow cytometer (Miltenyi Biotech). This method was entirely performed at the lab of Prof. Dr. Stefan Hüttelmaier. The stable cell lines were created using different constructs, as follows:

KG-1a cells	**CTRL**	**OE**	**KD**	**REC**
pLVX-shC	+	+	-	-
pLVX-shFAM30A	-	-	+	+
pLVX-EV	+	-	+	-
pLVX-FAM	-	+	-	+

KG-1a / K562 cells	**CTRL**	**OE**	**shM2/shA1**	**shM2/shA1 OE**
pLVX-shC	+	+	-	-
pLVX-shMSI2/hnRNPA1	-	-	+	+
pLVX-EV	+	-	+	-
pLVX-FAM	-	+	-	+

MV4-11 / ME-1 cells	CTRL	OE
pLVX-EV	+	-
pLVX-FAM	-	+

2.2.1.4.-Cell viability

Cell viability was assessed using the CellTiter-Glo® Luminescence Cell Viability Assay (Promega) following the manufacturer's instructions. The assay relies on measuring intracellular ATP concentration as an indicator of metabolically active cells. Cells were seeded at a concentration of 0.5-2 x 10^5 cells/mL with fresh media in a 24-wells plate and cell viability was determined after 48-72 hours after. Then, CellTiter-Glo® reagent was added to the cell culture medium with cells at a 1:1 ratio and incubated for 10 minutes. The resulting luminescence signal generated by the luciferase reaction, was detected using a GloMAX® luminometer (Promega). Another method to measure cell viability was determined by counting viable cell numbers (trypan blue exclusion, Bio-Rad) in the TC20 automated cell counter (Bio-Rad). This method was used for determination of cell expansion through time before seeding cells for colony-formation studies.

2.2.1.5.-Cell cycle analysis

For cell cycle analysis, cells were seeded at a concentration of 0.5 x 10^6 cells/mL in a 6-wells plate 48 hours prior analysis. Afterward, cells were collected and washed with PBS, and then incubated in cold 70% ethanol for 1 hour on ice. Subsequently, the cells were centrifuged, washed twice with PBS, and incubated with 200 µl RNaseA (Roth) in PBS (100 µg/ml) for 30 minutes at room temperature. After another centrifugation step, the cells were stained with DAPI (4',6-diaminidino-2-phenylindole; 1:5000). The stained cells were subjected to DNA content analysis using flow cytometry with a MACSQuant flow cytometer and the MACSQuantify Software (Miltenyi Biotec).

2.2.1.6.-Colony-forming cell (CFC) assay

Prior seeding out cells for colony-formation studies, cell expansion through a time-course (1 to 3 days) was calculated by counting viable cell numbers (trypan blue exclusion) in normal growth medium. Then, cells were seeded at 5000 cells/mL in 3 mL MethoCult™ (Stem Cell Technologies, H4434) in a 3.5 cm culture dish and placed in an incubator at 37 ºC and 5% CO_2. After 14 days of incubation, the number of colonies was manually counted and classified into each colony type following manufacturer´s instructions.

2.2.1.7.-Hypoxia

Prior seeding cells in hypoxia conditions, cells were washed with PBS and seeded at a cell density of 0.5 x 10^6 cells/mL in a T75 flask. Then, cells underwent hypoxic treatment (0.5% oxygen, 5% carbon dioxide) which was performed in a humidified incubator (Innova CO-170; New Brunswick Scientific) and cells were maintained for 72 hours. Afterwards, cell viability numbers were determined by trypan blue exclusion.

2.2.1.8.-Apoptosis assay

To measure apoptotic levels after treatment with chemotherapeutics, cells were seeded and incubated for 48 hours before analysis. Afterwards, cells were pelleted and washed twice with PBS and incubated with the staining agents Annexin-V-FITC (1:2000) and DAPI in a solution of PBS with 1% Bovine Serum Albumine (BSA) for 1 hour at 4ºC in the dark. Annexin-V-positive cells were used as early apoptosis and Annexin-V-positive, DAPI-positive cells as late apoptosis. Then, cells were washed again with PBS and plated in a 96-well plate for subsequent analysis of early and late apoptosis using flow cytometry with a MACSQuant flow cytometer and the MACSQuantify Software (Miltenyi Biotec).

2.2.1.9.-Luciferase reporter assay

In order to examine of how FAM30A repeats affect activation of the transcriptional activity of RUNX1 via MSI2, stable K562 cell clones with overexpression of FAM30A repeats in combination with MSI2 knockdowns were used for this experimental set up. K562 stable cell lines were seeded one day before transfection at a cell density of 2.5 x 10^5 cells/mL in a 24-well plate. A day after, cells were transiently transfected using DharmaFECT kb with pmirGLO Promo+Stopp*HindIII plasmid (empty vector) comprising two and three palindromic RUNX1 binding sites (TGTGG) in the promoter region of the Firefly luciferase gene. Another plasmid system used in this experiment was pmirGLO (empty vector) with RUNX1 P1 and P2 promoters next to the Firefly luciferase (see Figure 36A). Renilla luciferase on the same plasmid served as the normalization control. The luminescence values from Firefly and Renilla luciferases were determined 48 hours post transfection by DualGLO (Promega) according to the manufacturer's protocol

2.2.1.10.-Sucrose-based fractionation and polysome profiling

For the isolation of polysome fractions and polysome profiling, wild-type KG-1a or GDM-1 cells were grown at a concentration of 0.5-1 x 10^6 cells/mL. The day before, a sucrose gradient diluted in Gradient Lysis buffer (15-45%) was carefully layered in centrifuge tubes used for ultracentrifugation. To achieve that, five sucrose solutions were prepared (15%, 22.5%, 30%, 37.5% and 45%) and by taking the 45% solution and then carefully

overlaying each cushion with the next solution (2 mL) till achieving 10 mL total volume. Then the centrifuge tubes were placed overnight at 4ºC for an optimal diffusion of the gradient. A day after cells were pelleted and washed once with PBS, lysed with Gradient Lysis Buffer for 5 minutes and centrifuged in a tabletop microcentrifuge at 13000 rpm for 5 minutes. The supernatants were collected in new tubes and maintained on ice until further processing for polysome profiling. The equivalent of 10 million KG1-a/GDM-1 lysated cells was layered onto each 11 mL of 15-45% linear sucrose gradient. Gradients were then centrifuged (40000 rpm, 2 hours, 4ºC) in a SW40 Ti rotor. Then, gradient fractions (11 in total) were collected at 500 µl intervals from which 200 µl and 50 µl was used for downstream analysis of RNA and protein, respectively. For measurements of RNA transcripts, each isolated polysomal fraction was mixed with 1 mL of TRIzol and standard RNA procedure was followed. Unfractionated cell lysates served as input material for downstream analysis to assess the relative abundance of transcripts in each isolated fraction of the sucrose gradient.

2.2.1.11.-Subcellular fractionation

For the separation of wild-type MV4-11 and KG-1a cells in nuclear and cytoplasmic fractions, 3 x10^6 cells per conditions were harvested, washed with PBS and lysed in Cell Fractionation buffer including 100 µg/mL of digitonin for MV4-11 and 200 µg/mL for KG-1a cells (refer to 2.1.4.1 section). After 5 minutes incubation with digitonin, subcellular fractions were separated by centrifugation at 1000 g for 2 minutes and supernatant containing the cytoplasmic fraction was transferred to a new tube. The pellet was then subjected to another round of lysis in Gradient Lysis Buffer and centrifuged at 1000 g for 2 minutes. Afterwards, supernatant was discarded and the remaining pellet containing the nuclear fraction was lysed in Total Lysis Buffer. Downstream analysis was performed by Western Blotting confirming subcellular fractionation by detecting EEF2 in the cytoplasmic and PTB in the nuclear fraction.

2.2.1.12.-Compound testing

To assess cell viability in response to varying concentrations of chemotherapeutic agents (Cytarabine/Ara-Cytarabine and Daunorubicin), stable cell clones of KG-1a, MV4-11 and ME-1 cells were plated in a cell density ranging 1- 2.5 x 10^5 cells/mL in a 24-well plate. Beforehand, IC_{50} values were calculated in wild-type cell lines by using GraphPad Prism 8 and validated with published data in the same cell lines. Vehicle (water) or the respective compounds diluted in water (refer to 2.1.8 section) were added 24 hours post-seeding at indicated concentrations and cells were incubated at 37ºC and 5% CO_2. CellTiter-Glo was used 48-72 hours post-treatment for measuring cell viability, and instructions were followed according to the manufacturer´s protocol.

In the case of treating cells with chemical compounds used for blocking protein synthesis (emetine and cycloheximide), KG1-a stable cell lines were seeded 24 hours prior treatment at a higher cell density of 0.5 - 1 x 10^6 cells/mL in a 12-well plate. The day of treatment, cells were treated with vehicles (water or 0.1% DMSO, depending on the compound), 150 µM of emetine (in DMSO) and 100 µg/mL of cycloheximide (water) for 3 and 6 hours and cells were harvested for downstream protein analysis via Western blotting. CMYC was used as a positive control due to its high mRNA turnover rate and ACTB as a loading control for normalization.

2.2.1.13.- Flow cytometry and FACS staining

In order to analysis expression of cell surface protein markers, KG-1a stable cell lines were maintained at a cell density of 0.5 - 1 x 10^6 cells/mL. The equivalent of 0.5 x 10^6 cells was harvested per condition/ combination of antibody staining, washed once with PBS and centrifuged 2 minutes at 4000 rpm at a table microcentrifuge. Cells were incubated with the respective antibody (1:50 dilution) in PBS with 1% BSA for at least 30 minutes at 4ºC in the dark. Then, stained cells were washed once more with PBS with 1% BSA and plated in a 96-well plate for subsequent analysis in MACSQuant flow cytometer for the appropriate laser. Unstained KG-1a wild-type (no fluorescence) and unstained KG-1a stable cell clones (fluorescent with RFP and Crimson markers) were used for gating and positive staining.

2.2.1.14.- Immunofluorescence and microscopy

For examining whether subcellular localization of HNRNPA1 changes, 2.5 x 10^5 KG-1a stable cell clones were seeded onto glass coverslips which were previously treated with 40% (v/v) ethanol and 60% (v/v) hydrochloric acid. Before seeding the cells in a 24-well plate, coverslips were coated with Poly-D-Lysine (A3890401, Thermo Fisher Scientific) diluted in PBS (1:1000). Cells were incubated with the coverslips for 30 minutes at 37 ºC and 5% CO_2 and the plate was centrifuged at 500 rpm for 10 minutes to be certain of cell attachment to the coverslips. Standard immunofluorescence staining involved fixing cells with a 4% formaldehyde solution in PBS for 20 minutes at room temperature, followed by permeabilization of the membranes with 0.5% (w/v) Triton-X-100 in PBS for 4 minutes. Non-specific antibody binding was blocked with 5% BSA (w/v) in PBS for 45 minutes at room temperature. Primary (1:250 dilution) and fluorescently-labeled secondary antibody solutions (1:500) were prepared in 2% BSA in PBS (refer to 2.1.5 section) and then incubated for 1 hour at room temperature. For nucleus visualization, cells were stained with DAPI in PBS for 3 minutes. The coverslips were washed with water, dehydrated in 99% ethanol three times and mounted on an object slide using ProLong™ Diamond Antifade (Thermo Fisher, Cat# P36970) to preserve fluorescent signals. The Leica SP5X

or SP8X confocal microscopes were employed to capture images, utilizing a 63x oil objective. A white light laser was used to excite the fluorescent dye, and the emitted signals were captured by HyD detectors. Nucleus staining was visualized through DAPI excitation, achieved with a diode emitting light at a wavelength of 405 nm.

2.2.2.- Molecular biology methods

2.2.2.1.-Cloning

2.2.2.1.1.- DNA amplification by polymerase-chain reaction (PCR)

The process of cloning the FAM30A repeats into an overexpression plasmid (pLVX-RFP, TakaraBio) in order to generate stable cell lines involved first amplification of the desired sequence from cDNA of KG-1a cells. All desired sequences were amplified via a PCR reaction using Q5® High-Fidelity 2X Master Mix with corresponding oligonucleotides with flanking restriction sites (refer to Table 7, section 2.1.7). For the amplification, the following reaction mix was prepared:

cDNA	100 ng
10 mM dNTPs	0.5 µL
10 µM oligonucleotides (sense/antisense)	1 µL each
5x GC buffer (NEB)	4 µL
Q5® High-Fidelity 2X Master Mix	10 µL
Nuclease-free water	Add up to 20 µL

Afterwards, the reaction tubes were mixed and the amplification followed the adjusted program from the manufacturer´s instructions:

Step	Temperature	Time	
Initial Denaturation	98 ºC	30 seconds	
Denaturation	98 ºC	10 seconds	
Primer-annealing	55-65ºC	20 seconds	35-40 cycles
Elongation	72ºC	20-30 seconds	
Final Extension	72ºC	2 minutes	

The resulting amplified cDNA was mixed with 6x DNA loading dye (NEB) loaded onto a 1% agarose gel in 1x TAE buffer with ethidium bromide and separated through gel electrophoresis in 1x TAE buffer at 120 V. The Quick-load 2-log DNA-ladder (NEB) was used as size marker. Subsequent detection of the amplicon was achieved using UV light. Gel pieces containing amplicons of the expected nucleotide length were excised, and the DNA was extracted and purified using the Agarose Gel Extraction Kit, followed by

elution with ddH2O. The eluted PCR products underwent ligation into the linearized blunt-end vector pCR™-Blunt II - TOPO™, following the manufacturer's instructions. The solution containing the ligated plasmid was incubated with ultracompetent E. coli (TOP10) for 20 minutes at 4 °C. Transformation of the bacteria was carried out through heat shock at 42 °C for 1 minute, followed by a 10-minute resting phase on ice. Following transformation, the bacteria were cultured in 800 µL of LB growth medium for 1 hour at 37 °C and then spread onto LB-agar plates containing 50 mg/mL Kanamycin for overnight growth. The clones on the plates underwent testing for successful insertion in a colony-PCR reaction with OneTaq® 2X polymerase Master Mix using flanking oligonucleotides provided by manufacturer, following the program:

Step	Temperature	Time	
Initial Denaturation	98 ºC	30 seconds	
Denaturation	98 ºC	15 seconds	
Primer-annealing	55-65ºC	30 seconds	25-30 cycles
Elongation	68ºC	1 minute / kb	
Final Extension	72ºC	5 minutes	

Then, positive clones were subsequently cultivated in liquid LB-medium with 30 mg/mL Kanamycin at 37 °C overnight.

Plasmid DNA was isolated using the Plasmid Mini-Prep Kit from Jena Bioscience, following the manufacturer’s protocol. Correct sequence insertions were analyzed by Sanger Sequencing at Eurofins Genomics GmbH. The final step of cloning into the final expression plasmid involved enzymatic digestion of 2 µg of the empty expression plasmids (refer to Table 11) and the pCR™-Blunt II - TOPO™ containing the valid gene sequence, using 1 µL per restriction enzyme (New England Biolabs) in the corresponding buffer, followed by separation on agarose gel electrophoresis. Then, the gel-extracted DNA including the linearized expression plasmid and target gene sequence, was eluted in 30 µL ddH2O. A ligation reaction was then performed, involving 8 µL insert, 2 µL linearized vector, 1 µL T4 DNA ligase (New England Biolabs), 2 µL corresponding 10x ligation reaction buffer and ddH_20 up to a total volume of 20 µL. After a 2-hour incubation at room temperature, the ligation reactions were transformed into bacteria as described earlier. The subsequent steps, including plating, PCR testing, and liquid cultivation, followed the procedures outlined previously, with agar plates containing Kanamycin or Ampicillin for positive clone selection. Cloning of RUNX1 P1 and P2 promoters into the

final plasmids involved DNA amplification via PCR using genomic DNA from K562 cells as template (approx. 50-100 ng as template) and it followed the same protocol as described above.

2.2.2.1.2.- Cloning via oligonucleotide annealing

The cloning of shRNAs into the final plasmid utilized pLVX shC as the empty vector (see Table 11) and aimed to create constructs for establishing stable knockdowns of FAM30A, HNRNPA1, and MSI2. To initiate the process, oligonucleotides comprising the shRNA target sequences and flanking restrictions sites were annealed mixing 5 µL of sense and antisense oligonucleotides (100 µM each) and 10 µL nuclease-free water. Then, the mix was incubated at temperatures ranging from 95ºC and progressively decreasing to 65ºC for a total period of time of 10 minutes and afterwards cooled down at room temperature. Following the aforementioned protocol, the annealed oligonucleotides were ligated into the ultimate linearized plasmid. The cloning of palindromic RUNX1 binding sites into the pmirGLO Promo+Stopp*HindIII vector followed the same procedure, using oligonucleotide with flanking restriction sites and including palindromic RUNX1 binding sites. Due to unspecific annealing, two and three binding sites could be cloned (x2 and x3 RUNX1-BS) and both were used for downstream luciferase experiments.

2.2.2.2.-Isolation of total RNA and genomic DNA from cells

For cell culture samples, total RNA was extracted using guanidiniumthiocyanatphenol-chloroform extraction, as described by (Chomczynski & Sacchi, 1987). Cells were washed with ice-cold PBS, lysed in 1 ml TRIzol (refer to Table 2) and transferred into an RNase-free 1.5 ml reaction tube. The sample was frozen and stored at -80°C if no immediate use was required. Upon thawing, RNA was extracted by adding 200 µl chloroform, thoroughly mixing the sample for 15 sec by inverting, and incubating for 3 min at room temperature (RT). After centrifugation for 30 minutes at 13000x g and 4°C, the RNA-containing aqueous (upper) phase and transferred to a new RNase-free tube, mixed with 1 volume of ice-cold 2-isopropanol, vortexed and incubated for 20 minutes at RT for precipitation. The RNA was then precipitated by centrifugation at full speed for at least 45 minutes at 13000 x g and 4°C, washed twice with ice-cold 80% ethanol, and centrifuged for 10 min at 13000 x g and 4°C. After removing the ethanol with the air pump, the RNA was dried for a few minutes at room temperature and dissolved in nuclease-free water for 5 minutes. RNA concentration and quality was assessed using Tecan Spark® multimode microplate reader via measuring the ratio of absorbance (approx. 2) at wavelengths of 260 and 280 nm. Genomic DNA (gDNA) from human cell lines was extracted with the PureLink™ Genomic DNA-Minikit, following the manufacturer's instructions, and stored at -20°C or -80ºC for long-term storage.

2.2.2.3.-Reverse transcription

A reverse transcription reaction was conducted by random hexamer priming to synthesize complementary DNA (cDNA). The master mix used for the reaction was the following:

RNA (2 µg) diluted in ddH_20	
10 mM dNTPs	1 µL
5x M-MLV Reverse Transcriptase buffer (Promega, Cat #M531A)	4 µL
M-MLV Reverse Transcriptase (Promega, Cat #M1701, 200U/µL)	0.5 µL
RNase inhibitor (Promega, Cat #N2511, 2500 units)	0.25 µL
ddH_20	Add up to 20 µL

After thorough mixing, the reaction tube underwent an incubation process in a thermocycler (Eppendorf), adhering to the manufacturer's instructions. This involved maintaining the temperature at 42ºC for 2 hours, followed by the subsequent inactivation of the reverse transcriptase enzyme at 75ºC for 15 minutes. Then, the generated cDNA diluted in ddH_20 was used directly for RT-qPCR or stored at -20 ºC.

2.2.2.4.-Quantitative Real-Time PCR (RT-qPCR)

For the evaluation of mRNA relative abundances and expression changes among the different conditions, real-time quantitative PCR (RT-qPCR) was conducted using SYBRGreen I fluorescent dye. The amplification was performed using two different methods depending on the number of samples to be analyzed. For the first method, each well contained 2.5 µL of diluted (1:10 in ddH_20) cDNA and 2.5 µL of ORA™ qPCR Green ROX L Mix (highQu, Cat #QPD0101) with 0.4 µM of each oligonucleotide sense/ antisense (100mM) and the analysis was performed on a 384-well plate using LightCycler 480 II Roche device. The other method consisted on a reaction mix on each well of 5 µL of diluted cDNA and 5 µL of GoTaq® qPCR Master Mix 2X (Promega, Cat #A600A) with 0.4 µM of each oligonucleotide sense/ antisense (100mM) and the analysis was performed on a 96-well plate using BioRadi Cycler iQ device. The oligonucleotides used for the qRT-PCR (refer to Table 8) were designed with Primer-Blast from NCBI website and settings were optimized to have a length of amplicon of 100-250 nucleotides and annealing temperature 60ºC. The PCR conditions were the following:

Step	Temperature	Time	
Initial activation of polymerase	95 ºC	5 minutes	
Denaturation	95 ºC	10 seconds	
Oligonucleotides annealing	60 ºC	10 seconds	45 cycles
Elongation	72ºC	20 seconds	
Melting curve gradient*	55-95 ºC		

*This final step was performed to confirm that the oligonucleotides were binding efficiently and no by-products (i.e. double peaks) were produced during the amplification.

The quantification of newly synthesized double-stranded DNA involved assessing the intercalated fluorescent SYBRGreen dye. The determination of the cycle of quantification (Ct) was performed using the second derivative maximum. Calculations employing the ΔCt or the ΔΔCt method, elucidated changes in RNA expression levels relative to housekeeping genes (EEF2, ACT, PPIA) and control conditions (Livak & Schmittgen, 2001) using the following formula:

$$\text{Relative mRNA level} = 2^{(-((Ct\ \mathbf{X}\text{-}Ct\ \mathbf{N})\ \text{target condition} - (Ct\ \mathbf{X}\text{-}Ct\ \mathbf{N})\ \text{control condition}))}$$

X = target RNA; **N** = housekeeping RNA

2.2.2.5.- In vitro transcription

The DNA templates for CTRL (control RNA), FAM (1) (one FAM30A repeat) and MMut (one FAM30A repeat with mutated MSI2-binding sites) were amplified through oligonucleotide annealing with specific oligonucleotides including flanking sequences for binding of the T7 RNA polymerase (refer to Table 10) with the PCR program provided by manufacturer and using the following mix:

GoTaq® qPCR Master Mix 2X	60 µL
ddH_20	50 µL
Oligonucleotides (sense/antisense)	5 µL each

The PCR product was separated on a 2% agarose gel in TAE-buffer and DNA was extracted with Agarose Gel Extraction Kit (Jena Bioscience). This diluted DNA served as a template for subsequent *in vitro* transcription. In the case of FAM (6) (six FAM30A repeats), amplification was challenging so the pLVX-FAM vector was digested with XhoI for linearization and this was used as template for transcription as it includes a T7 polymerase binding sequence before the six repeats. Transcription was carried out using the diluted DNA, with an additional 20 µL of Transcription Optimized 5x Buffer, 10 µL of 100 mM DTT, 1 µL of 100 mM dNTPs (ATP, GTP, CTP), 0.3 µL of 100 mM unlabeled UTP, 1 µL RNAsin® Ribonuclease inhibitor, and 4 µL T7 RNA polymerase (all reagents acquired from Promega). In order to tag the newly synthesized RNA, 1 µL of 5mM biotinylated-UTP or Biotin-11-UTP (Jena Bioscience) was added to the reaction and ddH_20 was added up to a total volume of 100 µL, and transcription was carried out at 37 °C for 4 hours. After the reaction, the DNA template was digested by adding 5 µL RQ1-

DNase (Promega) at 37 °C for 20 minutes, and RNA was extracted with TRIzol (refer to 2.2.2.2 section).

2.2.2.6.- Biotinylated-RNA pulldown

To conduct RNA pulldowns, 15-20 µL of Dynabeads™ MyOne-Streptavidin-T1 (Thermo Fisher Scientific) per pulldown were placed into a reaction tube. The beads were washed three times with an equal volume of BW buffer (refer to Table 2) using a magnetic rack and then incubated with 0.5 - 1 µg of *in vitro* transcribed RNA. As a control, beads were incubated with no RNA for the beads control (namely BC) for 20 minutes at room temperature. Concurrently, cells were washed with PBS and lysed with Gradient Lysis buffer (Table 2) for 10 minutes and centrifuged for 5 minutes at 13000 rpm. The resulting supernatant was transferred to a new reaction tube, and 20 µL (1:10) of the lysate was used for the input sample for downstream protein analysis. Meanwhile, 200 µL of the cell lysate was mixed with each of the RNA-coupled streptavidin beads and incubated for 45 minutes at room temperature. Following this, pulldown samples were washed three times with 200 µL of Gradient Lysis buffer and separated on a magnetic rack to eliminate nonspecific proteins and RNA. For the analysis of associated proteins, the beads were resuspended in 50 µL of 1x NuPage™ LDS sample buffer (Thermo Fisher Scientific) diluted with Gradient Lysis Buffer and boiled at 95 °C for 5 minutes to elute proteins. Then, samples were then separated from the beads on a magnetic rack and subsequently analyzed for downstream protein analysis through SDS-PAGE and Western blotting.

2.2.3.- Protein-biochemical methods

2.2.3.1.- Protein extraction, SDS-PAGE and Western Blot

For the analysis of changes in protein abundances among the different samples and conditions, cells were harvested and pelletized by centrifugation at 4000 rpm for 3 minutes in a 1.5 mL centrifuge tube, washed once with PBS and the supernatant discarded. If the protein sample was not needed immediately, the pellet was frozed in liquid nitrogen prior to storage at -80 ºC.

In case the pellets were needed for protein expression analysis, total protein extraction was performed and cells were lysed in Total Lysis buffer (refer to Table 2) supplemented with 1 µL of benzonase or Turbo nuclease [Jena Bioscience, 250 U/µl]) for digestion of genomic DNA and protease and phosphatase inhibitor cocktails (1:1000) (Sigma-Aldrich). After a 10-minute incubation, the lysate was centrifuged at 13.000 rpm for 5 minute and the supernatant was then transferred to a new tube. Protein concentration was assessed using the colorimetric DC Protein Assay following manufacturer’s

instructions and then measuring absorbance at 700 nm with the GloMax® 96 well Microplate Luminometer. Protein samples were standardized to equal amounts via BSA quantification, mixed with 4x NuPage™ LDS sample buffer containing 50 mM DTT, and boiled at 95 °C for 5 minutes to achieve complete protein denaturation. The samples were loaded into a NuPAGE® Novex® 4-12% Bis-Tris protein gel and separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE) using the corresponding NuPAGE® MOPS or MES SDS-containing running buffer (Thermo Fisher Scientific). Immobilized proteins in the gel were transferred to a nitrocellulose membrane (Amersham, GE Healthcare) employing NuPAGE transfer buffer with 10% methanol. If necessary, transfer was assessed by Ponceau S staining. Subsequently, the membrane underwent a 1-hour blocking step with 5% (w/v) skimmed milk in PBS-T. Primary and fluorescently-labelled secondary antibody solutions (refer to Table 3 and 4) were prepared in 5% milk in PBS-T. Incubation with the primary antibody occurred for at least 2 hours at room temperature or overnight at 4°C, two washes with PBS-T followed by 1 hour incubation at room temperature for the secondary antibody and washed again three times with PBS-T. Fluorescent signals were then detected at 680 nm or 800 nm using the Odyssey Infrared Imaging System (LI-COR). Quantification of the protein levels based on intensities was carried out via Image Studio™ software (LI-COR) and relative protein amounts were determined based on loading controls and control conditions.

2.2.3.2.- RNA immunoprecipitation (RIP)

RNA co-immunoprecipitations were used to study the interaction of HNRNPA1 and MSI2 to its RNA targets. First, KG-1a cells (8×10^6 per condition) were harvested and washed with PBS and then lysed using Immunoprecipitation buffer (see Table 2). Inputs for either RNA or protein (20 µL) were extracted from the lysates. For RNA, these inputs were combined with 980 µL TRIZOL, while for protein, they were mixed with 7 µL of 4x NuPage™ LDS sample buffer containing 50 mM DTT.

Then, 10-15 µL of magnetic Protein G Dynabeads per condition (Life Technologies) were washed three times with 200 µL of Immunoprecipitation buffer and 5-10 µg/condition of anti-MSI2 (Bethyl) and anti-HNRNPA1 antibodies (see Table 3) were added and incubated for 15 minutes. The resulting supernatant with the protein lysate was incubated with the antibody-bound beads for 60 minutes at room temperature on a spinning-wheel. Following three washing steps with Immunoprecitation buffer, protein–RNA complexes were eluted using 100 µL of Immunoprecipitation buffer supplemented of 1% SDS and incubated at 65°C for 10 minutes and beads were separated in the magnetic rack. The supernatant containing the eluted protein-RNA complexes was transferred to a new tube and 20 µL was used for protein analysis and 80 µL for RNA analysis. Protein enrichment

was subsequently analyzed by Western blotting and co-purified RNAs were extracted using TRIzol and subjected to analysis by qRT-PCR.

2.2.3.3.- Quantitative proteomics by LC-MS/MS

For the analysis of associated proteins with FAM30 repeats, proteins were enriched using *in vitro* transcribed and biotinylated FAM30A repeats RNA compared to beads control with no RNA (BC) and a control RNA (CTRL) which is 500 base pairs before the repeats. The total cell lysate was extracted from 8 × 10^6 KG-1a wild-type cells lysed with Lysis Buffer for LC-MS/MS (see Table 2) suitable for mass spectrometry analysis followed by RNA pulldown as described above (see section 2.2.2.5). Following the third washing step with same lysis buffer, pulldown samples were maintained in the binding buffer. Subsequently, sample preparation and further analysis was carried out by Dr. Christian Ihling (AG Sinz, Institute for Pharmacy, Martin-Luther University) following the provided protocol by Dr. Ihling: pulldowns were prepared for liquid chromatography/tandem mass spectrometry (LC-MS/MS) following a modified FASP (filter-aided sample preparation) protocol (Wiśniewski et al., 2009). Protein samples were incubated and washed with 8 M urea in 50 mM HEPES ((4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 10 mM TCEP, pH 8.5). Washing steps were performed using 0.5 mL centrifuge filter units with a 30-kDa cutoff (Sigma Aldrich) and two centrifugation steps followed (14000 × g, 10 min). Afterwards, alkylation of proteins was performed with 50 mM iodoacetamide in 8 M urea, 50 mM HEPES, pH 8,5 at room temperature during 20 min in the dark. Samples were washed twice with 50 mM HEPES, pH 8.5 (18000 × g, 2 x 10 min) before they were incubated with 1 µg trypsin in 50 mM HEPES, pH 8.5 at 37 °C overnight. After enzymatic digestion, samples were acidified with trifluoro acetic acid (TFA) at 0.5 % (v/v) final concentration. Proteolytic peptide mixtures were analyzed by LC-MS/MS on an UltiMate 3000 RSLC nano-HPLC system coupled to an Orbitrap Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). Peptides were separated on reversed phase C18 columns (trapping column: Acclaim PepMap 100, 300 µm × 5 mm, 5 µm, 100 Å (Thermo Fisher Scientific); separation column: µPAC™ 75 cm C18 (Pharmafluidics). After desalting the samples on the trapping column, peptides were eluted and separated using a linear gradient from 3% to 35% B (solvent A: 0.1% (v/v) formic acid in water, solvent B: 0.08% (v/v) formic acid in acetonitrile) with a constant flow rate of 300 nl/min over 180 min. Data were acquired in data-dependent MS/MS mode with stepped higher-energy collision-induced dissociation (HCD) and normalized collision energies of 27%, 30%, and 33%. Each high-resolution full scan (m/z 375 to 1799, R = 140000 at m/z 200) in the orbitrap was followed by high-resolution product ion scans (R = 17500) of the 10 most intense signals in the full-scan

mass spectrum (isolation window 2 Th); the target value of the automated gain control was set to 3,000,000 (MS) and 200,000 (MS/MS), maximum accumulation times were set to 100 ms (MS) and 250 ms (MS/MS). Precursor ions with charge states <3 + and >8 + or were excluded from fragmentation. Dynamic exclusion was enabled (duration 60 seconds, window 3 ppm).

2.2.4.- Experiments in vivo

Human cell xenograft experiments were performed under complete supervision of Prof. Dr. David Vetrie (Wolfson Wohl Cancer Research Centre, University of Glasgow, The United Kingdom) and his laboratory staff. All mice were housed at the Beatson Research Unit (University of Glasgow, UK), and experimental protocols were approved by Karen Keeshan (Wolfson Wohl Cancer Research Centre, University of Glasgow, The United Kingdom). Before engraftment, KG-1a stable cell clones were cultured in RPMI with 20% FBS for five days for optimum cell growth at 37ºC and 5% CO_2. KG-1a cells (2×10^6 cells per mouse) were diluted in ice-cold PBS and transplanted via tail vein injection into mice 8-12-week-old sub-lethally irradiated or busulfan treated NRG-W41 immune-compromised mice (NOD.Cg-Rag1tm1Mom KitW-41J Il2rgtm1Wjl/EavJ), with those showing evidence of human cell engraftment randomized for treatment. After 28 days of treatment, all mice were euthanized, and bone marrow (BM) was harvested from the hind legs (ilia, femurs, and tibias). BM was filtered and stained for flow cytometry using human anti-CD34-BV510 and human anti-CD38-VioBlue (see table 5). Engrafted human cell immunophenotypes were determined by flow cytometry (BD, FACSVerse). To assess level of engraftment of human cells, cells were gated for APC-positive cells as KG-1a stable clones emit RFP/Crimson fluorescence and also using transplanted mice as negative control (with no fluorescence).

2.2.5.- Computational analysis

2.2.5.1.- Conservation analysis

To assess the phylogenetic conservation of FAM30A repeat across vertebrates, the sequence of a single FAM30A repeat (76 and 78 base pairs) was downloaded from NCBI gene website (https://www.ncbi.nlm.nih.gov/gene/) using its annotated sequence (NR_026800.2) and subsequently blasted against all taxonomic nucleotide databases using NCBI nblast tool (https://blast.ncbi.nlm.nih.gov/). Following the alignment of a single repeat through the NCBI nblast tool against various taxonomic nucleotide databases, the sequence exhibited alignment primarily with analogous sequences found in 21 primate species. Thus, the genome sequences of these species were downloaded from NCBI and ENSEMBL (https://www.ensembl.org): Aoutus nancymaae, Sapajus apella, Callithrix jacchus, Cenus capucinus, Cercocebus atys, Chlorocebus sabaeus,

Colobus angolensis, Macaca fascicularis, Macaca mulatta, Macaca nemestrrina, Mandrillus leucophaeus, Papio anubis, Piliocolobus tephrosceles, Rhinopithecus bieti, Rhinopithecus roxellana, Theropithecus gelada, Trachypithecus fracoisi, Hylobates moloch, Pan troglodytes, Gorilla gorilla and Homo sapiens.

Then, in order to confirm that only the repeats region of FAM30A is evolutionary conserved, a longer sequence including the flanking regions of the repeats was downloaded (500 base pairs upstream and downstream) and blasted against the primate genomes where FAM30A repeat was detected. After multiple sequence alignment using M-Coffee software (Wallace et al., 2006) and a conservation score was extracted from 0 (lowest) to 9 (highest) and the conservation along the 1500 base pairs was represented in a curve after locally weighted regression by Lowess algorithm using GraphPad 8.0.

2.2.5.2.- Bulk RNA-sequencing and Gene Set Enrichment Analysis (GSEA)

Total RNA-sequencing (bulk RNA-seq) utilized 1 µg of RNA from KG-1a stable cell clones (OE, CTRL, KD, REC) grown for five days in culture after transduction, puromycin selection and sorting (n=3). Novogene (Hong Kong) conducted library preparation and sequencing on an Illumina NextSeq platform. The initial processing was performed by Dr. Danny Misiak (group of Prof. Dr. Stefan Hüttelmaier). This involved the removal of low-quality read ends and residual sequencing adapters using Cutadapt (v2.8). Subsequently, reads were aligned to the human genome (UCSC GRCh38/hg38) using HiSat2 V2.1.0 (Kim et al., 2019), and FeatureCounts (v2.0) (Y. Liao et al., 2013) summarized gene-mapped reads. Annotations were based on Ensembl (GRCh38.102) (Yates et al., 2016). Differential gene expression (DE) analysis employed the R package edgeR (v3.34) (M. D. Robinson et al., 2010) with TMM (trimmed mean of M values) normalization on raw count data. RNA expression values were reported as FPKM (fragments per kilobase million mapped reads).

The Gene Set Enrichment Analyses (GSEA) were conducted using GSEA software (V3.0) (Subramanian) on pre-ranked lists, with a focus on Hallmark and Gene Ontology (Biology Processes) gene sets from MSigDB (v6.2) (Liberzon et al., 2015) as well as leukemic stem cell signatures. The ranking of all protein-coding genes was based on fold changes observed in OE vs CTRL, KD vs CTRL and REC vs KD as determined by RNA-seq. A permutation number of 1000 was implemented, and classical enrichment statistics were chosen for the analysis.

2.2.5.3.- MSI2-CLIP data

MSI2 (Musashi-2) CLIP peak data of K562 and NB4 cells were obtained from the Gene Expression Omnibus (GEO accession numbers: GSE62115; GSM2448651; GSE69583).

Peaks from the obtained 6 experiments were annotated to gene features (5' UTR, CDS and 3' UTR), summarized and visualized via an in-house R-script. The processing of data, analysis and visualization was performed by Dr. Danny Misiak (group of Prof. Dr. Stefan Hüttelmaier).

2.2.5.4.- Kaplan Meier analysis

Survival analysis for FAM30A, HNRNPA1 and MSI2 expression of patients was conducted using the Kaplan Meier plotter (Lánczky & Győrffy, 2021) integrating survival data from different gene datasets of Acute myeloid leukaemia patients (n=1608).

2.2.5.5.- Functional annotation clustering

The potential proteins bound by FAM30A repeats RNA versus beads control (no RNA) were clustered based on identified peptides (fold-change, $P < 0,05$) derived from LC-MS/MS analysis and uploaded onto the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (D. W. Huang et al., 2007). From the proteins significantly linked with FAM30A repeats RNA, a gene list was derived and underwent Gene Ontology (Biology Process) classification (http://david.abcc.ncifcrf.gov/). The most significant biological processed were selected and depicted.

2.2.5.6.- Softwares, databases, datasets and visualization

Western blot images were visualized, and the band intensities calculated using Image Studio™ Lite Quantification Software (LI-COR Biosciences). In case the orientation, brightness or contrast of the images needed adjustment, Adobe Photoshop or GIMP software were used. To visualize the sequencing files provided by Eurofins and to create plasmid maps, SnapGene software was used. The flow cytometry results were analyzed and graphs depicted using FlowJo™ v10.8 software (BD Life Sciences). Secondary structures of RNA were predicted and depicted using RNAFold web server (Kerpedjiev et al., 2015). Microsoft Word PowerPoint was used for including text boxes in the figures and GraphPad Prism 8 was utilized for making the graphs in the figures. The datasets used for MSI2-clip data were extracted from NCBI's Gene Expression Omnibus (GEO) databank: GSE62115 (n=1, K562 cells), GSE69583 (n=2, NB4 cells), GSM2448651 (n=3, K562 cells). The clip data was depicted and visualized using Integrative Genomics Viewer (IGV) software (J. T. Robinson et al., 2011). RNA expression data from various tissues and tumour types were downloaded from deepbase v3.0: https://rna.sysu.edu.cn/deepbase3/, ENCODE: https://genome.ucsc.edu/encode/; TCGA: https://www.cbioportal.org/study/ and R2 platform: https://r2.amc.nl/

For identification of RNA-binding protein motifs within the FAM30A sequence, the softwares RegRNA2.0: https://regrna2.mbc.nctu.edu.tw/ and RBPmap: https://rbpmap.technion.ac.il/overview.html were used.

All schematic illustrations used for the figures were created with BioRender (biorender.com) and are applicable for publication with the following licences: GE270LR0IV, YP270LQUEK, BH270LQVYF, OP270LQP8V, RP270LR1NT, EE270LQXEU, RE270LR504.

2.2.6.- Statistics

Unless specified differently, all experiments were conducted using a minimum of three independent biological replicates. Statistical analyses were conducted using GraphPad Prism software (V8.0) or Microsoft Excel. Statistical significance was determined using unpaired two-tailed Student's t-test, unless specified. Statistical analysis for phylogenetic conservation of the FAM30A repeats across vertebrates was performed using locally weighted regression of curve by Lowess algorithm in GraphPad. Gene expression correlations were assessed via Pearson correlation using the cBioPortal for Cancer Genomics website (cbioportal.org), or the Logrank test (Mantel–Cox) in GEPIA2 (gepia2.cancer-pku.cn) and Kaplan Meier Plotter (kmplot.com). The number of replicates is indicated in the respective figure legend.

3. RESULTS

3.1. Identification of FAM30A as a potential oncogenic lncRNA with functional relevance in AML LSCs

3.1.1.- Transcriptome-wide expression analysis of lncRNAs in AML

Over the past recent years, efforts in AML research have primarily concentrated on understanding the underlying genetics and biology of AML LSCs which are responsible for disease initiation, maintenance and chemoresistance in AML patients (Ishikawa et al., 2007; Lapidot et al., 1994; Shlush et al., 2014). Most of the studies have focused on identifying mRNA expression profiles of protein-coding genes and dissecting their molecular role in AML cell populations enriched for LSCs. In addition, further comparison of LSC-related gene signatures with normal haematopoietic stem cells (HSCs) gene signatures have led to the identification of core-enriched LSC transcriptional components that contribute to “stemness” and represent a useful tool to accurately predict initial therapy resistance (Elsayed et al., 2020; Eppert et al., 2011; N. Jung et al., 2015a, 2015b; S. W. K. Ng et al., 2016). Derived from these LSC-gene signatures, lncRNAs are being recognized as key drivers of stemness, cell differentiation and chemoresistance and AML LSC-associated lncRNA signatures have been recently described (Bill et al., 2019; Schwarzer et al., 2017). Nevertheless, the functional relevance from most of the identified lncRNAs remains unknown.

In order to identify previously uncharacterized lncRNAs with oncogenic potential in AML LSCs, a transcriptome-wide expression gene profiling was performed using publicly available sequencing datasets (Figure 8). In this case, the GEPIA2 web server (Gene Expression Profiling Interactive Analysis) was chosen essentially due to several reasons: because of its recent updates and the fact that it comprises both AML tumour and healthy BM samples from the TCGA and the GTEX databases, respectively (Tang et al., 2017). Moreover, the dataset employed in this platform (TCGA-LAML) includes a heterogeneous set of AML samples (including both paediatric and adult patients) with associated clinical attributes. First, expression data of AML patients (AML tumour samples = 173 samples, healthy BM = 70 samples) was downloaded and subjected to selection and a list of overexpressed differentially expressed genes (DEG) was derived based on median gene expression in tumor versus normal samples (Figure 8A, left panel). This was performed to generate a total list of potential oncogenes in AML. A comparison of numbers of DEG revealed approx. 8000 genes with an FDR ≤ 0.01 and Fold Change (FC) >2. The list of overexpressed genes was then compared with a matrix

generated with Ensembl GRCh38.100 including biotype and gene names to assign gene annotation (Figure 8A, right panel).

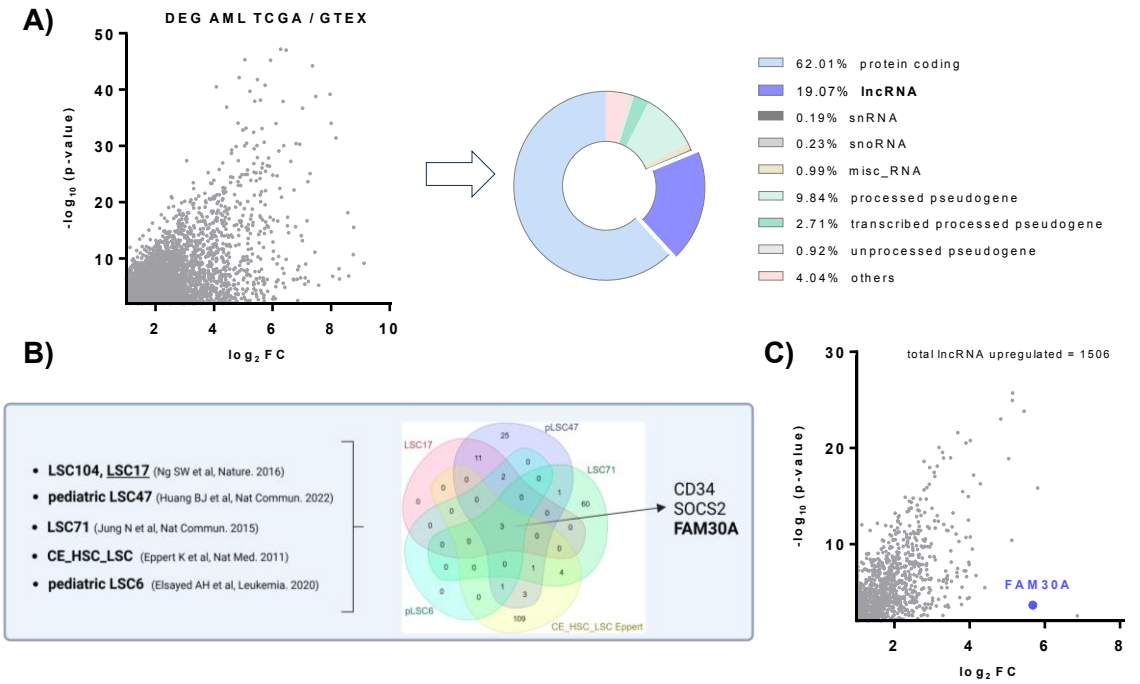


Figure 8. FAM30A is a highly upregulated lncRNA with oncogenic potential in AML LSCs. **A)** Representative graph of differentially expressed genes upregulated in AML samples (TCGA-LAML, PanCancerAtlas) when compared to healthy bone marrow samples (GTEX) (left panel) and pie chart describing gene annotation of upregulated genes (right panel). **B)** Schematic representation of the compendium of AML LSC gene signatures selected for the screening of potential oncogenic lncRNAs in AML LSCs. **C)** Representative graph of the list of upregulated lncRNAs with gene annotations derived from A. $Log_2FC = Log_2$ Fold Change of gene expression.

As expected, most of the upregulated DEG (62% and 4898 genes) were assigned to be protein-coding. The second most abundant gene class was lncRNAs (19,07% and 1506 genes) and the rest were other classes of non-coding RNAs such as small nuclear RNAs, small nucleolar RNAs and pseudogenes. This thesis focused on upregulated lncRNAs for further analysis as it is a vastly uncharacterized class of gene types in the AML LSC context.

The next step was to identify oncogenic lncRNAs related to stemness. As AML is a heterogenous type of leukaemia, “one size fits all” gene signature does not capture the cytogenetic landscape across a large cohort of AML samples. Thus, a screening was conducted across several established predictive and prognostic AML LSC gene signatures, encompassing both adult and paediatric AML. These signatures, which significantly correlate with the clinical outcomes of AML patients (i.e. response to chemotherapy), were also tested in independent cohorts (Figure 8B). When compared, there were only three genes present among all signatures: two well characterized protein-coding genes in AML LSCs (CD34 and SOCS2) and one lncRNA named

FAM30A (also called KIAA0125). Up to date, some studies using bioinformatic approaches (including the AML LSC signatures above mentioned) have only investigated FAM30A as an independent prognostic factor that contributes to worse clinical outcome in AML and myelodysplastic syndrome (MDS) (S. Guo et al., 2020; Hung et al., 2021; Y. H. Wang et al., 2021). Only one study has recently published preliminary studies investigating the potential relevance of FAM30A in AML (T. Zhang et al., 2023). However, on the mechanistic side, this study does not provide substantial insights into FAM30A function. This is why further validation *in vitro* and *in* vivo of FAM30A's role in AML is surely required. FAM30A was eligible for further in-depth studies as it was also one of the most overexpressed lncRNAs in AML when compared to normal samples with a 51-fold change and p-value of 2.29E-04 (Figure 8C).

Since FAM30A expression has been reported as a highly prognostic marker for AML and other myeloproliferative diseases, an analysis was conducted to determine whether its expression was ubiquitous across different tissues or specific to immune system-related organs. Therefore, RNA expression of this lncRNA was analyzed across various tissue and cancer types (Figure 9) using the database deepBase v3.0, which provides comprehensive expression profiles of lncRNAs based on TCGA and ENCODE datasets.

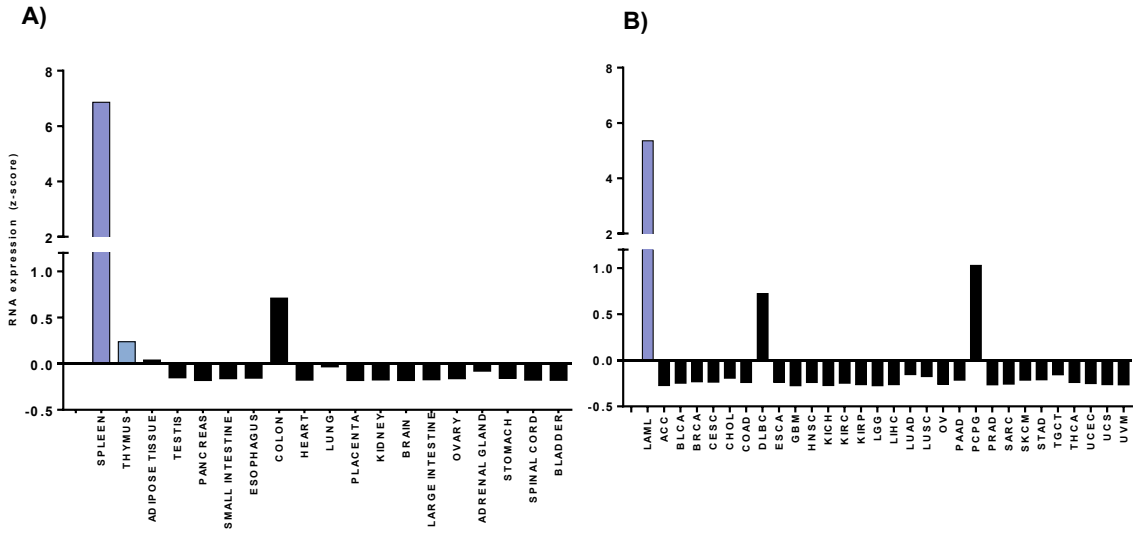


Figure 9. FAM30A expression is predominant in haematopoetic tissues and AML samples. **A)** RNA expression of FAM30A (ENSG00000226777) values normalized in terms of Z-score across a variety of normal tissues extracted from ENCODE database. **B)** RNA expression of FAM30A normalized in terms of Z-score across a variety of cancer tissues extracted from TCGA database.

As Figure 9A shows, FAM30A expression was the highest in spleen and to a lesser extent in the thymus, two organs which are part of the haemolymphatic system. Overall, there was low RNA expression in other tissues thus indicating a putative immune cell-specific function of this lncRNA. Interestingly, FAM30A expression levels were also

relatively high in colon tissue which is further delineated in the discussion section. On the other hand, across cancer types, Figure 9B shows the highest levels of FAM30A expression in AML samples (LAML) to a big extent when compared to other cancer types extracted from TCGA. In the other two tumour tissues where FAM30A expression was high to some extent were Diffuse large B-cell lymphoma (DLBC, a type of lymphoid neoplasm) and in pheochromocytoma and paraganglioma samples (PCPG, neuroendocrine tumors).

3.1.2.- FAM30A expression correlates with more aggressive karyotypes and negatively impacts overall survival in AML patients.

As shown in Figure 9, FAM30A expression levels were substantially high in the immune system and especially in lymphatic organs that contain mature immune cells (spleen and thymus). These lineage-committed cells derived from primitive haematopoietic progenitors residing in the bone marrow (HSCs) with the capacity of long-term self-renewal (Aguila et al., 1997). As other groups have previously suggested the potential relevance of FAM30A in AML LSCs, a comparison was conducted to examine the distribution of FAM30A expression in AML and other haematological malignancies versus healthy bone marrow. This aimed to confirm its clinical significance in this context (Figure 10A). Unfortunately, the deepBasev3.0 platform does not include other leukaemia subtypes and myeloproliferative diseases in order to analyze if FAM30A expression is restricted to AML patients only.

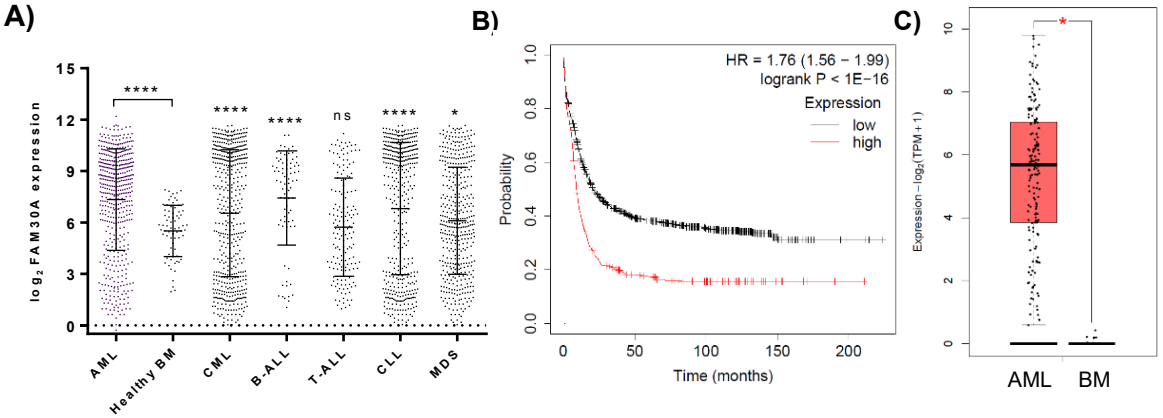


Figure 10. High FAM30A expression negatively impacts overall survival of AML patients. **A)** Representative graph with FAM30A RNA expression data in leukaemia subtypes extracted from MILE study from acute myeloid leukaemia (AML) patients (n= 516), healthy bone marrow (BM) (n= 70), chronic myeloid leukaemia (CML) (n = 497), B-cell acute lymphoblastic leukaemia (B-ALL) (n= 80), T-cell acute lymphoblastic leukaemia (n = 165), chronic lymphoblastic leukaemia (CLL) (n = 428) and myelodysplastic syndrome (MDS) (n = 426). **B)** Kaplan-Meier analysis of overall survival data plotted using the Kaplan-Meier plotter using the cut-off of median FAM30A expression for separating high and low FAM30A expression groups. **C)** RNA expression of FAM30A in AML primary samples shown in red (TCGA, n=173) compared to bone marrow (n=70) expressed in log_2 TPM (transcripts per million) from GEPIA.2 software.
-Continued on next page-

Statistical significance was determined by unpaired two-tailed student t-test in (A,C) and and by Log-rank test (Mantel-Cox) in (B). (****- p<0,0001; *** - p<0,001; **- p<0,01; *- p<0,05). Data is displayed as mean values and error bars represent SEM. ns = not significant. HR= Hazard ratio. AML= acute myeloid leukaemia. BM= bone marrow.

Therefore, gene expression data was extracted from the MILE study (MIcroarray Innovations in Leukemia, n= 2096 patients) that comprises gene expression data from accurately-diagnosed 16 leukaemia classes, MDS and non-leukaemia/normal bone marrow (BM) (Haferlach et al., 2010). This diagnostic dataset is extremely robust in clinics as it also includes different subclasses of AML with karyotypes. When compared to BM from healthy donors, FAM30A expression was significantly higher in AML samples (*P* = 4.93E-14) than other leukaemia types (Figure 10A). The other highly significant leukaemia types were chronic myeloid leukaemia (CML, *P* = 3.56E-05), B-cell acute lymphoblastic leukaemia (B-ALL, *P* = 3.09E-07) and chronic lymphoblastic leukaemia (CLL, *P* = 8.19E-07). Then, the correlation between FAM30A expression and clinical outcome was analyzed. The AML patients (n=1608) were divided into two groups by the median value of the FAM30A/KIAA0125 expression. As Figure 10B shows, high FAM30A expression was significantly correlated with worse overall survival. AML patients with higher FAM30A expression showed an overall survival of 9.1 months and patients with lower FAM30A expression 20.7 months (*P* = 1E-16, hazard ratio 1.76). In addition, FAM30A expression was significantly higher in AML primary samples compared to BM from healthy donours (Figure 10C).

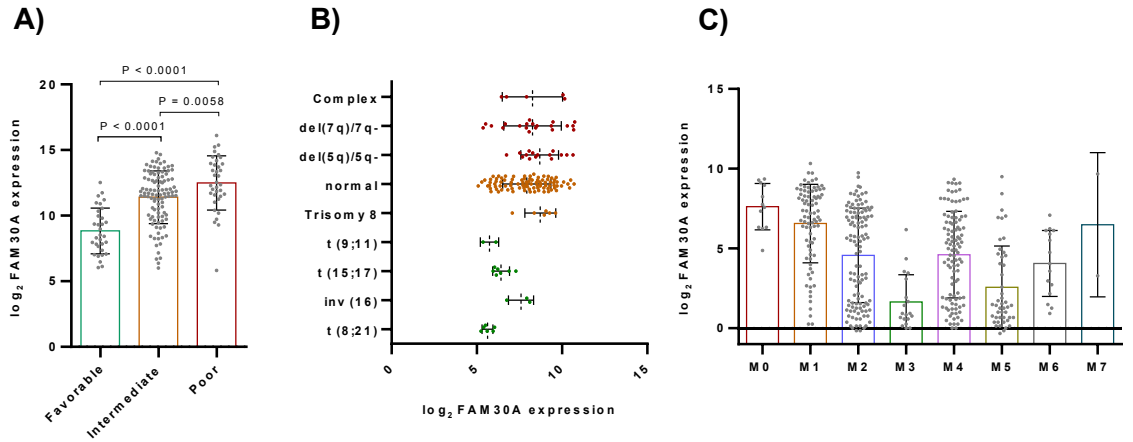


Figure 11. FAM30A expression correlates with aggressive AML karyotypes. A) FAM30A expression data calculated was extracted from cBIO-portal (TCGA) and each karyotype was assigned to be favorable (n=32), intermediate (n=105) or poor (n=36) prognosis according to current guidelines. **B)** FAM30A expression data from A) assigned to each of the karyotypes cytogenetically tested in the clinical study from TCGA. **C)** FAM30A expression data extracted from the R2 platform (Bohlander, n=422) with AML subtypes assigned according to the French-American-British (FAB) classification. Statistical significance was calculated using unpaired two-tailed student t-test. Data is displayed as mean values and error bars represent SEM.

Next, a comparison was made to determine if certain genetic alterations found in AML patients correlated with FAM30A expression in order to assess the clinical relevance of FAM30A in specific mutational signatures (Figure 11). Data from the cBioPortal (TCGA) were retrieved, as it includes detailed descriptions of the mutational status of each AML patient. AML karyotypes were then categorized into favourable, intermediate, and poor prognosis based on current disease classifications (Döhner et al., 2015) and it emerged that FAM30A expression was significantly higher in AML patients with worse prognosis compared to favourable prognosis (P = 3.8E-11). Unfavourable AML karyotypes such as complex aberrant karyotype (with multiple unrelated mutations), monosomal karyotypes del(5q) and del(7q) associate with rapid disease progression and refractory disease upon standard treatment regimens (Breems et al., 2008; Stölzel et al., 2016). On the other hand, FAM30A expression was considerably lower in patients with better prognosis and more differentiated and chemo-sensitive karyotypes such as t (8;21), inv (16) and t (15;17), as previously shown (Figure 11A and B) (Y. H. Wang et al., 2021). Furthermore, an analysis was conducted to determine if well-known mutations in AML oncogene drivers correlate with FAM30A expression (Supplemental Figure 1). AML patients with FLT3-ITD (present in 30% of all AML cases) and high EVI1 expression (8-10% cases) showed significantly higher FAM30A expression (P < 0.0001). Both mutations are features of aggressive leukaemia and high failure rates to achieve complete remission after chemotherapy treatments (DiNardo et al., 2020). Interestingly, FAM30A expression was also significantly correlated with NPM1+ mutations (25-30% cases of all AML patients), which is associated with better prognosis in AML patients.

From another perspective, FAM30A expression was also analyzed across different AML subtypes based on the widely used French-American-British (FAB) classification (Figure 11C). The FAB system defines the various subtypes depending on the cell type that the leukaemia developed from and its degree of maturity. Despite this classification is slightly outdated, available AML datasets still stratify patients using this system. Moreover, this classification is still used as the model to better assign cytomorphology, and myeloid features of the AML cell lines (Skopek et al., 2023). Among the different subtypes, the highest expression values were seen especially in the M0 subtype (acute myeloblastic leukaemia, minimally differentiated and 2-3% of cases) and in the M1 subtype (acute myeloblastic leukemia with minimal maturation and 15% of cases) (Vakiti & Mewawalla, 2023). In both M0 and M1 subtypes the differentiation blockade occurs at early myeloid stages thus explaining their highly undifferentiated states.

3.1.3.- FAM30A expression associates with AML LSC-enriched features

In most AML cases, LSCs are considered to reside within the CD34+/CD38− cell fraction even though depending on the nature and developmental stage of the disease, some LSCs may also reside in CD34+/CD38+ subsets or even in CD34− subfractions (Hope et al., 2004a). Considering that FAM30A has been previously reported in LSC signatures, an existing LSC gene signature was curated to sort for different AML LSC fractions, aiming to investigate if FAM30A expression is specific for certain AML LSC subpopulations. Eppert et *al.* thoroughly investigated different AML LSC fractions and compared mRNA expression profiles shared between functionally validated LSCs and HSCs fractions to generate a gene signature (CE_HSC_LSC) that includes highly significant independent predictors of patient survival which are enriched for “stemness” (Eppert et al., 2011). FAM30A expression data was extracted from this study and was first compared between functionally validated LSC and non-LSC fractions showing that FAM30A is significantly higher in LSCs, as it was previously reported (Figure 12A).

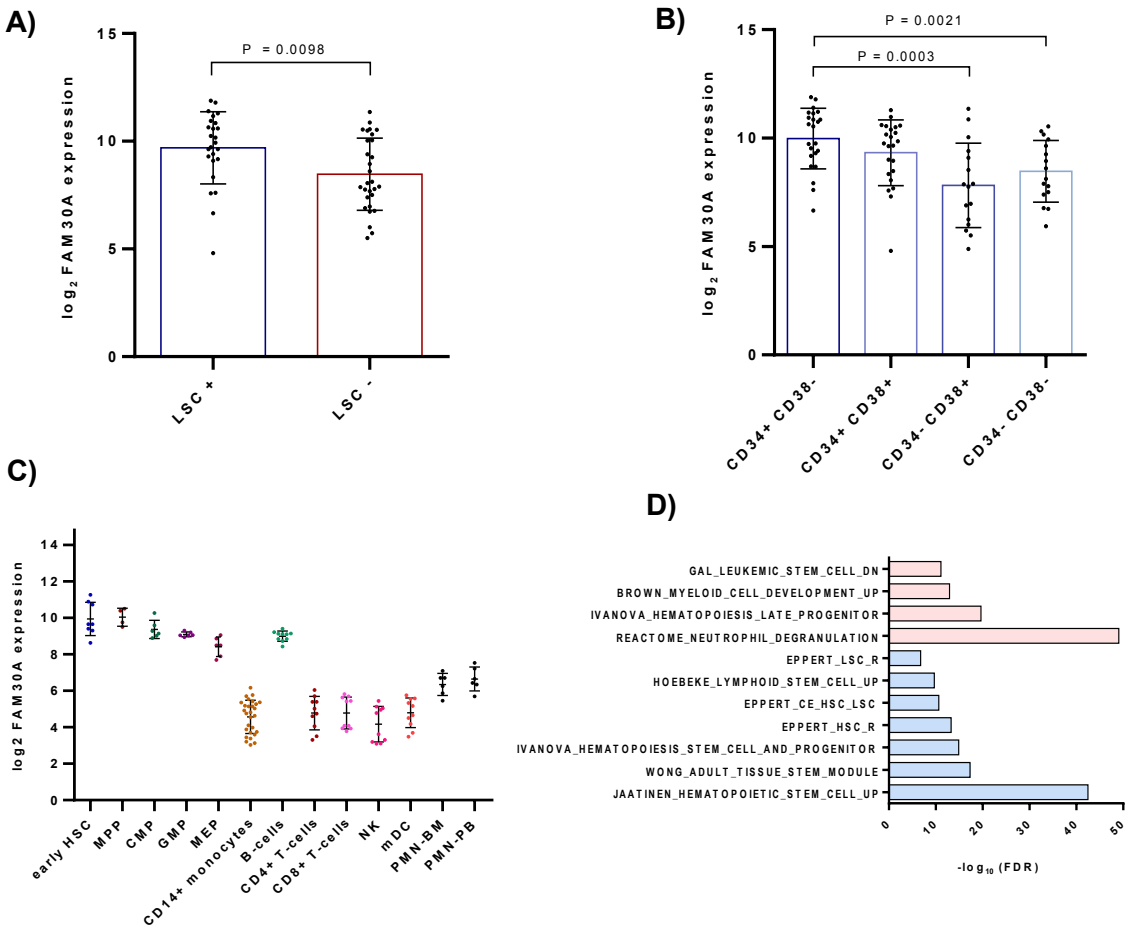


Figure 12. FAM30A is highly expressed in validated AML LSCs residing in the CD34+/CD38- fraction and correlates with LSC signatures. A) FAM30A expression values extracted from R2 platform from validated LSC (n=29) and non-LSC fractions (n=25) (Eppert et al., 2011).
-Continued on next page-

B) FAM30A expression data derived from validated LSC fractions with varying surface expression of CD34 and CD38 markers. **C)** FAM30A expression values extracted from (Rapin et al., 2014) in different haematopoietic cell lineages.**D)** Representative graph of GSEA analysis (C2 geneset) with the most significantly (cut-off $P < 0.0001$, Spearman correlation > 0.3) negatively correlated (red) and positively correlated (blue) genes with FAM30A. Correlation values were extracted from cBIO-portal (TCGA). Early HSC: early Hematopoietic Stem Cell, MPP: Multi-potent Progenitor, CMP: Common Myeloid Progenitor, GMP: Granulocyte/Monocyte Progenitor, MEP: Megakaryocyte/Erythroid Progenitor, NK: Natural Killer cell, mDC: myeloid Denditric Cells, PMN-BM: Polymorphonuclear cells from Bone Marrow, PMN–PB: Polymorphonuclear cells from Peripheral Blood. Statistical significance was calculated using unpaired two-tailed student t-test. Data is displayed as mean values and error bars represent SEM.

FAM30A expression levels were then analyzed in functionally validated LSC fractions (i.e. via xenotransplant assays) with varying expression of CD34 and CD38. Figure 12B shows that FAM30A expression is the highest in the CD34+/CD38- fraction most AML LSCs reside (Bonnet & Dick, 1997). FAM30A was significantly more expressed in CD34+CD38- when compared to CD34-CD38+ and CD34- CD38- fractions (4.49 and 2.85 times more, respectively). Additionally, the FAM30A expression pattern in normal haematopoietic lineages was investigated to determine whether FAM30A might potentially also play a role in HSCs (Figure 12C). Interestingly, FAM30A expression levels were also the highest in early HSCs and early myeloid-committed progenitors cells like MPPs, CMPs and heavily decreased along the differentiation process (FAM30A is expressed 2.4 and 30.7 times more in HSCs compared to CD14+ monocytes and dendritic cells, respectively). The only exception for the relatively high expression of FAM30A was in B-cells and the possible explanation will be detailed shortly after this section.

In support of this, GSEA analysis also indicated that the vast majority of leading-edge genes whose expression levels were the most significantly correlated with FAM30A are enriched for normal and LSC signatures (Figure 12D). On the other hand, the most negatively correlated genes are associated with more myeloid-related differentiation and down-regulation of LSC signatures suggesting a key role of FAM30A in the myeloid compartment.

3.2 Basic characterization of FAM30A

3.2.1.- Genomic localization of FAM30A

FAM30A (also known as KIAA0125 and less commonly annotated as HSPC053 or C14orf110) is a transcript with an associated gene type of non-coding RNA. It is located at the end of the long arm of the human chromosome 14 (14q32.33) in a region previously reported to be correlated with predisposition of myeloproliferative neoplasms (Di Giacomo et al., 2021; Venditti et al., 1997). Moreover, FAM30A gene is also

embedded in antisense orientation to the locus of Immunoglobulin heavy chain (variable region) (IGH) (Figure 13A).

The close proximity of the FAM30A gene to the IGH locus may elucidate the elevated expression levels in B-cells. This is because the IGH locus harbours genes crucial for antibody production, correlating with earlier research indicating significant upregulation of FAM30A post-vaccination (de Lima et al., 2019). Even though the length of the whole FAM30A transcript is 14664 nucleotides, the annotated and most validated processed transcript is 9643 base pairs (RefSeq NCBI: NR_026800 and Ensembl ID: ENST00000630242.2).

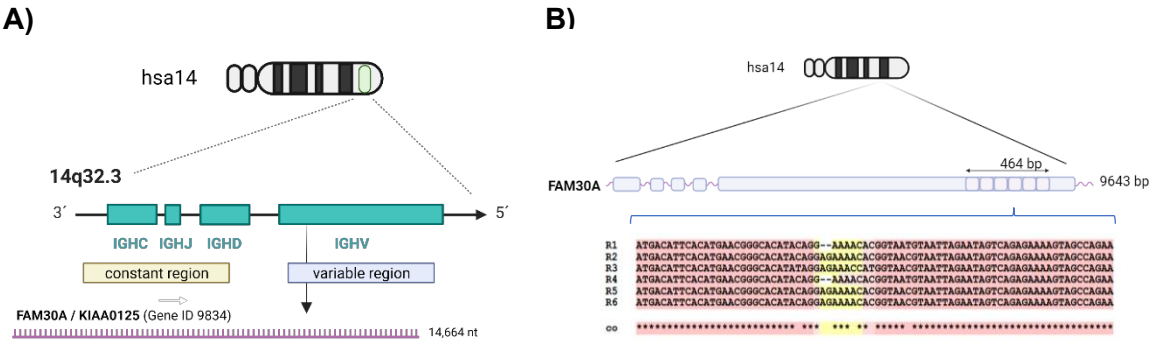


Figure 13. FAM30A transcript lies within a genomic region with predisposition for onset of myeloid leukaemias and its sequence comprises tandemly repeated motifs. A) The FAM30A gene within the IGH locus and length of the unprocessed RNA transcript based on the NCBI database (GRCh38.p14). **B)** Depiction of the processed and annotated FAM30A transcript (NR_026800.2) with a highlight on the region with the sequences repeated in tandem after multiple sequence alignment in T-Coffee software. Highlighted in red are conserved nucleotides (highest consistency score among the sequences) and in yellow are less conserved nucleotides.

In contrast to the six annotated exons depicted in the UCSC browser (using the prediction tool FANTOM5 for prediction of Transcription Starting Sites), RNA-seq data from CD34+ cord-blood cells was analyzed, where FAM30A is highly expressed and reads starting from exon 2 were detected. This makes a total of five expressed exons in haematopoietic cells where FAM30A will be further studied (Supplemental Figure 2).

Next, the RNA sequence of FAM30A was broadly analyzed. An interesting region was identified at the end of the long exon 5. This region is comprised of six almost identical sequences (two sequences of 76 bp and four sequences 78 bp) repeated in tandem (a total of 464 bp). From another perspective, these tandems are organized into two larger repeats, each composed of one 76-bp segment followed by two identical 78-bp repeats. Following multiple sequence alignment of the six sequences using T-Coffee software (Wallace et al., 2006), Figure 13B demonstrates the high similarity scores observed among them.

3.2.2.- Screening of FAM30A expression in AML cell lines and experimental approach for genetic studies

Considering the complexity and variety of AML karyotypes, it is particularly challenging to use a single cell line model that completely resembles the disease. In order to characterize FAM30A at the cellular and molecular level in an AML LSC-like cell line model, expression from genes included in the LSC17 signature (Ng et al., 2016) was measured in four available AML cell lines: MV4-11, MOLM-13, KG-1a and GDM-1. Two of them (MV4-11 and MOLM-13) are more differentiated AML blast-like type of cells and the other two resemble more AML LSC features (KG-1a and GDM-1). As expected, GDM-1 and KG-1a cell lines showed the best LSC17 gene expression profiles (Figure 14A).

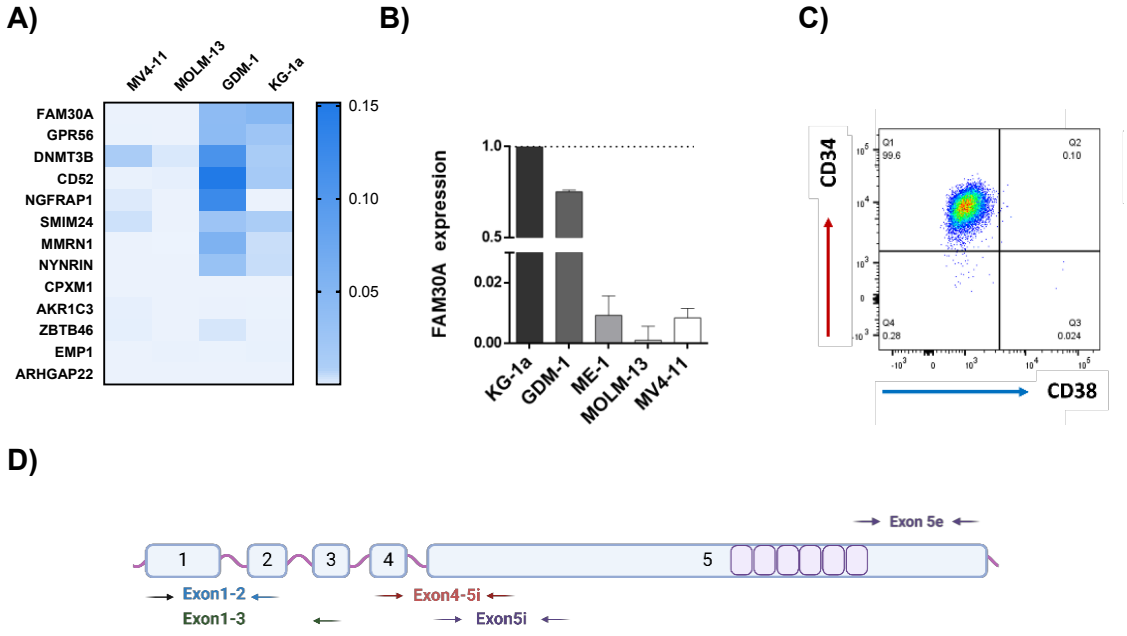


Figure 14. FAM30A expression in AML cell lines and experimental design used to study the biological functions of FAM30A. A) Heatmap displaying RNA expression data (expressed in dCT normalised to ACTB) of LSC17 genes calculated from qRT-PCR analysis in four AML cell lines. Only 13 genes are presented as the other 4 showed technical difficulties due to very low expression values in all cell lines. **B)** Relative FAM30A RNA expression values (normalized to ACTB and to KG-1a cells) derived from RT-qPCR in AML cell lines. **C)** FACS analysis of surface markers CD34 and CD38 in KG-1a cell line. **D)** Representation of the location of the different primers used for detection of FAM30A. Exon 5i = beginning of exon 5, Exon 5e = end of exon 5. (end) = endogenous total FAM30A detection, (ex) = exogenous repeats and ACTB as a housekeeper gene are depicted. Data is displayed as mean values and error bars represent SEM.

The KG-1a cell line was selected as the cell line model for further characterization studies of FAM30A due to several reasons. It showed the highest levels of FAM30A expression (Figure 14B), it is also defined as an M0/M1 AML subtype displaying very immature leukaemic stem cell features and commonly used as an AML LSC-like cell line (She et al., 2012; Siveen et al., 2017). Moreover, these cells do not spontaneously differentiate into granulocytes or macrophages, remarkably keeping its CD34+/CD38- immunophenotype (Figure 14C). Moreover, KG-1a cells also display high chemoresistance properties with a very complex karyotype and aggressive phenotype with several structural aberrations including NRAS and TP53 mutations (Matou-Nasri et al., 2022; Mrózek et al., 2003).

MV4-11 was also selected as an AML cell line that does not express FAM30A to study how exogenous expression of FAM30A affects its phenotypic features and chemoresistance.

3.2.3.- Identification of a highly conserved region within the FAM30A sequence

As mentioned above, NCBI and Ensembl databases displayed different expression patterns for the annotated exons in FAM30A when it was also compared to sequencing data in normal HSCs (Supplemental Figure 2). For that reason, the exon usage pattern of the expressed exons was determined to clarify which exons might be constitutively expressed in AML cells (Figure 15). As qRT-PCR analysis in Figure 15A shows, the particularly long exon 5 is the most constitutively expressed exon in AML cells (8.89 kb in length and accounting 94% of the total transcript).

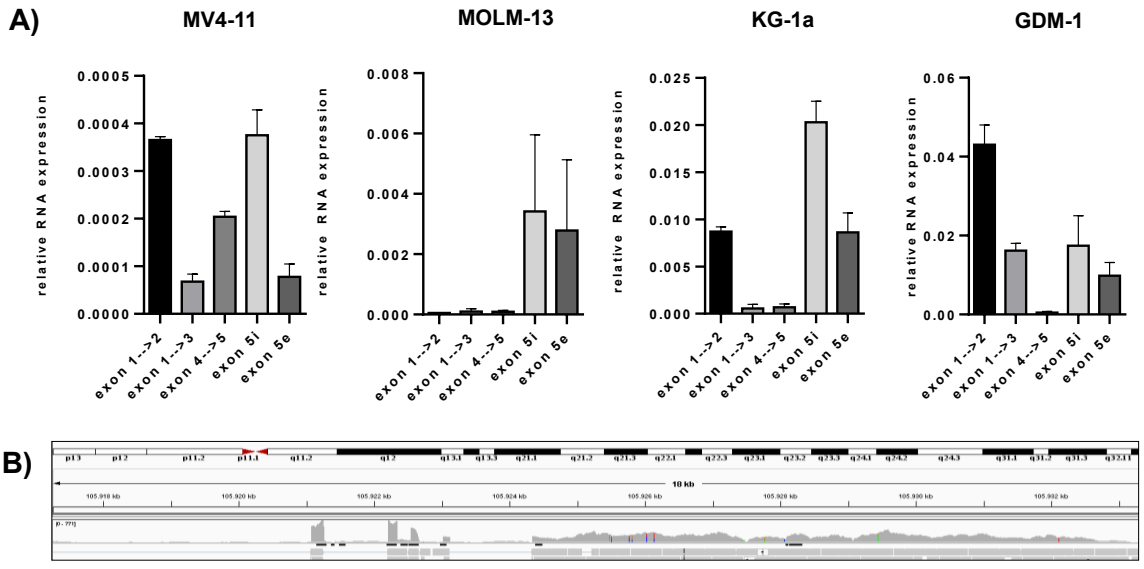


Figure 15. Determination of FAM30A exon usage in AML cell lines. A) qRT-PCR analysis of of FAM30A exons (dCT normalised to ACTB) using the five oligonucleotide combination described

-Continued on next page-

in Figure 14D. **B)** Read distribution of FAM30A from RNA-sequencing analysis displayed using IGV software. Data is displayed as mean values and error bars represent SEM.

In particular, KG-1a cells displayed the highest expression of exon 5. Among the other exons, exon 1 (196 bp) expression was relatively high and exons 2 (146bp), 3 (103 bp) and 4 (108 bp) did not show consistent expression levels among the cell lines. RNA-sequencing analysis performed in KG-1a cells also confirmed that all five described exons were clearly expressed (Figure 15B).

Exon 5 appeared to have a relevant function due to its constitutive expression across AML cell lines, with an especially high expression in KG-1a cells. Therefore, the RNA sequence of exon 5 was analyzed using RegRNA2.0 software to identify regulatory RNA elements such as conserved motifs for RNA-binding proteins. As previously described in Figure 13B, an interesting region within exon 5, including six highly similar sequences repeated in tandem, was identified. Although predicted RBP motifs were discovered sporadically along exon 5 (data not shown), most of them were located within each repeat, accounting for several putative RBP motifs (at least six) in each FAM30A molecule.

Due to the intriguing pattern of these tandemly repeated sequences in FAM30A, the evolutionary conservation of this repeat across vertebrates was investigated (Figure 16). The presence of FAM30A homologs was only detected in primates. Then, genome sequences of primates which are representative of different clades were downloaded and subjected to alignment. Figure 16A shows detection of the repeat (76 and/or 78 bp) in the different taxonomic families of primates. The presence of one repeat was detected in the family of higher primates (also known as Anthropoidea or Simiiformes) and an increased number of repeats was observed in the Hominoidea clade where chimpanzees (Pan troglodytes), gorillas (Gorilla gorilla) and humans (Homo sapiens) are located. The latter showed the highest number of repeats (six repeats) among all investigated species.

This posed the question if there was conservation of only the repeat and/or other parts of the RNA sequence. Thus, multiple sequence alignment was performed among primates for the sequences flanking the repeat region (500 bp downstream and upstream), including one repeat (complete alignment shown in Supplemental Figure 3). Interestingly, the region where the repeat begins shows the highest conservation score (Figure 16B). This underscores the relevance of the FAM30A sequence and prompted an in-depth examination of the molecular role of this repeat sequence, given the high evolutionary pressure to increase the number of repeats throughout the evolutionary process.

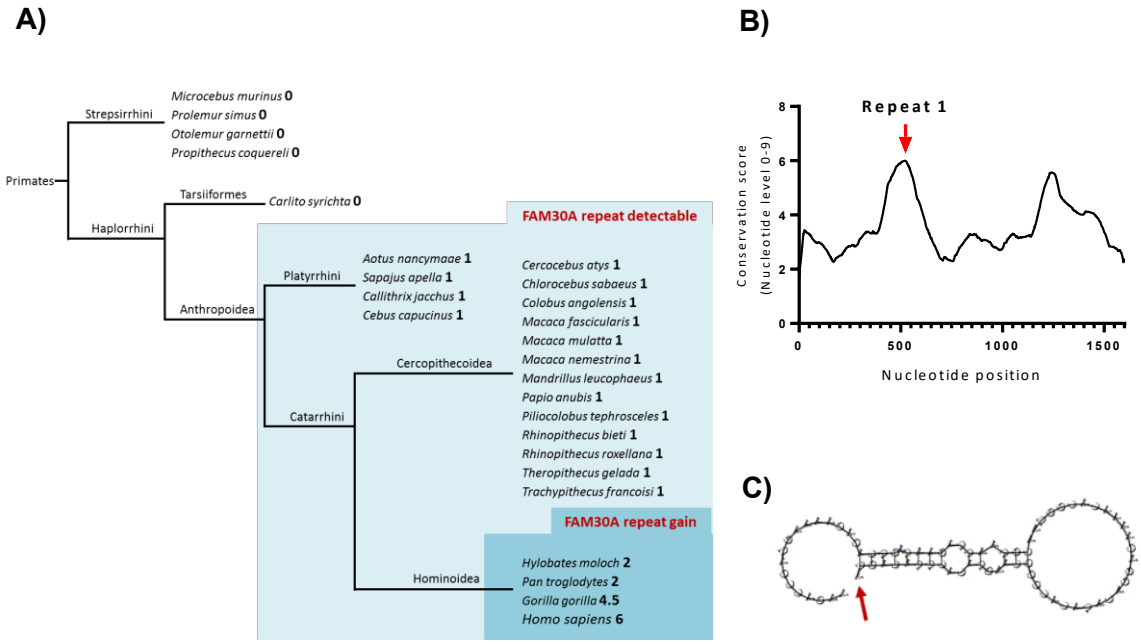


Figure 16. Analysis of FAM30A repeats sequences among primates. A) Representative taxonomic tree of primates with the species used for each clade. Highlighted in light blue is where at least one repeat was detected among the genome sequences and dark blue is where there more than one repeat detected. **B)** Conservation analysis depicted in terms of conservation score extracted from multiple sequence alignment (M-Coffee software) of a sequence range (500 bp upstream and downstream the repeat). For comparison, five primate genomes were extracted from NCBI and Ensembl databases. Statistical analysis was performed using locally weighted regression of curve by Lowess algorithm. Highlighted in red is the beginning of the Repeat 1 in the FAM30A sequence (see Figure 13B). **C)** Predicted secondary structure of the RNA sequence of one repeat (78bp) using RNAfold web server. The 5´end is highlighted with a red arrow.

3.2.4.- FAM30A is a predominantly cytoplasmic lncRNA

To gain insights into the functional role of FAM30A, its subcellular localization was initially studied. A digitonin-based cellular fractionation was performed extracting nuclear and cytoplasmic fractions. From these fractions, the abundance of FAM30A RNA in these subcellular compartments in KG-1a cells was measured (Figure 17A). ACTB and EEF2 mRNAs were used as control cytoplasmic RNAs and U6 and NEAT1 as control nuclear RNAs, respectively. In line with their protein-coding function, ACTB and EEF2 essentially located in the cytoplasm (more than 9%) whereas NEAT1 and U6 were more abundant in the nuclear fraction of the cells as they elicit nuclear-restricted RNA functions (Clemson et al., 2009). Surprisingly, FAM30A was predominantly located in the cytoplasm (91% in the cytoplasm and 9% in the nuclear fraction). Early studies indicated that most of the lncRNAs are generally nuclear and localized to chromatin (Derrien et al., 2012; Djebali et al., 2012) even though it is now becoming increasingly evident that many newly identified lncRNAs are exported to the cytoplasm where they exert their functions associated with proteins (Carlevaro-Fita et al., 2016).

In principle, cytoplasmic lncRNAs would have an increased chance of association with translation complexes and ultimately being translated. Therefore, even though FAM30A is predicted to be non-coding, its protein-coding potential was assessed using sucrose gradient-based fractionation, followed by targeted polysome profiling in KG-1a cells (Figure 17B).

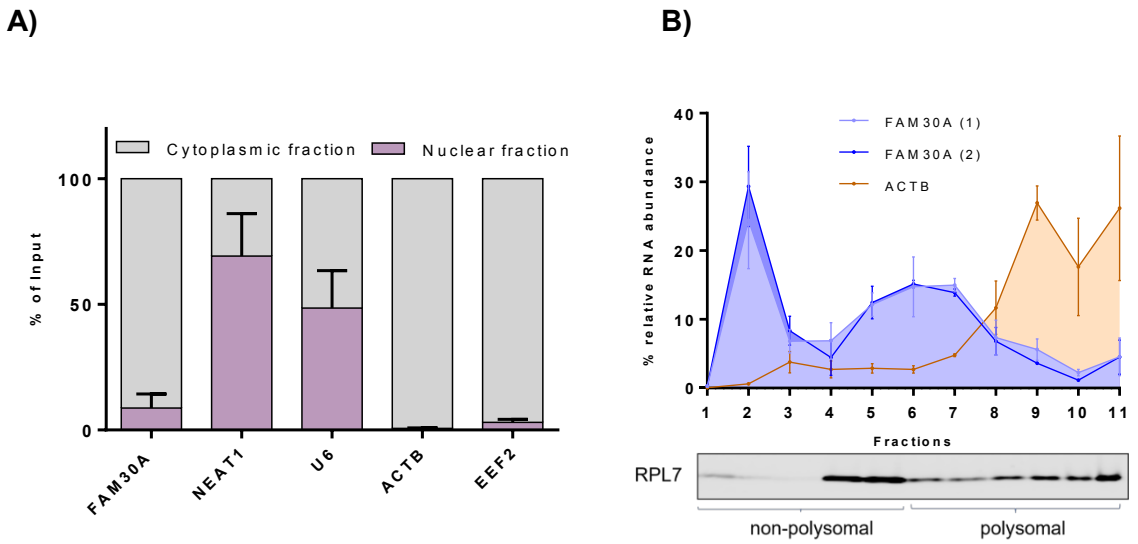


Figure 17. FAM30A is a cytoplasmic lncRNA. A) Relative RNA abundances in subcellular fractions (nuclear and cytoplasm) after digitonin treatment in KG-1a cells (n=3). NEAT1 and U6 were used as control nuclear RNAs and ACTB and EEF2 were used as control cytoplasm mRNAs. **B)** Targeted profiling of FAM30A and ACTB in KG-1a cells (n=3) in non-polysome associated (1-6) and polysomal fractions (7-11) isolated after sucrose-based gradient centrifugations. RPL7 is presented as a loading control for identification of the fractions. FAM30A (1) and FAM30A (2) represent different primers used for detection of FAM30A: Exon1-2 and Exon5i, respectively. Data is displayed as mean values and error bars represent SEM.

This method provides valuable information about the translational status of a particular RNA depending on its association with ribosomal complexes (heavy fractions) and it might also identify RNAs that are not translated (Panda et al., 2017; Tanese, 1997). The abundances of FAM30A and ACTB mRNA were analyzed on the isolated fractions of the gradient. Moreover, RNA abundances were compared to protein distribution of RPL7, indicating the different ribosomal complexes. Based on RPL7 distribution, the fractions were classified as follows: small complexes (1st to 4th fractions), monosomal (4th to 6th fraction) and polysomal fractions (7 th to 11th fraction). As expected, ACTB was enriched in the low and high molecular weight polysomal fractions (87% of the total abundance). On the other hand, FAM30A was enriched in the small complex and monosomal fractions (70% of the total). This indicates that FAM30A does not associate strongly with the translational machinery in polysomes and thereby it is likely to be not efficiently translated

at steady-state conditions in KG-1a cells. This experiment was also performed in another AML LSC-like cell line, GDM-1, and the results were largely consistent (data not shown).

In addition, the coding potential of FAM30A was also analyzed using coding potential scores from the established online tools phyloCSF (Lin et al., 2011) and the Coding Potential Calculator (CPC2.0) (Kong et al., 2007). Results show that the coding probability of FAM30A is relatively low when compared to known mRNAs (Supplemental Figure 4).

3.3. Cellular function of FAM30A

3.3.1.- FAM30A enhances AML progression and LSC-enriched gene signatures

Previous *in cellulo* functional studies have only indicated the cellular function of FAM30A as a tumour-suppresor in colorectal cancer (Yang, Zhao, et al., 2019) and inconclusive preliminary data suggested an oncogenic role of FAM30A in AML (T. Zhang et al., 2023). However, there is so far no clear molecular function of FAM30A acting as an oncogene, as previously suggested by studies correlating high FAM30A expression levels with worse clinical outcome. In order to decipher FAM30A´s leukaemogenic capacity in AML LSC-like cells, stable knockdowns (via shRNAs) in combination with exogenous expression of the repeats were performed in KG-1a cells. In this cell line, four conditions were phenotypically analysed throughout this thesis: control cells (CTRL or C), overexpression (OE; of the conserved repeat region of FAM30A), endogenous FAM30A knockdown (KD) and recovery (REC; KD plus OE). Unfortunately, exogenous expression of the complete FAM30A transcript (almost 10kb) was troublesome due to the difficulty to overexpress such a long transcript.

Also, the lack of information about FAM30A promoters made it difficult to design CRISPR-activation-based systems in order to endogenously activate transcription of FAM30A. Therefore, the decision was made to exogenously overexpress only the region with the six highly similar sequences repeated in tandem (hereinafter referred to as repeats or FAM30A repeats) as previously described (Figure 13B). The rationale for this experimental setup was to address the effects of depleting total endogenous FAM30A and to evaluate if certain stem cell-associated phenotypes could be rescued solely through exogenous overexpression of FAM30A repeats. Figure 18A shows the experimental genetic approach utilized in this thesis for gene silencing (using shRNAs) and overexpression constructs as well as the oligonucleotide combinations for detection of endogenous/exogenous FAM30A repeats.

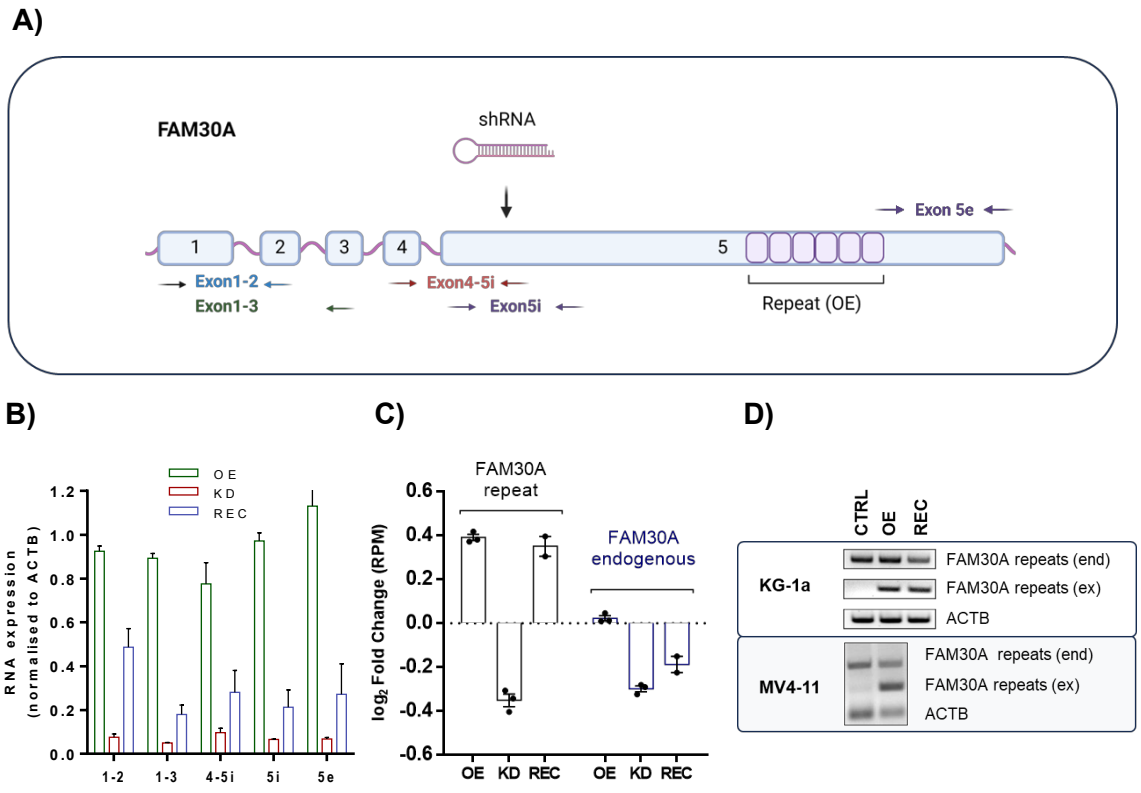


Figure 18. Quantification of RNA expression levels of endogenous FAM30A and its repeats in the stable cell lines used for RNA-seq analysis. A) Schematic depiction displaying the region where the shRNA targets FAM30A and the highlighted repeat region used for the overexpression construct. **B)** RT-qPCR analysis using oligonucleotides targeting different exons of FAM30A in the stable cell lines used. ACTB mRNA and CTRL cells were used for normalization for each condition. **C)** Quantification derived from RNA-sequencing performed in KG-1a stable cell lines. Data is expressed in terms of log_2 Fold Change of the reads per million (RPM) of FAM30A repeat and the flanking region before the repeat (FAM30A endogenous), normalized to shC EV. **D)** Semi-quantitative PCR showing endogenous FAM30A repeats (FAM30A repeats end) and exogenous repeat levels (FAM30A repeats ex) of CTRL, OE, and REC cells in KG-1a cells (higher panel) and CTRL and OE MV4-11 cells (lower panel) cell lines. ACTB was used as a loading control. Data is displayed as mean values and error bars represent SEM.

Quantification of the levels of exogenous repeats and endogenous FAM30A in the four conditions was also performed. RT-qPCR analysis showed that RNA expression levels of the different exons of FAM30A were uniformly downregulated upon FAM30A knockdown (KD) and slightly upregulated upon overexpression of the repeats (REC) (Figure 18B). This was also reflected after quantitatively assesing the raw read counts of endogenous FAM30A and the overexpressed repeats extracted from analysing the RNA-sequencing data (Figure 18C) and semi-quantitative PCR (Figure 18D).

First, the impact of FAM30A modulation on AML progression was assesed by a Gene Set Enrichment analysis (GSEA) meta-examination of gene sets correlated with AML signatures (Figure 19). Figure 19A depicts that knockdown of endogenous FAM30A

(KDvsC) showed a significant downregulation of genes (negative NES) involved in canonical AML pathways in both adult and paediatric AML.

In addition, overexpression of the repeats (OE vs C, REC vs KD) significantly correlated with upregulation of these oncogenic pathways (positive NES) (J. Chen et al., 2015). More importantly, among gene signatures affected by modulation of FAM30A expression, gene sets comprising AML LSC-specific gene signatures (EPPERT_LSC_R and EPPERT_CE_HSC_LSC, FDR < 0.001) were significantly enriched (Figure 19B). EPPERT_LSC_R gene signature comprises significantly upregulated genes in functionally defined AML LSCs and EPPERT_CE_HSC_LSC includes genes shared with normal HSCs which are highly expressed in AML LSCs versus their blasts counterparts (Eppert et al., 2011). This confirms the previous hypothesis that FAM30A, and more specifically its repeats, could be regulating leukemic stem cell characteristics.

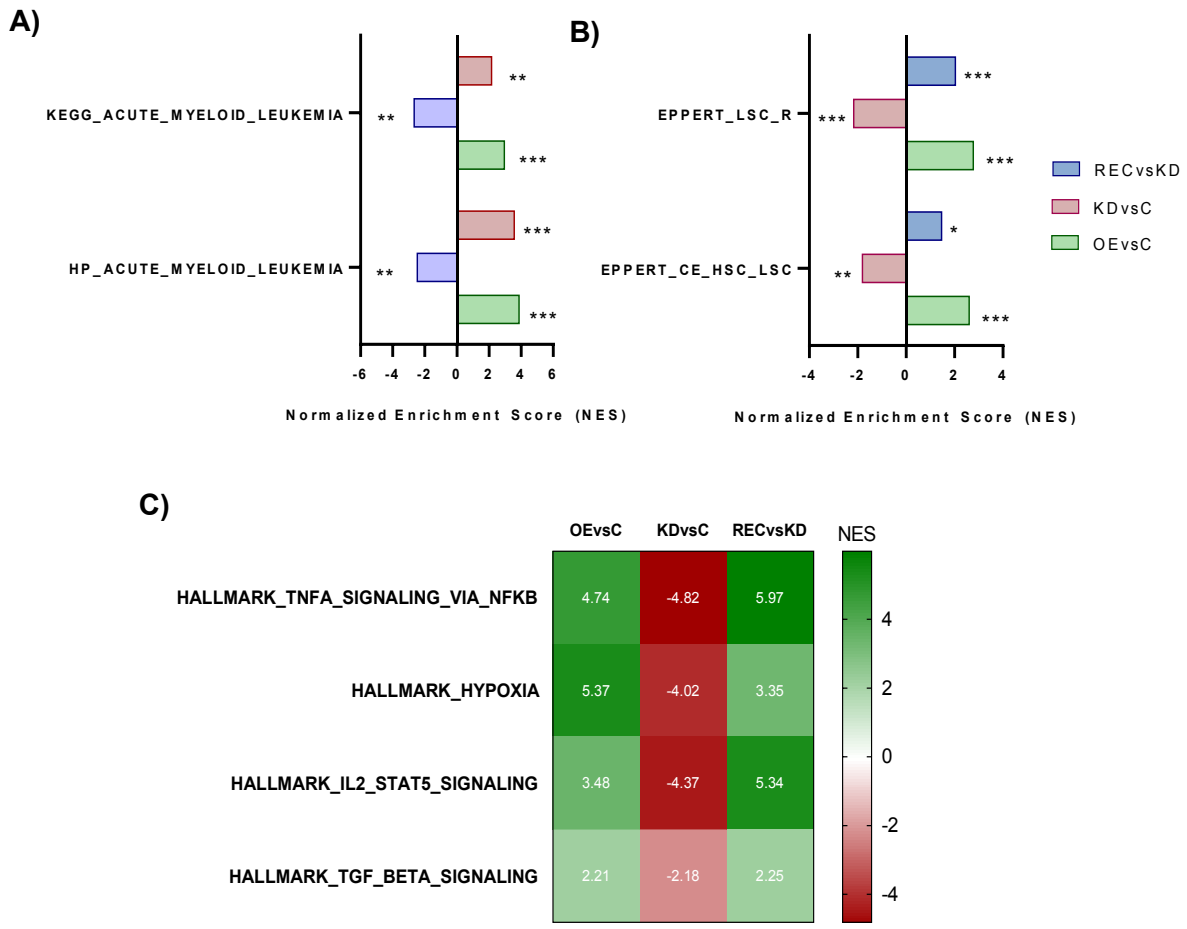


Figure 19. Transcriptomic analysis of AML-related signatures upon modulation of FAM30A expression. A) GSEA analysis (C2 dataset) performed in KG-1a stable cell lines showing enrichment of AML pathways. **B)** GSEA analysis showing enrichment of AML LSC-specific signatures in KG-1a stable cell lines.
-Continued on next page-

C) Heatmap representing the most significantly enriched pathways (HALLMARKS) between samples from GSEA analysis based on NES with a cut-off FDR<0.001 and NES>2. NES = Normalized Enrichment Score. * p <0.05, ** p <0.01, *** p <0.001.

This finding was also supported by GSEA analysis of transcriptional pathways (HALLMARKS) that shows significant enrichment (FDR<0.001) of pro-survival and maintenance pathways in AML LSCs such as TNF-α/NF-kB, hypoxia, IL2/STAT5 and TGF-β signaling pathways (Figure 19C) (Bruno et al., 2021; Guzman et al., 2001; Kagoya et al., 2014; Schepers et al., 2007; Tabe et al., 2013; Wingelhofer et al., 2018). Additionally, crucial receptors required to activate these pro-LSC pathways are also significantly correlated (FDR<0.01) with changes in expression of FAM30A and its repeats. For instance, PTCH1 (receptor involved in Hedgehog pathway) and CD81 (involved in leukaemia stemness) were significantly downregulated upon FAM30A knockdown (log_2FC= -2.07 and -1.89) and rescued upon ovexpression of FAM30A repeats (log_2FC= 0.43 and 0.55, respectively).

Altogether, these results suggest that transcriptional pathways and gene signatures that define and govern LSCs characteristics are downregulated upon FAM30A depletion and rescued by only exogenously expressing the repeats of FAM30A.

3.3.2.- Relevance of FAM30A and its highly conserved region in regulating stemness-related features in AML

Different phenotypic experiments were performed to assess the impact of FAM30A modulation on cell viability, resistance to hypoxic conditions, colony-formation potential and chemoresistance (Figure 20). Depletion of FAM30A (KD) greatly impacted cell viability in KG-1a cells (39% less compared to control) and overexpression of the repeats (OE, REC) increased cell viability when compared to their control counterparts (CTRL, KD). Overexpression of the repeats (OE) showed a 18% increase compared to control cells (OE vs CTRL) and a striking 193% increase in cell viability compared to FAM30A depletion (REC vs KD) (Figure 20A), indicating an pro-oncogenic effect of FAM30A and its repeats in this cell context. Analysis of cell cycle progression showed that overexpression of FAM30A repeats (OE, REC) significantly increased number of cells in G1 phase compared to their respective controls (CTRL, KD). In addition, depletion of FAM30A invoked a moderate but significant increase of cells in S phase which was recovered to control levels upon overexpression of the repeats (REC) (Figure 20B).

AML is initiated in the bone marrow where AML LSCs likely reside under local hypoxic niches where it competes with normal HSCs for similar microenvironmental support. Hypoxia-inducible Factor (HIF) pathway has been shown to be enhanced in AML LSCs when compared to healthy HSPCs which suggests that hypoxia contributes to AML LSC function and chemoresistance to maintain a quiescent state (Morrison & Scadden, 2014; Schito et al., 2017).

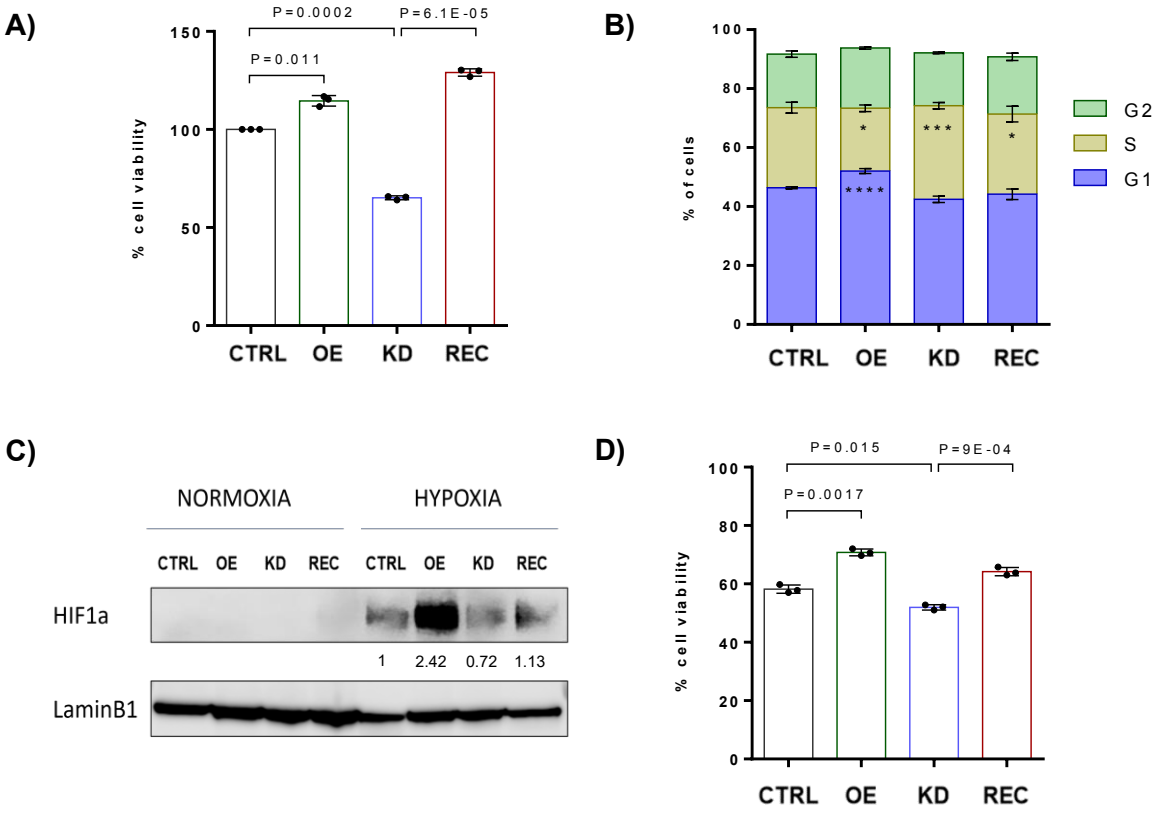


Figure 20. FAM30A affects cell viability and resistance to hypoxic conditions *in vitro* in AML LSC-like cells. A) Graph showing the percentage change of cell viability in KG-1a stable cell lines normalised to CTRL after 72 hours in liquid culture. **B)** Graph showing percentage of KG-1a stable cell lines in indicated cell cycle phases as determined by flow cytometry (n=6). **C)** Analysis of hypoxic conditions applied to KG1-a stable cell lines depicted by western blot analysis. HIF1a/HIF1α/HIF1alpha was used as hypoxia marker and LaminB1 as loading control under normoxic (21% O_2) and hypoxic (1% O_2) conditions. **D)** Graph showing percentage of cell viability after 72 hours in liquid culture under hypoxic conditions normalized to cells grown in normoxia (n=3). Significance was calculated using unpaired two-tailed unpaired student t-test and compared to CTRL for OE and KD for CTRL and REC (****- p<0,0001; *** - p<0,001; **- p<0,01; *- p<0,05). Data is displayed as mean values and error bars represent SEM.

As mentioned earlier, the hypoxia pathway is one of the most highly correlated pathways upon upregulation or downregulation of FAM30A and its repeats (Figure 19C). Therefore, the behaviour of KG-1a cells under hypoxic conditions was investigated upon modulation of FAM30A expression. HIF1α protein levels were assesed as indicator of the activation of the HIF pathway. Above 5% of oxygen (normoxia), the α subunit is rapidly degraded whereas under hypoxic conditions it is post-translationally stabilized and promotes

transcription of HIF target genes (Semenza, 2003). Figure 20C shows that upon hypoxic conditions, depletion of FAM30A (KD vs C) moderately affects HIF1α activation (28% less protein) and cell viability (*P* = 0.015) when compared to control cells. Interestingly, overexpression of the repeats (OE vs C, REC vs KD) greatly increases HIF1α activation compared to their respective control cells (2.42 and 1.57-fold more HIF1α protein). This translates into a significant increase in cell viability under hypoxia (13% and 8% more, respectively).

It is well established that AML LSCs represent the root of AML relapse as chemotherapeutic drugs (i.e. Cytarabine or Daunorubicin) can only eradict most of the AML blast cells but not LSCs (Hanekamp et al., 2018). Given that KG-1a cells serve as an effective model for studying chemoresistance properties in CD34+/CD38- AML LSC (Farge et al., 2017), the contribution of FAM30A and its repeats to chemoresistance was investigated (Figure 21).

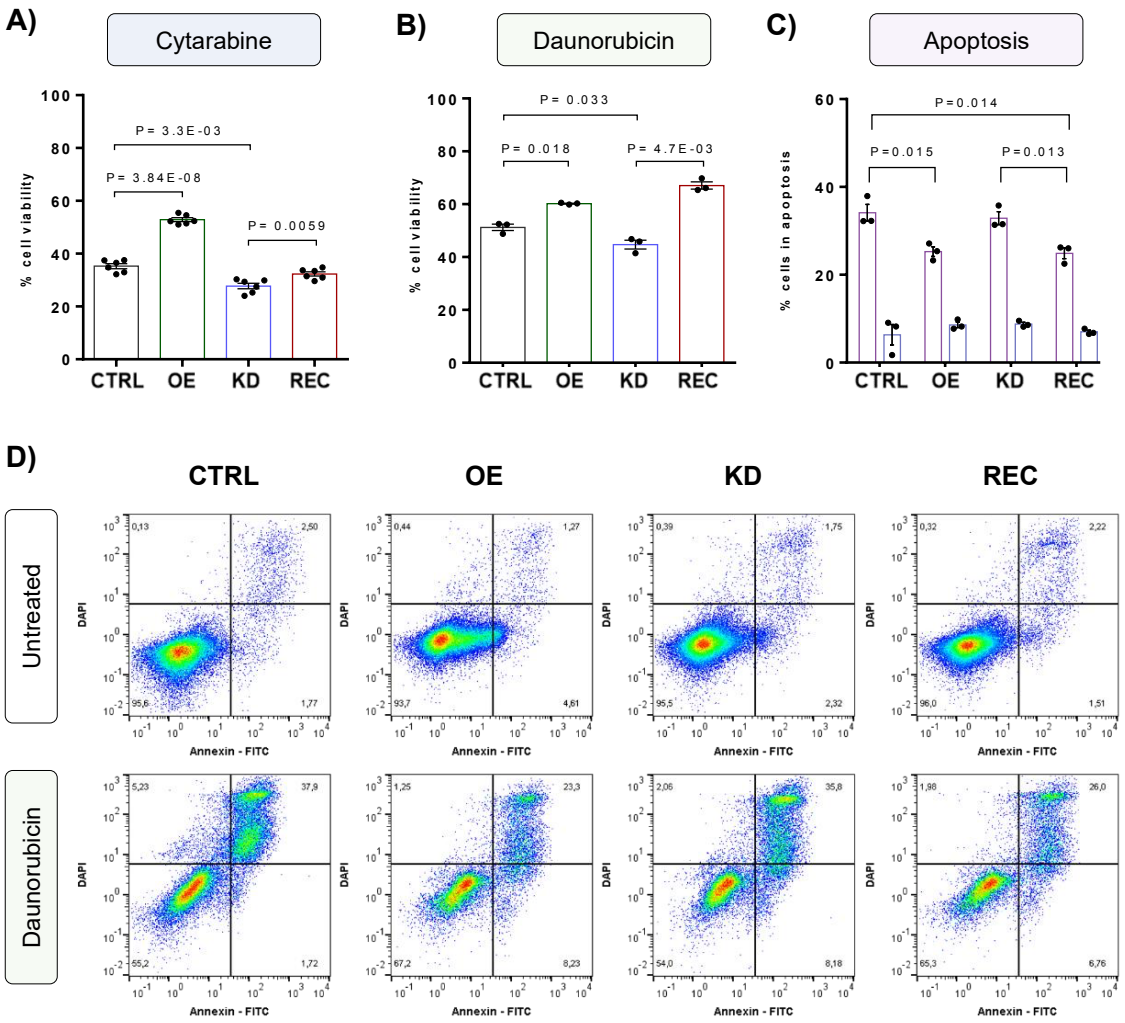


Figure 21. Modulation of FAM30A expression affects chemosensitivity in AML LSC-like cells. A) Graph depicting percentage of cell viability after treatment with Cytarabine (3 μM) for 48 hours in KG-1a stable cell lines normalized to untreated cells (n=6).
-Continued on next page-

B) Graph showing percentage of cell viability after treatment with Daunorubicin (0.5 µM) for 48 hours normalized to untreated cells (n=3). **C)** Graph depicting the quantification of apoptotic levels in KG-1a cells from B) after staining with Annexin-V and DAPI for detection of early and late apoptosis, respectively**. D)** Representative graphs from flow cytometry after staining with Annexin-V and DAPI from B). Significance was calculated using unpaired two-tailed student t-test and compared to CTRL (OE) and KD (REC). Data is displayed as mean values and error bars represent SEM.

Cytarabine arabinoside (Ara-C) and Daunorubicin were used as chemotherapeutic drugs and the IC_{50} values were calculated in KG-1a cells (± 3 µM and 0.5 µM, respectively) which were in line with previous studies (B. Chen et al., 2018). Figure 21A shows that upon overexpression of the repeats there is a substantial and significant increase in chemoresistance to Cytarabine when compared to control (OE vsC) and moderately in the recovery when compared to knockdown (REC vs KD) (17% and 6%, respectively). In adittion, FAM30A depletion also affected sensitivity of KG-1a cells to Ara-C (8% reduction). Aligning with these results, KG-1a cells treated with Daunorubicin also showed the same trend but a greater increase in chemoresistance was seen upon overexpression of the repeats when compared to the control (OE vs C) and knockdown (REC vs KD) (11% and 21% increase, respectively) (Figure 21B). Increased chemoresistance upon overexpression of the repeats (OE, REC) was also translated into lower apoptosis levels in FACS-based cytotoxicity assay compared their respective controls (Figure 21C, D). Altogether, these findings confirm the significance of FAM30A and its repeats, in protecting KG-1a cells from chemotherapy treatment.

3.3.3.- Modulation of FAM30A expression affects colony-formation potential and myeloid-lineage specification

Haematopoietic stem and progenitor cells (HSPCs) as well as LSCs posses the ability to proliferate and form colonies in a semi-solid medium that includes cytokines and growth factors that resembles the BM microenvironment. When cultured in the appropriate methylcellulose-based medium, individual progenitor cells (also called colony-formation units or CFUs) proliferate and differentiate to form distinguishable cell clusters containing their visually recognizable progeny. In the myeloid lineage, these progenitor cells can be multipotential (CFU-GM) capable of generating myeloid-restricted blood cells such as granulocytes (CFU-G) or monocytes/macrophages (CFU-M) (Allieri et al., 1990; Kronstein-Wiedemann & Tonn, 2019; Sarma et al., 2010). The clonogenic capacity of KG-1a cells was assesed upon modulation of expression of FAM30A and its repeats. This was performed by counting the number and types of colonies in a CFU-assay which provides information about the frequency and types of progenitor cells (Figure 22).

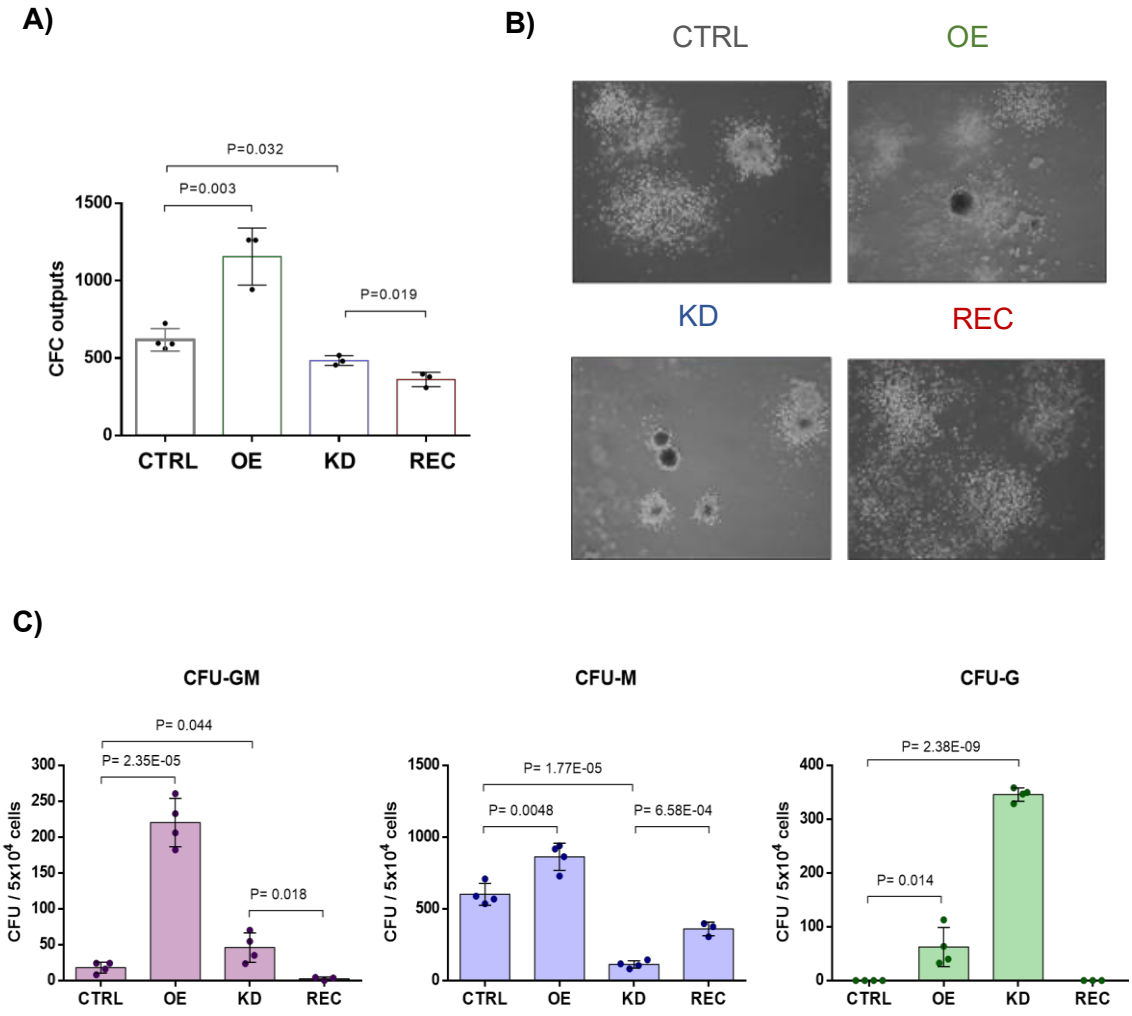


Figure 22. Effects of FAM30A expression changes on colony-forming abilities. A) Total number of colonies expressed in Colony-forming Cell (CFC) output values calculated from cell expansion rates in KG-1a stable cell lines cultured in methylcellulose-based medium for 14 days (n=4). **B)** Representative microscopy images at 40x magnification of KG-1a cells from A). **C)** Total numbers for each CFU type based on colony morphology (n=4) from A). Unpaired two-tailed student t-test was performed for statistical analysis. Data is displayed as mean values and error bars represent SEM.

First, the total number of colonies or colony-forming cell (CFC) outputs was calculated based on cell expansion in liquid culture (data not shown) to assess the leukaemogenic potential of each cell population (Figure 22A). Upon overexpression of the repeats, the CFC outputs were significantly increased by approx. 2-fold compared to the control (OE vs C), indicating an increased clonogenic capacity in these cells. On the other hand, cells with depleted FAM30A (KD) showed a significantly lower count of total colonies when compared to control (approx. 20% less colonies) which was not recovered upon overexpression of the repeats (REC). As FAM30A was affecting colony numbers, it was next determined if the colony types were changing. Thus, it was necessary to identify and quantify the different colony types of KG-1a cells based on general colony descriptions from StemCellTechnologies. Figure 22B (and Supplemental Figure 5)

shows the strong changes in the morphology of the colonies upon expression changes in FAM30A and its repeats. As expected, as KG-1a is a primitive myelomonocytic leukaemia cell line, colonies in the control group (CTRL) were myeloid-restricted with a few multipotential CFU-GM and mostly CFU-M showing monocytes and macrophages spreading over the plate (Hartman et al., 1994). Upon overexpression of the repeats (OE), the number of CFU-GM (dense core of cells with monocyte/macrophages at the periphery) greatly increased suggesting that this population includes more primitive and multipotent CFUs capable to generate more CFU-M and CFU-G colonies. This type of colonies were also significantly increased compared to the control (Figure 22C). However, endogenous FAM30A knockdown (KD) was translated into far less CFU-M colonies, a moderate increase in CFU-GM (P = 0.044) and a massive increase in CFU-G colonies (P = 2.38E-09). These results suggest that absence of FAM30A promotes the myeloid-comitted progenitor cell to differentiate towards the granulocytic path rather than the monocyte/macrophage lineage seen in the control. Interestingly, the REC population showed a strong increase in CFU-M colonies (macrophage-type of cells spreading in the plate) and no CFU-GM and CFU-G colonies were identified. This indicates that the REC cell population was differentiated and could not recover the more primitive type of cells seen before in the OE cell line. These results highlight the relevant role of FAM30A and its repeats in maintaining stemness and regulating myeloid-specific differentiation towards either monocyte/macrophage or granulocytes.

The next step was to validate these changes in the differentation process at the immunophenotype level in the four cell populations (Figure 23, Supplemental Figure 6). Thus, several surface markers expressed at different stages of the differentiation process were analyzed: CD34, CD38 and CD45 to confirm whether these AML LSC-like type of cells are expressing classical HSC markers. CD33 was used as a early myeloid progenitor marker which has been shown to be enriched in the LSC population having the most engraftment potential (Qin et al., 2021; Tambaro et al., 2021; Taussig et al., 2010). CD11B (also known as macrophage-1 antigen or ITGAM) was used as a differentiation marker as it is expressed in macrophages and granulocytes (Dziennis et al., 1995). HLA-DR or human leukocyte antigen-DR was used as it is a marker for functionally defined monocytes (Abeles et al., 2012; Joshi et al., 2023). Other myeloid differentation markers (for monocytes, macrophages and differentiated granulocytes) such as CD14, CD16, CD64 and CD86 were also used but data are not shown as expression levels were not readily detectable. Results from flow cytometry showed that the immunophenotype of the four populations were mainly CD34+, CD38-, CD45low.

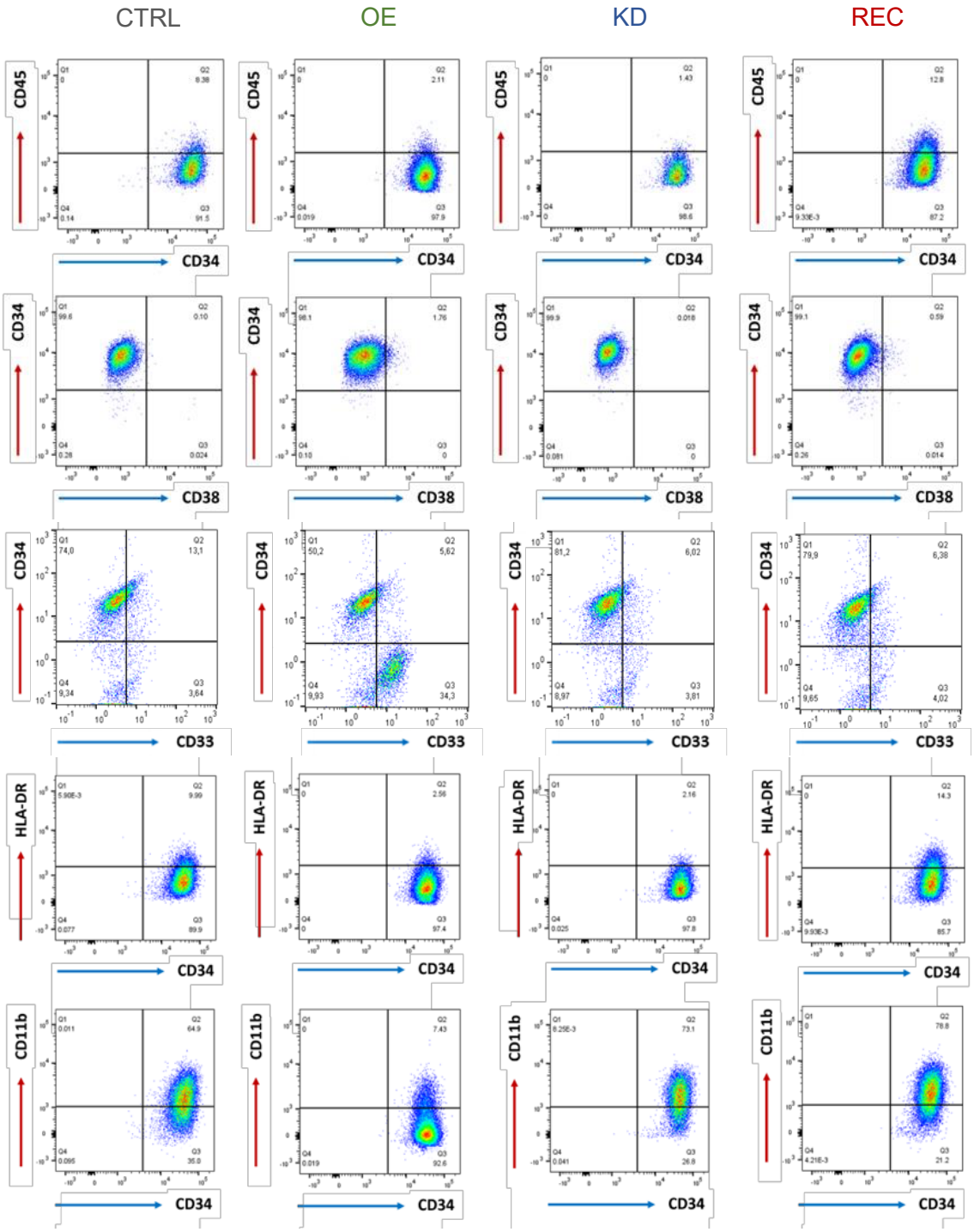


Figure 23. FACS analysis of early and myeloid markers upon modulation of expression of FAM30A and its repeats. Representative graphs from flow cytometry analysis of CD34, CD38, CD45, CD33, HLA-DR and CD11B in KG-1a stable cell lines after 5 days in liquid culture.

These are interesting findings as this immunophenotype has been functionally validated to have an enhanced colony formation and serial engraftment ability (Bonnet & Dick, 1997; Ishikawa et al., 2007). Staining with HLA-DR and CD11B showed that overexpression of FAM30A repeats (OE) decreases the expression of these myeloid differentiation markers compared to control (CTRL), especially of CD11B. In addition,

this OE population also showed a substantial increased expression of CD33, suggesting a more primitive immunophenotype with an early myeloid-commitment. Depletion of endogenous FAM30A (KD) showed a substantially increased expression of the CD11B marker but no increase in HLA-DR indicating a shift more towards granulocytic differentiation (as shown in Figure 21C). There was only a mild increase on CD45 expression in the cells overexpressing the repeats combined with the knockdown (REC vs KD). This was accompanied by an increased expression of HLA-DR in this population which suggests differentiation into monocytes/macrophages as previously seen visually as well in the CFU assays (Figure 22B).

As evidence has suggested, FAM30A and its repeats region confers certain stemness features in AML LSC-like cells. Therefore, the relevance of FAM30A repeats concerning chemoresistance and clonogenic potential in AML cell lines, which have barely detectable levels of FAM30A expression (MV4-11 and ME-1 cells), was investigated. MV4-11 and ME-1 cells were treated with Cytarabine and Daunorubicin at the IC_{50} values previously calculated (data not shown). Overexpression of the repeats (OE) significantly conferred increased chemoresistance compared to control (CTRL) in both cell lines and both drugs (i.e. 14% and 26% more resistance to Cytarabine and 4% and 14% to Daunorubicin in MV4-11 and ME-1, respectively). Strikingly, OE also increases cell viability in normal liquid culture when compared to control (approx 30-40% more in both cell lines) which suggests an overall increased leukaemogenic capacity due to the presence of the FAM30A repeats (Figure 24).

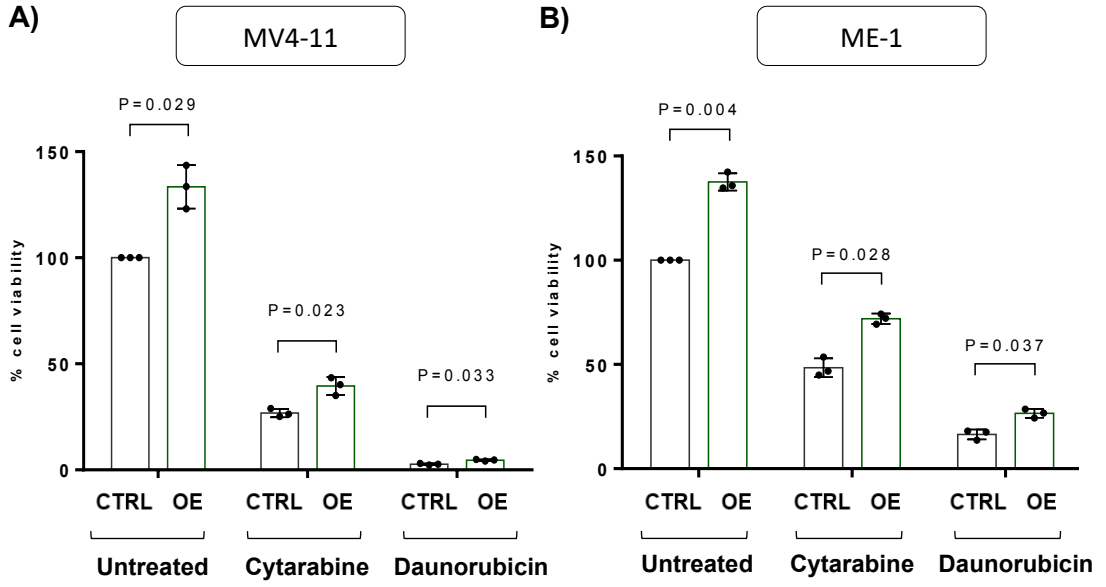


Figure 24. Overexpression of FAM30A repeats promotes chemoresistance in FAM30A^{low} AML cell lines. A) Graph representing percentage of cell viability of MV4-11 stable cell lines after 72 hours cultured in normal conditions (untreated), with Cytarabine (1 μM) and Daunorubicin (0.1 μM) and normalized to MV4-11 control cells (CTRL) untreated (n=3).
-Continued on next page-

B) The same protocol was performed in ME-1 cells as in A) (n=3). Data is displayed as mean values and error bars represent SEM. Unpaired two-tailed student t-test was performed for statistical analysis.

Moreover, to evaluate the effect of modulating FAM30A repeats on the leukaemogenic clonogenic potential of FAM30A^{low} AML cells, the colony number was determined in MV4-11 cells after 14 days in methylcellulose-based medium (Figure 25).

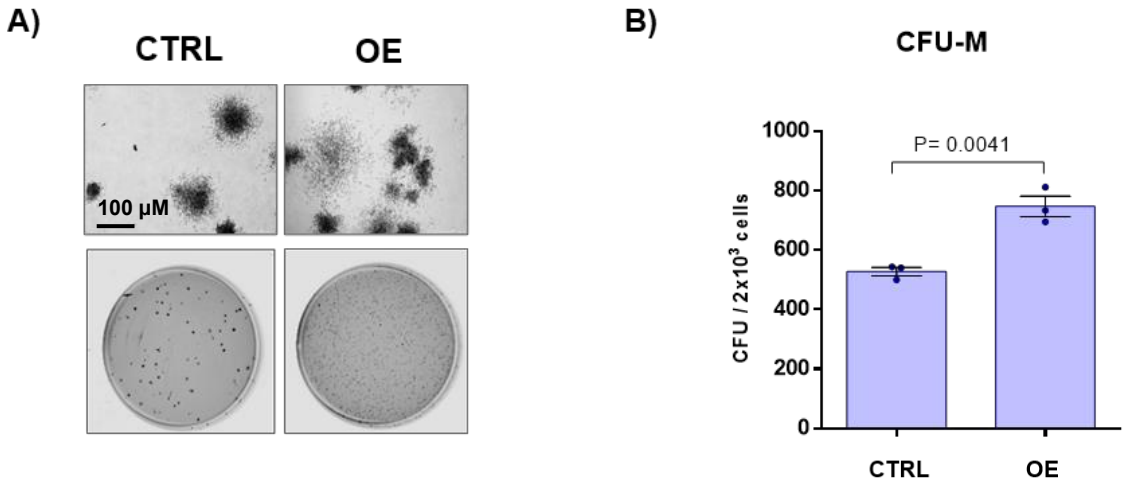


Figure 25. Overexpression of FAM30A repeats promotes clonogenic capacities in the FAM30A^{low} AML cell line MV4-11. A) Representative microscopy images at 40X magnification (upper) and the image of the plates (lower). **B)** Total numbers for CFU-M type based on colony morphology (n=3) from A). Data is displayed as mean values and error bars represent SEM. Unpaired two-tailed student t-test was performed for statistical analysis.

Due to technical reasons, this experiment could not be performed in ME-1 cells as they did not form discrete colonies. MV4-11 cells were used a model to evaluate if these cells that express early and late monocyte/macrophage markers (i.e. CD33, CD15, CD64 and HLA-DR) are able to restore certain stemness features. Figure 25A shows that, upon overexpression of FAM30A repeats (OE), the morphology of the colonies does not substantially change as MV4-11 mainly forms CFU-M with monocyte/macrophages. However, in this OE population there was a significant increase on colony number (1.54-fold more compared to control) though there was no apparent presence of CFU-GM that could indicate a restoration of stemness (i.e. more primitive cells). This finding was also confirmed via FACS analysis of stem cell markers (data not shown). Altogether, these results suggest that even though some of the effects of FAM30A repeats can be seen in MV4-11 cells, phenotypes associated with stemness are probably lost.

3.4. Molecular function of FAM30A

3.4.1.- LC-MS/MS to identify protein interactors of FAM30A

To further characterize the molecular function of FAM30A repeats, identifying interacting partners is essential for elucidating its relevance to "stemness" in the context of AML. Thus, RNA pulldowns were performed using biotinylated *in vitro* transcribed FAM30A repeats incubated with KG-1a cell lysates (Figure 26). After incubation and subsequent elution of proteins bound to the repeats, the pulldown samples were subjected to liquid chromatography with tandem mass spectrometry analysis (LC-MS/MS) for an unbiased approach to identify associated RNA-binding proteins (RBPs). The enriched proteins with FAM30A repeats over Bead Control (BC) are shown in Figure 26A.

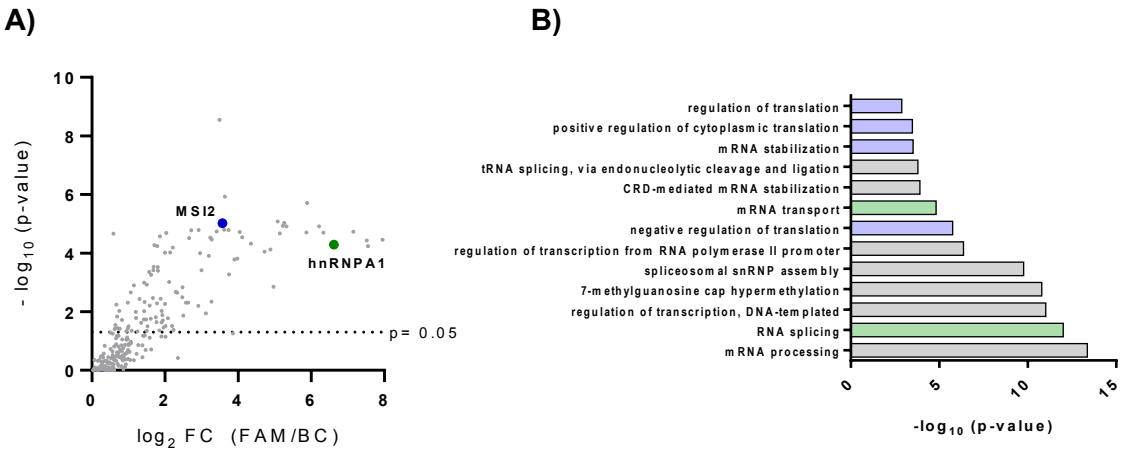


Figure 26. FAM30A repeats-associated proteins identified by RNA pulldown and LC-MS/MS analyses. A) Plot showing enriched proteins in FAM30A repeat pulldown (FAM) samples vs bead control (BC) based on identified peptides (Fold Change) in KG-1a cells (n=3). **B)** Functional clustering analysis using DAVID bioinformatic tool for significantly enriched proteins with FAM30A repeats vs bead control ($P < 0.05$). Highlighted in blue and green are biological processes associated to MSI2 and HNRNPA1, respectively.

Among the most significantly enriched proteins, two of them were marked as potential candidates for different reasons: Musashi-2 or MSI2 (Fold Change over BC = 11.87 ; *P* = 9.56E-06) and Heterogenous ribonucleoprotein A1 or HNRNPA1 (Fold Change = 98.32; *P* = 5.16E-05).

Even though HNRNPA1 is a well-known splicing factor, it's a multifunctional protein involved in a wide range of cellular processes such as gene transcription, splicing, mRNA export and translation (Jean-Philippe et al., 2013). This protein has been shown to abnormally shuttle to the cytoplasm in leukaemia cells and its abnormal high expression has also been reported in AML patients (J. Fang et al., 2017; S. Q. Li et al., 2020; Song et al., 2017). On the other side, MSI2 is a well-known cytoplasmic protein involved in AML progression through regulation of the oncogenic self-renewal program in LSCs and is recently being investigated as a therapeutic target for myeloid leukaemia (Byers et al.,

2011; Minuesa et al., 2019; Park et al., 2015). In addition, functional clustering of significantly enriched proteins was performed (cut-off $P < 0.05$) using DAVID bioinformatics software (D. W. Huang et al., 2007) (Figure 26B). As expected, the majority of the most significantly enriched biological processes (Gene Ontology terms) were involved in RNA processes such as splicing and regulation of transcription. This could indicate non-specific binding as these proteins reside in the nucleus and FAM30A is located in the cytoplasm (complete list of significantly enriched proteins FAM/BC provided in Supplementary Table 1). However, some of these processes could be also attributed to high binding of the HNRNP- family of proteins to FAM30A repeats (including HNRNPA1, highlighted in green). Other relevant functions associated with enriched proteins were related to regulation of cytoplasmic mRNA translation. Interestingly, MSI2 (highlighted in blue) has been shown by several studies to efficiently control the translation of crucial stemness factors in AML cells by directly interacting with their mRNAs (Hattori et al., 2017; Ito et al., 2010; Kharas et al., 2010).

RNA sequences of FAM30A repeats were also subjected to *in silico* analyses to identify putative RBPs motifs using regRNA2.0 and RBPmap softwares that comprise prediction of functionally validated RBP sites. Several predicted binding motifs were found for MSI2 and HNRNPA1 within the FAM30A repeats. This was in consonance with other previous studies that revealed the preferred binding sites for MSI2 (C. G. Bennett et al., 2016; Duggimpudi et al., 2018; Nguyen et al., 2020) and HNRNPA1 (Bruun et al., 2016). At least three and two binding sites were predicted to be located in each of the six repeats for MSI2 and HNRPA1, respectively (Figure 27).

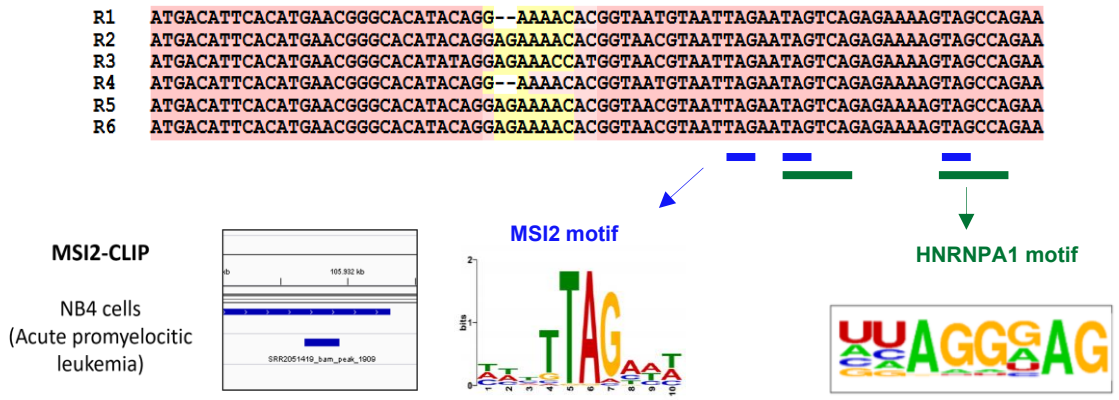


Figure 27. Predicted binding motifs for MSI2 and HNRNPA1 in FAM30A repeats. Depiction of the six repeats present in FAM30A sequence after multiple sequence alignment in T-Coffee software (in red highest score of sequence similarity and good score in yellow). Putative binding motifs are highlighted in blue (MSI2) and green (HNRNPA1) showing reported binding sites extracted from (Bruun et al., 2016; Duggimpudi et al., 2018). Left panel: IGV image of FAM30A repeats detected in MSI2-CLIP data from (Rentas et al., 2016).

In addition, publicly available MSI2-CLIP data also showed that FAM30A was captured in the repeat region in an AML cell line (Rentas et al., 2016).

3.4.2.- FAM30A associates with hnRNPA1 and MSI2 *in vitro*

Next, the association of MSI2 and HNRNPA1 to FAM30A repeats revealed in the LC-MS/MS experiment was validated via western blot analysis in two cell lines with high FAM30A expression: KG-1a and GDM-1 cells (Figure 28).

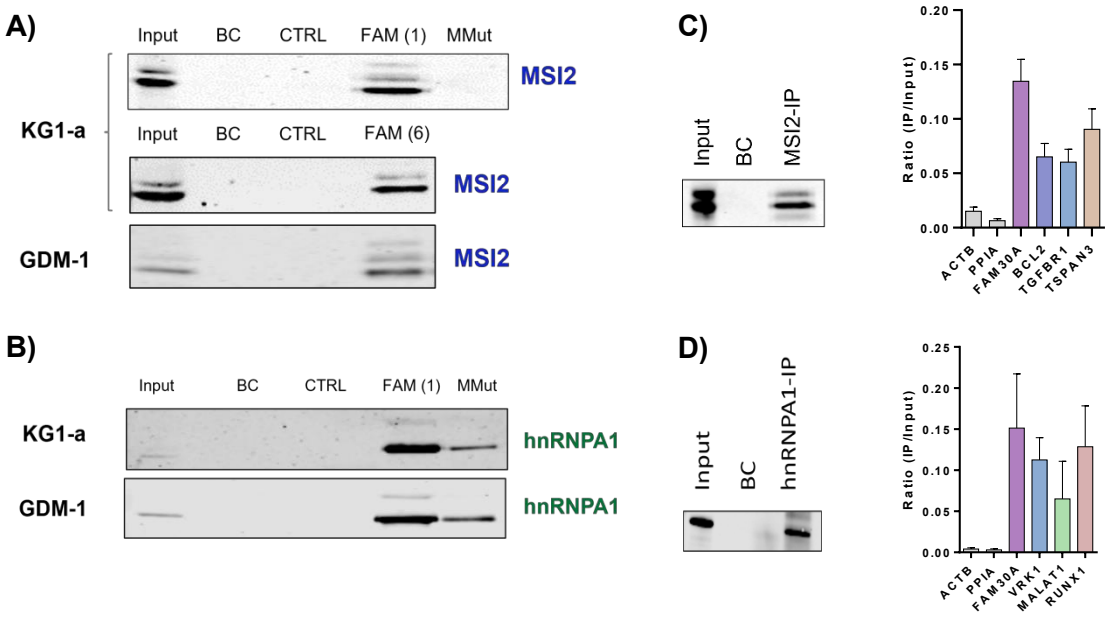


Figure 28. FAM30A repeats associate with MSI2 and HNRNPA1 *in vitro*. A, B) Western blot analysis for MSI2 (A) or HNRNPA1 (B) of RNA pulldown samples from KG-1a and GDM-1 cell lysates with the RNAs CTRL, one (FAM (1)) and six FAM30A repeats (FAM (6)) and mutated MSI2-binding sites (MMut) used as baits. **C, D)** Western blot analysis for validation of immunoprecipitated MSI2 (C) or HNRPA1 (D) protein and RT-qPCR analysis of associated RNAs (right) performed in KG-1a cells. Data is displayed as mean values and error bars represent SEM.

For this reason, the same experimental procedure was performed but also using an RNA sequence (CTRL) 500bp immediately prior to the repeat region (464bp) to confirm that binding of these proteins is specific to the repeats. In addition, two biotinylated *in vitro* transcribed wild-type FAM30A repeats sequences were used: one repeat (FAM(1)) and six repeats (FAM(6)) in order to examine if the two proteins had differential binding affinity depending on the number of putative binding sites. To validate the specific interaction site between FAM30A repeats and MSI2/HNRNPA1, one additional construct was generated. This version included FAM30A repeat whose putative MSI2-binding motifs were mutated TAG → AAA (MMut). Only mutation within the MSI2-binding motif was performed due to technical reasons to generate mutated constructs for HNRNPA1 plus MSI2-binding sites are included to a large extent within HNRPNA1 binding sites (see

Figure 27). After RNA pulldown, western blot analysis for each of the RNAs showed that MSI2 and HNRNPA1 strongly associate to one (FAM(1)) and the six repeats (FAM(6)) in both cell lines, suggesting directing binding. The association of MSI2 with the repeats is abolished when the MSI2-binding sites are mutated, indicating the interaction sites between them. Nonetheless, even though association of HNRNPA1 to the mutated version is still detected, the severely decreased signal suggests that mutation of most of the MSI2-binding site also affects the association between the repeats and HNRNPA1 (Figure 28B).

Moreover, this interaction was investigated in more detail by immunoprecipitation (IP) of MSI2 and HNRNPA1 in KG-1a cells and subsequent analysis of associated RNAs was performed via RT-qPCR. The succesful immunoprecipitation of endogenous MSI2 and HNRNPA1 was validated by western blot. Figures 28C and 28D show that both proteins strongly associate with FAM30A RNA. ACTB and PPIA were used as negative RNA controls and experimentally validated targets of MSI2 (Kwon et al., 2015; Park et al., 2014) and HNRNPA1 (Davis et al., 2021; Miao et al., 2022; Ryu et al., 2021) were used as positive controls. Binding ratios compared to the input showed that FAM30A RNA associates even more efficiently than well-known targets of both proteins.

3.4.3.- Effects of modulation of FAM30A expression on HNRNPA1 protein levels and downstream pathways.

Based on the *in vitro* association of FAM30A and HNRNPA1 and its potential relevance in the AML context, the correlation of their expression in AML patients was first investigated (data extracted from TCGA-cBioPortal). As Figure 29A shows, there was no significant correlation between FAM30A and HNRNPA1 expression. This does not imply there is no direct implication of FAM30A on HNRNPA1 molecular function as this protein is ubiquituosly expressed and there is no differential expression in normal HSCs and differentiated myeloid cells (R2 platform).

As previously mentioned, abnormal cytoplasmic localization of HNRNPA1 has been reported in leukaemia but these findings required validation in AML cell lines (Figure 29B). Results derived from cell fractionation studies indicate there is hardly any HNRNPA1 in the cytoplasmic fraction (less than 3% and 0.5% in MV4-11 and KG-1a cells, respectively). Moreover, immunostainings with an antibody against HNRNPA1 in KG-1a stable cells showed that there was no apparent difference in changes of cellular localization of HNRNPA1 protein that could suggest an interaction with FAM30A in the cytoplasm (data not shown).

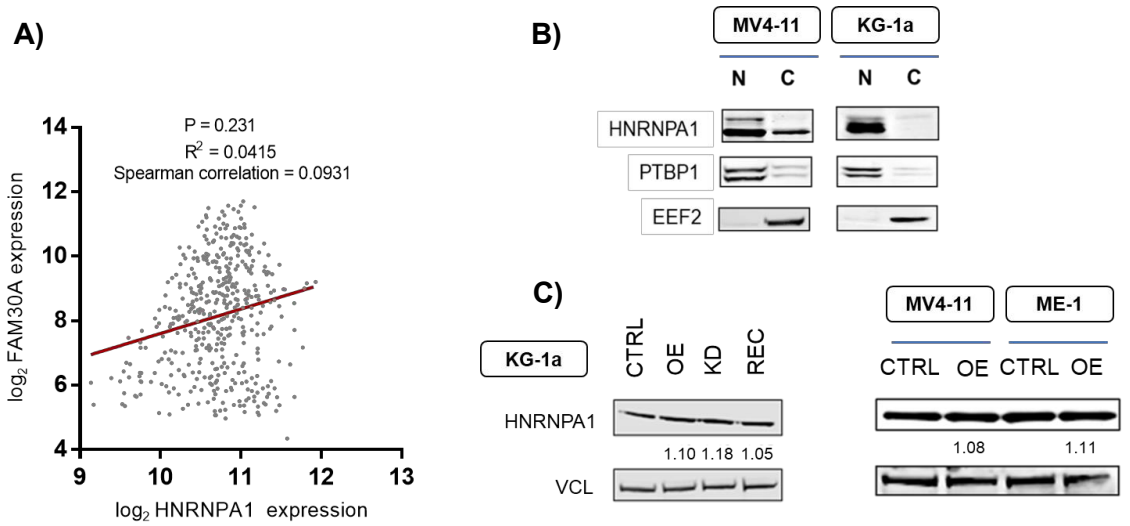


Figure 29. FAM30A repeats do not alter HNRNPA1 protein levels. A) Levels of FAM30A and HRNPA1 RNA expression expressed in log_2 RSEM (RNA-seq by Expectation-Maximization) extracted from cBioPortal, TCGA (PanCancerAtlas). Spearman´s correlation and p-values are extracted from regression line. **B)** Western blot analysis upon digitonin-based cell fractionation of nuclear (N) and cytoplasmic (C) fractions in KG-1a and MV4-11 wild-type cells depicting HRNPA1, PTBP1 (control for nuclear fraction) and EEF2 (control for cytoplasmic fractions) protein levels. **C)** Western blot analysis showing HRNPA1 protein levels in KG-1a, MV4-11 and ME-1 stable cell lines. VCL was used as loading control. (n=3).

In addition, the alteration of HNRNPA1 protein levels was analyzed upon modulation of FAM30A expression and its repeats (Figure 29C). Neither an endogenous depletion or FAM30A (KD) nor exogenous overexpression of the repeats (OE, REC) changed HRNPA1 protein levels in $FAM30A^{high}$ cells (KG-1a) and in $FAM30A^{low}$ cells (MV4-11, ME-1). Moreover, results from sucrose gradients showed that upon overexpression of the repeats, HNRNPA1 proteins still remains in the lower (free) fractions compared to control thus indicating again that HNRNPA1 protein unlikely resides in the cytoplasm (data not shown).

HNRNPA1 has been shown to heavily impact target gene splicing with subsequent haematopoietic defects independently of changes in its expression levels (J. Fang et al., 2017). Although no apparent effects on HNRNPA1 protein levels were observed, the potential downstream effects after manipulating the expression FAM30A and its repeats on HNRNPA1 function were investigated. In the AML context, HNRNPA1 has been found to be overexpressed and contribute to leukaemic phenotypes (Song et al., 2017). One of the reported mechanisms involving HNRNPA1 is to be a major player in controlling the splicing of the RUNX1 isoform pool (Davis et al., 2021). RUNX1 is one of the critical haematopoietic transcription factors that regulates HSCs fate through its three main protein isoforms (RUNX1A, B and C) which have antagonistic functions (Miyoshi et al.,

1995; Tanaka et al., 1995). Moreover, RUNX1 is one of the most commonly mutated genes in AML. In this regard, HNRNPA1 has been shown to be a potent repressor of RUNX1A formation via binding to the alternative exon 7A (Figure 30A). Thus, the relative fractions of RUNX1 isoforms were analyzed by RT-qPCR upon modulation of FAM30A expression and its repeats, along with concurrent depletion of HNRNPA1. This was performed to determine whether FAM30A repeats influence the RUNX1 splicing pattern via HNRNPA1.

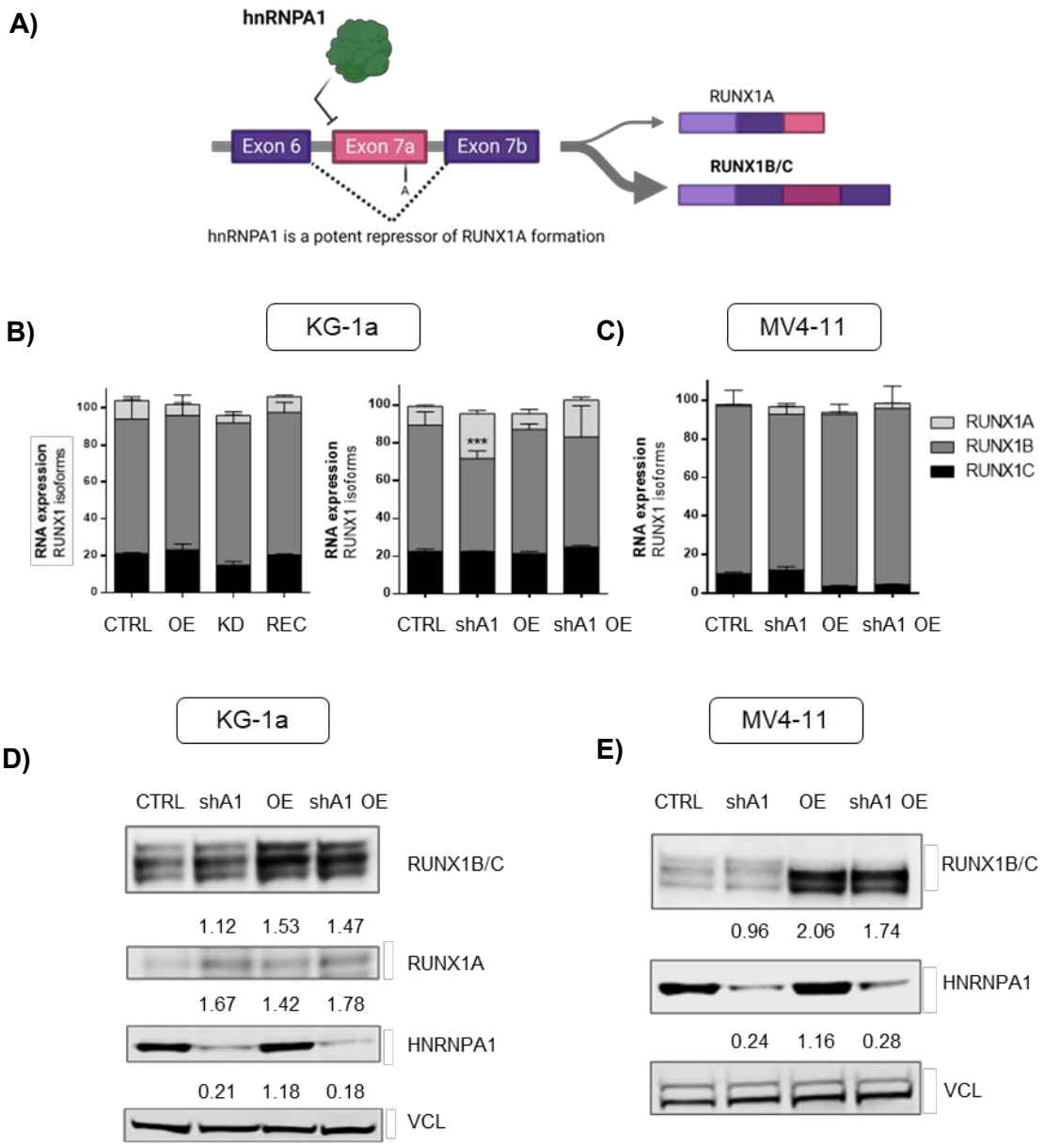


Figure 30. FAM30A repeats do not affect RUNX1 splicing pattern via HNRNPA1 but affect RUNX1 protein levels. A) Schematic overview of HNRNPA1 role on RUNX1 splicing, adapted from (Davis et al., 2021). **B)** Graph with fractions of RUNX1 isoforms after RT-qPCR analysis in KG-1a stable cell lines transduced with control (CTRL), overexpression of FAM30A repeats (OE), HNRNPA1 knockdown (shA1) and HNRNPA1 knockdown with overexpression of FAM30A repeats (shA1 OE) constructs (n=3). RNA expression levels were normalised to ACTB.
-Continued on next page-

C) RT-qPCR analysis in MV4-11 stable cell lines with control (CTRL), overexpression of FAM30A repeats (OE), HNRNPA1 knockdown (shA1) and HNRNPA1 knockdown with overexpression of FAM30A repeats (shA1 OE) constructs (n=3). **D)** Western blot analysis in KG-1a cells from B), right (n=3). **E)** Western blot analysis in MV4-11 cells from C). VCL was used as loading control (n=3). Significance was calculated using multiple unpaired student t-test (***- p<0,001). Data is displayed as mean values and error bars represent SEM.

Results in Figures 30B show that neither FAM30A nor its repeats were affecting substantially the distribution of RUNX1 isoforms in FAM30A^{high} cells (KG-1a). There was only a mild increase on RUNX1B isoform upon FAM30A knockdown (KD) and an increase on RUNX1C isoform upon overexpression of the repeats (REC) when compared to FAM30A knockdown (both not significant). As expected, HNRNPA1 knockdown (shA1) showed a significant increase on RUNX1A isoform but overexpression of FAM30A repeats in this setting (shA1 OE) did not enhance nor restore the original levels. Similar results were seen in FAM30A^{low} cells (MV4-11) with no significant changes in splicing of RUNX1 isoforms (Figure 30C), suggesting that there is an unlikely cooperation between FAM30A repeats and HNRNPA1 in regulating RUNX1 isoforms.

Strikingly, western blot analysis under the same conditions showed that upon FAM30A repeats overexpression (OE), RUNX1 total protein levels were considerably higher with a 1.5 and 2-fold increase compared to control in KG-1a and MV4-11 cells, respectively (Figures 30D, E). These results were independent of HNRNPA1 knockdown which suggests that there is an unexplored mechanism via FAM30A repeats that enhances RUNX1 protein levels, an event that will be further discussed later in this thesis.

3.4.4.- Characterization of the role of FAM30A on the oncogenic protein MSI2

Since MSI2 is a central regulator of cancer stem cell programs, along with the herein revealed *in vitro* association of FAM30A and MSI2 in AML LSC-like cells, further investigation into the potential oncogenic role of FAM30A in relation to MSI2 function in this context is required. First, FAM30A and MSI2 RNA expression levels were examined in AML patients using data extracted from cBioPortal (TCGA, PanCancerAtlas). Analysis determined that their correlation score was highly significant (Figure 31A) with MSI2 in the top25 most positively correlated genes with FAM30A. Moreover, MSI2 expression was strongly correlated with worse clinical outcome. AML patients with higher MSI2 expression had an overal survival of 15.5 months compared to 43.6 months ($P < 0.001$) from patients with lower MSI2 expression (Figure 31B). In this line, MSI2 protein and

RNA expression levels were significantly higher in AML primary samples compared to bone marrow samples from healthy patients (Figure 31C).

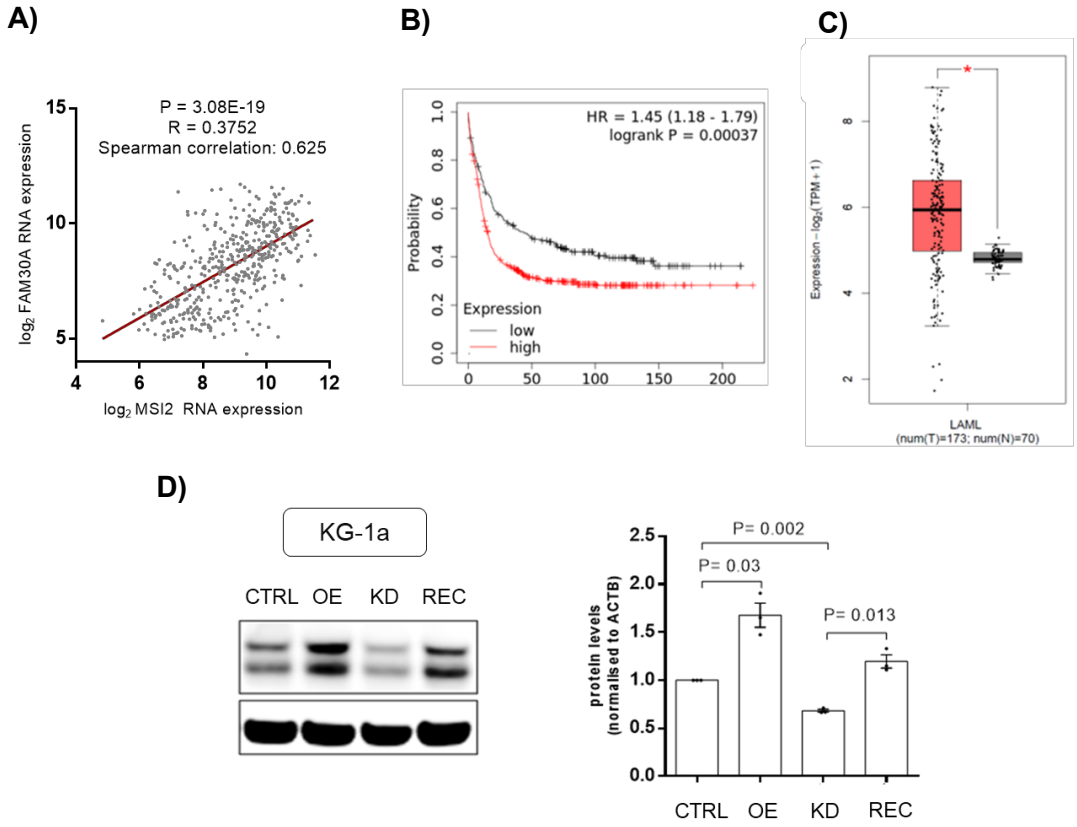


Figure 31. FAM30A and its repeats affect protein levels of the oncogenic MSI2 protein. A) Expression levels of FAM30A and MSI2 RNA (log_2 RSEM) extracted from cBioPortal, TCGA (PanCancerAtlas database). Spearman´s correlation and p-values are extracted from regression line. **B)** Kaplan-Meier analysis of overall survival data plotted using the Kaplan-Meier plotter using the cut-off of median MSI2 expression for separating high and low expression groups. **C)** RNA expression of MSI2 in AML primary samples shown in red (TCGA, n=173) compared to healthy bone marrow samples (n=70) expressed in log_2 TPM from GEPIA.2 software. **D)** Western blot analysis in KG-1a stable cell lines. Statistical significance was determined by unpaired two-tailed student t-test in (A) and and by Log-rank test (Mantel-Cox) in (B). (*- $p < 0,05$). Data is displayed as mean values and error bars represent SEM. ns = not significant. HR= Hazard ratio.

Subsequently, the effect of depleting endogenous FAM30A and overexpressing its repeats (MSI2-interacting region) on MSI2 protein levels was investigated. Results showed that exogenous overexpression of the repeats (OE) increased MSI2 protein levels (more than 60%) compared to control (CTRL). Conversely, FAM30A knockdown (KD) strongly decreased MSI2 protein levels by approx. 35% and these were restored to control levels after exogenous overexpression of the repeats (REC) (Figure 31D).

Genome-wide analysis of MSI2-deficient AML LSCs revealed MSI2´s primary role in controlling pathways that are critical for leukaemia stem cell maintenance as well as to contribute to normal HSC engraftment and cell fate decisions (Kharas et al., 2010; Park

et al., 2014, 2015). To confirm its implication in stemness, the MSI2 expression pattern in haematopoietic cells shows that its expression levels are the highest in the early HSCs and multipotent progenitors and progressively decreases along differentiation (Kharas et al., 2010). MSI2 RNA expression levels were analyzed in a panel of leukaemia cell lines, including those resembling AML LSC-like cell features KG-1a and GDM-1 cell lines. These showed the highest MSI2 expression levels, which also correlated with high FAM30A levels (Supplemental Figure 7). Interestingly, when screening FAM30A expression amongst AML primary samples, the group with the highest FAM30A levels also showed a significant correlation with MSI2 expression (*P* = 6.73E-07, TCGA PanCancer Atlas).

3.4.4.1.- Modulation of FAM30A and its repeats affects MSI2-targets at the post-transcriptional level

MSI2 protein is thought to function by binding mostly to the 3´ UTR – to a lower extent to CDS and 5´UTR – of target mRNAs. This usually involves MSI2-binding to the consensus sequence (UAG) to either enhance or repress the translation of mRNAs necessary for AML LSCs maintenance (Kawahara et al., 2008; Nguyen et al., 2020; Yeh et al., 2023) (Figure 32A). Some of the experimentally validated MSI2-targets are NUMB, TGFBR1, TSPAN3, SMAD3, BCL2, among others (Han et al., 2015; Kwon et al., 2015; Nishimoto & Okano, 2010). To establish the impact of FAM30A on MSI2 protein function, potential alterations in MSI2 targets were explored by modulating the expression of FAM30A and its repeats. Thus, in addition to generating stable KG-1a cell lines with FAM30A constructs for knockdown and overexpression of the repeats, K562 cells were subjected to MSI2 knockdown and FAM30A repeats overexpression. This approach was necessary because proper MSI2 knockdown could not be achieved in KG-1a cells.

Protein levels of NUMB and SMAD3, two targets where MSI2 exerts opposite effects at the translational level, were analyzed. MSI2 is known to bind to the 3' UTR of NUMB and repress its translation, whereas it efficiently promotes the translation of SMAD3 mRNA (Han et al., 2015; Nishimoto & Okano, 2010). Western blot analysis in KG-1a cells shows that overexpression of FAM30A repeats dramatically decreases NUMB protein levels and significantly increases SMAD3 thus overall indicating an increased MSI2 activity (also MSI2 protein levels were higher compared to control). Conversely, when FAM30A was depleted in KG-1a cells, it induced a reverse impact on NUMB and SMAD3 protein levels, a pattern also mirrored with MSI2 knockdown in K562 cells. Notably, the effects of the reduced MSI2 function (in both FAM30A or MSI2 knockdown settings) were restored and even exceeded control levels upon the overexpression of FAM30A repeats

in both cell lines thus suggesting a positive direct link between FAM30A and MSI2 function (Figure 32B).

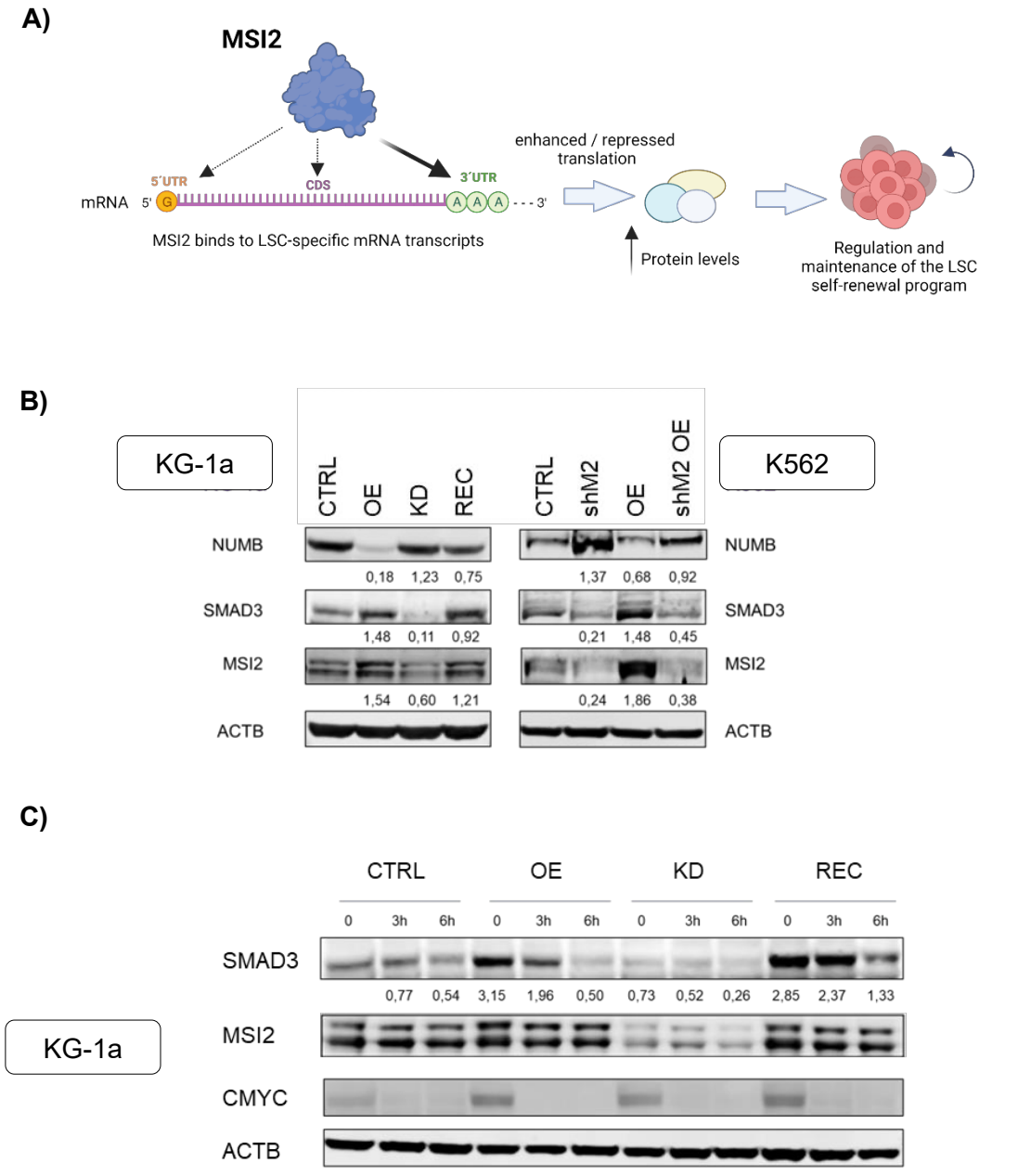


Figure 32. MSI2-targets are affected by modulation of FAM30A repeats expression. A) Schematic depiction of how MSI2 binds to its AML LSC-specific mRNA targets and subsequent maintains the self-renewal program. **B)** Western blot analyses of MSI2 targets -NUMB, SMAD3-, MSI2 and ACTB as loading control in stable KG-1a (left panel) and K562 stable cell lines (right panel) transduced with control (CTRL), overexpression of FAM30A repeats (OE), MSI2 knockdown (shM2), FAM30A knockdown (KD) and MSI2 knockdown (shM2 OE) or FAM30A knockdown (REC) with overexpression of FAM30A repeats constructs (n=3).
-Continued on next page-

C) Western Blot analysis of KG-1a stable cell lines transduced with control (CTRL), overexpression of FAM30A repeats (OE), FAM30A knockdown (KD) and FAM30A knockdown with overexpression of FAM30A repeats (REC) constructs (n=3) after treatment with emetine (0, 3 and 6 hours) (n=3). CMYC was used as a positive control and ACTB for loading control (n=3).

As there was a clear impact on MSI2-targets at the protein level, it was next investigated whether these variations on MSI2-targets stem from alterations in mRNA stability, translational efficiencies or protein degradation. Initially, the impact on translation was analyzed, considering the established association between MSI2 and translational control. Emetine is well-known protein synthesis inhibitor whose mode of action is through binding to the 40S small ribosomal subunit which irreversibly prevents the elongation of the tRNA-mRNA complex (Jiménez et al., 1977). After incubating KG-1a cells with emetine for 3 and 6 hours, noticeable effects on significantly sustained SMAD3 protein levels were observed upon the overexpression of FAM30A repeats (2.54-fold more in OE and 4.55-fold more in REC) when compared to their respective controls (CTRL, KD). Decreased protein levels in SMAD3 (33% and 50% less after 3 and 6 hours, respectively) were also observed upon FAM30A knockdown when compared to control cells (Figure 32C). In addition, available RNA-sequencing data from stable MSI2-knockdown experiments in leukaemia cells were analyzed. MSI2 inhibition (either by genetic depletion or using an MSI2 inhibitor named Ro 08-2750) in MOLM13 (AML) and K562 (CML) cells showed shared downregulated genes seen upon FAM30A knockdown (approx. 30 targets). Moreover, certain downstream targets upregulated upon overexpression of the repeats in KG-1a cells were also upregulated in MSI2-overexpressing HSCs in mouse (Supplemental Figure 8).

3.4.4.2.- FAM30A affects RUNX1 activation via a MSI2-dependent mechanism

Since previous findings demonstrated that changes in FAM30A expression affected RUNX1 protein but not RNA levels (Figure 30), an analysis was conducted to determine whether FAM30A differentially alters the pool of RUNX1 protein isoforms. Western blot analysis in KG-1a cells showed that overexpression of the repeats (OE) significantly increased RUNX1C (approx. 23% more) and RUNX1A (approx. 8%) as well as total RUNX1 protein levels by 2-fold compared to control. On the other hand, FAM30A knockdown (KD) showed the opposite trend as it greatly increased RUNX1B (approx. 25% more), decreased RUNX1C (approx. 14% less) and slightly increased RUNX1A (5% more) without significantly altering RUNX1 total protein levels. Overexpression of the repeats in the knockdown setting (REC) recovers original RUNX1B and RUNX1C protein levels seen in the control with a slight increase on RUNX1 total protein levels (Figure 33A, B). Interestingly, increased RUNX1C levels due to overexpression of the repeats

(OE, REC) were correlated with higher MSI2 protein levels compared to their controls (CTRL, KD). Results upon overexpression of the repeats in FAM30A^{low} cells (MV4-11) displayed a similar outcome as it increased both RUNX1B and RUNX1C protein levels (RUNX1A levels were not readily apparent) along with elevated MSI2 protein levels (Figure 33B).

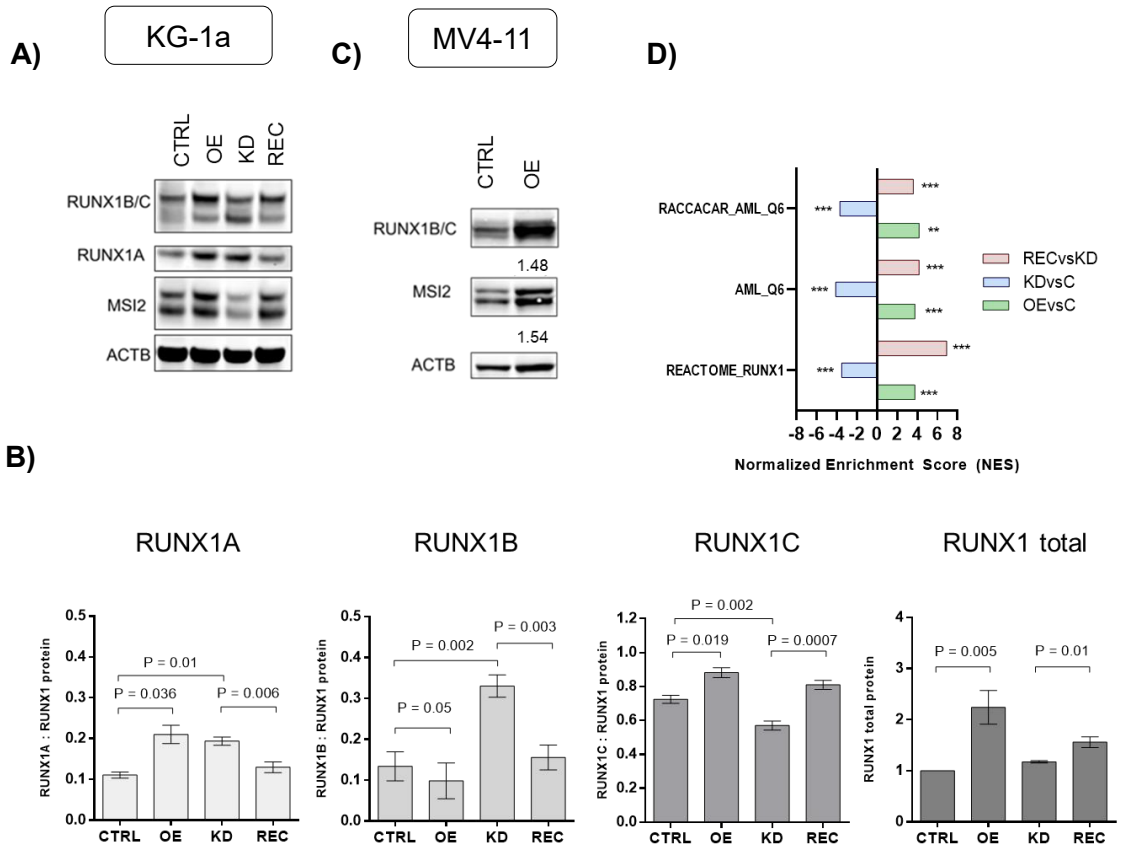


Figure 33. FAM30A and its repeats affect activation of the RUNX1 transcriptional network. **A)** Western blot analysis from KG-1a cells transfected with FAM30A constructs for knockdown and overexpression of the repeats (n=3). **B)** Quantification analysis of RUNX1 protein isoforms from A) normalized to total RUNX1 protein levels and against ACTB as loading control. **C**) Western blot analysis from MV4-11 cells transfected with FAM30A construct for overexpression of the repeats (n=3). ACTB was used as a loading control. **D)** GSEA analysis (C2 dataset) of RUNX1-related gene datasets in KG-1a cells transfected with FAM30A constructs for knockdown and overexpression of the repeats (n=3). OEvsC represents shC FAM vs shC EV; KDvsC is shFAM30A EV vs shC EV; RECvsKD is shFAM30A FAM vs shFAM30A EV. Significance was calculated using unpaired two-tailed student t-test (A, B) and in (C) shows FDR values. (***- p<0,001; **- p<0,01). Data is displayed as mean values and error bars represent SEM

As overexpression of FAM30A repeats was overall increasing total RUNX1 protein levels, RNA-sequencing data in stable KG-1a clones was examined in order to see whether the RUNX1 transcriptional network was also promoted. Indeed, GSEA analysis confirmed that transcriptional regulation by RUNX1 (REACTOME_RUNX1) was massively

increased upon overexpression of the repeats (OEvsC, RECvsKD). In the same line, transcripts with putative RUNX1 binding sites in their promoters (RACCACAR_AML_Q6 and AML_Q6 signatures) were significantly upregulated upon overexpression of the repeats. For instance, the direct RUNX1 target PU.1 (*SPI1* gene) was strongly upregulated upon overexpression of FAM30A repeats (logFC 2.59, FDR < 0.001). Conversely, FAM30A depletion significantly decreased the RUNX1 transcriptional network along with a downregulation of its target genes (Figure 33D). This suggests that the FAM30A repeats region plays a role in controlling protein levels of RUNX1 isoforms and ultimately affecting activation of the RUNX1 pathway.

MSI2 is known to bind thousands of target mRNAs mostly in the 3´UTR though differences in their translation efficiencies are restricted to a small fraction of these transcripts (Karmakar et al., 2022). Although there is no reported function of MSI2 directly regulating translation of RUNX1 mRNAs so far, this link was herein investigated in the AML context. First, publicly available MSI2 CLIP-seq data from three independent experiments conducted in NB4 (AML) and K562 (CML) cell lines were analyzed. MSI2 was heavily detected in all experiments at the 3´UTR of RUNX1B and RUNX1C (but not at the 3´UTR of RUNX1A) and also at the 5´UTR of RUNX1A and RUNX1B and overall MSI2 was detected in the CDS of all RUNX1 isoforms (Figure 34A).

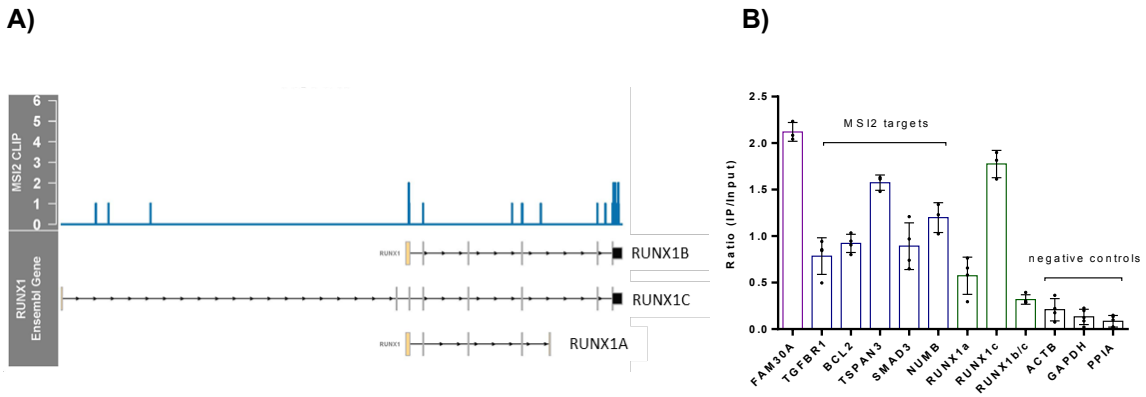


Figure 34. MSI2 differentially associates with RUNX1 isoforms. A) Schematic representation of regions where MSI2 binding was detected in RUNX1 isoforms, extracted from three publicly available and independent CLIP-seq experiments in two cell lines (NB4, K562 cells). Blue peaks represent the number of times that MSI2 CLIP sites were detected in the RUNX1 sequence (in 5´UTR, 3´UTR, CDS) (n=3). **B)** Western blot analysis after immunoprecipitation of endogenous MSI2 protein and subsequent RT-qPCR analysis of associated RNAs highlighting validated MSI2 targets (blue), RUNX1 isoforms (green) and negative controls (black) (n=3) performed in KG-1a cells. Data is displayed as mean values and error bars represent SEM.

These were interesting results as depending where MSI2 binds to its target mRNAs, it can either promote or inhibit its translation (Nguyen et al., 2020; Yeh et al., 2023). Since

there is mounting evidence suggesting a differential binding affinity of MSI2 to RUNX1 mRNAs, these results were validated *in vitro*. Immunoprecipitation of endogenous MSI2 protein and subsequent RT-qPCR analysis indicate that MSI2 heavily bind to RUNX1C mRNA and in a lesser extent to RUNX1A both in the same range as previously validated MSI2 mRNA targets (Figure 34B). Due to difficulties in designing specific primers for RUNX1B, primers targeting both RUNX1B and C isoforms were used. Since RUNX1B mRNA is much more abundant than RUNX1C in KG-1a cells (3:1 ratio in KG-1a wild-type cells), it could be inferred that MSI2 does not strongly bind to RUNX1B.

Given that FAM30A repeats promote MSI2 function (i.e. changes in MSI2-targets at the protein level) and MSI2 exhibits differential binding affinity to RUNX1 mRNAs, the changes in RUNX1 protein isoforms upon modulation of FAM30A expression were investigated when translation was blocked (Figure 35).

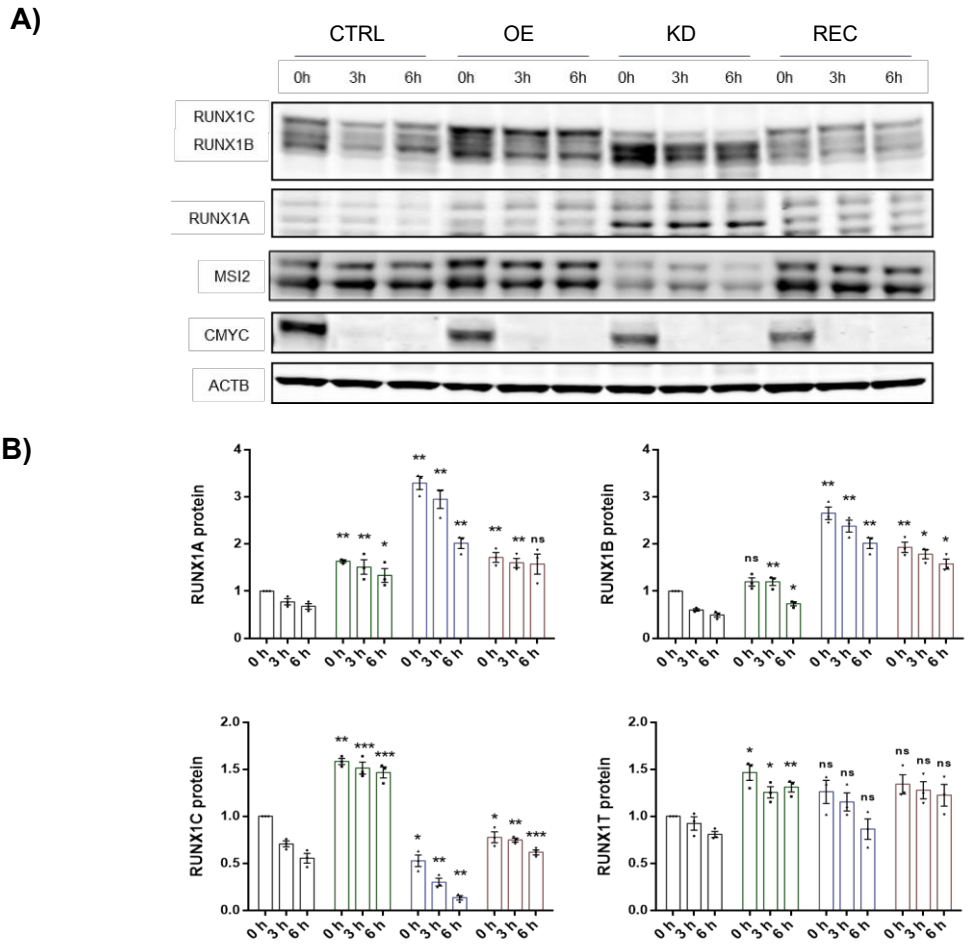


Figure 35. Modulation of expression of FAM30A and its repeats differentially affect protein turnover of RUNX1 protein isoforms. A) Western blot analysis of KG-1a cells transduced with FAM30A constructs after treatment with emetine (0, 3 and 6 hours). CMYC was used as a positive control and ACTB for loading control (n=3).
-Continued on next page-

B) Quantification analysis of RUNX1 protein isoforms after emetine treatment from A). Statistical significance was calculated using unpaired two-tailed student t-test OE vs CTRL, KD vs CTRL, REC vs KD normalised to each timepoint (***- p<0,001; **- p<0,01; *- p<0,05). Data is displayed as mean values and error bars represent SEM.

Even after 6 hours of emetine treatment, western blot analysis shows that overexpression of the repeats (OE, REC) maintained significantly high levels of RUNX1C isoforms (2,58 and 3,87-fold more, respectively) compared to their respective controls. Moreover, RUNX1B protein levels were mildly higher in OE (1,39-fold) but in REC showed the opposite trend (1,28-fold decrease). RUNX1A protein levels followed the same direction after emetine treatment as in RUNX1B. On the contrary, FAM30A depletion (KD) sustained high RUNX1A (2,98-fold) and RUNX1B (4,08-fold) protein levels when compared to controls after 6 hours of treatment. In this condition, RUNX1C protein levels were dramatically decreased when compared to control after 3 and 6 hours of treatment (58% and 76% less protein, respectively). Similar results were seen when incubating cells with cycloheximide, another protein synthesis inhibitor (data now shown). In summary, the data indicates that FAM30A repeats enhance RUNX1C protein isoform levels, in alignment with higher MSI2 protein levels. Although RUNX1A protein levels were moderately higher, the increased overall RUNX1 total protein levels in this condition can be explained by MSI2's high binding affinity to RUNX1C and, to a lesser extent, RUNX1A.

To investigate how endogenous RUNX1 activity is increased via the regulatory axis comprising FAM30A repeats and its partner MSI2, two different luciferase reporter constructs were generated and transfected into K562 cells. These were combined with stable K562 cell clones containing constructs for MSI2 knockdown and overexpression of FAM30A repeats (Figure 36A). This leukaemia cell line model was chosen because the luciferase intensity was more readily quantifiable compared to the low transfection rates observed in KG-1a cells. The first luciferase construct comprised palindromic RUNX1 binding motifs (TGTGG) in the minimal promoter region of a firefly luciferase gene to confirm whether overall RUNX1 activity is heightened. The other luciferase system consisted on the RUNX1 promoters (either P1 or P2) cloned next to the luciferase gene. Each of the P1 and P2 promoters lead to the production of RUNX1C and RUNX1A/B respectively. Moreover, they are known to bind to the promoter regions to control their own transcription (Martinez et al., 2016). The purpose of this construct was to elucidate whether increased RUNX1 activity was also connected to RUNX1 promoter usage. First, it was demonstrated via western blot analysis that MSI2 knockdown (shM2)

had a modest impact on RUNX1 protein levels, showing an approximate 20% reduction compared to the control (CTRL). These levels were restored to their original state when FAM30A repeats were overexpressed under this condition (shM2 OE). Furthermore, it was observed that overexpression of FAM30A repeats (OE) led to increased protein levels of both RUNX1B and RUNX1C (Figure 36B). The quantification of RUNX1A protein levels was challenging as its intensity was not easily detectable. Importantly, these results were validated by luciferase-based cell assays, which revealed that overexpression of FAM30A repeats in K562 cells led to a notable increase (approximately 30-50%) in RUNX1 transcriptional activity.

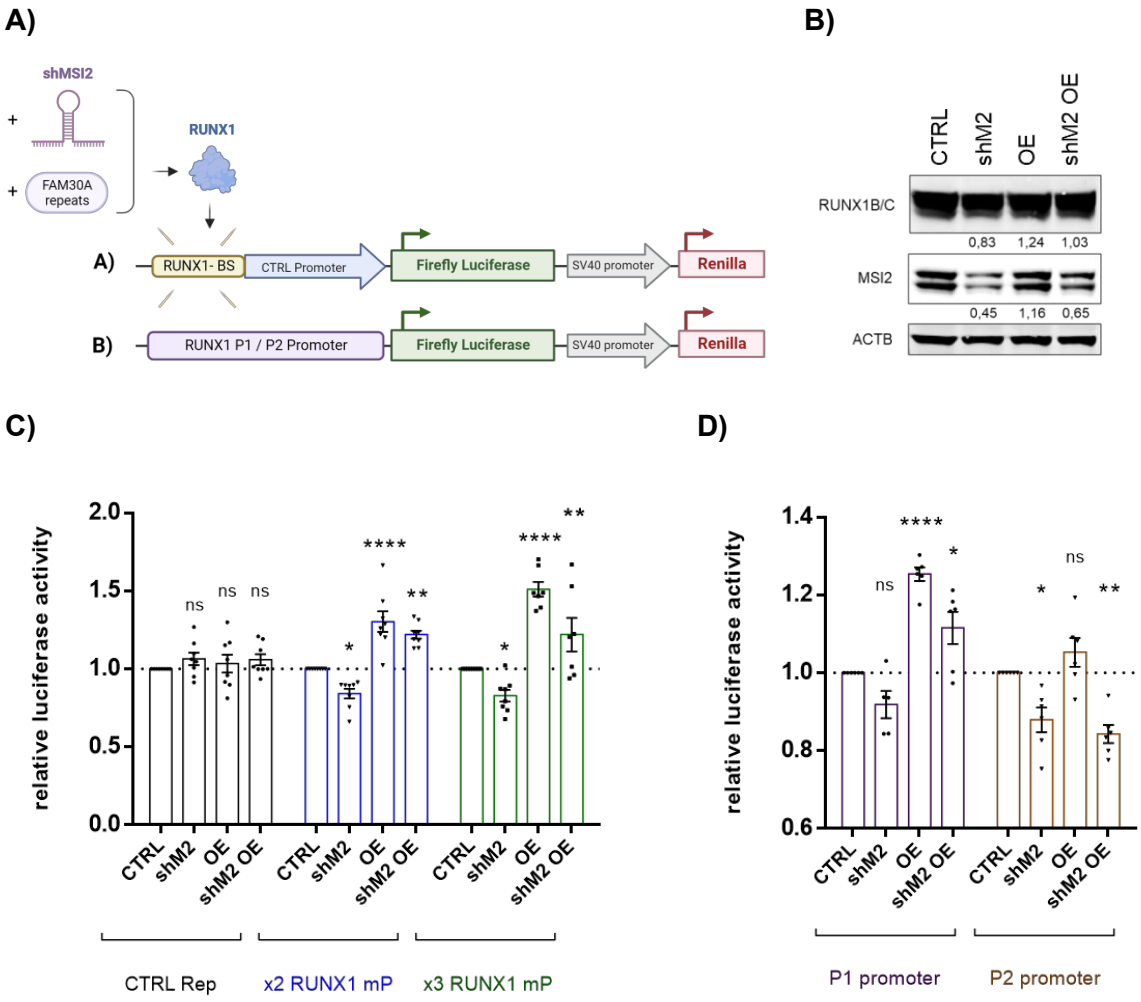


Figure 36. The increased RUNX1 transcriptional activity is FAM30A- and MSI2-dependent. **A)** Schematic depiction of the luciferase reporter constructs transiently transfected into K562 cells together with MSI2 knockdown and overexpression of the repeats and analysed after 48 hours. The upper luciferase construct (A) contains besides a minimal promoter (CTRL promoter), two (x2 RUNX1 BS) or three (x3 RUNX1 BS) RUNX1 binding sites cloned in the promoter region of the firefly luciferase. The lower constructs (B) contains the P1 or P2 promoters of RUNX1.
-Continued on next page-

The Renilla luciferase was used for normalization. **B)** Representative western blot of K562 stable cells transduced with control (CTRL), MSI2 knockdown (shM2), overexpression of FAM30A repeats (OE) and MSI2 knockdown with overexpression of FAM30A repeats (shM2 OE) constructs after 48 hours. ACTB was used as a loading control (n=3). **C)** Firefly luciferase activity was determined in K562 control (CTRL), MSI2 knockdown (shM2), overexpression of FAM30A repeats (OE) and MSI2-knockdown combined with overexpression of FAM30A repeats (shM2 OE) transduced with the three constructs to measure RUNX1 activation. Relative luciferase activity values from each condition were normalized to CTRL Rep and CTRL conditions. **D)** Firefly luciferase activity was determined in K562 cells transfected with RUNX1 promoters. Statistical significance was calculated using unpaired two-tailed student t-test OE vs CTRL, shM2 vs CTRL, shM2 OE vs shM2 normalised to each timepoint (****- p<0,0001; *** - p<0,001; **- p<0,01; *- p<0,05). Data is displayed as mean values and error bars represent SEM.

This was accompanied by a significant increase on activation of the P1 promoter but not the P2 and a mild upregulation RUNX1B and RUNX1C mRNAs (Figure 36C, D; Supplemental Figure 9). On the other hand, MSI2 depletion significantly decreased RUNX1 activation followed by a considerable decreased on the P2 promoter activity. RT-qPCR analysis showed a significant increase on RUNX1A mRNA but an overall moderate decrease on total RUNX1 mRNA. Decreased RUNX1 activation was considerable enhanced (56% more compared to MSI2 knockdown) through FAM30A repeats overexpression (shM2 OE) even though only P1 promoter activity was significantly elevated compared to MSI2 knockdown. This indicates that not only FAM30A repeats directly affect RUNX1C protein stability via MSI2, it also eventually enhances the activation of the P1 promoter of the RUNX1 gene that leads to production of RUNX1C isoform.

3.4.5.- Impact of FAM30A repeats on a therapeutically targetable AML LSC pathway

As mentioned above, modulation of the expression of FAM30A and its repeats greatly influences certain pro-survival AML LSC pathways, including the IL-2/ STAT5 pathway, which has been shown to be therapeutically targetable in AML (Wingelhofer et al., 2018). Particularly, overexpression of the repeats (OE, REC) was positively correlated with activation of this pathway (Figure 19C). Interleukin-2 receptor subunit alpha (IL2RA), also known as CD25 was among the massively upregulated mRNAs upon overexpression of the repeats (log FC 5.44, P = 0.00387). CD25 is an aberrant marker expressed on the surface of certain subsets of AML LSCs but not on normal HSCs (P. Liu et al., 2018) and its expression predicts adverse outcomes in AML patients (Fujiwara et al., 2017). Given that CD25 is transcriptionally controlled by STAT5, elevated CD25 expression in AML subpopulations subsets is indicative of STAT5 activation (Z. Guo et al., 2014). Thus, it was further investigated whether a correlation exists between the upregulation of CD25 and the activation of STAT5, as phosphorylation at Tyrosine 694 is essential for STAT5 activation (Gouilleux-Gruart et al., 1997; Matsumura et al., 1999). Additionally, it was examined whether concomitant suppression of CD25 and STAT5

selectively affects cell viability in KG-1a stable cell lines with overexpression of FAM30A repeats (Figure 37). As depicted in Figure 37A, overexpression of FAM30A repeats (OE) led to a substantial increase in CD25 protein expression (approx. 40% of the total population) compared to control cells (CTRL), where surface expression levels were not readily detectable. Noteworthy, activation of STAT5 was positively correlated with overexpression of the repeats (OE, REC) as phosphorylated-STAT5 was greatly increased when compared to their respective controls (CTRL, KD) while total STAT5 protein levels were slightly increased (Figure 37B).

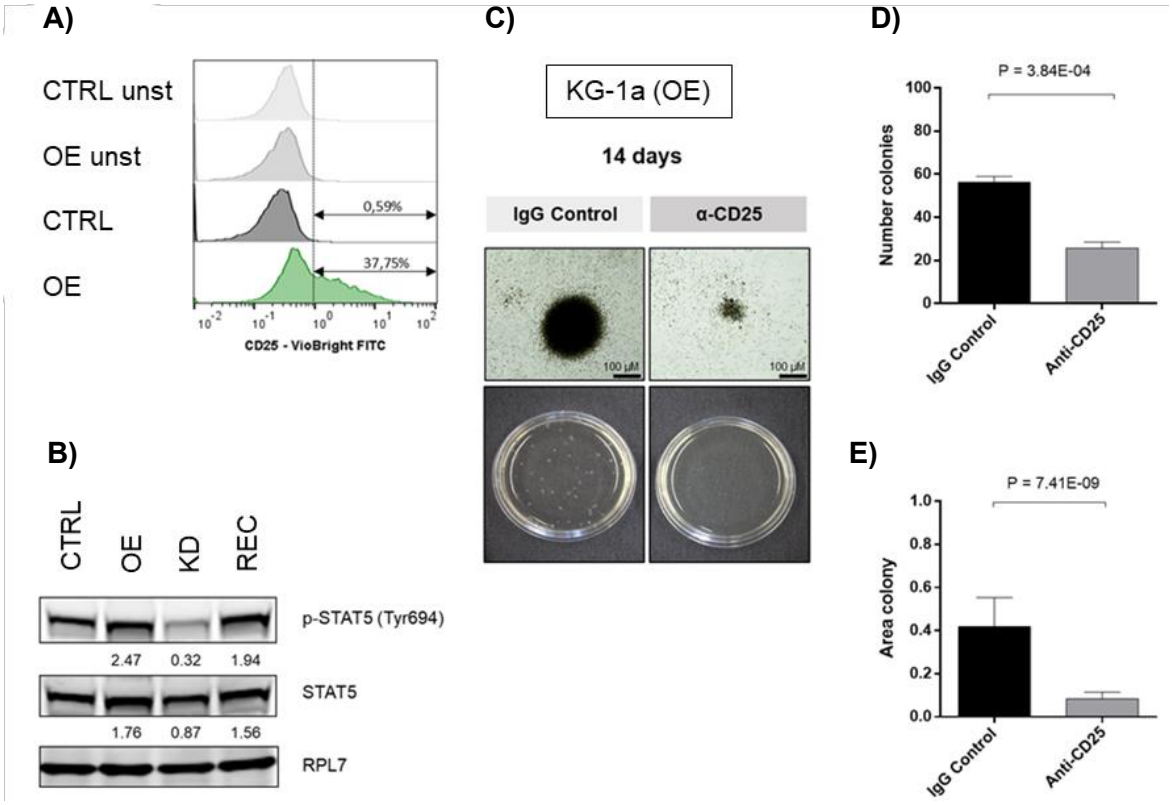


Figure 37. Overexpression of FAM30A repeats influences the therapeutically targetable IL2-STAT5 pathway by upregulating CD25. A) FACS analysis showing histogram of percentage of CD25-positive cells from KG-1a control (CTRL) and FAM30A repeats overexpression (OE) stable cell lines after 10 days in liquid culture and compared with their respective unstained (unst) controls. **B)** Western blot analysis of KG-1a stable cell lines for phosphorylated STAT5 at tyrosine 694 (p-STAT5 (Tyr694)) and total STAT5 protein. RPL7 was used as a loading control. n=2 **C)** Microscopy images of the KG1-a OE stable cell lines grown in methylcellulose after 14 days, which were pre-treated with 3 µg/mL of IgG and anti-CD25 (α-CD25 or BC-96) antibodies for 48 hours in liquid culture. **D)** Quantification of number of colonies from C). **E)** Area of the colonies in mm^2 from C). Data is displayed as mean values and error bars represent SEM.

Since it was previously demonstrated that overexpression of FAM30A repeats increased total CFC outputs compared to the control (Figure 22), it was further addressed whether targeting CD25 in this population would affect clonogenic potential. Recently, newly developed antibodies targeting CD25 have been shown promising results *in* vivo in specifically depleting CD25+ AML blasts (Flynn & Hartley, 2017; Pousse et al., 2023).

After prior incubation with α-CD25 (BC-96) for 48 hours, a striking decrease in CFC outputs was observed, not only in number of colonies (approx. 54% less) but also in the size of the colonies (Figure 37C, D, E) compared to a non-targeting control antibody (IgG). Accordingly, activation of STAT5 was assessed at the protein level after treatment with α-CD25 and compared to a STAT5 inhibitor (STAT5i) that targets the SH2 domain and disrupts STAT5 activation, dimerization, nuclear translocation, and STAT5-dependent downstream gene transcription (IC_{50} value extracted from (Wingelhofer et al., 2018)). Preliminary results showed that treatment with α-CD25 in the OE population mimicked the effects seen with the STAT5i at effectively decreasing phosphorylated levels of STAT5 (Supplemental Figure 10) when compared to control cells. However, STAT5i also impacted the total STAT5 protein levels in both cell lines, indicating a likely more robust overall effect. Thus, analysis of downstream effects from targeting this pathway by both inhibitors in this context should be performed as well as further experimental validation is required.

Altogether, these results point towards a potential therapeutically targetable pro-LSC pathway that it is activated through FAM30A and its repeats.

3.5 Potential clinical relevance of FAM30A in AML

Based on the findings presented here, substantial evidence is provided demonstrating that FAM30A exhibits oncogenic potential within the context of AML. However, it is noteworthy that there are currently no reported *in vivo* studies examining the impact of FAM30A on tumour growth in any cancer type. Further exploration is needed to investigate whether FAM30A influences the initial engraftment of AML LSCs cells in the bone marrow, a crucial aspect of LSCs that is essential for the subsequent progression of AML.

3.5.1.- PRELIMINARY RESULTS – Effects of FAM30A and its repeats in regulating leukaemic engraftment *in vivo*

In this study, a transgenic mouse model (in collaboration with Prof. Dr. David Vetrie from University of Glasgow) was used to investigate whether depletion of FAM30A (KD) in AML cells is able to suppress leukaemic engraftment of human AML LSC-like cells (KG-1a cells). In addition, we included control cells (CTRL) and FAM30A-depleted cells with overexpression of the repeats (REC) in order to examine if leukaemic engraftment potential is restored with FAM30A repeats.

After growing for 5 days in liquid culture, the stable KG-1a cell lines were trasplanted into immunocompromised mice and bone marrow and spleen was harvested after 30 days in order to analyse engraftment of leukaemic human cells in the bone marrow (Figure 31A). After quantification of APC+ positive human cells (as a mean of percentage of engrafted KG-1a human cells) in the bone marrow, no apparent engraftment of leukaemia cells was observed upon FAM30A depletion (KD) which was comparable with mice which had not been engrafted with fluorescently-labelled human cells (Supplemental Figure 11).

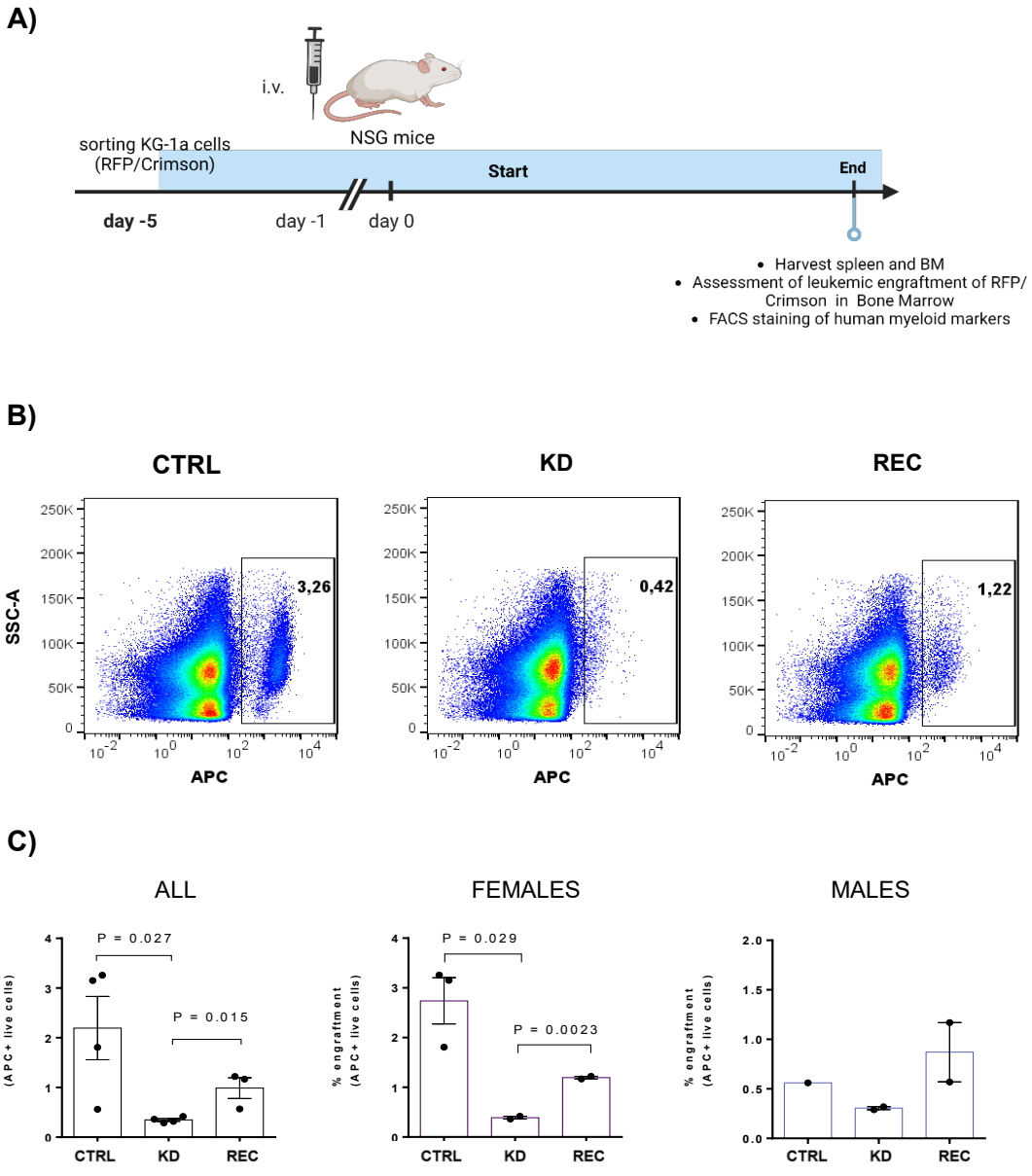


Figure 38. FAM30A depletion impedes leukaemic engraftment of AML LSC-like cells *in vivo* and it is partially restored upon overexpression of FAM30A repeats. A) Schematic depiction showing experimental design followed in the *in vivo* experiment performed in immunodeficient mice with KG-1a cells.

-Continued on next page-

B) Representative FACS images showing leukaemic engraftment in bone marrow after 30 days in KG-1a stable cell lines. Gating of APC-positive cells is based on mouse which have not been engrafted with fluorescent cells. **C)** Graphs showing quantitative analysis of leukaemic engraftment based on percentage of APC-positive cells in mouse bone marrow. Statistical significance was calculated using unpaired two-tailed student t-test. Data is displayed as mean values and error bars represent SEM. NSG: NOD scid gamma mouse.

However, engraftment was partially recovered – though 50% less than control- in cells which had overexpression of FAM30A repeats (REC) thus indicating again an oncogenic role of this region of the lncRNA in promoting leukaemic engraftment (Figure 38B,C). Clear changes in leukaemic engraftment were observed upon modulation of expression of FAM30A and its repeats. It was then further investigated whether the immunophenotype of the AML cells engrafting in the bone marrow across the three different cell populations correlates with worse engraftment potential (Figure 39).

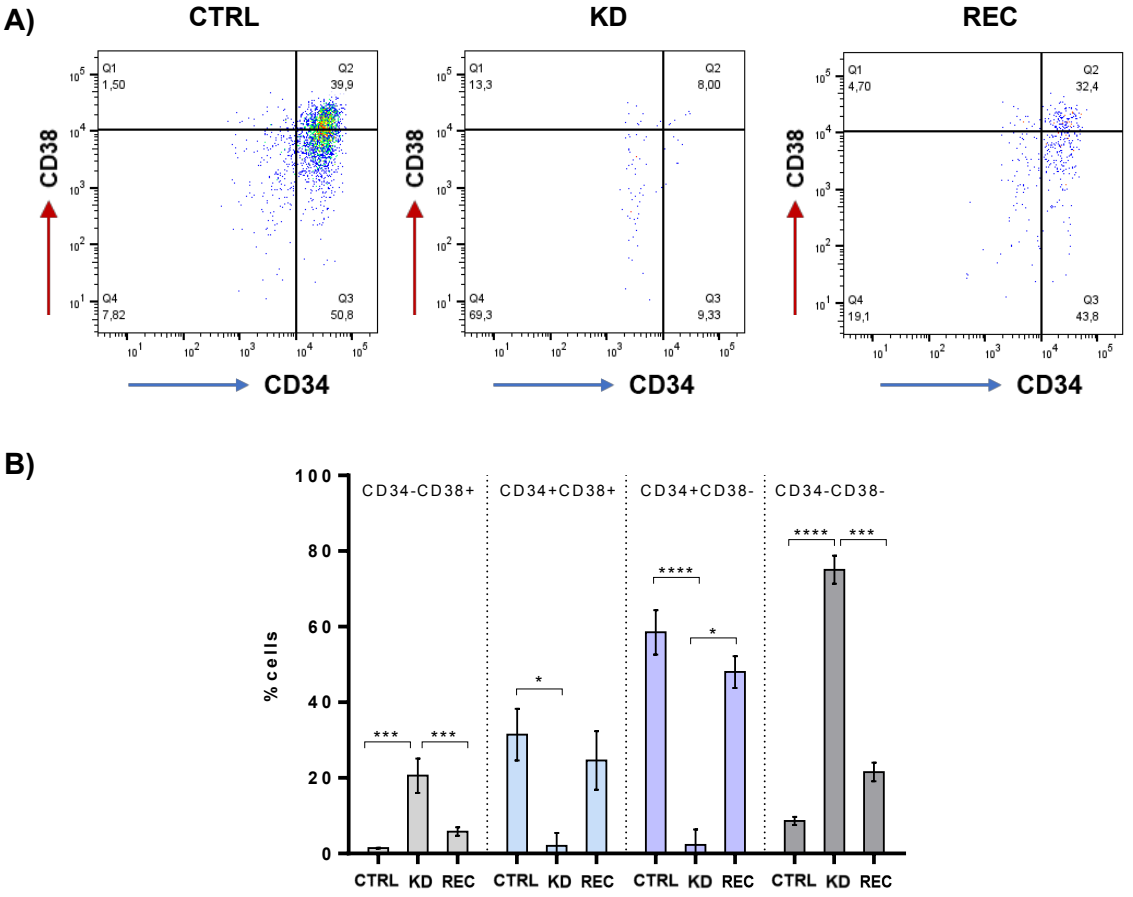


Figure 39. FAM30A depletion eradicates the engrafting CD34+CD38- population, with partial recovery observed upon the overexpression of FAM30A repeats. A) Representative FACS staining images for analysis of CD34 and CD38 markers in APC-positive cells in mouse bone marrow of the three cell populations trasplanted to immunocompromised mice. **B)** Quantification of percentage of cells from each of the quadrants from A). Statistical significance was calculated using unpaired two-tailed student t-test (****- p<0,0001; *** - p<0,001; **- p<0,01; *- p<0,05). Data is displayed as mean values and error bars represent SEM.

Although some reports have questioned this hypothesis, it is generally accepted that the predominant subset of human AML leukaemia-initiating cells are capable of engrafting in

SCID mice and forming abundant colony-forming progenitors that can differentiate into leukaemic blasts are identified as CD34+ CD38- (Blair & Sutherland, 2000; Bonnet & Dick, 1997; Lapidot et al., 1994).

The immunophenotype of KG-1a cells is CD34+ CD38- (see Figure 21C) and FACS analysis of human APC+ cells in the mouse bone marrow show that most of the engrafting population in control cells are CD34+CD38- (approx. 60%) and their more differentiated progeny CD34+CD38+ (30%) (Figure 39). The percentages of CD34+ cells in both populations are severely diminished upon FAM30A depletion as the few engrafting cells were mostly CD34- (approx. 80%), probably non-stem AML cells which can explain the significantly lower percentage of engraftment. Notably, it was observed that the more primitive CD34+ engrafting population is partially recovered upon overexpression of FAM30A repeats in FAM30A-depleted cells (REC) which indicates that overexpression of the repeats is directly correlated with better leukaemic engraftment thus highlighting the oncogenic potential of this region.

Please note that the experimental *in vivo* design is currently undergoing optimization, involving larger cohorts of mice, an increased number of cells injected per mouse, and an extended period of engraftment. This optimization aims to conclusively validate whether FAM30A depletion effectively eliminates leukaemic engraftment. Furthermore, the OE population would have to be included to assess whether the overexpression of FAM30A repeats induces a more aggressive phenotype.

4. DISCUSSION

One of the major factors contributing to the unfavourable prognosis of AML patients is the high disease relapse rate after following intense chemotherapy regimens. This recurrence is believed to be driven by the persistence of a subpopulation of chemotherapy-resistant leukaemia stem cells (LSCs), acting as the disease reservoir. Deciphering the expression and functionality of LSC-specific deregulated genes has marked a pivotal breakthrough in AML research as it enables the identification and targeting of LSCs, potentially leading to notable clinical improvements for AML patients (Elsayed et al., 2020; Eppert et al., 2011; S. W. K. Ng et al., 2016). Recently, aberrant long non-coding RNA (lncRNA) expression profiles and certain lncRNA signatures have been associated with clinical outcome in both paediatric and adult AML samples. Moreover, high expression of some of these lncRNAs was significantly correlated with LSC signatures and was proven *in vitro* and *in vivo* to be relevant for LSC activity (Al-Kershi et al., 2019; Bill et al., 2019; H. Luo et al., 2019).

In this particular context, FAM30A or KIAA0125 emerges as a prominent deregulated lncRNA with oncogenic potential in AML (Figure 8C). Although the precise molecular function in this disease has not been investigated, prior seminal studies have reported its prognostic role in other cancer types and diseases. For instance, upregulation of FAM30A was reported in ameloblastoma -a rare type of jaw tumour derived from the teeth- (Diniz et al., 2019) and laryngeal squamous cell carcinoma (Y. Shi et al., 2021). However, FAM30A has also been proposed as a tumour suppressor gene in diabetic pancreatic cancer (Yao et al., 2017), pancreatic ductal adenocarcinoma (Akrami et al., 2023; L. E. Shi et al., 2020), lung adenocarcinoma (G. Wu et al., 2021) and colorectal cancer (Yang, Zhao, et al., 2019). The latter one was aligning with results from this thesis where colon tissue had relatively high FAM30A expression levels (Figure 9A). This is likely because the colon tissue exhibits high degrees of infiltration with immune cells, a cell type where high FAM30A expression was observed. Nevertheless, FAM30A was also suggested to sponge hsa-miR-29b-3p, regulate BCL2 expression and influence PI3K-Akt signalling in colorectal cancer development (Yang, Zhao, et al., 2019).

FAM30A shows a tissue-specific expression pattern in the immune system, with high levels observed in the spleen (where its sequence was first described by Nagase et al. in 1995) and the thymus (Figure 9A). Consequently, the tissue and cell-type specific expression and other functional evidence strongly support the notion that immune-related lncRNAs possess unique features within the AML immune microenvironment (M. Zhang et al., 2024). This underscores the relevance of investigating the molecular role

of immune-related lncRNAs with differential expression in AML (i.e. FAM30A) which could uncover potential therapeutic targets and prognostic markers.

High FAM30A expression has also been correlated with inflammatory and chronic immune responses (D. S. de Lima et al., 2019; Sentis et al., 2023) and deregulated FAM30A expression has been found in other haematological malignancies apart from AML, such as MDS (Y. H. Wang et al., 2021) and paediatric acute lymphoid leukaemia (Alsuwaidi et al., 2021). This hypothesized oncogenic function of FAM30A in the haematological context aligns with the findings presented in this thesis, which demonstrated elevated FAM30A expression levels compared to bone marrow samples from healthy donours, with the highest levels observed in AML patients (Figure 10A). As part of a prognostic cuproptosis-based lncRNA signature for AML, only one study has preliminary explored the function of FAM30A in the AML context but the results warrant extensive further experimental validation (T. Zhang et al., 2023). Overall, the biological role and clinical relevance of how FAM30A regulates LSC activity as well as it contributes to AML progression remain largely inconclusive.

4.1.- FAM30A, a highly prognostic marker that correlates with LSC activity and worse clinical outcome in AML patients

Only bioinformatic analyses have investigated the potential implication of FAM30A as adverse risk factor so far in AML (Y. H. Wang et al., 2021). Even though high LSC frequencies are well known to be associated with worse clinical outcome in AML patients (Ng et al., 2016), the results herein presented confirmed FAM30A´s potential as an independent LSC-specific prognostic factor in predicting poor patient survival (Figure 10B) which aligns with the significantly higher FAM30A expression levels in patients with unfavourable karyotypes such as complex and EV1 or FLT3-ITD mutations (Figure 11A, B and Supplemental Figure 1). Interestingly, the prognostic relevance of FAM30A-asocciated gene mutations in AML patients is also confined to the stem cell compartment. The proto-oncogene ectopic viral integration site-1 (EVI1) encodes a transcription factor with stem cell-specific function essential for HSC fate and mutations in the EVI1 gene are commonly associated with poor response to treatment (Birdwell et al., 2021; Hinai & Valk, 2016). FLT3-ITD is a driver mutation that induces constitutive activation of STAT5 pathway and contributes to malignant transformation of HSCs by enhancing its self-renewal capacity and block in differentiation (Choudhary et al., 2007). In this regard, it would have been interesting to profile via single-cell RNA-sequencing (scRNA-seq) how FAM30A^{high} AML LSCs behave post-chemotherapy treatment and whether clonal

mutations are selectively acquired in this subpopulation. In this line, further information could have been extracted by sorting functionally validated FAM30A^{high} CD34+CD38- LSCs fractions from AML patients with specific mutational signatures as the data extracted from (Eppert et al., 2011) are only immunophenotype and functionally validated LSC fractions but the mutational status is not provided. This might have reinforced the suggested oncogenic function of FAM30A within distinct subpopulations of leukaemia-initiating cells, prompting the selection of a more tailored AML LSC-like cell line model to study FAM30A. On the other hand, expression of FAM30A was significantly lower in patients with more differentiated and less aggressive AML subtypes such as inv(16), t(15;17), NPM1 mutated or t(8;21), with particular interest drawn to the latter, a topic that will be further discussed.

After performing a comprehensive analysis of AML LSC-specific upregulated genes, high FAM30A expression levels were consistently observed in established gene signatures used for the isolation and identification of AML LSCs fractions (Figure 8B, C) (Elsayed et al., 2020; Eppert et al., 2011; N. Jung et al., 2015, 2015; S. W. K. Ng et al., 2016). Based on RNA expression data extracted from (Eppert et al., 2011), the highest FAM30A expression levels were significantly correlated with the experimentally defined and functional CD34+ CD38- fraction and slightly lower in the CD34+ CD38+ LSCs fraction. Both subpopulations are where most AML LSCs or leukaemia-initiating cells reside (~75%) (Bonnet & Dick, 1997; Eppert et al., 2011; Sarry et al., 2010) and FAM30A expression levels were the lowest in the CD34- fraction (Figure 12A, 5B). Despite the uncertainty about the nature of the clonal relationship between CD34+ and CD34- LSCs, it is generally accepted that the CD34- fraction originates downstream from the more immature CD34+ population, where the pre-leukaemic clone (pre-LSC) is even exclusively present (Quek et al., 2016). Moreover, the CD34- compartment of LSCs has been shown to be more therapy-sensitive, less immunogenic and leukaemogenic when trasplanted into the bone marrow niche of immunodeficient mice (Ishikawa et al., 2007). This implies a potential function of FAM30A within the more immature population of AML LSCs that exhibit higher self-renewal and chemoresistance potential (Eppert et al., 2011; Sarry et al., 2010). This also aligns with the highest expression profile of FAM30A in M0 and M1 subtypes of AML which are known to be present a differentiation blockade occurring at early myeloid stages thus their immature myeloid features (Figure 11C). Consistently, conventional chemotherapy regimens yield disappointing results in the M0 subtypes possibly due to the high frequency of chromosome abnormalities and complex karyotypes (Stasi et al., 1994; Venditti et al., 1997; Yokose et al., 1993).

High FAM30A expression levels are also seen in the most immature population of normal HSCs and progressively decreases upon differentiation of immune cells (Figure 12C). FAM30A expression also positively correlates with stem cell gene signatures -either HSC or LSC- and negatively correlates with differentiation of myeloid cells (Figure 12D), even though it is important to note that normal HSCs and LSCs share stem cell transcriptional programs which is potentially undistinguishable (Bonnet & Dick, 1997; Eppert et al., 2011). However, GSEA analysis upon stable up- or downregulated FAM30A expression in AML LSC-like cells, not only showed a significant positive correlation with changes in LSC-related specific gene profiles such as LSC_R (comprising genes more highly expressed in LSC fractions) but also correlated with specific HSC genes that are differentially expressed in LSCs (termed CE_HSC_LSC (Eppert et al., 2011)) (Figure 19B). Remarkably, FAM30A was part of both gene signatures from Eppert et *al* (Figure 8B). This implies that FAM30A could play a potential differential role in the malignant transformation of HSC to LSC: while it could be regulating HSCs maintenance and normal haematopoiesis, aberrant FAM30A activation could promote the emergence of AML LSCs and subsequent onset of the disease. This hypothesis has been validated for other LSC-specific lncRNAs in the AML context. For instance, HOTTIP activation was shown to abnormally alter the self-renewal capacity of HSCs via activation of HOXA9 and the Wnt/β-catenin pathway leading to AML progression (H. Luo et al., 2019). Similarly, the normally HSC-specific HOXA10-AS is an important regulator of maintenance of CD34+ HSCs and progenitor cells via regulation of myeloid differentiation but it is an oncogenic lncRNA in KMT2A-rearranged AML (Al-Kershi et al., 2019).

Based on this hypothesis, the herein presented study could have gained additional insights if FAM30A expression levels had been profiled in normal CD34+ CD38- HSCs from healthy donours, including more myeloid-committed progenitors such as GMP and other mature myeloid cells and compared those with their AML counterparts. For instance, scRNA-seq performed in primary AML samples (adult and paediatrict) and normal bone marrow could have better delineated whether FAM30A^{high} AML cells follow the same lineage identity as in normal haematopoiesis. Moreover, profiling FAM30A expression in pre-LSCs that are likely to become "primed" during clonal hematopoiesis (i.e. via screening of common gene mutations in epigenetic modifiers) could provide additional insights into the relevance of FAM30A before the emergence of fully competent AML LSCs. This would have also clarified whether these subpopulations of FAM30A^{high} AML LSCs give rise to most of their leukaemic progeny (AML blasts), even post-chemotherapy treatment. In fact, another set of samples that could be included in this

scRNA-seq setting would be analyzing FAM30A expression levels in patients before relapse and comparing them with short and long-term relapsed and primary refractory samples. In this line, by stratifying these immature FAM30A^{high} chemoresistant-AML populations and correlating them with their acquired mutations at relapse could represent a strategy to predict whether patients would poorly respond to chemotherapy.

This would have overall strengthened the hypothesis that FAM30A is differentially required in AML LSCs compared to normal HSCs. Despite these limitations, the significance of the stem cell-specific lncRNA FAM30A was addressed as a potential oncogenic driver in AML thus it was selected for further experimental characterization in AML cells.

4.2.- Biological and molecular properties of FAM30A and its highly conserved region

The FAM30A gene is located on chromosome 14 of the human genome (Figure 13A). Interestingly, chromosomal rearrangements (i.e. duplications, deletions or translocations) involving the genomic region where FAM30A is located -14q.32.33- are known to result in a predisposition to a number of adult-onset myeloproliferative neoplasms such as AML and it is associated with immature M0 and M1 acute myeloid leukaemias (Di Giacomo et al., 2021; Venditti et al., 1997). Another study also identified recurrent FAM30A amplifications more prevalent in ALL patients with gene fusions (Batista-Gomes et al., 2020). Other numerous haematological malignancies that were correlated with abnormalities within the 14q32 region include chronic B-cell leukemia (Bryant et al., 2023; Reindl et al., 2010) and multiple myeloma (Abaza et al., 2015; E. Tian et al., 2014).

As previously shown, FAM30A expression was relatively high in B-cells compared to other mature blood cells (Figure 12C) and correlated with prevalence of B-cell leukaemia. This could be explained by its genomic vicinity to immunoglobulin genes (FAM30A gene is in antisense orientation to the *IGH* locus) necessary for the production of the heavy chains of the human antibodies, predominantly generated by B-cells (D. Jung et al., 2006; Meffre et al., 2001). Moreover, FAM30A expression has been previously associated with B-lymphocyte-related genes as high FAM30A expression was correlated with high antibody production upon influenza vaccination (D. S. de Lima et al., 2019). Given that FAM30A expression is significantly higher in B-cell acute lymphoblastic leukaemia (B-ALL) compared to healthy bone marrow (Figure 10A), this poses the question of whether FAM30A has the same oncogenic significance in this context or if its

transcript is upregulated due to its genomic localization. Consequently, the potential oncogenic role of FAM30A in B-cell leukaemias could also represent a promising area of future research.

In order to study FAM30A at the molecular level, the KG-1a cell line was selected as it robustly resembles different AML LSC features: highly correlated expression with the LSC17 gene signature (Figure 14A), the CD34+CD38- immunophenotype (Figure 14C), as well as it belongs to the group of M0/M1 immature AML subtypes (where FAM30A also displayed the highest expression levels) and complex karyotype with high chemoresistance properties. In addition, FAM30A expression levels in KG-1a cells were the highest in a panel of AML cell lines (Figure 14B). Noteworthy, FAM30A sequence was first described as a cDNA expressed in the parental KG-1 cell line (M1 subtype) (Nagase et al., 1995). Determination of exon usage via RT-qPCR and RNA-sequencing analysis in KG-1a cells as well as in CD34+ cord blood cells showed a total of 5 exons (Figure 15, Supplemental Figure 2). This does not include an initial annotated exon which spans almost 800 bp, suggesting a different 5´ region not present in haematopoietic cells. A plausible reason would be that transcription of the FAM30A transcript starts earlier from an upstream promoter and this accounts for another less constitutively cell-type dependent expressed isoform. FAM30A may also be transcribed from various potentially complex promoters, such as bidirectional promoters that regulate other genes (i.e. genes transcribed from the IGH loci) and promoters located within the body of another gene. Thus, the precise identification of the FAM30A transcription start site (TSS) would have enabled the design of CRISPRa approaches to accurately perform targeted FAM30A gene activation (Chavez et al., 2015; Konermann et al., 2014). However, this strategy is likely to cause undesired perturbations of the neighboring genes, which may also lead to false identification of phenotypes (Goyal et al., 2017).

Regarding the conservation across various species, FAM30A emerges as a lncRNA present in primates while absent in rodents (and no sequences from individual exons aligned with any gene). Across the FAM30A sequence, a particularly intriguing region is located at the terminal part of exon 5. Within this region, there exists a series of putative RNA-binding protein (RBP) motifs, organized in a homologous segment spanning 76-78 base pairs which it is tandemly repeated six times. Throughout this thesis, this segment was identified as FAM30A repeats or simply repeats (as depicted in Figure 13B). Interestingly, this short tandem repeated region was already identified almost 30 years ago (Nagase et al., 1995) within the untranslated region of FAM30A but no function has been reported so far. Short tandem repeats have been structurally characterized by a sequential arrangement of repeat units, a large number of which lie within genes and

their regulatory regions. These genomic repetitive elements have been widely investigated since they play critical roles across the genome by driving evolution and regulating gene expression (X. Liao et al., 2023; N. M. Shah et al., 2023). Thus, they are an importance source of genetic variation with high polymorphisms rates that induce genome variability and they have also been associated with pathological conditions (Hannan, 2018; Ninomiya & Hirose, 2020). In particular, repeat-derived RNA sequences embedded in introns, untranslated regions or lncRNAs represent RNA functional elements important for protein-binding (Onoguchi et al., 2021) and functions of many lncRNAs depend on recruitment of multiple copies of specific RBPs to short tandem repeats. However, the role of these repeat-containing RNA sequences has only gained some attention in the recent years. For example, the AML-relevant lncRNA XIST comprises several tandem repeat regions relevant for recruiting transcriptional repressors and chromatin modifiers such as PRC2 or EZH2 (Elisaphenko et al., 2008; McHugh et al., 2015). However, the idea that such binding is essential for their function has been recently challenged, as many of these reported interactions may not occur *in vivo* (J. K. Guo et al., 2024).

Notably, the FAM30A repeats exhibit a distinctive feature. The present study implies that there exist a substantial evolutionary pressure to specifically increase the number of repeats (but not its flanking regions) over time, with the highest repeat counts of six observed in humans. These units appear to have resulted from a tandem sequence duplication that occurred at least more than 30 million years ago as the existence of at least one repeat is present for the first time in the order of higher primates (Anthropoidea) (Figure 16A, B) (Groves, 1998). Moreover, each of the repeats was predicted to form stem-loop structures (Figure 16C) which could serve as functional tandem domains for RBP binding, as other studies have reported before (Raposo et al., 2021; Tichon et al., 2016; Yap et al., 2018). This finding is compelling as regions from antisense lncRNAs that are enriched for translated ORFs or protein-RNA interactions are generally more conserved (Ruiz-Orera & Albà, 2019). However, no potential ORF within this particular region was detected, emphasizing the significance of this untranslated region for protein binding and potentially broader biological functions.

Even though FAM30A is annotated as a non-coding RNA and computationally predicted to be non-coding RNA by the PhyloCSF, CPAT and CPC2.0 tools gave it borderline scores likely due to the presence of potential ORFs within FAM30A sequences. A short open reading frame (sORF) was identified *in silico*, predicted to code for 168 amino acids (17 kDa) at the beginning of exon 5 (data not shown). The protein query did not align with any hypothetical proteins (UniProt) in other primates but it would be worth

investigating the coding potential of this protein and whether it has any function in the cell. In this context, one study that correlated FAM30A expression with neuronal degeneration in Alzheimer disease identified another sequence in FAM30A that could encode for a protein of 8 kDa (LOC9834) but no function has been reported so far (Uhrig et al., 2009). Interestingly, the gene encoding for this hypothetical protein was found to be hypomethylated in pancreatic cancer (Tan et al., 2009) thus suggesting potential oncogenic functions of FAM30A-derived peptides.

In principle, in line with the almost exclusive cytoplasmic localization of FAM30A at steady-state conditions, the nature of a cytoplasmic RNA would imply an increased chance of association with ribosomal complexes and ultimately being translated. Experimental validation in AML cell lines demonstrated that FAM30A resides in fractions not associated with heavy polysomes, where mRNAs typically reside. This provides the first evidence that FAM30A lacks protein-coding potential under steady-state conditions (Figure 17). However, this does not exclude the possibility that ribosomes transiently bind to FAM30A, as a substantial proportion of FAM30A -approx. 25% of the total transcript- was observed in the monosomal fractions (40+60S, 80S) with a minimal presence observed in the heavy polysomes. A possible explanation is that a substantial proportion of certain FAM30A transcripts are protein-coding and thereby bound to ribosomes. Preliminary, *in silico* prediction of lncRNA-encoded peptides using the LncPep tool was performed (data not shown). Several putative micropeptides ranging from 50-150 aminoacids with high coding probability (i.e. via Riboseq, CPC and CPAT) were detected. Very interestingly, increasing evidence suggests that a novel class of lncRNAs-derived micropeptides contribute to cancer development (Setrerrahmane et al., 2022). For instance, it was recently revealed that a lncRNA-derived micropeptide named APPLE (a peptide located in the endoplasmatic reticulum) is overexpressed in various subtypes of AML and confers poor prognosis. This micropeptide was found to be enriched in the ribosomes and regulates translation initiation by promoting PABPC1-eIF4G interaction and eventually supporting a pro-cancer translation program (L. Sun et al., 2021). In this regard, in-depth comparison of ribosome occupancy around putative initiation codons along FAM30A sequence could give further insights about the likelihood to generate FAM30A-encoded peptides. Subsequently, the translation of these putative FAM30A-encoded peptides should be evaluated *in vitro* using fluorescently tagged ORFs in both cultured cells and cell-free translation systems.

Another proposed hypothesis to FAM30A colocalizing with ribosomes is that cytoplasmic lncRNAs have been repored to sporadically engage with both light and heavy polysomes, potentially facilitating their degradation (Carlevaro-Fita et al., 2016) which could explain

the tightly regulated stem cell-specific expression pattern of FAM30A compared to other differentiated immune cells. Moreover, prior studies have suggested that the length of a lncRNA is associated with the number of ribosomes it recruits (Van Heesch et al., 2014), implying that longer lncRNAs like FAM30A are highly prone to spontaneously engage with the translational machinery.

Arguably the most attractive hypothesis is that FAM30A possesses functional roles in regulating translation. In fact, after incubation of *in vitro* transcribed FAM30A repeats with KG-1a cell lysate and subsequent pulldown followed by mass spectrometry, an enrichment of proteins involved in regulating translation was observed. Additionally, clustering analysis revealed the presence of proteins implicated in mRNA processing and stabilization in the cytoplasm (e.g. ATP-dependent RNA helicases) as well as nuclear proteins involved in RNA processes such as transcription or splicing, which were not included in subsequent analyses (Figure 26B, Supplementary Table 1). While the focus of this thesis was the identification of specific protein interactors of FAM30A repeats, there exists considerable potential for identifying numerous other protein binders along the FAM30A sequence. This could be accomplished through methods such as RNAse footprinting assay (i.e. Rfoot seq) which would enable the assessment of which regions of the lncRNA are bound by proteins, thereby shielding them from degradation.

In silico prediction of potential RBP motifs within FAM30A repeats (Figure 27) along with confirmation via mass spectrometry analyses (Figure 26), pulldowns and immunoprecipitation studies (Figure 28) revealed association of HNRNPA1 and MSI2 to FAM30A repeats, which were selected for further studies. Even though HNRNPA1 is a well-known splicing factor with nuclear localization at steady state, this protein has been shown to abnormally shuttle to the cytoplasm in leukaemia cells and its abnormal high expression was found in AML patients (J. Fang et al., 2017; S. Q. Li et al., 2020; Song et al., 2017). After thoughtful consideration, the decision not to explore HNRNPA1 as a potential candidate relevant to FAM30A-mediated function was undertaken for multiple reasons: (1) changes in FAM30A and its repeats does not seem to affect HNRNPA1 protein levels (Figure 29C) and (2) no observable shifts in the subcellular localization of HNRNPA1 were observed that might imply an interaction with FAM30A repeats in the cytoplasm (Figure 29B). In this context, significant changes in reported post-translational modifications (PTMs, e.g. phosphorylation) that would enable HNRNPA1-mediated mRNA transport and translation in the cytoplasm were reported before (Barrera et al., 2020; Clarke et al., 2021; Emdal et al., 2022). Proper LC-MS/MS analyses could here provide additional insights how FAM30A repeats could regulate HNRNPA1 PTMs and thereby its function. The only reported function of HNRNPA1 in the AML context (KG-1a

cells) is the control of post-transcriptional regulation of RUNX1 isoforms via splicing and that contribute to healthy and aberrant haematopoiesis (Davis et al., 2021). (3) Moreover, no significant changes in the RUNX1 mRNA isoform pool were observed upon changes in FAM30A expression in combination with depleted HNRNPA1 that could suggest a cooperative function (Figure 30).

4.3.- Unravelling the FAM30A-MSI2 axis potentially relevant for AML LSC activity

The impact of alterations in FAM30A repeats expression on MSI2 was evident, leading to MSI2 emerging as the prime candidate as the relevance of MSI2 as a key pro-LSC regulator in AML is well established. High expression of MSI2 correlates with worse clinical outcome in AML (Figure 31B), in line with previous studies (Byers et al., 2011; Kharas et al., 2010). Significant alterations in total MSI2 protein levels were observed following changes in the expression of FAM30A repeats across diverse myeloid leukemia cell lines (Figures 33A, B and 36B). In addition, mutations of the three MSI2-binding sites in one FAM30A repeat completely abrogated MSI2 association with FAM30A repeats. A single FAM30A repeat encompasses three potential MSI2-binding sites, characterized by UAG (Duggimpudi et al., 2018; Ruth Zearfoss et al., 2014) or UAG-containing motifs +/– additional flanking nucleotides (C. G. Bennett et al., 2016). While there are three putative MSI2-binding sites in each FAM30A repeat, the close proximity of two binding sites reduces the likelihood of simultaneous binding by two MSI2 proteins. However, this supports the hypothesis that either multiple RBP binding sites or enhanced oligomerization promotes RBP regulation. This was previously reported for other similar RBPs involved in translational control (i.e. IGF2BP1) (Moreno-Morcillo et al., 2020; Nielsen et al., 2004). Furthermore, other MSI2-binding sites (at least 8) were predicted to be distributed throughout exon 5 of FAM30A. Overall, this suggests that a single RNA molecule of FAM30A could have the capacity to potentially bind at least 12 MSI2 proteins. Notably, conservation studies presented in this thesis indicate significant evolutionary pressure to increase the number of FAM30A repeats containing MSI2-binding sites, whereas the flanking regions without these sites show no such pressure (Figure 16). Altogether, this underscores the potential physiological relevance of FAM30A repeats for MSI2 function in human normal haematopoiesis which then, when both dysregulated, can promote malignant processes in myeloid leukaemias.

The relevance and requirement of MSI2´s function in myeloid leukaemias have been thoroughly investigated by multiple research teams through a series of depletion and overexpression studies carried out in both mouse and human systems. MSI2 depletion

or inhibition was shown to strongly downregulate haematopoietic (De Andrés-Aguayo et al., 2011; Hope et al., 2004; Rentas et al., 2016) and leukaemic stem cell-renewal programs (Kwon et al., 2015; Minuesa et al., 2019; Nguyen et al., 2020; Park et al., 2015; Taggart et al., 2016). The potential collaboration between FAM30A and MSI2 in the stem cell compartment appears consistent with our findings from GSEA analysis, revealing that FAM30A depletion significantly correlated with the downregulation of HSC signatures (Figure 12D, Figure 19B), LSC signatures (Figure 19B), and pro-LSC signaling pathways, some of them reported to be controlled by MSI2 in AML, such as TGF-β/SMAD3 (Park et al., 2014) and TNF signalling via NF-kB (J. Zhao et al., 2021). Noteworthy, all of these LSC-specific signatures were significantly restored upon overexpression of the MSI2-interacting region of FAM30A.

MSI2 and FAM30A expression levels are highly correlated in AML samples (Figure 31A) and both show high stem cell-specific expression in normal early HSCs and GMPs, progressively decreasing upon myeloid differentiation (Hemaexplorer; BloodSpot) (De Andrés-Aguayo et al., 2011; Koechlein et al., 2016; Park et al., 2014). Overexpression of the MSI2-interacting region of FAM30A in KG-1a cells resulted on increased cell viability (Figure 27A), a substantial increase of more primitive CFU-GM (Figure 22) and more immature CD34+CD38- cells (Figure 23) compared to control. These findings further support the hypothesis that FAM30A enhances MSI2's pro-stemness function as previous studies have shown that enforced MSI2 expression induces robust self-renewal effects on early HSCs and AML LSCs, as evidenced by the enhanced colony-formation potential of early myeloid progenitors (Hattori et al., 2017; Rentas et al., 2016; Vu, Prieto, et al., 2017). To functionally validate the increased self-renewal capacity of AML cells due to the overexpression of FAM30A repeats, serial replating of colony formation assays should be performed on this cell population. The effects should then be compared to cells overexpressing MSI2 alone and MSI2-depleted cells combined with FAM30A repeats overexpression to further demonstrate the FAM30A/MSI2-dependent mechanism.

The effect on the clonogenic capacity was also consistently observed in MV4-11 cells - FAM30A^{low} AML cell line- (Figure 25) but could not be observed in all AML cell lines. This inconsistency can be attributed to the heterogenous expression of MSI2 and the distinct mutational dependencies among the AML cell lines (Supplemental Figure 7). MV4-11 cells (with FLT3-ITD mutation) have been previously shown to be MSI2-dependent for leukaemic growth as MSI2 depletion greatly diminished the colony-formation potential and concomitantly downregulated expression of oncogenic FLT3 by direct binding to its mRNA (Hattori et al., 2017). Interestingly, RNA-sequencing conducted on KG-1a cells

overexpressing the MSI2-interacting region of FAM30A repeats revealed a significant *de novo* increase in FLT3 mRNA expression (logFC: 5.45; FDR: 2E-48), despite the absence of MSI2-binding to FLT3 mRNA in control KG-1a cells after immunoprecipitation (data not displayed). This discrepancy could be attributed to the latent expression levels of FLT3 in this particular cell line. Given that AML patients exhibiting abnormal MSI2 upregulation are statistically more prone to harbouring the FLT3-ITD mutation (Park et al., 2015), this finding holds significant implications as FAM30A also correlates with this mutational signature, thus suggesting a potential cooperation of FAM30A and MSI2 within FLT3-ITD-mutated AML cells.

Moreover, constituve activation of FLT3 (i.e. FLT3-ITD) has been strongly linked to increased STAT5 activation in response to IL2 stimulation (Choudhary et al., 2007) which is commonly seen in LSCs of high-risk AML patients (Cook et al., 2014). The IL2/STAT5 pathway is therapeutically targetable and its inhibition efficiently targets CD34+ AML primary cells while sparing heathy CD34+ HSCs (Wingelhofer et al., 2018). The results presented in this thesis also reveals a significant correlation between FAM30A depletion and the downregulation of the IL2/STAT5 pathway, a correlation that is reversed upon enforced expression of the MSI2-interacting region of FAM30A (Figure 19C). Furthermore, the overexpression of FAM30A repeats induces *de novo* surface expression of the low affinity IL-2 receptor (IL2RA/CD25), heightened STAT5 activation and a substantial increase in colony-formation potential (Figure 30). Notably, CD25 overexpression has consistently been associated with poor therapy response and adverse outcomes in AML patients (Cerny et al., 2013; Du et al., 2019; Fujiwara et al., 2017; Gönen et al., 2012; Kageyama et al., 2018), prompting ongoing phase I trials to investigate the potential application of CD25-directed therapy in FLT3-ITD+ AML (clinical trials: NCT02588092, NCT00085150; (Flynn & Hartley, 2017)). The increased colony-formation potential observed in $CD25^{high}$ cells due to FAM30A repeats overexpression are significantly mitigated upon specific CD25-targeted inhibition (with concomitant STAT5 inhibition). This aligns with previous findings revealing that CD25 overexpression inhibits differentiation and promotes stem cell-related properties in AML LSCs (i.e. chemoresistance) required for leukaemogenesis (Nguyen et al., 2020). The herein proposed hypothesis is that FAM30A may facilitate MSI2-mediated upregulation of FLT3, leading to enhanced signaling that induces constitutive activation of STAT5. Consequently, this activation results in elevated transcription of CD25 (Z. Guo et al., 2014), establishing a positive feedback loop that confers a proliferative advantage to these cells. Altogether, these findings provide an illustrative example with therapeutic possibilities in the axis between FAM30A/MSI2 and *FLT3*-ITD/CD25/STAT5 signalling

pathway, suggesting their collaborative involvement in promoting agressiveness in AML LSCs.

Inhibiting MSI2 expression and function also greatly decreased proliferation, induced cell cycle arrest and decreased colony-formation potential in various AML cell lines and patient samples (Han et al., 2015; Hattori et al., 2017; Kharas et al., 2010) as well as in AML LSCs (Minuesa et al., 2019; Park et al., 2015) and CML cell lines (H. Zhang et al., 2014). Concomitantly, FAM30A knockdown mimicked the reduced leukaemia growth *in vitro* (Figure 20A) which is contradicted by the observed increase on S phase of the cell cycle (Figure 20B). These results may be misleading due to the broad approach used to analyze cell cycle phases, as DAPI staining alone cannot clearly distinguish between G1 and S phases. To address this limitation, one could consider using lentiviral transduction of fluorescence ubiquitin cell cycle indicator (FUCCI)-based reporter gene systems. These systems enable live and *in vivo* monitoring of cell cycle progression more accurately. However, if validated, this apparent acceleration on cell cycle progression in FAM30A-depleted cells could be explained by the increased lineage commitment in this setting, displaying more CFU-G (Figure 22) and expression of granulocytic marker (e.g. CD11B) (Figure 23) (Schnerch et al., 2012).

Modeling of CD34+ AML LSC in mice has shown that two main population coexist in most patients: the immature CD38- CD45RA+ population (similar to M0/M1 AML subtypes) and their progeny which is resembles a more GMP-like population (Goardon et al., 2011). Results from FAM30A knockdown could be explained with a study that correlated MSI2-knockdown with upregulation of a differentiation signature in GMPs (Kharas et al., 2010). Enforced expression of FAM30A repeats in this setting led to an accelerated differentiation towards the monocyte/macrophage phenotype, as indicated by surface markers (e.g. HLA-DR) and increase on CFU-M (Figures 22 and 23). Given that the parental KG-1 cell line comprises myeloblasts that typically differentiate into macrophages in response to certain cytokines (Koeffler et al., 1980; Skopek et al., 2023), it is proposed that FAM30A, and particularly its repeats, plays a pivotal role in determining cell fate within myeloid-committed LSC progenitors (i.e., GMP). Specifically, FAM30A appears to favor differentiation towards the monocyte/macrophage-committed lineage, while its absence leads to a shift in differentiation towards granulocytes. Nevertheless, further characterization of the immunophenotypes of the populations after FAM30A depletion/overexpression could strengthen this hypothesis (i.e. CD34, CD38 antibody stainings in combination with CD90, CD45RA for LMPP-like and CD123, CD45RA for GMP-like populations).

The findings presented here provide novel evidence that FAM30A plays a direct role in regulating murine bone marrow (BM) engraftment of AML cells. A series of complex interactions between adhesion molecules, cytokines and their receptors influence the “homing”, long-term engraftment (defined as ≥1% leukemic cells in murine BM) (Eppert et al., 2011; Lapidot et al., 1994) and proliferation of CD34+ LSCs in the BM niche to create a malignant niche inhospitable for normal HSCs (Colmone et al., 2008). Although this study did not specifically assess homing, depletion of FAM30A completely abrogated leukaemic engraftment after 6 weeks which suggests that these CD34+ CD38- LSC-like cells have lost their characteristic *in vivo* long-term disease-initating and propagating capacity (Figure 38) (Bonnet & Dick, 1997; Hope et al., 2004; Lapidot et al., 1994). This phenomenon was partially reversed by enforced expression of the MSI2-interacting region of FAM30A which was reflected in the increased number of engrafted CD34+CD38- and their more differentiated CD34+CD38+ AML cells (Figure 32). However, the engraftment levels were not completely restored by the recovery. This is probably explained by the slightly lower levels of FAM30A repeats compared to control (Figure 18B) and that only a small part of FAM30A was used for recovery studies. Nonetheless, additional experiments involving combination of FAM30A and/or MSI2 depletion and overexpression are necessary to validate this hypothesis. Interestingly, the results are consistent with previous experiments that revealed a significant loss in the engraftment potential of MSI2-depleted HSCs and AML LSCs which suggests that MSI2 plays a critical role after BM transplantation and long-term reconstitution (Kharas et al., 2010; Nguyen et al., 2020; Park et al., 2014). The pivotal role of FAM30A in the engraftment capacity of this population is highly relevant since AML LSCs predominantly inhabit the more immature, chemotherapy-resistant CD34+CD38- AML LSC subset. Moreover, these cells have been shown to exhibit enhanced self-renewal and engraftment capabilities in xenotransplantation assays (Eppert et al., 2011; Ishikawa et al., 2007; Sarry et al., 2010). Future *in vivo* experiments would include assessing whether FAM30A depletion completely abrogates disease onset over an extended period (i.e., survival analysis) and exploring the impact of FAM30A depletion on the engraftment of normal CD34+CD38- HSCs in the BM.

Another notable feature of MSI2 is its relevance in therapy-resistant AML stem cells (Spinler et al., 2020). It has been demonstrated that suppressing or inhibiting MSI2 significantly improves sensitivity to Daunorubicin (Han et al., 2015). Consequently, this study also reveals that depletion of FAM30A significantly elevates apoptotic levels and enhances sensitivity to chemotherapeutic agents such as Ara-Cytarabine and Daunorubicin (Figure 21). Importantly, these effects were reversed upon restoration of

the MSI2-interacting region of FAM30A, even in two FAM30A^{low} AML cell lines with different mutational background (Figure 24). These results confirmed previous predictive reports (F. Wang et al., 2018; Y. H. Wang et al., 2021) that correlated FAM30A expression with worse patients initial response to chemotherapy. Moreover, FAM30A expression levels were correlated with upregulation of ATP-dependent drug transporters important for chemoresistance in leukaemia (Marzac et al., 2011). Two additional lncRNAs have been found to be involved in increasing MSI2 expression and stabilizing important oncogenic mRNAs, leading to increased chemoresistance and stemness in gastric (Zhu et al., 2022) and liver cancer cells (T. Fang et al., 2017), mirroring the findings presented in this study.

Mechanistically, the central regulator MSI2 has been demonstrated to efficiently control the stem cell-specific translatome in AML cells operating in essential oncogenic signaling pathways (Duggimpudi et al., 2018; Ito et al., 2010; Minuesa et al., 2019; Vu, Prieto, et al., 2017). Based on the hypothesis that FAM30A repeats affect MSI2-driven translation of its target RNAs, an increased association of MSI2 with polysomal fractions should be observed in this setting. Although no increase in MSI2 protein levels in the polysomal fractions was observed upon overexpression of FAM30A repeats (data not shown), strong downstream effects at the protein level on certain MSI2 targets were noted. Even though MSI2 has been demonstrated to influence a broad mRNA interactome in AML LSCs, translational variations are evident in merely a subset of these mRNAs (Karmakar et al., 2022; Nguyen et al., 2020). This aligns with the herein presented results, as MSI2 was experimentally validated to bind to reported targets such as BCL2, TSPAN3, and TGFBR1 mRNAs (Figure 34B), yet no significant differences were observed at the protein level. This study also emphasizes the importance of FAM30A and its repeats, illustrating a consistent positive correlation with MSI2 activation at the molecular level. This correlation was demostrated based on the observed effects that align with the reported dual role of MSI2 as both translational activator (e.g. SMAD3) (Park et al., 2014, 2015) or repressor (e.g. NUMB) (Ito et al., 2010) to direct mRNA targets crucial for leukaemia progression (Figure 32B and Figure 34B). The alterations of MSI2-targets at the protein level were confirmed using an inhibitor of translation elongation while mRNA levels remained unaffected (Figure 32C). Interestingly, both of these pathways - Numb/Notch1 and TGFβ/SMAD3- have been suggested to be MSI2-dependent and LSC-specific as well as representing potential therapeutic targets in AML (Minuesa et al., 2019; Nguyen et al., 2020; Nishimoto & Okano, 2010).

The variability in MSI2-driven translation of its mRNA targets can be explained within the context of binding motifs, distinct MSI2 isoforms (M. Li et al., 2020; Wuebben et al.,

2012), phosphorylation status (MacNicol et al., 2017) or the recruitment of additional cofactors, offering further insights. Regarding the latter one, a study that investigated the MSI2-interactome identified the protein SYNCRIP as an indirect MSI2 interactor and demonstrated that both co-regulated the myeloid leukaemia stem-cell program. The evidence indicates that MSI2 and SYNCRIP share a set of stem cell-specific RNA targets critical for leukaemia progression without affecting normal haematopoietic development (Vu, Prieto, et al., 2017). Notably, mass spectrometry following pulldown revealed SYNCRIP as one of the highly enriched proteins bound to FAM30A repeats (log_2Fold-Change: 4.52; P = 4.22E-06). Should this hypothesis be substantiated, it could imply that FAM30A serves as a scaffold facilitating the function of both MSI2 and SYNCRIP, working synergistically to regulate the translation of stem cell-specific targets in AML. Further experiments in AML cells overexpressing FAM30A repeats in combination with MSI2 depletion/overexpression could globally identify candidate genes regulated by FAM30A/MSI2, such as measurement of newly synthesized proteins based on AHA incorporation. This approach would also confirm that the alteration of MSI2 targets at the protein level, due to overexpression of FAM30A repeats, is specifically due to changes in translational rates.

Altogether, the results strongly suggest that FAM30A repeats promote MSI2 function, indicated not only by increased MSI2 protein levels but aso enhancing MSI2-driven downstream effects. To the best of our knowledge, there have been no reports so far of a non-coding RNA associating to and directly influencing MSI2 protein function. Only DANCR, which is notably abundant in AML LSCs, and NEAT1 stands out as the sole lncRNAs known to regulate MSI2 expression in bladder and ovarian cancer, respectively (Zhan et al., 2018; C. Zhao, Zi, et al., 2021). Unfortunately, the mechanism by which the interaction of FAM30A repeats positively influences MSI2 function could not be mechanistically addressed, though some possible scenarios are proposed in this thesis. The most scenario is that FAM30A serves as a scaffold for the MSI2 protein to bind to its target mRNAs and exert its translational control, involving the formation of a ternary complex (FAM30A/MSI2/target mRNAs). This hypothesis could be validated using chemiluminescent Electrophoresis Mobility Shift Assays (EMSA) to assess MSI2-RNA complexes *in vitro*, as described by Köhn et al. 2010 and Minuesa et al. 2019. By adding increasing concentrations of recombinant MSI2-tagged protein incubated with *in vitro* transcribed biotinylated FAM30A repeats and/or validated MSI2-RNA target oligos labelled with different fluorescent markers, this approach could evaluate whether the aforementioned ternary complex is formed. An additional possibility is that FAM30A repeats simultaneously enhance MSI2 protein stability either by (1) recruiting proteins

that perform post-translational modifications on MSI2 (e.g. phosphorylation at the C-terminal region (MacNicol et al., 2017)), which could be assessed via mass spectrometry analyses in FAM30A repeat-overexpressing cells (OE, REC) compared to controls (CTRL, KD), or (2) shielding MSI2 from the protein degradation machinery. A different scenario is that FAM30A may act as a miRNA sponge. Similarly, the lncRNAs DANCR and NEAT1 have previously been proposed to sequester miRNAs that negatively regulate MSI2 levels. Some miRNAs, such as miR193a-5p, were predicted *in silico* to bind FAM30A (data not shown) so further investigations are required to validate miRNA binding to FAM30A.

4.4.- Influence of FAM30A on RUNX1 transcriptional activation: the role of the differential MSI2-driven post-transcriptional regulation of RUNX1 isoforms

It is noteworthy that FAM30A has recently been identified as a significant mediator gene influencing the impact of t(8;21)/RUNX1-RUNX1T1 fusions on survival in AML patients (Hornung et al., 2018). Patients with RUNX1 point mutations, typically associated with poor outcomes, show higher levels of FAM30A expression, whereas FAM30A expression is significantly lower in those with the t(8;21)/RUNX1-RUNX1T1 translocation (Y. H. Wang et al., 2021). Furthermore, FAM30A is the most significantly downregulated gene in AML cells lacking the t(8;21) translocation (Ptasinska et al., 2012). These findings suggest that FAM30A may be involved in or serves as surrogate of cellular processes that influence survival of AML patients with RUNX1 alterations. While there is no direct evidence of a connection between FAM30A and RUNX1, a study has suggested the possibility of RUNX1 directly regulating the transcription of the FAM30A gene as several RUNX1-binding TGTGG core sequences were identified within the upstream sequence of FAM30A (Y. H. Wang et al., 2021), indicating the potential presence of a self-regulatory loop involving FAM30A and RUNX1.

This potential association between FAM30A and RUNX1 was also clearly apparent in the present study, as changes in FAM30A expression not only influenced total RUNX1 protein levels in AML and CML cell lines (Figure 33A, B and 36B), but also positively correlated with RUNX1-transcriptional activation based on RNA-sequencing analysis (Figure 33C). Moreover, overexpression of FAM30A repeats in an MSI2 knockdown setting transfected with luciferase reporter constructs confirmed that RUNX1 activation was both FAM30A- and MSI2-dependent in a leukaemia cell line (Figure 36C). Even though there has not been any direct link between MSI2 and RUNX1 so far, analysis of

available CLIP-seq data in different leukaemia cell lines (Figure 34A) and further confirmation via immunoprecipitation studies (Figure 34B) demonstrated that MSI2 is more heavily associated with the RUNX1C isoform (3´UTR) and to a lesser extent with the RUNX1A (5´UTR) and RUNX1B. Following treatment with a translation inhibitor, we herein unveiled a direct positive regulation between FAM30A repeats, increased MSI2 protein and substantially elevated RUNX1C whereas FAM30A absence decreased MSI2 and favoured RUNX1B (Figure 35). These observations were reflected in the favoured activation of P1-promoter (RUNX1C-specific) upon increased FAM30A repeats/MSI2, explained by the reported transcriptional auto-regulation of P1-promoter (Ferrell et al., 2015; Martinez et al., 2016) (Figure 36). Nevertheless, MSI2 knockdown decreased RUNX1 activity that was only reflected on P2 promoter thus it cannot be ruled out that FAM30A could influence additional factors involved in the differential downstream activation of RUNX1 promoters. Further investigations into the hypothesized FAM30A/MSI2/RUNX1 axis could be conducted in RUNX1-mutated AML cell lines that require RUNX1 expression for survival (such as Kasumi-1, ME-1, and MOLM-13) (Ben-Ami et al., 2013; Wilkinson et al., 2013). These studies would also help to determine whether increased MSI2 function due to FAM30A overexpression and the associated phenotypes seen in this study are independent of RUNX1 mutational status.

Translation initiation is considered the rate-limiting step of translation and it is regulated by *cis*-acting elements, such as the length of the 3´UTR or secondary structures within the 5´UTR (Merrick, 1992; Szostak & Gebauer, 2013). A study demonstrated that alternative RUNX1 promoter usage coupled to translation control is essential for regulation of the temporally and cell type-specific RUNX1 isoform pool (Pozner et al., 2000). RUNX1B and RUNX1C isoforms share the same 3´UTR and they only differ on their N-terminal sequences which is indicative of alternative translation initiation (Goyama et al., 2019). The 5´UTR of RUNX1B is remarkably long and comprises several GC-rich islands forming highly stable stem-loop structures and other non-canonical initiation codons (Levanon et al., 2001) that makes the translation of this isoform highly inefficient (Kozak, 1989, 1991) whereas the shorter 5´UTR of RUNX1C directs its efficient translation. Even though MSI2 is known to preferentially bind to the 3´UTR of the target mRNA and repress its translation (Ito et al., 2010; Kawahara et al., 2008), MSI2 can also promote translation of its target through 5´UTR-dependent processes (Vu, Prieto, et al., 2017; Yeh et al., 2023). Thus, a possibility would be that high FAM30A repeats increases MSI2 binding to the 3´UTR of RUNX1C and forms a structural loop that recruits the translational machinery to initiate the efficient translation at the 5´UTR. Nevertheless, further investigations are needed to explore how potential structural

disparities within these untranslated regions may impact the differential binding of MSI2 to RUNX1C as opposed to RUNX1B. To investigate how MSI2 regulates different RUNX1 mRNA isoforms differently (i.e. mRNA stability or translation), the distinct 5´- and 3´-UTRs of these isoforms could be cloned into luciferase reporter constructs. These constructs could then be transfected into K562 cells to evaluate how stable high levels of FAM30A repeats and MSI2, in combination with MSI2 depletion, affect luciferase expression levels.

RUNX1 expression is significantly higher in primary LSC+ cells from AML patients (Ng et al., 2016) and recently RUNX1 has been shown to be critical for LSC maintenance (Wesely et al., 2020). Although RUNX1B and RUNX1C share a high degree of structural similarity, they serve distinct and specific roles in adult haematopoiesis. RUNX1C has been demonstrated to possess a greater binding capacity to its target genes (Telfer & Rothenberg, 2001), and its expression correlates with proliferative CD34+ subpopulations (Challen & Goodell, 2010; Ferrell et al., 2015). Consequently, it has been suggested that RUNX1C may be the stem cell-specific isoform in AML, as overexpression of RUNX1C in AML blasts has been shown to confer a phenotype more akin to haematopoietic stem cells (HSCs) (Wesely et al., 2020). This study showed that RUNX1 knockdown decreased LSC frequency and colony-formation potential within the CD34+ fraction as well as it abrogated their ability to engraft. Moreover, RUNX1C function has also been recognized to be relevant within the monocyte/macrophage compartment as it increases CFU-GM and CFU-M types (Navarro-Montero et al., 2017), a phenomenon also observed upon overexpression of FAM30A repeats (Figure 22). These effects may be achieved by regulating the expression of crucial haematopoietic transcription factors like GATA1, PU.1, as well as other cytokines and receptors essential for myeloid-committed progenitors, such as CSF1 (M-CSF) and CSF2 (GM-CSF) and their receptors (all showing significance with $\log_2FC > 1$ in the OE condition). Therefore, the increase in RUNX1 transcriptional activation resulting from the overexpression of the MSI2-interacting region of FAM30A is likely attributed to the heightened levels and activity of RUNX1C, as this isoform holds relevance in the AML LSC context, unlike RUNX1B. However, whether RUNX1C overexpression mirrors the phenotypes upon overexpression FAM30A repeats requires further validation. Additionally, AML patients with FLT3-ITD mutations exhibit elevated RUNX1 levels, and studies have shown that increased RUNX1 expression collaborates with FLT3-ITD to drive leukaemia progression (Behrens et al., 2017; Salarpour et al., 2017). Moreover, RUNX1 is a cofactor of STAT5 in AML (Fan et al., 2023) and together with FOXP3, they drive the transcription of the CD25 gene (Rudra et al., 2009). Altogether, this implies a potential link between

FAM30A, MSI2 and increased RUNX1 activation that could work in conjunction with FLT3-ITD and IL2/STAT5 pathway to drive AML progression.

4.5.- Targeting the FAM30A-MSI2 interaction: a promising therapeutic avenue to eradicate AML LSCs

Given the central role of MSI2 in regulating key biological processes that are critical for cancer initiation, progression, and drug resistance, targeting MSI2 RNA-binding activity has emerged as a promising therapeutic strategy in several types of solid tumors (Lan et al., 2018; Yeh et al., 2023; X. Zhang et al., 2022) and notably in hematological malignancies (Minuesa et al., 2019; Sureda-Gómez et al., 2022). Noteworthy, a large chemical screen identified the druggability of MSI2 by developing a small molecule called Ro 08–2750 (Ro) that directly interacts with MSI2-RRM1 and competes for its RNA-binding capacity and displaces MSI2-target mRNAs in biochemical and cellular assays in leukaemia cell lines (Minuesa et al., 2014). Treatment with Ro has been shown to specifically alter MSI2-specific oncogenic targets without impacting global translation. As a result, there is an increase in differentiation, apoptosis, and a decrease in disease burden in a aggressive murine AML model (Minuesa et al., 2019). Despite the potential of these MSI2 antagonists, their specificity and selectivity have not been fully characterized. Furthermore, the extent of off-target effects, particularly through interactions with other RNA-binding proteins (Walters et al., 2023), remains to be elucidated.

The relevance of targeting *in vivo* certain stem cell-specific lncRNAs in AML has already been shown to have an impact on LSC features (Bill et al., 2019; H. Luo et al., 2019; X. Y. Ma et al., 2015). Due to the potential clinical relevance of FAM30A repeats in promoting LSC activity (i.e. colony-formation potential, chemoresistance, engraftment), a promising possibility would be designing Locked Nucleic Acid (LNA) Gapmers to silence endogenous FAM30A or to design specific small molecules that bind to FAM30A repeats and disrupt FAM30A-MSI2 interaction. While further research is needed to explore the biological significance of FAM30A in CD34+ HSPCs, as well as to confirm that FAM30A-induced effects are solely driven by enhanced MSI2/RUNX1 activity, the targeting of the FAM30A/MSI2 axis holds considerable promise as a therapeutic approach to target AML LSCs particularly as combination regimens (i.e. with chemotherapeutic agents). Although requiring further validation, the probable occurrence of FAM30A and MSI2 binding occurs via the RRM1 domain suggests the potential for Ro

or Gapmers to disrupt this interaction and subsequent downstream oncogenic activation (Numb/Notch1, TGFβ/SMAD3). Another possibility is that MSI2 is tightly bound to FAM30A repeats under steady-state conditions, shielding MSI2-RRM1 from being targeted by Ro. Future experiments could involve treating FAM30A-depleted KG-1a cells (KD) with Ro, alongside overexpression of FAM30A repeats in the same setting (REC). This approach would help determine whether targeting MSI2 RNA-binding capacity in the absence or presence of FAM30A repeats affects the phenotypic outcomes described in this thesis, as well as MSI2-targets at the protein level. Additionally, by comparing these results with stable MSI2 depletion, we can assess whether the observed effects are MSI2-dependent. The proposed differential necessity of both MSI2 (Nguyen et al., 2020) and potentially FAM30A in AML LSCs presents a pathway for targeted therapy for AML patients. However, further studies are required to confirm whether targeting this oncogenic regulatory axis also safeguards normal haematopoiesis.

In summary, this study uncovers an important role for the stem-cell specific FAM30A in in defining characteristic features of AML LSCs (stemness, chemoresistance, engraftment) that is mechanistically driven by the interaction of the conserved FAM30A repeat region with MSI2 and its downstream effects on oncogenic factors critical for leukaemia progression. Moreover, this interaction is influential not only on canonical AML LSC signatures but also impacts the transcriptional activation of RUNX1, highlighting its multifaceted role in LSC maintenance (Figure 40, next page).

This underscores the need for further exploration into the regulatory mechanisms governed by FAM30A, not only within the AML context but also in normal haematopoiesis.

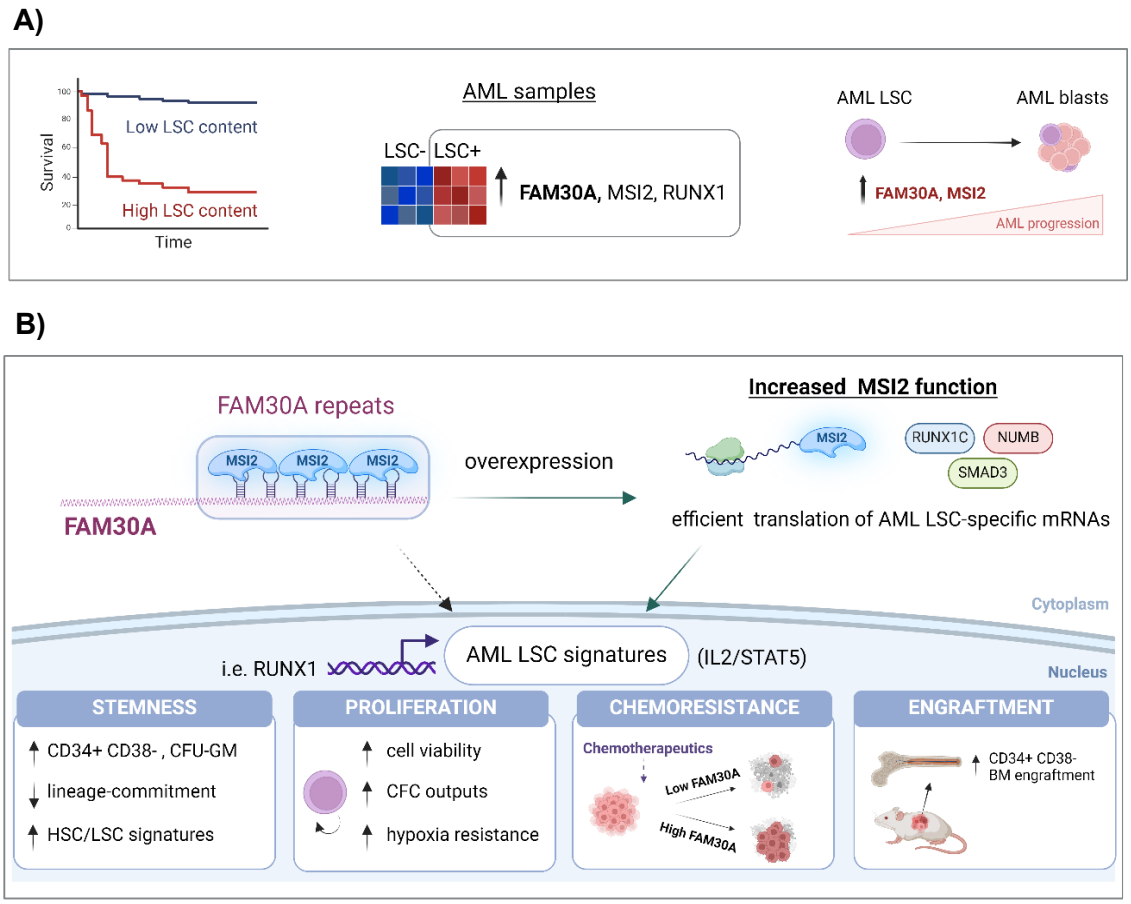


Figure 40. The relevant role of FAM30A and its partner MSI2 in AML leukaemia stem cells. **A)** Left panel: representative Kaplan-Meier showing how high LSC content negatively impacts survival of AML patients (Ng et al., 2016). Middle panel: microarray analysis of functionally sorted LSC fractions where there is high expression levels of FAM30A, MSI2 and in a lesser extent RUNX1, when compared to non-LSC fractions (Eppert et al., 2011). Right panel: hypothesis of the differential requirement of FAM30A and MSI2 (Kharas et al., 2010; Nguyen et al., 2020; Park et al., 2015) in AML LSCs necessary for disease onset. **B)** The data presented in this thesis supports an important role of how FAM30A influences the hallmarks of AML LSC activity (stemness, sustained proliferation, chemoresistance and increased engraftment) that is mechanistically driven by the novel direct binding of the pro-LSC regulator MSI2 to the highly conserved region of FAM30A. Consequently, there is an enhanced MSI2-driven translation of stem cell-specific mRNAs (i.e. RUNX1C), thereby inducing upregulation of AML LSC signatures (i.e. RUNX1, IL2/STAT5) essential for leukaemia progression.

5. SUMMARY

Acute myeloid leukaemia (AML) is the most common leukaemia in adults and it is characterized by the uncontrolled proliferation of malignant myeloid precursors in the bone marrow (Döhner et al., 2015). High relapse and refractoriness rates after induction chemotherapy still contribute to the unfavourable prognosis of AML patients (Ferrara & Schiffer, 2013). This recurrence is thought to be sustained by a heterogenous subset of chemotherapy-resistant leukaemia stem cells (LSCs), serving as the disease reservoir (Misaghian et al., 2009). Understanding the molecular function of genes specifically deregulated in LSCs represents a significant breakthrough in AML research, offering a better understanding of LSC biology necessary to develop novel therapies to eradicate the LSCs. Long non-coding RNAs (lncRNAs) are increasingly emerging as pivotal prognostic markers and potential targets for cancer therapy in AML (Bill et al., 2019). For many years, high FAM30A expression has consistently been associated with specific gene signatures linked to AML LSCs activity and poor prognosis in both adult (Eppert et al., 2011; S. W. K. Ng et al., 2016) and pediatric AML (Elsayed et al., 2020; B. J. Huang et al., 2022). However, the molecular role of FAM30A in this context remains unknown.

For the fist time, this study provides curcial insights into the role of FAM30A and its highly conserved repeats region in regulating LSC activity in the AML context. FAM30A depletion correlated with downregulation of canonical AML LSC signatures, decreased cell viability, increased apoptosis and sensitivity chemotherapeutics compared to control. This also led to decreased colony-formation potential, increased granulocytic differentiation and, importantly, abrogated leukaemic engraftment in murine bone marrow (BM) by 6 weeks post-transplant. Conversely, overexpression of FAM30A repeats in this setting resulted in increased stemness, proliferation, chemoresistance and murine BM engraftment compared to control. These findings could be explained mechanistically by association of the pro-LSC regulator MSI2 to FAM30A repeats. Additionally, changes in the expression of FAM30A and its repeats within stable KG-1a cells were found to influence MSI2-mediated post-transcriptional regulation of its stem cell-specific mRNAs which are critical for leukaemia progression. These investigations further unveiled a correlation between differential MSI2-binding to RUNX1 isoforms and variations in RUNX1 activity, a transcription factor commonly mutated in AML. Notably, we also confirm that RUNX1 transcriptional activation is both FAM30A- and MSI2-dependent thus implying a novel positive feedback loop involving FAM30A, MSI2 and RUNX1 relevant for AML LSCs maintenance. We propose that a further exploration of this mechanism is warranted as a means of defining a therapeutically tractable pathway to target AML LSCs.

6. REFERENCES

Abaza, H. M. H., Youssef, S. R., Saad, A. A., Kamal, G. M., Hegazy, M. G. A., Ibrahim, R. I., & Annaka, L. M. H. (2015). Detection of 14q32 rearrangements in multiple myeloma, using simultaneous FISH analysis combined with immunofluorescence. *Hematology/Oncology and Stem Cell Therapy*, *8*(2), 56–63. https://doi.org/10.1016/J.HEMONC.2015.04.002

Abdel-Wahab, O., & Levine, R. L. (2013). Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood*, *121*(18), 3563–3572. https://doi.org/10.1182/BLOOD-2013-01-451781

Abeles, R. D., McPhail, M. J., Sowter, D., Antoniades, C. G., Vergis, N., Vijay, G. K. M., Xystrakis, E., Khamri, W., Shawcross, D. L., Ma, Y., Wendon, J. A., & Vergani, D. (2012). CD14, CD16 and HLA-DR reliably identifies human monocytes and their subsets in the context of pathologically reduced HLA-DR expression by CD14(hi) /CD16(neg) monocytes: Expansion of CD14(hi) /CD16(pos) and contraction of CD14(lo) /CD16(pos) monocytes in acute liver failure. *Cytometry. Part A : The Journal of the International Society for Analytical Cytology*, *81*(10), 823–834. https://doi.org/10.1002/CYTO.A.22104

Adams, B. D., Parsons, C., Walker, L., Zhang, W. C., & Slack, F. J. (2017). Targeting noncoding RNAs in disease. *The Journal of Clinical Investigation*, *127*(3), 761–771. https://doi.org/10.1172/JCI84424

Aguila, H. L., Akashi, K., Domen, J., Gandy, K. L., Lagasse, E., Mebius, R. E., Morrison, S. J., Shizuru, J., Strober, S., Uchida, N., Wright, D. E., & Weissman, I. L. (1997). From stem cells to lymphocytes: biology and transplantation. *Immunological Reviews*, *157*, 13–40. https://doi.org/10.1111/J.1600-065X.1997.TB00971.X

Akashi, K., Traver, D., Miyamoto, T., & Weissman, I. L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, *404*(6774), 193–197. https://doi.org/10.1038/35004599

Akrami, S., Tahmasebi, A., Moghadam, A., Ramezani, A., & Niazi, A. (2023). Integration of mRNA and protein expression data for the identification of potential biomarkers associated with pancreatic ductal adenocarcinoma. *Computers in Biology and Medicine*, *157*, 106529. https://doi.org/10.1016/J.COMPBIOMED.2022.106529

Alaggio, R., Amador, C., Anagnostopoulos, I., Attygalle, A. D., Araujo, I. B. de O., Berti, E., Bhagat, G., Borges, A. M., Boyer, D., Calaminici, M., Chadburn, A., Chan, J. K. C., Cheuk, W., Chng, W. J., Choi, J. K., Chuang, S. S., Coupland, S. E., Czader, M., Dave, S. S., … Xiao, W. (2022). The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*, *36*(7), 1720–1748. https://doi.org/10.1038/S41375-022-01620-2

Al-Kershi, S., Bhayadia, R., Ng, M., Verboon, L., Emmrich, S., Gack, L., Schwarzer, A., Strowig, T., Heckl, D., & Klusmann, J. H. (2019). The stem cell–specific long noncoding RNA HOXA10-AS in the pathogenesis of KMT2A-rearranged leukemia. *Blood Advances*, *3*(24), 4252. https://doi.org/10.1182/BLOODADVANCES.2019032029

Allieri, M., Douay, L., Deloux, J., Smadja, N., Najman, A., & Gorin, N. (1990). The role of methylcellulose on colony growth of human myeloid leukemic progenitors (AML-CFU). *Experimental Hematology*.

Alsuwaidi, L., Hachim, M., & Senok, A. (2021). Novel Markers in Pediatric Acute Lymphoid Leukemia: The Role of ADAM6 in B Cell Leukemia. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/FCELL.2021.706129

Alvarez-Dominguez, J. R., Hu, W., Yuan, B., Shi, J., Park, S. S., Gromatzky, A. A., Van Oudenaarden, A., & Lodish, H. F. (2014). Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood*, *123*(4), 570–581. https://doi.org/10.1182/BLOOD-2013-10-530683

Andrews, R. G., Singer, J. W., & Bernstein, I. D. (1989). Precursors of colony-forming cells in humans can be distinguished from colony-forming cells by expression of the CD33 and CD34 antigens and light scatter properties. *The Journal of Experimental Medicine*, *169*(5), 1721. https://doi.org/10.1084/JEM.169.5.1721

Aparicio-Prat, E., Arnan, C., Sala, I., Bosch, N., Guigó, R., & Johnson, R. (2015). DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. *BMC Genomics*, *16*(1), 1–15. https://doi.org/10.1186/S12864-015-2086-Z/FIGURES/6

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., Bloomfield, C. D., Cazzola, M., & Vardiman, J. W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, *127*(20), 2391–2405. https://doi.org/10.1182/BLOOD-2016-03-643544

Aref, S., Azmy, E., Ibrahim, L., Sabry, M., & Agdar, M. El. (2020). Prognostic value of CD25/CD123 pattern of expression in acute myeloid leukemia patients with normal cytogenetic. *Leukemia Research Reports*, *13*, 100203. https://doi.org/10.1016/J.LRR.2020.100203

Barmpouti, A., Höglund, M., Johansson, B., Lassen, C., Nilsson, P.-G., Hagemeijer, A., Mitelman, F., & Fioretos, T. (2003). A Novel Gene, MSI2, Encoding a Putative RNA-binding Protein Is Recurrently Rearranged at Disease Progression of Chronic Myeloid Leukemia and Forms a Fusion Gene with HOXA9 as a Result of the Cryptic t(7;17)(p15;q23). *Cancer Research*, *63*(6), 1202–1206. https://portal.research.lu.se/en/publications/a-novel-gene-msi2-encoding-a-putative-rna-binding-protein-is-recu

Barrera, A., Ramos, H., Vera-Otarola, J., Fernández-Garciá, L., Angulo, J., Olguín, V., Pino, K., Mouland, A. J., & López-Lastra, M. (2020). Post-translational modifications of hnRNP A1 differentially modulate retroviral IRES-mediated translation initiation. *Nucleic Acids Research*, *48*(18), 10479. https://doi.org/10.1093/NAR/GKAA765

Bassett, A. R., Akhtar, A., Barlow, D. P., Bird, A. P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A. C., Gingeras, T. R., Haerty, W., Higgs, D. R., Miska, E. A., & Ponting, C. P. (2014). Considerations when investigating lncRNA function in vivo. *ELife*, *3*(August2014), 1–14. https://doi.org/10.7554/ELIFE.03058

Batista-Gomes, J. A., Mello, F. A. R., De Oliveira, E. H. C., De Souza, M. P. C., Wanderley, A. V., Da Costa Pantoja, L., Dos Santos, N. P. C., Khayat, B. C. M., & Khayat, A. S. (2020). Identifying novel genetic alterations in pediatric acute lymphoblastic leukemia based on copy number analysis. *Molecular Cytogenetics*, *13*(1), 1–8. https://doi.org/10.1186/S13039-020-00491-5/TABLES/5

Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A. M., & Peault, B. (1992). Isolation of a candidate human hematopoietic stem-cell population. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(7), 2804. https://doi.org/10.1073/PNAS.89.7.2804

Beermann, J., Kirste, D., Iwanov, K., Lu, D., Kleemi, F., Kumarswamy, R., Schimmel, K., Bär, C., & Thum, T. (2018). A large shRNA library approach identifies lncRNA Ntep as an essential regulator of cell proliferation. *Cell Death and Differentiation*, *25*(2), 307. https://doi.org/10.1038/CDD.2017.158

Bellissimo, D. C., & Speck, N. A. (2017). RUNX1 Mutations in Inherited and Sporadic Leukemia. *Frontiers in Cell and Developmental Biology*, *5*(DEC). https://doi.org/10.3389/FCELL.2017.00111

Ben-Ami, O., Friedman, D., Leshkowitz, D., Goldenberg, D., Orlovsky, K., Pencovich, N., Lotem, J., Tanay, A., & Groner, Y. (2013). Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Reports*, *4*(6), 1131–1143. https://doi.org/10.1016/J.CELREP.2013.08.020

Ben-Bassat, H., Korkesh, A., Voss, R., Leizerowitz, R., & Polliack, A. (1982). Establishment and characterization of a new permanent cell line (GDM-1) from a patient with myelomonoblastic leukemia. *Leukemia Research*, *6*(6). https://doi.org/10.1016/0145-2126(82)90055-8

Bencomo-Alvarez, A. E., Rubio, A. J., Gonzalez, M. A., & Eiring, A. M. (2019). Energy Metabolism and Drug Response in Myeloid Leukaemic Stem Cells. *British Journal of Haematology*, *186*(4), 524. https://doi.org/10.1111/BJH.16074

Bennett, C. G., Riemondy, K., Chapnick, D. A., Bunker, E., Liu, X., Kuersten, S., & Yi, R. (2016a). Genome-wide analysis of Musashi-2 targets reveals novel functions in governing epithelial cell migration. *Nucleic Acids Research*, *44*(8), 3788–3800. https://doi.org/10.1093/NAR/GKW207

Bennett, C. G., Riemondy, K., Chapnick, D. A., Bunker, E., Liu, X., Kuersten, S., & Yi, R. (2016b). Genome-wide analysis of Musashi-2 targets reveals novel functions in governing epithelial cell migration. *Nucleic Acids Research*, *44*(8), 3788. https://doi.org/10.1093/NAR/GKW207

Bennett, J. M., Catovsky, D., Daniel, M. -T, Flandrin, G., Galton, D. A. G., Gralnick, H. R., & Sultan, C. (1976). Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *British Journal of Haematology*, *33*(4), 451–458. https://doi.org/10.1111/J.1365-2141.1976.TB03563.X

Bennett, J. M., Catovsky, D., Daniel, M. -T, Flandrin, G., Galton, D. A. G., Gralnick, H. R., & Sultan, C. (1991). Proposal for the recognition of minimally differentiated acute myeloid leukaemia (AML-MO). *British Journal of Haematology*, *78*(3), 325–329. https://doi.org/10.1111/J.1365-2141.1991.TB04444.X

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., & Sultan, C. (1985). Criteria for the diagnosis of acute leukemia of megakaryocyte lineage (M7). A report of the French-American-British Cooperative Group. *Annals of Internal Medicine*, *103*(3), 460–462. https://doi.org/10.7326/0003-4819-103-3-460

Bester, A. C., Lee, J. D., Chavez, A., Lee, Y. R., Nachmani, D., Vora, S., Victor, J., Sauvageau, M., Monteleone, E., Rinn, J. L., Provero, P., Church, G. M., Clohessy, J. G., & Pandolfi, P. P. (2018). An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance. *Cell*, *173*(3), 649-664.e20. https://doi.org/10.1016/J.CELL.2018.03.052

Bick, A. G., Weinstock, J. S., Nandakumar, S. K., Fulco, C. P., Bao, E. L., Zekavat, S. M., Szeto, M. D., Liao, X., Leventhal, M. J., Nasser, J., Chang, K., Laurie, C., Burugula, B. B., Gibson, C. J., Lin, A. E., Taub, M. A., Aguet, F., Ardlie, K., Mitchell, B. D., … Natarajan, P. (2020). Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature 2020 586:7831*, *586*(7831), 763–768. https://doi.org/10.1038/s41586-020-2819-2

Bill, M., Nicolet, D., Kohlschmidt, J., Walker, C. J., Mrózek, K., Eisfeld, A. K., Papaioannou, D., Rong-Mullins, X., Brannan, Z., Kolitz, J. E., Powell, B. L., Archer, K. J., Dorrance, A. M., Carroll, A. J., Stone, R. M., Byrd, J. C., Garzon, R., & Bloomfield, C. D. (2020). Mutations associated with a 17-gene leukemia stem cell score and the score's prognostic relevance in the context of the European LeukemiaNet classification of acute myeloid leukemia. *Haematologica*, *105*(3), 721–729. https://doi.org/10.3324/HAEMATOL.2019.225003

Bill, M., Papaioannou, D., Karunasiri, M., Kohlschmidt, J., Pepe, F., Walker, C. J., Walker, A. E., Brannan, Z., Pathmanathan, A., Zhang, X., Mrózek, K., LaRocco, A., Volinia, S., Bloomfield, C. D., Garzon, R., & Dorrance, A. M. (2019a). Expression and functional relevance of long non-coding RNAs in acute myeloid leukemia stem cells. *Leukemia*, *33*(9), 2169–2182. https://doi.org/10.1038/S41375-019-0429-5

Bill, M., Papaioannou, D., Karunasiri, M., Kohlschmidt, J., Pepe, F., Walker, C. J., Walker, A. E., Brannan, Z., Pathmanathan, A., Zhang, X., Mrózek, K., LaRocco, A., Volinia, S., Bloomfield, C. D., Garzon, R., & Dorrance, A. M. (2019b). Expression and functional relevance of long non-coding RNAs in acute myeloid leukemia stem cells. *Leukemia*, *33*(9), 2169–2182. https://doi.org/10.1038/S41375-019-0429-5

Birdwell, C., Fiskus, W., Kadia, T. M., DiNardo, C. D., Mill, C. P., & Bhalla, K. N. (2021). EVI1 dysregulation: impact on biology and therapy of myeloid malignancies. *Blood Cancer Journal 2021 11:3*, *11*(3), 1–14. https://doi.org/10.1038/s41408-021-00457-9

Blair, A., & Sutherland, H. J. (2000). Primitive acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo lack surface expression of c-kit (CD117). *Experimental Hematology*, *28*(6), 660–671. https://doi.org/10.1016/S0301-472X(00)00155-7

Boisset, J. C., Van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., & Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, *464*(7285), 116–120. https://doi.org/10.1038/NATURE08764

Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, *3*(7), 730–737. https://doi.org/10.1038/NM0797-730

Bosman, M. C. J., Schuringa, J. J., & Vellenga, E. (2016). Constitutive NF-κB activation in AML: Causes and treatment strategies. *Critical Reviews in Oncology/Hematology*, *98*, 35–44. https://doi.org/10.1016/J.CRITREVONC.2015.10.001

Bowman, R. L., Busque, L., & Levine, R. L. (2018). Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies. *Cell Stem Cell*, *22*(2), 157–170. https://doi.org/10.1016/J.STEM.2018.01.011

Boyd, A. L., Aslostovar, L., Reid, J., Ye, W., Tanasijevic, B., Porras, D. P., Shapovalova, Z., Almakadi, M., Foley, R., Leber, B., Xenocostas, A., & Bhatia, M. (2018). Identification of Chemotherapy-Induced Leukemic-Regenerating Cells Reveals a Transient Vulnerability of Human AML Recurrence. *Cancer Cell*, *34*(3), 483-498.e5. https://doi.org/10.1016/J.CCELL.2018.08.007

Breems, D. A., Van Putten, W. L. J., De Greef, G. E., Van Zelderen-Bhola, S. L., Gerssen-Schoorl, K. B. J., Mellink, C. H. M., Nieuwint, A., Jotterand, M., Hagemeijer, A., Beverloo, H. B., & Löwenberg, B. (2008). Monosomal karyotype in acute myeloid leukemia: a better indicator of poor prognosis than a complex karyotype. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *26*(29), 4791–4797. https://doi.org/10.1200/JCO.2008.16.0259

Bruno, S., Mancini, M., De Santis, S., Monaldi, C., Cavo, M., & Soverini, S. (2021). The Role of Hypoxic Bone Marrow Microenvironment in Acute Myeloid Leukemia and Future Therapeutic Opportunities. *International Journal of Molecular Sciences*, *22*(13). https://doi.org/10.3390/IJMS22136857

Bruun, G. H., Doktor, T. K., Borch-Jensen, J., Masuda, A., Krainer, A. R., Ohno, K., & Andresen, B. S. (2016). Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biology*, *14*(1), 1–19. https://doi.org/10.1186/S12915-016-0279-9/FIGURES/5

Bryant, D., Smith, L., Rogers-Broadway, K. R., Karydis, L., Woo, J., Blunt, M. D., Forconi, F., Stevenson, F. K., Goodnow, C., Russell, A., Humburg, P., Packham, G., Steele, A. J., & Strefford, J. C. (2023). Network analysis reveals a major role for 14q32 cluster miRNAs in determining transcriptional differences between IGHV-mutated and unmutated CLL. *Leukemia 2023 37:7*, *37*(7), 1454–1463. https://doi.org/10.1038/s41375-023-01918-9

Bullinger, L., Döhner, K., & Dohner, H. (2017). Genomics of Acute Myeloid Leukemia Diagnosis and Pathways. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *35*(9), 934–946. https://doi.org/10.1200/JCO.2016.71.2208

Busch, B., Bley, N., Müller, S., Glaß, M., Misiak, D., Lederer, M., Vetter, M., Strauß, H. G., Thomssen, C., & Hüttelmaier, S. (2016). The oncogenic triangle of HMGA2, LIN28B and IGF2BP1 antagonizes tumor-suppressive actions of the let-7 family. *Nucleic Acids Research*, *44*(8), 3845–3864. https://doi.org/10.1093/NAR/GKW099

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T., & Rodewald, H. R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, *518*(7540), 542–546. https://doi.org/10.1038/NATURE14242

Byers, R. J., Currie, T., Tholouli, E., Rodig, S. J., & Kutok, J. L. (2011). MSI2 protein expression predicts unfavorable outcome in acute myeloid leukemia. *Blood*, *118*(10), 2857–2867. https://doi.org/10.1182/BLOOD-2011-04-346767

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D. B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., Von Paleske, L., Renders, S., Wünsche, P., Zeisberger, P., Brocks, D., Gu, L., Herrmann, C., Haas, S., Essers, M. A. G., Brors, B., Eils, R., … Trumpp, A. (2014). Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell*, *15*(4), 507–522. https://doi.org/10.1016/J.STEM.2014.07.005

Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927. https://doi.org/10.1101/GAD.17446611

Calvo Sánchez, J., & Köhn, M. (2021). Small but Mighty-The Emerging Role of snoRNAs in Hematological Malignancies. *Non-Coding RNA*, *7*(4). https://doi.org/10.3390/NCRNA7040068

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A., & Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA (New York, N.Y.)*, *22*(6), 867–882. https://doi.org/10.1261/RNA.053561.115

Cerny, J., Yu, H., Ramanathan, M., Raffel, G. D., Walsh, W. V., Fortier, N., Shanahan, L., O'Rourke, E., Bednarik, J., Barton, B., Kroll-Desrosiers, A., Hao, S., Woda, B., Hutchinson, L., M. Evens, A., Rosmarin, A. G., & Nath, R. (2013). Expression of

CD25 independently predicts early treatment failure of acute myeloid leukaemia (AML). *British Journal of Haematology*, *160*(2), 262–266. https://doi.org/10.1111/BJH.12109

Challen, G. A., & Goodell, M. A. (2010). Runx1 isoforms show differential expression patterns during hematopoietic development but have similar functional effects in adult hematopoietic stem cells. *Experimental Hematology*, *38*(5), 403–416. https://doi.org/10.1016/J.EXPHEM.2010.02.011

Challen, G. A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J. S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., Liang, S., Lu, Y., Darlington, G. J., Meissner, A., Issa, J. P. J., Godley, L. A., Li, W., & Goodell, M. A. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genetics*, *44*(1), 23–31. https://doi.org/10.1038/NG.1009

Chandel, N. S., Jasper, H., Ho, T. T., & Passegué, E. (2016). Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. *Nature Cell Biology*, *18*(8), 823–832. https://doi.org/10.1038/NCB3385

Chang, K., Marran, K., Valentine, A., & Hannon, G. J. (2012). RNAi in Cultured Mammalian Cells Using Synthetic siRNAs. *Cold Spring Harbor Protocols*, *2012*(9), 957. https://doi.org/10.1101/PDB.PROT071076

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J., & Church, G. M. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nature Methods 2015 12:4*, *12*(4), 326–328. https://doi.org/10.1038/nmeth.3312

Chen, B., Lee, J. B., Kang, H., Minden, M. D., & Zhang, L. (2018). Targeting chemotherapy-resistant leukemia by combining DNT cellular therapy with conventional chemotherapy. *Journal of Experimental & Clinical Cancer Research : CR*, *37*(1). https://doi.org/10.1186/S13046-018-0756-9

Chen, J., Li, C., Zhu, Y., Sun, L., Sun, H., Liu, Y., Zhang, Z., & Wang, C. (2015). Integrating GO and KEGG terms to characterize and predict acute myeloid leukemia-related genes. *Hematology (Amsterdam, Netherlands)*, *20*(6), 336–342. https://doi.org/10.1179/1607845414Y.0000000209

Chen, X., Clarence Yan, C., Luo, C., Ji, W., Zhang, Y., & Dai, Q. (2015). Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. *Scientific Reports*, *5*. https://doi.org/10.1038/SREP11338

Chen, Z. H., Wang, W. T., Huang, W., Fang, K., Sun, Y. M., Liu, S. R., Luo, X. Q., & Chen, Y. Q. (2017). The lncRNA HOTAIRM1 regulates the degradation of PML-RARA oncoprotein and myeloid cell differentiation by enhancing the autophagy pathway. *Cell Death and Differentiation*, *24*(2), 212. https://doi.org/10.1038/CDD.2016.111

Cheng, H., Zheng, Z., & Cheng, T. (2020). New paradigms on hematopoietic stem cell differentiation. *Protein & Cell*, *11*(1), 34–44. https://doi.org/10.1007/S13238-019-0633-0

Cheshier, S. H., Morrison, S. J., Liao, X., & Weissman, I. L. (1999). In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 3120–3125. https://doi.org/10.1073/PNAS.96.6.3120

Cheson, B. D., Cassileth, P. A., Head, D. R., Schiffer, C. A., Bennett, J. M., Bloomfield, C. D., Brunning, R., Gale, R. P., Grever, M. R., Keating, M. J., Sawitsky, A., Stass, S., Weinstein, H., & Woods, W. G. (1990). Report of the National Cancer Institute-sponsored workshop on definitions of diagnosis and response in acute myeloid leukemia. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *8*(5), 813–819. https://doi.org/10.1200/JCO.1990.8.5.813

Chou, M. Y., Underwood, J. G., Nikolic, J., Luu, M. H., & Black, D. L. (2000). Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Molecular cell, 5(6), 949–957. https://doi.org/10.1016/s1097-2765(00)80260-9

Choudhary, C., Brandts, C., Schwable, J., Tickenbrock, L., Sargin, B., Ueker, A., Böhmer, F. D., Berdel, W. E., Müller-Tidow, C., & Serve, H. (2007). Activation mechanisms of STAT5 by oncogenic Flt3-ITD. *Blood*, *110*(1), 370–374. https://doi.org/10.1182/BLOOD-2006-05-024018

Chuang, M. K., Chiu, Y. C., Chou, W. C., Hou, H. A., Tseng, M. H., Kuo, Y. Y., Chen, Y., Chuang, E. Y., & Tien, H. F. (2015). An mRNA expression signature for prognostication in de novo acute myeloid leukemia patients with normal karyotype. *Oncotarget*, *6*(36), 39098–39110. https://doi.org/10.18632/ONCOTARGET.5390

Clarke, J. P., Thibault, P. A., Salapa, H. E., & Levin, M. C. (2021). A Comprehensive Analysis of the Role of hnRNP A1 Function and Dysfunction in the Pathogenesis of Neurodegenerative Disease. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/FMOLB.2021.659610

Clarkson, B., Ohkita, T., Ota, K., & Fried, J. (1967). Studies of cellular proliferation in human leukemia. I. Estimation of growth rates of leukemic and normal hematopoietic cells in two adults with acute leukemia given single injections of tritiated thymidine. *The Journal of Clinical Investigation*, *46*(4), 506–529. https://doi.org/10.1172/JCI105553

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., & Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular Cell*, *33*(6), 717–726. https://doi.org/10.1016/J.MOLCEL.2009.01.026

Collins, A., Littman, D. R., & Taniuchi, I. (2009). RUNX proteins in transcription factor networks that regulate T-cell lineage choice. *Nature Reviews. Immunology*, *9*(2), 106–115. https://doi.org/10.1038/NRI2489

Colmone, A., Amorim, M., Pontier, A. L., Wang, S., Jablonski, E., & Sipkins, D. A. (2008). Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science (New York, N.Y.)*, *322*(5909), 1861–1865. https://doi.org/10.1126/SCIENCE.1164390

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry*, 162(1), 156–159. https://doi.org/10.1006/abio.1987.9999

Cook, A. M., Li, L., Ho, Y., Lin, A., Li, L., Stein, A., Forman, S., Perrotti, D., Jove, R., & Bhatia, R. (2014). Role of altered growth factor receptor-mediated JAK2 signaling in growth and maintenance of human acute myeloid leukemia stem cells. *Blood*, *123*(18), 2826–2837. https://doi.org/10.1182/BLOOD-2013-05-505735

Coombs, C. C., Zehir, A., Devlin, S. M., Kishtagari, A., Syed, A., Jonsson, P., Hyman, D. M., Solit, D. B., Robson, M. E., Baselga, J., Arcila, M. E., Ladanyi, M., Tallman, M. S., Levine, R. L., & Berger, M. F. (2017). Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell*, *21*(3), 374-382.e4. https://doi.org/10.1016/J.STEM.2017.07.010

Copley, M. R., Babovic, S., Benz, C., Knapp, D. J. H. F., Beer, P. A., Kent, D. G., Wohrer, S., Treloar, D. Q., Day, C., Rowe, K., Mader, H., Kuchenbauer, F., Humphries, R. K., & Eaves, C. J. (2013). The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nature Cell Biology*, *15*(8), 916–925. https://doi.org/10.1038/NCB2783

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J., Majeti, R., & Chang, H. Y. (2016a). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics 2016 48:10*, *48*(10), 1193–1203. https://doi.org/10.1038/ng.3646

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J., Majeti, R., & Chang, H. Y. (2016b). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics 2016 48:10*, *48*(10), 1193–1203. https://doi.org/10.1038/ng.3646

Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C., & Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2548–2553. https://doi.org/10.1073/PNAS.1324297111/-/DCSUPPLEMENTAL/PNAS.201324297SI.PDF

Corley, M., Burns, M. C., & Yeo, G. W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell*, *78*(1), 9–29. https://doi.org/10.1016/J.MOLCEL.2020.03.011

Corral, J., Lavenir, I., Impey, H., Warren, A. J., Forster, A., Larson, T. A., Bell, S., McKenzie, A. N. J., King, G., & Rabbitts, T. H. (1996). An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell*, *85*(6), 853–861. https://doi.org/10.1016/S0092-8674(00)81269-6

Costa, F. F. (2005). Non-coding RNAs: new players in eukaryotic biology. *Gene*, *357*(2), 83–94. https://doi.org/10.1016/J.GENE.2005.06.019

Costales, M. G., Childs-Disney, J. L., Haniff, H. S., & Disney, M. D. (2020). How We Think about Targeting RNA with Small Molecules. *Journal of Medicinal Chemistry*, *63*(17), 8880. https://doi.org/10.1021/ACS.JMEDCHEM.9B01927

Crooke, S. T., Liang, X. H., Baker, B. F., & Crooke, R. M. (2021). Antisense technology: A review. *The Journal of Biological Chemistry*, *296*. https://doi.org/10.1016/J.JBC.2021.100416

Cui, C., Wang, Y., Gong, W., He, H., Zhang, H., Shi, W., & Wang, H. (2021). Long Non-Coding RNA LINC00152 Regulates Self-Renewal of Leukemia Stem Cells and Induces Chemo-Resistance in Acute Myeloid Leukemia. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/FONC.2021.694021/FULL

Curtis, J. E., Messner, H. A., Hasselback, R., Elhakim, T. M., & McCulloch, E. A. (1984). Contributions of host- and disease-related attributes to the outcome of patients with acute myelogenous leukemia. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *2*(4), 253–259. https://doi.org/10.1200/JCO.1984.2.4.253

Davis, A. G., Einstein, J. M., Zheng, D., Jayne, N. D., Fu, X. D., Tian, B., Yeo, G. W., & Zhang, D. E. (2021). A CRISPR RNA-binding protein screen reveals regulators of RUNX1 isoform generation. *Blood Advances*, *5*(5), 1310–1323. https://doi.org/10.1182/BLOODADVANCES.2020002090

De Andrés-Aguayo, L., Varas, F., Kallin, E. M., Infante, J. F., Wurst, W., Floss, T., & Graf, T. (2011). Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice. *Blood*, *118*(3), 554–564. https://doi.org/10.1182/BLOOD-2010-12-322081

de Beauchamp, L., Himonas, E., & Helgason, G. V. (2022). Mitochondrial metabolism as a potential therapeutic target in myeloid leukaemia. *Leukemia*, *36*(1). https://doi.org/10.1038/S41375-021-01416-W

De Bruijn, M., & Dzierzak, E. (2017). Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood*, *129*(15), 2061–2069. https://doi.org/10.1182/BLOOD-2016-12-689109

De Clara, E., Gourvest, M., Ma, H., Vergez, F., Tosolini, M., Dejean, S., Demur, C., Delabesse, E., Recher, C., Touriol, C., Martelli, M. P., Falini, B., Brousset, P., & Bousquet, M. (2017). Long non-coding RNA expression profile in cytogenetically normal acute myeloid leukemia identifies a distinct signature and a new biomarker in NPM1-mutated patients. *Haematologica*, *102*(10), 1718–1726. https://doi.org/10.3324/HAEMATOL.2017.171645

de Grouw, E. P. L. M., Raaijmakers, M. H. G. P., Boezeman, J. B., van der Reijden, B. A., van de Locht, L. T. F., de Witte, T. J. M., Jansen, J. H., & Raymakers, R. A. P. (2006). Preferential expression of a high number of ATP binding cassette transporters in both normal and leukemic CD34+CD38- cells. *Leukemia*, *20*(4), 750–754. https://doi.org/10.1038/SJ.LEU.2404131

De Haan, G., & Lazare, S. S. (2018). Aging of hematopoietic stem cells. *Blood*, *131*(5), 479–487. https://doi.org/10.1182/BLOOD-2017-06-746412

de Lima, D. S., Cardozo, L. E., Maracaja-Coutinho, V., Suhrbier, A., Mane, K., Jeffries, D., Silveira, E. L. V., Amaral, P. P., Rappuoli, R., de Silva, T. I., & Nakaya, H. I. (2019). Long noncoding RNAs are involved in multiple immunological pathways in response to vaccination. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 17121–17126. https://doi.org/10.1073/PNAS.1822046116

de Lima, M., Roboz, G. J., Platzbecker, U., Craddock, C., & Ossenkoppele, G. (2021). AML and the art of remission maintenance. *Blood Reviews*, *49*. https://doi.org/10.1016/J.BLRE.2021.100829

Delás, M. J., Jackson, B. T., Kovacevic, T., Vangelisti, S., Munera Maravilla, E., Wild, S. A., Stork, E. M., Erard, N., Knott, S. R. V., & Hannon, G. J. (2019). lncRNA Spehd Regulates Hematopoietic Stem and Progenitor Cells and Is Required for Multilineage Differentiation. *Cell Reports*, *27*(3), 719. https://doi.org/10.1016/J.CELREP.2019.03.080

Delmer, A., Marie, J. P., Thevenin, D., Cadiou, M., Viguie, F., & Zittoun, R. (1989). Multivariate analysis of prognostic factors in acute myeloid leukemia: value of clonogenic leukemic cell properties. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *7*(6), 738–746. https://doi.org/10.1200/JCO.1989.7.6.738

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigó, R. (2012a). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. https://doi.org/10.1101/GR.132159.111

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigó, R. (2012b). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. https://doi.org/10.1101/GR.132159.111

Deschler, B., & Lübbert, M. (2006). Acute myeloid leukemia: epidemiology and etiology. *Cancer*, *107*(9), 2099–2107. https://doi.org/10.1002/CNCR.22233

Deveson, I. W., Hardwick, S. A., Mercer, T. R., & Mattick, J. S. (2017). The Dimensions, Dynamics, and Relevance of the Mammalian Noncoding Transcriptome. *Trends in Genetics : TIG*, *33*(7), 464–478. https://doi.org/10.1016/J.TIG.2017.04.004

DeWolf, S., & Tallman, M. S. (2020). How I treat relapsed or refractory AML. *Blood*, *136*(9), 1023–1032. https://doi.org/10.1182/BLOOD.2019001982

Di Giacomo, D., La Starza, R., Gorello, P., Pellanera, F., Kalender Atak, Z., De Keersmaecker, K., Pierini, V., Harrison, C. J., Arniani, S., Moretti, M., Testoni, N., De Santis, G., Roti, G., Matteucci, C., Bassan, R., Vandenberghe, P., Aerts, S., Cools, J., Bornhauser, B., … Mecucci, C. (2021). 14q32 rearrangements deregulating BCL11B mark a distinct subgroup of T-lymphoid and myeloid immature acute leukemia. *Blood*, *138*(9), 773. https://doi.org/10.1182/BLOOD.2020010510

Dick, J. E. (2003). Breast cancer stem cells revealed. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 3547–3549. https://doi.org/10.1073/PNAS.0830967100/ASSET/6F9D3687-FC14-4A54-AEF1-88DE436A96F5/ASSETS/GRAPHIC/PQ0830967002.JPEG

DiNardo, C. D., Jonas, B. A., Pullarkat, V., Thirman, M. J., Garcia, J. S., Wei, A. H., Konopleva, M., Döhner, H., Letai, A., Fenaux, P., Koller, E., Havelange, V., Leber, B., Esteve, J., Wang, J., Pejsa, V., Hájek, R., Porkka, K., Illés, Á., … Pratz, K. W. (2020). Azacitidine and Venetoclax in Previously Untreated Acute Myeloid Leukemia. *The New England Journal of Medicine*, *383*(7), 617–629. https://doi.org/10.1056/NEJMOA2012971

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., … Dipersio, J. F. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, *481*(7382), 506–510. https://doi.org/10.1038/NATURE10738

Diniz, M. G., França, J. A., Vilas-Boas, F. A. S., de Souza, F. T. A., Calin, G. A., Gomez, R. S., de Sousa, S. F., & Gomes, C. C. (2019). The long noncoding RNA KIAA0125 is upregulated in ameloblastomas. *Pathology, Research and Practice*, *215*(3), 466–469. https://doi.org/10.1016/J.PRP.2018.12.030

Disney, M. D., Winkelsas, A. M., Velagapudi, S. P., Southern, M., Fallahi, M., & Childs-Disney, J. L. (2016). Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chemical Biology*, *11*(6), 1720–1728. https://doi.org/10.1021/ACSCHEMBIO.6B00001

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108. https://doi.org/10.1038/NATURE11233

Döhner, H., Wei, A. H., Appelbaum, F. R., Craddock, C., DiNardo, C. D., Dombret, H., Ebert, B. L., Fenaux, P., Godley, L. A., Hasserjian, R. P., Larson, R. A., Levine, R. L., Miyazaki, Y., Niederwieser, D., Ossenkoppele, G., Röllig, C., Sierra, J.,

Stein, E. M., Tallman, M. S., … Löwenberg, B. (2022). Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood*, *140*(12), 1345–1377. https://doi.org/10.1182/BLOOD.2022016867

Döhner, H., Weisdorf, D. J., & Bloomfield, C. D. (2015a). Acute Myeloid Leukemia. *The New England Journal of Medicine*, *373*(12), 1136–1152. https://doi.org/10.1056/NEJMRA1406184

Döhner, H., Weisdorf, D. J., & Bloomfield, C. D. (2015b). Acute Myeloid Leukemia. *New England Journal of Medicine*, *373*(12), 1136–1152. https://doi.org/10.1056/NEJMRA1406184

Doulatov, S., Notta, F., Laurenti, E., & Dick, J. E. (2012). Hematopoiesis: a human perspective. *Cell Stem Cell*, *10*(2), 120–136. https://doi.org/10.1016/J.STEM.2012.01.006

Draper, J. E., Sroczynska, P., Tsoulaki, O., Leong, H. S., Fadlullah, M. Z. H., Miller, C., Kouskoff, V., & Lacaud, G. (2016). RUNX1B Expression Is Highly Heterogeneous and Distinguishes Megakaryocytic and Erythroid Lineage Fate in Adult Mouse Hematopoiesis. *PLoS Genetics*, *12*(1). https://doi.org/10.1371/JOURNAL.PGEN.1005814

Drexler, H. G., Matsuo, Y., & MacLeod, R. A. F. (2004). Malignant hematopoietic cell lines: In vitro models for the study of erythroleukemia. *Leukemia Research*, *28*(12), 1243–1251. https://doi.org/10.1016/j.leukres.2004.03.022

Du, W., He, J., Zhou, W., Shu, S., Li, J., Liu, W., Deng, Y., Lu, C., Lin, S., Ma, Y., He, Y., Zheng, J., Zhu, J., Bai, L., Li, X., Yao, J., Gu, S., Li, H., Guo, A., … Hu, D. (2019). High IL2RA mRNA expression is an independent adverse prognostic biomarker in core binding factor and intermediate-risk acute myeloid leukemia. *Journal of Translational Medicine*, *17*(1), 191. https://doi.org/10.1186/S12967-019-1926-Z

DuBridge, R. B., Tang, P., Hsia, H. C., Leong, P.-M., Miller, J. H., & Calos, M. P. (1987). Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Molecular and Cellular Biology*, *7*(1), 379–387. https://doi.org/10.1128/MCB.7.1.379-387.1987

Duggimpudi, S., Kloetgen, A., Maney, S. K., Münch, P. C., Hezaveh, K., Shaykhalishahi, H., Hoyer, W., McHardy, A. C., Lang, P. A., Borkhardt, A., & Hoell, J. I. (2018). Transcriptome-wide analysis uncovers the targets of the RNA-binding protein MSI2 and effects of MSI2's RNA-binding activity on IL-6 signaling. *The Journal of Biological Chemistry*, *293*(40), 15359–15369. https://doi.org/10.1074/JBC.RA118.002243

Dziennis, S., Van Etten, R. A., Pahl, H. L., Morris, D. L., Rothstein, T. L., Blosch, C. M., Perlmutter, R. M., & Tenen, D. G. (1995). The CD11b Promoter Directs High-Level Expression of Reporter Genes in Macrophages in Transgenic Mice. *Blood*, *85*(2), 319–329. https://doi.org/10.1182/BLOOD.V85.2.319.319

Ebralidze, A. K., Guibal, F. C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M. A., Petkova, V., Rosenbauer, F., Huang, G., Dayaram, T., Klupp, J., O'Brien, K. B., Will, B., Hoogenkamp, M., Borden, K. L. B., Bonifer, C., & Tenen, D. G. (2008). PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes & Development*, *22*(15), 2085–2092. https://doi.org/10.1101/GAD.1654808

Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nature Reviews. Genetics*, *2*(12), 919–929. https://doi.org/10.1038/35103511

Elisaphenko, E. A., Kolesnikov, N. N., Shevchenko, A. I., Rogozin, I. B., Nesterova, T. B., Brockdorff, N., & Zakian, S. M. (2008). A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements. *PLOS ONE*, *3*(6), e2521. https://doi.org/10.1371/JOURNAL.PONE.0002521

Elsayed, A. H., Rafiee, R., Cao, X., Raimondi, S., Downing, J. R., Ribeiro, R., Fan, Y., Gruber, T. A., Baker, S., Klco, J., Rubnitz, J. E., Pounds, S., & Lamba, J. K. (2020). A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. *Leukemia*, *34*(3), 735–745. https://doi.org/10.1038/S41375-019-0604-8

Emdal, K. B., Palacio-Escat, N., Wigerup, C., Eguchi, A., Nilsson, H., Bekker-Jensen, D. B., Rönnstrand, L., Kazi, J. U., Puissant, A., Itzykson, R., Saez-Rodriguez, J., Masson, K., Blume-Jensen, P., & Olsen, J. V. (2022). Phosphoproteomics of primary AML patient samples reveals rationale for AKT combination therapy and p53 context to overcome selinexor resistance. *Cell Reports*, *40*(6). https://doi.org/10.1016/J.CELREP.2022.111177

Eppert, K., Takenaka, K., Lechman, E. R., Waldron, L., Nilsson, B., Van Galen, P., Metzeler, K. H., Poeppl, A., Ling, V., Beyene, J., Canty, A. J., Danska, J. S., Bohlander, S. K., Buske, C., Minden, M. D., Golub, T. R., Jurisica, I., Ebert, B. L., & Dick, J. E. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. *Nature Medicine*, *17*(9), 1086–1093. https://doi.org/10.1038/NM.2415

Fabre, M. A., de Almeida, J. G., Fiorillo, E., Mitchell, E., Damaskou, A., Rak, J., Orrù, V., Marongiu, M., Chapman, M. S., Vijayabaskar, M. S., Baxter, J., Hardy, C., Abascal, F., Williams, N., Nangalia, J., Martincorena, I., Campbell, P. J., McKinney, E. F., Cucca, F., … Vassiliou, G. S. (2022). The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature 2022 606:7913*, *606*(7913), 335–342. https://doi.org/10.1038/s41586-022-04785-z

Fan, A. C., Nakauchi, Y., Bai, L., Azizi, A., Nuno, K. A., Zhao, F., Köhnke, T., Karigane, D., Cruz-Hernandez, D., Reinisch, A., Khatri, P., & Majeti, R. (2023). RUNX1 loss renders hematopoietic and leukemic cells dependent on IL-3 and sensitive to JAK inhibition. *Journal of Clinical Investigation*, *133*(19). https://doi.org/10.1172/JCI167053

Fang, J., Bolanos, L. C., Choi, K., Liu, X., Christie, S., Akunuru, S., Kumar, R., Wang, D., Chen, X., Greis, K. D., Stoilov, P., Filippi, M. D., Maciejewski, J. P., Garcia-Manero, G., Weirauch, M. T., Salomonis, N., Geiger, H., Zheng, Y., & Starczynowski, D. T. (2017). Ubiquitination of hnRNPA1 by TRAF6 links chronic innate immune signaling with myelodysplasia. *Nature Immunology*, *18*(2), 236–245. https://doi.org/10.1038/NI.3654

Fang, T., Lv, H., Wu, F., Wang, C., Li, T., Lv, G., Tang, L., Guo, L., Tang, S., Cao, D., Wu, M., Yang, W., & Wang, H. (2017). Musashi 2 contributes to the stemness and chemoresistance of liver cancer stem cells via LIN28A activation. *Cancer Letters*, *384*, 50–59. https://doi.org/10.1016/J.CANLET.2016.10.007

Farge, T., Saland, E., de Toni, F., Aroua, N., Hosseini, M., Perry, R., Bosc, C., Sugita, M., Stuani, L., Fraisse, M., Scotland, S., Larrue, C., Boutzen, H., Féliu, V., Nicolau-Travers, M. L., Cassant-Sourdy, S., Broin, N., David, M., Serhan, N., … Sarry, J. E. (2017). Chemotherapy Resistant Human Acute Myeloid Leukemia Cells are Not Enriched for Leukemic Stem Cells but Require Oxidative Metabolism. *Cancer Discovery*, *7*(7), 716. https://doi.org/10.1158/2159-8290.CD-16-0441

Ferrara, F., & Schiffer, C. A. (2013). Acute myeloid leukaemia in adults. *Lancet (London, England)*, *381*(9865), 484–495. https://doi.org/10.1016/S0140-6736(12)61727-9

Ferrell, P. I., Xi, J., Ma, C., Adlakha, M., & Kaufman, D. S. (2015). The RUNX1 +24 enhancer and P1 promoter identify a unique subpopulation of hematopoietic progenitor cells derived from human pluripotent stem cells. *Stem Cells (Dayton, Ohio)*, *33*(4), 1130–1141. https://doi.org/10.1002/STEM.1940

Fisher, C. L., Pineault, N., Brookes, C., Helgason, C. D., Ohta, H., Bodner, C., Hess, J. L., Humphries, R. K., & Brock, H. W. (2010). Loss-of-function Additional sex combs like 1 mutations disrupt hematopoiesis but do not cause severe myelodysplasia or leukemia. *Blood*, *115*(1), 38–46. https://doi.org/10.1182/BLOOD-2009-07-230698

Flynn, M. J., & Hartley, J. A. (2017). The emerging role of anti-CD25 directed therapies as both immune modulators and targeted agents in cancer. *British Journal of Haematology*, *179*(1), 20–35. https://doi.org/10.1111/BJH.14770

Forouzanfar, M., Lachinani, L., Dormiani, K., Nasr-Esfahani, M. H., Gure, A. O., & Ghaedi, K. (2020). Intracellular functions of RNA-binding protein, Musashi1, in stem and cancer cells. *Stem Cell Research & Therapy*, *11*(1). https://doi.org/10.1186/S13287-020-01703-W

Fujiwara, S. I., Muroi, K., Yamamoto, C., Hatano, K., Okazuka, K., Sato, K., Oh, I., Ohmine, K., Suzuki, T., & Ozawa, K. (2017). CD25 as an adverse prognostic factor in elderly patients with acute myeloid leukemia. *Hematology (Amsterdam, Netherlands)*, *22*(6), 347–353. https://doi.org/10.1080/10245332.2016.1276240

Gaidzik, V. I., Teleanu, V., Papaemmanuil, E., Weber, D., Paschka, P., Hahn, J., Wallrabenstein, T., Kolbinger, B., Köhne, C. H., Horst, H. A., Brossart, P., Held, G., Kündgen, A., Ringhoffer, M., Götze, K., Rummel, M., Gerstung, M., Campbell, P., Kraus, J. M., … Döhner, H. (2016). RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. *Leukemia*, *30*(11), 2160–2168. https://doi.org/10.1038/LEU.2016.126

Gal, H., Amariglio, N., Trakhtenbrot, L., Jacob-Hirsh, J., Margalit, O., Avigdor, A., Nagler, A., Tavor, S., Ein-Dor, L., Lapidot, T., Domany, E., Rechavi, G., & Givol, D. (2006). Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells. *Leukemia*, *20*(12), 2147–2154. https://doi.org/10.1038/SJ.LEU.2404401

Gao, S., Zhou, B., Li, H., Huang, X., Wu, Y., Xing, C., Yu, X., & Ji, Y. (2018). Long noncoding RNA HOTAIR promotes the self-renewal of leukemia stem cells through epigenetic silencing of p15. *Experimental Hematology*, *67*, 32-40.e3. https://doi.org/10.1016/J.EXPHEM.2018.08.005

Gary Gilliland, D., & Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia. *Blood*, *100*(5), 1532–1542. https://doi.org/10.1182/BLOOD-2002-02-0492

Garzon, R., Volinia, S., Papaioannou, D., Nicolet, D., Kohlschmidt, J., Yan, P. S., Mrózek, K., Bucci, D., Carroll, A. J., Baer, M. R., Wetzler, M., Carter, T. H., Powell, B. L., Kolitz, J. E., Moore, J. O., Eisfeld, A. K., Blachly, J. S., Blum, W., Caligiuri, M. A., … Bloomfield, C. D. (2014). Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18679–18684. https://doi.org/10.1073/PNAS.1422050112/SUPPL_FILE/PNAS.1422050112.ST08.DOCX

Gasparini, C., Celeghini, C., Monasta, L., & Zauli, G. (2014). NF-κB pathways in hematological malignancies. *Cellular and Molecular Life Sciences : CMLS*, *71*(11), 2083–2102. https://doi.org/10.1007/S00018-013-1545-4

Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., … McCarroll, S. A. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England Journal of Medicine*, *371*(26), 2477–2487. https://doi.org/10.1056/NEJMOA1409405

Gerritsen, M., Yi, G., Tijchon, E., Kuster, J., Schuringa, J. J., Martens, J. H. A., & Vellenga, E. (2019). RUNX1 mutations enhance self-renewal and block granulocytic differentiation in human in vitro models and primary AMLs. *Blood Advances*, *3*(3), 320–332. https://doi.org/10.1182/BLOODADVANCES.2018024422

Ghozi, M. C., Bernstein, Y., Negreanu, V., Levanon, D., & Groner, Y. (1996). Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(5), 1935–1940. https://doi.org/10.1073/PNAS.93.5.1935

Gialesaki, S., Bräuer-Hartmann, D., Issa, H., Bhayadia, R., Alejo-Valle, O., Verboon, L., Schmell, A. L., Laszig, S., Regényi, E., Schuschel, K., Labuhn, M., Ng, M., Winkler, R., Ihling, C., Sinz, A., Glaß, M., Hüttelmaier, S., Matzk, S., Schmid, L., … Klusmann, J. H. (2023). RUNX1 isoform disequilibrium promotes the development of trisomy 21–associated myeloid leukemia. *Blood*, *141*(10), 1105–1118. https://doi.org/10.1182/BLOOD.2022017619

Gibson, C. J., Lindsley, R. C., Tchekmedyian, V., Mar, B. G., Shi, J., Jaiswal, S., Bosworth, A., Francisco, L., He, J., Bansal, A., Morgan, E. A., Lacasce, A. S., Freedman, A. S., Fisher, D. C., Jacobsen, E., Armand, P., Alyea, E. P., Koreth, J., Ho, V., … Ebert, B. L. (2017). Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell

Transplantation for Lymphoma. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *35*(14), 1598–1605. https://doi.org/10.1200/JCO.2016.71.6712

Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K. A., Rout, R., Chaudhury, S., Gilkes, A., Knapper, S., Beldjord, K., Begum, S., Rose, S., Geddes, N., Griffiths, M., Standen, G., … Vyas, P. (2011). Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell*, *19*(1), 138–152. https://doi.org/10.1016/J.CCR.2010.12.012

Goldberg, A. D., Atallah, E., Rizzieri, D., Walter, R. B., Chung, K. Y., Spira, A., Stock, W., Tallman, M. S., Cruz, H. G., Boni, J., Havenith, K. E. G., Chao, G., Feingold, J. M., Wuerthner, J., & Solh, M. (2020). Camidanlumab tesirine, an antibody-drug conjugate, in relapsed/refractory CD25-positive acute myeloid leukemia or acute lymphoblastic leukemia: A phase I study. *Leukemia Research*, *95*. https://doi.org/10.1016/J.LEUKRES.2020.106385

Goldfarb, A. N. (2009). Megakaryocytic Programming by a Transciptional Regulatory Loop: A Circle Connecting RUNX1, GATA-1, and P-TEFb. *Journal of Cellular Biochemistry*, *107*(3), 377. https://doi.org/10.1002/JCB.22142

Gönen, M., Sun, Z., Figueroa, M. E., Patel, J. P., Abdel-Wahab, O., Racevskis, J., Ketterling, R. P., Fernandez, H., Rowe, J. M., Tallman, M. S., Melnick, A., Levine, R. L., & Paietta, E. (2012a). CD25 expression status improves prognostic risk classification in AML independent of established biomarkers: ECOG phase 3 trial, E1900. *Blood*, *120*(11), 2297. https://doi.org/10.1182/BLOOD-2012-02-414425

Gönen, M., Sun, Z., Figueroa, M. E., Patel, J. P., Abdel-Wahab, O., Racevskis, J., Ketterling, R. P., Fernandez, H., Rowe, J. M., Tallman, M. S., Melnick, A., Levine, R. L., & Paietta, E. (2012b). CD25 expression status improves prognostic risk classification in AML independent of established biomarkers: ECOG phase 3 trial, E1900. *Blood*, *120*(11), 2297–2306. https://doi.org/10.1182/BLOOD-2012-02-414425

Gourvest, M., De Clara, E., Wu, H. C., Touriol, C., Meggetto, F., De Thé, H., Pyronnet, S., Brousset, P., & Bousquet, M. (2021). A novel leukemic route of mutant NPM1 through nuclear import of the overexpressed long noncoding RNA LONA. *Leukemia 2021 35:10*, *35*(10), 2784–2798. https://doi.org/10.1038/s41375-021-01307-0

Goyal, A., Myacheva, K., Groß, M., Klingenberg, M., Duran Arqué, B., & Diederichs, S. (2017). Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic Acids Research*, *45*(3), e12. https://doi.org/10.1093/NAR/GKW883

Goyama, S., Schibler, J., & Mulloy, J. C. (2019). Alternative translation initiation generates the N-terminal truncated form of RUNX1 that retains hematopoietic activity. *Experimental Hematology*, *72*, 27–35. https://doi.org/10.1016/J.EXPHEM.2019.01.008

Griessinger, E., Vargaftig, J., Horswell, S., Taussig, D. C., Gribben, J., & Bonnet, D. (2018). Acute myeloid leukemia xenograft success prediction: Saving time. *Experimental Hematology*, *59*, 66. https://doi.org/10.1016/J.EXPHEM.2017.12.002

Groves, C. (1998). Systematics of tarsiers and lorises. *Primates*, *39*(1), 13–27. https://doi.org/10.1007/BF02557740

Guan, Y., & Hogge, D. E. (2000). Proliferative status of primitive hematopoietic progenitors from patients with acute myelogenous leukemia (AML). *Leukemia*, *14*(12), 2135–2141. https://doi.org/10.1038/SJ.LEU.2401975

Guo, C. J., Ma, X. K., Xing, Y. H., Zheng, C. C., Xu, Y. F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G. G., Yang, L., & Chen, L. L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*, *181*(3), 621-636.e22. https://doi.org/10.1016/J.CELL.2020.03.006

Guo, H., Ma, O., Speck, N. A., & Friedman, A. D. (2012). Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis. *Blood*, *119*(19), 4408–4418. https://doi.org/10.1182/BLOOD-2011-12-397091

Guo, J. K., Blanco, M. R., Walkup, W. G., Bonesteele, G., Urbinati, C. R., Banerjee, A. K., Chow, A., Ettlin, O., Strehle, M., Peyda, P., Amaya, E., Trinh, V., & Guttman, M. (2024). Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. *Molecular Cell*, *84*(7), 1271-1289.e12. https://doi.org/10.1016/J.MOLCEL.2024.01.026

Guo, S., Li, B., Xu, X., Wang, W., Wang, S., Lv, T., & Wang, H. (2020). Construction of a 14-lncRNA risk score system predicting survival of children with acute myelocytic leukemia. *Experimental and Therapeutic Medicine*, *20*(2), 1521. https://doi.org/10.3892/ETM.2020.8846

Guo, Z., Wang, A., Zhang, W., Levit, M., Gao, Q., Barberis, C., Tabart, M., Zhang, J., Hoffmann, D., Wiederschain, D., Rocnik, J., Sun, F., Murtie, J., Lengauer, C., Gross, S., Zhang, B., Cheng, H., Patel, V., Schio, L., … Huang, S. M. A. (2014). PIM inhibitors target CD25-positive AML cells through concomitant suppression of STAT5 activation and degradation of MYC oncogene. *Blood*, *124*(11), 1777–1789. https://doi.org/10.1182/BLOOD-2014-01-551234

Gutschner, T., Baas, M., & Diederichs, S. (2011). Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Research*, *21*(11), 1944. https://doi.org/10.1101/GR.122358.111

Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., Young, G., Lucas, A. B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J. L., Root, D. E., & Lander, E. S. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, *477*(7364), 295–300. https://doi.org/10.1038/NATURE10398

Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, *482*(7385), 339–346. https://doi.org/10.1038/NATURE10887

Guzman, M. L., Neering, S. J., Upchurch, D., Grimes, B., Howard, D. S., Rizzieri, D. A., Luger, S. M., & Jordan, C. T. (2001). Nuclear factor-kappaB is constitutively activated in primitive human acute myelogenous leukemia cells. *Blood*, *98*(8), 2301–2307. https://doi.org/10.1182/BLOOD.V98.8.2301

Haferlach, T., Kohlmann, A., Wieczorek, L., Basso, G., Te Kronnie, G., Béné, M. C., De Vos, J., Hernández, J. M., Hofmann, W. K., Mills, K. I., Gilkes, A., Chiaretti, S., Shurtleff, S. A., Kipps, T. J., Rassenti, L. Z., Yeoh, A. E., Papenhausen, P. R., Liu, W. M., Williams, P. M., & Foà, R. (2010). Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *28*(15), 2529–2537. https://doi.org/10.1200/JCO.2009.23.4732

Han, Y., Ye, A., Zhang, Y., Cai, Z., Wang, W., Sun, L., Jiang, S., Wu, J., Yu, K., & Zhang, S. (2015). Musashi-2 Silencing Exerts Potent Activity against Acute Myeloid Leukemia and Enhances Chemosensitivity to Daunorubicin. *PLoS ONE*, *10*(8). https://doi.org/10.1371/JOURNAL.PONE.0136484

Hanekamp, D., Denys, B., Kaspers, G. J. L., te Marvelde, J. G., Schuurhuis, G. J., De Haas, V., De Moerloose, B., de Bont, E. S., Zwaan, C. M., de Jong, A., Depreter, B., Lammens, T., Philippé, J., Cloos, J., & van der Velden, V. H. J. (2018). Leukaemic stem cell load at diagnosis predicts the development of relapse in young acute myeloid leukaemia patients. *British Journal of Haematology*, *183*(3), 512–516. https://doi.org/10.1111/BJH.14991

Hannan, A. J. (2018). Tandem repeats mediating genetic plasticity in health and disease. *Nature Reviews. Genetics*, *19*(5), 286–298. https://doi.org/10.1038/NRG.2017.115

Harada, Y., Inoue, D., Ding, Y., Imagawa, J., Doki, N., Matsui, H., Yahata, T., Matsushita, H., Ando, K., Sashida, G., Iwama, A., Kitamura, T., & Harada, H. (2013). RUNX1/AML1 mutant collaborates with BMI1 overexpression in the development of human and murine myelodysplastic syndromes. *Blood*, *121*(17), 3434–3446. https://doi.org/10.1182/BLOOD-2012-06-434423

Hartman, K. R., LaRussa, V. F., Rothwell, S. W., Atolagbe, T. O., Ward, F. T., & Klipple, G. (1994). Antibodies to Myeloid Precursor Cells in Autoimmune Neutropenia. *Blood*, *84*(2), 625–631. https://doi.org/10.1182/BLOOD.V84.2.625.625

Hattori, A., McSkimming, D., Kannan, N., & Ito, T. (2017). RNA binding protein MSI2 positively regulates FLT3 expression in myeloid leukemia. *Leukemia Research*, *54*, 47–54. https://doi.org/10.1016/J.LEUKRES.2017.01.015

Havens, M. A., & Hastings, M. L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Research*, *44*(14), 6549. https://doi.org/10.1093/NAR/GKW533

Hentze, M. W., Castello, A., Schwarzl, T., & Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature Reviews. Molecular Cell Biology*, *19*(5), 327–341. https://doi.org/10.1038/NRM.2017.130

Heuser, M., Freeman, S. D., Ossenkoppele, G. J., Buccisano, F., Hourigan, C. S., Ngai, L. L., Tettero, J. M., Bachas, C., Baer, C., Béné, M. C., Bücklein, V., Czyz, A., Denys, B., Dillon, R., Feuring-Buske, M., Guzman, M. L., Haferlach, T., Han, L., Herzig, J. K., … Cloos, J. (2021). 2021 Update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD Working Party. *Blood*, *138*(26), 2753. https://doi.org/10.1182/BLOOD.2021013626

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Reports*, *11*(7), 1110–1122. https://doi.org/10.1016/J.CELREP.2015.04.023

Higuchi, M., O'Brien, D., Kumaravelu, P., Lenny, N., Yeoh, E. J., & Downing, J. R. (2002). Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell*, *1*(1), 63–74. https://doi.org/10.1016/S1535-6108(02)00016-8

Hinai, A. A., & Valk, P. J. M. (2016). Review: Aberrant EVI1 expression in acute myeloid leukaemia. *British Journal of Haematology*, *172*(6), 870–878. https://doi.org/10.1111/BJH.13898

Ho, M. M., Hogge, D. E., & Ling, V. (2008). MDR1 and BCRP1 expression in leukemic progenitors correlates with chemotherapy response in acute myeloid leukemia. *Experimental Hematology*, *36*(4), 433–442. https://doi.org/10.1016/J.EXPHEM.2007.11.014

Hoogenkamp, M., Krysinska, H., Ingram, R., Huang, G., Barlow, R., Clarke, D., Ebralidze, A., Zhang, P., Tagoh, H., Cockerill, P. N., Tenen, D. G., & Bonifer, C. (2007). The Pu.1 Locus Is Differentially Regulated at the Level of Chromatin Structure and Noncoding Transcription by Alternate Mechanisms at Distinct Developmental Stages of Hematopoiesis. *Molecular and Cellular Biology*, *27*(21), 7425. https://doi.org/10.1128/MCB.00905-07

Hope, K. J., Jin, L., & Dick, J. E. (2004a). Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nature Immunology*, *5*(7), 738–743. https://doi.org/10.1038/NI1080

Hope, K. J., Jin, L., & Dick, J. E. (2004b). Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nature Immunology*, *5*(7), 738–743. https://doi.org/10.1038/NI1080

Hosen, N., Park, C. Y., Tatsumi, N., Oji, Y., Sugiyama, H., Gramatzki, M., Krensky, A. M., & Weissman, I. L. (2007). CD96 is a leukemic stem cell-specific marker in human acute myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(26), 11008–11013. https://doi.org/10.1073/PNAS.0704271104

Hourigan, C. S., Gale, R. P., Gormley, N. J., Ossenkoppele, G. J., & Walter, R. B. (2017). Measurable residual disease testing in acute myeloid leukaemia. *Leukemia*, *31*(7), 1482–1490. https://doi.org/10.1038/LEU.2017.113

Hu, N., Chen, L., Wang, C., & Zhao, H. (2019). MALAT1 knockdown inhibits proliferation and enhances cytarabine chemosensitivity by upregulating miR-96 in acute myeloid leukemia cells. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, *112*. https://doi.org/10.1016/J.BIOPHA.2019.108720

Huang, B. J., Smith, J. L., Farrar, J. E., Wang, Y. C., Umeda, M., Ries, R. E., Leonti, A. R., Crowgey, E., Furlan, S. N., Tarlock, K., Armendariz, M., Liu, Y., Shaw, T. I., Wei, L., Gerbing, R. B., Cooper, T. M., Gamis, A. S., Aplenc, R., Kolb, E. A., … Meshinchi, S. (2022). Integrated stem cell signature and cytomolecular risk determination in pediatric acute myeloid leukemia. *Nature Communications 2022 13:1*, *13*(1), 1–11. https://doi.org/10.1038/s41467-022-33244-6

Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., Stephens, R., Baseler, M. W., Lane, H. C., & Lempicki, R. A. (2007). The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology*, *8*(9), 1–16. https://doi.org/10.1186/GB-2007-8-9-R183/TABLES/3

Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Growney, J. D., Shivdasani, R. A., Gilliland, D. G., Speck, N. A., Nimer, S. D., & Tenen, D. G. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nature Genetics*, *40*(1), 51–60. https://doi.org/10.1038/NG.2007.7

Hung, S. Y., Lin, C. C., Hsu, C. L., Yao, C. Y., Wang, Y. H., Tsai, C. H., Hou, H. A., Chou, W. C., & Tien, H. F. (2021). The expression levels of long non-coding RNA KIAA0125 are associated with distinct clinical and biological features in myelodysplastic syndromes. *British Journal of Haematology*, *192*(3), 589–598. https://doi.org/10.1111/BJH.17231

Ichikawa, M., Asai, T., Saito, T., Yamamoto, G., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Hirai, H., Ogawa, S., & Kurokawa, M. (2004). AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nature Medicine*, *10*(3), 299–304. https://doi.org/10.1038/NM997

Imperato, M. R., Cauchy, P., Obier, N., & Bonifer, C. (2015). The RUNX1-PU.1 axis in the control of hematopoiesis. *International Journal of Hematology*, *101*(4), 319–329. https://doi.org/10.1007/S12185-015-1762-8

Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N., Fukata, M., Miyamoto, T., Lyons, B., Ohshima, K., Uchida, N., Taniguchi, S., Ohara, O., Akashi, K., Harada, M., & Shultz, L. D. (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nature Biotechnology*, *25*(11), 1315–1321. https://doi.org/10.1038/NBT1350

Ito, T., Kwon, H. Y., Zimdahl, B., Congdon, K. L., Blum, J., Lento, W. E., Zhao, C., Lagoo, A., Gerrard, G., Foroni, L., Goldman, J., Goh, H., Kim, S. H., Kim, D. W., Chuah, C., Oehler, V. G., Radich, J. P., Jordan, C. T., & Reya, T. (2010). Regulation of myeloid leukaemia by the cell-fate determinant Musashi. *Nature*, *466*(7307), 765–768. https://doi.org/10.1038/NATURE09171

Ivey, A., Hills, R. K., Simpson, M. A., Jovanovic, J. V., Gilkes, A., Grech, A., Patel, Y., Bhudia, N., Farah, H., Mason, J., Wall, K., Akiki, S., Griffiths, M., Solomon, E., McCaughan, F., Linch, D. C., Gale, R. E., Vyas, P., Freeman, S. D., … Grimwade, D. (2016). Assessment of Minimal Residual Disease in Standard-Risk AML. *The New England Journal of Medicine*, *374*(5), 422–433. https://doi.org/10.1056/NEJMOA1507471

Iwasaki, M., Liedtke, M., Gentles, A. J., & Cleary, M. L. (2015). CD93 Marks a Non-Quiescent Human Leukemia Stem Cell Population and Is Required for Development of MLL-Rearranged Acute Myeloid Leukemia. *Cell Stem Cell*, *17*(4), 412–421. https://doi.org/10.1016/J.STEM.2015.08.008

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T. R., Prensner, J. R., Evans, J. R., Zhao, S., Poliakov, A., Cao, X., Dhanasekaran, S. M., Wu, Y. M., Robinson, D. R., Beer, D. G., Feng, F. Y., Iyer, H. K., & Chinnaiyan, A. M. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics*, *47*(3), 199–208. https://doi.org/10.1038/NG.3192

Jagani, Z., Mora-Blanco, E. L., Sansam, C. G., McKenna, E. S., Wilson, B., Chen, D., Klekota, J., Tamayo, P., Nguyen, P. T. L., Tolstorukov, M., Park, P. J., Cho, Y. J., Hsiao, K., Buonamici, S., Pomeroy, S. L., Mesirov, J. P., Ruffner, H., Bouwmeester, T., Luchansky, S. J., … Dorsch, M. (2010). Loss of the tumor suppressor Snf5 leads to aberrant activation of the Hedgehog-Gli pathway. *Nature Medicine*, *16*(12), 1429–1434. https://doi.org/10.1038/NM.2251

Jagannathan-Bogdan, M., & Zon, L. I. (2013). Hematopoiesis. *Development (Cambridge, England)*, *140*(12), 2463. https://doi.org/10.1242/DEV.083147

Jaiswal, S., & Ebert, B. L. (2019). Clonal hematopoiesis in human aging and disease. *Science (New York, N.Y.)*, *366*(6465). https://doi.org/10.1126/SCIENCE.AAN4673

Janga, H., Aznaourova, M., Boldt, F., Damm, K., Grünweller, A., & Schulte, L. N. (2018). Cas9-mediated excision of proximal DNaseI/H3K4me3 signatures confers robust silencing of microRNA and long non-coding RNA genes. *PLOS ONE*, *13*(2), e0193066. https://doi.org/10.1371/JOURNAL.PONE.0193066

Jean-Philippe, J., Paz, S., & Caputi, M. (2013). hnRNP A1: the Swiss army knife of gene expression. *International Journal of Molecular Sciences*, *14*(9), 18999–19024. https://doi.org/10.3390/IJMS140918999

Jiang, Z., Liu, T., Wang, Y., Li, J., Guo, & L., & Guo, L. (2024). *Effect of lncRNA XIST on acute myeloid leukemia cells via miR-142-5p-PFKP axis*. https://doi.org/10.1080/16078454.2024.2306444

Jilani, I., Estey, E., Huh, Y., Joe, Y., Manshouri, T., Yared, M., Giles, F., Kantarjian, H., Cortes, J., Thomas, D., Keating, M., Freireich, E., & Albitar, M. (2002). Differences in CD33 intensity between various myeloid neoplasms. *American Journal of Clinical Pathology*, *118*(4), 560–566. https://doi.org/10.1309/1WMW-CMXX-4WN4-T55U

Jiménez, A., Carrasco, L., & Vázquez, D. (1977). Enzymic and nonenzymic translocation by yeast polysomes. Site of action of a number of inhibitors. *Biochemistry*, *16*(21), 4727–4730. https://doi.org/10.1021/BI00640A030

Jin, L., Hope, K. J., Zhai, Q., Smadja-Joffe, F., & Dick, J. E. (2006). Targeting of CD44 eradicates human acute myeloid leukemic stem cells. *Nature Medicine*, *12*(10), 1167–1174. https://doi.org/10.1038/NM1483

Jin, L., Lee, E. M., Ramshaw, H. S., Busfield, S. J., Peoppl, A. G., Wilkinson, L., Guthridge, M. A., Thomas, D., Barry, E. F., Boyd, A., Gearing, D. P., Vairo, G., Lopez, A. F., Dick, J. E., & Lock, R. B. (2009). Monoclonal antibody-mediated targeting of CD123, IL-3 receptor alpha chain, eliminates human acute myeloid leukemic stem cells. *Cell Stem Cell*, *5*(1), 31–42. https://doi.org/10.1016/J.STEM.2009.04.018

Joaquina Delás, M., Sabin, L. R., Dolzhenko, E., Knott, S. R. V., Maravilla, E. M., Jackson, B. T., Wild, S. A., Kovacevic, T., Stork, E. M., Zhou, M., Erard, N., Lee, E., Kelley, D. R., Roth, M., Barbosa, I. A. M., Zuber, J., Rinn, J. L., Smith, A. D., & Hannon, G. J. (2017). lncRNA requirements for mouse acute myeloid leukemia and normal differentiation. *ELife*, *6*. https://doi.org/10.7554/ELIFE.25607

Joshi, I., Carney, W. P., & Rock, E. P. (2023). Utility of monocyte HLA-DR and rationale for therapeutic GM-CSF in sepsis immunoparalysis. *Frontiers in Immunology*, *14*. https://doi.org/10.3389/FIMMU.2023.1130214

Jung, D., Giallourakis, C., Mostoslavsky, R., & Alt, F. W. (2006). Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annual Review of Immunology*, *24*, 541–570. https://doi.org/10.1146/ANNUREV.IMMUNOL.23.021704.115830

Jung, N., Dai, B., Gentles, A. J., Majeti, R., & Feinberg, A. P. (2015a). An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nature Communications*, *6*. https://doi.org/10.1038/NCOMMS9489

Jung, N., Dai, B., Gentles, A. J., Majeti, R., & Feinberg, A. P. (2015b). An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nature Communications*, *6*. https://doi.org/10.1038/NCOMMS9489

Kageyama, Y., Miwa, H., Arakawa, R., Tawara, I., Ohishi, K., Masuya, M., Nakase, K., & Katayama, N. (2018). Expression of CD25 fluctuates in the leukemia-initiating cell population of CD25-positive AML. *PLoS ONE*, *13*(12). https://doi.org/10.1371/JOURNAL.PONE.0209295

Kagoya, Y., Yoshimi, A., Kataoka, K., Nakagawa, M., Kumano, K., Arai, S., Kobayashi, H., Saito, T., Iwakura, Y., & Kurokawa, M. (2014). Positive feedback between NF-κB and TNF-α promotes leukemia-initiating cell capacity. *The Journal of Clinical Investigation*, *124*(2), 528–542. https://doi.org/10.1172/JCI68101

Kalev-Zylinska, M. L., Horsfield, J. A., Flores, M. V. C., Postlethwait, J. H., Vitas, M. R., Baas, A. M., Crosier, P. S., & Crosier, K. E. (2002). Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. *Development (Cambridge, England)*, *129*(8), 2015–2030. https://doi.org/10.1242/DEV.129.8.2015

Kantarjian, H., Kadia, T., DiNardo, C., Daver, N., Borthakur, G., Jabbour, E., Garcia-Manero, G., Konopleva, M., & Ravandi, F. (2021). Acute myeloid leukemia: current progress and future directions. *Blood Cancer Journal*, *11*(2). https://doi.org/10.1038/S41408-021-00425-3

Karmakar, S., Ramirez, O., Paul, K. V., Gupta, A. K., Kumari, V., Botti, V., De Los Mozos, I. R., Neuenkirchen, N., Ross, R. J., Karanicolas, J., Neugebauer, K. M., & Pillai, M. M. (2022). Integrative genome-wide analysis reveals EIF3A as a key downstream regulator of translational repressor protein Musashi 2 (MSI2). *NAR Cancer*, *4*(2). https://doi.org/10.1093/NARCAN/ZCAC015

Kawahara, H., Imai, T., Imataka, H., Tsujimoto, M., Matsumoto, K., & Okano, H. (2008). Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. *The Journal of Cell Biology*, *181*(4), 639–653. https://doi.org/10.1083/JCB.200708004

Kelly, L. M., & Gilliland, D. G. (2002). Genetics of myeloid leukemias. *Annual Review of Genomics and Human Genetics*, *3*, 179–198. https://doi.org/10.1146/ANNUREV.GENOM.3.032802.115046

Kerpedjiev, P., Hammer, S., & Hofacker, I. L. (2015). Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics (Oxford, England)*, *31*(20), 3377–3379. https://doi.org/10.1093/BIOINFORMATICS/BTV372

Kharas, M. G., & Lengner, C. J. (2017). Stem cells, cancer and MUSASHI in blood and guts. *Trends in Cancer*, *3*(5), 347. https://doi.org/10.1016/J.TRECAN.2017.03.007

Kharas, M. G., Lengner, C. J., Al-Shahrour, F., Bullinger, L., Ball, B., Zaidi, S., Morgan, K., Tam, W., Paktinat, M., Okabe, R., Gozo, M., Einhorn, W., Lane, S. W., Scholl, C., Fröhling, S., Fleming, M., Ebert, B. L., Gilliland, D. G., Jaenisch, R., & Daley, G. Q. (2010a). Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. *Nature Medicine*, *16*(8), 903–908. https://doi.org/10.1038/NM.2187

Kharas, M. G., Lengner, C. J., Al-Shahrour, F., Bullinger, L., Ball, B., Zaidi, S., Morgan, K., Tam, W., Paktinat, M., Okabe, R., Gozo, M., Einhorn, W., Lane, S. W., Scholl, C., Fröhling, S., Fleming, M., Ebert, B. L., Gilliland, D. G., Jaenisch, R., & Daley, G. Q. (2010b). Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. *Nature Medicine 2010 16:8*, *16*(8), 903–908. https://doi.org/10.1038/nm.2187

Kikushige, Y., Shima, T., Takayanagi, S. I., Urata, S., Miyamoto, T., Iwasaki, H., Takenaka, K., Teshima, T., Tanaka, T., Inagaki, Y., & Akashi, K. (2010). TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. *Cell Stem Cell*, *7*(6), 708–717. https://doi.org/10.1016/J.STEM.2010.11.014

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kirstetter, P., Schuster, M. B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E., Theilgaard-Mönch, K., Månsson, R., Pedersen, T. Å., Pabst, T., Schrock, E., Porse, B. T., Jacobsen, S. E. W., Bertone, P., Tenen, D. G., & Nerlov, C. (2008). Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. *Cancer Cell*, *13*(4), 299–310. https://doi.org/10.1016/J.CCR.2008.02.008

Koechlein, C. S., Harris, J. R., Lee, T. K., Weeks, J., Fox, R. G., Zimdahl, B., Ito, T., Blevins, A., Jung, S. H., Chute, J. P., Chourasia, A., Covert, M. W., & Reya, T. (2016). High-resolution imaging and computational analysis of haematopoietic cell dynamics in vivo. *Nature Communications 2016 7:1*, *7*(1), 1–14. https://doi.org/10.1038/ncomms12169

Koeffler, H. P., Billing, R., Lusis, A. J., Sparkes, R., & Golde, D. W. (1980). An undifferentiated variant derived from the human acute myelogenous leukemia cell line (KG-1). *Blood*, *56*(2), 265–273. https://doi.org/10.1182/blood.v56.2.265.265

Kondo, M., Weissman, I. L., & Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, *91*(5), 661–672. https://doi.org/10.1016/S0092-8674(00)80453-5

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O., & Zhang, F. (2014). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature 2014 517:7536*, *517*(7536), 583–588. https://doi.org/10.1038/nature14136

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., & Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, *35*(Web Server issue). https://doi.org/10.1093/NAR/GKM391

Koreth, J., Schlenk, R., Kopecky, K. J., Honda, S., Sierra, J., Djulbegovic, B. J., Wadleigh, M., DeAngelo, D. J., Stone, R. M., Sakamaki, H., Appelbaum, F. R., Döhner, H., Antin, J. H., Soiffer, R. J., & Cutler, C. (2009). Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: systematic review and meta-analysis of prospective clinical trials. *JAMA*, *301*(22), 2349–2361. https://doi.org/10.1001/JAMA.2009.813

Kozak, M. (1989). The scanning model for translation: an update. *The Journal of Cell Biology*, *108*(2), 229–241. https://doi.org/10.1083/JCB.108.2.229

Köhn, M., Lederer, M., Wächter, K., & Hüttelmaier, S. (2010). Near-infrared (NIR) dye-labeled RNAs identify binding of ZBP1 to the noncoding Y3-RNA. RNA (New York, N.Y.), 16(7), 1420–1428. https://doi.org/10.1261/rna.2152710

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: intimations of translational control. *The Journal of Cell Biology*, *115*(4), 887–903. https://doi.org/10.1083/JCB.115.4.887

Kreso, A., & Dick, J. E. (2014). Evolution of the cancer stem cell model. *Cell Stem Cell*, *14*(3), 275–291. https://doi.org/10.1016/J.STEM.2014.02.006

Kronstein-Wiedemann, R., & Tonn, T. (2019). Colony Formation: An Assay of Hematopoietic Progenitor Cells. *Methods in Molecular Biology (Clifton, N.J.)*, *2017*, 29–40. https://doi.org/10.1007/978-1-4939-9574-5_3

Kudinov, A. E., Karanicolas, J., Golemis, E. A., & Boumber, Y. (2017). Musashi RNA-Binding Proteins as Cancer Drivers and Novel Therapeutic Targets. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *23*(9), 2143–2153. https://doi.org/10.1158/1078-0432.CCR-16-2728

Kuo, Y. H., Landrette, S. F., Heilman, S. A., Perrat, P. N., Garrett, L., Liu, P. P., Le Beau, M. M., Kogan, S. C., & Castilla, L. H. (2006). Cbf beta-SMMHC induces distinct abnormal myeloid progenitors able to develop acute myeloid leukemia. *Cancer Cell*, *9*(1), 57–68. https://doi.org/10.1016/J.CCR.2005.12.014

Kwon, H. Y., Bajaj, J., Ito, T., Blevins, A., Konuma, T., Weeks, J., Lytle, N. K., Koechlein, C. S., Rizzieri, D., Chuah, C., Oehler, V. G., Sasik, R., Hardiman, G., & Reya, T. (2015). Tetraspanin 3 Is Required for the Development and Propagation of Acute Myelogenous Leukemia. *Cell Stem Cell*, *17*(2), 152–164. https://doi.org/10.1016/J.STEM.2015.06.006

Lagadinou, E. D., Sach, A., Callahan, K., Rossi, R. M., Neering, S. J., Minhajuddin, M., Ashton, J. M., Pei, S., Grose, V., O'Dwyer, K. M., Liesveld, J. L., Brookes, P. S., Becker, M. W., & Jordan, C. T. (2013). BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. *Cell Stem Cell*, *12*(3), 329–341. https://doi.org/10.1016/J.STEM.2012.12.013

Lan, L., Liu, H., Smith, A. R., Appelman, C., Yu, J., Larsen, S., Marquez, R. T., Wu, X., Liu, F. Y., Gao, P., Gowthaman, R., Karanicolas, J., De Guzman, R. N., Rogers, S., Aubé, J., Neufeld, K. L., & Xu, L. (2018). Natural product derivative Gossypolone inhibits Musashi family of RNA-binding proteins. *BMC Cancer*, *18*(1), 1–14. https://doi.org/10.1186/S12885-018-4704-Z/FIGURES/7

Lánczky, A., & Győrffy, B. (2021). Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. *Journal of Medical Internet Research*, *23*(7). https://doi.org/10.2196/27633

Langabeer, S. E., Gale, R. E., Rollinson, S. J., Morgan, G. J., & Linch, D. C. (2002). Mutations of the AML1 gene in acute myeloid leukemia of FAB types M0 and M7. *Genes, Chromosomes & Cancer*, *34*(1), 24–32. https://doi.org/10.1002/GCC.10031

Langston, A. W., & Gudas, L. J. (1992). Identification of a retinoic acid responsive enhancer 3' of the murine homeobox gene Hox-1.6. *Mechanisms of Development*, *38*(3), 217–227. https://doi.org/10.1016/0925-4773(92)90055-O

Lansdorp, P. M., Sutherland, H. J., & Eaves, C. J. (1990). Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. *The Journal of Experimental Medicine*, *172*(1), 363. https://doi.org/10.1084/JEM.172.1.363

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., & Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature 1994 367:6464*, *367*(6464), 645–648. https://doi.org/10.1038/367645a0

Laverdière, I., Boileau, M., Neumann, A. L., Frison, H., Mitchell, A., Ng, S. W. K., Wang, J. C. Y., Minden, M. D., & Eppert, K. (2018). Leukemic stem cell signatures identify novel therapeutics targeting acute myeloid leukemia. *Blood Cancer Journal*, *8*(6). https://doi.org/10.1038/S41408-018-0087-2

Lennox, K. A., & Behlke, M. A. (2016). Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Research*, *44*(2), 863–877. https://doi.org/10.1093/NAR/GKV1206

Levanon, D., Glusman, G., Bangsow, T., Ben-Asher, E., Male, D. A., Avidan, N., Bangsow, C., Hattori, M., Taylor, T. D., Taudien, S., Blechschmidt, K., Shimizu, N., Rosenthal, A., Sakaki, Y., Lancet, D., & Groner, Y. (2001). Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. *Gene*, *262*(1–2), 23–33. https://doi.org/10.1016/S0378-1119(00)00532-1

Levantini, E., Lee, S., Radomska, H. S., Hetherington, C. J., Alberich-Jorda, M., Amabile, G., Zhang, P., Gonzalez, D. A., Zhang, J., Basseres, D. S., Wilson, N. K., Koschmieder, S., Huang, G., Zhang, D. E., Ebralidze, A. K., Bonifer, C., Okuno, Y., Gottgens, B., & Tenen, D. G. (2011). RUNX1 regulates the CD34 gene in haematopoietic stem cells by mediating interactions with a distal regulatory element. *The EMBO Journal*, *30*(19), 4059–4070. https://doi.org/10.1038/EMBOJ.2011.285

Li, M., Li, A. Q., Zhou, S. L., Lv, H., Wei, P., & Yang, W. T. (2020a). RNA-binding protein MSI2 isoforms expression and regulation in progression of triple-negative breast cancer. *Journal of Experimental & Clinical Cancer Research : CR*, *39*(1). https://doi.org/10.1186/S13046-020-01587-X

Li, M., Li, A. Q., Zhou, S. L., Lv, H., Wei, P., & Yang, W. T. (2020b). RNA-binding protein MSI2 isoforms expression and regulation in progression of triple-negative breast cancer. *Journal of Experimental & Clinical Cancer Research : CR*, *39*(1). https://doi.org/10.1186/S13046-020-01587-X

Li, S. Q., Liu, J., Zhang, J., Wang, X. L., Chen, D., Wang, Y., Xu, Y. M., Huang, B., Lin, J., Li, J., & Wang, X. Z. (2020). Transcriptome profiling reveals the high incidence of hnRNPA1 exon 8 inclusion in chronic myeloid leukemia. *Journal of Advanced Research*, *24*, 301. https://doi.org/10.1016/J.JARE.2020.04.016

Li, Z., Wang, F., Tian, X., Long, J., Ling, B., Zhang, W., Xu, J., & Liang, A. (2021). HCK maintains the self-renewal of leukaemia stem cells via CDK6 in AML. *Journal of Experimental & Clinical Cancer Research : CR*, *40*(1). https://doi.org/10.1186/S13046-021-02007-4

Liao, X., Zhu, W., Zhou, J., Li, H., Xu, X., Zhang, B., & Gao, X. (2023). Repetitive DNA sequence detection and its role in the human genome. *Communications Biology 2023 6:1*, *6*(1), 1–21. https://doi.org/10.1038/s42003-023-05322-y

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10), e108–e108. https://doi.org/10.1093/NAR/GKT214

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, *1*(6), 417. https://doi.org/10.1016/J.CELS.2015.12.004

Liersch, R., Müller-Tidow, C., Berdel, W. E., & Krug, U. (2014). Prognostic factors for acute myeloid leukaemia in adults--biological significance and clinical use. *British Journal of Haematology*, *165*(1), 17–38. https://doi.org/10.1111/BJH.12750

Lin, M. F., Jungreis, I., & Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics (Oxford, England)*, *27*(13). https://doi.org/10.1093/BIOINFORMATICS/BTR209

Lindsley, R. C., Mar, B. G., Mazzola, E., Grauman, P. V., Shareef, S., Allen, S. L., Pigneux, A., Wetzler, M., Stuart, R. K., Erba, H. P., Damon, L. E., Powell, B. L., Lindeman, N., Steensma, D. P., Wadleigh, M., DeAngelo, D. J., Neuberg, D., Stone, R. M., & Ebert, B. L. (2014). Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood*, *125*(9), 1367–1376. https://doi.org/10.1182/BLOOD-2014-11-610543

Lis, R., Karrasch, C. C., Poulos, M. G., Kunar, B., Redmond, D., Duran, J. G. B., Badwe, C. R., Schachterle, W., Ginsberg, M., Xiang, J., Tabrizi, A. R., Shido, K., Rosenwaks, Z., Elemento, O., Speck, N. A., Butler, J. M., Scandura, J. M., & Rafii, S. (2017). Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature*, *545*(7655), 439–445. https://doi.org/10.1038/NATURE22326

Liu, L., Vujovic, A., Deshpande, N. P., Sathe, S., Anande, G., Chen, H. T. T., Xu, J., Minden, M. D., Yeo, G. W., Unnikrishnan, A., Hope, K. J., & Lu, Y. (2022). The splicing factor RBM17 drives leukemic stem cell maintenance by evading nonsense-mediated decay of pro-leukemic factors. *Nature Communications*, *13*(1). https://doi.org/10.1038/S41467-022-31155-0

Liu, P., Jiang, H., Che, M., Fu, R., Wang, H., Li, L., Zhang, W., Tao, J., Gao, S., & Shao, Z. (2018). Abnormal CD25 expression on hematopoietic cells in myelodysplastic syndromes. *Leukemia Research*, *67*, 12–16. https://doi.org/10.1016/J.LEUKRES.2017.11.010

Liu, X., Zhang, Q., Zhang, D. E., Zhou, C., Xing, H., Tian, Z., Rao, Q., Wang, M., & Wang, J. (2009). Overexpression of an isoform of AML1 in acute leukemia and its potential role in leukemogenesis. *Leukemia*, *23*(4), 739–745. https://doi.org/10.1038/LEU.2008.350

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, *25*(4), 402–408. https://doi.org/10.1006/METH.2001.1262

Long, N. A., Golla, U., Sharma, A., & Claxton, D. F. (2022). Acute Myeloid Leukemia Stem Cells: Origin, Characteristics, and Clinical Implications. *Stem Cell Reviews and Reports*, *18*(4), 1211–1226. https://doi.org/10.1007/S12015-021-10308-6

Loughran, S. J., Haas, S., Wilkinson, A. C., Klein, A. M., & Brand, M. (2020). Lineage commitment of hematopoietic stem cells and progenitors: insights from recent single cell and lineage tracing technologies. *Experimental Hematology*, *88*, 1–6. https://doi.org/10.1016/J.EXPHEM.2020.07.002

Lozzio, C. B., & Lozzio, B. B. (1975). Human Chronic Myelogenous Leukemia Cell-Line With Positive Philadelphia Chromosome. *Blood*, *45*(3), 321–334. https://doi.org/10.1182/BLOOD.V45.3.321.321

Lu, C., Ward, P. S., Kapoor, G. S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C. R., Khanin, R., Figueroa, M. E., Melnick, A., Wellen, K. E., Oğrourke, D. M., Berger, S. L., Chan, T. A., Levine, R. L., Mellinghoff, I. K., & Thompson, C. B. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature 2012 483:7390*, *483*(7390), 474–478. https://doi.org/10.1038/nature10860

Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology 2007 8:6*, *8*(6), 479–490. https://doi.org/10.1038/nrm2178

Luo, H., Zhu, G., Xu, J., Lai, Q., Yan, B., Guo, Y., Fung, T. K., Zeisig, B. B., Cui, Y., Zha, J., Cogle, C., Wang, F., Xu, B., Yang, F. C., Li, W., So, C. W. E., Qiu, Y., Xu, M., & Huang, S. (2019). HOTTIP lncRNA Promotes Hematopoietic Stem Cell Self-Renewal Leading to AML-like Disease in Mice. *Cancer Cell*, *36*(6), 645-659.e8. https://doi.org/10.1016/J.CCELL.2019.10.011

Luo, M., Jeong, M., Sun, D., Park, H. J., Rodriguez, B. A. T., Xia, Z., Yang, L., Zhang, X., Sheng, K., Darlington, G. J., Li, W., & Goodell, M. A. (2015). Long Non-coding RNAs Control Hematopoietic Stem Cell Function. *Cell Stem Cell*, *16*(4), 426. https://doi.org/10.1016/J.STEM.2015.02.002

Ma, X. Y., Wang, J. H., Wang, J. L., Ma, C. X., Wang, X. C., & Liu, F. S. (2015). Malat1 as an evolutionarily conserved lncRNA, plays a positive role in regulating proliferation and maintaining undifferentiated status of early-stage hematopoietic cells. *BMC Genomics*, *16*(1), 1–11. https://doi.org/10.1186/S12864-015-1881-X/FIGURES/7

Ma, X., Zhang, W., Zhao, M., Li, S., Jin, W., & Wang, K. (2020). Oncogenic role of lncRNA CRNDE in acute promyelocytic leukemia and NPM1-mutant acute myeloid leukemia. *Cell Death Discovery 2020 6:1*, *6*(1), 1–13. https://doi.org/10.1038/s41420-020-00359-y

Maamar, H., Cabili, M. N., Rinn, J., & Raj, A. (2013). linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes & Development*, *27*(11), 1260. https://doi.org/10.1101/GAD.217018.113

MacNicol, M. C., Cragle, C. E., McDaniel, F. K., Hardy, L. L., Wang, Y., Arumugam, K., Rahmatallah, Y., Glazko, G. V., Wilczynska, A., Childs, G. V., Zhou, D., & MacNicol, A. M. (2017a). Evasion of regulatory phosphorylation by an alternatively spliced isoform of Musashi2. *Scientific Reports 2017 7:1*, *7*(1), 1–17. https://doi.org/10.1038/s41598-017-11917-3

MacNicol, M. C., Cragle, C. E., McDaniel, F. K., Hardy, L. L., Wang, Y., Arumugam, K., Rahmatallah, Y., Glazko, G. V., Wilczynska, A., Childs, G. V., Zhou, D., & MacNicol, A. M. (2017b). Evasion of regulatory phosphorylation by an alternatively spliced isoform of Musashi2. *Scientific Reports 2017 7:1*, *7*(1), 1–17. https://doi.org/10.1038/s41598-017-11917-3

Marchese, F. P., Raimondi, I., & Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biology*, *18*(1). https://doi.org/10.1186/S13059-017-1348-2

Martinez, M., Hinojosa, M., Trombly, D., Morin, V., Stein, J., Stein, G., Javed, A., & Gutierrez, S. E. (2016). Transcriptional Auto-Regulation of RUNX1 P1 Promoter. *PLOS ONE*, *11*(2), e0149119. https://doi.org/10.1371/JOURNAL.PONE.0149119

Marzac, C., Garrido, E., Tang, R., Fava, F., Hirsch, P., de Benedictis, C., Corre, E., Lapusan, S., Lallemand, J. Y., Marie, J. P., Jacquet, E., & Legrand, O. (2011). ATP Binding Cassette transporters associated with chemoresistance: transcriptional profiling in extreme cohorts and their prognostic impact in a cohort of 281 acute myeloid leukemia patients. *Haematologica*, *96*(9), 1293. https://doi.org/10.3324/HAEMATOL.2010.031823

Matou-Nasri, S., Najdi, M., AlSaud, N. A., Alhaidan, Y., Al-Eidi, H., Alatar, G., AlWadaani, D., Trivilegio, T., AlSubait, A., AlTuwaijri, A., Abudawood, M., & Almuzzaini, B. (2022). Blockade of p38 MAPK overcomes AML stem cell line KG1a resistance to 5-Fluorouridine and the impact on miRNA profiling. *Plos One*, *17*(5), e0267855–e0267855. https://doi.org/10.1371/JOURNAL.PONE.0267855

Matsuo, Y., MacLeod, R. A. F., Uphoff, C. C., Drexler, H. G., Nishizaki, C., Katayama, Y., Kimura, G., Fujii, N., Omoto, E., Harada, M., & Orita, K. (1997). Two acute monocytic leukemia (AML-M5a) cell lines (MOLM-13 and MOLM-14) with interclonal phenotypic heterogeneity showing MLL-AF9 fusion resulting from an occult chromosome insertion, ins(11;9)(q23;p22p23). *Leukemia*, *11*(9), 1469–1477. https://doi.org/10.1038/SJ.LEU.2400768

Matsuzaki, Y., Kinjo, K., Mulligan, R. C., & Okano, H. (2004). Unexpectedly efficient homing capacity of purified murine hematopoietic stem cells. *Immunity*, *20*(1), 87–93. https://doi.org/10.1016/S1074-7613(03)00354-6

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K., & Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–236. https://doi.org/10.1038/NATURE14443

McKerrell, T., Park, N., Moreno, T., Grove, C. S., Ponstingl, H., Stephens, J., Crawley, C., Craig, J., Scott, M. A., Hodkinson, C., Baxter, J., Rad, R., Forsyth, D. R., Quail, M. A., Zeggini, E., Ouwehand, W., Varela, I., Vassiliou, G. S., & Understanding Society Scientific Group. (2015). Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Reports*, *10*(8), 1239–1245. https://doi.org/10.1016/J.CELREP.2015.02.005

Meffre, E., Milili, M., Blanco-Betancourt, C., Antunes, H., Nussenzweig, M. C., & Schiff, C. (2001). Immunoglobulin heavy chain expression shapes the B cell receptor repertoire in human B cell development. *Journal of Clinical Investigation*, *108*(6), 879. https://doi.org/10.1172/JCI13051

Mendez, L. M., & Patnaik, M. M. (2023). Clonal Hematopoiesis: Origins and determinants of evolution. *Leukemia Research*, *129*. https://doi.org/10.1016/J.LEUKRES.2023.107076

Menegatti, S., Potts, B., Garcia-Alegria, E., Paredes, R., Lie-A-Ling, M., Lacaud, G., & Kouskoff, V. (2021). The RUNX1b Isoform Defines Hemogenic Competency in Developing Human Endothelial Cells. *Frontiers in Cell and Developmental Biology*, *9*, 812639. https://doi.org/10.3389/FCELL.2021.812639/BIBTEX

Mer, A. S., Lindberg, J., Nilsson, C., Klevebring, D., Wang, M., Grönberg, H., Lehmann, S., & Rantalainen, M. (2018). Expression levels of long non-coding RNAs are prognostic for AML outcome. *Journal of Hematology and Oncology*, *11*(1), 1–13. https://doi.org/10.1186/S13045-018-0596-2/FIGURES/7

Merrick, W. C. (1992). Mechanism and regulation of eukaryotic protein synthesis. *Microbiological Reviews*, *56*(2), 291–315. https://doi.org/10.1128/MR.56.2.291-315.1992

Metzeler, K. H., Hummel, M., Bloomfield, C. D., Spiekermann, K., Braess, J., Sauerland, M. C., Heinecke, A., Radmacher, M., Marcucci, G., Whitman, S. P., Maharry, K., Paschka, P., Larson, R. A., Berdel, W. E., Büchner, T., Wörmann, B., Mansmann, U., Hiddemann, W., Bohlander, S. K., & Buske, C. (2008). An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood*, *112*(10), 4193–4201. https://doi.org/10.1182/BLOOD-2008-02-134411

Metzeler, K. H., Maharry, K., Kohlschmidt, J., Volinia, S., Mrózek, K., Becker, H., Nicolet, D., Whitman, S. P., Mendler, J. H., Schwind, S., Eisfeld, A. K., Wu, Y. Z., Powell, B. L., Carter, T. H., Wetzler, M., Kolitz, J. E., Baer, M. R., Carroll, A. J., Stone, R. M., … Bloomfield, C. D. (2013). A stem cell-like gene expression signature associates with inferior outcomes and a distinct microRNA expression profile in adults with primary cytogenetically normal acute myeloid leukemia. *Leukemia*, *27*(10), 2023–2031. https://doi.org/10.1038/LEU.2013.181

Meyers, J., Yu, Y., Kaye, J. A., & Davis, K. L. (2013). Medicare fee-for-service enrollees with primary acute myeloid leukemia: an analysis of treatment patterns, survival, and healthcare resource utilization and costs. *Applied Health Economics and Health Policy*, *11*(3), 275–286. https://doi.org/10.1007/S40258-013-0032-2

Miao, H., Wu, F., Li, Y., Qin, C., Zhao, Y., Xie, M., Dai, H., Yao, H., Cai, H., Wang, Q., Song, X., & Li, L. (2022). MALAT1 modulates alternative splicing by cooperating with the splicing factors PTBP1 and PSF. *Science Advances*, *8*(51). https://doi.org/10.1126/SCIADV.ABQ7289/SUPPL_FILE/SCIADV.ABQ7289_TABLES_S1_AND_S2.ZIP

Michaelis, L. C. (2018). Cytotoxic therapy in acute myeloid leukemia: not quite dead yet. *Hematology. American Society of Hematology. Education Program*, *2018*(1), 51–62. https://doi.org/10.1182/ASHEDUCATION-2018.1.51

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O., & Rebar, E. J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology*, *25*(7), 778–785. https://doi.org/10.1038/NBT1319

Minuesa, G., Albanese, S. K., Xie, W., Kazansky, Y., Worroll, D., Chow, A., Schurer, A., Park, S. M., Rotsides, C. Z., Taggart, J., Rizzi, A., Naden, L. N., Chou, T., Gourkanti, S., Cappel, D., Passarelli, M. C., Fairchild, L., Adura, C., Glickman, J. F., … Kharas, M. G. (2019). Small-molecule targeting of MUSASHI RNA-binding activity in acute myeloid leukemia. *Nature Communications*, *10*(1). https://doi.org/10.1038/S41467-019-10523-3

Minuesa, G., Antczak, C., Shum, D., Radu, C., Bhinder, B., Li, Y., Djaballah, H., & Kharas, M. (2014). A 1536-well fluorescence polarization assay to screen for modulators of the MUSASHI family of RNA-binding proteins. *Combinatorial Chemistry & High Throughput Screening*, *17*(7), 596–609. https://doi.org/10.2174/1386207317666140609122714

Misaghian, N., Ligresti, G., Steelman, L. S., Bertrand, F. E., Bäsecke, J., Libra, M., Nicoletti, F., Stivala, F., Milella, M., Tafuri, A., Cervello, M., Martelli, A. M., & McCubrey, J. A. (2009). Targeting the leukemic stem cell: the Holy Grail of leukemia therapy. *Leukemia : Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, *23*(1), 25. https://doi.org/10.1038/LEU.2008.246

Miyoshi, H., Ohira, M., Shimizu, K., Mitani, K., Hirai, H., Imai, T., Yokoyama, K., Soceda, E., & Ohkl, M. (1995). Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. *Nucleic Acids Research*, *23*(14), 2762. https://doi.org/10.1093/NAR/23.14.2762

Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., & Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(23), 10431. https://doi.org/10.1073/PNAS.88.23.10431

Mohrin, M., Bourke, E., Alexander, D., Warr, M. R., Barry-Holson, K., Le Beau, M. M., Morrison, C. G., & Passegué, E. (2010). Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell*, *7*(2), 174–185. https://doi.org/10.1016/J.STEM.2010.06.014

Monia, B. P., Lesnik, E. A., Gonzalez, C., Lima, W. F., McGee, D., Guinosso, C. J., Kawasaki, A. M., Dan Cook, P., & Freier, S. M. (1993). Evaluation of 2′-modified oligonucleotides containing 2′-deoxy gaps as antisense inhibitors of gene expression. *Journal of Biological Chemistry*, *268*(19), 14514–14522. https://doi.org/10.1016/S0021-9258(19)85268-7

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M. E., Vasanthakumar, A., Patel, J., Zhao, X., Perna, F., Pandey, S., Madzo, J., Song, C., Dai, Q., He, C., Ibrahim, S., Beran, M., Zavadil, J., … Levine, R. L. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*, *20*(1), 11. https://doi.org/10.1016/J.CCR.2011.06.001

Moreno-Morcillo, M., Francisco-Velilla, R., Embarc-Buh, A., Fernández-Chamorro, J., Ramón-Maiques, S., & Martinez-Salas, E. (2020). Structural basis for the dimerization of Gemin5 and its role in protein recruitment and translation control. *Nucleic Acids Research*, *48*(2), 788. https://doi.org/10.1093/NAR/GKZ1126

Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews. Genetics*, *15*(6), 423–437. https://doi.org/10.1038/NRG3722

Morrison, S. J., & Scadden, D. T. (2014a). The bone marrow niche for haematopoietic stem cells. *Nature*, *505*(7483), 327–334. https://doi.org/10.1038/NATURE12984

Morrison, S. J., & Scadden, D. T. (2014b). The bone marrow niche for haematopoietic stem cells. *Nature*, *505*(7483), 327–334. https://doi.org/10.1038/NATURE12984

Morrison, S. J., & Weissman, I. L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity*, *1*(8), 661–673. https://doi.org/10.1016/1074-7613(94)90037-X

Moshaver, B., van Rhenen, A., Kelder, A., van der Pol, M., Terwijn, M., Bachas, C., Westra, A. H., Ossenkoppele, G. J., Zweegman, S., & Schuurhuis, G. J. (2008). Identification of a small subpopulation of candidate leukemia-initiating cells in the side population of patients with acute myeloid leukemia. *Stem Cells (Dayton, Ohio)*, *26*(12), 3059–3067. https://doi.org/10.1634/STEMCELLS.2007-0861

Mrózek, K., Tanner, S. M., Heinonen, K., & Bloomfield, C. D. (2003). Molecular cytogenetic characterization of the KG-1 and KG-1a acute myeloid leukemia cell lines by use of spectral karyotyping and fluorescence in situ hybridization. *Genes, Chromosomes & Cancer*, *38*(3), 249–252. https://doi.org/10.1002/GCC.10274

Müller-Mcnicoll, M., & Neugebauer, K. M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nature Reviews. Genetics*, *14*(4), 275–287. https://doi.org/10.1038/NRG3434

Nagase, T., Seki, N., Tanaka, A., Ishikawa, K. I., & Nomura, N. (1995). Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *2*(4), 167–174. https://doi.org/10.1093/DNARES/2.4.167

National Cancer Institute. (2019). *SEER Cancer Stat Facts*. National Institutes of Health. https://seer.cancer.gov/statfacts/

Navarro-Montero, O., Ayllon, V., Lamolda, M., López-Onieva, L., Montes, R., Bueno, C., Ng, E., Guerrero-Carreno, X., Romero, T., Romero-Moya, D., Stanley, E., Elefanty, A., Ramos-Mejia, V., Menendez, P., & Real, P. J. (2017). RUNX1c Regulates Hematopoietic Differentiation of Human Pluripotent Stem Cells Possibly in Cooperation with Proinflammatory Signaling. *Stem Cells (Dayton, Ohio)*, *35*(11), 2253–2266. https://doi.org/10.1002/STEM.2700

Neelamraju, Y., Gonzalez-Perez, A., Bhat-Nakshatri, P., Nakshatri, H., & Janga, S. C. (2018). Mutational landscape of RNA-binding proteins in human cancers. *RNA Biology*, *15*(1), 115. https://doi.org/10.1080/15476286.2017.1391436

Nemeth, K., Bayraktar, R., Ferracin, M., & Calin, G. A. (2024). Non-coding RNAs in disease: from mechanisms to therapeutics. *Nature Reviews. Genetics*, *25*(3), 211–232. https://doi.org/10.1038/S41576-023-00662-1

Ng, S. W. K., Mitchell, A., Kennedy, J. A., Chen, W. C., McLeod, J., Ibrahimova, N., Arruda, A., Popescu, A., Gupta, V., Schimmer, A. D., Schuh, A. C., Yee, K. W., Bullinger, L., Herold, T., Görlich, D., Büchner, T., Hiddemann, W., Berdel, W. E., Wörmann, B., … Wang, J. C. Y. (2016). A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*, *540*(7633), 433–437. https://doi.org/10.1038/NATURE20598

Ng, S. W. K., Murphy, T., King, I., Zhang, T., Mah, M., Lu, Z., Stickle, N., Ibrahimova, N., Arruda, A., Mitchell, A., Mai, M., He, R., Madala, B. S., Viswanatha, D. S., Dick, J. E., Chan, S., Virtanen, C., Minden, M. D., Mercer, T., … Wang, J. C. Y. (2022). A clinical laboratory-developed LSC17 stemness score assay for rapid risk assessment of patients with acute myeloid leukemia. *Blood Advances*, *6*(3), 1064–1073. https://doi.org/10.1182/BLOODADVANCES.2021005741

Nguyen, D. T. T., Lu, Y., Chu, K. L., Yang, X., Park, S. M., Choo, Z. N., Chin, C. R., Prieto, C., Schurer, A., Barin, E., Savino, A. M., Gourkanti, S., Patel, P., Vu, L. P., Leslie, C. S., & Kharas, M. G. (2020). HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. *Nature Communications 2020 11:1*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-15814-8

Niebuhr, B., Kriebitzsch, N., Fischer, M., Behrens, K., Günther, T., Alawi, M., Bergholz, U., Müller, U., Roscher, S., Ziegler, M., Buchholz, F., Grundhoff, A., & Stocking, C. (2013). Runx1 is essential at two stages of early murine B-cell development. *Blood*, *122*(3), 413–423. https://doi.org/10.1182/BLOOD-2013-01-480244

Nieke, S., Yasmin, N., Kakugawa, K., Yokomizo, T., Muroi, S., & Taniuchi, I. (2017). Unique N-terminal sequences in two Runx1 isoforms are dispensable for Runx1 function. *BMC Developmental Biology*, *17*(1), 1–12. https://doi.org/10.1186/S12861-017-0156-Y/FIGURES/5

Nielsen, J., Kristensen, M. A., Willemoës, M., Nielsen, F. C., & Christiansen, J. (2004). Sequential dimerization of human zipcode-binding protein IMP1 on RNA: a cooperative mechanism providing RNP stability. *Nucleic Acids Research*, *32*(14), 4368–4376. https://doi.org/10.1093/NAR/GKH754

Ninomiya, K., & Hirose, T. (2020). Short Tandem Repeat-Enriched Architectural RNAs in Nuclear Bodies: Functions and Associated Diseases. *Non-Coding RNA*, *6*(1). https://doi.org/10.3390/NCRNA6010006

Nishimoto, Y., & Okano, H. (2010). New insight into cancer therapeutics: Induction of differentiation by regulating the Musashi/Numb/Notch pathway. *Cell Research 2010 20:10*, *20*(10), 1083–1085. https://doi.org/10.1038/cr.2010.122

Niu, J., Peng, D., & Liu, L. (2022). Drug Resistance Mechanisms of Acute Myeloid Leukemia Stem Cells. *Frontiers in Oncology*, *12*. https://doi.org/10.3389/FONC.2022.896426

Oehler, V. G., Walter, R. B., Cummings, C., Sala-Torra, O., Stirewalt, D. L., Fang, M., Radich, J. P., & Wood, B. (2010). CD52 Expression In Leukemic Stem/Progenitor Cells. *Blood*, *116*(21), 2743–2743. https://doi.org/10.1182/BLOOD.V116.21.2743.2743

Okuda, T., Van Deursen, J., Hiebert, S. W., Grosveld, G., & Downing, J. R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, *84*(2), 321–330. https://doi.org/10.1016/S0092-8674(00)80986-1

Olson, O. C., Kang, Y. A., & Passegué, E. (2020). Normal Hematopoiesis Is a Balancing Act of Self-Renewal and Regeneration. *Cold Spring Harbor Perspectives in Medicine*, *10*(12), a035519. https://doi.org/10.1101/CSHPERSPECT.A035519

Onoguchi, M., Zeng, C., Matsumaru, A., & Hamada, M. (2021). Binding patterns of RNA-binding proteins to repeat-derived RNA sequences reveal putative functional RNA elements. *NAR Genomics and Bioinformatics*, *3*(3). https://doi.org/10.1093/NARGAB/LQAB055

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. *Nature Reviews. Genetics*, *1*(1), 57–64. https://doi.org/10.1038/35049577

Orkin, S. H., & Zon, L. I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. In *Cell* (Vol. 132, Issue 4, pp. 631–644). https://doi.org/10.1016/j.cell.2008.01.025

Ossenkoppele, G., & Schuurhuis, G. J. (2016). MRD in AML: does it already guide therapy decision-making? *Hematology. American Society of Hematology. Education Program*, *2016*(1), 356–365. https://doi.org/10.1182/ASHEDUCATION-2016.1.356

Othus, M., Appelbaum, F. R., Petersdorf, S. H., Kopecky, K. J., Slovak, M., Nevill, T., Brandwein, J., Larson, R. A., Stiff, P. J., Walter, R. B., Tallman, M. S., Stenke, L., & Erba, H. P. (2015). Fate of patients with newly diagnosed acute myeloid leukemia who fail primary induction therapy. *Biology of Blood and Marrow Transplantation : Journal of the American Society for Blood and Marrow Transplantation*, *21*(3), 559–564. https://doi.org/10.1016/J.BBMT.2014.10.025

Othus, M., Kantarjian, H., Petersdorf, S., Ravandi, F., Godwin, J., Cortes, J., Pierce, S., Erba, H., Faderl, S., Appelbaum, F. R., & Estey, E. (2014). Declining rates of treatment-related mortality in patients with newly diagnosed AML given 'intense' induction regimens: a report from SWOG and MD Anderson. *Leukemia*, *28*(2), 289–292. https://doi.org/10.1038/LEU.2013.176

Pabst, C., Bergeron, A., Lavallée, V. P., Yeh, J., Gendron, P., Norddahl, G. L., Krosl, J., Boivin, I., Deneault, E., Simard, J., Imren, S., Boucher, G., Eppert, K., Herold, T., Bohlander, S. K., Humphries, K., Lemieux, S., Hébert, J., Sauvageau, G., & Barabé, F. (2016). GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. *Blood*, *127*(16), 2018–2027. https://doi.org/10.1182/BLOOD-2015-11-683649

Palacios, F., Yan, X. J., Ferrer, G., Chen, S. S., Vergani, S., Yang, X., Gardner, J., Barrientos, J. C., Rock, P., Burack, R., Kolitz, J. E., Allen, S. L., Kharas, M. G., Abdel-Wahab, O., Rai, K. R., & Chiorazzi, N. (2021). Musashi 2 influences chronic lymphocytic leukemia cell survival and growth making it a potential therapeutic target. *Leukemia*, *35*(4), 1037. https://doi.org/10.1038/S41375-020-01115-Y

Palanichamy, J. K., Tran, T. M., Howard, J. M., Contreras, J. R., Fernando, T. R., Sterne-Weiler, T., Katzman, S., Toloue, M., Yan, W., Basso, G., Pigazzi, M., Sanford, J. R., & Rao, D. S. (2016). RNA-binding protein IGF2BP3 targeting of oncogenic transcripts promotes hematopoietic progenitor proliferation. *The Journal of Clinical Investigation*, *126*(4), 1495–1511. https://doi.org/10.1172/JCI80046

Panda, A., Martindale, J., & Gorospe, M. (2017). Polysome Fractionation to Analyze mRNA Distribution Profiles. *Bio-Protocol*, *7*(3). https://doi.org/10.21769/BIOPROTOC.2126

Pandolfi, A., Barreyro, L., & Steidl, U. (2013). Concise review: preleukemic stem cells: molecular biology and clinical implications of the precursors to leukemia stem cells. *Stem Cells Translational Medicine*, *2*(2), 143–150. https://doi.org/10.5966/SCTM.2012-0109

Pang, K. C., Frith, M. C., & Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics : TIG*, *22*(1), 1–5. https://doi.org/10.1016/J.TIG.2005.10.003

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., Potter, N. E., Heuser, M., Thol, F., Bolli, N., Gundem, G., Van Loo, P., Martincorena, I., Ganly, P., Mudie, L., McLaren, S., O'Meara, S., Raine, K., Jones, D. R., … Campbell, P. J. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. *The New England Journal of Medicine*, *374*(23), 2209–2221. https://doi.org/10.1056/NEJMOA1516192

Papaioannou, D., Nicolet, D., Volinia, S., Mrózek, K., Yan, P., Bundschuh, R., Carroll, A. J., Kohlschmidt, J., Blum, W., Powell, B. L., Uy, G. L., Kolitz, J. E., Wang, E. S., Eisfeld, A. K., Orwick, S. J., Lucas, D. M., Caligiuri, M. A., Stone, R. M., Byrd, J. C., … Bloomfield, C. D. (2017). Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia. *Haematologica*, *102*(8), 1391–1400. https://doi.org/10.3324/HAEMATOL.2017.166215

Paris, J., Morgan, M., Campos, J., Spencer, G. J., Shmakova, A., Ivanova, I., Mapperley, C., Lawson, H., Wotherspoon, D. A., Sepulveda, C., Vukovic, M., Allen, L., Sarapuu, A., Tavosanis, A., Guitart, A. V., Villacreces, A., Much, C., Choe, J., Azar, A., … Kranc, K. R. (2019). Targeting the RNA m6A Reader YTHDF2 Selectively Compromises Cancer Stem Cells in Acute Myeloid Leukemia. *Cell Stem Cell*, *25*(1), 137. https://doi.org/10.1016/J.STEM.2019.03.021

Park, S. M., Deering, R. P., Lu, Y., Tivnan, P., Lianoglou, S., Al-Shahrour, F., Ebert, B. L., Hacohen, N., Leslie, C., Daley, G. Q., Lengner, C. J., & Kharas, M. G. (2014). Musashi-2 controls cell fate, lineage bias, and TGF-β signaling in HSCs. *The Journal of Experimental Medicine*, *211*(1), 71–87. https://doi.org/10.1084/JEM.20130736

Park, S. M., Gönen, M., Vu, L., Minuesa, G., Tivnan, P., Barlowe, T. S., Taggart, J., Lu, Y., Deering, R. P., Hacohen, N., Figueroa, M. E., Paietta, E., Fernandez, H. F., Tallman, M. S., Melnick, A., Levine, R., Leslie, C., Lengner, C. J., & Kharas, M. G. (2015). Musashi2 sustains the mixed-lineage leukemia-driven stem cell regulatory program. *The Journal of Clinical Investigation*, *125*(3), 1286–1298. https://doi.org/10.1172/JCI78440

Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C., & Weissman, I. L. (2005). Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *The Journal of Experimental Medicine*, *202*(11), 1599–1611. https://doi.org/10.1084/JEM.20050967

Patel, J. P., Gönen, M., Figueroa, M. E., Fernandez, H., Sun, Z., Racevskis, J., Van Vlierberghe, P., Dolgalev, I., Thomas, S., Aminova, O., Huberman, K., Cheng, J., Viale, A., Socci, N. D., Heguy, A., Cherry, A., Vance, G., Higgins, R. R., Ketterling, R. P., … Levine, R. L. (2012). Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *The New England Journal of Medicine*, *366*(12), 1079–1089. https://doi.org/10.1056/NEJMOA1112304

Pearce, D. J., Taussig, D., Zibara, K., Smith, L. L., Ridler, C. M., Preudhomme, C., Young, B. D., Rohatiner, A. Z., Lister, T. A., & Bonnet, D. (2006). AML engraftment in the NOD/SCID assay reflects the outcome of AML: implications for our understanding of the heterogeneity of AML. *Blood*, *107*(3), 1166–1173. https://doi.org/10.1182/BLOOD-2005-06-2325

Pepe, F., Bill, M., Papaioannou, D., Karunasiri, M., Walker, A., Naumann, E., Snyder, K., Ranganathan, P., Dorrance, A., & Garzon, R. (2022). Targeting Wnt signaling in acute myeloid leukemia stem cells. *Haematologica*, *107*(1), 307. https://doi.org/10.3324/HAEMATOL.2020.266155

Percival, M. E., Lai, C., Estey, E., & Hourigan, C. S. (2017). Bone marrow evaluation for diagnosis and monitoring of acute myeloid leukemia. *Blood Reviews*, *31*(4), 185–192. https://doi.org/10.1016/J.BLRE.2017.01.003

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W., Guilak, F., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods*, *10*(10), 973–976. https://doi.org/10.1038/NMETH.2600

Pfirrmann, M., Ehninger, G., Thiede, C., Bornhäuser, M., Kramer, M., Röllig, C., Hasford, J., & Schaich, M. (2012). Prediction of post-remission survival in acute myeloid leukaemia: a post-hoc analysis of the AML96 trial. *The Lancet. Oncology*, *13*(2), 207–214. https://doi.org/10.1016/S1470-2045(11)70326-6

Pich, O., Reyes-Salazar, I., Gonzalez-Perez, A., & Lopez-Bigas, N. (2022). Discovering the drivers of clonal hematopoiesis. *Nature Communications 2022 13:1*, *13*(1), 1–12. https://doi.org/10.1038/s41467-022-31878-0

Pimkin, M., Kossenkov, A. V., Mishra, T., Morrissey, C. S., Wu, W., Keller, C. A., Blobel, G. A., Lee, D., Beer, M. A., Hardison, R. C., & Weiss, M. J. (2014). Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Research*, *24*(12), 1932–1944. https://doi.org/10.1101/GR.164178.113

Pousse, L., Korfi, K., Medeiros, B. C., Berrera, M., Kumpesa, N., Eckmann, J., Hutter, I. K., Griesser, V., Karanikas, V., Klein, C., & Amann, M. (2023). CD25 targeting with the afucosylated human IgG1 antibody RG6292 eliminates regulatory T cells and CD25+ blasts in acute myeloid leukemia. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/FONC.2023.1150149

Pozner, A., Goldenberg, D., Negreanu, V., Le, S.-Y., Elroy-Stein, O., Levanon, D., & Groner, Y. (2000). Transcription-coupled translation control of AML1/RUNX1 is mediated by cap- and internal ribosome entry site-dependent mechanisms. *Molecular and Cellular Biology*, *20*(7), 2297–2307. https://doi.org/10.1128/MCB.20.7.2297-2307.2000

Qin, H., Yang, L., Chukinas, J. A., Shah, N., Tarun, S., Pouzolles, M., Chien, C. D., Niswander, L. M., Welch, A. R., Taylor, N., Tasian, S. K., & Fry, T. J. (2021). Systematic preclinical evaluation of CD33-directed chimeric antigen receptor T cell immunotherapy for acute myeloid leukemia defines optimized construct design. *Journal for Immunotherapy of Cancer*, *9*(9). https://doi.org/10.1136/JITC-2021-003149

Qiu, Y., Xu, M., & Huang, S. (2021). Long noncoding RNAs: emerging regulators of normal and malignant hematopoiesis. *Blood*, *138*(23), 2327–2336. https://doi.org/10.1182/BLOOD.2021011992

Quek, L., Otto, G. W., Garnett, C., Lhermitte, L., Karamitros, D., Stoilova, B., Lau, I. J., Doondeea, J., Doondeea, B., Kennedy, A., Metzner, M., Goardon, N., Ivey, A., Allen, C., Gale, R., Davies, B., Sternberg, A., Killick, S., Hunter, H., … Vyas, P.

(2016). Genetically distinct leukemic stem cells in human CD34- acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. *The Journal of Experimental Medicine*, *213*(8), 1513–1535. https://doi.org/10.1084/JEM.20151775

Quivoron, C., Couronné, L., Della Valle, V., Lopez, C. K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M. H., Godley, L., Opolon, P., Tilly, H., Solary, E., Duffourd, Y., Dessen, P., Merle-Beral, H., Nguyen-Khac, F., Fontenay, M., … Bernard, O. A. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell*, *20*(1), 25–38. https://doi.org/10.1016/J.CCR.2011.06.003

Ran, D., Schubert, M., Pietsch, L., Taubert, I., Wuchter, P., Eckstein, V., Bruckner, T., Zoeller, M., & Ho, A. D. (2009). Aldehyde dehydrogenase activity among primary leukemia cells is associated with stem cell features and correlates with adverse clinical outcomes. *Experimental Hematology*, *37*(12), 1423–1434. https://doi.org/10.1016/J.EXPHEM.2009.10.001

Ran, D., Shia, W. J., Lo, M. C., Fan, J. B., Knorr, D. A., Ferrell, P. I., Ye, Z., Yan, M., Cheng, L., Kaufman, D. S., & Zhang, D. E. (2013). RUNX1a enhances hematopoietic lineage commitment from human embryonic stem cells and inducible pluripotent stem cells. *Blood*, *121*(15), 2882–2890. https://doi.org/10.1182/BLOOD-2012-08-451641

Rapaport, F., Neelamraju, Y., Baslan, T., Hassane, D., Gruszczynska, A., Robert de Massy, M., Farnoud, N., Haddox, S., Lee, T., Medina-Martinez, J., Sheridan, C., Thurmond, A., Becker, M., Bekiranov, S., Carroll, M., Moses Murdock, H., Valk, P. J. M., Bullinger, L., D'Andrea, R., … Garrett-Bakelman, F. E. (2021). Genomic and evolutionary portraits of disease relapse in acute myeloid leukemia. *Leukemia 2021 35:9*, *35*(9), 2688–2692. https://doi.org/10.1038/s41375-021-01153-0

Rapin, N., Bagger, F. O., Jendholm, J., Mora-Jensen, H., Krogh, A., Kohlmann, A., Thiede, C., Borregaard, N., Bullinger, L., Winther, O., Theilgaard-Mönch, K., & Porse, B. T. (2014). Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients. *Blood*, *123*(6), 894–904. https://doi.org/10.1182/BLOOD-2013-02-485771

Raposo, A. C., Casanova, M., Gendrel, A. V., & da Rocha, S. T. (2021). The tandem repeat modules of Xist lncRNA: a swiss army knife for the control of X-chromosome inactivation. *Biochemical Society Transactions*, *49*(6), 2549–2560. https://doi.org/10.1042/BST20210253

Reindl, L., Bacher, U., Dicker, F., Alpermann, T., Kern, W., Schnittger, S., Haferlach, T., & Haferlach, C. (2010). Biological and clinical characterization of recurrent 14q deletions in CLL and other mature B-cell neoplasms. *British Journal of Haematology*, *151*(1), 25–36. https://doi.org/10.1111/J.1365-2141.2010.08299.X

Rentas, S., Holzapfel, N. T., Belew, M. S., Pratt, G. A., Voisin, V., Wilhelm, B. T., Bader, G. D., Yeo, G. W., & Hope, K. J. (2016). Musashi-2 attenuates AHR signalling to expand human haematopoietic stem cells. *Nature*, *532*(7600), 508–511. https://doi.org/10.1038/NATURE17665

Reya, T., Morrison, S. J., Clarke, M. F., & Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, *414*(6859), 105–111. https://doi.org/10.1038/35102167

Richard-Carpentier, G., & DiNardo, C. D. (2019). Single-agent and combination biologics in acute myeloid leukemia. *Hematology. American Society of Hematology. Education Program*, *2019*(1), 548–556. https://doi.org/10.1182/HEMATOLOGY.2019000059

Riether, C., Schürch, C. M., & Ochsenbein, A. F. (2015). Regulation of hematopoietic and leukemic stem cells by the immune system. *Cell Death and Differentiation*, *22*(2), 187–198. https://doi.org/10.1038/CDD.2014.89

Rinaldi, C., & Wood, M. J. A. (2018). Antisense oligonucleotides: the next frontier for treatment of neurological disorders. *Nature Reviews. Neurology*, *14*(1), 9–22. https://doi.org/10.1038/NRNEUROL.2017.148

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/NBT.1754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/BIOINFORMATICS/BTP616

Rodrigues, A. C. B. da C., Costa, R. G. A., Silva, S. L. R., Dias, I. R. S. B., Dias, R. B., & Bezerra, D. P. (2021). Cell signaling pathways as molecular targets to eliminate AML stem cells. *Critical Reviews in Oncology/Hematology*, *160*. https://doi.org/10.1016/J.CRITREVONC.2021.103277

Rosenbauer, F., & Tenen, D. G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. *Nature Reviews. Immunology*, *7*(2), 105–117. https://doi.org/10.1038/NRI2024

Rothenberg-Thurley, M., Amler, S., Goerlich, D., Köhnke, T., Konstandin, N. P., Schneider, S., Sauerland, M. C., Herold, T., Hubmann, M., Ksienzyk, B., Zellmeier, E., Bohlander, S. K., Subklewe, M., Faldum, A., Hiddemann, W., Braess, J., Spiekermann, K., & Metzeler, K. H. (2018). Persistence of pre-leukemic clones during first remission and risk of relapse in acute myeloid leukemia. *Leukemia*, *32*(7), 1598. https://doi.org/10.1038/S41375-018-0034-Z

Rozen, E. J., Ozeroff, C. D., & Allen, M. A. (2023). RUN(X) out of blood: emerging RUNX1 functions beyond hematopoiesis and links to Down syndrome. *Human Genomics 2023 17:1*, *17*(1), 1–20. https://doi.org/10.1186/S40246-023-00531-2

Rudra, D., Egawa, T., Chong, M. M. W., Treuting, P., Littman, D. R., & Rudensky, A. Y. (2009). Runx-CBFbeta complexes control expression of the transcription factor Foxp3 in regulatory T cells. *Nature Immunology*, *10*(11), 1170–1177. https://doi.org/10.1038/NI.1795

Ruiz-Orera, J., & Albà, M. M. (2019). Conserved regions in long non-coding RNAs contain abundant translation and protein-RNA interaction signatures. *NAR Genomics and Bioinformatics*, *1*(1). https://doi.org/10.1093/NARGAB/LQZ002

Ruth Zearfoss, N., Deveau, L. M., Clingman, C. C., Schmidt, E., Johnson, E. S., Massi, F., & Ryder, S. P. (2014). A Conserved Three-nucleotide Core Motif Defines Musashi RNA Binding Specificity. *The Journal of Biological Chemistry*, *289*(51), 35530. https://doi.org/10.1074/JBC.M114.597112

Ryu, H. G., Jung, Y., Lee, N., Seo, J. Y., Kim, S. W., Lee, K. H., Kim, D. Y., & Kim, K. T. (2021). Hnrnp a1 promotes lung cancer cell proliferation by modulating vrk1 translation. *International Journal of Molecular Sciences*, *22*(11). https://doi.org/10.3390/IJMS22115506/S1

Saha, S., Murmu, K. C., Biswas, M., Chakraborty, S., Basu, J., Madhulika, S., Kolapalli, S. P., Chauhan, S., Sengupta, A., & Prasad, P. (2019). Transcriptomic Analysis Identifies RNA Binding Proteins as Putative Regulators of Myelopoiesis and Leukemia. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/FONC.2019.00692

Saito, Y., Kitamura, H., Hijikata, A., Tomizawa-Murasawa, M., Tanaka, S., Takagi, S., Uchida, N., Suzuki, N., Sone, A., Najima, Y., Ozawa, H., Wake, A., Taniguchi, S., Shultz, L. D., Ohara, O., & Ishikawa, F. (2010). Identification of Therapeutic Targets for Quiescent, Chemotherapy-Resistant Human Leukemia Stem Cells. *Science Translational Medicine*, *2*(17), 17ra9. https://doi.org/10.1126/SCITRANSLMED.3000349

Saito, Y., Uchida, N., Tanaka, S., Suzuki, N., Tomizawa-Murasawa, M., Sone, A., Najima, Y., Takagi, S., Oki, Y., Wake, A., Taniguchi, S., Shultz, L. D., & Ishikawa, F. (2010). Induction of cell cycle entry eliminates human leukemia stem cells in a mouse model of AML. *Nature Biotechnology*, *28*(3), 275–280. https://doi.org/10.1038/NBT.1607

Sakakibara, S. I., Nakamura, Y., Satoh, H., & Okano, H. (2001). Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *21*(20), 8091–8107. https://doi.org/10.1523/JNEUROSCI.21-20-08091.2001

Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, *146*(3), 353–358. https://doi.org/10.1016/J.CELL.2011.07.014

Sarma, N. J., Takeda, A., & Yaseen, N. R. (2010). Colony forming cell (CFC) assay for human hematopoietic cells. *Journal of Visualized Experiments : JoVE*, *46*. https://doi.org/10.3791/2195

Sarry, J. E., Murphy, K., Perry, R., Sanchez, P. V., Secreto, A., Keefer, C., Swider, C. R., Strzelecki, A. C., Cavelier, C., Récher, C., Mansat-De Mas, V., Delabesse, E., Danet-Desnoyers, G., & Carroll, M. (2010). Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *The Journal of Clinical Investigation*, *121*(1), 384–395. https://doi.org/10.1172/JCI41495

Sasaki, K., Ravandi, F., Kadia, T. M., DiNardo, C. D., Short, N. J., Borthakur, G., Jabbour, E., & Kantarjian, H. M. (2021). De novo acute myeloid leukemia: A population-based study of outcome in the United States based on the Surveillance, Epidemiology, and End Results (SEER) database, 1980 to 2017. *Cancer*, *127*(12), 2049–2061. https://doi.org/10.1002/CNCR.33458

Sato, H., Wheat, J. C., Steidl, U., & Ito, K. (2016). DNMT3A and TET2 in the Pre-Leukemic Phase of Hematopoietic Disorders. *Frontiers in Oncology*, *6*(AUG), 187. https://doi.org/10.3389/FONC.2016.00187

Sawai, C. M., Babovic, S., Upadhaya, S., Knapp, D. J. H. F., Lavin, Y., Lau, C. M., Goloborodko, A., Feng, J., Fujisaki, J., Ding, L., Mirny, L. A., Merad, M., Eaves, C. J., & Reizis, B. (2016). Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals. *Immunity*, *45*(3), 597–609. https://doi.org/10.1016/J.IMMUNI.2016.08.007

Schepers, H., Van Gosliga, D., Wierenga, A. T. J., Eggen, B. J. L., Schuringa, J. J., & Vellenga, E. (2007). STAT5 is required for long-term maintenance of normal and leukemic human stem/progenitor cells. *Blood*, *110*(8), 2880–2888. https://doi.org/10.1182/BLOOD-2006-08-039073

Schito, L., Rey, S., & Konopleva, M. (2017). Integration of hypoxic HIF-α signaling in blood cancers. *Oncogene*, *36*(38), 5331–5340. https://doi.org/10.1038/ONC.2017.119

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., & Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular Cell*, *65*(1), 25–38. https://doi.org/10.1016/J.MOLCEL.2016.11.029

Schnerch, D., Yalcintepe, J., Schmidts, A., Becker, H., Follo, M., Engelhardt, M., & Wäsch, R. (2012). Cell cycle control in acute myeloid leukemia. *American Journal of Cancer Research*, *2*(5), 508. /pmc/articles/PMC3433102/

Schnittger, S., Dicker, F., Kern, W., Wendland, N., Sundermann, J., Alpermann, T., Haferlach, C., & Haferlach, T. (2011). RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. *Blood*, *117*(8), 2348–2357. https://doi.org/10.1182/BLOOD-2009-11-255976

Schuurhuis, G. J., Heuser, M., Freeman, S., Béne, M. C., Buccisano, F., Cloos, J., Grimwade, D., Haferlach, T., Hills, R. K., Hourigan, C. S., Jorgensen, J. L., Kern, W., Lacombe, F., Maurillo, L., Preudhomme, C., van der Reijden, B. A., Thiede, C., Venditti, A., Vyas, P., … Ossenkoppele, G. J. (2018). Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood*, *131*(12), 1275–1291. https://doi.org/10.1182/BLOOD-2017-09-801498

Schwarzer, A., Emmrich, S., Schmidt, F., Beck, D., Ng, M., Reimer, C., Adams, F. F., Grasedieck, S., Witte, D., Käbler, S., Wong, J. W. H., Shah, A., Huang, Y., Jammal, R., Maroz, A., Jongen-Lavrencic, M., Schambach, A., Kuchenbauer, F., Pimanda,

J. E., … Klusmann, J. H. (2017). The non-coding RNA landscape of human hematopoiesis and leukemia. *Nature Communications*, *8*(1). https://doi.org/10.1038/S41467-017-00212-4

Semenza, G. L. (2003). Targeting HIF-1 for cancer therapy. *Nature Reviews. Cancer*, *3*(10), 721–732. https://doi.org/10.1038/NRC1187

Sentis, G., Loukogiannaki, C., Malissovas, N., Nikolopoulos, D., Manolakou, T., Flouda, S., Grigoriou, M., Banos, A., Boumpas, D. T., & Filia, A. (2023). A network-based approach reveals long non-coding RNAs associated with disease activity in lupus nephritis: key pathways for flare and potential biomarkers to be used as liquid biopsies. *Frontiers in Immunology*, *14*. https://doi.org/10.3389/FIMMU.2023.1203848

Setrerrahmane, S., Li, M., Zoghbi, A., Lv, X., Zhang, S., Zhao, W., Lu, J., Craik, D. J., & Xu, H. (2022). Cancer-related micropeptides encoded by ncRNAs: Promising drug targets and prognostic biomarkers. *Cancer Letters*, *547*, 215723. https://doi.org/10.1016/J.CANLET.2022.215723

Shah, A., Andersson, T. M. L., Rachet, B., Björkholm, M., & Lambert, P. C. (2013). Survival and cure of acute myeloid leukaemia in England, 1971-2006: a population-based study. *British Journal of Haematology*, *162*(4), 509–516. https://doi.org/10.1111/BJH.12425

Shah, N. M., Jang, H. J., Liang, Y., Maeng, J. H., Tzeng, S. C., Wu, A., Basri, N. L., Qu, X., Fan, C., Li, A., Katz, B., Li, D., Xing, X., Evans, B. S., & Wang, T. (2023). Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nature Genetics*, *55*(4), 631–639. https://doi.org/10.1038/S41588-023-01349-3

She, M., Niu, X., Chen, X., Li, J., Zhou, M., He, Y., Le, Y., & Guo, K. (2012). Resistance of leukemic stem-like cells in AML cell line KG1a to natural killer cell-mediated cytotoxicity. *Cancer Letters*, *318*(2), 173–179. https://doi.org/10.1016/J.CANLET.2011.12.017

Shi, L. E., Shang, X., Nie, K. C., Lin, Z. Q., Wang, M., Huang, Y. Y., & Zhu, Z. Z. (2020). Identification of hub genes correlated with the pathogenesis and prognosis in Pancreatic adenocarcinoma on bioinformatics methods. *Translational Cancer Research*, *9*(8), 4550. https://doi.org/10.21037/TCR-19-2873

Shi, Y., Yang, D., & Qin, Y. (2021). Identifying prognostic lncRNAs based on a ceRNA regulatory network in laryngeal squamous cell carcinoma. *BMC Cancer*, *21*(1), 1–15. https://doi.org/10.1186/S12885-021-08422-2/FIGURES/9

Shimony, S., Stahl, M., & Stone, R. M. (2023). Acute myeloid leukemia: 2023 update on diagnosis, risk-stratification, and management. *American Journal of Hematology*, *98*(3), 502–526. https://doi.org/10.1002/AJH.26822

Shin, J. Y., Hu, W., Naramura, M., & Park, C. Y. (2014). High c-Kit expression identifies hematopoietic stem cells with impaired self-renewal and megakaryocytic bias. *The Journal of Experimental Medicine*, *211*(2), 217–231. https://doi.org/10.1084/JEM.20131128

Shivarov, V., & Bullinger, L. (2014). Expression profiling of leukemia patients: key lessons and future directions. *Experimental Hematology*, *42*(8), 651–660. https://doi.org/10.1016/J.EXPHEM.2014.04.006

Shizuru, J. A., Negrin, R. S., & Weissman, I. L. (2005). Hematopoietic stem and progenitor cells: clinical and preclinical regeneration of the hematolymphoid system. *Annual Review of Medicine*, *56*, 509–538. https://doi.org/10.1146/ANNUREV.MED.54.101601.152334

Shlush, L. I., Mitchell, A., Heisler, L., Abelson, S., Ng, S. W. K., Trotman-Grant, A., Medeiros, J. J. F., Rao-Bhatia, A., Jaciw-Zurakowsky, I., Marke, R., McLeod, J. L., Doedens, M., Bader, G., Voisin, V., Xu, C., McPherson, J. D., Hudson, T. J., Wang, J. C. Y., Minden, M. D., & Dick, J. E. (2017). Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature*, *547*(7661), 104–108. https://doi.org/10.1038/NATURE22993

Shlush, L. I., Zandi, S., Mitchell, A., Chen, W. C., Brandwein, J. M., Gupta, V., Kennedy, J. A., Schimmer, A. D., Schuh, A. C., Yee, K. W., McLeod, J. L., Doedens, M., Medeiros, J. J. F., Marke, R., Kim, H. J., Lee, K., McPherson, J. D., Hudson, T. J., Brown, A. M. K., … Dick, J. E. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature 2014 506:7488*, *506*(7488), 328–333. https://doi.org/10.1038/nature13038

Short, N. J., Rytting, M. E., & Cortes, J. E. (2018). Acute myeloid leukaemia. *Lancet (London, England)*, *392*(10147), 593–606. https://doi.org/10.1016/S0140-6736(18)31041-9

Sill, H., Olipitz, W., Zebisch, A., Schulz, E., & Wölfler, A. (2011). Therapy-related myeloid neoplasms: pathobiology and clinical characteristics. *British Journal of Pharmacology*, *162*(4), 792–805. https://doi.org/10.1111/J.1476-5381.2010.01100.X

Siveen, K. S., Uddin, S., & Mohammad, R. M. (2017). Targeting acute myeloid leukemia stem cell signaling by natural products. *Molecular Cancer*, *16*(1), 1–12. https://doi.org/10.1186/S12943-016-0571-X/TABLES/1

Skopek, R., Palusińska, M., Kaczor-Keller, K., Pingwara, R., Papierniak-Wyglądała, A., Schenk, T., Lewicki, S., Zelent, A., & Szymański, Ł. (2023). Choosing the Right Cell Line for Acute Myeloid Leukemia (AML) Research. *International Journal of Molecular Sciences*, *24*(6). https://doi.org/10.3390/ijms24065377

Slack, F. J., & Chinnaiyan, A. M. (2019). The Role of Non-coding RNAs in Oncology. *Cell*, *179*(5), 1033–1055. https://doi.org/10.1016/J.CELL.2019.10.017

Smith, L. G., Weissman, I. L., & Heimfeld, S. (1991). Clonal analysis of hematopoietic stem-cell differentiation in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(7), 2788–2792. https://doi.org/10.1073/PNAS.88.7.2788

Somervaille, T. C. P., Matheny, C. J., Spencer, G. J., Iwasaki, M., Rinn, J. L., Witten, D. M., Chang, H. Y., Shurtleff, S. A., Downing, J. R., & Cleary, M. L. (2009). Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell*, *4*(2), 129–140. https://doi.org/10.1016/J.STEM.2008.11.015

Sommerkamp, P., Renders, S., Ladel, L., Hotz-Wagenblatt, A., Schönberger, K., Zeisberger, P., Przybylla, A., Sohn, M., Zhou, Y., Klibanski, A., Cabezas-Wallscheid, N., & Trumpp, A. (2019). The long non-coding RNA Meg3 is dispensable for hematopoietic stem cells. *Scientific Reports 2019 9:1*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-38605-8

Song, L., Lin, H. S., Gong, J. N., Han, H., Wang, X. S., Su, R., Chen, M. T., Shen, C., Ma, Y. N., Yu, J., & Zhang, J. W. (2017). microRNA-451-modulated hnRNP A1 takes a part in granulocytic differentiation regulation and acute myeloid leukemia. *Oncotarget*, *8*(33), 55453. https://doi.org/10.18632/ONCOTARGET.19325

Spinler, K., Bajaj, J., Ito, T., Zimdahl, B., Hamilton, M., Ahmadi, A., Koechlein, C. S., Lytle, N., Kwon, H. Y., Anower-E-Khuda, F., Sun, H., Blevins, A., Weeks, J., Kritzik, M., Karlseder, J., Ginsberg, M. H., Park, P. W., Esko, J. D., & Reya, T. (2020). A stem cell reporter based platform to identify and target drug resistant stem cells in myeloid leukemia. *Nature Communications 2020 11:1*, *11*(1), 1–15. https://doi.org/10.1038/s41467-020-19782-x

Stasi, R., Del Poeta, G., Venditti, A., Masi, M., Stipa, E., Dentamaro, T., Cox, C., Dallapiccola, B., & Papa, G. (1994). Analysis of Treatment Failure in Patients With Minimally Differentiated Acute Myeloid Leukemia (AML-M0). *Blood*, *83*(6), 1619–1625. https://doi.org/10.1182/BLOOD.V83.6.1619.1619

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2020). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology 2020 22:2*, *22*(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Stölzel, F., Mohr, B., Kramer, M., Oelschlägel, U., Bochtler, T., Berdel, W. E., Kaufmann, M., Baldus, C. D., Schäfer-Eckart, K., Stuhlmann, R., Einsele, H., Krause, S. W., Serve, H., Hänel, M., Herbst, R., Neubauer, A., Sohlbach, K., Mayer, J., Middeke, J. M., … Ehninger, G. (2016). Karyotype complexity and prognosis in acute myeloid leukemia. *Blood Cancer Journal*, *6*(1), e386. https://doi.org/10.1038/BCJ.2015.114

Strauss, L. C., Brovall, C., Fackler, M. J., Schwartz, J. F., Shaper, J. H., Loken, M. R., & Civin, C. I. (1986). Antigenic analysis of hematopoiesis. IV. The My-11 hematopoietic cell surface antigen is expressed by myelomonocytic and lymphoid, but not erythroid, progenitor cells. *Experimental Hematology*, *14*(10), 935–945.

Sugimura, R., Jha, D. K., Han, A., Soria-Valles, C., Da Rocha, E. L., Lu, Y. F., Goettel, J. A., Serrao, E., Rowe, R. G., Malleshaiah, M., Wong, I., Sousa, P., Zhu, T. N., Ditadi, A., Keller, G., Engelman, A. N., Snapper, S. B., Doulatov, S., & Daley, G. Q. (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature 2017 545:7655*, *545*(7655), 432–438. https://doi.org/10.1038/nature22370

Sun, H., Sun, Y., Chen, Q., & Xu, Z. (2020). LncRNA KCNQ1OT1 contributes to the progression and chemoresistance in acute myeloid leukemia by modulating Tspan3 through suppressing miR-193a-3p. *Life Sciences*, *241*. https://doi.org/10.1016/J.LFS.2019.117161

Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y. J., Klein, A., Hofmann, O., & Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. *Nature*, *514*(7522), 322–327. https://doi.org/10.1038/NATURE13824

Sun, L., Wang, W., Han, C., Huang, W., Sun, Y., Fang, K., Zeng, Z., Yang, Q., Pan, Q., Chen, T., Luo, X., & Chen, Y. (2021). The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. *Molecular Cell*, *81*(21), 4493-4508.e9. https://doi.org/10.1016/J.MOLCEL.2021.08.033

Sun, T., Dong, L., Guo, Y., Zhao, H., & Wang, M. (2022). Revealing key lncRNAs in cytogenetically normal acute myeloid leukemia by reconstruction of the lncRNA–miRNA–mRNA network. *Scientific Reports*, *12*(1). https://doi.org/10.1038/S41598-022-08930-6

Sun, Y., Chen, B. R., & Deshpande, A. (2018). Epigenetic regulators in the development, maintenance, and therapeutic targeting of acute myeloid leukemia. *Frontiers in Oncology*, *8*(FEB), 320797. https://doi.org/10.3389/FONC.2018.00041/BIBTEX

Sureda-Gómez, M., Balsas, P., Rodríguez, M.-L., Nadeu, F., Bolòs, A. De, Eguileor, Á., Kulis, M., Castellano, G., Giné, E., Demajo, S., Martín-Subero, J. I., Jares, P., Campo, E., & Amador, V. (2022). P1295: DECIPHERING THE ROLE OF MSI2 AS A REGULATOR OF MANTLE CELL LYMPHOMA STEM-LIKE PROPERTIES. *HemaSphere*, *6*(Suppl), 1180–1181. https://doi.org/10.1097/01.HS9.0000848044.34774.94

Szostak, E., & Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. *Briefings in Functional Genomics*, *12*(1), 58–65. https://doi.org/10.1093/BFGP/ELS056

Tabe, Y., Shi, Y. X., Zeng, Z., Jin, L., Shikami, M., Hatanaka, Y., Miida, T., Hsu, F. J., Andreeff, M., & Konopleva, M. (2013). TGF-β-Neutralizing Antibody 1D11 Enhances Cytarabine-Induced Apoptosis in AML Cells in the Bone Marrow Microenvironment. *PloS One*, *8*(6). https://doi.org/10.1371/JOURNAL.PONE.0062785

Taggart, J., Ho, T. C., Amin, E., Xu, H., Barlowe, T. S., Perez, A. R., Durham, B. H., Tivnan, P., Okabe, R., Chow, A., Vu, L., Park, S. M., Prieto, C., Famulare, C., Patel, M., Lengner, C. J., Verma, A., Roboz, G., Guzman, M., … Kharas, M. G. (2016). MSI2 is required for maintaining activated myelodysplastic syndrome stem cells. *Nature Communications*, *7*. https://doi.org/10.1038/NCOMMS10739

Takahashi, S. (2011). Current findings for recurring mutations in acute myeloid leukemia. *Journal of Hematology & Oncology*, *4*. https://doi.org/10.1186/1756-8722-4-36

Takebe, N., Harris, P. J., Warren, R. Q., & Ivy, S. P. (2011). Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nature Reviews. Clinical Oncology*, *8*(2), 97–106. https://doi.org/10.1038/NRCLINONC.2010.196

Takei, H., & Kobayashi, S. S. (2019). Targeting transcription factors in acute myeloid leukemia. *International Journal of Hematology*, *109*(1), 28–34. https://doi.org/10.1007/S12185-018-2488-1/FIGURES/3

Tambaro, F. P., Singh, H., Jones, E., Rytting, M., Mahadeo, K. M., Thompson, P., Daver, N., DiNardo, C., Kadia, T., Garcia-Manero, G., Chan, T., Shah, R. R., & Wierda, W. G. (2021). Autologous CD33-CAR-T cells for treatment of relapsed/refractory acute myelogenous leukemia. *Leukemia*, *35*(11), 3282–3286. https://doi.org/10.1038/S41375-021-01232-2

Tamma, R., Ribatti, D., Ranieri, G., & Kumar Vashist, Y. (2017). Bone Niches, Hematopoietic Stem Cells, and Vessel Formation. *International Journal of Molecular Sciences 2017, Vol. 18, Page 151*, *18*(1), 151. https://doi.org/10.3390/IJMS18010151

Tan, A. C., Jimeno, A., Lin, S. H., Wheelhouse, J., Chan, F., Solomon, A., Rajeshkumar, N. V., Rubio-Viqueira, B., & Hidalgo, M. (2009). Characterizing DNA methylation patterns in pancreatic cancer genome. *Molecular Oncology*, *3*(5–6), 425–438. https://doi.org/10.1016/J.MOLONC.2009.03.004

Tanaka, T., Tanaka, K., Ogawa, S., Kurokawa, M., Mitani, K., Nishida, J., Shibata, Y., Yazaki, Y., & Hirai, H. (1995). An acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. *The EMBO Journal*, *14*(2), 341–350. https://doi.org/10.1002/J.1460-2075.1995.TB07008.X

Tanese, N. (1997). Small-scale density gradient sedimentation to separate and analyze multiprotein complexes. *Methods (San Diego, Calif.)*, *12*(3), 224–234. https://doi.org/10.1006/METH.1997.0475

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*, *45*(W1), W98–W102. https://doi.org/10.1093/NAR/GKX247

Taussig, D. C., Vargaftig, J., Miraki-Moud, F., Griessinger, E., Sharrock, K., Luke, T., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S. G., Lister, T. A., Gribben, J. G., & Bonnet, D. (2010). Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. *Blood*, *115*(10), 1976–1984. https://doi.org/10.1182/BLOOD-2009-02-206565

Taylor, J., Xiao, W., & Abdel-Wahab, O. (2017). Diagnosis and classification of hematologic malignancies on the basis of genetics. *Blood*, *130*(4), 410–423. https://doi.org/10.1182/BLOOD-2017-02-734541

Telfer, J. C., & Rothenberg, E. V. (2001a). Expression and function of a stem cell promoter for the murine CBFalpha2 gene: distinct roles and regulation in natural killer and T cell development. *Developmental Biology*, *229*(2), 363–382. https://doi.org/10.1006/DBIO.2000.9991

Telfer, J. C., & Rothenberg, E. V. (2001b). Expression and function of a stem cell promoter for the murine CBFalpha2 gene: distinct roles and regulation in natural killer and T cell development. *Developmental Biology*, *229*(2), 363–382. https://doi.org/10.1006/DBIO.2000.9991

Terwijn, M., Feller, N., van Rhenen, A., Kelder, A., Westra, G., Zweegman, S., Ossenkoppele, G., & Schuurhuis, G. J. (2009). Interleukin-2 receptor alpha-chain (CD25) expression on leukaemic blasts is predictive for outcome and level of residual disease in AML. *European Journal of Cancer (Oxford, England : 1990)*, *45*(9), 1692–1699. https://doi.org/10.1016/J.EJCA.2009.02.021

Terwijn, M., Zeijlemaker, W., Kelder, A., Rutten, A. P., Snel, A. N., Scholten, W. J., Pabst, T., Verhoef, G., Löwenberg, B., Zweegman, S., Ossenkoppele, G. J., & Schuurhuis, G. J. (2014). Leukemic stem cell frequency: a strong biomarker for clinical outcome in acute myeloid leukemia. *PloS One*, *9*(9). https://doi.org/10.1371/JOURNAL.PONE.0107587

Testa, U., Riccioni, R., Militi, S., Coccia, E., Stellacci, E., Samoggia, P., Latagliata, R., Mariani, G., Rossini, A., Battistini, A., Lo-Coco, F., & Peschle, C. (2002). Elevated expression of IL-3Ralpha in acute myelogenous leukemia is associated with enhanced blast proliferation, increased cellularity, and poor prognosis. *Blood*, *100*(8), 2980–2988. https://doi.org/10.1182/BLOOD-2002-03-0852

The Cancer Genome Atlas Program (TCGA) - NCI. (n.d.). Retrieved 9 April 2024, from https://www.cancer.gov/ccg/research/genome-sequencing/tcga

Tian, B., & Manley, J. L. (2016). Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology 2016 18:1*, *18*(1), 18–30. https://doi.org/10.1038/nrm.2016.116

Tian, E., Sawyer, J. R., Heuck, C. J., Zhang, Q., van Rhee, F., Barlogie, B., & Epstein, J. (2014). In Multiple Myeloma, 14q32 Translocations Are Non-Random Chromosomal Fusions Driving High Expression Levels of the Respective Partner Genes. *Genes, Chromosomes & Cancer*, *53*(7), 549. https://doi.org/10.1002/GCC.22165

Tichon, A., Gil, N., Lubelsky, Y., Solomon, T. H., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., & Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nature Communications 2016 7:1*, *7*(1), 1–10. https://doi.org/10.1038/ncomms12209

Tiessen, I., Abildgaard, M. H., Lubas, M., Gylling, H. M., Steinhauer, C., Pietras, E. J., Diederichs, S., Frankel, L. B., & Lund, A. H. (2019). A high-throughput screen identifies the long non-coding RNA DRAIC as a regulator of autophagy. *Oncogene*, *38*(26), 5127–5141. https://doi.org/10.1038/S41388-019-0783-9

Tomlinson, B., & Lazarus, H. M. (2017). Enhancing acute myeloid leukemia therapy - monitoring response using residual disease testing as a guide to therapeutic decision-making. *Expert Review of Hematology*, *10*(6), 563–574. https://doi.org/10.1080/17474086.2017.1326811

Trinh, B. Q., Ummarino, S., Zhang, Y., Ebralidze, A. K., Bassal, M. A., Nguyen, T. M., Heller, G., Coffey, R., Tenen, D. E., van der Kouwe, E., Fabiani, E., Gurnari, C., Wu, C. S., Angarica, V. E., Yang, H., Chen, S., Zhang, H., Thurm, A. R., Marchi, F., … Tenen, D. G. (2021). Myeloid lncRNA LOUP mediates opposing regulatory effects of RUNX1 and RUNX1-ETO in t(8;21) AML. *Blood*, *138*(15), 1331–1344. https://doi.org/10.1182/BLOOD.2020007920

Tsai, C. H., Yao, C. Y., Tien, F. M., Tang, J. L., Kuo, Y. Y., Chiu, Y. C., Lin, C. C., Tseng, M. H., Peng, Y. L., Liu, M. C., Liu, C. W., Yao, M., Lin, L. I., Chou, W. C., Chen, C. Y., Hou, H. A., & Tien, H. F. (2019). Incorporation of long non-coding RNA expression profile in the 2017 ELN risk classification can improve prognostic prediction of acute myeloid leukemia patients. *EBioMedicine*, *40*, 240–250. https://doi.org/10.1016/J.EBIOM.2019.01.022

Tuval, A., & Shlush, L. I. (2019). Evolutionary trajectory of leukemic clones and its clinical implications. *Haematologica*, *104*(5), 872–880. https://doi.org/10.3324/HAEMATOL.2018.195289

Udomsakdi, C., Eaves, C. J., Swolin, B., Reid, D. S., Barnett, M. J., & Eaves, A. C. (1992). Rapid decline of chronic myeloid leukemic cells in long-term culture due to a defect at the leukemic stem cell level. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(13), 6192. https://doi.org/10.1073/PNAS.89.13.6192

Uhrig, M., Ittrich, C., Wiedmann, V., Knyazev, Y., Weninger, A., Riemenschneider, M., & Hartmann, T. (2009). New Alzheimer Amyloid β Responsive Genes Identified in Human Neuroblastoma Cells by Hierarchical Clustering. *PLOS ONE*, *4*(8), e6779. https://doi.org/10.1371/JOURNAL.PONE.0006779

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature Reviews. Genetics*, *17*(10), 601–614. https://doi.org/10.1038/NRG.2016.85

Upadhaya, S., Reizis, B., & Sawai, C. M. (2018). New genetic tools for the in vivo study of hematopoietic stem cell function. *Experimental Hematology*, *61*, 26. https://doi.org/10.1016/J.EXPHEM.2018.02.004

Vakiti, A., & Mewawalla, P. (2023). Acute Myeloid Leukemia. *StatPearls*. https://www.ncbi.nlm.nih.gov/books/NBK507875/

Van Heesch, S., Van Iterson, M., Jacobi, J., Boymans, S., Essers, P. B., De Bruijn, E., Hao, W., MacInnes, A. W., Cuppen, E., & Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biology*, *15*(1), 1–12. https://doi.org/10.1186/GB-2014-15-1-R6/FIGURES/5

Van Rhenen, A., Van Dongen, G. A. M. S., Kelder, A. Le, Rombouts, E. J., Feller, N., Moshaver, B., Walsum, M. S. Van, Zweegman, S., Ossenkoppele, G. J., & Schuurhuis, G. J. (2007). The novel AML stem cell associated antigen CLL-1 aids in discrimination between normal and leukemic stem cells. *Blood*, *110*(7), 2659–2666. https://doi.org/10.1182/BLOOD-2007-03-083048

Vasconcelos, F. C., de Souza, P. S., Hancio, T., de Faria, F. C. C., & Maia, R. C. (2021). Update on drug transporter proteins in acute myeloid leukemia: Pathological implication and clinical setting. *Critical Reviews in Oncology/Hematology*, *160*. https://doi.org/10.1016/J.CRITREVONC.2021.103281

Vasseur, L., Fenwarth, L., Lambert, J., de Botton, S., Figeac, M., Villenet, C., Heiblig, M., Dumas, P. Y., Récher, C., Berthon, C., Lemasle, E., Lebon, D., Lambert, J., Terré, C., Celli-Lebras, K., Dombret, H., Preudhomme, C., Cheok, M., Itzykson, R., & Duployez, N. (2023). LSC17 score complements genetics and measurable residual disease in acute myeloid leukemia: an ALFA study. *Blood Advances*, *7*(15), 4024–4034. https://doi.org/10.1182/BLOODADVANCES.2023010155

Venditti, A., Del Poeta, G., Buccisano, F., Tamburini, A., Cox, M. C., Stasi, R., Bruno, A., Aronica, G., Maffei, L., Suppo, G., Simone, M. D., Forte, L., Cordero, V., Postorino, M., Tufilli, V., Isacchi, G., Masi, M., Papa, G., & Amadori, S. (1997). Minimally Differentiated Acute Myeloid Leukemia (AML-M0): Comparison of 25 Cases With Other French-American-British Subtypes. *Blood*, *89*(2), 621–629. https://doi.org/10.1182/BLOOD.V89.2.621

Venkatraman, A., He, X. C., Thorvaldsen, J. L., Sugimura, R., Perry, J. M., Tao, F., Zhao, M., Christenson, M. K., Sanchez, R., Yu, J. Y., Peng, L., Haug, J. S., Paulson, A., Li, H., Zhong, X. B., Clemens, T. L., Bartolomei, M. S., & Li, L. (2013). Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. *Nature*, *500*(7462), 345–349. https://doi.org/10.1038/NATURE12303

Verma, K., Zang, T., Gupta, N., Penning, T. M., & Trippier, P. C. (2016). Selective AKR1C3 Inhibitors Potentiate Chemotherapeutic Activity in Multiple Acute Myeloid Leukemia (AML) Cell Lines. *ACS Medicinal Chemistry Letters*, *7*(8), 774–779. https://doi.org/10.1021/ACSMEDCHEMLETT.6B00163/SUPPL_FILE/ML6B00163_SI_001.PDF

Versluis, J., Pandey, M., Flamand, Y., Erika Haydu, J., Belizaire, R., Faber, M., Vedula, R. S., Charles, A., Copson, K. M., Shimony, S., Rozental, A., Bendapudi, P. K., Wolach, O., Griffiths, E. A., Thompson, J. E., Stone, R. M., DeAngelo, D. J., Neuberg, D., Luskin, M. R., … Coleman Lindsley, R. (2022). Prediction of life-threatening and disabling bleeding in patients with AML receiving intensive induction chemotherapy. *Blood Advances*, *6*(9), 2835. https://doi.org/10.1182/BLOODADVANCES.2021006166

Vetrie, D., Helgason, G. V., & Copland, M. (2020). The leukaemia stem cell: similarities, differences and clinical prospects in CML and AML. *Nature Reviews. Cancer*, *20*(3), 158–173. https://doi.org/10.1038/S41568-019-0230-9

Vitali, C., Bassani, C., Chiodoni, C., Fellini, E., Guarnotta, C., Miotti, S., Sangaletti, S., Fuligni, F., De Cecco, L., Piccaluga, P. P., Colombo, M. P., & Tripodo, C. (2015). SOCS2 controls proliferation and stemness of hematopoietic cells under stress conditions and its deregulation marks unfavorable acute leukemias. *Cancer Research*, *75*(11), 2387–2399.

https://doi.org/10.1158/0008-5472.CAN-14-3625/651810/AM/SOCS2-CONTROLS-PROLIFERATION-AND-STEMNESS-OF

Vosberg, S., & Greif, P. A. (2019). Clonal evolution of acute myeloid leukemia from diagnosis to relapse. *Genes, Chromosomes & Cancer*, *58*(12), 839–849. https://doi.org/10.1002/GCC.22806

Vu, L. P., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., Mackay, M., Schulman, J., Famulare, C., Patel, M., Klimek, V. M., Garrett-Bakelman, F. E., Melnick, A., Carroll, M., Mason, C. E., Jaffrey, S. R., & Kharas, M. G. (2017). The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nature Medicine*, *23*(11), 1369–1376. https://doi.org/10.1038/NM.4416

Vu, L. P., Prieto, C., Amin, E. M., Chhangawala, S., Krivtsov, A., Calvo-Vidal, M. N., Chou, T., Chow, A., Minuesa, G., Park, S. M., Barlowe, T. S., Taggart, J., Tivnan, P., Deering, R. P., Chu, L. P., Kwon, J. A., Meydan, C., Perales-Paton, J., Arshi, A., … Kharas, M. G. (2017). Functional screen of MSI2 interactors identifies an essential role for SYNCRIP in myeloid leukemia stem cells. *Nature Genetics*, *49*(6), 866–875. https://doi.org/10.1038/NG.3854

Vujovic, A., de Rooij, L., Chahi, A. K., Chen, H. T., Yee, B. A., Loganathan, S. K., Liu, L., Chan, D. C. H., Tajik, A., Tsao, E., Moreira, S., Joshi, P., Xu, J., Wong, N., Balde, Z., Jahangiri, S., Zandi, S., Aigner, S., Dick, J. E., … Hope, K. J. (2023). In Vivo Screening Unveils Pervasive RNA-Binding Protein Dependencies in Leukemic Stem Cells and Identifies ELAVL1 as a Therapeutic Target. *Blood Cancer Discovery*, *4*(3), 181–207. https://doi.org/10.1158/2643-3230.BCD-22-0086/716497/AM/IN-VIVO-SCREENING-UNVEILS-PERVASIVE-RNA-BINDING

Wallace, I. M., O'Sullivan, O., Higgins, D. G., & Notredame, C. (2006). M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Research*, *34*(6), 1692–1699. https://doi.org/10.1093/NAR/GKL091

Walter, R. B., Ofran, Y., Wierzbowska, A., Ravandi, F., Hourigan, C. S., Ngai, L. L., Venditti, A., Buccisano, F., Ossenkoppele, G. J., & Roboz, G. J. (2021). Measurable residual disease as a biomarker in acute myeloid leukemia: theoretical and practical considerations. *Leukemia*, *35*(6), 1529–1538. https://doi.org/10.1038/S41375-021-01230-4

Walters, K., Sajek, M. P., Murphy, E., Issaian, A., Baldwin, A., Harrison, E., Daniels, M., Reisz, J. A., Hansen, K., D'Alessandro, A., & Mukherjee, N. (2023). Small-molecule Ro-08-2750 interacts with many RNA-binding proteins and elicits MUSASHI2-independent phenotypes. *RNA*, *29*(10), 1458–1470. https://doi.org/10.1261/RNA.079605.123/-/DC1

Wang, C., Li, L., Li, M., Wang, W., Liu, Y., & Wang, S. (2020). Silencing long non-coding RNA XIST suppresses drug resistance in acute myeloid leukemia through down-regulation of MYC by elevating microRNA-29a expression. *Molecular Medicine*, *26*(1), 1–11. https://doi.org/10.1186/S10020-020-00229-4/FIGURES/6

Wang, D., Zeng, T., Lin, Z., Yan, L., Wang, F., Tang, L., Wang, L., Tang, D., Chen, P., & Yang, M. (2020). Long non-coding RNA SNHG5 regulates chemotherapy resistance through the miR-32/DNAJB9 axis in acute myeloid leukemia. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, *123*. https://doi.org/10.1016/J.BIOPHA.2019.109802

Wang, E., Lu, S. X., Pastore, A., Chen, X., Imig, J., Chun-Wei Lee, S., Hockemeyer, K., Ghebrechristos, Y. E., Yoshimi, A., Inoue, D., Ki, M., Cho, H., Bitner, L., Kloetgen, A., Lin, K. T., Uehara, T., Owa, T., Tibes, R., Krainer, A. R., … Aifantis, I. (2019). Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. *Cancer Cell*, *35*(3), 369-384.e7. https://doi.org/10.1016/J.CCELL.2019.01.010

Wang, F., Tian, X., Zhou, J., Wang, G., Yu, W., Li, Z., Fan, Z., Zhang, W., & Liang, A. (2018). A three-lncRNA signature for prognosis prediction of acute myeloid leukemia in patients. *Molecular Medicine Reports*, *18*(2), 1473. https://doi.org/10.3892/MMR.2018.9139

Wang, H., Li, W., Guo, R., Sun, J., Cui, J., Wang, G., Hoffman, A. R., & Hu, J. F. (2014). An intragenic long noncoding RNA interacts epigenetically with the RUNX1 promoter and enhancer chromatin DNA in hematopoietic malignancies. *International Journal of Cancer*, *135*(12), 2783–2794. https://doi.org/10.1002/IJC.28922

Wang, Q., Stacy, T., Binder, M., Marín-Padilla, M., Sharpe, A. H., & Speck, N. A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(8), 3444–3449. https://doi.org/10.1073/PNAS.93.8.3444

Wang, X. Q. D., & Dostie, J. (2017). Reciprocal regulation of chromatin state and architecture by HOTAIRM1 contributes to temporal collinear HOXA gene activation. *Nucleic Acids Research*, *45*(3), 1091. https://doi.org/10.1093/NAR/GKW966

Wang, Y. H., Lin, C. C., Hsu, C. L., Hung, S. Y., Yao, C. Y., Lee, S. H., Tsai, C. H., Hou, H. A., Chou, W. C., & Tien, H. F. (2021). Distinct clinical and biological characteristics of acute myeloid leukemia with higher expression of long noncoding RNA KIAA0125. *Annals of Hematology*, *100*(2), 487–498. https://doi.org/10.1007/S00277-020-04358-Y

Wang, Y., Liu, Y., Malek, S. N., Zheng, P., & Liu, Y. (2011). Targeting HIF1α eliminates cancer stem cells in hematological malignancies. *Cell Stem Cell*, *8*(4), 399–411. https://doi.org/10.1016/J.STEM.2011.02.006

Warner, J. K., Wang, J. C. Y., Hope, K. J., Jin, L., & Dick, J. E. (2004). Concepts of human leukemic development. *Oncogene*, *23*(43), 7164–7177. https://doi.org/10.1038/SJ.ONC.1207933

Washietl, S., Kellis, M., & Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Research*, *24*(4), 616–628. https://doi.org/10.1101/GR.165035.113

Wattad, M., Weber, D., Döhner, K., Krauter, J., Gaidzik, V. I., Paschka, P., Heuser, M., Thol, F., Kindler, T., Lübbert, M., Salih, H. R., Kündgen, A., Horst, H. A., Brossart, P., Götze, K., Nachbaur, D., Köhne, C. H., Ringhoffer, M., Wulf, G., … Schlenk, R.

F. (2017). Impact of salvage regimens on response and overall survival in acute myeloid leukemia with induction failure. *Leukemia*, *31*(6), 1306–1313. https://doi.org/10.1038/LEU.2017.23

Weissman, I. L. (2000). Translating Stem and Progenitor Cell Biology to the Clinic: Barriers and Opportunities. *Science*, *287*(5457), 1442–1446. https://doi.org/10.1126/SCIENCE.287.5457.1442

Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., … Wilson, R. K. (2012). The origin and evolution of mutations in acute myeloid leukemia. *Cell*, *150*(2), 264–278. https://doi.org/10.1016/J.CELL.2012.06.023

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B. S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C., Sheng, Y., Wang, Y., Wunderlich, M., Zhang, B., Dore, L. C., Su, R., Deng, X., Ferchen, K., Li, C., … Chen, J. (2018). METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m6A Modification. *Cell Stem Cell*, *22*(2), 191-205.e9. https://doi.org/10.1016/J.STEM.2017.11.016

Wesely, J., Kotini, A. G., Izzo, F., Luo, H., Yuan, H., Sun, J., Georgomanoli, M., Zviran, A., Deslauriers, A. G., Dusaj, N., Nimer, S. D., Leslie, C., Landau, D. A., Kharas, M. G., & Papapetrou, E. P. (2020). Acute Myeloid Leukemia iPSCs Reveal a Role for RUNX1 in the Maintenance of Human Leukemia Stem Cells. *Cell Reports*, *31*(9). https://doi.org/10.1016/J.CELREP.2020.107688

Westervelt, P., Lane, A. A., Pollock, J. L., Oldfather, K., Holt, M. S., Zimonjic, D. B., Popescu, N. C., DiPersio, J. F., & Ley, T. J. (2003). High-penetrance mouse model of acute promyelocytic leukemia with very low levels of PML-RARalpha expression. *Blood*, *102*(5), 1857–1865. https://doi.org/10.1182/BLOOD-2002-12-3779

Wiedenheft, B., Sternberg, S. H., & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, *482*(7385), 331–338. https://doi.org/10.1038/NATURE10886

Wilkinson, A. C., Ballabio, E., Geng, H., North, P., Tapia, M., Kerry, J., Biswas, D., Roeder, R. G., Allis, C. D., Melnick, A., de Bruijn, M. F. T. R., & Milne, T. A. (2013). RUNX1 is a key target in t(4;11) leukemias that contributes to gene activation through an AF4-MLL complex interaction. *Cell Reports*, *3*(1), 116–127. https://doi.org/10.1016/J.CELREP.2012.12.016

Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lió, P., MacDonald, H. R., & Trumpp, A. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, *135*(6), 1118–1129. https://doi.org/10.1016/J.CELL.2008.10.048

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E., Pimanda, J. E., De Bruijn, M. F. T. R., & Göttgens, B. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, *7*(4), 532–544. https://doi.org/10.1016/J.STEM.2010.07.016

Wingelhofer, B., Maurer, B., Heyes, E. C., Cumaraswamy, A. A., Berger-Becvar, A., De Araujo, E. D., Orlova, A., Freund, P., Ruge, F., Park, J., Tin, G., Ahmar, S., Lardeau, C. H., Sadovnik, I., Bajusz, D., Keseru, G. M., Grebien, F., Kubicek, S., Valent, P., … Moriggl, R. (2018). Pharmacologic inhibition of STAT5 in acute myeloid leukemia. *Leukemia 2018 32:5*, *32*(5), 1135–1146. https://doi.org/10.1038/s41375-017-0005-9

Winkle, M., El-Daly, S. M., Fabbri, M., & Calin, G. A. (2021). Noncoding RNA therapeutics — challenges and potential solutions. *Nature Reviews Drug Discovery 2021 20:8*, *20*(8), 629–651. https://doi.org/10.1038/s41573-021-00219-z

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359–362. https://doi.org/10.1038/NMETH.1322

Wu, G., Wang, Q., Zhu, T., Fu, L., Li, Z., Wu, Y., & Zhang, C. (2021). Identification and Validation of Immune-Related LncRNA Prognostic Signature for Lung Adenocarcinoma. *Frontiers in Genetics*, *12*, 681277. https://doi.org/10.3389/FGENE.2021.681277/FULL

Wu, Z., Gao, S., Zhao, X., Chen, J., Keyvanfar, K., Feng, X., Kajigaya, S., & Young, N. S. (2019). Long noncoding RNAs of single hematopoietic stem and progenitor cells in healthy and dysplastic human bone marrow. *Haematologica*, *104*(5), 894–906. https://doi.org/10.3324/HAEMATOL.2018.208926

Wuebben, E. L., Mallanna, S. K., Cox, J. L., & Rizzino, A. (2012). Musashi2 Is Required for the Self-Renewal and Pluripotency of Embryonic Stem Cells. *PLOS ONE*, *7*(4), e34827. https://doi.org/10.1371/JOURNAL.PONE.0034827

Wunderlich, M., Chou, F. S., Link, K. A., Mizukawa, B., Perry, R. L., Carroll, M., & Mulloy, J. C. (2010). AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. *Leukemia*, *24*(10), 1785–1788. https://doi.org/10.1038/LEU.2010.158

Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., Ozenberger, B. A., Welch, J. S., Link, D. C., Walter, M. J., Mardis, E. R., Dipersio, J. F., Chen, F., Wilson, R. K., Ley, T. J., & Ding, L. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature Medicine*, *20*(12), 1472–1478. https://doi.org/10.1038/NM.3733

Yan, H., Wang, Z., Sun, Y., Hu, L., & Bu, P. (2021). Cytoplasmic NEAT1 Suppresses AML Stem Cell Self-Renewal and Leukemogenesis through Inactivation of Wnt Signaling. *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)*, *8*(22). https://doi.org/10.1002/ADVS.202100914

Yanagisawa, K., Horiuchi, T., & Fujita, S. (1991). Establishment and Characterization of a New Human Leukemia Cell Line Derived From M4E0. *Blood*, *78*(2), 451–457. https://doi.org/10.1182/BLOOD.V78.2.451.451

Yang, Y., Dai, W., Sun, Y., & Zhao, Z. (2019). Long non-coding RNA linc00239 promotes malignant behaviors and chemoresistance against doxorubicin partially via activation of the PI3K/Akt/mTOR pathway in acute myeloid leukaemia cells. *Oncology Reports*, *41*(4), 2311–2320. https://doi.org/10.3892/OR.2019.6991

Yang, Y., Zhao, Y., Hu, N., Zhao, J., & Bai, Y. (2019). lncRNA KIAA0125 functions as a tumor suppressor modulating growth and metastasis of colorectal cancer via Wnt/β-catenin pathway. *Cell Biology International*, *43*(12), 1463. https://doi.org/10.1002/CBIN.11196

Yao, K., Wang, Q., Jia, J., & Zhao, H. (2017). A competing endogenous RNA network identifies novel mRNA, miRNA and lncRNA markers for the prognosis of diabetic pancreatic cancer. *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine*, *39*(6). https://doi.org/10.1177/1010428317707882

Yap, K., Mukhina, S., Zhang, G., Tan, J. S. C., Ong, H. S., & Makeyev, E. V. (2018). A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. *Molecular Cell*, *72*(3), 525-540.e13. https://doi.org/10.1016/J.MOLCEL.2018.08.041

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., … Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, *44*(D1), D710–D716. https://doi.org/10.1093/NAR/GKV1157

Yeh, D. W., Zhao, X., Siddique, H. R., Zheng, M., Choi, H. Y., Machida, T., Narayanan, P., Kou, Y., Punj, V., Tahara, S. M., Feldman, D. E., Chen, L., & Machida, K. (2023). MSI2 promotes translation of multiple IRES-containing oncogenes and virus to induce self-renewal of tumor initiating stem-like cells. *Cell Death Discovery 2023 9:1*, *9*(1), 1–15. https://doi.org/10.1038/s41420-023-01427-9

Yildirim, E., Kirby, J. E., Brown, D. E., Mercier, F. E., Sadreyev, R. I., Scadden, D. T., & Lee, J. T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell*, *152*(4), 727–742. https://doi.org/10.1016/J.CELL.2013.01.034

Yokose, N., Ogata, K., Ito, T., Miyake, K., An, E., Inokuchi, K., Yamada, T., Gomi, S., Tanabe, Y., Ohki, I., Kuwabara, T., Hasegawa, S., Shinohara, T., Dan, K., & Nomura, T. (1993). Chemotherapy for minimally differentiated acute myeloid leukemia (AML-M0). A report on five cases and review of the literature. *Annals of Hematology*, *66*(2), 67–70. https://doi.org/10.1007/BF01695886

Zappulla, D. C., & Cech, T. R. (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(27), 10024–10029. https://doi.org/10.1073/PNAS.0403641101

Zhan, Y., Chen, Z., Li, Y., He, A., He, S., Gong, Y., Li, X., & Zhou, L. (2018). Long non-coding RNA DANCR promotes malignant phenotypes of bladder cancer cells by modulating the miR-149/MSI2 axis as a ceRNA. *Journal of Experimental & Clinical Cancer Research : CR*, *37*(1). https://doi.org/10.1186/S13046-018-0921-1

Zhang, H., Tan, S., Wang, J., Chen, S., Quan, J., Xian, J., Zhang, S. S., He, J., & Zhang, L. (2014). Musashi2 modulates K562 leukemic cell proliferation and apoptosis involving the MAPK pathway. *Experimental Cell Research*, *320*(1), 119–127. https://doi.org/10.1016/J.YEXCR.2013.09.009

Zhang, M., Zhang, L. L., Yi, L. B., Tu, X. N., Zhou, Y., Li, D. Y., Xue, H. C., Li, Y. X., & Zheng, Z. Z. (2024). Comprehensive analysis of immune-related lncRNAs in AML patients uncovers potential therapeutic targets and prognostic biomarkers. *Heliyon*, *10*(9). https://doi.org/10.1016/j.heliyon.2024.e30616

Zhang, T., Liao, D., & Hu, Y. (2023). Cuproptosis-related lncRNAs forecast the prognosis of acute myeloid leukemia. *Translational Cancer Research*, *12*(5), 1175–1195. https://doi.org/10.21037/TCR-22-2526/COIF)

Zhang, X., Lian, Z., Padden, C., Gerstein, M. B., Rozowsky, J., Snyder, M., Gingeras, T. R., Kapranov, P., Weissman, S. M., & Newburger, P. E. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood*, *113*(11), 2526–2534. https://doi.org/10.1182/BLOOD-2008-06-162164

Zhang, X., Su, K., Liu, Y., Zhu, D., Pan, Y., Ke, X., & Qu, Y. (2022). Small Molecule Palmatine Targeting Musashi-2 in Colorectal Cancer. *Frontiers in Pharmacology*, *12*, 793449. https://doi.org/10.3389/FPHAR.2021.793449/BIBTEX

Zhao, C., Wang, Y., Tu, F., Zhao, S., Ye, X., Liu, J., Zhang, J., & Wang, Z. (2021). A Prognostic Autophagy-Related Long Non-coding RNA (ARlncRNA) Signature in Acute Myeloid Leukemia (AML). *Frontiers in Genetics*, *12*. https://doi.org/10.3389/FGENE.2021.681867

Zhao, C., Zi, P., & Zhou, D. (2021). Long non-coding RNA NEAT1/miR-320b/MSI2 axis regulates cisplatin resistance in ovarian cancer. *Archives of Medical Science*. https://doi.org/10.5114/AOMS/131566

Zhao, J., Xiu, Y., Fu, L., Dong, Q., Borcherding, N., Wang, Y., Li, Q., De Silva, N. S., Klein, U., Boyce, B. F., & Zhao, C. (2021). TIFAB accelerates MLL-AF9-Induced acute myeloid leukemia through upregulation of HOXA9. *IScience*, *24*(12). https://doi.org/10.1016/J.ISCI.2021.103425

Zhao, R., Fu, J., Zhu, L., Chen, Y., & Liu, B. (2022). Designing strategies of small-molecule compounds for modulating non-coding RNAs in cancer therapy. *Journal of Hematology & Oncology 2022 15:1*, *15*(1), 1–19. https://doi.org/10.1186/S13045-022-01230-6

Zheng, R., & Small, D. (2005). Mutant FLT3 signaling contributes to a block in myeloid differentiation. *Leukemia & Lymphoma*, *46*(12), 1679–1687. https://doi.org/10.1080/10428190500261740

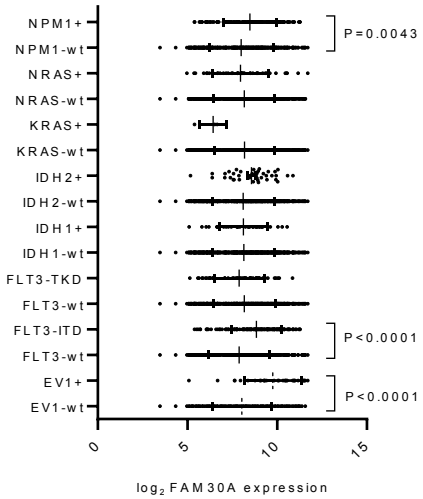
Zhong, F., Yao, F., Cheng, Y., Liu, J., Zhang, N., Li, S., Li, M., Huang, B., & Wang, X. (2022). m6A-related lncRNAs predict prognosis and indicate immune microenvironment in acute myeloid leukemia. *Scientific Reports 2022 12:1*, *12*(1), 1–17. https://doi.org/10.1038/s41598-022-05797-5

Zhu, Y., He, J., Li, Z., & Yang, W. (2023). Cuproptosis-related lncRNA signature for prognostic prediction in patients with acute myeloid leukemia. *BMC Bioinformatics*, *24*(1), 1–17. https://doi.org/10.1186/S12859-023-05148-9/FIGURES/7

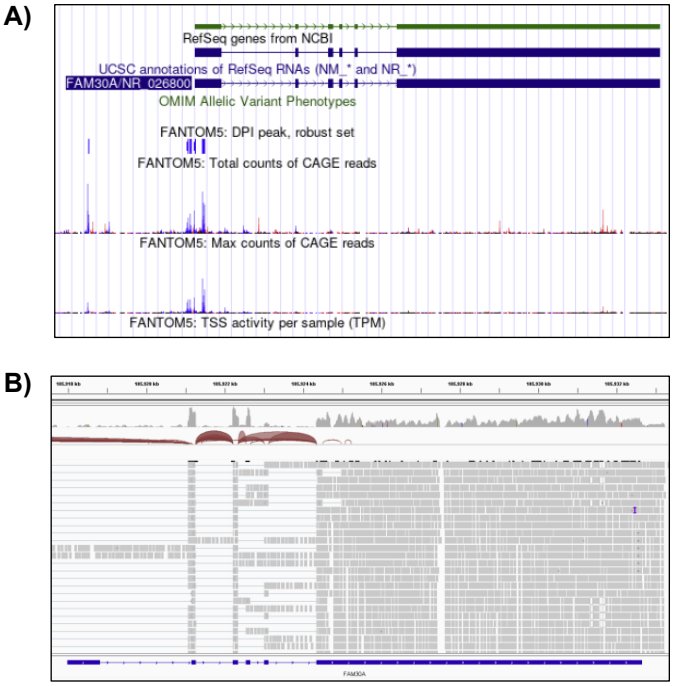
Zhu, Y., Zhou, B., Hu, X., Ying, S., Zhou, Q., Xu, W., Feng, L., Hou, T., Wang, X., Zhu, L., & Jin, H. (2022). LncRNA LINC00942 promotes chemoresistance in gastric cancer by suppressing MSI2 degradation to enhance c-Myc mRNA stability. *Clinical and Translational Medicine*, *12*(1). https://doi.org/10.1002/CTM2.703

7. APPENDIX

7.1.- Supplementary data



Supplemental Figure 1. FAM30A expression correlates with worse mutation patterns in AML. RNA expression levels of FAM30A expressed in terms of log_2 in the most common mutations found among AML patients. Data extracted from R2 platform (Bohlander, n=422). + represents mutations in the gene, wt = wild-type; ITD= internal tandem duplication; TKD= mutation in the tyrosine kinase domain. Statistical significance was calculated using unpaired two-tailed student t-test.

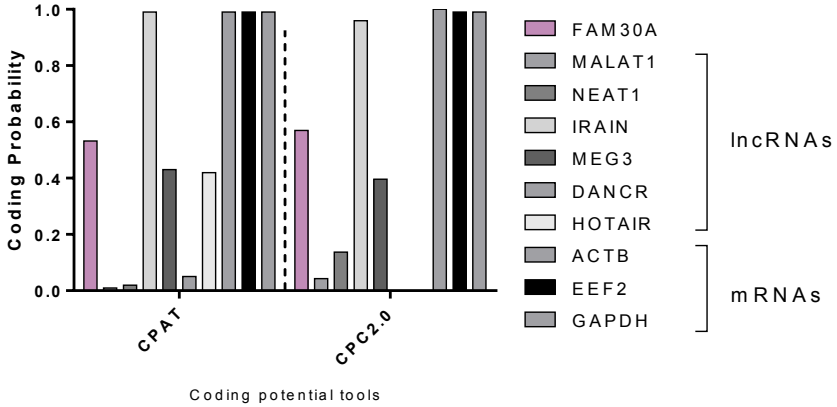


Supplemental Figure 2. Exon usage of FAM30A. A) Visualization of the most constitutively expressed exons in FAM30A shown from UCSC Genome browser (Refseq FAM30A / NR_026800) and showing results from FANTOM5 for prediction of Transcription Starting Sites (TSS).

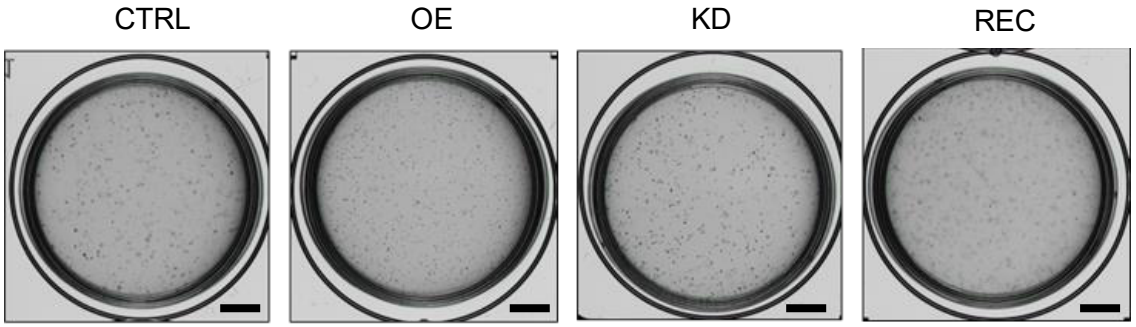
-Continued on next page-

B) Depiction from IGV software from the RNA-sequencing performed in normal CD34+ cord-blood cells. The image shows the main five exons of FAM30A with the highest raw reads.

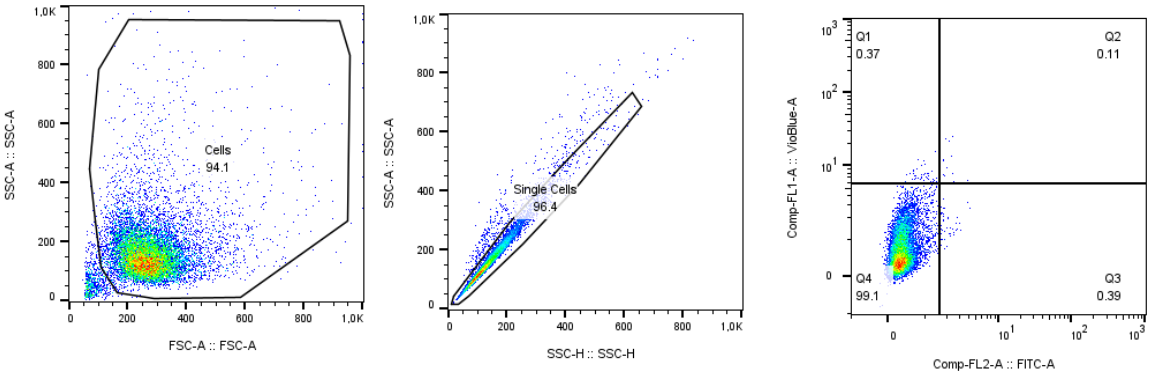
Supplemental Figure 3. Multiple sequence alignment of FAM30A repeat and its flanking region in human and primates. Results from multiple sequence using M-Coffee software. Upper legend shows conservation score in the species used for the alignment. Hs= homo sapiens; Hm= Hylobates moloch; Pa= Papio Anubis; Ca= Cercocebus atys; An= Aoutus nancymaae.



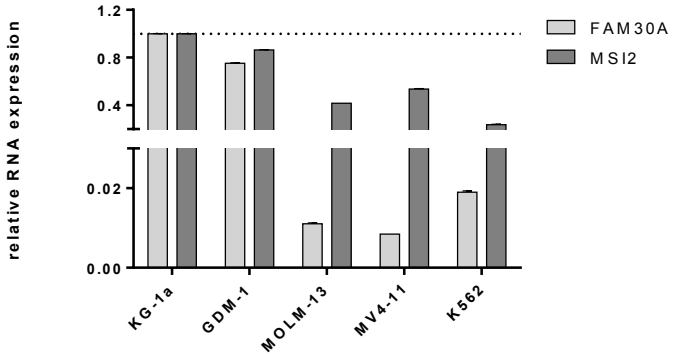
Supplemental Figure 4. Protein-coding probabilities of FAM30A and other lncRNAs and mRNAs. Graph depicting coding probabilities of FAM30A and established cytoplasmic or nuclear long non-coding RNAs (lncRNAs) and compared with mRNAs with known protein-coding potential. All RNA sequences were extracted from NCBI website and aligned using two protein-coding bioinformatic tools: Coding-Potential Assessment Tool (CPAT) and Coding Potential Calculator2.0 (CPC2.0).



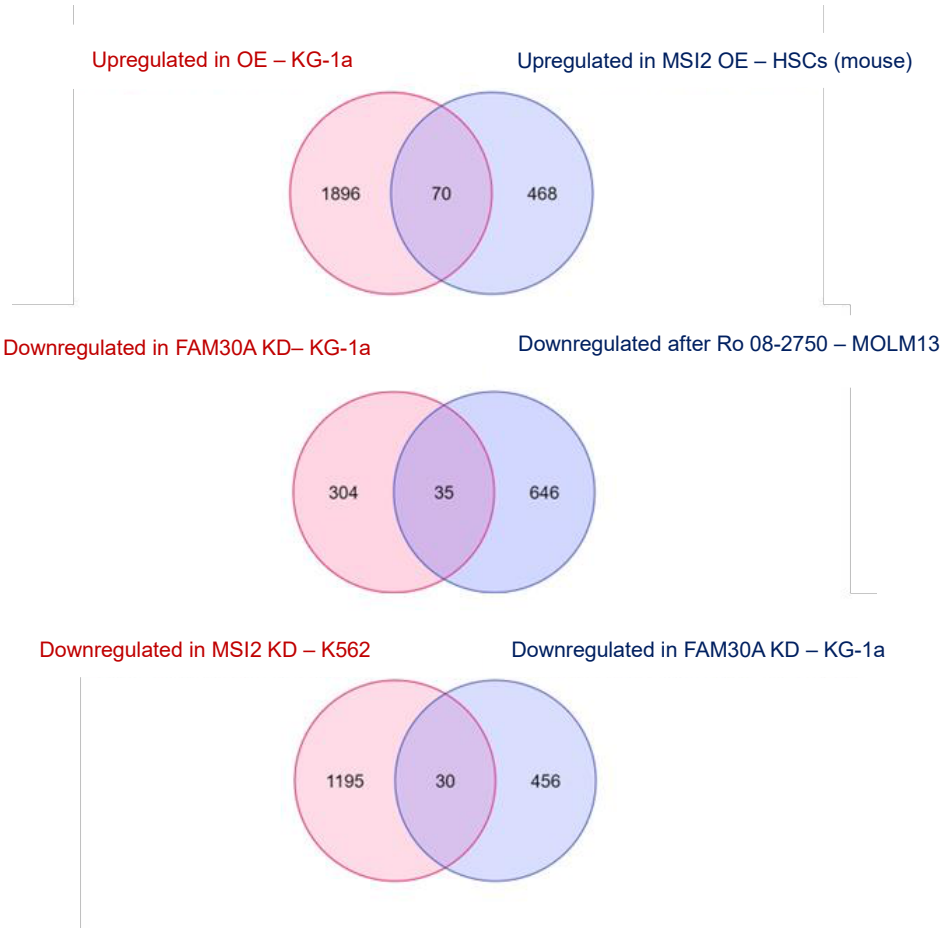
Supplemental Figure 5. Colony-formation assay after modulation of expression of FAM30A and its repeats. Images of the plates in the different conditions captured with Licor-Oddysey. Scale bar represents 500 mm.



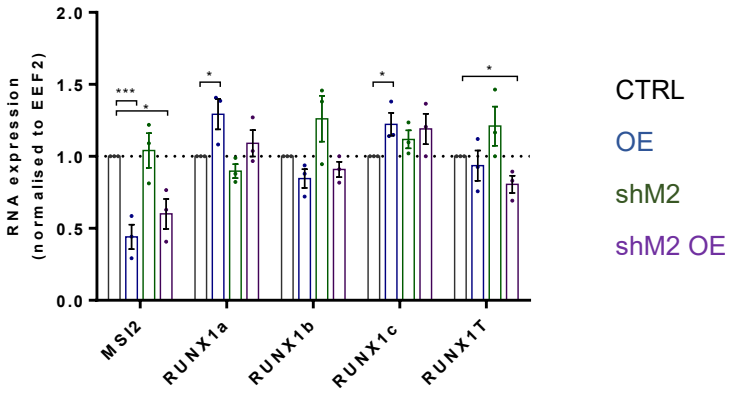
Supplemental Figure 6. Representative gating strategy used for FACS staining. Images after analysis in Flow-Jo software for living cells (left), singlets (middle) and for negative staining (right).



Supplemental Figure 7. FAM30A and MSI2 RNA expression levels in the leukaemia cell lines used in this thesis. RNA expression levels relative to ACTB and normalized against KG-1a cell line that showed the highest expression values for both RNAs. Data is shown in terms of mean and error bars represent SEM.

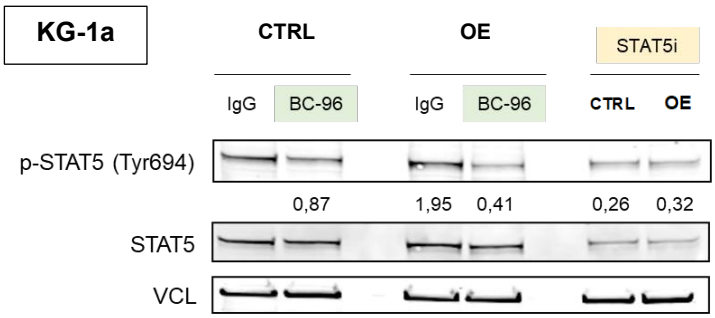


Supplemental Figure 8. Altering the expression levels of FAM30A and MSI2 impacts only a small subset of common downstream targets. Venn diagrams showing number of significantly down- and upregulated targets upon modulation of FAM30A expression (and its repeats) in KG-1a cells and MSI2 knockdown or overexpression in different leukaemia cell lines or *in vivo* experiments. RNA-sequencing and microarrays data for MSI2 was derived from publicly available datasets extracted from (Minuesa et al., 2019; Duggimpudi et al., 2018; De Andrés-Aguayo et al., 2011). Ro 08-2750 = MSI2 inhibitor.

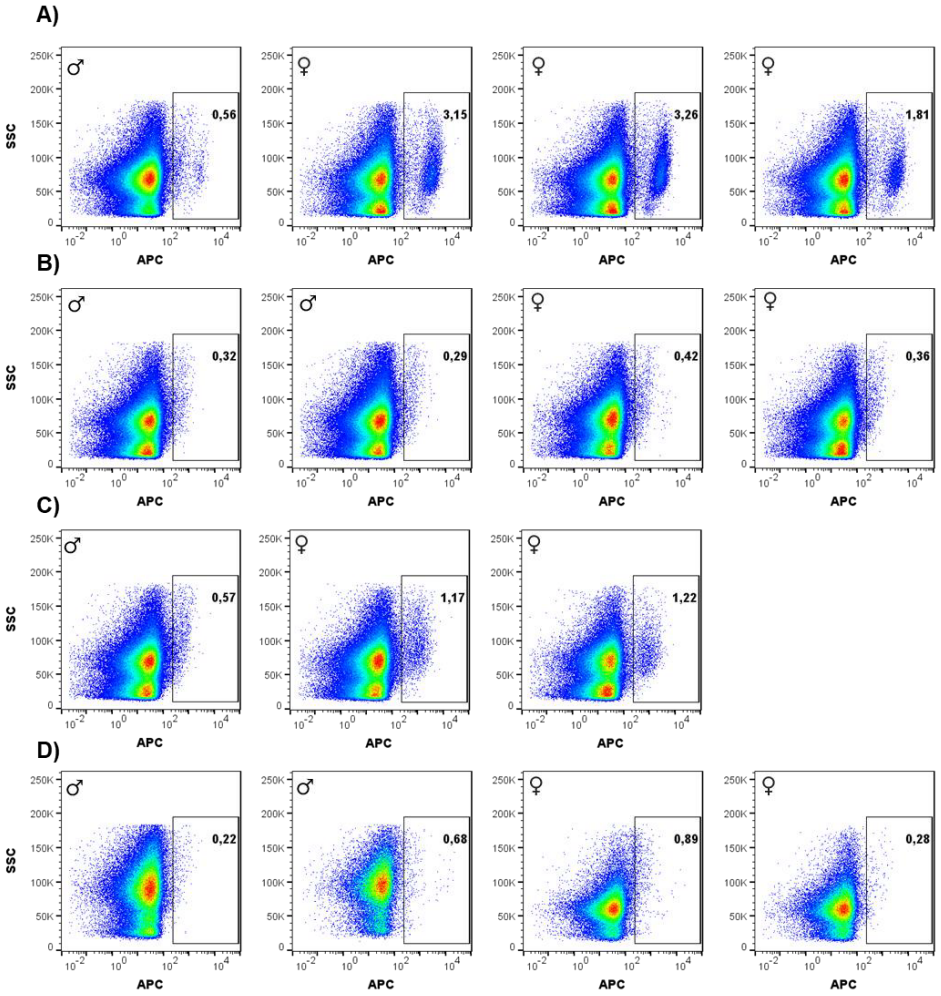


Supplemental Figure 9. Alterations of RUNX1 isoforms at the RNA level remain unchanged after MSI2 knockdown. Graph depicting RT-qPCR analysis upon stable MSI2 knockdown in combination with FAM30A repeats overexpression in K562 cells. CTRL: control cells, shM2: MSI2 knockdown, OE: FAM30A repeats overexpression, shM2 OE: MSI2 knockdown combined with FAM30A repeats overexpression.
-Continued on next page-

Statistical significance was calculated using unpaired two-tailed student t-test. (*** - p<0,001; **- p<0,01; *- p<0,05). Data is displayed as mean values and error bars represent SEM.



<u>Supplemental Figure 10</u>. Comparison of α-CD25 treatment and STAT5 inhibition on STAT5 activation upon modulation of FAM30A repeats. Western blot analysis in stable KG-1a cells: shC EV (control cells) and shC FAM (FAM30A repeats overexpression) treated with 3 μg/mL of control IgG, BC-96 (α-CD25) and 10 μM of STAT5i (STAT5 inhibitor (Wingelhofer et al., 2018)) for 48 hours. VCL was used as a loading control (n=1).



<u>Supplemental Figure 11.</u> FACS staining of mice used for the study of how FAM30A and it repeats affect engraftment in the bone marrow.

-Continued on next page-

Images after analysis with Flow-Jo software of all mice used for the engraftment of KG-1a cells, highlighting percentage of APC-positive cells as percentage of engraftment of human cells. **A)** control (CTRL), **B)** FAM30A knockdown (KD), **C)** FAM30A knockdown and overexpression of FAM30A repeats (REC), **D)** no engraftment of human cells (Unst).

Supplementary Table 1. Mass spectrometry analysis of significantly enriched pulled-down proteins with FAM30A repeats. List of significantly pulled down proteins detected after RNA-pulldown using *in vitro* transcribed FAM30A repeats as bait and compared to no RNA (BC) and a flanking region of FAM30A outside of the repeats (EV). The cut off used: $\log_2$ (abundance ratio) >3 and adjusted p-value < 0.001 vs BC and adjusted p-value < 0.05 vs EV).

DESCRIPTION	Abundance (FAM) / (BC)	p-value: vs (BC)	p-value: vs (EV)
ATP-dependent RNA helicase A DHX9	6,15	5,24E-11	0,000748
Phosphatidylinositol 4-phosphate 5-kinase type-1 alpha PIP5K1A	5,04	5,24E-11	0,000935
RNA-binding protein 28 RBM28	4,15	5,24E-11	0,00152
Zinc finger protein 101 ZNF101	6,83	5,24E-11	0,008986
Serine/arginine-rich splicing factor 9 SRSF9	3,49	2,83E-09	3,95E-06
ATP-dependent DNA/RNA helicase DHX36 DHX36	5,06	1,18E-06	8,54E-05
Target of EGR1 protein 1 TOE1	6,19	1,18E-06	0,000436
5'-3' exoribonuclease 2 XRN2	3,64	1,18E-06	0,00707
Ribosome biogenesis protein BRX1 homolog BRIX1	5,89	1,94E-06	0,006528
Nuclear fragile X mental retardation-interacting protein 1 NUFIP1	4,26	2,17E-06	0,001497
DNA-directed RNA polymerase, mitochondrial POLRMT	3,16	2,19E-06	0,000173
Double-stranded RNA-specific editase 1 ADARB1	4,76	2,4E-06	0,004724
Zinc finger protein ZXDC ZXDC	3,72	2,81E-06	0,006404
Zinc finger protein 335 ZNF335	6,23	3,92E-06	0,027132
Interferon-induced, double-stranded RNA-activated protein kinase EIF2AK2	4,99	6,65E-06	0,000132
Zinc finger protein 22 ZNF22	5,08	8,28E-06	0,029334
rRNA 2'-O-methyltransferase fibrillarin FBL	3,35	8,42E-06	0,016086
Transcriptional activator protein Pur-beta PURB	5,26	9,28E-06	6,26E-05
RNA-binding protein Musashi homolog 2 MSI2	3,57	9,57E-06	0,005332
Interleukin enhancer-binding factor 3 ILF3	5,13	9,63E-06	0,007458
Hepatocyte growth factor HGF	3,3	9,67E-06	0,000953
Ribonucleases P/MRP protein subunit POP1 POP1	3,36	1,03E-05	0,000123
CDKN2A-interacting protein CDKN2AIP	5,22	1,17E-05	0,008337
Heterogeneous nuclear ribonucleoprotein A0 HNRNPA0	6,21	1,23E-05	0,000114
RNA-binding motif protein, X chromosome RBMX	5,32	1,23E-05	0,004136
ATP-dependent RNA helicase DHX30 DHX30	3,75	1,23E-05	0,027621
Heterogeneous nuclear ribonucleoprotein U-like protein 2 HNRNPUL2	3,61	1,31E-05	0,002005
Lysine-rich nucleolar protein 1 KNOP1	4,44	1,45E-05	0,002492
Nucleolar RNA helicase 2 DDX21	5,19	1,49E-05	0,004756
Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase ALG3	3,62	1,58E-05	8,11E-05
Interleukin enhancer-binding factor 2 ILF2	5,46	1,6E-05	0,013791
Nucleolysin TIAR TIAL1	3,03	1,62E-05	0,000397
rRNA N6-adenosine-methyltransferase ZCCHC4 ZCCHC4	4,93	1,62E-05	0,012895
Zinc finger protein 180 ZNF180	3,74	1,62E-05	0,026189
Zinc finger CCCH domain-containing protein 8 ZC3H8	4,74	1,66E-05	0,008753

Zinc finger protein 652 ZNF652	4,79	1,66E-05	0,014006
39S ribosomal protein L15, mitochondrial MRPL15	3,41	1,84E-05	0,026004
Heterogeneous nuclear ribonucleoprotein L HNRNPL	7,15	1,86E-05	6,36E-05
Zinc finger protein 324A ZNF324	4,04	1,91E-05	0,002717
U1 small nuclear ribonucleoprotein A SNRPA	5,87	1,99E-05	0,001343
Double-stranded RNA-binding protein Staufen homolog 1 STAU1	4,17	1,99E-05	0,022655
Activator of basal transcription 1 ABT1	4,4	2,02E-05	0,029861
Nucleoside diphosphate kinase, mitochondrial NME4	6,33	2,02E-05	0,038175
GTP-binding protein 2 GTPBP2	4,02	2,07E-05	0,033891
Chromodomain-helicase-DNA-binding protein 4 CHD4	5,14	2,12E-05	9,1E-05
Interferon-inducible double-stranded RNA-dependent protein kinase activator A PRKRA	3,35	2,55E-05	0,000813
Mediator of RNA polymerase II transcription subunit 19 MED19	3,4	2,56E-05	0,002356
PHD finger protein 19 PHF19	3,44	2,59E-05	0,001905
Lupus La protein SSB	4,73	2,61E-05	0,004241
S1 RNA-binding domain-containing protein 1 SRBD1	4,61	2,67E-05	0,032996
ATP-dependent RNA helicase DDX55 DDX55	3,06	2,81E-05	9,76E-05
Zinc finger protein 121 ZNF121	4,11	2,87E-05	0,000592
Casein kinase I isoform delta CSNK1D	3,24	2,96E-05	0,000997
SAFB-like transcription modulator SLTM	3,28	3,46E-05	0,007515
Zinc finger protein OZF ZNF146	7,95	3,53E-05	0,000183
ATP-dependent RNA helicase DDX50 DDX50	3,41	3,53E-05	0,00135
Heterogeneous nuclear ribonucleoprotein A3 HNRNPA3	7,52	3,68E-05	0,000371
MKI67 FHA domain-interacting nucleolar phosphoprotein NIFK	3,72	4E-05	0,004192
Ribosome production factor 2 homolog RPF2	4,14	4,5E-05	0,009136
U1 small nuclear ribonucleoprotein C SNRPC	4,35	4,72E-05	0,013233
Heterogeneous nuclear ribonucleoprotein A1 HNRNPA1	6,62	5,16E-05	0,000138
Probable RNA-binding protein 19 RBM19	4,1	5,41E-05	0,006029
THAP domain-containing protein 11 THAP11	3,3	5,54E-05	0,004141
Band 4.1-like protein 5 EPB41L5	4,02	5,69E-05	0,001217
Heterogeneous nuclear ribonucleoproteins A2/B1 HNRNPA2B1	7,55	5,8E-05	0,00027
Terminal uridylyltransferase 7 TUT7	3,56	6,64E-05	0,012671
Nucleolar MIF4G domain-containing protein 1 NOM1	3,42	6,99E-05	0,039197
Beta-1,3-galactosyltransferase 6 B3GALT6	3,12	7,07E-05	0,004886
Peptidyl-tRNA hydrolase 2, mitochondrial PTRH2	4,88	7,5E-05	0,000114
Transcriptional activator protein Pur-alpha PURA	4,72	9E-05	0,001456
H/ACA ribonucleoprotein complex subunit 2 NHP2	3,3	9,54E-05	0,020359
Polyadenylate-binding protein 2 PABPN1	3,2	0,000124	0,006059
Histone H1x H1FX	3,72	0,000149	0,044075
Non-histone chromosomal protein HMG-14 HMGN1	3,98	0,000154	7,28E-05
PIN2/TERF1-interacting telomerase inhibitor 1 PINX1	4,07	0,000159	0,012531
Cold-inducible RNA-binding protein CIRBP	3,9	0,000166	0,000114
Splicing factor, arginine/serine-rich 19 SCAF1	3,18	0,000176	0,020947
Surfeit locus protein 6 SURF6	3,61	0,000225	0,004242
Ras-responsive element-binding protein 1 RREB1	3	0,000386	0,000131
3'-5' exoribonuclease 1 ERI1	4,07	0,000411	0,011435

TRAF2 and NCK-interacting protein kinase TNIK	3,75	0,000537	0,012206
Kinesin-like protein KIFC1 KIFC1	3,12	0,000552	0,039342

7.2.- List of Figures

7.3.- List of Tables

7.4.- List of abbreviations

3´UTR: 3´untranslated region; **5´UTR:** 5´untranslated region; **ABC:** ATP-binding cassettes; **Allo-HSCT:** allogeneic haematopoietic stem cell transplant; **ACTB**: beta actin; **AKR1C3:** aldo-keto reductase family 1 member C3; **AML LSC**: acute myeloid leukaemia leukaemia stem cells; **AML:** acute myeloid leukaemia; **Ara-C/AraC**: cytosine arabinoside; SOs: antisense oligonucleotides; **ARHGAP22:** Rho GTPase-activating protein 22; **ASXL1**: Additional Sex Combs-Like 1; **ATP:** Adenosine triphosphate; **B-ALL:** B-cell acute lymphoblastic leukaemia; **BCL2:** b-cell lymphoma 2 protein; **BCR-ABL1**: breakpoint cluster region protein and Tyrosine-protein kinase ABL1; **BM:** bone marrow; **bp:** base pairs; **CBFβ:** core-binding factor subunit beta; **CD_:** cluster of differentiation; **CDKN1A:** cyclin-dependent kinase inhibitor 1A, **CDK6**: cyclin-dependent kinase 6; **CEBPA:** CCAAT/enhancer-binding protein alpha protein; **ceRNA:** competitive endogenous RNA; **CFC:** colony-forming cell; **CFU:** colony-forming unit; **CH:** clonal haematopoiesis; **CLIP:** cross-linking and immunoprecipitation; **CLL:** chronic lymphocytic leukemia; **CLP:** common lymphoid progenitor; **cm:** centimetre; **CMP:** common myeloid progenitor; **CN-AML**: cytogenetically normal acute myeloid leukaemia; **CPXM1:** carboxypeptidase X, M14 family member 1; **CR:** complete remission; **CRISPR-Cas:** clustered regularly interspaced short palindromic repeats; **CRNDE:** Colorectal Neoplasia Differentially Expressed; **CSC:** cancer stem cell; **DANCR:** Differentiation antagonizing nonprotein coding RNA; **ddH2O:** double- distilled water; **DEG:** Differential Gene Expression; **DMEM:** Dulbecco's Modified Eagle Medium; **DNA:** deoxyribonucleic acid; **DNR:** daunorubicin; **DNAJB9:** DnaJ homolog subfamily B member 9; **DNTM3B:** DNA Methyltransferase 3 Beta; **dNTPs:** Deoxynucleotide triphosphates; **DSMZ:** Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH; **ETV6:** translocation-Ets-leukemia virus protein 6; **ELN:** European LeukaemiaNet; **EIF4A:** Eukaryotic translation initiation factor 4 A; **EIF4G:** Eukaryotic translation initiation factor 4 G; **EMP1:** Epithelial membrane protein 1; **EEF2:** Elongation factor 2; **FAB:** French-American-British; **FACS:** fluorescence-activated cell sorting; **FAM30A:** Family With Sequence Similarity 30 Member A; **FBS:** fetal bovine serum; **FL:** follicular lymphoma; **FLT3:** fms like tyrosine kinase 3; **FLT3-ITD:** FLT3- internal tandem duplication; **gDNA:** genomic DNA; **GEPIA:** Gene Expression Profiling Interactive Analysis; **GFI1:** growth factor independent 1; **GM-CSF:** granulocyte-macrophage colony-stimulating factor; **GMP:** granulocyte-macrophage progenitor; **GO:** gemtuzumab-ozogamicin; **GPR56**: G protein-coupled receptor 56; **GSEA**: gene set enrichment analysis; **GTEX:** Genotype-Tissue Expression Program; **HITS-CLIP:** high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation; **HOXA9:** Homeobox protein Hox-A9; **HOXBLINC:** HOXB locus-associated long non-coding RNA; **HOTAIR:** HOX transcript antisense intergenic RNA; **HSC:** haematopoietic stem cell; **HSPC**: haematopoietic stem and progenitor cells; **HyperTRIBE:** Targets of RNA-binding proteins Identified By Editing; **IDH1/2:** Isocitrate dehydrogenase 1; **IKZF2:** IKAROS Family Zinc Finger 2; **IL3:** interleukin-3; **IRES:** internal ribosomal re-entry signal; **kb:** kilobases; **KH:** k Homology; **IP:** immunoprecipitation; **KCNQ1OT1:** KCNQ1 Opposite Strand/Antisense Transcript 1; **KRAS:** Kirsten rat sarcoma virus; **JUNB:** transcription factor jun-B; **LDAC:** low-dose cytarabine; **lncRNA**:

long non-coding RNA; **LOUP**: lncRNA originating from upstream regulatory element of SPI1; **LSC**: leukaemic stem cell; **LTC-IC:** long-term initiating cell; **LT-HSC:** long-term haematopoietic stem cell; **M-CSF:** macrophage colony-stimulating factor; **MALAT1:** metastasis associated lung adenocarcinoma transcript 1; **MCL1:** myeloid leukemia cell differentiation protein 1; **MDR:** multi-drug resistance; **MDS:** myelodysplastic syndrome; **MEP:** megakaryocyte-erythroid progenitor; **miRNA:** microRNA; **mL**: milliliter; **mg:** milligram; **MLL:** mixed-lineage leukaemia; **MLL-AF9**: Myeloid/lymphoid or mixed-lineage leukemia translocated to chromosome 3 protein; **MLV:** Murine Leukaemia Viruses; **MMRN1:** Multimerin-1; **MPN:** myeloproliferative neoplasm; **MPO:** myeloperoxidase; **MPP:** multipotent progenitor; **MPO:** myeloperoxidase; **mRNA:** messenger RNA; **MSI2:** Mushashi-2; **m6A:** N6-Methyladenosine, **ncRNA:** non-coding RNA; **NF- kB**: nuclear factor kappa-light-chain-enhancer of activated B-cells; **NEAT1:** Nuclear Enriched Abundant Transcript 1; **NK-cells:** natural killer cells; **NOD/SCID:** nonobese diabetic/severe combined immunodeficiency; **NOTCH:** neurogenic locus notch homolog protein 1; **NSG: NOD** scid gamma mouse; **nt**: nucleotides; **NUMB:** NUMB endocytic adaptor protein; **OXPHOS:** oxidative phosphorylation; **OS:** overall survival; **pAML:** paediatric AML; **PARP1:** Poly [ADP-ribose] polymerase 1; **PB:** peripheral blood; **PCR:** polymerase chain reaction; **PDX:** patient-derived xenograft; **PML-RARA:** translocation involving the promyelocytic leukemia gene and retinoic acid receptor alpha; **Pol II:** RNA polymerase II; **PPIA:** Peptidyl-prolyl cis-trans isomerase A; **pre-LSC:** pre-leukaemic stem cell; ; **PTBP1:** Polypyrimidine Tract Binding Protein 1; **PTM:** post-translational modification; **RAD21:** Double-strand-break repair protein rad21 homolog; **RT-qPCR:** quantitative real-time polymerase chain reaction; **RBD:** RNA-binding domain; **RBP:** RNA-binding protein; **RNA:** ribonucleic acid; **RNAi:** RNA interference; **RNP:** ribonucleoprotein; **ROS:** reactive oxygen species; **RPL7:** ribosomal protein like 7; rpm: revolutions per minute; **RPMI**: Roswell Park Memorial Institute media; **RRM:** RNA-recognition motif; **rRNA:** ribosomal RNA; **RT-qPCR:** quantitative reverse transcription PCR; **RUNX1:** Runt-related transcription factor 1; **RUNXOR**: RUNX1 overlapping RNA; **SDS:** sodium dodecyl sulfate; **SEM:** standard error of the mean; **SF3B1**: Splicing factor 3B subunit 1; **shRNA:** short-hairpin RNA; **siRNA:** small interfering RNA; **SMAD3:** Mothers against decapentaplegic homolog 3; **SMC1A:** Structural maintenance of chromosomes protein 1A; **SMIM25**: Small integral membrane protein 25; **SNHG5:** small nucleolar RNA host gene 5; **snRNA:** small nuclear RNA; **snoRNA:** small nucleolar RNA; **SOCS2:** Suppressor of cytokine signaling 2; **SRSF2:** Splicing factor, arginine/serine-rich 2; **ST-HSC**: short-term haematopoietic stem cell; **STAT3**: signal transducer and activator of transcription 3; **SYNCRIP:** Synaptotagmin Binding Cytoplasmic RNA Interacting Protein; **TAD:** transactivation domain; **TAL1:** T-cell acute lymphocytic leukaemia protein 1; **T-ALL:** T-cell acute lymphoblastic leukemia; **TCGA:** the cancer genome atlas; **TET2:** Tet methylcytosine dioxygenase 2; **TF:** transcription factor; **TGFβ:** Transforming growth factor beta; **TGFβR1:** Transforming growth factor beta receptor 1; **TM4SF1:** Transmembrane 4 L6 family member 1; **TNF:** tumour necrosis factor; **tRNA:** transfer RNA; **TP53:** tumour potein 53; **TSPN3:** tetraspanin3; **UKH:** Universitätsmedizin Halle; **UTR:** untranslated region; **UTX/MLL:**translocation of Ubiquitously transcribed tetratricopeptide repeat, X chromosome with mixed-leukaemia protein; **U2AF1**: Splicing factor

U2AF 35 kDa subunit; **VCL:** vinculin; **v/v**: volume per volume; **VRK1:** Serine/threonine-protein kinase VRK1; **WHO:** World health organization; **w/v:** weight per volume; **XIST:** X-inactive specific transcript; **ZBTB46:** Zinc finger and BTB domain-containing protein 46; **ZRSR2:** U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit-related protein 2; **µg:** microgram

7.4.- List of publications and presentations

Published articles:

Zorn, P. & **Calvo Sánchez, J.**, Alakhras, T., Schreier, B., Gekle, M., Hüttelmaier, S., & Köhn, M. (2024). Rbfox1 controls alternative splicing of focal adhesion genes in cardiac muscle cells. *Journal of Molecular Cell Biology*, 16(1), mjae003. https://doi.org/10.1093/jmcb/mjae003

Piersimoni, L., Abd El Malek, M., Bhatia, T., Bender, J., Brankatschk, C., **Calvo Sánchez, J.**, Dayhoff, G. W., Di Ianni, A., Figueroa Parra, J. O., Garcia-Martinez, D., Hesselbarth, J., Köppen, J., Lauth, L. M., Lippik, L., Machner, L., Sachan, S., Schmidt, L., Selle, R., Skalidis, I., Sorokin, O., … Uversky, V. N. (2022). Lighting up Nobel Prize-winning studies with protein intrinsic disorder. *Cellular and Molecular Life Sciences*, 79(8), 449. https://doi.org/10.1007/s00018-022-04468-y

Calvo Sánchez, J., & Köhn, M. (2021). Small but Mighty -The Emerging Role of snoRNAs in Hematological Malignancies. *Non-coding RNA*, 7(4), 68. https://doi.org/10.3390/ncrna7040068

Articles pending for submission:

Calvo Sánchez J, Sinz, A., Hüttelmaier, S., Vetrie, D., Köhn, M. Unravelling the molecular role of the long non-coding RNA FAM30A in regulating acute myeloid leukemia stem cell dynamics. (2024).

Ritchie, J., Cendejas Orozco, M., S Binshaya, A., Stevens T., **Calvo Sánchez, J.**, Nasvytis, P., Kyi Lai Yin Swe, K., Kritalug T., Yin Yuen C., Dunn K., M. Michie A., Helgason G. H., Scott M. T., Vetrie D. D2 class dopamine receptors regulate key survival pathways in CML stem cells. (2024).

7.5.- Erklärung

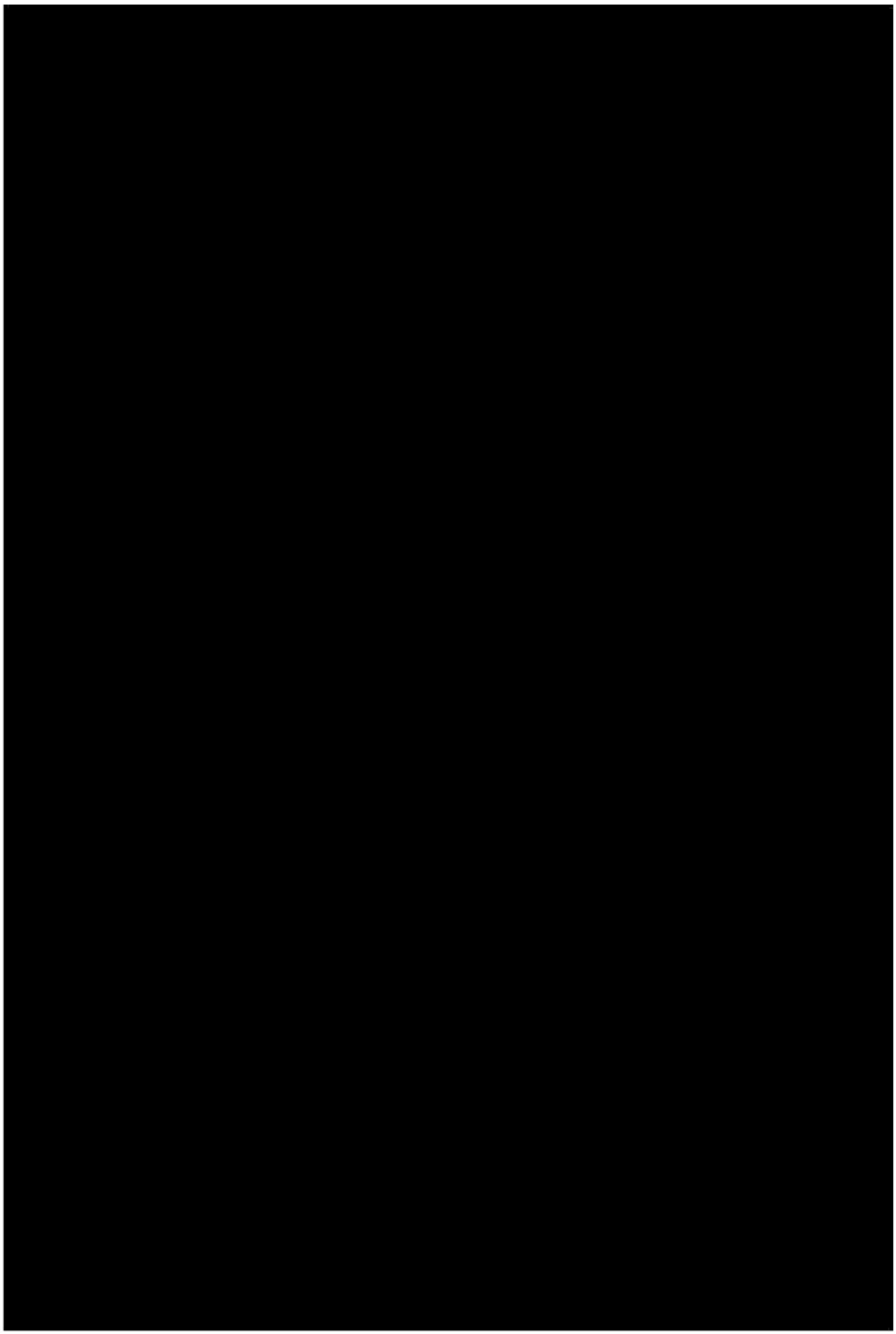
Hiermit versichere ich, dass ich meine Dissertationsschrift selbstständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen habe ich als solche kenntlich gemacht.

Mit der vorliegenden Arbeit bewerbe ich mich erstmals um die Erlangung des Doktorgrades.

Halle (Saale), 09.07.2024

Jaime Calvo Sánchez

7.6.- Acknowledgements



7.7.- Curriculum Vitae

Personal information

Name	**Jaime Calvo Sánchez**
Date and place of birth	[redacted]
Address	

Professional experience

10.2018 – 10.2024	**Research assistant** Doctoral thesis (see below) **Faculty of Medicine; Institute of Molecular Medicine** **Martin-Luther-Universität Halle-Wittenberg** **Germany**
03.2018 - 09.2018	**Research assistant** MSc project (see below) **Wolfson Wohl Cancer Research Centre** **University of Glasgow** **United Kingdom**
01.2017-07.2017	**Student internship** BSc Project (see below) Prof. Dr. Yulia Sidorova **Institute of Biotechnology – University of Helsinki** **Helsinki** **Finland**
2014, 2015	**Student internship** Project: investigation into heat shock proteins. Laboratory Macromolecular Structure Prof. Dr. Jose M Valpuesta)

National Centre of Biotechnology (CNB-CSIC)

Madrid

Spain

Education

10.2018 - **Doctoral studies (Dr. rer. nat.)**

Molecular cell biology, RNA biology, acute myeloid leukaemia

Doctoral thesis: Unravelling the molecular role of the long non-coding RNA FAM30A in modulating acute myeloid leukaemia dynamics

Faculty of Medicine; Institute of Molecular Medicine

Martin-Luther-Universität Halle-Wittenberg

Germany

09.2017 - 09.2018 **Master of Science (MSc)**

Molecular oncology; translational research

Master thesis: Modulation of dopaminergic signalling components in Jurl-MK1 cells: a mechanistic model to eradicate chronic myeloid leukaemia stem cells.

Wolfson Wohl Cancer Research Centre

University of Glasgow

United Kingdom

10.2010 - 09.2013 **Bachelor of Science (BSc)**

Biotechnology (Human Health Sciences)

Bachelor thesis: Selection and *in vitro* characterization of GDNF-related compounds as potential agents agains Parkinson Disease. (at University of Helsinki)

Universidad Politécnica de Madrid (UPM)

Spain

09.1998 - 06.2013 **Primary, secondary and high school**

Colegio Garcilaso de la Vega, Madrid, Spain.

Colegio San Agustín, Madrid, Spain.

Relevant courses and skills

10.2022	Leadership in the Scientific Environment
03.2022	FELASA B Certificate – Laboratory Animal Training
08.2020	Good Manufacturing Practice (GMP) basic course
04.2018	Introduction to Statistics and R programming