UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

DIPARTIMENTO DI MEDICINA E CHIRURGIA

DOTTORATO IN EMATOLOGIA SPERIMENTALE
XXVIII CICLO

CHARACTERIZATION OF NEW ONCOGENES IDENTIFIED THROUGH NGS-BASED ANALYSIS OF LEUKEMIAS: SETBP1 AND ETS2-ERG

Dr. Marco Peronaci

Coordinatore: *Prof. Carlo Gambacorti Passerini*

Tutor: *Dr.ssa Sara Redaelli*
Dr. Rocco Piazza

Anno Accademico 2015-2016

To my parents

Abstract

The improvements in sequencing technology led to the development of 'Next Generation Sequencing' (NGS) technologies. Several NGS approaches exist. Whole genome sequencing (WGS) and whole exome sequencing (WES) allow the identification of genomic alterations such as small insertions/deletions, point mutations and structural variants. Whole transcriptome sequencing (RNA-Seq) permits to quantify gene expression profiles and to detect alternative splicing and fusion transcripts. Recently, by using WES on atypical chronic myeloid leukemia (aCML) samples, our group identified recurrent mutations in SETBP1 gene [1]; also, by using RNA-Seq on acute myeloid leukemia (AML), we identified a new fusion gene: ETS2-ERG [2].

The aim of this work is to characterize both SETBP1 and ETS2-ERG to gain further insight about their function and role in the pathogenesis of aCML and APL, respectively.

In aCML, SETBP1 mutations disrupt a degron binding site, leading to decreased protein degradation. This leads to an increased amount of SETBP1 protein interacting with its natural ligand SET, which in turn acts inhibiting the protein phosphatase 2A (PP2A) oncosuppressor. Interestingly, the SETBP1 mutational cluster affected in aCML is highly conserved and the same mutations were also observed in the Schinzel-Giedion syndrome (SGS). However, the inhibition of the PP2A by SET, the only known interactor of SETBP1, does not explain the phenotype of SGS. To further characterize the role of SETBP1 protein, 293 Flp-In isogenic cellular models expressing the empty vector or the wild type (WT) or mutated (G870S) form of SETBP1 were established. In these models, SETBP1 was fused with a V5 tag. Chromatin immunoprecipitation sequencing experiments (ChIP-Seq) performed against V5 confirmed the binding of SETBP1 to DNA, both for the WT and G870S forms. In addition, RNA-Seq experiments were performed. The comparison between ChIP-Seq and RNA-Seq data has allowed us to identify 130 genes presenting both the binding of SETBP1 to their promoter region and transcriptional upregulation. Together these data suggest a role for SETBP1 as a transcriptional activator. Co-immunoprecipitation (Co-IP) experiments in transiently transfected 293T cells coupled with mass spectrometry (MS) analysis were performed to identify potential interactors of SETBP1. MS analysis led to the identification of the host cell factor 1 (HCF1), a component of the SET1/KMT2A COMPASS-like complex. Independent validation by western blot and fluorescence resonance energy transfer (FRET) confirmed the direct binding of HCF1 to SETBP1. Further independent experiments

confirmed the Co-IP of SET1/KMT2A and PHF8 with SETBP1. SET1/KMT2A is a core component of COMPASS-like complex and possesses H3K4 methyltransferase activity, whereas PHF8 possesses H4K20 demethylase activity. Both marks are associated with actively transcribed genes. Taken together, we have shown that SETBP1 protein is able to act as a transcriptional activator recruiting the HCF1/KMT2A/PHF8 complex.

In a previous study, comparing cytogenetic analysis and RNA-Seq to detect chromosomal abnormalities on AML patient samples, a new fusion between the ETS2 and ERG genes was reported. The patient carrying this fusion was affected by acute promyelocytic leukemia (APL) and did not respond to therapy with retinoic acid. The role of the ETS2-ERG fusion is not known. To gain insight about the functional role of ETS2-ERG fusion in APL two cellular models were established. HL-60 cells were stable transfected with retroviral empty vector or with a vector carrying the fusion gene. This vector also carries the GFP as a positive selection marker. HL-60 cells carrying the ETS2-ERG fusion treated with retinoic acid showed a decrease in the expression at membrane level of the differentiation marker CD11b. This suggests that the ETS2-ERG fusion is able to impair the differentiation of APL cells upon retinoic acid treatment.

Table of content

List of figures

List of tables

Introduction

1.1 Next-Generation Sequencing

The ‘chain-terminator’ DNA sequencing method was described by Frederick Sanger in 1977 and it is commonly known as Sanger sequencing [3]. A decade later, this technique was improved, leading to the automation of the sequencing process. The automated Sanger method dominated the field of genome sequencing for almost two decades and it became the gold-standard in genomic research, achieving important accomplishments, such as the completion of the Human Genome Project (HGP) that resulted in the sequencing of the first human genome [4, 5]. After years of improvements, this ‘first-generation’ technique allows to achieve a read-length of ~1000 bp as well as a per-base ‘raw’ accuracy of 99.999% [6]. Due to intrinsic limitations of Sanger sequencing, such as the high operating costs, the limited throughput and the time consuming protocols [7], researchers demanded for faster and cheaper technologies. This led to the development of newer ‘second-generation’ sequencing methods, commonly known as Next Generation Sequencing (NGS). However, even if the acronym NGS is commonly used, the term ‘second-generation’ sequencing is more correct given that ‘third-generation’ sequencing technologies are already commercially available. A general overview of the past and present technologies, as well as those that are actually under development, is provided in Table 1.

Manufacturer	Amplification	Detection	Chemistry	URL
Commercial				
Illumina	Clonal	Optical	Sequencing by synthesis	**http://www.illumina.com**
Oxford Nanopore	Single molecule	Nanopore	Nanopore	**http://www.nanoporetech.com**
Pacific Biosciences	Single molecule	Optical	Sequencing by synthesis	**http://www.pacb.com**
ThermoFisher Ion Torrent	Clonal	Solid state	Sequencing by synthesis	**http://www.thermofisher.com/us/en/home/brands/ion-torrent.html**
Precommercial				
Quantum Biosystems	Single molecule	Nanogate	Nanogate	**http://www.quantumbiosystems.com**
Base4	Single molecule	Optical	Pyrophosphorolysis	**http://base4.co.uk**
GenapSys (GENIUS)	Clonal	Solid state	Sequencing by synthesis	**http://www.genapsys.com**
QIAGEN (GeneReader)	Clonal	Optical	Sequencing by synthesis	**http://www.qiagen.com**
Roche Genia	Single molecule	Solid state	Nanopore	**http://geniachip.com**
Postcommercial				
Roche 454 (GS FLX)	Clonal	Optical	Sequencing by synthesis	**http://www.454.com**
Helicos BioSciences (Heliscope)	Single molecule	Optical	Sequencing by synthesis	—
Dover (Polonator)	Clonal	Optical	Sequencing by ligation	—
ThermoFisher Applied Biosystems (SOLiD)	Clonal	Optical	Sequencing by ligation	**http://www.thermofisher.com/us/en/home/brands/applied-biosystems.html**
Complete Genomics	Clonal	Optical	Sequencing by ligation	**http://www.completegenomics.com**

Table 1 Summary of NGS platforms.

The platforms listed as precommercial have been announced, but not formally launched. The platforms listed as postcommercial are no longer sold as new instruments by their company. Dashes indicate that no website is available. Adapted from Levy et al., 2016 [8].

In general, all the different sequencing platforms can be divided along three axes. The first axis is represented by the detection of single-molecule or clonally amplified DNA. The second one is constituted by the detection system used, which can be based on optical or non-optical methods. The third axis is the use of indirect sequencing technologies, such as the sequencing-by-synthesis (SBS) and sequencing-by-ligation (SBL) methods, or the direct measurement of the DNA molecules [8]. Even if built upon different technologies, all of these NGS platforms are based on a combination of template preparation, sequencing and imaging, and genome and assembly

methods [9]. For the purpose of this thesis, only a general overview of the Illumina technology will be discussed.

In 2006, Solexa (California, USA), acquired by Illumina in 2007, introduced in the market the Genome Analyzer (GA) sequencer that was able to generate 1 gigabase (Gb) of data per run. In comparison, the current Illumina NGS instruments range from the MiniSeq® System to the HiSeq X series that are able to produce from 1.8-7.5 Gb to 16-18 terabases (Tb) per run, respectively [10]. Illumina platforms are based on the concept of SBS [11] and their general workflow is composed of four basic steps: library preparation, cluster generation, sequencing and data analysis.

Library preparation

In this step, the DNA or cDNA is randomly broken in smaller fragments by different techniques, such as sonication or transposons-based methods. Fragment sizes usually range from 100 to 800 bp. This random fragmentation is followed by 5' and 3' repair and adapter ligation. After that, a short PCR reaction, usually limited to 6-10 cycles of amplification, is performed in order to increase the amount of the starting material. Recently, the concept of 'multiplexing' has emerged. In addition to adapters, a unique identifier, or index, is ligated to each DNA fragment that compose a library. In this way, different libraries are tagged with different indexes and can be pulled and sequenced together. This indexing system has dramatically reduced both the time and the cost of sequencing.

Cluster generation

In this step, the library is loaded into a flow-cell, a planar surface similar to a microscope glass slide that contains immobilized oligonucleotides complementary to the adapters [12]. The single-strand DNA (ssDNA) template is hybridized to the flow-cell and amplified by solid-state bridge amplification. In this particular amplification system, the ssDNA binds to two different immobilized oligonucleotides, forming an arch structure. Once this structure is formed, the amplification takes place, resulting in a double strand DNA (dsDNA) structure. The dsDNA is denatured, forming two ssDNA. At this point, the entire process restarts, with the free end of ssDNA that interacts with nearby primers to form arch structures. During the multiple amplification cycles, all the amplicons remain attached to the immobilized primers so that each original ssDNA molecule gives rise to different clusters consisting of ~1000 clonal amplicons (Figure 1).

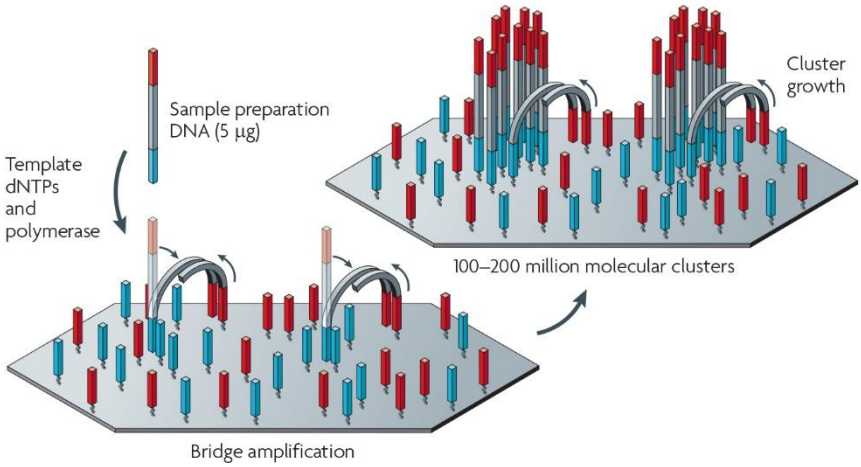


Figure 1 Illumina/Solexa amplification technology.

The template molecules bind to the immobilized oligonucleotides on the flow-cell (priming) and a single-stranded, single-molecule template is generated. The subsequent step is the bridge amplification of the immobilized templates with the adjacent primers in order to form clusters. Adapted from Metzker *et al.*, 2010 [9].

Sequencing

The Illumina sequencing technology is based on a proprietary reversible terminator-based method. It uses four nucleotides modified in two different ways: i) they possess a cleavable terminator at the 3' hydroxyl position; ii) each of the four nucleotides is equipped with one of four different cleavable fluorescent labels. The sequencing process starts with the hybridization of a sequencing primer complementary to the adapter region of ssDNA. Next, a cyclic process, composed of nucleotide incorporation, fluorescence imaging and cleavage, takes place [13]. The polymerase incorporates one labeled nucleotide. After that, the laser light excites fluorescent labels and the emission of each cluster is detected in four channels. Lastly, both the fluorescent label and terminator moiety are cleaved and the entire process restarts with the addiction of another nucleotide (Figure 2).

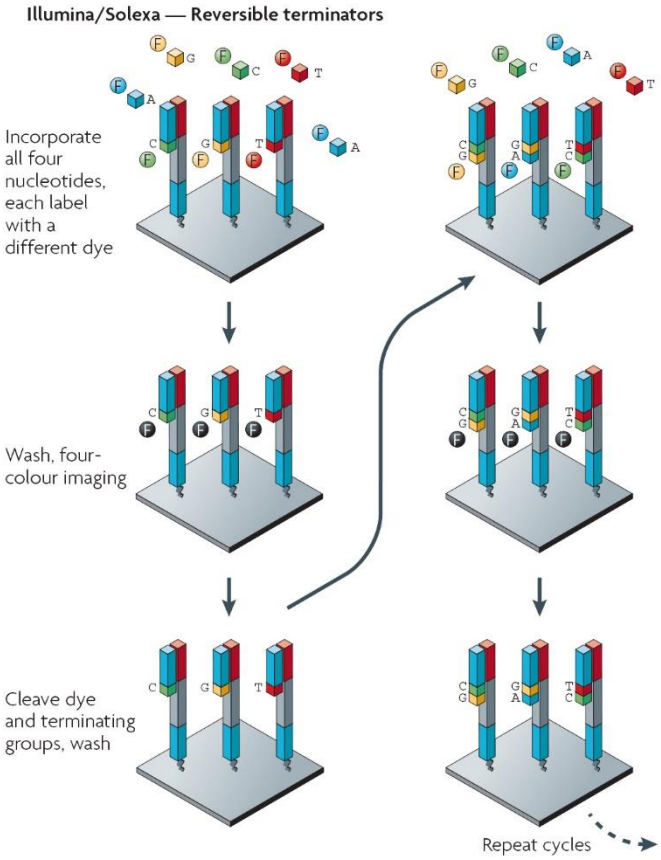


Figure 2 Illumina/Solexa sequencing strategy.

The sequencing-by-synthesis Illumina technology is based on a proprietary reversible terminator chemistry. Fluorescently labeled, 3' blocked nucleotides are incorporated in the template clusters (shown as single molecules for illustrative purposes). After an image acquisition step, the fluorescent labels are removed and the 3'-OH of the blocked nucleotides is regenerated, allowing the incorporation of another reversible terminator in a cyclic process. Adapted from Metzker *et al.*, 2010 [9].

The result is the generation of one 'read' for each cluster sequenced. Illumina technology allows the sequencing of each run as 'single read' or 'paired-end'. The single read method involves the sequencing of only one end of the DNA fragments, whereas the paired-end system permits the sequencing from both ends of the fragments. The use of paired-end reads permits a better alignment to the reference genome and, in addition, allows the detection of indels, which otherwise is not possible with single reads data [14].

Data analysis

NGS experiments generate an enormous amount of data. This represent a big challenge for data management, storage and analysis. The first step in NGS data analysis is to convert the

acquired images containing fluorescent signals into reads, a process termed ‘base calling’ for which different algorithms were developed during the years [15-17]. During base calling these algorithms evaluate different parameters, such as fluorescence intensity, background noise and non-specific signals in order to assign a quality score (Q-score) to each nucleotide identified. The Q-score is an index that represents the probability of an error in base calling [18]. The raw Illumina data output consists in per-cycle BCL base call binary files [19]. This binary BCL format is incompatible with the major part of the open source analysis software. Therefore, BCL files are converted in the universally accepted FASTQ format. FASTQ are text-based files that store the sequence reads and the quality score of each base, both encoded with a single ASCII character. Each entry of the FASTQ file represents a read and it is composed of four elements (Figure 3).

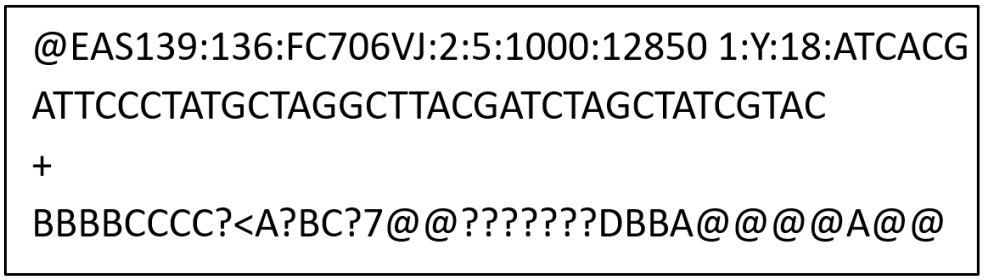


Figure 3 Example of FASTQ entry (read).

Each FASTQ entry is represented by four elements. From the top of the image to the bottom: i) element 1, sequence identifier; ii) element 2, sequence; iii) element 3, spacer and optional repeat of sequence identifier; iiii) element 4, quality line.

The next step is usually the alignment of the reads to a reference genome, except for *de novo* sequencing, in which a reference sequence is not available. For this purpose, several mapping software were developed during the years. Some of them, such as Bowtie [20], Bowtie 2 [21] and BWA [22], fall in the category of the genome aligners, since they are not able to manage splicing events. As a result, these tools are not able to map reads spanning splice junctions on the reference genome. On the contrary, other aligner tools, such as TopHat [23], TopHat2 [24] and STAR [25], are able to manage reads containing splice junctions and, for this reason, they fall in the category of the ‘splicing-aware’ aligners [26]. The data output of these aligners is usually presented in the Sequence Alignment/Map (SAM) format, which contains the reads mapped to the reference genome. The SAM format is a TAB-delimited text-based file, introduced by Heng Li [27], composed of an optional header section, containing information regarding the sequencing

platform, and an alignment section, at least composed of 11 mandatory field [28]. The human-readable SAM format usually is converted into the Binary Alignment/Map (BAM) format. As the name suggests, the BAM file is a binary compressed format that reduces the huge SAM file size providing, at the same time, a faster access to the data. Following the alignment step, the BAM file containing the mapped reads is further processed and used for different analysis such as variant detection, gene expression profiling, indels identification and new transcripts discovery.

1.2 NGS applications

Next-generation sequencing has revolutionized the field of both medical and biomedical research. This technology allows researchers to explore genomic, transcriptomic and epigenetic landscapes in a high-throughput approach impossible before. In addition, it can help clinicians in the therapeutic choice, contributing to the so-called 'precision medicine'. Precision medicine is a term to identify a new approach in the treatment of medical disorders that takes in account the variability in individual genetic, environmental factors and life-style [29]. The most common applications include whole-genome sequencing (WGS), whole-exome sequencing (WES), RNA sequencing (RNA-Seq) and chromatin immunoprecipitation-sequencing (ChIP-Seq).

1.2.1 WGS

WGS consists in the sequencing of an entire genome. Historically, the classical example of WGS approach is the sequencing of the first human genome by the HGP group [4, 5]. This project was based on the traditional Sanger sequencing technology and it was completed in more than ten years. WGS provides a comprehensive view of an individual's genome or cancer genome. Indeed, with a single assay it is possible to obtain a complete map of single nucleotide variations (SNVs), indels, copy number variations (CNVs) and complex structural rearrangements [30]. In the last years, the cost of sequencing has dropped dramatically, breaking the amount of US$1,000 per human WGS [31]. However, this cost represents only the one necessary for the sequence production. The huge computational resources and the big data storage systems required increase the cost of WGS to a level that is actually not compatible for routine genetic and biological studies [32].

1.2.2 WES

The whole-exome sequencing is the most widespread and used NGS application and it is based on the capture and sequencing of the entire gene-coding portion (exome) of the genome. The exome represents roughly the 1% of human genome [33], but it contains approximately 85% of disease-causing mutations [34]. The sequencing of the protein-coding regions of the genome represents a cost-effective alternative to WGS. Indeed, in comparison to WGS, if we assume a fixed-sequencing cost for the NGS study, the use of WES allows to increase the number of samples sequenced in a single run, resulting in an increase of the breadth of the study and in an enhanced depth of coverage [35]. Moreover, WES generates a lower amount of data than WGS, resulting in a faster data analysis. The initial capturing and enrichment steps of the exonic gDNA are two of the most critical and, for this purpose, several commercial kits are available [36]. However, none of these kits cover completely all the exons listed in the Consensus Coding Sequencing project (CCDS) [37], RefSeq [38] or Ensembl [39] databases [32].

1.2.3 RNA-Seq

Other than the genome, NGS technologies allow the sequencing of the entire transcriptome. The transcriptome is the complete set of RNAs contained in a cell, from protein-coding messenger RNA (mRNA) to non-coding RNA such as transfer RNA (tRNA), ribosomal (rRNA) and other non-coding RNA (ncRNA) [40]. First of all, by sequencing the entire set of RNAs, RNA-Seq gives an unbiased view of the transcriptome profile and allows for the quantification of mRNA. The quantification of the mRNA can provide a measure of the gene expression. In comparison to the previously hybridization-based technologies such as the microarrays, RNA-Seq is not limited neither in the number of hybridization probes nor in species specific probes. In addition to gene expression profiles, by mapping RNA-Seq reads onto the genome it is possible to detect new transcripts as well as mRNA isoforms derived by alternative splicing. Moreover, it allows the detection of new fusion genes, therefore being an important tool both in cancer research and in cancer diagnostic, due to its increased sensibility in comparison to traditional cytogenetic analysis [41].

1.2.4 ChIP-Seq

The ChIP-Seq approach merges the standard ChIP technique with the NGS, allowing the sequencing of the nucleic acids bound to the immunoprecipitated protein of interest. It is a useful

technique in order to study DNA-binding proteins, histone modifications and nucleosome, leading to a better understanding of both gene regulation and epigenetic mechanisms [42].

1.3 Hematopoiesis

Hematopoiesis is a highly regulated process that results in the formation of all the blood components. It is widely accepted that it is a hierarchical process in which, at the top, resides a pluripotent hematopoietic stem cell (HSC) that possesses both self-renewing and differentiation capabilities. However, the HSCs are not the first cells to emerge in the hematopoietic system. The establishment and specification of a definitive HSCs population is preceded, during the embryogenesis, by different waves of hematopoiesis. These waves are defined as 'primitive' and 'definitive' hematopoiesis and they appear in different anatomical location during the embryo development [43]. The primitive wave originates in the blood island of yolk sack [44]. During this first wave, primitive erythroblasts (EryP-CFC) differentiate from mesoderm [45]. These nucleated EryP-CFC go through different maturation steps directly in the blood stream, in a semi-synchronous manner, leading to the formation of primitive erythroid cells [46]. Moreover, megakaryocyte (Meg-CFC) and macrophage (Mac-CFC) precursors emerge in concomitance with EryP-CFC in the yolk sack [47, 48]. The definitive hematopoiesis emerges immediately after these events and can be further divided into two waves [49]. During the first, erythromyeloid progenitors (EMPs) arise from the hemogenic endothelium. This latter is composed of a small subset of endothelial cells able to change their fate towards hematopoietic progenitors [50]. EMPs still emerge in the yolk sack, but then they enter into the blood stream and migrate to the fetal liver, where they expand and produce definitive enucleated erythrocytes and also granulocytes [51, 52]. Subsequently, in the second wave of the definitive hematopoiesis, HSCs emerge from the aorta-gonad-mesonephros (AGM) region of the growing embryo. Then, HSCs migrate to the fetal hematopoietic organs such as the liver and the spleen, where they proliferate increasing in number [53, 54]. Finally, just before the birth, HSCs migrate to the bone marrow (BM) where they remain throughout adulthood [55]. Therefore, BM is the major site of hematopoiesis during the adult life.

The hematopoietic system is divided in two different branches: lymphoid and myeloid. Both of these compartments originate from HSCs in a process that involves the loss of multipotency in favor of differentiation [56]. Based on their ability to self-renew HSCs are subdivided in long-term HSCs (LT-HSC), short-term HSCs (ST-HSCs) and multipotent progenitors (MPPs). LT-HSCs give rise to ST-HSCs and, in irradiated hosts, are able to sustain the repopulation of BM in the long-term [57]. In contrast, ST-HSCs can only transiently support the production of myeloid and lymphoid cells [58, 59]. MPPs originate through ST-HSCs

differentiation and retain the ability to differentiate in all the blood cells types, but they lack the capacity of self-renewal [60]. In mice, all these cells are identified by surface markers as the Lin$^-$ c-kit^{hi} Sca-1$^+$ (LSK or KLS) fraction. Moreover, HSCs and MPPs can also be distinguished by surface markers expression and are defined as Lin$^-$ c-kit^{hi} Sca-1$^+$ Flt-3$^-$ or Lin$^-$ c-kit^{hi} Sca-1$^+$ Flt-3$^+$, respectively [59, 61, 62]. In the standard model of hematopoiesis (Figure 4), the distinction between the myeloid and lymphoid lineages shows up very early during the differentiation process [63].

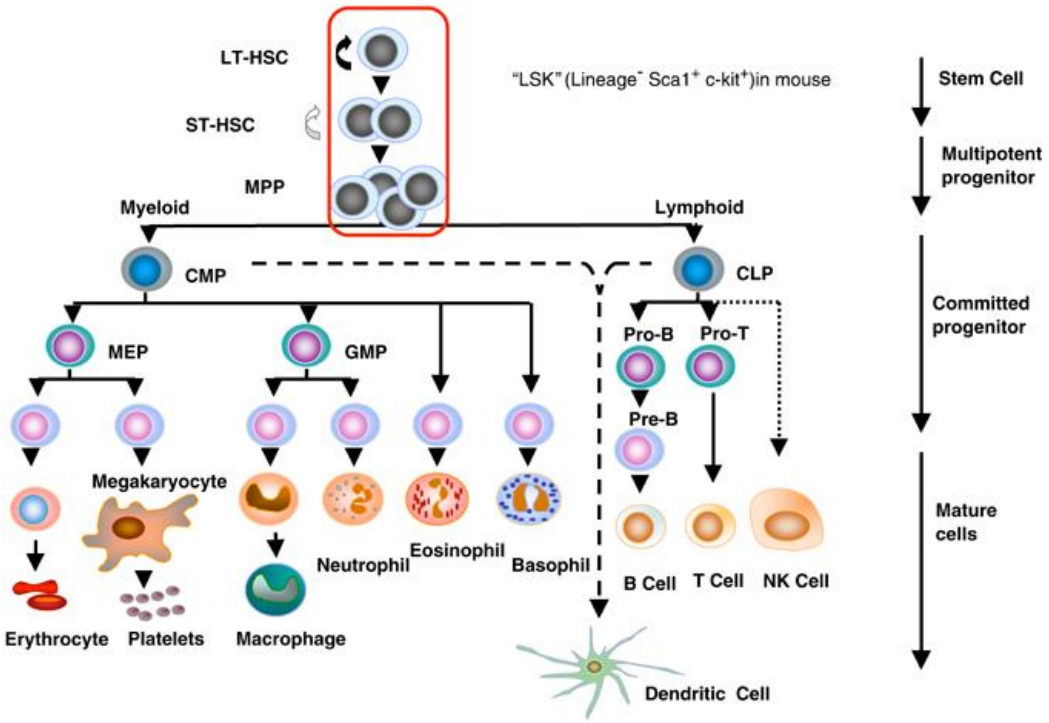


Figure 4 Schematic representation of the classical model of hematopoiesis.

LT-HSC, long-term repopulating HSC; ST-HSC, short-term repopulating HSC; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte-macrophage progenitor. Adapted from Larsson *et al.*, 2005 [64].

Indeed, two different oligopotent progenitors differentiate from MPPs, the common myeloid progenitor (CMP) [65] and the common lympoid progenitor (CLP) [66]. Myeloid-committed CMPs give rise to megakaryocyte-erythrocyte progenitors (MEPs) and granulocyte-macrophage progenitors (GMPs). Lineage-restricted erythroid progenitors (EryPs) and

megakaryocytic precursors (MkPs) originate downstream to MEPs, leading to terminally differentiated erythrocytes and megakaryocytes/platelets, respectively. GMPs can generate neutrophils, eosinophils, basophils, and mast cells as well [67]. On the other side, CLPs give rise to pro-B and pro-T lymphoid precursors, which in turn differentiate into mature B and T cells, respectively, as well as natural killer (NK) lineage cells [62]. The origin of dendritic cells (DCs) is still controversial. Although DCs originate mainly from CMPs, under certain conditions they can originate also from lymphoid progenitors [68, 69]. It is important to remember that this model, although still valid, is rather simplified.

1.4 Leukemia

The term leukemia was used for the first time in 1847 by Rudolf Ludwig Karl Virchow (1821-1902) [70] and it derives from the Greek words *λευκός* (*leukós*, "white") and *αἷμα* (*haîma*, "blood").

Leukemias are liquid tumors characterized by the clonal proliferation of white blood cells. Depending on the cell lineage involved, leukemia can be classified as "myeloid" or "lymphoid". Additionally, leukemia can also be categorized as "acute" or "chronic". Acute leukemias are characterized by immature cells, termed blasts, that quickly proliferate. Due to the rapid proliferation and the undifferentiated state of leukemic cells, acute leukemias progress rapidly and are generally fatal within few months, if not treated. On the contrary, chronic leukemias are usually characterized by a low percentage of blasts and more mature cells that retain some of their normal functions, resulting in a slower progression and in a disease that can last years.

1.5 The 2016 WHO classification of myeloid neoplasm and acute myeloid leukemia

The World Health Organization has combined for the first time the genetic features of myeloid malignancies with their morphology, cytochemistry and immunophenotype as well as clinical and biological features. The first WHO classification of myeloid neoplasm was done in 2001 and later revised in 2008. The actual 2016 WHO classification of myeloid neoplasm includes 5 major entities: the acute myeloid leukemia (AML), the myeloid/lymphoid neoplasm with eosinophilia and rearrangement of *PDGFRA*, *PDGFRB*, or *FGFR1*, or with *PCM1-JAK2*, the myelodysplastic syndromes (MDS), the myeloproliferative neoplasm (MPN) and the overlapping category of the MDS/MPN. This thesis work is focused on the atypical chronic myeloid leukemia (aCML), BCR-ABL$^-$ and on the acute promyelocytic leukemia with PML-RARA (APL), which belong to the MDS/MPN and AML entities, respectively.

1.6 Myeloproliferative neoplasms

Myeloproliferative neoplasms (MPNs) are clonal disorders characterized by an increased proliferation of the myeloid cells that maintain all the maturation steps [71].

In 1951, William Dameshek classified chronic myeloproliferative neoplasms for the first time [72]. In his work, he described all the myeloproliferative disorders (MPD), such as chronic myeloid leukemia (CML), polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF). Moreover, the erythroleukemia, which nowadays does not belong to MPN, was described. Dameshek discovered that all these diseases were able to evolve towards a blastic phase, typical of acute leukemia [73]. The abnormal myeloproliferation is caused by clonal rearrangements or mutations in genes encoding for tyrosine kinase proteins, resulting in a constitutively active signal transduction. In some cases, such as the *BCR-ABL1* fusion gene in CML, the genetic abnormality is used as a diagnostic feature. In other cases, genetic abnormalities, such as JAK2 and KIT mutations, are not specific of a single MPN, but only of a clonal proliferation [74]. For example, the JAK2 V617F is the most common mutation in the BCR-ABL$^-$ MPN syndromes. This mutation is present in more than 90% of patients affected by PV, but also in patients with PMF or ET. So, JAK2 V617F mutation is not specific for a single MPN [75].
MPN diseases are listed below in Table 2

Myeloproliferative neoplasms (MPN)
Chronic myeloid leukemia (CML), *BCR-ABL1*$^+$
Chronic neutrophilic leukemia (CNL)
Polycythemia vera (PV)
Primary myelofibrosis (PMF)
PMF, prefibrotic/early stage
PMF, overt fibrotic stage
Essential thrombocythemia (ET)
Chronic eosinophilic leukemia, not otherwise specified (NOS)
MPN, unclassifiable
Mastocytosis

Table 2 The 2016 WHO classification of myeloproliferative neoplasm.
Adapted from Arber *et al*., 2016.[76]

1.7 Myelodysplastic syndromes

Myelodysplastic syndromes (MDS) are clonal disorders of the hematopoietic cell precursors characterized by cytopenia, dysplasia of one or more myeloid lineages, inefficient hemopoiesis, increased apoptosis and evolution towards AML. A blast percentage lower than 20% in bone marrow (BM) and peripheral blood (PB) is the parameter used for the diagnosis of MDS. Above this threshold, the disease is classified as AML.

Peripheral cytopenia is an essential feature of MDS. However, the WHO classification is based on the degree of dysplasia and blast percentages. Consequently, in the 2016 WHO revision terms such as 'refractory anemia' and 'refractory cytopenia' have been replaced with MDS followed by the correct identifiers: single vs multi-lineage dysplasia, ring sideroblast, excess blast or del(5q) cytogenetic abnormality. The most common mutated genes in MDS are *SF3B1*, *TET2*, *SRSF2*, *ASXL1*, *DNMT3A*, *RUNX1*, *U2AF1*, *TP53*, and *EZH2* [76].
The MDS are listed in the table below

Myelodysplastic syndromes (MDS)
MDS with single lineage dysplasia
MDS with ring sideroblasts (MDS-RS)
MDS-RS and single lineage dysplasia
MDS-RS and multilineage dysplasia
MDS with multilineage dysplasia
MDS with excess blasts
MDS with isolated del(5q)
MDS, unclassifiable
Provisional entity: Refractory cytopenia of childhood
Myeloid neoplasms with germ line predisposition

Table 3 The 2016 WHO of myelodysplastic syndromes.
Adapted from Arber *et al*., 2016 [76].

1.8 Myelodysplastic/myeloproliferative neoplasms

Myelodysplastic/myeloproliferative neoplasms (MDS/MPN) are a group of rare clonal myeloid neoplasias, characterized, at the onset, by the presence of both myelodysplastic and myeloproliferative features [77, 78].

Nevertheless, differently from MDS, in MDS/MPN diseases there is no lack of leukocytes and monocytes in the peripheral blood.

To the MDS/MPN group belong the chronic myelomonocytic leukemia (CMML), the juvenile myelomonocytic leukemia (JMML), atypical chronic myeloid leukemia (aCML), and MDS/MPN-unclassifiable. MDS/MPN have been recently extensively characterized in their molecular alterations. The *SRFSF2*, *TET2* and/or *ASXL1* genes are mutated in more than 80% of CMML cases. Moreover, *ASXL1* has been incorporated as a poor prognostic factor for CMML together with the classical diagnostic parameters. Other mutations occurring at lower frequencies affect the *RUNX1*, *CBL*, *EZH2*, *SETBP1*, *NRAS/KRAS* genes [76].

1.8.1 Atypical chronic myeloid leukemia (aCML)

Atypical chronic myeloid leukemia is a hematological clonal disorder included in the MDS/MPN group according to the 2016 WHO classification. Historically, it was described as a myeloid neoplasm similar to chronic myeloid leukemia (CML), but lacking the pathognomonic Philadelphia chromosome.

The Philadelphia chromosome is the typical genetic alteration of CML, caused by the reciprocal translocation involving the *ABL1* gene, on the chromosome 9, and the *BCR* gene, on chromosome 22. The resulting fusion gene *BCR-ABL* encodes for a chimeric protein BCR-ABL in which the tyrosine kinase activity of ABL is deregulated. As a consequence, multiple pathways are altered, such as the MAPK, the PI3K-Akt and the protein kinase (PK)C/PKD signaling, leading to increased cell survival and proliferation.

The diagnostic criteria (Table 1) published in the 2016 WHO defines the aCML as a pathology characterized by leukocytosis with the presence of immature myeloid precursor in the peripheral blood (≥10% of leukocytes), dysgranulopoiesis, low or absent basophils (<2% of leukocytes) and monocytes (<10% of leukocytes), less than 20% of blasts in the blood and BM, absence of *PDGFRA*, *PDGFRB*, or *FGFR1* rearrangement, or *PCM1-JAK*. Moreover, aCML diagnosis must not meet the WHO criteria for BCR^{+}-$ABL1^{+}$ CML, PMF, PV or ET.

aCML diagnostic criteria
• PB leukocytosis due to increased numbers of neutrophils and their precursors (promyelocytes, myelocytes, metamyelocytes) comprising ≥10% of leukocytes)
• Dysgranulopoiesis, which may include abnormal chromatin clumping
• No or minimal absolute basophilia; basophils usually <2% of leukocytes
• No or minimal absolute monocytosis; monocytes <10% of leukocytes
• Hypercellular BM with granulocytic proliferation and granulocytic dysplasia, with or without dysplasia in the erythroid and megakaryocytic lineages
• <20% blasts in the blood and BM
• No evidence of *PDGFRA*, *PDGFRB*, or *FGFR1* rearrangement, or *PCM1-JAK2*
• Not meeting WHO criteria for *BCR-ABL1*$^+$ CML, PMF, PV, or ET*

Table 4 Diagnostic criteria for aCML according to 2016 WHO classification.
Adapted from Arber *et al.*, 2016 [76].

aCML is considered a rare neoplasm with an incidence estimated between 1-2% of BCR-ABL$^+$ leukemias. In contrast to CML, which has become a highly treatable cancer after the introduction of the tyrosine kinase inhibitor (TKI) Imatinib [79], aCML presents an overall survival of 12.4 months [80]. Moreover, the 40% of patients affected by this hematological malignancy progresses to AML [77].

1.8.1.1 Molecular aspects

The molecular aspects of aCML are not well defined and our knowledge regarding the molecular events leading to this rare neoplasm is still poor. However, in the last years, many efforts were made in order to identify specific genomic lesions for this pathology. In 2013, Piazza *et al.* identified SETBP1 as recurrent mutated in aCML in about 25% of patients [1]. This was the first evidence of a recurrently somatic lesion in aCML that was also confirmed by other independent studies [81]. In a group of chronic neutrophilic leukemia (CNL) and aCML patients, Maxson *et al.* identified recurrent mutations in the *CSF3R* gene in approximately 40% of patients [82]. However, different subsequent studies reported this mutation with lower frequency in aCML [83]. Recently, our group identified recurrent mutation also in the *ETNK1* gene in 8.8% of aCML patients. [84] Additionally, recurrent somatic mutations in *NRAS*, *KRAS*, *JAK2*, *EZH2*, *TET2*, *IDH2* and *CBL* can be detected, although with lower frequency [1, 81, 85, 86].

1.8.1.2 SETBP1

The SET binding protein 1 (SETBP1), previously known as SEB, is located on the chromosome 18q21.1 and encodes for a large protein of 1,596 amino acids with a predicted molecular weight of 170 KDa. Even if its structure is unknown it is reported to possess three AT-hooks (amino acids 584-596; 1,016-1,028; 1,451-1,463), a SKI-homologous region (amino acids 706-917), so called for its similarity with the oncogene SKI, a SET binding domain (amino acids 1,292-1,488) and a repeat domain (amino acids 1,520-1,543)(Figure 5) [1, 87].

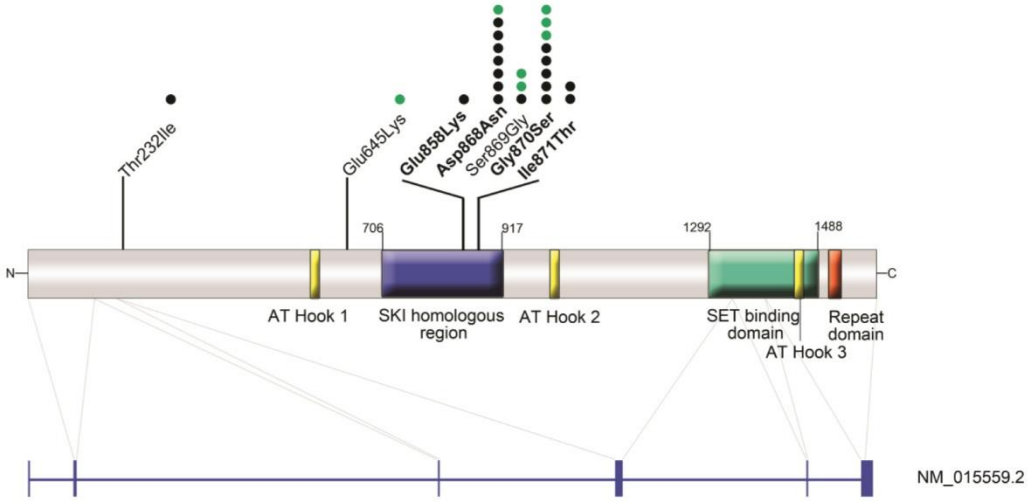


Figure 5 Schematic view of SETBP1 structure and its associated mutations.

SETBP1 numbering is referred to the NCBI reference sequence NM_015559.2. The isoform A of SETBP1 (1,592 amino acid) is encoded by five exons (blue bars). The SETBP1 sequence contains three AT hook domains (amino acids 584-596; 1,016-1,028; 1,451-1,463), a SKI homologous region (amino acids 706-917), a SET-binding domain (amino acids 1,292-1,488) and a repeat domain (amino acids 1,520-1,543). The black circles represent alterations found in aCML whereas the green circles represent alterations found in other diseases. Somatic variants are indicated in bold. Adapted from Piazza *et al.*, 2013 [1].

SETBP1 has a predominantly nuclear localization and its only known interactor is SET [87], which in turn is described as a potent inhibitor of the tumor suppressor protein phosphatase 2A (PP2A) [88]. In a single patient affected by T-cell acute lymphoblastic leukemia (T-ALL) it is reported an in frame fusion consisting in the exon 12 of nucleoporin 98 (*NUP98*) fused to the exon 5 of *SETBP1*, however the function of this fusion protein is still unknown [89]. In some cases of acute myeloid leukemia (AML) *SETBP1* is overexpressed because of the translocation t(12;18)(p13;q12) involving the ETS variant 6 (*ETV6*) gene. In the same work, Cristobal *et al.*

described for the first time a novel leukemogenic mechanism by which the overexpression of SETBP1 was able to protect SET from protease cleavage, leading to the formation of a SETBP1-SET-PP2A complex resulting in PP2A inhibition and subsequently proliferation of leukemic cells (Figure 6) [90].

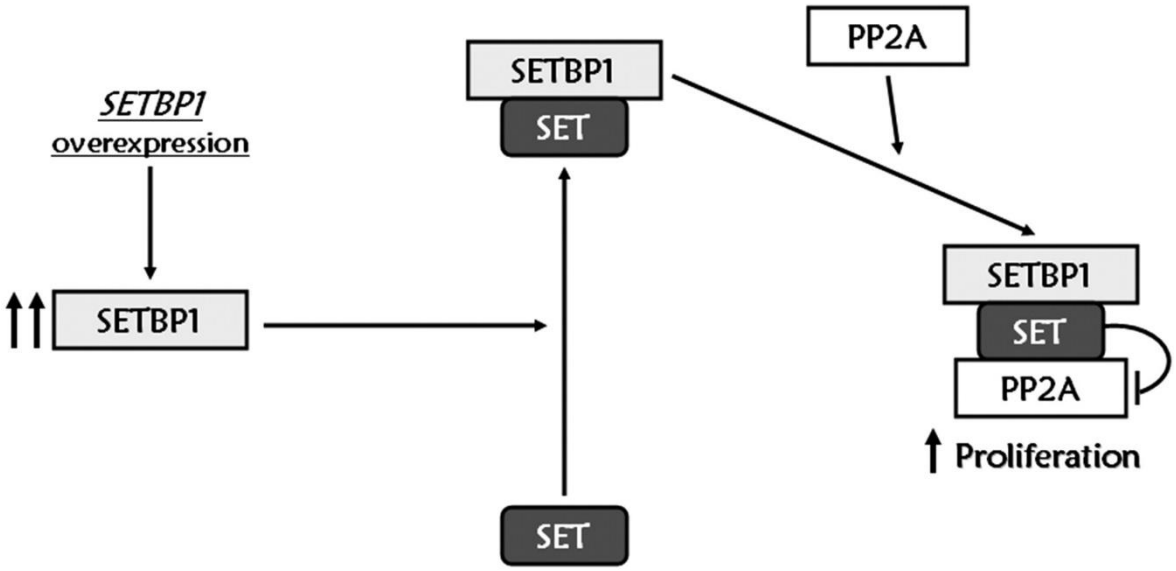


Figure 6 SETBP1-mediated PP2A inhibition.

In AML patients, SETBP1 overexpression protects SET from protease cleavage, leading to the formation of a SETBP1-SET-PP2A complex. SET exerts its function by inhibiting the tumor suppressor PP2A. Adapted from Cristóbal *et al.*, 2010 [90].

In 2013, Piazza *et al.* identified SETBP1 as recurrently mutated in aCML. Subsequent studies confirmed the presence of SETBP1 mutations in other hematological malignancies, such as CMML and JMML [91, 92]. The major part of the mutations are located in the SKI homologous region, between the amino acids 868 and 871 (Figure 5) [1]. The authors demonstrated that these mutations are located within a degron, a substrate region that contains a binding site for the E3 ubiquitin ligase beta-TrCP. The latter is a protein involved in the processes of protein degradation. The disruption of this degron causes that SETBP1 is no longer bound by the beta-TrCP, resulting in increased level of SETBP1 due to protein accumulation and the consequent increase of stabilized SET, which in turn inhibits PP2A.

Interestingly, SETBP1 mutations cluster in a highly conserved region and the same set of alterations was previously observed as germline mutations in the Schinzel-Giedion syndrome (SGS) [93]. SGS is a rare pathology characterized by severe mental retardation, distinctive facial features, multiple congenital malformations such as skeletal abnormalities, genitourinary and renal malformations and cardiac defects, and an increased risk to develop tumors [94, 95]. However, it is not known how these mutations are involved in the SGS phenotype. SETBP1 has been reported as involved in myeloid leukemia development. Indeed, SETBP1 overexpression confers self-renewal capability to myeloid progenitors *in vitro* by interacting with the homeobox A9 (HOXA9) and homeobox A10 (HOXA10) promoters [96]. Moreover, increased mRNA levels of SETBP1 were also detected in chronic myeloid leukemia (CML) patients that present concomitant high levels of HOXA9 and HOXA10. In another study, Vishwakarma *et al.* demonstrated that SETBP1 is able to interact with the runt related transcription factor 1 (Runx1) promoter, resulting in Runx1 downregulation. [97] RUNX1 is a well known negative regulator of the HSC self-renewal and its loss is known to be associated with a partial block of myeloid differentiation [98, 99].

However, a complete overview regarding the physiological role of SETBP1 is still lacking, so further investigations are required to better comprehend the complexity of its functions and consequently the effects of its mutations.

1.9 Acute myeloid leukemia (AML)

Acute myeloid leukemia is a neoplasia of the hematopoietic tissue that originates from the clonal transformation of a hematopoietic cell through the acquisition of chromosomal rearrangements and multiple genetic mutations [100]. This hematological malignancy is characterized by uncontrolled proliferation of leukemic cells, named blasts, which accumulate into the bone marrow. This leads to a compromised production of erythrocytes, granulocytes and platelets, causing anemia, thrombocytopenia and neutropenia. In AML, somatic mutations occur either at the level of multipotent cells or even in more committed progenitors, causing both proliferation and differentiation alterations [101]. AML is the most frequent leukemia in the adults, with an incidence of 3-4/100.000 cases.

1.9.1 Classification

AML can be divided into two main groups: i) primary or *de novo* AML, which occurs in patients that have never been exposed to risk factors; ii) secondary AML, which occurs in patients previously exposed to cancer agents or affected by other hemopathies [102].

Since the neoplastic transformation can occur at different stages during the differentiation process of HSC, different AML subtypes have been identified. The first system developed to classify AML malignancies is the French-American-British (FAB) classification. By using morphological, cytochemical and immunophenotypical features FAB recognizes 8 groups of AML based on the cell in which the mutation occurs and on the differentiation stage (Figure 7, Figure 8)

FAB classification of acute myeloblastic leukaemia

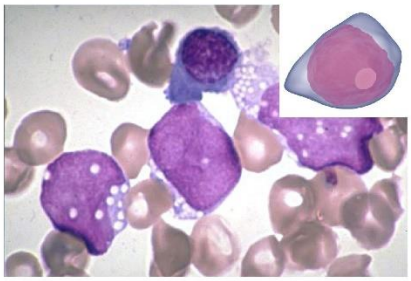


Photo courtesy of: Acute myeloid leukemia pathophysiology, 2012

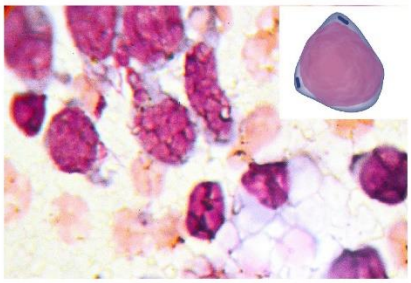
M0 Acute myeloblastic leukaemia with minimal differentiation

Morphology:

Can resemble LLA-L2 blasts. Medium-sized blasts, rounded nucleus, fine chromatin, basophilic non-granular cytoplasm, prominent nucleoli.

Immunophenotype

- CD13 +
- CD33 +
- CD11b +
- CD11c +
- CD14 +
- CD15 +



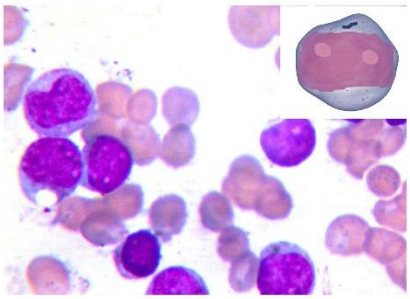
M1 Acute myeloblastic leukaemia without maturation

Morphology:

Medium-sized blasts with high nucleo:cytoplasm (n:c) ratio, rounded nuclei with immature, dispersed chromatin with one or more prominent nucleoli. Blasts can show fine azurophilic granulation or isolated Auer rods in the cytoplasm in 5% to 10% of cases

Immunophenotype

- MPO +
- CD13 +
- CD33 +
- CD117+
- CD34 +/-



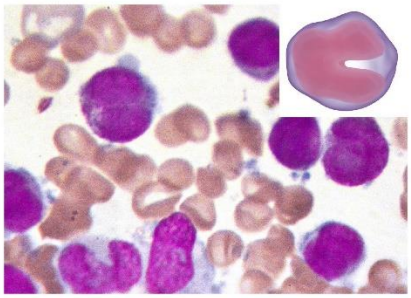
M2 Acute myeloblastic leukaemia with maturation

Morphology:

Small to medium-sized blasts with high nucleo: cytoplasm (n:c) ratio and rounded nuclei sometimes located in a corner of the cytoplasm. The nucleus shows dispersed, immature chromatin with one or more nucleoli. The cytoplasm is basophilic and can contain traces of primary azurophilic granulation or isolated Auer rods.

Immunophenotype

- MPO +
- CD34 +/-
- CD13 +
- CD15 +
- HLA-DR +/-
- Sudan black +
- CD117 +/-



M3 Promyelocytic leukaemia

Morphology:

Abundant, intensely azurophilic granulation. The nucleus is usually monocytic in appearance (reniform) and is either irregular or bilobed with a deep cleft. Scarcely basophilic cytoplasm due to the proliferation of azurophilic granulation. Some atypical promyelocytes also contain elongated or splinter-shaped crystalline cytoplasmic inclusions specific to this type of leukaemia.These usually form clumps, but differ from Auer rods in that they show a tubular substructure on electronic microscopy.

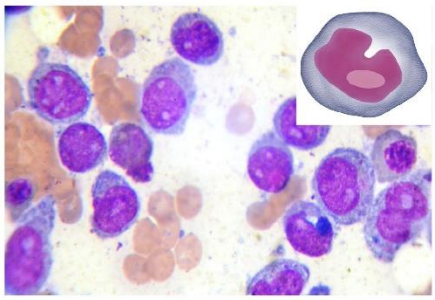
Immunophenotype

- CD13 +
- CD33 +
- HLA-DR -
- CD34 -

Figure 7 FAB classification of acute myeloid leukemia.

Adapted from Ladines-Castro *et al.*, *2016* [103].

FAB classification of acute myeloblastic leukaemia



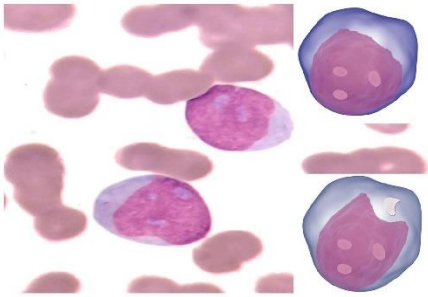
M4 Acute myelomonocytic leukaemia

Morphology:

Large blasts, moderate nucleo:cytoplasm (n:c) ratio and variable basophilia. The nucleus may be rounded, kidney-shaped or irregular. Nucleoli are usually prominent.

Immunophenotype

- CD13 +
- CD15 +
- CD33 +
- CD11b +
- CD11c +
- CD14 +
- CD64 +
- CD4 +

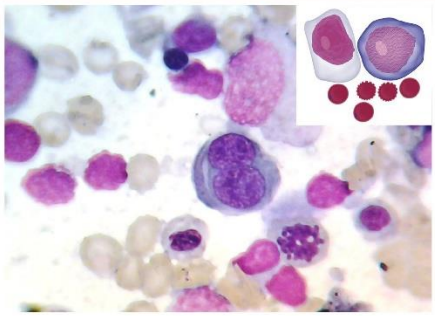


M5 Acute monocytic leukaemia

M5a acute monoblastic leukaemia:
Large blasts with rounded nucleus and dispersed, immature chromatin (1-3 nucleoli) and moderately large and intensely basophilic cytoplasm. The cytoplasm may show some Auer rods and/or prolongations and granulations.
M5b acute monocytic leukaemia
Promonocytes have a rounded or kidney-shaped nucleus with a less basophilic cytoplasm that is more highly granulated than monoblasts and contains some vacuoles. A findings of erythrophagocytosis together with monocytic blasts suggests a t(8;16) translocation.

Immunophenotype

- CD14 +
- CD68 +
- CD4 +
- CD11c +
- HLA-DR +
- CD64 +

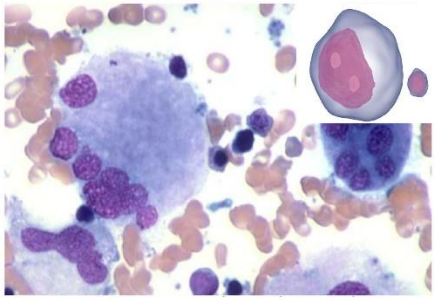


M6 Acute erythroid leukaemia

M6a erythroid leukaemia with proliferation of mixed blasts:
Over 50% erythroid precursors and around 30% myeloblasts. Morphology of erythrocytes in peripheral blood is greatly changed, with schistocytes, "pincered" or mushroom-shaped cells, and spiculated echinocyte and acanthocyte cells.
M6b pure erythroid leukaemia:
Erythroids make up 80% of bone marrow cells, with less than 3% myeloid cells. Erythrocytes in peripheral blood consist of macrocytes, basophilic stippling, Howell-Jolly bodies or Cabot rings.

Immunophenotype

- CD13 +
- CD33 +
- CD15 +
- Glycophorin A +
- Glycophorin C +



M7 Acute megakaryocytic leukaemia

Morphology:

Highly immature, polymorphic blasts. The nucleus is eccentric with dispersed, reticulated chromatin and 1-3 prominent nucleoli. The cytoplasm is non-granular, basophilic, and very similar in appearance to platelets, with pseudopods or granulations. Micromegakaryocytes and fragments of megakarioblasts are seen in peripheral blood (giant platelets, some highly degranulated).

Immunophenotype

- CD41 +
- CD61 +
- CD42 +
- CD13 +
- CD33 +
- CD34 +

Figure 8 FAB classification of acute myeloid leukemia (continued).

Adapted from Ladines-Castro *et al.*, *2016* [103].

In particular, AML are divided in:

- M0 or undifferentiated AML. It constitutes about 3% of the total AML cases and the blasts are not differentiated. Phenotypically, the M0-AML expresses early myeloid differentiation antigens as well as lymphoid markers [104].
- M1 or myeloblastic AML with minimal differentiation. It represents about 15-20% of AML. Phenotypically it is similar to M0-AML.
- M2 or myeloblastic AML with differentiation. It is about 25-30% of cases. The blasts carry maturation features. This group is associated with a translocation that involves chromosome 8 and chromosome 21 t(8;21)(q21;q22). This translocation results in acute myeloid leukemia 1 (AML1)/RUNX1 and eight-twenty one (ETO)/MTG8 fusion, leading to the formation of AML1-ETO fusion protein [105]. Phenotypically, blasts carry myeloid antigens, such as CD13, CD 15, CD33, CD34 and CD117.
- M3 or acute promyelocytic leukemia (APL): it is about 10-15% of cases. It is characterized by a hypergranulated cell proliferation. These cells are blocked at the promyelocytic stage. In more than 90% of cases, the diagnosis is based on the presence of the chromosomal translocation t(15;17)(q22;q21). This aberration causes the formation of the *PML-RARA* fusion gene, involving the promyelocytic leukemia gene (*PML*) and retinoid acid receptor alpha (*RARA*) [106]. APL cells express high level of CD13, CD33, CD64, but also CD34.
- M4 or myelomonocytic AML, with granulocytic and monocytic maturation. The incidence is about 20%. It is characterized by some granulocytic cells such as from promyelocytic to granulocytes, at different stage of differentiation, and monocytes such as from monoblasts to monocytes. These cells carry CD11b , CD11c, CD33, CD64, CD87 e CD116 antigens.
- M5 or monoblastic (M5a) and monocytic (M5b). It is about 10% of AML cases, it is characterized by cells (monoblasts, promyelocites, monocytes) at various differentiation stages. Phenotypically, these cells express CD11b, CD11c, CD33, CD64, CD87 and CD116 markers. In particular, the most immature form (M5a) is negative for CD14, while M5b is positive.
- M6 or erythrocytic AML, or acute erythroleukemia, represents about 5% of cases. It presents the coexistence of myeloid blasts and abnormal erythroblasts into the bone marrow. The abnormal expansion of the erythroid compartment is detected by the presence of CD45-, myeloperoxidase (MPO) negative as well as myeloid antigens negative cells, that express instead CD71 and CD105 antigens. Blasts are similar to ones belonging to the other subtypes.

- M7 or megacaryoblastic AML: it is a rare form, prevalently found in down syndrome patients. The cells are similar to immature megakaryocytes and express CD41, CD61, CD42a and CD42b markers.

The FAB classification is still commonly used to classify AML; however, it does not take in account many of the genetic factors that are known to affect the prognosis. As previously described, the WHO classification implemented the previous criteria by adding cytogenetic and molecular features. In the 2016 WHO classification there are different categories: i) AML with recurrent genetic abnormalities; ii) AML with myelodysplasia-related changes; iii) therapy-related myeloid neoplasm; iiii) AML not otherwise specified. A complete overview of the 2016 WHO for acute myeloid leukemia is in Table 5.

Acute myeloid leukemia (AML) and related neoplasms
AML with recurrent genetic abnormalities
AML with t(8;21)(q22;q22.1);*RUNX1-RUNX1T1*
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);*CBFB-MYH11*
APL with *PML-RARA*
AML with t(9;11)(p21.3;q23.3);*MLLT3-KMT2A*
AML with t(6;9)(p23;q34.1);*DEK-NUP214*
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2, MECOM*
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);*RBM15-MKL1*
Provisional entity: AML with BCR-ABL1
AML with mutated *NPM1*
AML with biallelic mutations of *CEBPA*
Provisional entity: AML with mutated RUNX1
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, NOS
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/monocytic leukemia
Pure erythroid leukemia
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Transient abnormal myelopoiesis (TAM)
Myeloid leukemia associated with Down syndrome

Table 5 The 2016 WHO revision of acute myeloid leukemia.

Adapted from Arber *et al., 2016* [76].

1.9.2 Acute promyelocytic leukemia (APL)

Acute promyelocytic leukemia, classified as the M3-AML subtype in the FAB classification and as APL with PML-RARA in the 2016 WHO classification, is characterized by unique morphologic, cytogenetic and molecular features [107, 108]. Immunophenotipically, APL cells express high levels of CD13, CD33 and CD19, and very low levels of CD7, CD11b and CD14 [109, 110]. Cytogenetically, APL is characterized by a balanced reciprocal translocation involving the chromosomes 15 and 17, t(15;17)(q22;q21). This rearrangement results in a PML-RARA fusion gene.

PML belongs to a family of proteins implicated in the tumor suppressor function and genome stability control. In particular, PML controls the induction of p53-mediated apoptosis, acting as a tumor suppressor gene. Moreover, it is involved in the formation of the PML oncogenic domains (POD) subnuclear stuctures, also named nuclear domain 10 (ND10), which are involved in tumor suppression, transcriptional regulation, apoptosis and protein degradation [111-113].

RARA is a member of the retinoic acid (RA) nuclear receptor family. In the absence of all-trans retinoic acid (ATRA), RARA binds to the retinoic acid responsive elements (RARE) of target genes as a heterodimer in association with retinoic X receptor (RXR), resulting in the recruitment of corepressor molecules, inhibiting gene transcription. Under physiological conditions, ATRA binding to RARA results in a conformational change of RARA that leads to corepressors release and transcriptional coactivators recruitment. The PML-RARA fusion acts with a dominant mechanism on RARA, forming homodimers, which in turn recruit corepressors. PML-RARA fusion is insensitive to physiological concentration of ATRA and, once bound to RARE, acts as a constitutive transcriptional repressor. PML-RARA and its corepressors are able to inhibit the function of downstream genes implicated in differentiation, leading to a block of differentiation at the stage of promyelocytes (Figure 9). In addition, PML-RARA protein is able to inhibit the PML function that, as previously described, acts as a tumor suppressor gene. Therefore, the fusion gene acts as dominant negative effect against both PML and RARA.

In the treatment of APL, pharmacological concentrations of ATRA are able to induce a conformational change of the multifunctional complexes linked to PML-RARA, causing the specific corepressors release. Consequently, PML-RARA normal function is restored, leading to a physiological hematopoietic differentiation (Figure 9) [114]. However, ATRA therapy can lead to side effects, such as in the case of the retinoic acid syndrome [115], characterized by unexplained fever, weight gain, dyspnea with pulmonary infiltrates, pleuropericardial effusion, hypotension, and renal failure, and also to frequent secondary ATRA resistances in a fraction of patients [106, 116]. In the last ten years, the new therapeutic agent arsenic trioxide (ATO) has been

introduced in APL patients treatment. ATO induces PML-RARA degradation, acting on the PML pathway and causing the degradation of PML-RARA oncoprotein [117].

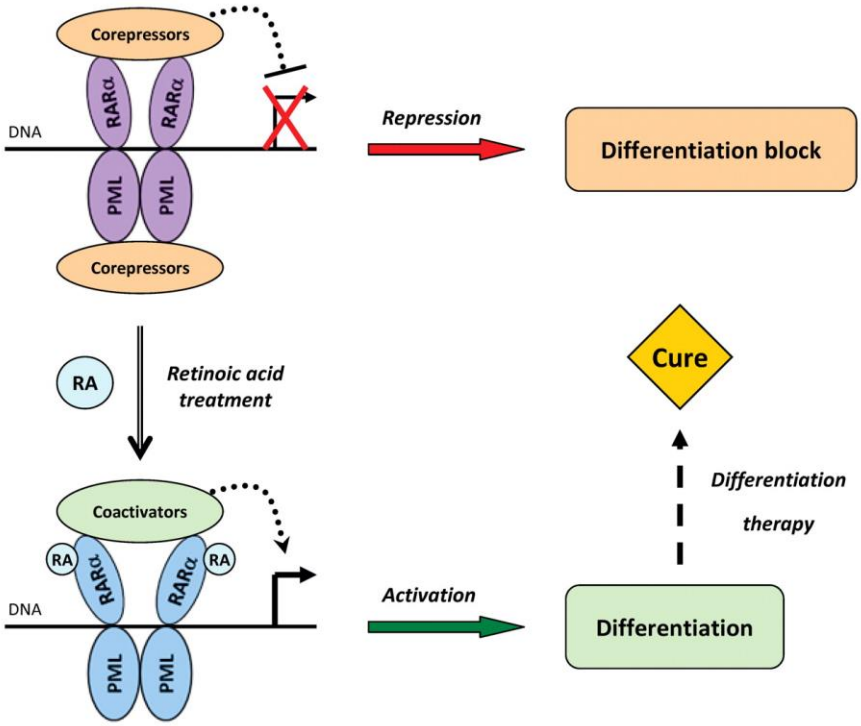


Figure 9 Model of PML-RARA transcriptional repression.

PML-RARA forms homodimers through PML coiled-coils domains and binds to RARE elements on DNA through RARA, recruiting corepressor molecules and leading to a differentiation block. Pharmacological concentrations of ATRA cause PML-RARA to switch to an active conformation. Coactivators are recruited, resulting in the transcriptional activation of genes involved in differentiation processes and ultimately leading to cure.

The presence of this fusion gene is extremely important for the therapy. Indeed, all-trans retinoic acid (ATRA) and/or arsenic trioxide (ATO) treatments are effective only on PML/RARA [118-120].

1.9.2.1 ETS2-ERG fusion

In 2013, Piazza *et al.* identified by RNA-Seq analysis in an APL patient (AML002) the new fusion gene *ETS2-ERG* [121].

The patient affected by the AML-M3 combined with the *ETS2-ERG* fusion received standard ATRA treatment and chemotherapy. This treatment leads to complete remission in approximately 80-90% of cases [122].However, after an initial therapy response, the patient relapsed and did not respond to any further treatment (ATO-ATRA combination, autologous bone marrow transplantation and allogenic bone marrow transplantation). The patient died 28 months after the diagnosis.

The ETS2-ERG fusion (Figure 10) derives from the cryptic in-frame translocation t(21;21)(q22;q22), which involves the *ETS2* and *ERG* genes.

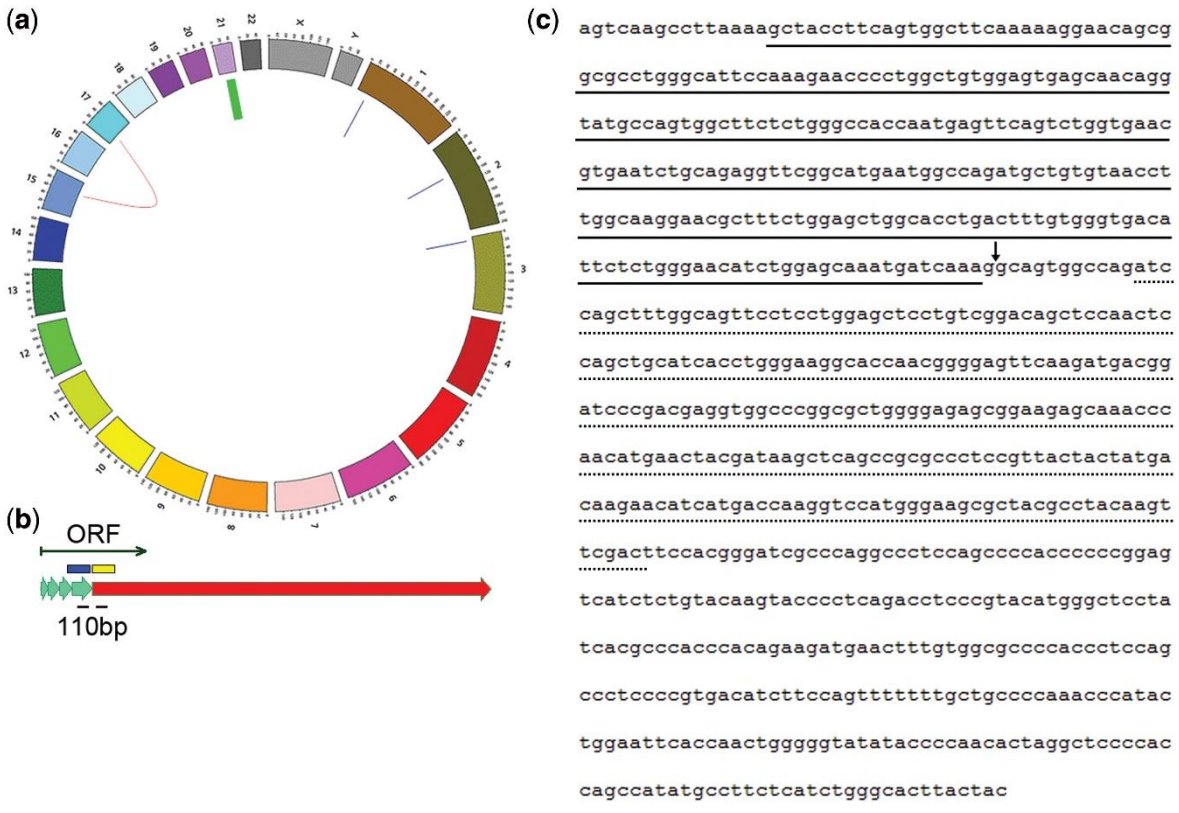


Figure 10 ETS2-ERG fusion in patient AML002.

A) Inside the circle, the red arch represents the *PML-RARA* translocation and the thick green bar represents the intrachromosomal *ETS2-ERG* fusion. B) Schematic representation of the ETS2-ERG fusion. Green arrows indicates the *ETS2* exons while the red arrow corresponds to the *ERG* exon. The thin black arrow indicates the opening reading frame (ORF). The blue and yellow boxes represent the ETS2 pointed domain (PNT) and the ETS domain of ERG, respectively. C) The thick black line marks the sequence corresponding

to the PNT domain of ETS while the dotted line indicates the ETS domain of ERG. The small black arrow indicates the breakpoint region. Adapted from Piazza *et al., 2012* [121].

As described above *ETS2* and *ERG* are located on chromosome 21. They both belong to the ETS family of transcription factors (TFs). This family comprises at least 28 members in humans [123], characterized by an evolutionary conserved 80-amino acid DNA-binding domain, named ETS domain, containing 4 tryptophan repeats, that binds the DNA sequence 5' - GGA(A/T) - 3' [124]. This class of TFs can act as positive or negative regulator of gene expression and its members are expressed ubiquitously or in tissue specific patterns [125]. Together, they regulate a multitude of processes such as development, proliferation, differentiation, apoptosis, migration, invasion and angiogenesis in various normal cells types as well as in cancer cells [126-128].

ETS2 is a gene encoding a transcription factor that controls different genes involved in apoptosis process and cell growth. It is characterized by an N-terminal pointed (PNT) domain, responsible for protein-protein interaction and a C-terminal ETS domain, responsible for DNA-binding, as previously described. It is a proto-oncogene involved in telomerase regulation. Indeed, it is known that is involved in breast cancer cell proliferation, interacting with both the human telomerase reverse transcriptase (*hTERT*) and *c-MYC* genes [129]. Gene expression alteration involving *ETS2* are also found in Down syndrome pathogenesis [130]. *ETS2* overexpression is reported in megakaryocytic AML. It regulates hematopoietic lineage and transcription factors genes involved in erythropoiesis and megakaryopoiesis. Moreover, it is responsive to some chemotherapy agents, such as daunorubicin and cytosine arabinoside (ara-C) [131]. Recently, it has also been described that ATO is able to upregulate *ETS2* expression [132].

ERG is an oncogene encoding for a TF and it is physiologically expressed in HSC and in the megakaryoblastic lineage. As *ETS2,* it contains both a PNT and ETS domains. In erytrholeukemic cell lines, *ERG* expression starts just after the megakaryocytic differentiation, and its expression leads to a phenotypical switch from the erythroid to megakaryocytic lineage [133]. It is known to be involved in the definitive hematopoiesis and in the physiological maintenance of HSC [134]. Moreover, its overexpression in AML is associated with a poor prognosis [135]. It is also expressed during the initial stages of T- and B-cells differentiation [136]. *ERG* participates in different chromosomal translocation, leading to the formation of oncogenic fusion proteins in different type of tumors, such as in the Ewing's sarcoma (EWS-ERG), in prostate cancer (TMPRRS2-ERG) and AML (FUS-ERG) [137]. The broad spectrum of action of *ERG* results in

a strong oncogenic activity upon its alteration and its contribution in cell differentiation, development and neoplastic transformation is still under study.

Aim

In the last decade, NGS technologies led to increasing discoveries in the field of cancer genomic. Our research group applied, in the last years, several NGS techniques to different hematological malignancies. By means of whole exome sequencing, we found *SETBP1* gene as recurrent mutated in a subset of aCML patients [1], whereas through RNA-Sequencing we identified the new fusion gene *ETS2-ERG* in acute promyelocytic leukemia [121]. Overall, the focus of my research was to better characterize the biological role of these two alterations.

The first part of this work was aimed to gain further insight the role and function of SETBP1 mutations. In aCML, SETBP1 mutations disrupt a degron binding site for the beta-TrCP, a key element of the proteasome machinery, leading to protein accumulation. After that, mutated SETBP1 acts as its physiological counterpart, stabilizing SET, which in turn inhibits the tumor suppressor PP2A. Interestingly, the same mutations were identified in the Schinzel-Giedion syndrome, a non-neoplastic pathology that presents a higher incidence to develop tumors [93]. It is reported that SETBP1 possesses three AT-hook domains, which are known to be involved in the binding to gDNA. Therefore, it is reasonable to presume that other mechanisms, beside the SETBP1-SET-PP2a complex formation, exist.

With a combined use of ChIP-Seq and RNA-Seq techniques we explored the interaction of SETBP1 with gDNA as well as its effects on transcriptional regulation, while by mass spectrometry analysis and Co-IP experiments we identified novel SETBP1 interactors.

The second part of this work is related to the ETS2-ERG fusion observed in an APL patient that did not respond neither to ATRA nor to ATRA-ATO combination therapies. The complete remission for this subtype of AML upon treatment with ATRA is approximately 80-90%. However, no information regarding the biological role of the ETS2-ERG fusion is available in literature. To investigate about a possible involvement of the ETS2-ERG fusion in the failed response to ATRA therapy, we generated a stable cell line model expressing this chromosomal alteration and we studied the effect on cell differentiation in response to ATRA treatment by evaluating the expression of the differentiation marker CD11b.

Materials and Methods

3.1 Cell Lines

Human embryonic kidney (HEK) 293T, 293 Flp-In™ and TF-1 cell lines were purchased from ATCC (Mannas, Virginia, USA), Thermo Fisher Scientific (Waltham, MA, USA) and DSMZ (Braunschweig, Germany) respectively. HL-60 cell line was purchased from DMSZ (Braunschweig, Germany). 293T and 293 Flp-In™ were cultured in DMEM (Euroclone) medium supplemented with 10% FBS (Euroclone), 2 mmol/L L-glutamine, 100 U/mL penicillin (Euroclone), 100 µg/mL streptomycin (Euroclone), 20 mmol/L HEPES (Euroclone). TF-1 and HL-60 cell lines were cultured in RPMI 1640 (Euroclone), with 10% FBS (Euroclone), 2 mmol/L L-glutamine, 100 U/mL penicillin (Euroclone), 100 µg/mL streptomycin (Euroclone), 20 mmol/L HEPES (Euroclone). All the cells lines were incubated at 37° C with 5% CO_2 and routinely screened for mycoplasma (GATC Biotech AG, Constance, Germany).

3.2 SETBP1 plasmids and cell lines generation

Transient 293T transfectants were produced using the pcDNA6.2-SETBP1wt, pcDNA6.2-SETBP1-G870S or empty vector using jetPrime (Polyplus-transfection® SA, Illkirch, France) reagent accordingly to manufacturer's protocol. The pcDNA6.2-SETBP1wt and pcDNA6.2-SETBP1-G870S vectors were previously created in our laboratory.[1] Briefly, the long isoform (NM_015559.2) of SETBP1 cDNA was PCR amplified from a plasmid (SC114671, OriGene Technologies, Rockville, MD) and cloned in the pENTR™1A Gateway entry vector (Thermo Fisher Scientific). To obtain the SETBP1-G870S encoding plasmid, pENTR1A-SETBP1wt was mutagenized using the PfuUltra High Fidelity Polymerase (Agilent Technologies, Santa Clara, CA). After the digestion with DpnI (Roche, Monza, Italy), 2 µl of the product was used for the transformation of TOP10 competent cells (Life Technology). Plasmids were recovered and the presence of the G870S mutation was verified by Sanger sequencing (Eurofins, MWG Operon, Ebersberg, Germany). The two final destination vectors were obtained subcloning SETBP1wt or SETBP1-G870S into the pcDNA6.2/N-EmGFP-DEST (Thermo Fisher Scientific) destination vector using the Gateway clonase system (Thermo Fisher Scientific).

Stable 293 Flp-In Empty, Flp-In SETBP1wt and 293 Flp-In SETBP1-G870S were obtained cotransfecting pOG44 (Thermo Fisher Scientific) and pcDNA5™/FRT-SETBP1 vectors with FuGENE6 reagent (Promega, WI, USA) following manufacturer's instruction and they were selected adding 100 µg/ml Hygromycin to the grow medium. To obtain the pcDNA5™/FRT-SETBP1 vectors, the sequences encoding for SETBP1wt or SETBP1-G870S, were excised from

the pENTR1A by double-digestion of KpnI and XhoI restriction sites and subcloned in the pcDNA5^{TM}/FRT vector.

Stable TF-1 Empty, TF-1 SETBP1wt and TF-1 SETBP1-G870S were generated by means of retroviral transduction. Once excised from the previously generated pENTR1A-SETBP1 vectors, SETBP1wt and SETBP1-G870S amplicons were cloned into the MIGR1-EGFP vector in the XhoI restriction site. Recombinant retroviruses were produced in the Phoenix packaging cell line, which was transfected using FuGENE6 (Promega). The surnatant containing retroviral particles was collected after 3 days of culture and 5×10^4 cells were transduced by spin infection in retroviral surnatant with 4 µg/mL of polybrene (Sigma-Aldrich) and resuspended in complete medium after 48 h. To select stable transfectants, the cells were sorted for GFP+ expression using a FACSAria flow cytometer.

The plasmids carrying the deletion of the HBM site (aa 991-994) within the SETBP1 sequence were generated by mutagenizing the pcDNA6.2 vectors (SETBP1wt or SETBP1-G870S) using the PfuUltra High Fidelity Polymerase enzyme (QuickChange II XL Site-Directed Mutagenesis Kit, Stratagene, La Jolla, CA) as described above. For the site-directed mutagenesis, the following forward and reverse primers were used:

- SETBP1_HCF1del_for:
  5’ – CAGCATTTTTCGGATTAATTTTCCGGTGCCATATATCCAGTATG – 3’
- SETBP1_HCF1del_rev:
  5’ – CATACTGGATATATGGCACCGGAAAATTAATCCGAAAAATGCTG – 3’

The plasmid pCGN-HCF1-fl (Addgene plasmid #53309) was a kindly gift from Winship Herr [138].

3.3 Chromatin Immunoprecipitation sequencing (ChIP-Seq)

For ChIP-Seq experiments, the proteins were cross-linked to DNA with 0.4% of formaldehyde. Subsequently, cells were lysed and sonicated by means of a Bioruptor sonicator system (Diagenode, SA, USA). The fragmented chromatin was immunoprecipitated using histone 3 lysine 4 tri-methylation (H3K4me3) (ab8580, Abcam, UK), histone 3 lysine 4 di-methylation (H3K4me2) (C15410035C, Diagenode), histone 3 lysine 36 tri-methylation (H3K36me3) (ab9050, Abcam), histone 3 lysine 27 acetylation (H3K27Ac) (ab4729, Abcam), histone 3 lysine 9 acetylation (H3K9Ac) (39137, Active Motif, CA, USA) and histone 4 lysine 20 methylation (H4K20me1) (Mab147-010, Diagenode) antibodies or with the Anti-V5 Agarose Affinity Gel

(Sigma-Aldrich). The immunoprecipitated chromatin was reverse-crosslinked and DNA was purified using the QIAquick PCR Purification Kit (QIAGEN) The DNA libraries preparation and the subsequent sequencing were performed by a commercial service (Galseq, Monza, Italy), using the TrueSeq ChIP library prep kit IP-202-1012 (Illumina, San Diego, CA) and an Illumina HiSeq 2500 sequencer in single read mode.

The data obtained by ChIP-Seq analysis were validated by means of SYBR-Green real-time quantitative PCR (qRT-PCR). All the qRT-PCR experiments were performed in triplicate amplifying the immunoprecipitated DNA with specific primer pairs and using the input DNA as normalization control.

For ChIP-Seq data validation, the following primer pairs were used:

- BMP5_Fw: 5' – CAACCCTGCTGGGAAAGAAGAG – 3'
  BMP5_Rw: 5' – TCATCAAGCTAACTTAGGCACAAC – 3'
- NFE2L2_Fw: 5' – AACCAGAAGAATACAATCCCAATG – 3'
  NFE2L2_Rw: 5' – AAGAAGTTTCTGCTCATCCTTTGTAG – 3'
- PDED4_Fw: 5' – CCTTGAGCCAACCTTCTCCTTC – 3'
  PDED4_Rw: 5' – CACCCAAAGACATGACCAACCTC – 3'
- SKIDA1_Fw: 5' – TTCAAGTATCACGTTACTGTTTGC – 3'
  SKIDA1_Rw: 5' – GTCACTTATTCAGCCACGCAGAC – 3'
- COBLL1_Fw: 5' – TCTAATTGGTGGCAGGTTTAAGC – 3'
  COBLL1_Rw: 5' – TGTCTGTCAGGTGTAAAGAATCATC – 3'
- RUNX1P2_Fw: 5' – CCTATGCAAACGAGCTGAGG – 3'
  RUNX1P2_Rw: 5' – GCTCTATGAATGAGAGTGCCTG – 3'
- MECOM_Fw: 5' – CTCCCAAATGTCTTAATCGTGTCG – 3'
  MECOM_Rw: 5' – TTCGGACCCTTTGGCTAGATTGTG – 3'

3.4 Co-immunoprecipitation (Co-IP)

The pcDNA6.2-SETBP1wt, SETBP1-G870S or the empty vector were transduced in HEK 293T cells. After 48 h, 10^8 cells were harvested and washed in ice-cold PBS. Nuclear proteins were extracted using the following optimized protocol. Briefly, cells were incubated 10 minutes on ice in Buffer 1 (Pipes pH8.0 5 mM, KCl 85 mM, NP-40 0.5%) with the addition of protease inhibitors (Halt™ Protease Inhibitor Cocktail, Thermo Fisher Scientific) and homogenized with a douncer homogenizer. The cells lysates were centrifuged 10 minutes at 2500 rpm and the surnatants were discarded. The pellets were resuspended in Buffer 2 (Tris-HCl pH 8.0 50 mM,

SDS 0.1%, Deoxicholate 0.5%) plus proteases inhibitor and sonicated with Bioruptor Next Gen (Diagenode) to disrupt gDNA. After centrifugation at 13.000 rpm for 10 minutes, the surnatant (nuclear fraction) was quantified by standard Bradford assay. One µg of total proteins were incubated, under rotation, with 100 µL of Anti-V5 Agarose Affinity Gel (Sigma-Aldrich) overnight at 4° C. Next, after 3 washes with ice-cold PBS supplemented with protease inhibitors, the proteins were recovered using elution buffer (7M Urea, 2M Thiourea, 4% CHAPS) or Laemmli buffer for mass spectrometry (MS) or western blot analysis, respectively. For immunoblot analysis, the following antibodies were used.

Primary antibodies: anti-V5 (ab27671 330 Abcam, Cambridge, UK), anti-HCF1 (A301-399A Bethyl Laboratories, TX, USA), anti-MLL1 (14689 Cell Signaling Technology, MA, USA), anti-PHF8 (A301-772A, Bethyl Laboratories). Secondary antibodies anti-mouse or anti-rabbit conjugated with HRP were used (Biorad Hercules).

3.5 Proteomics data analysis

The proteomic data analysis was kindly performed by the Meilahti Clinical Proteomics Core Facility, Department of Biochemistry and Developmental Biology, Faculty of Medicine, University of Helsinki, Helsinki, Finland. Briefly, the protein fragmentation was obtained by digesting samples in Amicon Ultra-0.5 centrifugal filters (Merck Millipore, MA, USA) using the filter aided sample preparation method (FASP) [139]. The obtained peptides were separated by means of the nanoAcquity UPLC system (Waters, Milano, Italy). The instrument was equipped with a 5- µm Symmetry C18 trapping column, 180 µm × 20 mm, reverse-phase (Waters), followed by an analytical 1.7-µm, 75 µm × 250 mm BEH-130 C18 reversed-phase column (Waters), in single-pump trapping mode. The protein identifications were performed using the ProteinLynx Global Server (PLGS v3.0) [140] and the queries search were performed against the UniProt human protein database (release_07072015, x 71907 entries) using Ion Accounting algorithm ad using the following parameters: peptide and fragment tolerance: automatic, maximum protein mass: 500 kDa, minimum fragments ions matches per protein: 7, minimum fragment ions matches per peptide: 3, minimum peptide matches per protein: 1, primary digest reagent: trypsin, missed cleavages allowed: 2, fixed modification: carbamidomethylation C, variable modifications: deamidation (N, Q), oxidation of Methionine (M) and false discovery rate (FDR) < 4 %.

3.6 Immunofluorescence microscopy

293T cells were transfected with pcDNA6.2-SETBP1 or empty vectors as previously described and seeded on coverslips previously treated with poly-D-lysine at 0.1 mg/mL concentration. After washing, the cells were fixed in p-formaldehyde in 0.12M sodium phosphate buffer, pH 7.4, and incubated 1 h with primary antibodies in GBD buffer. Then, after 1 hour staining with conjugated secondary anti IgG antibody followed by washing steps, the coverslips were mounted on glass slides.

3.7 Acceptor photobleaching Resonance Energy Transfer (FRET)

All the FRET experiments were performed using the laser-induced acceptor photobleaching method [141]. In our experimental setup the FRET couples analyzed were made up of transiently transfected GFP-fused SETBP1wt or mutated forms in combination with transiently transfected Flag-tagged bTrCP or endogenous HCF1. 48 hours after transfection 293T cells were labeled with proper primary and secondary antibodies and imaged for FRET analysis. For FRET investigations the GFP signal was used as donor fluorochrome and Alexa Fluor 555-conjugated secondary antibodies represents the acceptor fluorochrome. Briefly, three images were acquired before bleaching in the 488 nm and the 561 nm channels using the line-by-line sequential mode without any averaging steps to reduce basal bleaching. Bleaching of the acceptor was performed within region of interest (ROI) identified in the nuclear areas with a positive colocalization between the FRET couples using 30 pulses of a full power 20 mW 561-nm laser line (each pulse 1.28 μsec/pixel). After bleaching, seven images were acquired in the same channels without any time delay to obtain a full curve. All the parameters, such as the number of bleaching steps, the laser intensity and the acquisition settings, were the same for each experiment. The quantification of FRET signals was performed by measuring the average intensities of ROIs in the donor and acceptor fluorochrome channels before and after bleaching. The ImageJ software (http://rsbweb.nih.gov/ij/) was used for the analysis. A distinct unbleached ‘sentinel’ ROI, approximately of the same size of the bleached ROI, was measured in parallel to detect any change of fluorescence intensities not caused by FRET during the measurements and all the results were normalized according to the background bleaching recorded in the sentinel. Proper controls were performed to verify that no artefacts were generated in the emission spectra in the experimental setup due to sample overheating. For each experimental condition, twenty measurements from three different experiments were recorded.

3.8 RNA-Seq

To generate RNA libraries 1 µg of total RNA extracted from 5×10^6 cells using TRIzol (Thermo Fisher Scientific) was used. The quality of RNA was checked using a Tape Station instrument (Agilent Technologies). To avoid over-representation of 3'ends, only high-quality RNA with a RNA Integrity Number (RIN) ≥ 8 was used. RNA was processed according to the TruSeq Stranded mRNA Library Prep Kit protocol. The libraries were sequenced on an Illumina HiSeq 3000. Image processing and basecall were carried out using the Illumina Real Time Analysis Software. FASTQ files were aligned to the human genome (GRCh38/hg39) by using the splice junction mapper STAR36 in conjunction with the corresponding splice junctions Ensembl GTF annotation, using the following parameters:
--runThreadN 8 --outReadsUnmapped Fastx --outFilterType BySJout --outSAMattributes NH HI AS nM MD --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --alignIntronMin 20 –alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignTranscriptsPerReadNmax 100000 --quantMode TranscriptomeSAM GeneCounts --limitBAMsortRAM 16620578182 --outSAMtype BAM SortedByCoordinate --chimSegmentMin 20 --chimJunctionOverhangMin 10.

3.9 Differential expression analysis

The gene counts output obtained by the --quantModeGeneCounts parameter was used to calculate gene expression. All the subsequent steps were performed using dedicated R scripts. These scripts were automatically generated through metaprogramming, using a dedicated tool (StarCounts2DESeq2) developed by Dr. Rocco Piazza using C# programming language. Differential expression analyses were performed using DESeq2 [142]. The analysis of differentially expressed genes characterized by the co-occurrence of both H3K4me2 and H3K9Ac marks was performed with the following filtering criteria: H3K4me2 log2FC > 0.5; H3K9ac log2FC > 0.5; RNA p-value <0.05.

3.10 Quantitative Real-Time PCR (qRT-PCR)

Total RNA was extracted with TRIzol reagent (Thermo Fisher Scientific) with standard protocol. After quantification by A260 and quality evaluation by the A260/A280 ratio, 1µg of RNA was treated with DNAseI Amplification Grade (Thermo Fisher Scientific) to avoid gDNA contamination and reverse transcribed using Reverse Transcription Reagents (Thermo Fisher

Scientific). The qRT-PCR reaction was performed using TaqMan® Brilliant II QPCR Master Mix (Agilent Technologies, Santa Clara, CA, USA) on a Stratagene-MX3005P (Agilent technologies) accordingly to manufacturer's specification. The GUSB housekeeping gene was used as an internal reference. The following TaqMan® Gene Expression Assays (Thermo Fisher Scientific) were used: *SKIDA1* (Hs01096520), *NFE2L2* (Hs00975961), *PDED4* (Hs03988495), *FBXO8* (Hs00942619), *CEP44* (Hs00604612), *COBLL1* (Hs01117513), *BMP5* (Hs00234930), *ERBB4* (Hs00955525), *CDKN1B* (Hs01597588), *SETBP1* (Hs00210203), *RUNX1* (Hs01021971), *MECOM* (Hs00602795).

3.11 *MECOM* analysis

In order to identify direct targets of *MECOM* among the SETBP1 DEGs, previously published RNA-Seq and ChIP-Seq datasets in the context of *MECOM* overexpression were analyzed. The resulting gene set was functionally annotated with FunRich [143] and filtered in accordance with their expression within the bone marrow (Hypergeometric enrichment test).

3.12 AML patients and ETS2-ERG fusion screening

A cohort of 20 patients, affected by newly diagnosed or relapsed APL and not responding or partially responding to medical treatment with ATRA, was included in the study. Mononuclear cells derived from bone marrow (BM) aspirates or peripheral blood (PB) and RNA samples were collected from different clinical centers. The samples were obtained from patients after informed consent, according to the Declaration of Helsinki. The Institutional Review Board of our University approved the study. Total RNA was extracted using the RNeasy total RNA isolation kit (QIAGEN, Hilden, Germany) accordingly to manufacturer's protocol. Optional treatment with RNase-Free DNase Set (QIAGEN) was done to avoid gDNA contamination. RNA recovery was determined by A260 and its quality assessed by A260/280 ratio. An amount of 0.5-1 µg of RNA was reverse transcribed using Reverse Transcription Reagents (Thermo Fisher Scientific). To check for ETS2-ERG fusion transcripts cDNA was analyzed by PCR reaction followed by nested PCR. The *GUSB* gene was used as retrotranscription control. The following primers pairs were used:

- ETS2_FOR_3: 5' – TTCGGAATCAAGAATATGGACCAGG – 3'
  ERG_Rev_3: 5' – GAGCTTATCGTAGTTCATGTTGGGT – 3'

- ETS2_FOR_nested: 5’ – CCACCAATGAGTTCAGTC – 3’
  ERG_Rev_nested: 5’ – TGTTGGGTTTGCTCTTCC – 3’
- GUSB_FOR: 5’ – GAAAATACGTGGTTGGAGAGCTCATT – 3’
  GUSB_Rev: 5’ – CCGAGTGAAGATCCCCTTTTTA – 3’

The analyzed samples were obtained from: ULSS 6 Ospedale di VICENZA, Vicenza, Italy; AOU Policlinico di Modena, Modena, Italy; ASST Spedali Civili, Brescia, Italy; Fondazione PTV-Policlinico Tor Vergata, Roma, Italy.

3.13 ETS2-ERG plasmids and cell line generation

For ETS2-ERG plasmids generation, cDNA from patient AML002 was used as substrate for PCR amplification. BglII restriction site and Kozak consensus sequence were introduced artificially at the 5’ position of ETS2-ERG, whereas V5-tag and XhoI restriction site were introduced at the 3’position using the following forward and reverse primers:

- ETS2-ERG_BglII_for: 5' – AATAAGATCTGCAGGATGAATGATTTCGGAATCAA G – 3'
- ETS2-ERG_V5_XhoI_rev: 5 ' – TATTCTCGAGTTACGTAGAATCGAGACCGAGGA GAGGGTTAGGGATAGGCTTCCCTTCGAAGGGCCCTCTAGACTCTAGGTAGTA AGTGCCCAGATGAGA – 3'

The amplicon was introduced by restriction enzyme/ligase cloning in the pMSCV-IRES-GFP II (pMIGII) vector that was a gift from Dario Vignali (Addgene plasmid #52107).[144]
HL-60 Empty and HL-60-ETS2-ERG were produced by retroviral infection as previously described. All the cell lines were sorted for GFP expression at least two times using a FACSAria flow cytometer. Protein expression was confirmed by western blot analysis with the primary anti-V5 antibody (ab27671 330 Abcam, Cambridge, UK). Actin (Sigma-Aldrich) was used as loading control. Secondary antibodies anti-mouse or anti-rabbit conjugated with HRP were used (Biorad Hercules, CA, USA).

3.13 HL-60 cells differentiation assay and flow cytometric analysis

The differentiation of HL-60 Empty and ETS2-ERG was assessed by flow cytometry detection of the cellular surface differentiation marker CD11b. Briefly, 300.000 cells/mL were seeded in 6-well plates and treated with ATRA at 1 μM. DMSO (Euroclone) was used as the control. After 72 h incubation, 1×10^6 cells were harvested, washed twice in PBS and incubated 30 minutes at 4° C with PE-conjugated CD11b antibody (BD Biosciences, San Jose, CA). After that, the cells were washed twice, resuspended in PBS and 20.000 events were acquired with a BD FACSCantoI (BD Bioscience) flow cytometer. The percentage of CD11b positive cells was calculated using the '*Histogram Subtraction Tool'* of the FCS Express 4 software (De Novo Software, CA, USA).

ATRA was purchased by Sigma-Aldrich, St. Louis, MO, and dissolved in DMSO (Euroclone, Milano, Italy) at 10 mM concentration, aliquoted and stored at -20° C protected from the light until use.

Results

4.1 SETBP1 protein interacts with genomic DNA

As previously described, mutations in SETBP1 are observed in aCML as well as in other hematological malignancies and in the SGS syndrome [1, 92, 94]. In pathological condition, the increased inhibitory effect on the tumor suppressor PP2A, due to accumulation of SETBP1, explains the oncogenic properties of SETBP1 [90]. However, it is possible that SETBP1 could act through other mechanism, due to the presence of three A/T-hook domain within its sequence [93]. Indeed, it is described in literature that A/T-hook domains are key elements of DNA-binding proteins involved in DNA-binding [145].

To investigate the hypothesis that SETBP1 protein could act as a DNA-binding protein we created stable isogenic cellular models using 293 Flp-In™ cell line. The Flp-In™ system allows the integration of a gene of interest (GOI) at a specific genomic *locus*. As a result, the GOI is expressed as a single integrated copy, mimicking a physiological condition. We generated three different cell lines expressing SETBP1 wild type or mutated at position G870S, as well as Empty lines. Firstly, the expression of SETBP1wt or SETBP1-G870S was assessed both at mRNA and protein levels (Figure 11). Accordingly to previously published data, we observed a similar expression of SETBP1wt and SETBP1-G870S at transcript level but not at protein level, with the G870S line showing protein accumulation due to decreased protein degradation [1].

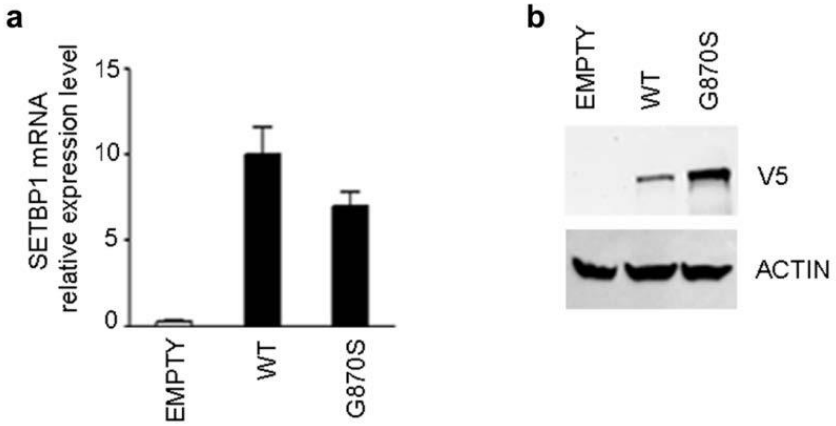


Figure 11 Characterization of 293 Flp-In stable cell lines.

A) qRT-PCR performed on 293 Flp-In cells expressing SETBP1wt (WT), mutated (G870S) or Empty vector (EMPTY). GUSB gene was used as internal reference. B) Western Blot analysis of 293 Flp-In total cells lysates. The anti-V5 antibody recognizes the SETBP1-V5 tagged form. Actin was used as loading control.

Next, we performed ChIP-Seq experiments against the V5-tag for both the Flp-In SETBP1wt and SETBP1-G870S cell lines. In cells expressing SETBP1wt and SETBP1-G870S we observed the presence of broad peaks that occurred mainly in A/T-rich regions. The mean peaks A/T content for SETBP1wt/G870S was 65.8%, whereas it was 59% in the whole human genome ($p<0.001$; Figure 12).

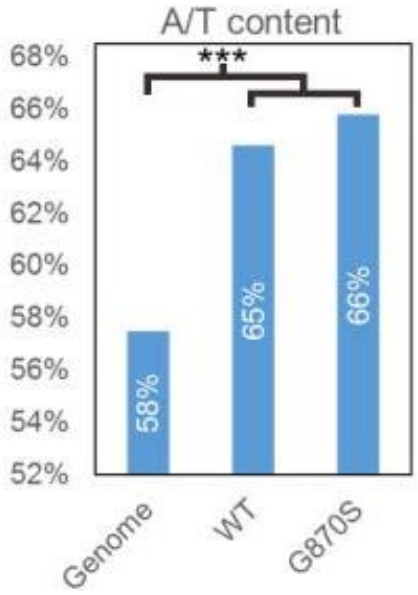


Figure 12 SETBP1 bound gDNA A/T content.

A/T content comparison between SETBP1wt (WT) and SETBP1-G870S (G870S) binding regions and the reference genome. *** $p<0.001$.

Furthermore, a *de novo* motif discovery search performed against SETBP1 sequence, identified a consensus binding site AAAATAA/T ($p=0.002$), which is largely overlapping with the consensus binding motif of high motility group AT-hook 1 (HMGA1) gene (AAAATA; http://hocomoco.aoutosome.ru/motif/HMGA1_HUMAN.H10MO.D) (Figure 13) [146].

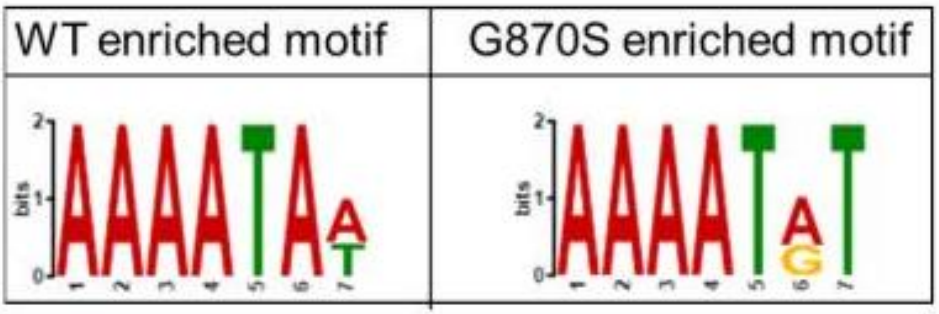


Figure 13 SETBP1 de novo motif discovery.

SETBP1 consensus binding site revealed by de novo motif discovery analysis using the MEME software [147].

HMGA1 is described as a chromatin remodeling protein [148] that is able to bind the minor groove of DNA A/T-rich DNA sequences. Moreover, it is characterized by the presence of three AT-hook DNA domains [149]. These data suggest that also SETBP1 could bind to gDNA through its AT-hook DNA binding motifs. ChIP-Seq analysis identified a total of 386 genomic regions bound by SETBP1wt and 2722 bound by SETBP1-G870S. In addition, peak distribution analysis revealed that the 58% and 65% of these regions, respectively for SETBP1wt and SETBP1-G870S, were located within a distance of +/-10 kilobases (Kbs) from each transcription start site (TSS) (Figure 14).

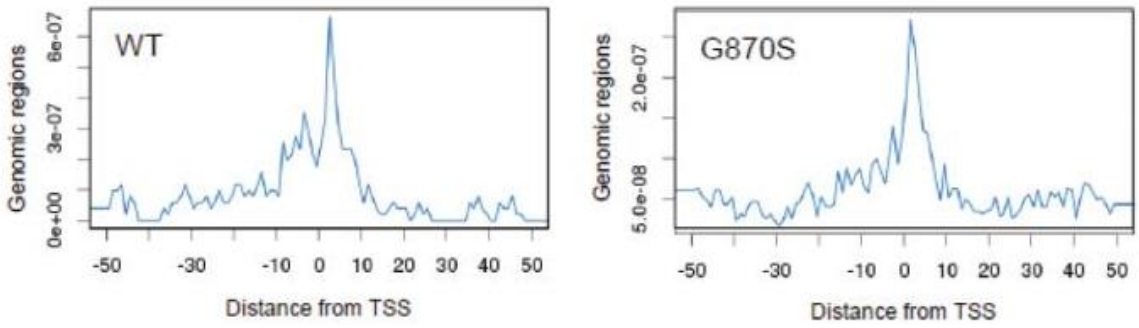


Figure 14 Peak distribution density of SETBP1 bound region.

Peak distribution density analysis of SETBP1wt and SETBP1G870S bound gDNA according to the distance from gene transcription start sites.

This indicates that both SETBP1wt and SETBP1-G870S preferentially interact with the gDNA around promoter regions, suggesting that SETBP1 may function as a transcription factor (TF). In fact, it is known that TFs bind to promoter regions, starting the recruitment of an active initiation complex around TSS [150]. However, beyond the promoter regions, SETBP1 is able to bind also enhancers and exonic, intronic and intergenic regions (Figure 15). In comparison to SETBP1wt, ChIP-Seq experiments showed an increase in the number of peaks referred to SETBP1-G870S for all the regions analyzed (Figure 15), which is possible explained by the effect of the G870S mutation, leading to SETBP1 accumulation as previously described [1].

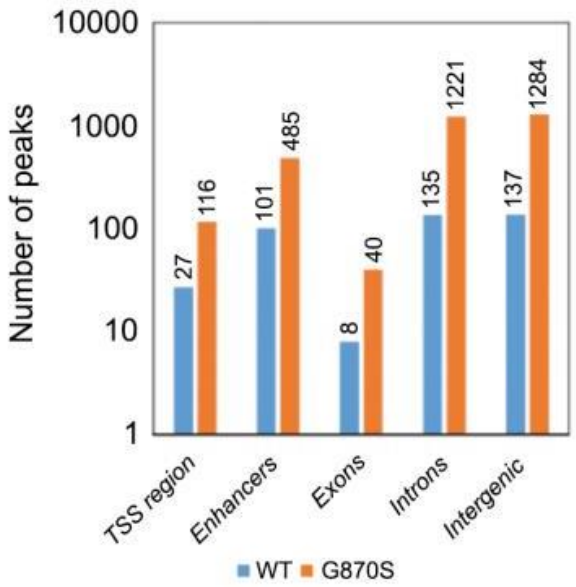


Figure 15 Peak quantitation in the different genomic regions.

Peak counts showing the binding of SETBP1wt and SETBP1-G870S to different genomic regions.

Following peak annotation, we identified 305 genes bound to SETBP1-G870S.
A subsequent Gene Ontology (GO) functional enrichment analysis revealed a robust enrichment of terms referred to multiple developmental processes such as bone, cartilage and organ development as well as mesenchymal cell differentiation (Figure 16).

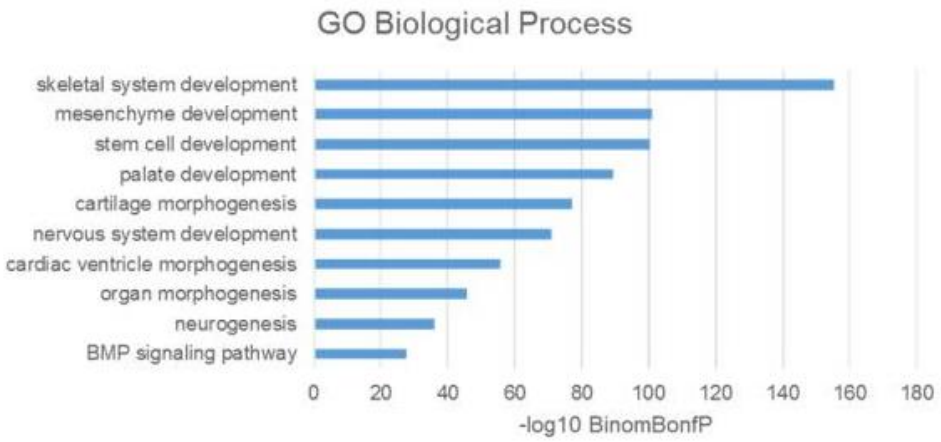


Figure 16 Gene Ontology (GO) on SETBP1 target genes.

GO biological process enrichment performed on the genes bound by SETBP1 showing enrichment in several developmental processes.

4.2 Effect of SETBP1-G870S at transcriptional level.

In order to assess the effects of SETBP1-G870S gene at transcriptional level, we performed RNA-Seq experiments. A differential expression analysis between Flp-In^{TM} cells expressing the Empty vector and cells expressing the SETBP1-G870S gene, identified the presence of 2687 differentially expressed genes (DEGs), 1157 upregulated and 1530 downregulated (data not shown). Both up- and down-regulated genes were intersected with the genes bound by SETBP1 in promoter regions (identified with ChIP-Seq experiment), revealing 130 upregulated and only 16 downregulated co-occurring genes (89%; $p=1 \times 10^{-6}$; Figure 17).

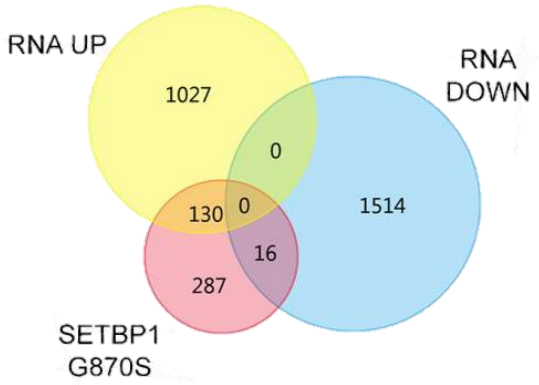


Figure 17 SETBP1 target genes and their differential expression in SETBP1-G870S cells.

Venn diagram showing the number of differentially expressed genes being directly bound by SETBP1 within their promoter region (red circle).

This marked preference towards the gene upregulation suggests that SETBP1 is more likely acting as a positive inducer of gene expression. These data were confirmed also by qRT-PCR analysis (Figure 18).

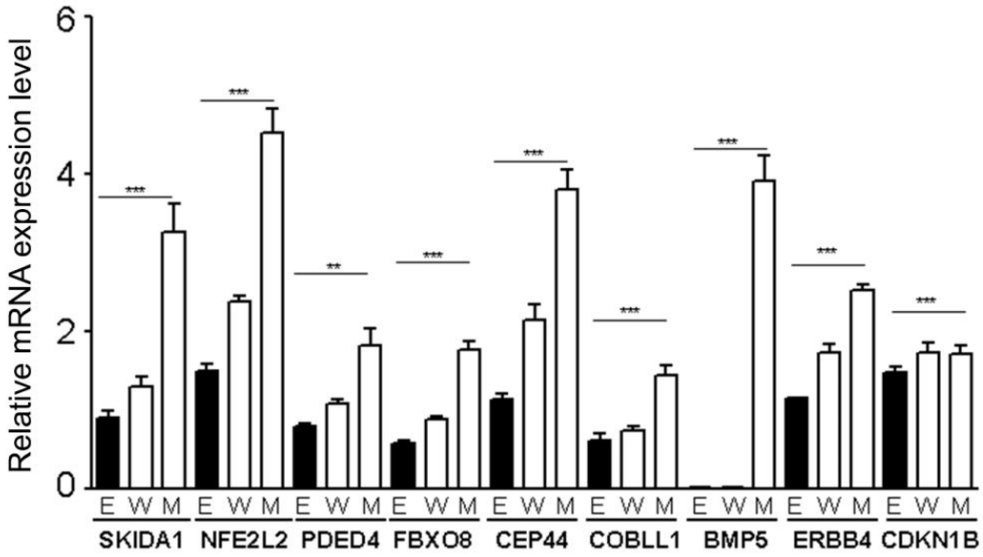


Figure 18 Expression of a subset of DEGs identified by RNA-Seq.

qRT-PCR analysis of some of the genes identified by the RNA-Seq analysis showing significant upregulation in SETBP1-G870S samples for all the genes tested. *** $p<0.001$. E (Empty), W (SETBP1wt), M (SETBP1-G870S).

A GO functional enrichment analysis performed on the DEGs obtained from the RNA-Seq analysis, showed a significant enrichment of GO terms relative to both cell differentiation and tissue development (Figure 19).

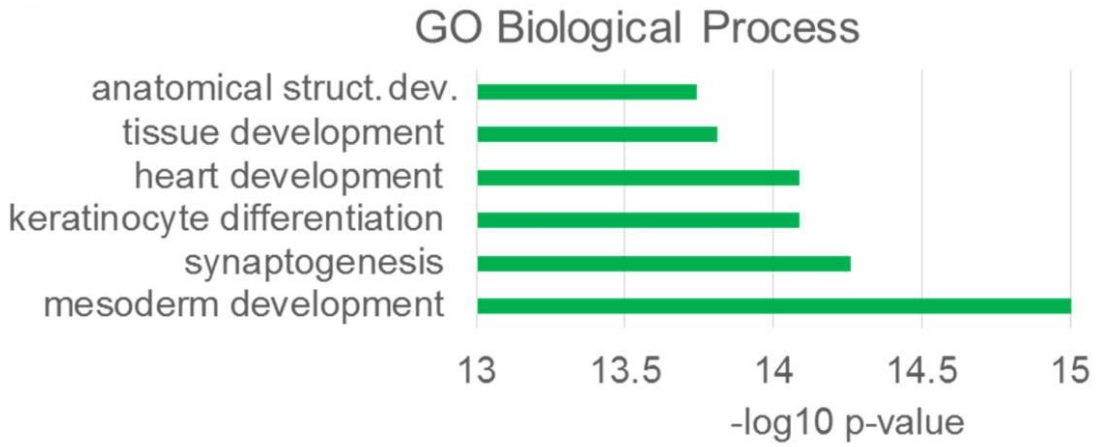


Figure 19 GO biological process functional enrichment of differentially expressed genes resulting from G870S mutation.

The differentially expressed genes in the context of the SETBP1-G870S mutation result in enrichment of development-related GO biological processes.

4.3 MECOM and its direct targets are upregulated by SETBP1

MDS1 and EVI1 complex locus (*MECOM*), previously known as *EVI1*, was one of the genes identified as upregulated in the SETBP1-G870S Flp-In™ model by the differential expression RNA-Seq analysis (Figure 20a). *MECOM* upregulation was also confirmed by qRT-PCR (Figure 20b).

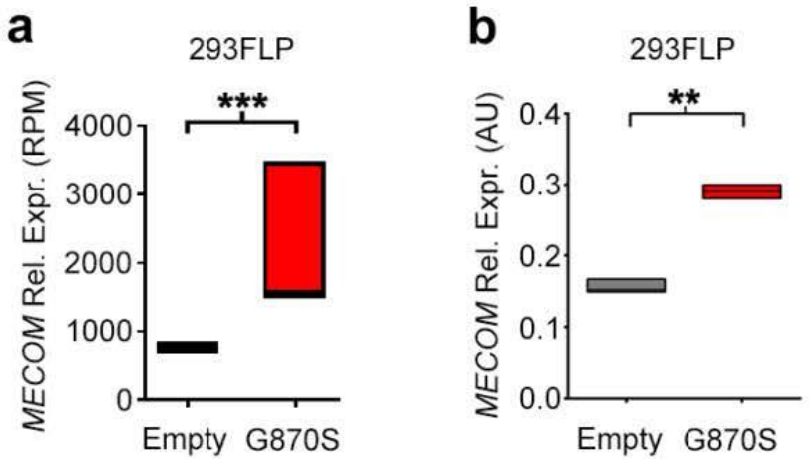


Figure 20 MECOM expression in 293 Flp-In cells.

A) Box-plot showing the RNA-Seq differential expression analysis of *MECOM* in the 293 Flp-In Empty/SETBP1-G870S cell model. The top and bottom of the box represent the first and third quartile. The internal line represents the median. B) qRT-PCR analysis of *MECOM* expression in the 293 Flp-In Empty/SETBP1-G870S cell model. ** p<0.01, *** p<0.001.

To confirm this data also in a hematopoietic cellular model, we transduced the human myeloid TF-1 cell line with a retrovirus expressing either SETBP1-G870S or the Empty vector and we assessed *MECOM* expression at mRNA level by qRT-PCR. As expected, we observed an increase of *MECOM* expression in the SETBP1-G870S model in comparison to the Empty control (Figure 21).

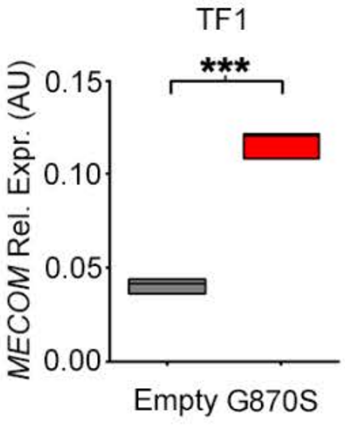


Figure 21 MECOM expression in TF1 cells.

qRT-PCR analysis of *MECOM* expression in TF1 cell line transduced with an empty vector or with a vector encoding SETBP1-G870S. The top and bottom of the bo represents the first and third quartile. The internal line represents the median. *** p<0.001

A peak distribution analysis of ChIP-Seq data revealed an enrichment in the promoter region of *MECOM* for both SETBP1-G870S and SETBPwt, with a more marked effect for SETBP1-G870S (Figure 22), thus suggesting a potential direct effect of SETBP1 on the upregulation of *MECOM*. To gain further insight into the mechanism of SETBP1-mediated activation of gene expression, we performed ChIP-Seq analysis of a set of histone marks, known to be associated with transcriptional activation, such as H3K4me2, H3K4me3, H3K9ac, H3K27ac and H3K36me3. The different histone methylation tracks were superimposed to the SETBP1 ChIP-Seq tracks previously aligned to the reference human genome hg19, observing enriched peaks for these histone marks around the promoter of *MECOM*, except for the H3K36me3 mark (Figure 22). Further, to confirm the binding of SETBP1 to *MECOM* promoter, we performed independent ChIP experiments against the V5 tag in the established Flp-In cellular models, followed by a qRT-PCR analysis with specific primers targeting the *MECOM* promoter in the SETBP1-G870S binding region. As expected, we observed an enrichment of this *MECOM* promoter region in SETBP-G870S samples in comparison to Empty control (box of Figure 22).

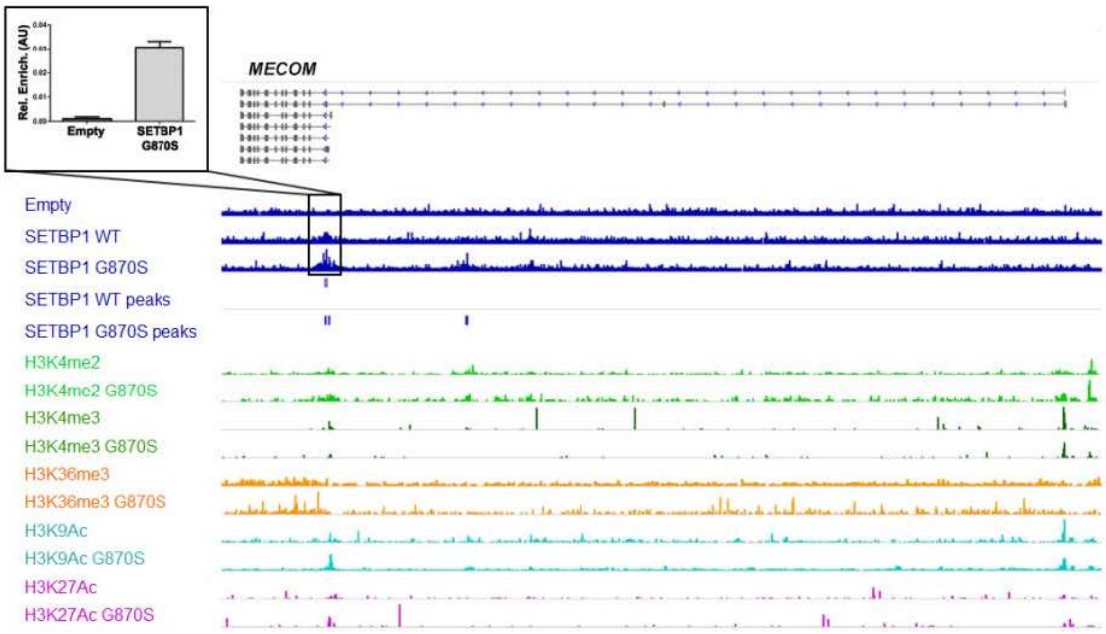


Figure 22 Analysis of MECOM locus occupancy in comparison to SETBP1 and histone marks.

SETBP1 ChIP-Seq coverage track and peak alignment to the hg19 reference genome (blue track) superimposed to the different histone methylation ChIP-Seq coverage tracks, within the *MECOM* locus. The boxed histogram represents an independent ChIP experiment performed against the V5 tag in Flp-In cells followed by a qRT-PCR directed against the predicted SETBP1-G870S binding locus on the *MECOM* promoter.

Subsequently, we wanted to investigate whether the dysregulation of *MECOM* could affect the expression of its downstream target genes (Table 6).

Gene symbol	Description	log2 FC 293FLP G870S/Empty
ATPIF1	ATPase inhibitory factor 1	0.417046544
BUB1	BUB1 mitotic checkpoint serine/threonine kinase	0.55302509
CD63	CD63 molecule	0.580584785
CDC25C	cell division cycle 25C	0.633394407
CDKN3	cyclin-dependent kinase inhibitor 3	0.433961059
CIT	citron rho-interacting serine/threonine kinase	0.439699554
DHRS1	dehydrogenase/reductase (SDR family) member 1	0.483302518
FKBP15	FK506 binding protein 15, 133kDa	0.474183165
GINS4	GINS complex subunit 4 (Sld5 homolog)	0.542144085
GMNN	geminin, DNA replication inhibitor	0.48310151
HAUS1	HAUS augmin-like complex, subunit 1	0.458641317
HMOX1	heme oxygenase 1	0.71242341
KDM1B	lysine (K)-specific demethylase 1B	0.828246828
KIF20A	kinesin family member 20A	0.373195269
ME2	malic enzyme 2, NAD(+)-dependent, mitochondrial	0.823516525
MFF	mitochondrial fission factor	0.487486993
NAGK	N-acetylglucosamine kinase	0.603971934
RSU1	Ras suppressor protein 1	1.004837381
STIL	SCL/TAL1 interrupting locus	0.428445113
TPI1	triosephosphate isomerase 1	0.550514412
TSPAN13	tetraspanin 13	0.376906023
TYMS	thymidylate synthetase	0.409222189
WIPI1	WD repeat domain, phosphoinositide interacting 1	0.749784455

Table 6 Downstream target genes of MECOM.

RNA-Seq analysis of MECOM target genes in Flp-In SETBP1-G870S/Empty models. Expression values are calculated as the log2 of the G870S/Empty Fold Change.

To this purpose, we analyzed the RNA-Seq data derived from three independent clones of Flp-In SETBP1-G870S cells as well as three Empty control clones. As a result we observed that also the

downstream targets of *MECOM* were significantly upregulated in SETBP1-G870S samples in comparison with Empty control (p=0.0001; Figure 23, Table 6).

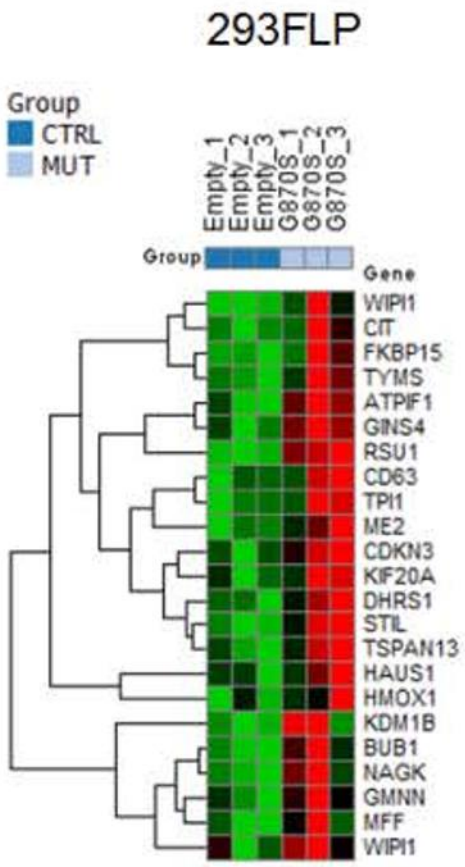


Figure 23 Differential expression of MECOM target genes in Flp-In cells.

Gene expression heatmap of MECOM target genes in 3 Empty/SETBP1-G870S Flp-In clones.

Moreover, in order to confirm these results in patient-derived samples we collected 32 aCML patients, 11 positive and 21 negative for SETBP1 somatic mutations, and we investigated the expression level of *MECOM* by RNA-Seq and qRT-PCR analyses. As expected, we observed a significant increase of *MECOM* at mRNA level in patients with mutated SETBP1 in comparison to the patients with SETBP1 wild type (p=0.0002; Figure 24a). This data was also supported by qRT-PCR (Figure 24b).

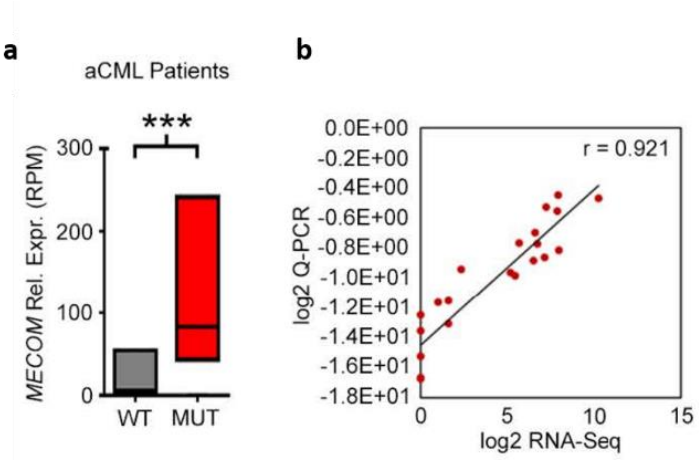


Figure 24 MECOM expression in aCML patients.

A) Differential MECOM expression as Read counts Per Million of mapped reads (RPM) in 32 aCML patients carrying WT (n=21) or mutated (n=11) SETBP1. The top and bottom of the box represent the first and third quartile. The internal line represents the median. B) Linear correlation MECOM expression as assessed by RNA-Seq (X axis) and qRT-PCR (Y axis). r represents the Pearson linear correlation coefficient. *** $p<0.001$

Analysis of MECOM downstream targets, done on the same aCML set revealed a significant upregulation of *MECOM* direct targets in patients positive for SETBP1 mutation ($p<0.0001$; Figure 25).

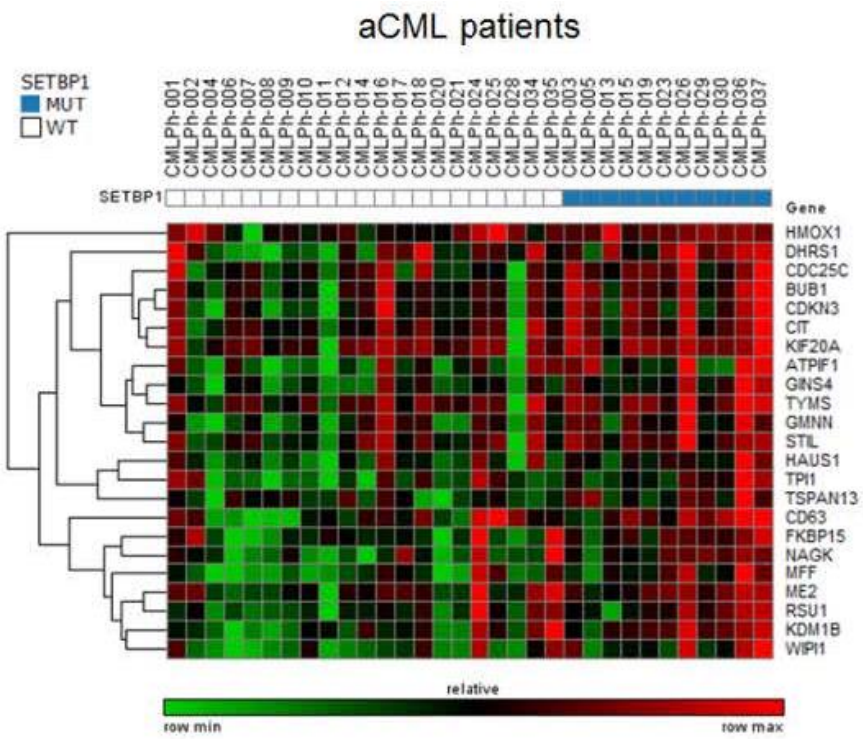


Figure 25 Expression of MECOM target genes in aCML patients.

Gene expression heatmap of *MECOM* target genes in 32 aCML patients carrying WT (n=21) or mutated SETBP (n=11). *MECOM* downstream targets are upregulated in SETBP1^{+} patients.

4.4 SETBP1 has a role in the epigenetic modulation of gene expression.

Previous studies demonstrated that A/T-hook containing proteins are often part of large chromatin remodeling complexes [151-153]. To determine whether SETBP1 binding to DNA was able to lead to epigenetic changes we took advantage of the previously generated ChIP-Seq data. First of all, we performed a peak distribution analysis for the histone marks H3K4me2, H3K4me3, H3K9ac, H3K27ac and H3K36me3. As reported before, these histone marks are known to be associated with transcriptional activation. In line with their expression profiles reported in literature, a peak enrichment analysis revealed an enrichment in promoter regions for all the marks tested, except H3K36me3 [154]. Indeed, the latter is known to be expressed in gene bodies (Figure 26) [155].

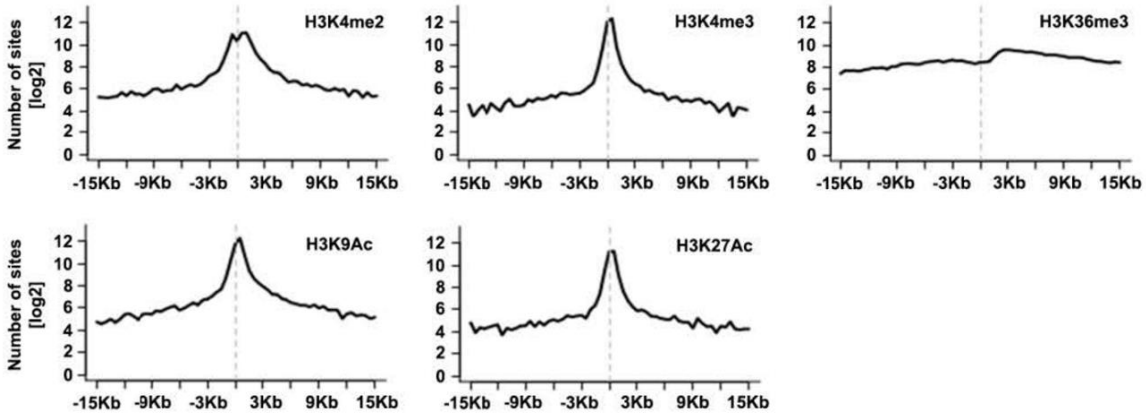


Figure 26 Peak distribution analysis of histone marks.

Histone modification ChIP-Seq peak distribution density according to their distance from gene transcription start site. ChIP were performed against SETBP1.

Next, we wanted to test if there was a correlation between the regions bound by SETBP1-G870S and the increase of these histone marks. For this purpose, we compared the differential DNA binding of SETBP1, with each histone modification differential enrichment, both expressed as fold change between G870S and Empty control. As a result we observed a significant correlation between SETBP1 promoters occupancy and increase of the H3K4me2 and H3K9ac histone marks ($p=1 \times 10^{-4}$ and $p=1.2 \times 10^{-16}$ respectively; Figure 27a-b).

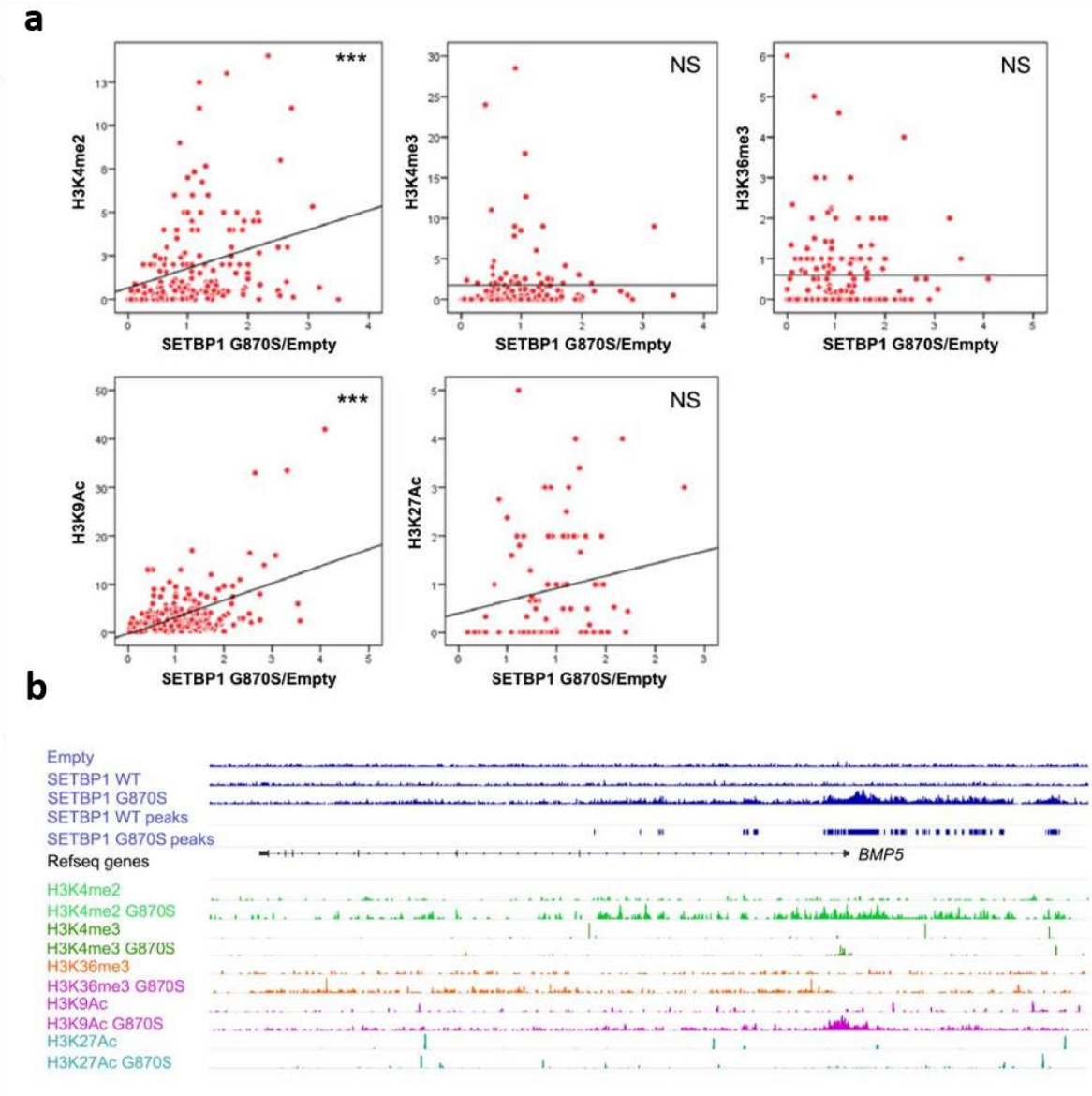


Figure 27 SETBP1-mediated epigenetic modulation.

A) The epigenetic changes resulting from the presence of SETBP1-G870S are expressed as function of SETBP1 differential DNA binding (G870S/Empty fold change in X axis) versus histone modification differential enrichment (G870S/Empty fold change in Y axis). B) SETBP1 ChIP-Seq coverage track and peak alignment to the hg19 reference genome are superimposed to the different histone methylation ChIP-Seq coverage tracks within the *BMP5* locus.

Using previously generated RNA-Seq data we checked the transcriptional profile of the subset of genes showing both SETBP1 binding to their promoter regions and the increase of the H3K4me2 and H3K9Ac marks. For this specific subset of genes, we found significant transcriptional upregulation, as shown in Figure 28 and Table 7. Together, these data suggest that

SETBP1 could be part of a nucleosome-remodeling complex, resulting in chromatin decondensation, and transcriptional activation.

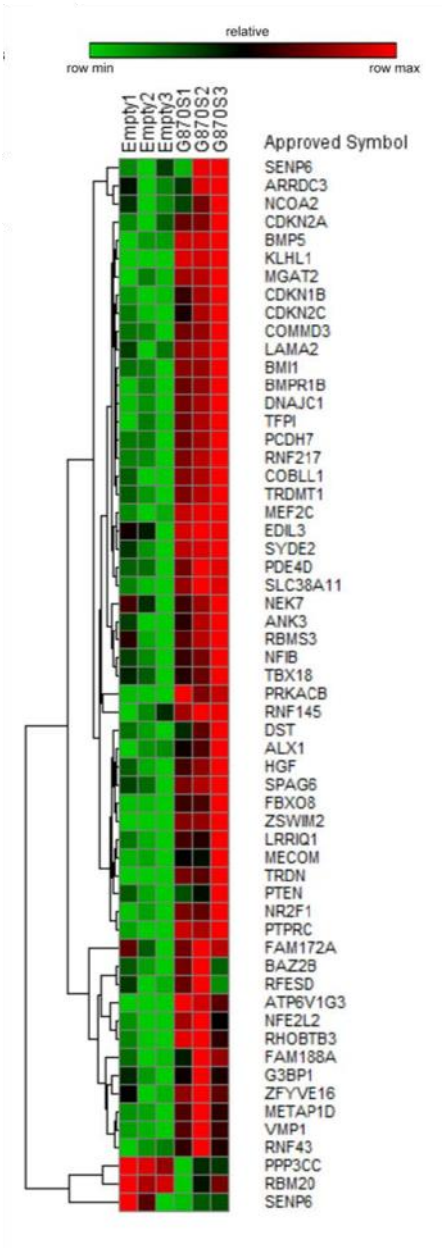


Figure 28 Differential expression of SETBP1-target presenting epigenetic modulation.

Gene expression heatmap of the subset of SETBP1 targets harboring increased H3K4me2 and H3K9ac activation marks.

Gene	log2FC ChIP-Seq SETBP1	log2FC ChIP-Seq H3K4me2	log2FC ChIP-Seq H3K4me3	log2FC ChIP-Seq H3K36me3	log2FC ChIP-Seq H3K9ac	log2FC ChIP-Seq H3K27ac	log2FC RNA-Seq
ALX1	4.54	1.43	1.85	0.55	4.66	3.00	1.89
ANK3	6.00	11.00		0.00	0.80		0.52
ARRDC3	2.29	0.88	9.00		1.07	0.50	0.91
ATP6V1G3	11.33	0.86	0.00	0.73	8.13	5.00	1.79
BAZ2B	2.37	1.99	0.51	0.33	1.63	3.25	0.92
BMI1	1.77	1.21	1.93		14.12	1.00	1.52
BMP5	8.26	0.65	0.15	1.29	13.04	5.38	6.14
BMPR1B	3.84	2.00	0.50	0.62	3.46	2.12	1.26
CDKN1B	1.70	0.95	0.92	1.03	3.52	2.50	0.76
CDKN2A	3.21	0.89	1.48	0.58	5.91		0.66
CDKN2C	1.33	2.92	1.17	0.16	2.43	0.50	0.63
COBLL1	3.76	2.55	0.48	2.06	4.89	8.12	1.07
COMMD3	2.73	1.37	4.08	2.83	5.36	1.52	1.84
DNAJC1	4.03	8.00		1.00	2.67		1.62
DST	1.52	0.61	0.44	0.50	6.08	1.50	0.71
DUSP16	1.44	0.58		0.25	1.70		0.36
EDIL3	2.50	1.36	1.56		2.79	2.15	1.67
FAM172A	2.70	2.33		0.00	0.96	1.00	0.38
FAM188A	3.33	1.00		0.00	14.00		0.66
FBXO8	2.96	3.52	0.02	2.03	6.56	0.38	1.58
G3BP1	2.20	15.00	0.50	0.25	1.29		0.35
HGF	8.43	1.01	0.50	2.75	5.63	3.00	2.65
KLHL1	10.79	3.00	0.13	0.50	3.88		5.40
LAMA2	7.31	1.60		0.00	2.75	9.00	2.50
LRRIQ1	8.48	3.46	0.30	0.98	5.08	0.81	2.17
MECOM	2.44	0.60	0.18	1.00	2.95	0.12	1.30
MEF2C	2.15	2.03	7.24	0.00	5.38		1.71
METAP1D	2.95	3.38	1.13	0.83	2.59	0.00	0.72
MGAT2	3.67	5.00			2.20		0.76
NCOA2	5.17	4.16	1.50	0.19	3.26	1.00	0.42
NEK7	2.83	1.00		0.00	1.17		0.52
NFE2L2	3.14	1.26	3.59	3.00	1.71	2.67	0.83
NFIB	2.05	2.75	0.17	0.00	0.83		0.44
NR2F1	6.25	5.00		1.25	5.86		2.12
PCDH7	4.40	0.80	0.34	0.53	2.80	0.00	1.06
PDE4D	3.31	1.95	2.52	0.79	2.90	1.19	1.02
PPP3CC	5.00	0.67		1.00	2.75	0.00	-0.87
PRKACB	4.53	2.67		0.17	2.91	0.00	0.91
PTEN	1.66	1.09	2.79	1.03	2.90		0.67
PTPRC	10.16	1.38	0.00	0.28	4.18	0.88	5.93
RBM20	2.43	1.29	0.54	1.00	0.79	1.00	-0.63
RBMS3	4.53	2.50	0.17	0.36	2.30	0.00	1.27

RFESD	3.57	2.25			3.32	0.00	0.62
RHOBTB3	7.00	0.72	3.26	2.44	2.40	3.38	1.48
RNF145	2.57	0.93	3.32	2.00	1.94	2.50	0.32
RNF217	4.09	2.55	0.13	0.98	4.90	2.50	1.67
RNF43	1.60	1.54	7.35	0.04	1.70	1.50	0.53
SENP6	3.44	0.65	0.54	0.80	0.97	0.55	-0.39
SLC38A11	3.44	1.87	0.25	0.71	5.20	6.67	2.68
SPAG6	3.69	1.22	1.83	0.67	1.92	3.50	2.33
SYDE2	3.00	4.67	3.00	0.25	3.09	2.50	1.10
TBX18	2.11	3.43	2.25	1.22	5.02		0.46
TFPI	3.72	4.04	0.96	0.98	5.89	1.91	1.63
TRDMT1	2.94	1.34	3.00	0.50	3.26		1.25
TRDN	8.85	1.46	1.69	0.47	8.12	1.23	1.31
VMP1	4.25	1.50	0.83	0.00	4.50	4.83	0.73
ZFYVE16	2.33	1.25		0.00	2.75		0.69
ZSWIM2	8.95	1.04	3.00	0.00	8.50	1.30	3.33

Table 7 List of the subset of SETBP1 targets presenting increased epigenetic marks.

Expression values from ChIP-Seq data and RNA-Seq data relative to the subset of genes targeted by SETBP1.

4.5 SETBP1 protein interacts with the SET1/KMT2A compass-like complex

To investigate about the possible mechanism of action explaining SETBP1-mediated gene regulation we performed a series of Co-IP experiments coupled with mass spectrometry (MS) in collaboration with the Meilahti Clinical Proteomics Core Facility, University of Helsinki, Helsinki, Finland. We performed three independent Co-IP experiments directed against the V5-tag in transiently transfected cells with either SETBP1-G870S or SETBP1wt cells. The same cells transfected with the Empty vector were used as control. One of the proteins identified by the MS was host cell factor C1 (HCF1) (Table 8).

Gene Symbol	Protein.Entry	Media di protein.MatchedProductIntenSum Empty	Media di protein.MatchedProductIntenSum MUT	Media di protein.MatchedProductIntenSum WT
PPP2R2B	2ABB_HUMAN	0	34518.5	27239
PPP2R2D	2ABD_HUMAN	0	44193	46254
ACAD9	ACAD9_HUMAN	0	17649.5	30453.5
ACTR1A	ACTZ_HUMAN	0	39299.5	35716

ADD3	ADDG_HUMAN	0	12518	81514.5
ADNP	ADNP_HUMAN	0	145452.7	114200.7
ADRM1	ADRM1_HUMAN	0	25020.5	28348
AKAP8L	AKP8L_HUMAN	0	62594.67	48711.33
ANP32B	AN32B_HUMAN	0	53586	63143
ANP32E	AN32E_HUMAN	0	49033.33	45989.33
ANXA6	ANXA6_HUMAN	0	37179.5	17636
APRT	APT_HUMAN	0	135554	54946.5
ARL1	ARL1_HUMAN	0	16284	12479
ARL2	ARL2_HUMAN	0	30248.5	29682.5
ASNA1	ASNA_HUMAN	0	27054	26361
ATP12A	AT12A_HUMAN	0	111153	28439
ATP1B3	AT1B3_HUMAN	0	15611	22288.5
ATG4B	ATG4B_HUMAN	0	14151	0
ACOT7	BACH_HUMAN	0	62862.5	41828
BCCIP	BCCIP_HUMAN	0	66662.33	43754.67
BLVRA	BIEA_HUMAN	0	36068.33	41127.5
BYSL	BYST_HUMAN	0	16142	97051.5
CAP2	CAP2_HUMAN	0	35220.67	41119.5
CAPZB	CAPZB_HUMAN	0	15442	20227.5
CAPZA2	CAZA2_HUMAN	0	49883.33	42859.67
C2orf47	CB047_HUMAN	0	41906	11165
CBX1	CBX1_HUMAN	0	409574	126236.7
CCAR1	CCAR1_HUMAN	0	48831	34598
CDC73	CDC73_HUMAN	0	93140	43815
CDK14	CDK14_HUMAN	0	29499	49607
CDK15	CDK15_HUMAN	0	35299	23706
CDK16	CDK16_HUMAN	0	36918.67	0
CDK4	CDK4_HUMAN	0	36375	0
CDK5	CDK5_HUMAN	0	38658	25714.5
CDK6	CDK6_HUMAN	0	37265	23706
CENPV	CENPV_HUMAN	0	223293	110889
CHAMP1	CHAP1_HUMAN	0	59181.5	19077
SLC25A12	CMC1_HUMAN	0	29099	46583.67
SLC25A13	CMC2_HUMAN	0	59202.67	42572.67
NCAPG	CND3_HUMAN	0	19225	15606
COIL	COIL_HUMAN	0	26087.33	23816
COPA	COPA_HUMAN	0	33238	45924.67
COPG1	COPG1_HUMAN	0	66115.67	59906.67
COPG2	COPG2_HUMAN	0	16994	13169.5
CPVL	CPVL_HUMAN	0	143321	154640.7
CSNK2A3	CSK23_HUMAN	0	89162.67	82268.67
CSNK2B	CSK2B_HUMAN	0	27697.67	31295.5
COPS4	CSN4_HUMAN	0	27377.67	29388
CUL1	CUL1_HUMAN	0	34313	48277.67

CUL4A	CUL4A_HUMAN	0	26229.67	247991
CUL5	CUL5_HUMAN	0	29225	15372
DNAAF5	DAAF5_HUMAN	0	14972	19670.5
DCTN5	DCTN5_HUMAN	0	68310	54533
DDX20	DDX20_HUMAN	0	22475	30205
DNAJC10	DJC10_HUMAN	0	104587	52456
DYNLRB1	DLRB1_HUMAN	0	43935	21922.5
DNAJB1	DNJB1_HUMAN	0	33411	21719
DRG1	DRG1_HUMAN	0	34441	18907
DYNLL1	DYL1_HUMAN	0	48660.5	32133.5
DNM2	DYN2_HUMAN	0	42951.67	49636.67
DNM3	DYN3_HUMAN	0	33443	43378
DHFR	DYR_HUMAN	0	38607	37427
EIF2AK2	E2AK2_HUMAN	0	39281.33	43143.33
EPB41L3	E41L3_HUMAN	0	32290	20623
GFM1	EFGM_HUMAN	0	26477	25837
EHD1	EHD1_HUMAN	0	33213	21786
EHD4	EHD4_HUMAN	0	28222.5	20398
EIF2B2	EI2BB_HUMAN	0	38343	24351
EIF2B5	EI2BE_HUMAN	0	19102.5	14591
EIF3F	EIF3F_HUMAN	0	66463	72590.33
EIF3K	EIF3K_HUMAN	0	27447.5	56218
HSP90B2P	ENPLL_HUMAN	0	54060	44164
ESF1	ESF1_HUMAN	0	36409.5	38305
ESYT1	ESYT1_HUMAN	0	70373.67	56647.67
G3BP2	G3BP2_HUMAN	0	168594.3	138683.7
ACBD3	GCP60_HUMAN	0	23503	17755.33
GET4	GET4_HUMAN	0	17635.5	10775
GRB2	GRB2_HUMAN	0	29015	23110
GSTK1	GSTK1_HUMAN	0	12012	18411.5
GTPBP1	GTPB1_HUMAN	0	194264	49582
HCFC1	HCFC1_HUMAN	0	57672.67	42486.67
HDAC2	HDAC2_HUMAN	0	197618.7	174321.7
HDDC2	HDDC2_HUMAN	0	15691.5	6511
HMGA1	HMGA1_HUMAN	0	25990	0
HMGB2	HMGB2_HUMAN	0	262355.7	160423
IDH3B	IDH3B_HUMAN	0	83023	35244
IDH2	IDHP_HUMAN	0	28944	38505.5
KPNA4	IMA3_HUMAN	0	475983.7	362478.3
ISOC2	ISOC2_HUMAN	0	39131	22851.5
IWS1	IWS1_HUMAN	0	27411	0
CMPK1	KCY_HUMAN	0	32120	20148.5
KHDRBS3	KHDR3_HUMAN	0	34504	36698
KIF5C	KIF5C_HUMAN	0	35450.33	31719.33

RPS6KA1	KS6A1_HUMAN	0	19261	17830.5
RPS6KA3	KS6A3_HUMAN	0	44949.5	42736
LARP4	LARP4_HUMAN	0	41221	34904
LMNA	LMNA_HUMAN	0	51807.5	32888
LRRC40	LRC40_HUMAN	0	36817	24832
MAK16	MAK16_HUMAN	0	41801.33	30904
MAPRE1	MARE1_HUMAN	0	40452.5	28347
MBD3	MBD3_HUMAN	0	119150.7	64399.67
MAT2A	METK2_HUMAN	0	30589	0
MAPK1	MK01_HUMAN	0	28711	18738
MAP2K2	MP2K2_HUMAN	0	35574.5	41571.5
MRE11A	MRE11_HUMAN	0	34294	22453
NAA11	NAA11_HUMAN	0	38576.5	41714
NACAP1	NACP1_HUMAN	0	34408	23094.67
NDUFA8	NDUA8_HUMAN	0	24336	16597
NDUFA12	NDUAC_HUMAN	0	16165	17828.5
NDUFA13	NDUAD_HUMAN	0	34654	72502
NDUFS2	NDUS2_HUMAN	0	33669	34882.67
NDUFS4	NDUS4_HUMAN	0	28610.5	35662
NDUFS6	NDUS6_HUMAN	0	12960	20178.5
NDUFS8	NDUS8_HUMAN	0	45389	28231.5
NDUFV1	NDUV1_HUMAN	0	40284.33	43691.33
NDUFV2	NDUV2_HUMAN	0	46345.67	40991.33
NELFB	NELFB_HUMAN	0	37411	30740.33
NKRF	NKRF_HUMAN	0	27803	0
NOC4L	NOC4L_HUMAN	0	20804	0
NUP133	NU133_HUMAN	0	87598	35054
NUP214	NU214_HUMAN	0	80587.5	31398.33
NUMA1	NUMA1_HUMAN	0	98116.67	157543.3
NUP35	NUP53_HUMAN	0	25005.5	14711
NVL	NVL_HUMAN	0	27566	18790.5
NXF1	NXF1_HUMAN	0	58531.5	75569
DBT	ODB2_HUMAN	0	45386	51577.33
PDHA2	ODPAT_HUMAN	0	57024	27250
OPA1	OPA1_HUMAN	0	13503	57271.5
LEPRE1	P3H1_HUMAN	0	37970.5	7181
P4HA1	P4HA1_HUMAN	0	79100.67	60444.67
TP53	P53_HUMAN	0	273607	176178.5
PYCR1	P5CR1_HUMAN	0	38761	52192.33
PYCR2	P5CR2_HUMAN	0	47100	23980

GATAD2A	P66A_HUMAN	0	184042.5	98523
GATAD2B	P66B_HUMAN	0	89158	57718
PAF1	PAF1_HUMAN	0	34159	29818.33
PPY	PAHO_HUMAN	0	35896	39905
PAK2	PAK2_HUMAN	0	101066	43110.5
PDCD6	PDCD6_HUMAN	0	134649.7	126505
POLDIP3	PDIP3_HUMAN	0	30790	34309.5
PFKM	PFKAM_HUMAN	0	32823	30888
PTGES2	PGES2_HUMAN	0	24407.5	16136
PLRG1	PLRG1_HUMAN	0	25610	20621.33
PPP1CA	PP1A_HUMAN	0	33361.33	36377.67
PPP1CB	PP1B_HUMAN	0	35082.67	42514.5
PPP1CC	PP1G_HUMAN	0	30676.67	45741.5
PPIH	PPIH_HUMAN	0	42015.33	40755.5
PPT1	PPT1_HUMAN	0	72455	77588.67
PFN2	PROF2_HUMAN	0	42171.5	52104.5
PRPF31	PRP31_HUMAN	0	71377.5	44086.33
PRPS1L1	PRPS3_HUMAN	0	59074	36103.5
PSMC4	PRS6B_HUMAN	0	239264.3	187698
PSMA8	PSA7L_HUMAN	0	19610	29880.5
PSMD10	PSD10_HUMAN	0	40169.5	34401.33
PSMD7	PSMD7_HUMAN	0	49548	13067
PURA	PURA_HUMAN	0	27861.67	46308.5
CTPS2	PYRG2_HUMAN	0	28991	0
RAE1	RAE1L_HUMAN	0	62869.67	114692
RBM15	RBM15_HUMAN	0	79678.33	142833.7
RCC1	RCC1_HUMAN	0	56584.67	33664.67
RCN1	RCN1_HUMAN	0	52128	25904
UPF1	RENT1_HUMAN	0	44278.33	49734.33
RBP1	RET1_HUMAN	0	25967	109807
RFC4	RFC4_HUMAN	0	15068	19456.5
RBFOX1	RFOX1_HUMAN	0	22247.5	15636
RBFOX2	RFOX2_HUMAN	0	18371	15636
RHEB	RHEB_HUMAN	0	33501	32987.67
RHOA	RHOA_HUMAN	0	23573	0
RPL38	RL38_HUMAN	0	52036.5	62030
RPLP1	RLA1_HUMAN	0	460310	405468
MRPL23	RM23_HUMAN	0	27451	35009.5
MRPL24	RM24_HUMAN	0	21669	19182
MRPL37	RM37_HUMAN	0	24206	30729
MRPL41	RM41_HUMAN	0	27048.5	23950
POLR2E	RPAB1_HUMAN	0	35816.5	28396
POLR1C	RPAC1_HUMAN	0	50641.67	60380.67
POLR2C	RPB3_HUMAN	0	72516.33	63421.33

RPRD1A	RPR1A_HUMAN	0	36348	22676.33
RPRD1B	RPR1B_HUMAN	0	109241	61064
RPS10	RS10_HUMAN	0	85562.5	56404
RPS13	RS13_HUMAN	0	30625	41142.67
RPS23	RS23_HUMAN	0	37522	43213
RPS27L	RS27L_HUMAN	0	85147	155312.5
RPS29	RS29_HUMAN	0	47446	17765.5
MRPS10	RT10_HUMAN	0	19970	18521.5
MRPS27	RT27_HUMAN	0	39253	49939
SAMHD1	SAMH1_HUMAN	0	33612	107355.3
SEC23A	SC23A_HUMAN	0	42050	48129
SEC31A	SC31A_HUMAN	0	21573	27512
SCAMP3	SCAM3_HUMAN	0	40497.5	24641.5
SCFD1	SCFD1_HUMAN	0	24026.33	27084.5
SEC63	SEC63_HUMAN	0	36870.5	29680.67
SEPT2	SEPT2_HUMAN	0	53777	59972
SEPT6	SEPT6_HUMAN	0	36177	14647
SF3A2	SF3A2_HUMAN	0	16734	28927.5
SIN3A	SIN3A_HUMAN	0	71121.5	0
SMAP	SMAP_HUMAN	0	34876	31009.5
SMCHD1	SMHD1_HUMAN	0	47719.67	31442
SMARCC1	SMRC1_HUMAN	0	55960.5	19205
SMARCAD1	SMRCD_HUMAN	0	27322	14443
SNRNP40	SNR40_HUMAN	0	29432	31099.67
SNX9	SNX9_HUMAN	0	33676	38258
SPCS2	SPCS2_HUMAN	0	24951	69389
SUPT5H	SPT5H_HUMAN	0	78555	60933.67
SUPT6H	SPT6H_HUMAN	0	51140.5	13670
SRP9	SRP09_HUMAN	0	19117	35743
SRSF11	SRS11_HUMAN	0	40751	22372.5
SSR4	SSRD_HUMAN	0	53953	41159.33
STAT1	STAT1_HUMAN	0	17270	12159
STK24	STK24_HUMAN	0	24593.33	25880
STK25	STK25_HUMAN	0	28519	19068.5
STK4	STK4_HUMAN	0	190411.5	0
QARS	SYQ_HUMAN	0	33392.67	58435.5
VCP	TERA_HUMAN	0	33229.5	69513.5
TIMM23	TIM23_HUMAN	0	20785	21916
TIMM50	TIM50_HUMAN	0	66775	55585
TMED2	TMED2_HUMAN	0	39877	28299.67
TMED5	TMED5_HUMAN	0	29876	29872.5

TNPO2	TNPO2_HUMAN	0	26352	35457.67
TOMM22	TOM22_HUMAN	0	57624.67	60895.67
TPD52L2	TPD54_HUMAN	0	27798	53261
TRMT5	TRM5_HUMAN	0	31176.33	25916.67
PUS1	TRUA_HUMAN	0	38906.5	37841.33
TWF1	TWF1_HUMAN	0	41211	28254
YY1	TYY1_HUMAN	0	16243	8736
UBE2I	UBC9_HUMAN	0	49000.5	50360
UBE2O	UBE2O_HUMAN	0	41909.5	20521
UBE2S	UBE2S_HUMAN	0	20328	23054.5
USP7	UBP7_HUMAN	0	160245.3	115961
UQCRFS1	UCRI_HUMAN	0	13877	19086
UBE2NL	UE2NL_HUMAN	0	22957	29527.5
UHRF1	UHRF1_HUMAN	0	160309	112772.3
NAE1	ULA1_HUMAN	0	37714	18966
USMG5	USMG5_HUMAN	0	27269.5	29085
UTP18	UTP18_HUMAN	0	38445.5	35482
VAPA	VAPA_HUMAN	0	123915	37359
VAPB	VAPB_HUMAN	0	34012.5	35993
ATP6V1A	VATA_HUMAN	0	33490	23443.33
HDLBP	VIGLN_HUMAN	0	101696	71784.33
WDR11	WDR11_HUMAN	0	32027	42109.5
YME1L1	YMEL1_HUMAN	0	37261	23790
YTHDC1	YTDC1_HUMAN	0	24557.5	7796
YTHDF1	YTHD1_HUMAN	0	31395.33	29838.33
ZC3H18	ZCH18_HUMAN	0	86692	32142
ZNF622	ZN622_HUMAN	0	27830	34286.67
SETBP1	SETBP_HUMAN	103020	25969045	24745688
KPNA3	IMA4_HUMAN	7395	596893.3	465424.3
SET	SET_HUMAN	28399	1164562	820881.7
MCM4	MCM4_HUMAN	17307.67	674226.7	466937.3
CHORDC1	CHRD1_HUMAN	6657	212475.7	170643.3
SETSIP	SETLP_HUMAN	26685.67	769510.7	613111
TUBB6	TBB6_HUMAN	37418.33	1057519	1090478
TUBB4A	TBB4A_HUMAN	113322	2679126	2348520
TUBB3	TBB3_HUMAN	146719.7	3289254	3002798
SMC3	SMC3_HUMAN	28676.67	609198.7	341852.3
RALY	RALY_HUMAN	8773	183651	100136
WDR12	WDR12_HUMAN	6979.5	138395.7	77276.33
-	TBB8L_HUMAN	17093.33	320519.7	455015.7
RPS20	RS20_HUMAN	14197	262316	246463.5
RPN2	RPN2_HUMAN	21403.5	392529	310402.3
NDUFS3	NDUS3_HUMAN	10014	183211	170187

EIF3M	EIF3M_HUMAN	7378	134915	107477
ARCN1	COPD_HUMAN	9809.667	171563	164993
MTA2	MTA2_HUMAN	6485	108143	41358.5
CAP1	CAP1_HUMAN	26690.5	441160	332758
EPB41	41_HUMAN	7279	119324	21787
CDK1	CDK1_HUMAN	21502	343963	204345.3
MYH11	MYH11_HUMAN	14000	223753.5	38859
COPE	COPE_HUMAN	8480	135449.7	143307.3
DDX39A	DX39A_HUMAN	75008.67	1171038	410365.3
SRM	SPEE_HUMAN	7359	114249.3	77989.33
AHSA1	AHSA1_HUMAN	26268	402230.3	339827.3
PSMC3	PRS6A_HUMAN	13047.33	195904.3	201107
RCN2	RCN2_HUMAN	17910	268421	203537.3
SMC1A	SMC1A_HUMAN	25862.67	381480.7	201400
PRPF40A	PR40A_HUMAN	9401	138498.7	95363.33
RSF1	RSF1_HUMAN	17108	251562.7	31540.67
INA	AINX_HUMAN	17438	254457	181512
ZNF326	ZN326_HUMAN	21156.33	308585.3	252449
KPNA6	IMA7_HUMAN	7424	107948.3	72302
CCT3	TCPG_HUMAN	89151.33	1296184	1036694
NEFH	NFH_HUMAN	13689.67	198684.3	151161.3
KPNA1	IMA5_HUMAN	8810	125402.7	86430.67
NEFL	NFL_HUMAN	36570	519723	436757.7
VIM	VIME_HUMAN	460187.7	6524796	5103435
PGD	6PGD_HUMAN	6094	85357.5	56715
MCM7	MCM7_HUMAN	26374.33	364277.7	285215
MCM3	MCM3_HUMAN	26262.67	360998	256137.7
PPP2R2A	2ABA_HUMAN	8374	113006.3	78550
RAB1B	RAB1B_HUMAN	30073.67	393072.7	225094
HNRNPUL2	HNRL2_HUMAN	11266	145849.7	78725
KPNA5	IMA6_HUMAN	7424	95007.67	70356.33
DNAJA3	DNJA3_HUMAN	13144	167456.3	186161
NEFM	NFM_HUMAN	63497.33	807310	596953
TUBB	TBB5_HUMAN	416738.3	5297060	4871386
HMGB1	HMGB1_HUMAN	30376	384873	210761.3
RPS14	RS14_HUMAN	31642.33	399372.7	652751.7
EIF6	IF6_HUMAN	6957	87112.67	58435
TUBB4B	TBB4B_HUMAN	361318	4479203	3966326
VDAC3	VDAC3_HUMAN	36786.67	453254.7	560883.7
ALDH18A1	P5CS_HUMAN	17930	220731.7	181454.7
RAB1A	RAB1A_HUMAN	29554	353240	200531
XRN2	XRN2_HUMAN	9594.5	114074.7	21891

MCM5	MCM5_HUMAN	20088.33	237466	212279
TMED10	TMEDA_HUMAN	7085	82189.67	66863.33
SRRT	SRRT_HUMAN	20211.33	232390	134934.7
CBX3	CBX3_HUMAN	44015.5	504922	228672.7
WDR43	WDR43_HUMAN	9765	110892.3	26308.5
TUBA4A	TBA4A_HUMAN	288192	3256873	2860694
TUBB2B	TBB2B_HUMAN	322194.7	3625785	3352666
TUBB2A	TBB2A_HUMAN	321919.3	3601421	3330060
ATP5B	ATPB_HUMAN	180281.7	2012583	1736336
TUBB1	TBB1_HUMAN	20103.67	224056.7	295019.7
TCP1	TCPA_HUMAN	111310.3	1234833	1067342
NAP1L1	NP1L1_HUMAN	47871	530906.7	674023
LTV1	LTV1_HUMAN	6809	75193.33	55739.67
NAA10	NAA10_HUMAN	7863	86492.67	94496.67
EIF4A2	IF4A2_HUMAN	95098.33	1043098	1116545
SF3B3	SF3B3_HUMAN	21567.5	233309.3	210200
C1QBP	C1QBP_HUMAN	282347	2996427	2855408
HNRNPF	HNRPF_HUMAN	142640	1489857	1428802
DNTTIP2	TDIF2_HUMAN	17989	185188.3	78489
TUBA8	TBA8_HUMAN	213945.7	2187442	2061742
PHB2	PHB2_HUMAN	117542.3	1198514	1516679
FAM98A	FA98A_HUMAN	13338	134792.7	89926.33
MAP4	MAP4_HUMAN	6724	67677	41943
NAMPT	NAMPT_HUMAN	11933	119968	130214.7
FARSB	SYFB_HUMAN	9434	93771.5	50319.67
TARDBP	TADBP_HUMAN	16026.33	153372.7	167197.7
KPNA2	IMA1_HUMAN	38243	362039.7	320845
EIF3CL	EIFCL_HUMAN	23472.33	221429.5	83983.33
NHP2L1	NH2L1_HUMAN	14170	133571	93825.33
EIF3C	EIF3C_HUMAN	23554.67	221589	83983.33
CCT6A	TCPZ_HUMAN	160872.3	1513278	1054051
CSNK2A1	CSK21_HUMAN	14666.67	137734.7	145479.3
TRAP1	TRAP1_HUMAN	62525.67	580330.7	454350.7
NDUFS1	NDUS1_HUMAN	17186.33	158923.3	148132
DNAJA1	DNJA1_HUMAN	19856.67	183353.7	150629
SNX1	SNX1_HUMAN	7129	65671.5	12490
MCM6	MCM6_HUMAN	25506.67	234341	137320
USO1	USO1_HUMAN	11081.5	101742	65011
DES	DESM_HUMAN	33628.33	306625	257971.3
HMGB1P1	HGB1A_HUMAN	30862.5	280876.7	150403
G3BP1	G3BP1_HUMAN	18084	163401.7	170409.3
FH	FUMH_HUMAN	7698	69516.5	62441
ETF1	ERF1_HUMAN	19675.33	176229.3	147720.3

CBX5	CBX5_HUMAN	18709	167460	116918
TOMM40	TOM40_HUMAN	19297.33	172714	162294
CCT4	TCPD_HUMAN	82809	739153.7	543193
DDX1	DDX1_HUMAN	65594.67	583683.3	486512.3
FUBP3	FUBP3_HUMAN	12281	108683.5	62862.5
EIF3B	EIF3B_HUMAN	21640.33	191226.3	219196
SF3B6	SF3B6_HUMAN	8020	70484.67	76137
NDUFA5	NDUA5_HUMAN	8468	74403.33	65944.67
DDX3Y	DDX3Y_HUMAN	38488.67	336858.3	360176
PSMD1	PSMD1_HUMAN	19650	170896.3	153861.7
PSMC5	PRS8_HUMAN	30999.67	269034.3	240109.7
NUP155	NU155_HUMAN	8391	72746.33	61156.67
AATF	AATF_HUMAN	11596	100426.5	48696.67
CCT5	TCPE_HUMAN	60553	521542.3	481824
RPLP0	RLA0_HUMAN	143877	1232469	959791
SND1	SND1_HUMAN	13721.67	116644.7	100397
RPS6KA1	PSPC1_HUMAN	14757	125091	78143.67
TFAM	TFAM_HUMAN	10596	89747.5	47415
CCT2	TCPB_HUMAN	67682.67	571970.7	525201
PSMC1	PRS4_HUMAN	8689	73273.33	88798
ATP5A1	ATPA_HUMAN	264114.3	2218192	2158893
DDX3X	DDX3X_HUMAN	46545	390080	426583
WARS	SYWC_HUMAN	9097.5	76204	71565
TUBB8	TBB8_HUMAN	73458.67	614014	635478
RPS17L	RS17L_HUMAN	9002	74913	93036.5
IGF2BP1	IF2B1_HUMAN	83259	691118	525596.7
RBBP4	RBBP4_HUMAN	109211.7	904045.7	526647.7
RBM14	RBM14_HUMAN	46282.33	381379	390591.7
RPS17	RS17_HUMAN	9002	73991.5	73585
ARF5	ARF5_HUMAN	12707	103922.7	100092.3
SRSF10	SRS10_HUMAN	10109	82659	37146
HSPA8	HSP7C_HUMAN	470063.7	3840920	3013178
RPN1	RPN1_HUMAN	47170	384849.7	343322.3
DNAJA2	DNJA2_HUMAN	40950.33	333297.3	252886.3
PSMD14	PSDE_HUMAN	15733.5	126169.3	103041.3
PGAM5	PGAM5_HUMAN	7268	58036	40920
CFL2	COF2_HUMAN	32690	260711.7	227961.7
SSRP1	SSRP1_HUMAN	118708.3	942419.7	581927
HSPA1L	HS71L_HUMAN	441382.3	3492078	2993229
RPS3	RS3_HUMAN	178690.7	1397987	1285411
CLTC	CLH1_HUMAN	51274.33	397116.3	401136.3
VDAC1	VDAC1_HUMAN	167905.3	1296826	1542011

ILF2	ILF2_HUMAN	38481	297084	283187.7
VDAC2	VDAC2_HUMAN	114436.3	882995.7	1047095
CPSF7	CPSF7_HUMAN	9381.5	72382.33	64377.33
PPIB	PPIB_HUMAN	20492	155594	51445
GAPDH	G3P_HUMAN	32178	243635	195108.3
SLC1A5	AAAT_HUMAN	13646.33	102968.7	92097.33
IMPDH2	IMDH2_HUMAN	20876	157499.3	136699.3
TRIM28	TIF1B_HUMAN	62264	468848	303788.7
GARS	SYG_HUMAN	9715	72854	73313.33
HSPA2	HSP72_HUMAN	246751.7	1841621	1461748
IGF2BP2	IF2B2_HUMAN	20216.33	150602.7	117043.3
HSDL2	HSDL2_HUMAN	15979	118995	145849
EIF4A1	IF4A1_HUMAN	267156	1986039	2034221
HSP90B1	ENPL_HUMAN	119617	888393	714734.7
ARF1	ARF1_HUMAN	28892.33	213351.3	191345.3
DDX39B	DX39B_HUMAN	88230	651284.3	491823.3
ARF3	ARF3_HUMAN	28892.33	212738	191345.3
EPB41L2	E41L2_HUMAN	12978.67	95532	81738.33
FARSA	SYFA_HUMAN	8458	62145	63965.33
EEF2	EF2_HUMAN	76632.33	562800.7	481426.3
RAB1C	RAB1C_HUMAN	26883.33	197058.7	167279
HSPA6	HSP76_HUMAN	209311.3	1531472	1518667
DNAJB11	DJB11_HUMAN	9286.5	67852	37689.5
DHX9	DHX9_HUMAN	192431	1405447	989814.7
WDR5	WDR5_HUMAN	8295	60402.5	64059.33
CFL1	COF1_HUMAN	82435	600093.7	651587
UQCRC2	QCR2_HUMAN	22627.67	164289.7	119722
HYOU1	HYOU1_HUMAN	10831.5	78636	84435
AGPS	ADAS_HUMAN	8756	63525.33	53618.67
SFXN1	SFXN1_HUMAN	20914.33	149519.3	117870.3
CALU	CALU_HUMAN	7262	51580	57690
AGK	AGK_HUMAN	12780	90479	98866
IGF2BP3	IF2B3_HUMAN	24128	170552	120170
PDHB	ODPB_HUMAN	52515.67	367533	371512.3
PARP1	PARP1_HUMAN	139819.3	968922.7	858850.3
TUFM	EFTU_HUMAN	74319.33	514910	387207.7
RPL9P7	RL9_HUMAN	38402	265958	415757.7
LUC7L3	LC7L3_HUMAN	12070.5	83242	50897.67
MCM2	MCM2_HUMAN	29209.67	201273	234271
GCN1L1	GCN1L_HUMAN	12550	86441	48795
AIFM1	AIFM1_HUMAN	19031.5	131061	191203
TUBA1B	TBA1B_HUMAN	988442	6776185	6015180
GRWD1	GRWD1_HUMAN	22318	152568	144964
CTPS1	PYRG1_HUMAN	14933.67	101837.3	98273
NAP1L4	NP1L4_HUMAN	18725	127403.3	137517.7

RPS5	RS5_HUMAN	13116.5	89118.5	145475
HSPA1A	HS71A_HUMAN	1131622	7686152	6547003
HSPA1B	HS71B_HUMAN	1131622	7686152	6546831
NACA	NACAM_HUMAN	60779.67	412553	342302
NUDC	NUDC_HUMAN	42316.67	286988	252928
SEPT9	SEPT9_HUMAN	7889	53424.33	47409.67
XRCC6	XRCC6_HUMAN	136934	924945.3	841144.3
FASN	FAS_HUMAN	116865	787378	1152199
TUBA1A	TBA1A_HUMAN	988534	6653296	5913187
NPEPPS	PSA_HUMAN	14735	98698.67	85511.67
RAN	RAN_HUMAN	240791.3	1611634	1432128
CIRH1A	CIR1A_HUMAN	9483	63247.5	70686
EIF2S3	IF2G_HUMAN	30533.33	203182.7	185932.7
NSUN2	NSUN2_HUMAN	21010.67	139546.7	109122.3
CCT8	TCPQ_HUMAN	72748.67	481911.7	478248.7
RBBP7	RBBP7_HUMAN	112460	744584.7	488115
XRCC5	XRCC5_HUMAN	110960.7	734092	568462.3
VPS26A	VP26A_HUMAN	11094	73378	40106.33
ARF4	ARF4_HUMAN	14934	98732.33	100505.3
RUVBL1	RUVB1_HUMAN	59711.33	393785.7	363209.3
STUB1	CHIP_HUMAN	22286.5	146726.7	87787.33
COPB2	COPB2_HUMAN	9920.5	65116	54718.67
PDCD6IP	PDC6I_HUMAN	32777.33	214624.3	209599
TUBA1C	TBA1C_HUMAN	985853.7	6453417	5749653
EIF3G	EIF3G_HUMAN	59877.67	388095.3	437371.7
MT-CO2	COX2_HUMAN	37491.5	241635	196901.7
COPB1	COPB_HUMAN	12647	81375.33	76833.67
U2AF2	U2AF2_HUMAN	17055	109601	62897
RPS27	RS27_HUMAN	14431	91582	129027.3
KCTD12	KCD12_HUMAN	10629.5	67405.33	46899.33
TRA2B	TRA2B_HUMAN	22205	140657.7	163990.3
PCBP1	PCBP1_HUMAN	84055.67	526328	538596.3
PCBP2	PCBP2_HUMAN	95437.33	597592.7	636273.3
NACA	NACA_HUMAN	53524	334605.7	279964.7
COX7A2	CX7A2_HUMAN	30563	190915.3	215922.3
HSPA4	HSP74_HUMAN	28230	176151.7	29178.33
PKM	KPYM_HUMAN	231066.3	1440473	1427279
RPS16	RS16_HUMAN	22229	138497	223610.3
LONP1	LONM_HUMAN	9415	58604.5	52006
DDX5	DDX5_HUMAN	209865.7	1302081	1241287
RAB8A	RAB8A_HUMAN	24142.5	149750	32510
HNRNPM	HNRPM_HUMAN	315075.3	1939453	1740864
CDC5L	CDC5L_HUMAN	16080.5	98884	56649.33
HNRNPA0	ROA0_HUMAN	87275.33	536456	663178
GART	PUR2_HUMAN	41569.33	254389.3	186071.3

DCTN1	DCTN1_HUMAN	10569	64650	69575
RPS7	RS7_HUMAN	67236	410187.7	513911.3
HNRNPH2	HNRH2_HUMAN	295905	1798988	1723662
PRMT5	ANM5_HUMAN	9515.5	57774.33	61433.33
ANP32A	AN32A_HUMAN	16074	96997.33	99456.67
TALDO1	TALDO_HUMAN	17318	104464.5	103362.5
LARS	SYLC_HUMAN	34419	207166	203405
PCBP3	PCBP3_HUMAN	56519.67	339933.3	319404
OAT	OAT_HUMAN	21188.5	127090	118368.7
HNRNPH1	HNRH1_HUMAN	459006.3	2745329	2629967
SRSF3	SRSF3_HUMAN	117740	702900.7	593200.7
ATAD3B	ATD3B_HUMAN	36717.67	218959.7	177090
TOP1	TOP1_HUMAN	16789.5	99922	61520
EPRS	SYEP_HUMAN	41187.67	244929.3	203316
RPA3	RFA3_HUMAN	9262	54847	29502
DIS3	RRP44_HUMAN	15349.5	90714.67	88869
RPS15A	RS15A_HUMAN	30085	177640.3	205729
ALDH1B1	AL1B1_HUMAN	15593	91699.67	77121.67
EIF2S1	IF2A_HUMAN	12935	76036.33	67900
DDX17	DDX17_HUMAN	204431	1195133	1102195
EIF3D	EIF3D_HUMAN	17691.5	102796.7	110518.3
EIF2S3L	IF2GL_HUMAN	21733	126122.3	122085.7
TIMM44	TIM44_HUMAN	11915.5	68941.67	78922.33
PFKL	PFKAL_HUMAN	8951	51335.67	32381
ERLIN2	ERLN2_HUMAN	15629.33	89611.33	79802.33
RAB6A	RAB6A_HUMAN	10305.33	59028.5	64305
SNRPA1	RU2A_HUMAN	22638	129619	129013.7
ELAVL1	ELAV1_HUMAN	44547	254054.7	232660.7
NDUFA10	NDUAA_HUMAN	14082	80259.67	22696
RPLP0P6	RLA0L_HUMAN	148732.3	846174	679273.7
SLC25A31	ADT4_HUMAN	38216.5	216752	132068.5
ATAD3A	ATD3A_HUMAN	55068.33	312169.7	298123
TUBA3E	TBA3E_HUMAN	823690.7	4664854	4050410
TUBA3D	TBA3C_HUMAN	895254.7	5060104	4371220
ETFA	ETFA_HUMAN	13863.5	78011.67	61692.33
HSPA7	HSP77_HUMAN	158767.3	889067.3	897240.7
PSMC2	PRS7_HUMAN	28506.33	158611.7	172467
RPLP2	RLA2_HUMAN	158404	878215.7	687998.3
ECM29	ECM29_HUMAN	12567	69094.33	74313.67
HADHB	ECHB_HUMAN	25308.67	138485	139044
TBL3	TBL3_HUMAN	19480	106387.7	60309
ACIN1	ACINU_HUMAN	29951.67	161822.3	99167.67
SAE1	SAE1_HUMAN	13670	73811.5	50971.33
LDHA	LDHA_HUMAN	143089.3	768642.7	640619

MARS	SYMC_HUMAN	11073.5	59179.67	50290.33
ERLIN1	ERLN1_HUMAN	6076	32042.67	43219
EIF4A3	IF4A3_HUMAN	130431.3	684419.7	654779.3
PRPS1	PRPS1_HUMAN	13559.5	70550.67	56194.33
HNRNPU	HNRPU_HUMAN	356581	1854588	1512780
KARS	SYK_HUMAN	12058.5	62593.67	67271
RPL30	RL30_HUMAN	24252	125353.5	163665.5
LMNB2	LMNB2_HUMAN	48516.33	250408.3	108632.3
RAB5C	RAB5C_HUMAN	16014.67	82279.33	79850.33
HSP90AB1	HS90B_HUMAN	964529	4941199	4021460
SUPT16H	SP16H_HUMAN	293363	1497142	944317.3
KPNB1	IMB1_HUMAN	37786.67	192147.7	211405
RUVBL2	RUVB2_HUMAN	101299.3	514706	527238
MATR3	MATR3_HUMAN	101319.7	514092	426884.3
IARS	SYIC_HUMAN	31163.33	157951	172285.7
PPP2R1A	2AAA_HUMAN	13267	67201.67	82974
CDC37	CDC37_HUMAN	25956.5	130810	90350
AP2B1	AP2B1_HUMAN	7966	40104.33	46439.67
HNRNPK	HNRPK_HUMAN	132107.7	664086.3	649015.3
YTHDF2	YTHD2_HUMAN	11087	55566.33	61393.33
BCLAF1	BCLF1_HUMAN	57092.33	286057	133810.7
MSH6	MSH6_HUMAN	8562	42888	27676
RPS19	RS19_HUMAN	56029.67	280400.5	179327.7
DIAPH1	DIAP1_HUMAN	6976	34841	40607.5
SPATA5	SPAT5_HUMAN	8156	40448.33	41047
CCT7	TCPH_HUMAN	73950.33	365867	390763
RPL21	RL21_HUMAN	23004.5	113739.7	117563.7
IPO5	IPO5_HUMAN	25364.33	124227	170921.7
SERPINH1	SERPH_HUMAN	7347	34848.33	71252
RPL22	RL22_HUMAN	76655	363323.3	651981.7
IMMT	MIC60_HUMAN	73852.33	349630.7	402059.3
RTCB	RTCB_HUMAN	68967.67	324486.3	353202.3
LDHC	LDHC_HUMAN	25119	116354.5	227148
PHB	PHB_HUMAN	185272	853764.7	1086412
EIF3A	EIF3A_HUMAN	28798	131727	212464
ATP1A4	AT1A4_HUMAN	7421	33621	39617
PTBP1	PTBP1_HUMAN	66063	295019.7	423028
EFTUD2	U5S1_HUMAN	37414.67	165017.7	187216
TSFM	EFTS_HUMAN	6441	27970	50268.5
EIF2B3	EI2BG_HUMAN	6537	27821	36304
EIF3E	EIF3E_HUMAN	20913	88046	105845
DDX21	DDX21_HUMAN	87496.67	366892.7	597117
NUP210	PO210_HUMAN	6465	25845.5	55217

SRP54	SRP54_HUMAN	14753.5	58296	94428.67
NCAPH	CND2_HUMAN	13585	53022.5	71602.5
VARS	SYVC_HUMAN	15656	60995.33	138192
RPS24	RS24_HUMAN	15273	57255.33	88604
STK26	STK26_HUMAN	10991	40577.5	113523
MRPS22	RT22_HUMAN	13517.5	47205.67	80013.33
ABCF1	ABCF1_HUMAN	8748	30318	94099
XPO5	XPO5_HUMAN	6835	22723	34969
IGKV4-1	KV401_HUMAN	169396	529785.7	1361295
-	KV403_HUMAN	169396	529785.7	1361295
-	KV404_HUMAN	169396	529588.3	1361295
-	KV402_HUMAN	178620	530037	1361295
TLE3	TLE3_HUMAN	7040	20421.5	126837.7
TPD52	TPD52_HUMAN	8139	22478	94962.5
RPL23	RL23_HUMAN	44345	117072.3	281762.7
CDK13	CDK13_HUMAN	17459.5	44394.67	123523.5
PRPF4B	PRP4B_HUMAN	13674	32079	89119.67
HIST1H2AA	H2A1A_HUMAN	483173.7	1121672	2963711
UTP6	UTP6_HUMAN	8704	20036	170254
LARP1	LARP1_HUMAN	16199	33287	82689
LRRC47	LRC47_HUMAN	6928	13170	35485.5
HIST2H2AB	H2A2B_HUMAN	458379.3	795841.7	2448836
UTP15	UTP15_HUMAN	8524	13956	58413
PPP2R2C	2ABG_HUMAN	0	0	15153
AP1M1	AP1M1_HUMAN	0	0	20444.5
AP2A2	AP2A2_HUMAN	0	0	21941.5
CCDC88B	CC88B_HUMAN	0	0	186635
CLPB	CLPB_HUMAN	0	0	75548.5
DDX47	DDX47_HUMAN	0	0	16429.5
ESYT2	ESYT2_HUMAN	0	0	28406.5
GFAP	GFAP_HUMAN	95776.5	0	625387
PRKCA	KPCA_HUMAN	0	0	26630.5
LMAN1	LMAN1_HUMAN	0	0	42243.5
MTA3	MTA3_HUMAN	0	0	30367.33
NOP10	NOP10_HUMAN	0	0	25478
RCL1	RCL1_HUMAN	0	0	16801.5
MRPL1	RM01_HUMAN	0	0	15473.5
MRPL49	RM49_HUMAN	0	0	16441
SNRPN	RSMN_HUMAN	7261	0	58898
TOR1AIP1	TOIP1_HUMAN	0	0	21126

Table 8 Proteins identified by MS analysis.

HCF1 is a core protein of the SET1/KMT2A complex, involved in transcriptional regulation and H3K4 mono-methylation and di-methylation [156]. This is in line with our previous data, so we performed anti-V5 Co-IP experiment in 293T cells transiently co-transfected with HCF1 and either SETBP1-G870S or Empty vector, followed by western blot to confirm the data obtained by MS analysis. As expected, we found HCF1 protein immunoprecipitating together with SETBP1-G870S (Figure 29a). To validate this data we also performed independent Co-IP experiments targeting the HCF1 protein. This countercheck showed SETBP1-G870S protein immunoprecipitating together with HCF1 (Figure 29b).

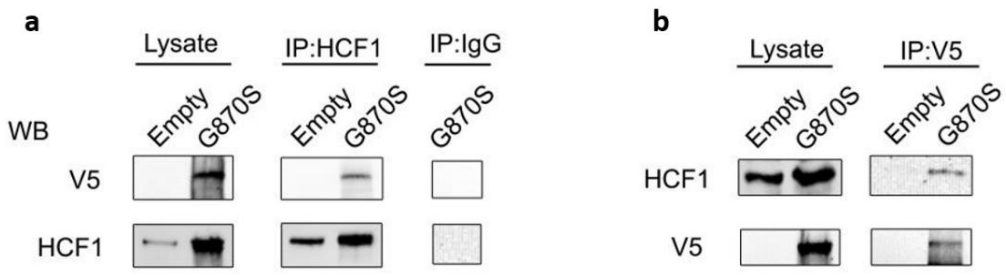


Figure 29 SETBP1 binding to HCF1.

A) Co-IP against HCF1 protein in 293T cells co-transfected with either SETBP1-G870S or Empty vector and with HCF1, showing immunoprecipitation of SETBP1. B) Co-IP against the V5 tag of SETBP1 resulting in the immunoprecipitation of HCF1.

Performing an *in-silico* linear motif analysis with the Eukaryotic Linear Motif (ELM) software (http://elm.eu.org) [157] on the SETBP1 sequence, we observed the presence of a putative HCF1 binding motif (HBM), occurring at position 991-994 of SETBP1 (Figure 30).

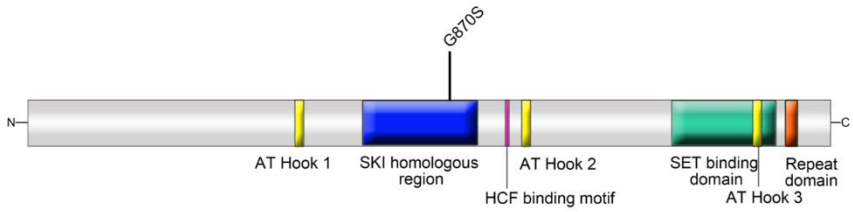


Figure 30 Schematic view of HBM in SETBP1 sequence.

SETBP1 (NM_015559.2) is characterized by an HCF binding motif occurring at position 991-994.

In order to confirm that SETBP1 interacts through this linear domain with HCF1, we used site-directed mutagenesis to delete HBM in pcDNA6.2-SETBP1wt and -SETBP1-G870S plasmids. The deletion of HBM (ΔHBM) was confirmed by Sanger sequencing (Figure 31).

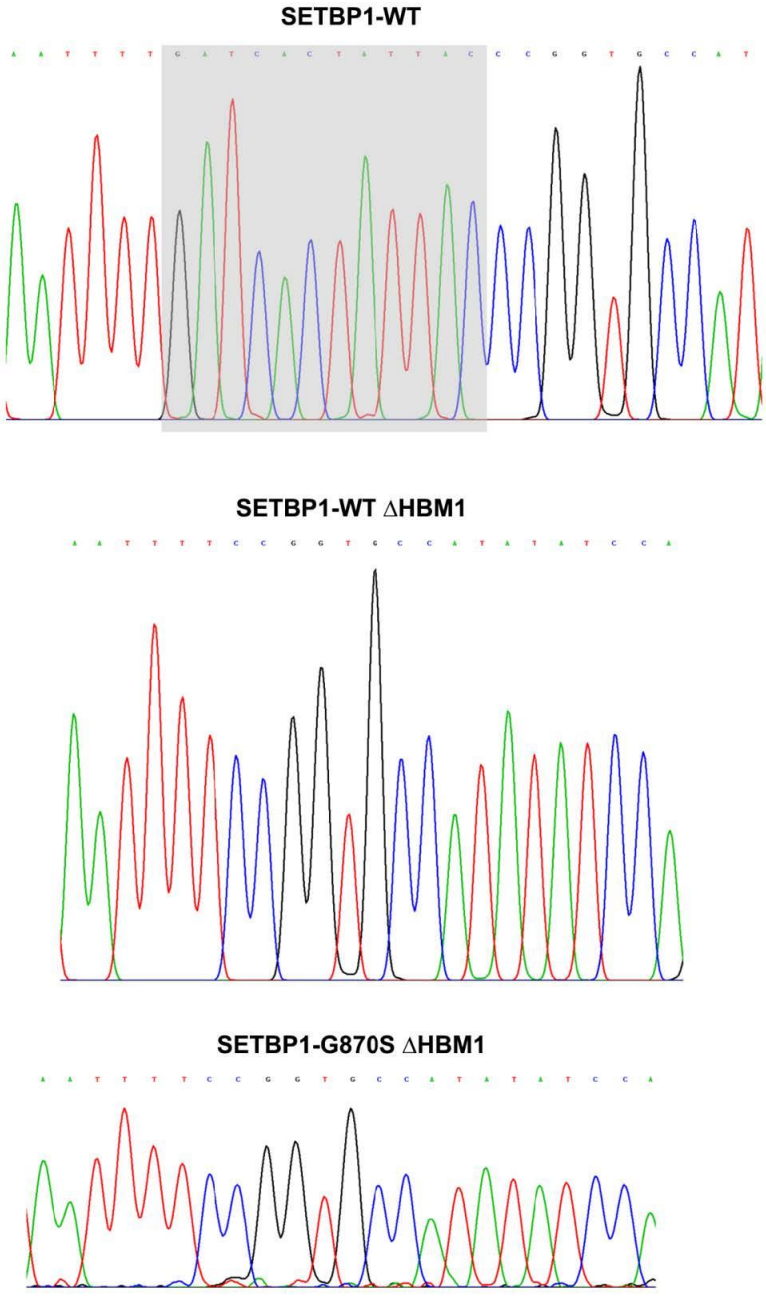


Figure 31 Sanger sequencing of pcDNA6.2 SETBP1.

WT (upper panel), WT ΔHBM (central) and G870S ΔHBM (lower panel) are shown. The shadowed area indicates the deletion of 12 nucleotides, corresponding to the HBM (aa 991-994).

Plasmids carrying the deletion were subsequently used to perform FRET experiments. FRET permits to measure the proximity of two molecules within a distance of 10 nm, because of the radiationless energy transfer between donor and acceptor fluorochromes. In fact, if the distance between the two molecules is less than 10 nm, the donor transfers part of its energy to the acceptor, which in turn starts to emit fluorescence. As expected, we observed that the deletion of HBM region was able to abrogate the interaction between SETBP1 and HCF1 (Figure 32a). Next, to assess that this abrogation was due to ΔHBM and not to a complete inactivation of SETBP1, we performed a similar set of experiments looking for the binding of beta-TrCP. The latter was used as a positive control because it is known to interact with the wild-type form of SETBP1 but not with the mutated one. Notably, we observed that ΔHBM did not alter the binding properties of SETBP1 towards the beta-TrCp (Figure 32b), confirming that the general binding function of SETBP1 is not altered. Taken together, these experiments suggest a direct interaction between SETBP1 and HCF1.

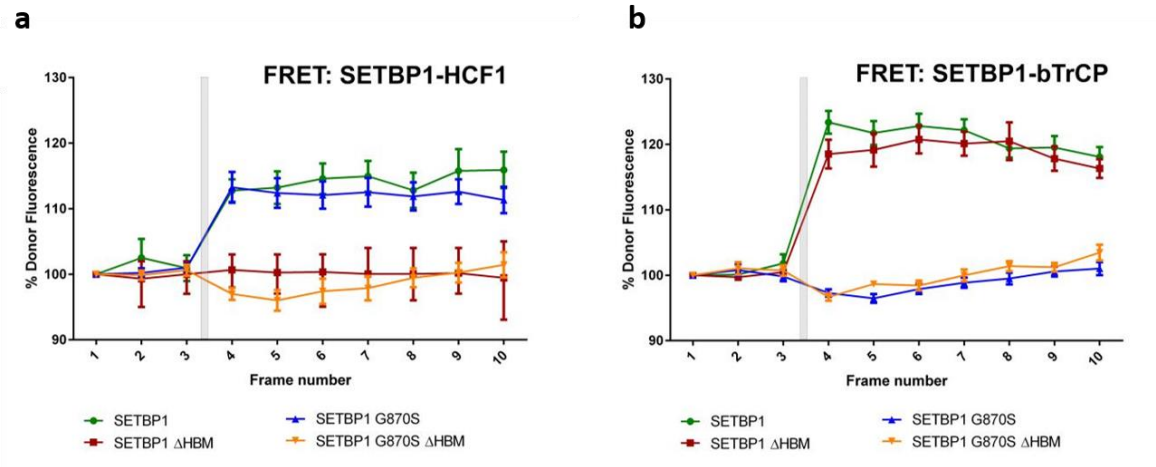


Figure 32 SETBP1-HCF1 interaction.

Positive FRET signal was recorded in both couples of HCF1 and SETBP1wt or SETP1-G870S (panel A, green and blue lines), conversely no FRET signal was recorded for HCF1 and SETBP1 ΔHBM or G870S ΔHBM (panel B, red and orange lines). FRET between bTrCP and SETBP1 variants was assayed to demonstrate that ΔHBM did not modify the known SETBP1- bTrCP interaction (panel B). Acceptor photobleaching was performed after the third acquired frame and indicated with gray bars in both the graphs.

HCF1 is known to be a component of the SET1/KMT2A-complex proteins associated with set 1 (COMPASS), which in turn is involved in transcriptional activation. Together with our previous results regarding SETBP1-mediated gene upregulation, we supposed that KMT2A could be part of the SETBP1 complex. To verify this hypothesis, we performed Co-IP experiments against V5-tag as previously described. As expected, KMT2A co-immunoprecipitated with SETBP1-G870S, suggesting the direct interaction of SETBP1, with this complex (Figure 33).

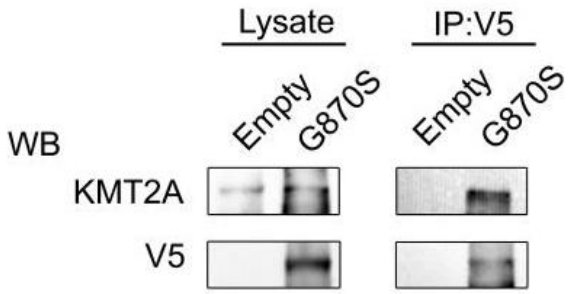


Figure 33 KMT2A interaction with SETBP1.

Co-IP experiment were performed against the V5 tag in transiently transfected 293T cells with SETBP1wt or Empty vector and blotted with anti-KMT2A antibody.

It is known that the PHF8 lysine demethylase can bind di- and tri-methylated H3K4 in the context of KMT2A complexes thanks to the presence of a PHD finger domain and exerts its activity on H4K20, resulting in its demethylation and transcriptional activation. Initially, to check that also PHF8 is part of in the SETBP1-mediated transcriptional complex we performed ChIP experiments against the H4K20me1 marks. By qRT-PCR we detected a significant decrease in all of the genes tested, suggesting that PHF8 is part of SETBP1 complex (Figure 34)

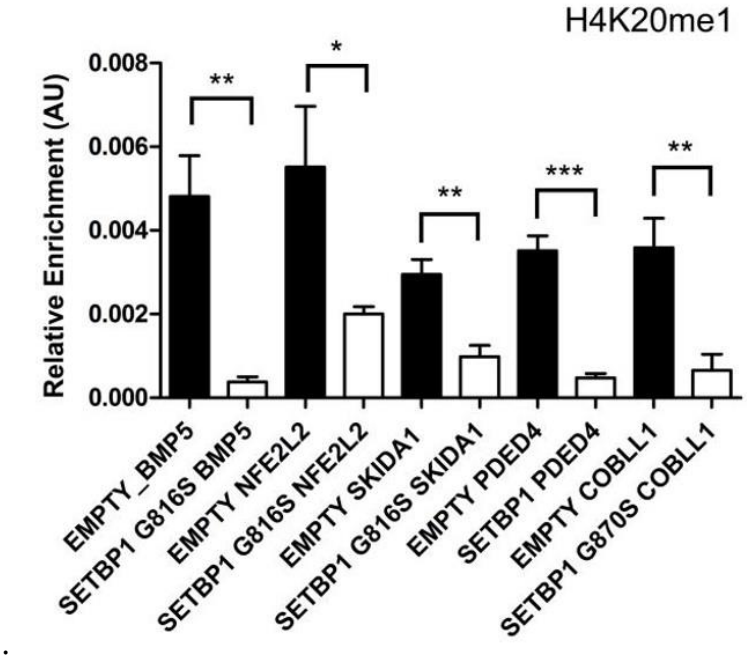


Figure 34 SETBP1/PHF8-mediated H4K20 demethylation.

ChIP experiments against H4K20me1 followed by qRT-PCR on a set of SETBP1 target genes showing a decrease of the H4K20me1 marks. * $p<0.05$ ** $p<0.01$ *** $p<0.001$

As we observed a reduction in the H4K20me1 mark for all the genes tested we wanted to verify that the possible cause of this decrease was the demethylating activity of PHF8. To test this hypothesis, we performed Co-IP experiments against V5-tag in transiently transfected 293T with SETBP1-G870S or the Empty vector. As expected, we observed the co-immunopreciptiation of SETBP1 and PHF8, confirming that also PHF8 is part of the SETBP1 complex (Figure 35).

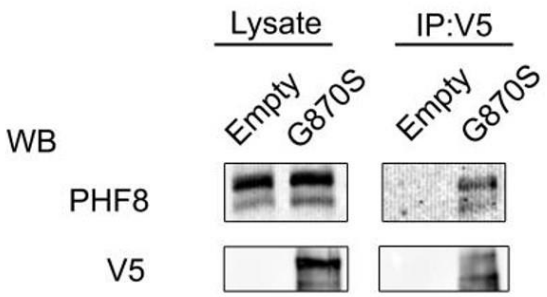


Figure 35 PHF8 partecipates in the SETBP1/KMT2A/HCF1-complex.

Co-IP experiments in 293T cells transiently transfected with either Empty vector or SETBP1-G870S results in the immunoprecipitation of PHF8 protein.

4.6 ETS2-ERG fusion altered response to ATRA in HL-60 cells

HL-60 cell line is a well known cellular model for differentiations studies. These cells, that largely resembles promyelocytes can be induced to differentiate into neutrophil-like cells upon treatment with micromolar doses of ATRA, as shown by the increase in expression of the neutrophilic marker CD11b [158].

To investigate the hypothesis that the ETS2-ERG fusion could results in an altered response to the ATRA differentiating agent, we created a suitable model with the ATRA-responsive cell line HL-60. To this aim, cells were transduced with either a plasmid expressing ETS2-ERG or the Empty vector and the expression of ETS2-ERG fusion protein was confirmed by western blot analysis (Figure 36).

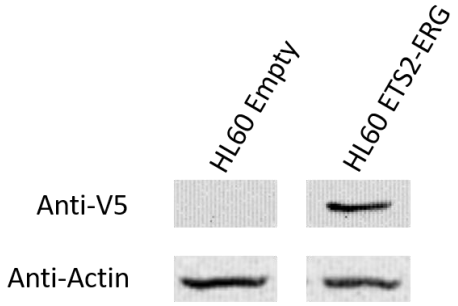


Figure 36 ETS2-ERG expression by HL60 cell line

Western blot analysis showing the expression at protein level of the ETS2-ERG fusion using an anti-V5 antibody. 100 µg of total lysate was loaded. Anti-actin was used as loading control.

After that, we checked the differentiation potential toward neutrophil-like cells of HL-60 expressing the ETS2-ERG fusion protein upon ATRA treatment for 72h at a dosage of 1 µM, in comparison to the cells transduced with the Empty vector. The differentiation state was evaluated by the expression of the differentiation marker CD11b (Figure 37) in 6 independent experiment. As a result, we observed that the CD11b positive percent of HL60 cells carrying the ETS2-ERG fusion gene was significantly lower in comparison to the cells transduced with the Empty vector (28.34% ± 3.68% vs 47.21% ± 4.89%, N=6, p=0.012; Figure 38).

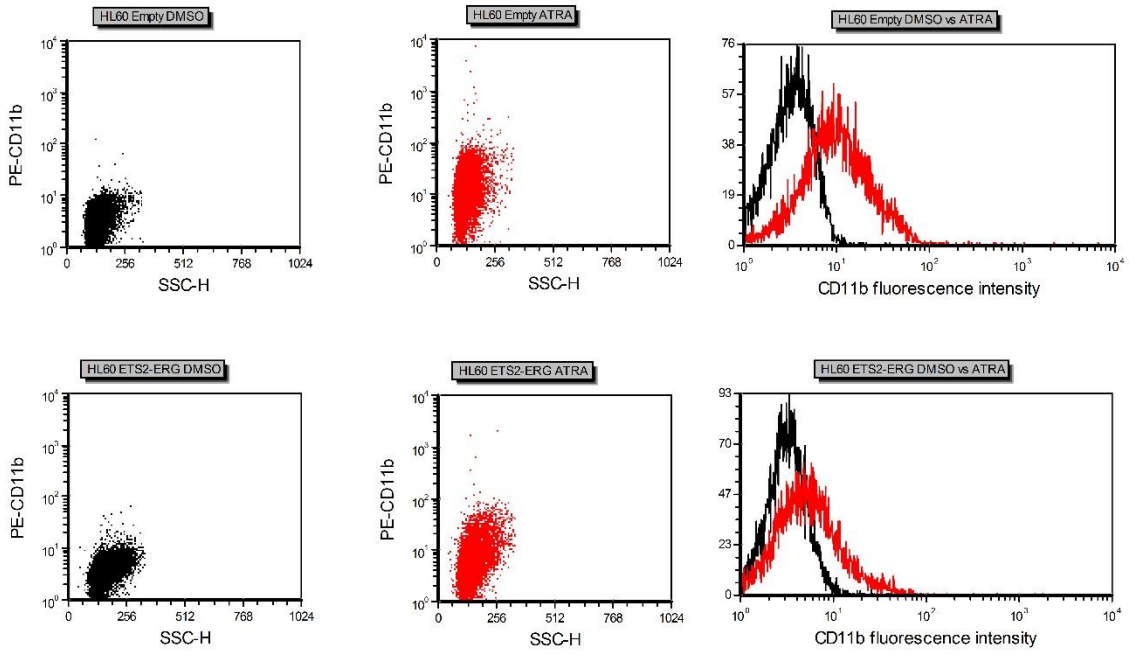


Figure 37 CD11b expression upon ATRA treatment.

Representative experiments showing the differentiation level in HL60 cells expressed as fluorescence intensity of the CD11b marker. The red color represents cells treated with ATRA 1 µM, whereas the black color represents the control cells treated with DMSO at the same concentration. Upper boxes show the differentiation level of cells transduced with the Empty vector. Bottom boxes show the differentiation of cells carrying the ETS2-ERG fusion gene.

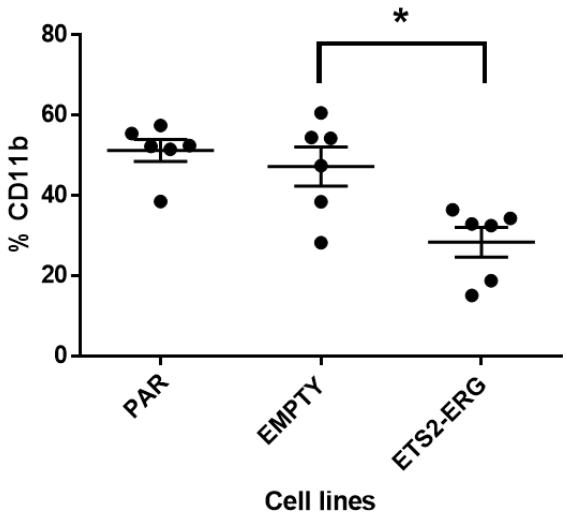


Figure 38 ETS2-ERG impairs the differentiation potential of HL60 cells.

Parental, empty (control) and ETS2-ERG HL-60 treated with 1 µM ATRA for 72h. CD11b positive percentages are 51,24% ± 2,71%, 47.21% ± 4.89%, 28.34% ± 3.68%, for PAR, EMPTY and ETS2-ERG, respectively. Values represent the mean ± SEM. PAR= HL-60 parental cells. CD11b expression was assessed by flow cytometry. *$p<0.05$

4.7 ETS2-ERG fusion in APL patients

In order to assess if the ETS2-ERG somatic fusion could be a recurrent event in APL patients we planned to screen a subset of APL patients by PCR and nested PCR analyses. Due to the impaired differentiation upon ATRA treatement observed in our cellular model expressing the ETS2-ERG fusion, we focused on patients that did not respond or had a very low response to ATRA treatment. We screened 20 patients by PCR and nested-PCR. Unfortunately, all the patients tested were negative for the ETS2-ERG fusion (data not shown).

Discussion

In the last few years, NGS technologies have revolutionized the field of cancer research. Different NGS applications allow to investigate, in a high-throughput manner, the genomic landscape of tumors, leading to the identification of the 'driver' mutations responsible for the cancer progression. In addition, these applications allow to explore cancer alterations at both transcriptomic and epigenomic level. The combined use of these different approaches can help scientists to obtain a global view of the cancer genome at a high-resolution level.

Our research group identified recurrent somatic mutations of *SETBP1* in aCML patients using whole-exome sequencing.[1] Moreover, using an in-house developed bioinformatic tool, we were able to detect through RNA-Seq the presence of the new fusion *ETS2-ERG* in a patient affected by APL.[2, 121] Due to little information about SETBP1 function and the complete lack of knowledge regarding the biological role of the ETS2-ERG fusion protein, we decided to gain further insight about these two proteins.

Historically, cancer research was focused on the study of genetic alterations. Indeed, tumors have been considered for years a set of diseases driven mainly by the accumulation of genetic mutations, which were retained the main cause of cancer [159]. However, this paradigm now has been expanded.

It is broadly accepted that the epigenetic control of gene expression plays a key role in cancer progression. The term 'epigenetics' is referred to any process inheritable by cell division, leading to changes in gene expression without altering the DNA sequence. The epigenetic machinery comprehends different mechanisms to control transcriptional activity, such as DNA methylation, histone modifications, nucleosome remodeling and RNA-mediated targeting [160]. Several authors showed evidence that SETBP1 is able to bind to gDNA [96, 97], demonstrating its involvement in transcriptional regulation. However, the mechanism through which it exerts its activity was not known.

In this study, we provide different lines of evidence that show that SETBP1 binds to DNA and recruits a HCF1/KMT2A/PHF8 complex, acting as a transcriptional activator.

Our ChIP-Seq data demonstrated that SETBP1 binds to promoter regions but also to other regions of gDNA such as enhancer, exonic, intronic and intergenic region. Although these results are difficult to explain, it is possible that SETBP1 could act at different regulatory levels within the genome. For example, intron-binding proteins are known to be involved in splicing regulation [161].

By Co-IP, MS and FRET analysis we showed that SETBP1 is able to interact directly with HCF1 to recruit KMT2A.

KMT2A, also known as mixed-lineage leukemia 1 (MLL1) is a well-known protein involved in epigenetic regulation during development and hematopoiesis. It has been identified as the mammalian ortholog of *Drosophila Trithorax* (*trx*) and belongs to the Trithorax group (TrxG) of proteins [162]. It exerts its action as an histone-modifying protein, leading to chromatin modifications through its H3K4 methyltransferase activity. The *in vivo* functions of KMT2A has been extensively studied. For example, it has been observed that the homozygous knockout of *Kmt2a* is embryonically lethal in mice whereas heterozygous knockout leads to growth retardation, hematopoietic defects and skeletal malformations [163]. In zebrafish embryos, the knockdown of Kmt2a resulted in hematopoietic defects and the same phenotype was observed in mammals [163]. Moreover, Kmt2a is described as essential in neural development in zebrafish and it is also an essential regulator of complex behaviors in mice [164, 165]. Taken globally, these data point to a critical role of KMT2A in controlling the differentiation process of bone, nervous and hematopoietic systems. This is in line with our functional enrichment analyses, where we observed a strong over-representation of development-related processes as well as enrichment for ontologies linked to cell differentiation. Notably, the same SETBP1 mutational hotspot found in aCML was previously observed in the Schinzel-Giedion syndrome, a severe germline, *de novo* disorder characterized by multiorgan development abnormalities, mental retardation and increased risk of cancer [95]. It is plausible that SETBP1-mediated transcriptional deregulation could play an important role in the onset of the SGS. Further *in vivo* studies on a SETBP1 animal model will allow to analyze in depth the correlation between mutated SETBP1 and SGS phenotype.

KMT2A is a core protein of the complex proteins associated with set1 (COMPASS)-like. COMPASS-like complexes perform the bulk of H3K4 mono- and di-methylation, and they are involved in the regulation of development-specific genes [166]. According that KMT2A possesses a strong H3 mono- and di-methylation but weak trimethylation activity, no significant H3K4me3 enrichment could be found in promoters occupied by SETBP1. The intersection between SETBP1 promoter occupancy and transcriptome analysis, together with the increase in H3K9Ac and direct interaction of SETBP1 with PHF8 demethylase, revealed that SETBP1 is part of an activator complex. The exact explanation of this complex histone pattern is still unclear, due to our limited knowledge about the functions of the H3K4me2 mark. Nevertheless, several lines of evidence suggest that H3K4me2 is strongly enriched in lineage-specific promoters [167-169]. Again, our observation correlates well with the abundance of genes associated with multi-organ development

among the SETBP1-regulated genes, and it is interesting in the context of SGS, whose hallmark is the presence of multi-organ development abnormalities.

In the literature, it is known that PHF8 is able to bind di- and tri-methylated H3K4 in the context of KMT2A [170] and that the demethylase activity of PHF8 results in the H4K20/H3K9 demethylation [171], in the recruitment of the RNA polymerase II (RNAPII) and activation of transcription [172]. In line with these data, our experiments identified PHF8 as part of the SETBP1 activator complex and we detected a significant decrease in the H4K20me1 mark for all of the tested SETBP1 target genes. Interestingly, PHF8 is known to interact with the RNA polymerase II (RNAPII). Notably, PHF8 is also described as responsible for brain and craniofacial development in zebrafish [173] and mutations in this gene are associated with X-linked mental retardation in humans [174]. Together, these observations suggest a link between an altered PHF8 activity and SGS phenotype.

MECOM gene encodes for a zinc finger transcription factor involved in several normal developmental processes as well as in leukomogenesis, by recruiting both coactivator and corepressor [175]. It is expressed in HSCs, where it plays a critical role in the hematopoiesis and in HSCs self-renewal [176]. The oncogenic role of *MECOM* is highlighted by the evidence that it is often overexpressed, as a result of chromosomal translocation, in MDS and in approximately 10% of AML cases, representing a poor prognostic factor [177]. Moreover, it is also able to modulate other transcription factors acting on the c-Jun N-terminal kinase (JNK)-mediated signaling pathway to inhibit cells apoptosis. Interestingly, *MECOM* is one of the genes identified as dysregulated in our *in vitro* analysis as well as in SETBP1$^+$ aCML patients. Its identification as a direct target of SETBP1 transcriptional complex highlights a potential critical role of *MECOM* and/or its downstream effectors in the oncogenesis of SETBP1$^+$ malignancies. However, further studies will be required to confirm these findings.

The identification of a new mechanism of action of SETBP1 could also affect the therapy of SETBP1$^+$ neoplasm in the future. Unfortunately, no selective drug directly targeting SETBP1 is currently available. However, some alternative strategies might be evaluated. For example, the MM-401 compound is described as an inhibitor of the KMT2A methyltransferase activity by blocking the KMT2A-WDR5 interaction [178], since WDR5 is described as a component of COMPASS-like complexes [166]. Another strategy could be the use of the identified HBM site in the SETBP1 sequence to develop a peptidomimetic drug, therefore inhibiting the SETBP1-HCF1 interaction. However, it should be considered that targeting other factors rather than mutated

SETBP1 could be deleterious, since also the physiological activity of these proteins will be blocked. Since we observed an increase in H3K9Ac mark, another therapeutic approach could be the use of histone acetyltransferase (HAT) inhibitors. In fact, this class of compounds seems promising for the treatment of different cancers [179]. For example, it is reported that HAT inhibitors PU139 and PU141 are effective in neuroblastoma, at least in mouse models. Moreover, PU139 is shown to synergize with doxorubicin, an anthracycline compound with antineoplastic activity already used in induction therapy of some myeloid malignancies [180].

In conclusion, our data identify a mechanism of action of SETBP1 as transcriptional activator (Figure 39), beyond its original SET-binding activity, and shed light on possible new therapeutic targets for the SETBP1^{+} neoplasms.

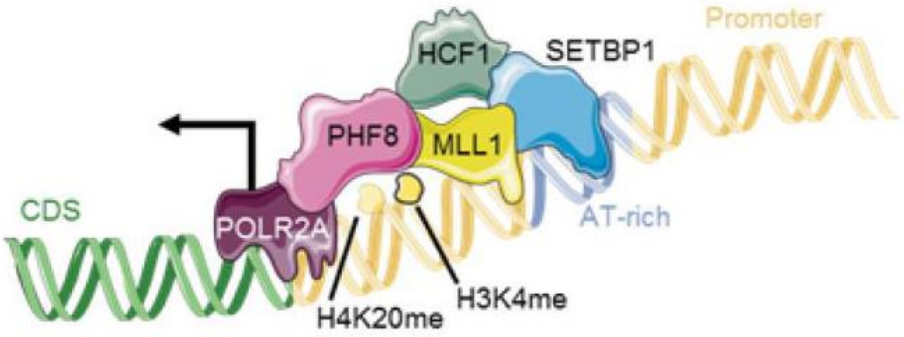


Figure 39 Proposed model of SETBP1-mediated epigentic regulation.

SETBP1 recruits a HCF1/KMT2A/PHF8 complex to induce transcriptional activation.
In the picture KMT2A is represented using its alternative name MLL1.

The new fusion gene ETS2-ERG was discovered in a patient (AML002) affected by APL. However, no information is available in literature regarding this fusion. ATRA, the acid form of the vitamin A, is able to induce differentiation in various cell lines and it is used in the therapy of APL [106, 181]. Notably, the patient carrying this fusion did not respond to ATRA treatment. Therefore, we hypothesized that ETS2-ERG could exert its biological effects by impairing the

differentiation process in APL. In order to better characterize this chromosomal abnormality, we generated a HL-60 cell line model overexpressing the ETS2-ERG fusion and we assessed the differentiation level upon ATRA administration. As expected, we observed a significant decrease in the differentiation antigen CD11b expression in cells carrying the ETS2-ERG fusion, strongly supporting our hypothesis. However, the main pitfall of our model is that the HL-60 cell line does not carry the characteristic APL fusion gene PML-RARA. In addition, we screened APL patients in order to verify if this chromosomal translocation is a recurrent event among patients unresponsive to ATRA therapy. Unfortunately, all the patients tested so far were negative for the ETS2-ERG fusion, indicating that it is a rare event. In conclusion, our results suggest that the ETS2-ERG fusion could play a significant role in inducing resistance against ATRA in APL patients. However, more studies are needed to validate our observations, especially accomplished in a cellular model carrying the PML-RARA fusion.

Bibliography

1. Piazza, R., et al., *Recurrent SETBP1 mutations in atypical chronic myeloid leukemia.* Nat Genet, 2013. **45**(1): p. 18-24.
2. Piazza, R., et al., *RNA-seq is a valuable complement of conventional diagnostic tools in newly diagnosed AML patients.* Am J Hematol, 2015. **90**(12): p. E227-8.
3. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
4. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
5. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
6. Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1135-45.
7. Chaisson, M., P. Pevzner, and H. Tang, *Fragment assembly with short reads.* Bioinformatics, 2004. **20**(13): p. 2067-74.
8. Levy, S.E. and R.M. Myers, *Advancements in Next-Generation Sequencing.* Annu Rev Genomics Hum Genet, 2016. **17**: p. 95-115.
9. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
10. *An Introduction to Next-Generation Sequencing Technology*. 2016; Available from: http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
11. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-9.
12. Holt, R.A. and S.J. Jones, *The new paradigm of flow cell sequencing.* Genome Res, 2008. **18**(6): p. 839-46.
13. Metzker, M.L., *Emerging technologies in DNA sequencing.* Genome Res, 2005. **15**(12): p. 1767-76.
14. Nakazato, T., T. Ohta, and H. Bono, *Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive.* PLoS One, 2013. **8**(10): p. e77910.
15. Rougemont, J., et al., *Probabilistic base calling of Solexa sequencing data.* BMC Bioinformatics, 2008. **9**: p. 431.
16. Whiteford, N., et al., *Swift: primary data analysis for the Illumina Solexa sequencing platform.* Bioinformatics, 2009. **25**(17): p. 2194-9.
17. Das, S. and H. Vikalo, *Base calling for high-throughput short-read sequencing: dynamic programming solutions.* BMC Bioinformatics, 2013. **14**: p. 129.
18. *Understanding Illumina Quality Scores*. 2014; Available from: http://www.illumina.com/documents/products/technotes/technote_understanding_quality_scores.pdf.
19. *bcl2fastq Conversion*. 2013; Available from: http://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq_letterbooklet_15038058brpmi.pdf.
20. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
21. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
22. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

23. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
24. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.
25. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
26. Benjamin, A.M., et al., *Comparing reference-based RNA-Seq mapping methods for non-human primate data.* BMC Genomics, 2014. **15**: p. 570.
27. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
28. *Sequence Alignment/Map Format Specification*. 2016; Available from: https://samtools.github.io/hts-specs/SAMv1.pdf.
29. Giardino, A., et al., *Role of Imaging in the Era of Precision Medicine.* Acad Radiol, 2017.
30. Meldrum, C., M.A. Doyle, and R.W. Tothill, *Next-generation sequencing for cancer diagnostics: a practical perspective.* Clin Biochem Rev, 2011. **32**(4): p. 177-95.
31. Park, S.T. and J. Kim, *Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing.* Int Neurourol J, 2016. **20**(Suppl 2): p. S76-83.
32. Chaitankar, V., et al., *Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research.* Prog Retin Eye Res, 2016. **55**: p. 1-31.
33. Ma, Y., et al., *Applications of Next-generation Sequencing in Systemic Autoimmune Diseases.* Genomics Proteomics Bioinformatics, 2015. **13**(4): p. 242-9.
34. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19096-101.
35. Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies.* Nat Rev Genet, 2016. **17**(6): p. 333-51.
36. Warr, A., et al., *Exome Sequencing: Current and Future Perspectives.* G3 (Bethesda), 2015. **5**(8): p. 1543-50.
37. Farrell, C.M., et al., *Current status and new features of the Consensus Coding Sequence database.* Nucleic Acids Res, 2014. **42**(Database issue): p. D865-72.
38. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.
39. Cunningham, F., et al., *Ensembl 2015.* Nucleic Acids Res, 2015. **43**(Database issue): p. D662-9.
40. Lindberg, J. and J. Lundeberg, *The plasticity of the mammalian transcriptome.* Genomics, 2010. **95**(1): p. 1-6.
41. Conesa, A., et al., *A survey of best practices for RNA-seq data analysis.* Genome Biol, 2016. **17**(1): p. 13.
42. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology.* Nat Rev Genet, 2009. **10**(10): p. 669-80.
43. Mikkola, H.K. and S.H. Orkin, *The journey of developing hematopoietic stem cells.* Development, 2006. **133**(19): p. 3733-44.
44. Palis, J. and M.C. Yoder, *Yolk-sac hematopoiesis: the first blood cells of mouse and man.* Exp Hematol, 2001. **29**(8): p. 927-36.
45. Palis, J., et al., *Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse.* Development, 1999. **126**(22): p. 5073-84.
46. Palis, J., *Primitive and definitive erythropoiesis in mammals.* Front Physiol, 2014. **5**: p. 3.
47. McGrath, K.E., J.M. Frame, and J. Palis, *Early hematopoiesis and macrophage development.* Semin Immunol, 2015. **27**(6): p. 379-87.
48. Tober, J., et al., *The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis.* Blood, 2007. **109**(4): p. 1433-41.
49. Clements, W.K. and D. Traver, *Signalling pathways that control vertebrate haematopoietic stem cell specification.* Nat Rev Immunol, 2013. **13**(5): p. 336-48.
50. Yoder, M.C., *Inducing definitive hematopoiesis in a dish.* Nat Biotechnol, 2014. **32**(6): p. 539-41.

51. McGrath, K.E., et al., *Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo.* Cell Rep, 2015. **11**(12): p. 1892-904.
52. Gomez Perdiguero, E., et al., *Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors.* Nature, 2015. **518**(7540): p. 547-51.
53. Christensen, J.L., et al., *Circulation and chemotaxis of fetal hematopoietic stem cells.* PLoS Biol, 2004. **2**(3): p. E75.
54. O'Neill, H.C., et al., *Spleen as a site for hematopoiesis of a distinct antigen presenting cell type.* Stem Cells Int, 2011. **2011**: p. 954275.
55. Ciriza, J., et al., *The migration of hematopoietic progenitors from the fetal liver to the fetal bone marrow: lessons learned and possible clinical applications.* Exp Hematol, 2013. **41**(5): p. 411-23.
56. Kondo, M., et al., *Biology of hematopoietic stem cells and progenitors: implications for clinical application.* Annu Rev Immunol, 2003. **21**: p. 759-806.
57. Seita, J. and I.L. Weissman, *Hematopoietic stem cell: self-renewal versus differentiation.* Wiley Interdiscip Rev Syst Biol Med, 2010. **2**(6): p. 640-53.
58. Morrison, S.J. and I.L. Weissman, *The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype.* Immunity, 1994. **1**(8): p. 661-73.
59. Christensen, J.L. and I.L. Weissman, *Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14541-6.
60. Morrison, S.J., et al., *Identification of a lineage of multipotent hematopoietic progenitors.* Development, 1997. **124**(10): p. 1929-39.
61. Kondo, M., *Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors.* Immunol Rev, 2010. **238**(1): p. 37-46.
62. Lai, A.Y. and M. Kondo, *T and B lymphocyte differentiation from hematopoietic stem cell.* Semin Immunol, 2008. **20**(4): p. 207-12.
63. Fiedler, K. and C. Brunner, *Mechanisms Controlling Hematopoiesis*. 2012: INTECH Open Access Publisher.
64. Larsson, J. and S. Karlsson, *The role of Smad signaling in hematopoiesis.* Oncogene, 2005. **24**(37): p. 5676-92.
65. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages.* Nature, 2000. **404**(6774): p. 193-7.
66. Kondo, M., I.L. Weissman, and K. Akashi, *Identification of clonogenic common lymphoid progenitors in mouse bone marrow.* Cell, 1997. **91**(5): p. 661-72.
67. Iwasaki, H. and K. Akashi, *Myeloid lineage commitment from the hematopoietic stem cell.* Immunity, 2007. **26**(6): p. 726-40.
68. Manz, M.G., et al., *Dendritic cell potentials of early lymphoid and myeloid progenitors.* Blood, 2001. **97**(11): p. 3333-41.
69. Traver, D., et al., *Development of CD8alpha-positive dendritic cells from a common myeloid progenitor.* Science, 2000. **290**(5499): p. 2152-4.
70. Virchow, R.L.K., *Leukemie, Gesammelte Abhandlungen zur wissen schaftlichen medicin.* Frankfurt: Meidinger Sohn & comp. 1856. 190-212.
71. Bennett, J.M., et al., *Proposals for the classification of the myelodysplastic syndromes.* Br J Haematol, 1982. **51**(2): p. 189-99.
72. Dameshek, W., *Some speculations on the myeloproliferative syndromes.* Blood, 1951. **6**(4): p. 372-5.
73. Vardiman, J.W., N.L. Harris, and R.D. Brunning, *The World Health Organization (WHO) classification of the myeloid neoplasms.* Blood, 2002. **100**(7): p. 2292-302.
74. Vardiman, J.W., et al., *The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes.* Blood, 2009. **114**(5): p. 937-51.
75. Tefferi, A., et al., *FIP1L1-PDGFRA and c-kit D816V mutation-based clonality studies in systemic mast cell disease associated with eosinophilia.* Haematologica, 2004. **89**(7): p. 871-3.

76. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia.* Blood, 2016. **127**(20): p. 2391-405.
77. Orazi, A. and U. Germing, *The myelodysplastic/myeloproliferative neoplasms: myeloproliferative diseases with dysplastic features.* Leukemia, 2008. **22**(7): p. 1308-19.
78. Tiu, R.V. and M.A. Sekeres, *Making sense of the myelodysplastic/myeloproliferative neoplasms overlap syndromes.* Curr Opin Hematol, 2014. **21**(2): p. 131-40.
79. Gambacorti-Passerini, C., et al., *Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib.* J Natl Cancer Inst, 2011. **103**(7): p. 553-61.
80. Wang, S.A., et al., *Atypical chronic myeloid leukemia is clinically distinct from unclassifiable myelodysplastic/myeloproliferative neoplasms.* Blood, 2014. **123**(17): p. 2645-51.
81. Meggendorfer, M., et al., *SETBP1 mutations occur in 9% of MDS/MPN and in 4% of MPN cases and are strongly associated with atypical CML, monosomy 7, isochromosome i(17)(q10), ASXL1 and CBL mutations.* Leukemia, 2013. **27**(9): p. 1852-60.
82. Maxson, J.E., et al., *Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML.* N Engl J Med, 2013. **368**(19): p. 1781-90.
83. Pardanani, A., et al., *CSF3R T618I is a highly prevalent and specific mutation in chronic neutrophilic leukemia.* Leukemia, 2013. **27**(9): p. 1870-3.
84. Gambacorti-Passerini, C.B., et al., *Recurrent ETNK1 mutations in atypical chronic myeloid leukemia.* Blood, 2015. **125**(3): p. 499-503.
85. Zoi, K. and N.C. Cross, *Molecular pathogenesis of atypical CML, CMML and MDS/MPN-unclassifiable.* Int J Hematol, 2015. **101**(3): p. 229-42.
86. Gotlib, J., et al., *The new genetics of chronic neutrophilic leukemia and atypical CML: implications for diagnosis and treatment.* Blood, 2013. **122**(10): p. 1707-11.
87. Minakuchi, M., et al., *Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET.* Eur J Biochem, 2001. **268**(5): p. 1340-51.
88. Li, M., A. Makkinje, and Z. Damuni, *The myeloid leukemia-associated protein SET is a potent inhibitor of protein phosphatase 2A.* J Biol Chem, 1996. **271**(19): p. 11059-62.
89. Panagopoulos, I., et al., *Fusion of NUP98 and the SET binding protein 1 (SETBP1) gene in a paediatric acute T cell lymphoblastic leukaemia with t(11;18)(p15;q12).* Br J Haematol, 2007. **136**(2): p. 294-6.
90. Cristobal, I., et al., *SETBP1 overexpression is a novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute myeloid leukemia.* Blood, 2010. **115**(3): p. 615-25.
91. Shiba, N., et al., *SETBP1 mutations in juvenile myelomonocytic leukaemia and myelodysplastic syndrome but not in paediatric acute myeloid leukaemia.* Br J Haematol, 2014. **164**(1): p. 156-9.
92. Makishima, H., et al., *Somatic SETBP1 mutations in myeloid malignancies.* Nat Genet, 2013. **45**(8): p. 942-6.
93. Hoischen, A., et al., *De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.* Nat Genet, 2010. **42**(6): p. 483-5.
94. Lehman, A.M., et al., *Schinzel-Giedion syndrome: report of splenopancreatic fusion and proposed diagnostic criteria.* Am J Med Genet A, 2008. **146A**(10): p. 1299-306.
95. Schinzel, A. and A. Giedion, *A syndrome of severe midface retraction, multiple skull anomalies, clubfeet, and cardiac and renal malformations in sibs.* Am J Med Genet, 1978. **1**(4): p. 361-75.
96. Oakley, K., et al., *Setbp1 promotes the self-renewal of murine myeloid progenitors via activation of Hoxa9 and Hoxa10.* Blood, 2012. **119**(25): p. 6099-108.
97. Vishwakarma, B.A., et al., *Runx1 repression by histone deacetylation is critical for Setbp1-induced mouse myeloid leukemia development.* Leukemia, 2016. **30**(1): p. 200-8.
98. Ichikawa, M., et al., *AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis.* Nat Med, 2004. **10**(3): p. 299-304.
99. Growney, J.D., et al., *Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype.* Blood, 2005. **106**(2): p. 494-504.

100. Betz, B.L. and J.L. Hess, *Acute myeloid leukemia diagnosis in the 21st century.* Arch Pathol Lab Med, 2010. **134**(10): p. 1427-33.
101. De Kouchkovsky, I. and M. Abdul-Hay, *'Acute myeloid leukemia: a comprehensive review and 2016 update'.* Blood Cancer J, 2016. **6**(7): p. e441.
102. Lowenberg, B., J.R. Downing, and A. Burnett, *Acute myeloid leukemia.* N Engl J Med, 1999. **341**(14): p. 1051-62.
103. Ladines-Castro, W., et al., *Morphology of leukaemias.* Revista Médica del Hospital General de México, 2016. **79**(2): p. 107-113.
104. Venditti, A., et al., *Minimally differentiated acute myeloid leukemia (AML-M0): comparison of 25 cases with other French-American-British subtypes.* Blood, 1997. **89**(2): p. 621-9.
105. Licht, J.D., *AML1 and the AML1-ETO fusion protein in the pathogenesis of t(8;21) AML.* Oncogene, 2001. **20**(40): p. 5660-79.
106. Wang, Z.Y. and Z. Chen, *Acute promyelocytic leukemia: from highly fatal to highly curable.* Blood, 2008. **111**(5): p. 2505-15.
107. Ferrara, F. and C.A. Schiffer, *Acute myeloid leukaemia in adults.* Lancet, 2013. **381**(9865): p. 484-95.
108. Lo-Coco, F., et al., *Acute promyelocytic leukemia: recent advances in diagnosis and management.* Semin Oncol, 2008. **35**(4): p. 401-9.
109. Dong, H.Y., et al., *Flow cytometry rapidly identifies all acute promyelocytic leukemias with high specificity independent of underlying cytogenetic abnormalities.* Am J Clin Pathol, 2011. **135**(1): p. 76-84.
110. Di Noto, R., P. Mirabelli, and L. Del Vecchio, *Flow cytometry analysis of acute promyelocytic leukemia: the power of 'surface hematology'.* Leukemia, 2007. **21**(1): p. 4-8.
111. Nakahara, T. and P.F. Lambert, *Induction of promyelocytic leukemia (PML) oncogenic domains (PODs) by papillomavirus.* Virology, 2007. **366**(2): p. 316-29.
112. Mattsson, K., et al., *Proteins associated with the promyelocytic leukemia gene product (PML)-containing nuclear body move to the nucleolus upon inhibition of proteasome-dependent protein degradation.* Proc Natl Acad Sci U S A, 2001. **98**(3): p. 1012-7.
113. Kawai, T., S. Akira, and J.C. Reed, *ZIP kinase triggers apoptosis from nuclear PML oncogenic domains.* Mol Cell Biol, 2003. **23**(17): p. 6174-86.
114. Ablain, J. and H. de The, *Revisiting the differentiation paradigm in acute promyelocytic leukemia.* Blood, 2011. **117**(22): p. 5795-802.
115. Sanz, M.A. and P. Montesinos, *How we prevent and treat differentiation syndrome in patients with acute promyelocytic leukemia.* Blood, 2014. **123**(18): p. 2777-82.
116. Zeidan, A.M. and S.D. Gore, *New strategies in acute promyelocytic leukemia: moving to an entirely oral, chemotherapy-free upfront management approach.* Clin Cancer Res, 2014. **20**(19): p. 4985-93.
117. de The, H., M. Le Bras, and V. Lallemand-Breitenbach, *The cell biology of disease: Acute promyelocytic leukemia, arsenic, and PML bodies.* J Cell Biol, 2012. **198**(1): p. 11-21.
118. Grimwade, D., et al., *Acute promyelocytic leukemia: a paradigm for differentiation therapy.* Cancer Treat Res, 2010. **145**: p. 219-35.
119. Mi, J.Q., et al., *How to manage acute promyelocytic leukemia.* Leukemia, 2012. **26**(8): p. 1743-51.
120. Lo-Coco, F. and E. Ammatuna, *Front line clinical trials and minimal residual disease monitoring in acute promyelocytic leukemia.* Curr Top Microbiol Immunol, 2007. **313**: p. 145-56.
121. Piazza, R., et al., *FusionAnalyser: a new graphical, event-driven tool for fusion rearrangements discovery.* Nucleic Acids Res, 2012. **40**(16): p. e123.
122. Park, J., et al., *Emerging new approaches for the treatment of acute promyelocytic leukemia.* Ther Adv Hematol, 2011. **2**(5): p. 335-52.
123. Cooper, C.D., J.A. Newman, and O. Gileadi, *Recent advances in the structural molecular biology of Ets transcription factors: interactions, interfaces and inhibition.* Biochem Soc Trans, 2014. **42**(1): p. 130-8.

124. Shaikhibrahim, Z., et al., *ERG is specifically associated with ETS-2 and ETV-4, but not with ETS-1, in prostate cancer.* Int J Mol Med, 2012. **30**(5): p. 1029-33.
125. Maroulakou, I.G. and D.B. Bowe, *Expression and function of Ets transcription factors in mammalian development: a regulatory network.* Oncogene, 2000. **19**(55): p. 6432-42.
126. Oikawa, T., *ETS transcription factors: possible targets for cancer therapy.* Cancer Sci, 2004. **95**(8): p. 626-33.
127. Hashiya, N., et al., *In vivo evidence of angiogenesis induced by transcription factor Ets-1: Ets-1 is located upstream of angiogenesis cascade.* Circulation, 2004. **109**(24): p. 3035-41.
128. Ciau-Uitz, A., et al., *ETS transcription factors in hematopoietic stem cell development.* Blood Cells Mol Dis, 2013. **51**(4): p. 248-55.
129. Xu, D., et al., *Ets2 maintains hTERT gene expression and breast cancer cell proliferation by interacting with c-Myc.* J Biol Chem, 2008. **283**(35): p. 23567-80.
130. Wolvetang, E.W., et al., *The chromosome 21 transcription factor ETS2 transactivates the beta-APP promoter: implications for Down syndrome.* Biochim Biophys Acta, 2003. **1628**(2): p. 105-10.
131. Ge, Y., et al., *The role of the proto-oncogene ETS2 in acute megakaryocytic leukemia biology and therapy.* Leukemia, 2008. **22**(3): p. 521-9.
132. Ge, H.Y., et al., *VEGFA Expression Is Inhibited by Arsenic Trioxide in HUVECs through the Upregulation of Ets-2 and miRNA-126.* PLoS One, 2015. **10**(8): p. e0135795.
133. Rainis, L., et al., *The proto-oncogene ERG in megakaryoblastic leukemias.* Cancer Res, 2005. **65**(17): p. 7596-602.
134. Taoudi, S., et al., *ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification.* Genes Dev, 2011. **25**(3): p. 251-62.
135. Baldus, C.D., et al., *Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses overexpression of APP, ETS2, and ERG genes.* Proc Natl Acad Sci U S A, 2004. **101**(11): p. 3915-20.
136. Anderson, M.K., et al., *Precise developmental regulation of Ets family transcription factors during specification and commitment to the T cell lineage.* Development, 1999. **126**(14): p. 3131-48.
137. Sotoca, A.M., et al., *The oncofusion protein FUS-ERG targets key hematopoietic regulators and modulates the all-trans retinoic acid signaling pathway in t(16;21) acute myeloid leukemia.* Oncogene, 2016. **35**(15): p. 1965-76.
138. Wilson, A.C., et al., *The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein.* Cell, 1993. **74**(1): p. 115-25.
139. Scifo, E., et al., *Proteomic analysis of the palmitoyl protein thioesterase 1 interactome in SH-SY5Y human neuroblastoma cells.* J Proteomics, 2015. **123**: p. 42-53.
140. Makela, J., et al., *Peroxisome proliferator-activated receptor-gamma coactivator-1alpha mediates neuroprotection against excitotoxic brain injury in transgenic mice: role of mitochondria and X-linked inhibitor of apoptosis protein.* Eur J Neurosci, 2016. **43**(5): p. 626-39.
141. Busnelli, M., et al., *Analysis of GPCR dimerization using acceptor photobleaching resonance energy transfer techniques.* Methods Enzymol, 2013. **521**: p. 311-27.
142. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
143. Pathan, M., et al., *FunRich: An open access standalone functional enrichment and interaction network analysis tool.* Proteomics, 2015. **15**(15): p. 2597-601.
144. Holst, J., et al., *Generation of T-cell receptor retrogenic mice.* Nat Protoc, 2006. **1**(1): p. 406-17.
145. Aravind, L. and D. Landsman, *AT-hook motifs identified in a wide variety of DNA-binding proteins.* Nucleic Acids Res, 1998. **26**(19): p. 4413-21.
146. Kulakovskiy, I.V., et al., *HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models.* Nucleic Acids Res, 2016. **44**(D1): p. D116-25.
147. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.
148. Reeves, R. and L. Beckerbauer, *HMGI/Y proteins: flexible regulators of transcription and chromatin structure.* Biochim Biophys Acta, 2001. **1519**(1-2): p. 13-29.

149. Huth, J.R., et al., *The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif.* Nat Struct Biol, 1997. **4**(8): p. 657-65.
150. Wang, X., et al., *High-resolution human core-promoter prediction with CoreBoost_HM.* Genome Res, 2009. **19**(2): p. 266-75.
151. Bourachot, B., M. Yaniv, and C. Muchardt, *The activity of mammalian brm/SNF2alpha is dependent on a high-mobility-group protein I/Y-like DNA binding domain.* Mol Cell Biol, 1999. **19**(6): p. 3931-9.
152. Cairns, B.R., et al., *Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains.* Mol Cell, 1999. **4**(5): p. 715-23.
153. Xiao, H., et al., *Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions.* Mol Cell, 2001. **8**(3): p. 531-43.
154. Dong, X. and Z. Weng, *The correlation between histone modifications and gene expression.* Epigenomics, 2013. **5**(2): p. 113-6.
155. Hahn, M.A., et al., *Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks.* PLoS One, 2011. **6**(4): p. e18844.
156. Rao, R.C. and Y. Dou, *Hijacked in cancer: the KMT2 (MLL) family of methyltransferases.* Nat Rev Cancer, 2015. **15**(6): p. 334-46.
157. Dinkel, H., et al., *ELM 2016--data update and new functionality of the eukaryotic linear motif resource.* Nucleic Acids Res, 2016. **44**(D1): p. D294-300.
158. Carrigan, S.O., et al., *Neutrophil differentiated HL-60 cells model Mac-1 (CD11b/CD18)-independent neutrophil transepithelial migration.* Immunology, 2005. **115**(1): p. 108-17.
159. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
160. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy.* Cell, 2012. **150**(1): p. 12-27.
161. Benecke, A.G. and S. Eilebrecht, *RNA-Mediated Regulation of HMGA1 Function.* Biomolecules, 2015. **5**(2): p. 943-57.
162. Djabali, M., et al., *A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias.* Nat Genet, 1992. **2**(2): p. 113-8.
163. Yu, B.D., et al., *Altered Hox expression and segmental identity in Mll-mutant mice.* Nature, 1995. **378**(6556): p. 505-8.
164. Huang, Y.C., et al., *The epigenetic factor Kmt2a/Mll1 regulates neural progenitor proliferation and neuronal and glial differentiation.* Dev Neurobiol, 2015. **75**(5): p. 452-62.
165. Jakovcevski, M., et al., *Neuronal Kmt2a/Mll1 histone methyltransferase is essential for prefrontal synaptic plasticity and working memory.* J Neurosci, 2015. **35**(13): p. 5097-108.
166. Ford, D.J. and A.K. Dingwall, *The cancer COMPASS: navigating the functions of MLL complexes in cancer.* Cancer Genet, 2015. **208**(5): p. 178-91.
167. Pekowska, A., et al., *A unique H3K4me2 profile marks tissue-specific gene regulation.* Genome Res, 2010. **20**(11): p. 1493-502.
168. Zhang, J., J. Parvin, and K. Huang, *Redistribution of H3K4me2 on neural tissue specific genes during mouse brain development.* BMC Genomics, 2012. **13 Suppl 8**: p. S5.
169. Meng, N., R. Machiraju, and K. Huang, *Identify Critical Genes in Development with Consistent H3K4me2 Patterns across Multiple Tissues.* IEEE/ACM Trans Comput Biol Bioinform, 2015. **12**(5): p. 1104-11.
170. Sanchez, R. and M.M. Zhou, *The PHD finger: a versatile epigenome reader.* Trends Biochem Sci, 2011. **36**(7): p. 364-72.
171. Asensio-Juan, E., C. Gallego, and M.A. Martinez-Balbas, *The histone demethylase PHF8 is essential for cytoskeleton dynamics.* Nucleic Acids Res, 2012. **40**(19): p. 9429-40.
172. Fortschegger, K., et al., *PHF8 targets histone methylation and RNA polymerase II to activate transcription.* Mol Cell Biol, 2010. **30**(13): p. 3286-98.
173. Qi, H.H., et al., *Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development.* Nature, 2010. **466**(7305): p. 503-7.

174. Laumonnier, F., et al., *Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate.* J Med Genet, 2005. **42**(10): p. 780-6.
175. Wieser, R., *The oncogene and developmental regulator EVI1: expression, biochemical properties, and biological functions.* Gene, 2007. **396**(2): p. 346-57.
176. Kataoka, K. and M. Kurokawa, *Ecotropic viral integration site 1, stem cell self-renewal and leukemogenesis.* Cancer Sci, 2012. **103**(8): p. 1371-7.
177. Barjesteh van Waalwijk van Doorn-Khosrovani, S., et al., *High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients.* Blood, 2003. **101**(3): p. 837-45.
178. Cao, F., et al., *Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia.* Mol Cell, 2014. **53**(2): p. 247-61.
179. Heerboth, S., et al., *Use of epigenetic drugs in disease: an overview.* Genet Epigenet, 2014. **6**: p. 9-19.
180. Gajer, J.M., et al., *Histone acetyltransferase inhibitors block neuroblastoma cell growth in vivo.* Oncogenesis, 2015. **4**: p. e137.
181. Congleton, J., et al., *ATRA-induced HL-60 myeloid leukemia cell differentiation depends on the CD38 cytosolic tail needed for membrane localization, but CD38 enzymatic activity is unnecessary.* Exp Cell Res, 2011. **317**(7): p. 910-9.