

Epigenetic and Transcriptional Dysregulation in Acute Myeloid Leukemia

By

Fatma Eryildiz

A DISSERTATION

Presented to the
Cancer Biology graduate program
and the Oregon Health & Science University
School of Medicine

in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

March 2022

School of Medicine
Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

Fatma Eryildiz

has been approved

Jeffrey Tyner (Mentor)

Andrew A. Adey (Committee Chair)

Elie Traer (Committee Member)

Hisham Mohammed (Committee Member)

Anupriya Agarwal (Examination Member)

Table of Contents

Abstract	1
CHAPTER 1.....	2
Epigenetic dysregulation in acute myeloid leukemia	2
1.1. Epigenetic hallmarks and disease control	2
1.2. High-throughput platforms to analyze epigenome: single cell epigenomics.....	9
1.3. Epigenetic heterogeneity is a hallmark of AML.....	11
Abstract	14
CHAPTER 2.....	16
Chromatin dynamics in acute myeloid leukemia subsets defined by genetic and drug response diversity.....	16
2.1. Introduction	16
2.2. Results.....	18
2.2.1. Single cell chromatin accessibility profiles from DNMT3A, TET2 and ASXL1 mutant primary patients.....	18
2.2.2. Epigenetic heterogeneity across primary patients with selected mutations and differential drug responses	21
2.2.3. Identifying transcription factor associated changes in AML tumors	24
2.2.4. Drug resistance specific chromatin changes upon JQ1 treatment in OCI-AML3 line	27
2.3. Summary and Discussion	29

2.4. Materials and Methods.....	31
Patient Samples.....	31
Cell Cultures	31
Ex vivo functional drug screens	31
Sci-ATAC-seq assay and sequencing	32
Abstract.....	40
CHAPTER 3.....	41
The Identification of ZNF687 variants in Acute Myeloid Leukemia	41
3.1. Introduction	41
3.2. Epigenetic control of transcriptional gene regulation via transcription factors	43
3.3. Zinc finger proteins in health and disease	44
3.4. Identification of ZNF687 germline mutations	45
3.5. Evidence for ZNF687 involvement in DNA damage response	47
3.6. Summary	52
3.7. Results.....	53
3.7.1. A cohort of primary patients from BeatAML suggests a pathogenic somatic frameshift mutation in <i>ZNF687</i>	53
3.7.2. ZNF687 expression is significantly upregulated in primary AML patients	56
3.7.3. ZNF687 frameshift causes loss of a functional C2H2 superfamily domain and related functional features	60

3.7.4. Expression of frameshift variant leads to a change in subcellular localization of ZNF687 and frameshift protein sequesters in cytoplasm.....	69
3.7.5. Differential expression identifies transcriptional changes in frameshift mutated cell lines.....	73
3.8. Materials and Methods.....	82
CHAPTER 4.....	89
Summary and Discussion.....	89
References.....	95

List of Figures

Figure 1. Waddington's Epigenetic Landscape	3
Figure 2. Chromatin state transitions	4
Figure 3. Panoramic view of acute myeloid leukemia	12
Figure 4. Cohort of AML patient samples and CD34+ primary bone-marrow derived cells from healthy controls	19
Figure 5. An overview of the AML patient cohort	20
Figure 6. Clustering of cells using cistopic clustering.....	22
Figure 7. Transcriptomic profiles of AML patients sequenced via sci-ATAC-seq.....	23
Figure 8. Orchestrate of distinct transcription factors for human hematopoiesis.....	25
Figure 9. chromvar generated TF motif deviation scores across all the cells plotted on UMAP... 26	
Figure 10. sciATAC on parental and resistant JQ1 cell lines	28
Figure 11 - ZNF687-dependent GCT/PDB formation	48
Figure 12 - Hypothetical illustration of ZNF687 involvement in bone cell proliferation as the downstream target of NFkB.	50
Figure 13 - Depiction of the recruitment of the NuRD complex to its genome-wide binding sites	51
Figure 14. Carriers of frameshift mutation in BeatAML cohort.....	53
Figure 15. ZNF687 recurrent frameshift mutation was first identified in a cohort of AML patients by whole exome sequencing and confirmed by Sanger sequencing	54
Figure 16 Frameshift mutation found in %1 of BeatAML patients.....	55
Figure 17. ZNF687 mRNA expression does not correlate with mutational status	57
Figure 18. ZNF687 mRNA is upregulated specifically in AML patients	58

Figure 19. ZNF687 mRNA expression does not correlate with clinical severity of the disease.....	59
Figure 20. Illustration of ZNF687 WT, FS and 937 mutations at mRNA and protein sequence	62
Figure 21. Multiple sequence alignment of missense and frameshift mutations along with WT ZNF687	63
Figure 22. Nostradamus predicts nuclear localization signals.....	64
Figure 23. In-silico analysis indicates that FS indel causes the loss of zf-C2H2_11 superfamily domain in native proteins	64
Figure 24. NetPhos3.1 is used to predict serine, threonine, and tyrosine phosphorylation sites	65
Figure 25. Bayesian Discriminant Method is used to predict high confidence ubiquitination sites in native ZNF687 protein	66
Figure 26. 3D folding of native ZNF687 protein was predicted in silico by the Alphafold2 software package with default parameters	67
Figure 27. AlphaFold predicted 3D folding of ZNF687 FS, ZNF687 P937R aligned with WT	68
Figure 28. ZNF687 frameshift generates a truncated protein product which migrates due to decreased molecular weight in HEK293T whole cell lysates.....	70
Figure 29. The expression of frameshift alters nuclear and cytoplasmic ratio of ZNF687 by sequestering almost exclusively in cytoplasm compared to WT	71
Figure 30. Immunocytochemical staining of transfected HEK293T cells	72
Figure 31. Scatter plot of Principal component analysis.....	75
Figure 32. Heat map of the expression patterns of ZNF687 WT, Empty, FS and 937 expressing HEK293T cells.....	76
Figure 33. GO BP ENSEMBL 92 was used to analyze fold-change values of all genes in database	77
Figure 34. KEGG Release 86.1 was used to analyze fold-change values of all genes in database.	78
Figure 35. Reactome-V64 was used to analyze fold-change values of all genes in database.....	79

Figure 36. GO MF ENSEMBL 92 was used to analyze fold-change values of all genes in database	80
Figure 37. Differential expression analysis showed the significant downregulation of KMT2D, a major enhancer mono-methyltransferase	81
Figure 38. Hypothetical mechanistic model of a putative link between ZNF687 and KMT2D	81

List of Tables

Table 1.1. NIB Buffer preparation	33
Table 1.2. 4X TAPS Buffer preparation	33
Table 1.3. Transposase neutralization buffer preparation	34
Table 1.4. PCR preparation	34
Table 1.5. PCR steps	35
Table 1.6. Custom sequencing primers and dilution	36
Table 3.1. ZNF687 Mutagenesis and Whole ORF sequencing Primers	87

Abbreviations

AML Acute myeloid leukemia

NGS Next generation sequencing

CGIs CpG islands

CML Chronic myeloid leukemia

MTHFR Methylenetetrahydrofolate reductase

MBDs Methyl-binding domain proteins

DMRs Differentially methylated regions

DNMTs DNA methyltransferases

H3K9me Histone H3 lysine 9 methylation

LSD1/KDM1 lysine specific demethylase 1

CH clonal hematopoiesis

TETs ten–eleven translocation enzymes

ncRNAs Non-coding RNAs

snRNAs small non-coding RNAs

lncRNAs long non-coding RNAs

ASXL1 Additional sex combs-like gene

PRC2 Polycomb repressive complex 2

PRC1 Polycomb repressive complex 1

H2AK119ub monoubiquitinated histone H2A at lysine 119

ATP Adenosine triphosphate

BET Bromodomain extraterminal

DAPI 4,6-Diamidino-2-phenylindole

DNA Deoxyribonucleic acid

PCA Principal component analysis

RNA Ribonucleic acid

ATAC-seq Assay for transposase-accessible chromatin

scATAC-seq Single-cell ATAC sequencing

sci-ATAC-seq Single-cell combinatorial indexed ATAC sequencing

NIB Nuclei Isolation buffer

TAPS [tris(hydroxymethyl)methylamino]propanesulfonic acid

DAPI 4,6-Diamidino-2-phenylindole

FACS Fluorescence-activated cell sorting

FANS flow assisted nuclei sorting

t-SNE t-stochastic neighborhood embedding

UMAP Uniform approximation and projection method

TF Transcription Factor

DA Differentially accessible

TCGA The Cancer Genome Atlas

TF Transcription factor

WT Wildtype

FS frameshift

PDB Paget's disease of bone

GCT Giant cell tumor of bone

NuRD Nucleosome remodeling and histone deacetylase

ChIP-seq Chromatin immunoprecipitation sequencing

DSBs Double stranded breaks

HR homologous recombination

SNPs single-nucleotide polymorphism

DEG Differentially expressed genes

NLS nuclear localization signals

HCC hepatocellular carcinoma

AI Artificial Intelligence

KEGG Kyoto Encyclopedia of Genes and Genomes

GO BP Gene Ontology Biological Process

GO MF Gene Ontology Molecular Function

Acknowledgements

I am very grateful to both Dr. Jeff Tyner and Dr. Andrew Adey who introduced me to the field and supported me throughout this journey to the unknown. Jeff's knowledge on the subject, commitment to mentoring combined with Andrew's remarkable curiosity continues to be inspirational for my life. Being an international student brings a lot of challenges but also opportunities and I am thankful for all the growth and learning opportunities. Grown up with a cultural awareness that teaching is sacred, I feel indebted to these great mentors for helping me to grow academically. They were there to provide support whenever I needed.

I am very thankful for the committee members Elie Traer, Hisham Mohamed and Anupriya Agarwal whose insights, continued support has enriched my PhD journey.

This thesis would not have been possible without the unrelenting and unconditional support of my dear father, and my husband. My dad, who supported my education abroad, deserves a huge THANK YOU. I truly thank my husband whom I can't wait to sail for new shores with indefinite love, adventure, and respect. Their encouragement, understanding, and love have kept me in check during the trials and tribulation of this PhD. *Sizleri çok seviyorum.*

Outside the pristine labs of the Knight institute, my extended family of friends have enriched my OHSU experience. I am monumentally indebted to my foster family Laura and George Fendel. Their company, friendship and support have made me stay up on my feet. Portland would not have been the same without them. Last but not least, Portland, the beautiful city, exactly what I could imagine

growing up, rainy, cloudy, more gloomy, less promising: You will always have a special place in my heart.

Abstract

Acute myeloid leukemia (AML) was one of the first cancers to be sequenced at the level of the whole genome and found to be highly complex among all hematologic malignancies. Molecular profiling of AML through targeted sequencing panels has become monumental in risk-stratifying AML patients and guiding clinicians seeking for tailored therapies for their patients in the context of precision oncology. Also, these cutting-edge tools have been helping in advancing our understanding of genetic and epigenetic determinants of AML.

Therefore, the extensive high-resolution genomic data generated to characterize AML have been instrumental in revealing the tremendous biological complexity of the disease, reflected in part by mutational, clonal, and epigenetic heterogeneity. This is further complicated by the primitive nonleukemic state of clonal hematopoiesis that is associated with an increased risk of developing a hematologic malignancy [1].

Here in this work, I explored epigenetic dysregulation AML by (1) Chromatin dynamics in acute myeloid leukemia subsets defined by genetic and drug response diversity [Chapter 2], (2) The Identification of ZNF687 variants in Acute Myeloid Leukemia [Chapter 3].

CHAPTER 1

Epigenetic dysregulation in acute myeloid leukemia

1.1. Epigenetic hallmarks and disease control

The epigenetic landscape model was introduced by Conrad Waddington in 1957 to describe the landscape of cellular development when there was very limited mechanistic understanding of it. A ball rolling down from the top of Waddington's Mountain to the bottom of a valley passing through a succession of valleys, illustrating the progressive restriction of cell differentiation during development (Figure 1) [2], [3].

In the modern era, the epigenetics is a broader term which is built "beyond/above" the DNA, a unit of universal information, that enforces cellular identities through several mechanistically complex processes. Epigenetic mechanisms stabilize gene expression programs to create an adaptive chromatin state to stimulate or repress gene function that is written in DNA. The definition of this chromatin state has been expanded by Dr. Brad Bernstein and Dr. Andy Feinberg. Dr. Bernstein's elegant model describes epigenetic marks to define a flexible energy wall that one cell needs to jump over to be in a different chromatin state reminiscent of Gibbs' Free energy diagram for enzymatic reactions (Figure 2) [4]. The plasticity is high when the energy bar is too low that practically allows a cell to easily move from one state to another, and vice versa.

With the advent of technology, especially with the next generation sequencing technologies (NGS), the level and state of these enzymatic modifications can be assayed to allow epigenetic, genome-wide chromatin profiling in healthy and diseased settings. Further, epigenetic information can be analyzed along with the genetic information to project the trajectories of the cells.

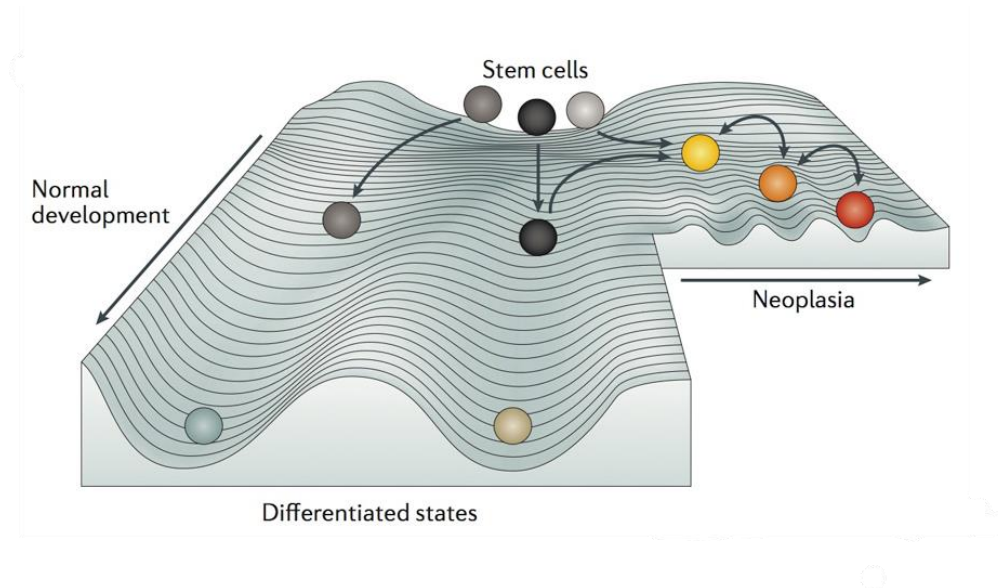


Figure 1. Waddington's Epigenetic Landscape.¹

These epigenetic events can occur in different ways such as DNA and RNA methylations, histone modifications and non-coding RNAs.

i. Methylation of DNA

¹ Reprinted/adapted by permission from Springer Nature: [Nature Reviews Genetics, 2016][3]

The link between the chemical modification of DNA and gene regulation was first introduced in the 1970s [5]. Since then, it has been the most extensively studied epigenetic mechanism that predominantly occurs in CpG islands (CpGIs) which renders about 60-80% [6] of them methylated by addition of a methyl (CH_3) group to the 5th carbon of

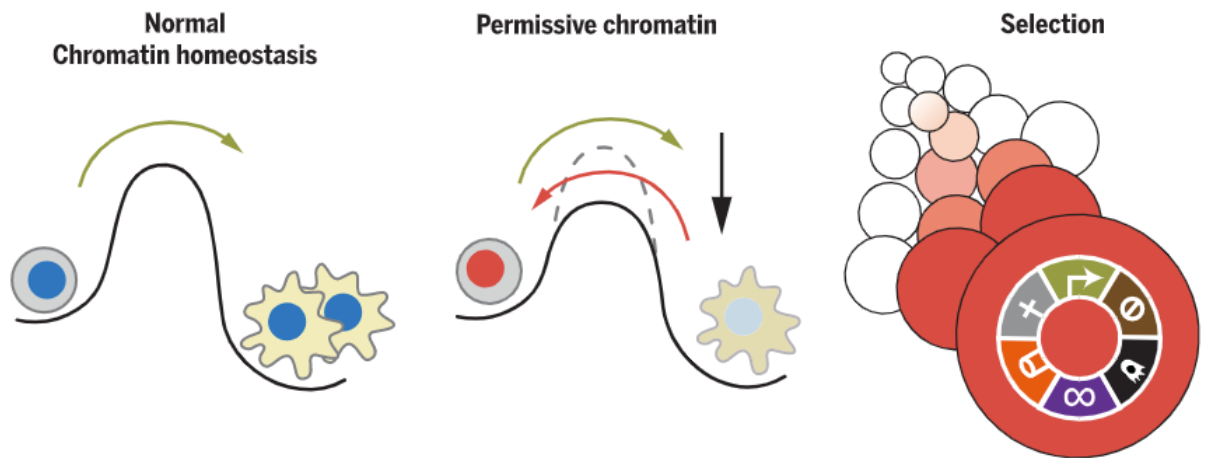


Figure 2. Chromatin state transitions.

2

cytosine in CpG dinucleotides. These CpGs are highly concentrated around gene regulatory elements like promoters, enhancers, and insulators. Although methylation is particularly important for mammalian development such as X chromosome inactivation in females, differentiation, genomic imprinting, silencing of genome repetitive elements, protection against the integration of viral sequences, transcriptional regulation, lineage specification and maintenance of cellular identity [7]–[10], I would like to detail here its significance in cancer.

² Reprinted/adapted by permission from Springer Nature: [Nature Reviews Genetics, 2016][3]

When the normal patterns of DNA methylation are disrupted, the cells manifest aberrantly high and low levels of these methylation marks relative to healthy control tissues, referred as hypermethylation and hypomethylation, respectively. In literature, the first methylation aberration is described as hypomethylation of human tumors [11]. Up until the development of sequencing technologies with high resolution and precision, it has been mostly associated with genome instability and aneuploidy, reactivation of transposable elements and loss of imprinting [6], [11], [12]. The cancer gene specific hypomethylation has been described in several cancer types [13]–[16]. One interesting study showed that two common polymorphisms in methylenetetrahydrofolate reductase (MTHFR) gene, a key enzyme in the regulation of folate metabolism, leads to global hypomethylation. The presence of these polymorphisms in individuals leads to reduced enzymatic activity of MTHFR and increased colorectal cancer prevalence according to dietary conditions of the studied populations. The incidence was only lower in patients consuming high dietary methionine which increases methylation content [17]–[19].

Although cancer-linked hypomethylation of regulatory regions remains to be elucidated, hypermethylation patterns in cancer and particularly in AML have been more frequently studied. Hypermethylation can silence tumor-suppressor genes by either blocking their transcription or recruiting protein complexes with high affinity for methylated DNA, methyl-binding domain proteins (MBDs) [20], [21].

Hypermethylation of promoters accompanied by a genome-wide loss of methylation is considered as a hallmark of cancer [22]. In the field of cancer epigenetics, the researchers identified these special sites where the DNA methylation is aberrant than “normal” and named as differentially methylated regions (DMRs).

A comprehensive large-scale study elucidating aberrant methylation patterns in AML displayed that these patterns can be classified under distinct subtypes. In this study [23],

they defined 16 subgroups by looking at the methylation profiles of 344 AML patients and matched these subgroups with clinical outcomes based on their methylation profiles. In addition to defining methylation subgroups, they also identified a panel of genes that show consistent and aberrant DNA methylation where most of them are hypermethylated [23].

From an enzymatic standpoint, methylation is catalyzed by a group of critically important enzymes, DNA methyltransferases (DNMTs). They can either maintain an existing methylation status (DNMT1) during replication or de-novo methylate previously unmethylated DNA (DNMT3a and DNMT3b recruited by DNMT3L) regardless of the cell replication status [24]. In addition to DNMTs, chromatin associated enzymes play critical role in regulation of DNA methylation. Histone H3 lysine 9 methylation (H3K9me) and lysine specific demethylase 1 (LSD1, also known as KDM1) regulates the methylation of DNMT1 gene to ensure its stability and fidelity of global DNA methylation during DNA replication [7]. Many DNMT3A mutations in AML has been identified in several patients matching with an intermediate-risk cytogenetic profile [25]. DNMT3A is a crucial gene for normal hematopoiesis so that, its deletion impaired differentiation and enhanced self-renewal in hematopoietic stem cells [26], [27] which is in line that it is one of the most significantly mutated gene in AML patients. About %20 of the patients carries at least one somatic mutation in DNMT3A gene and it is one of the leading factors in clonal hematopoiesis (CH) [28]. The presence of DNMT3A mutations found in other myeloid disorders (e.g. Myelodysplastic syndromes and myeloproliferative neoplasms) presents a proof that these occur early in clonal evolution [29]. Recent studies profiling DNMT3A mutations showed that DNMT3A-R882H mutation acts in a dominant negative manner where the mutant protein inhibits DNA methylation of wt-DNMT3A [25], [30].

DNA methylation can also be reversed (demethylation) by ten–eleven translocation (TET) family enzymes, namely TET1, TET2 and TET3. These enzymes work to oxidize

methylated cytosines to convert it into 5-hydroxymethylcytosine to revert the effects of gene repression and restore expression [6].

DNA methylation eraser enzyme TET2 is one of the other most significantly dysregulated gene in myeloid malignancies and T cell lymphomas. As in the case of DNMT3A, there are reports associating TET2 with poorer disease outcomes and prognosis in patients [31]–[33] and it is implicated as an early oncogenic mutation [34]–[36] displaying that it is needed for proper differentiation and HSC self-renewal [31], [37], [38].

ii. Histone modifications and chromatin remodeling

Histone tail modifications occur as posttranslational modifications of histone proteins. These modifications affect chromatin structure/conformation to facilitate or block the access of genes to transcription factors and regulatory proteins. These can happen in several forms; however, histone methylation and acetylation are the most common mechanisms that have a key role in dynamic modulation of gene expression [6].

Additional sex combs-like gene (*ASXL1*) is one frequently mutated epigenetic modifier in AML and people with CH and it was found to be mutated in 15-25% of patients with MDS and in 10-15% of patients with myeloproliferative neoplasms and AML [39]. Clinical studies consistently associated *ASXL1* mutations with poor prognosis and reduced overall survival [40]. *ASXL1* functions in both transcriptional activation and repression of genes involved in either differentiation or proliferation, specifically in self-renewal and proper differentiation of hematopoietic stem/progenitor cells.

ASXL1 gene encodes a protein product that is known to interact with polycomb repressive complex PRC2 that downregulates HOXA cluster genes. *ASXL1* also interacts with BAP1 complex that is part of the polycomb repressive deubiquitinase complex (PR-DUB). BAP1

plays a role to deubiquitinate monoubiquitinated histone H2A at lysine 119 (H2AK119ub) in concert with polycomb repressive complex PRC1. The interaction between BAP1 and wild-type ASXL1 results in repression of gene transcription. In addition to PRC2, and BAP1, ASXL1 interacts with EZH2, EED, and SUZ12 to promote trimethylation of H3 at lysine 27 (H3K27me3). Altogether, ASXL1 is an integral protein in various chromatin remodeling processes and histone modifications [41].

Mutations in the *ASXL1* gene are in the last exon, which encodes a protein domain involved in chromatin remodeling and methylated lysine binding [39], [42]. Most frequently, these mutations are frameshift and non-sense mutations that lead to C-terminal truncation and loss-of-function, reduced ASXL1 expression [39], eventually resulting in a myeloid differentiation bias that is strongly associated with leukemic progression [42].

Loss of ASXL-1 expression in myeloid neoplasm seems to result in reduced H3K27me3 at specific target loci, possibly through inhibition of PRC2 recruitment, and consequent overexpression of leukemia-promoting genes [39].

Double sided effects of loss of ASXL1 expression, therefore, has been associated with global dysregulation of HOX cluster genes expression [39], [43] and it has a role in both suppressing and activating HOX gene expression [44]. Although, how it contributes to leukemogenesis remains to be elucidated.

Due to the direct link between ASXL1 and BAP1, loss-of-function mutations in ASXL-1 affect epigenetic regulation through interruption of ubiquitin removal from specific histone lysine residues [29], [43]. Wang et al. demonstrated a frameshift mutation in ASXL1 leads to the expression of a truncated protein and it is more effective in stabilizing BAP1 and its targets than its wild-type counterpart does [45]. Accordingly, ASXL1 mutations and stabilized BAP1 led to the upregulation of genes encoding regulators of metabolism and

development which is interpreted by the authors that these two genes should induce similar cellular and transcriptional responses as BAP1 mutations have not been observed in AML but it is frequently mutated in solid tumors [45], [46]. Therefore, ASXL1 provides an alternative mechanism in leukemias to BAP1 mutation in solid tumors.

iii. Non-coding RNAs

Non-coding RNAs (ncRNAs) constitute more than 70% of the sizeable portion of the human genome [47] and exemplifies critical regulator roles; miRNAs, an example of small non-coding RNAs (snRNAs), in ensuring homeostasis of organisms and down-regulating majority of protein-coding genes' expression and long non-coding RNAs (lncRNAs) with more complicated roles as chromatin regulators, enhancers etc. [48]–[50].

lncRNAs can directly bind to chromatin-modifying enzymes to regulate gene expression by either facilitating the access of chromatin regulator proteins to targeted genomic regions or by blocking the access of chromatin modifiers to specific genomic sites. As an example, lncRNA TARID (*TCF21 antisense RNA inducing demethylation*) couples with GADD45A and regulates the activity of DNA demethylation in cancer cells to regulate transcriptional activity [51], [52].

1.2. High-throughput platforms to analyze epigenome: single cell epigenomics

An improved understanding of clinical benefits of rapid detection of epigenomic features with functional implementations enabled a plethora of tools characterizing these features

to be developed. There are several features which are in the development focus of these high throughput monitoring tools.

i. DNA modifications

The primary DNA modification monitored in tumor cells is methylation and methylation profiling employs two generic strategies, global assessment of methylation and evaluation of a panel of genes that include known candidate oncogenes. So far, there are several studies in AML that show a link between global methylation profiles and distinct chromosomal rearrangements.

ii. Histone modifications

The analysis of posttranslational modifications of the core histones in AML have been empowered by the development of ChIP (chromatin immunoprecipitation) that can identify these histone modifications at specific gene promoters.

ChIP coupled with real-time PCR allows the identification of histone modification patterns of single promoters and enhancers and the possibility of covering a region of several kilobases with interspaced amplicons has allowed researchers to define binding patterns of TFs and of regions of histone modifications with high sensitivity and specificity [53].

iii. DNA Accessibility

Eukaryotic DNA is compacted into the nucleus through interactions between DNA and histone proteins to form chromatin [54]. Generally, the chromatin and the nucleosome creates a significant obstacle to processes, as factors are physically unable to occupy regions on DNA that are occupied by histone proteins [55]–[57].

Regions of open chromatin, however, are accessible to DNA-binding proteins and are often found at regulatory regions of the genome [58], [59]. Identifying regions of the genome that are accessible therefore provides important information for genomic regulatory regions, such as enhancers, promoters, and insulators as well as describing the nucleosome structure of known regulatory regions of the genome [59].

Genomic methods used to examine chromatin accessibility have traditionally been based on preferential enzymatic digestion or modification of accessible DNA to DNA that is protected by bound histone proteins or transcription factors [60], [61].

More recently, the assay for transposase accessibility and deep sequencing (ATAC-seq) is employed as a powerful technology to assess accessible chromatin. ATAC-seq utilizes the use of a hyperactive Tn5 transposase to insert sequencing adapters into open regions of chromatin to then sequence those regions through next generation sequencing [62].

ATAC-seq has successfully been used to assess chromatin accessibility in single cells [63]–[65]. Howard Chang's and William Greenleaf's groups have published single cell accessibility studies in a several primary human cancers using ATAC-seq [66], [67].

Single cell monitoring can provide benefits such as detecting rare cell populations with clinical relevance that can easily be shaded by bulk methods and deciphering a full picture of highly heterogenous sub clonal populations presenting unique combinations of epigenomic features [68]–[70].

1.3. Epigenetic heterogeneity is a hallmark of AML

Next generation sequencing efforts have displayed a whole repertoire of genetic aberrations that causes great clonal and mutational complexity of AML [71].

Aberrant epigenetic patterning is a hallmark of cancer [72]. Especially in myeloid malignancies, more than 70% of those patients with AML, myelodysplastic syndromes, and asymptomatic people with clonal hematopoiesis harbor at least one mutation in genes that encode epigenetic regulators [46], [51], [52]. Many of these mutations were found in DNA methylation-related genes, chromatin modifiers or cohesin-complex genes.

Some of these mutations such as DNMT3A and ASXL1 are represented with a high allelic frequency and the studies on the mutant proteins expressed by these genes showed that the alterations in those genes result in broader changes to the epigenetic status of cells rather than just affecting their canonical targets [46]. This notion highlights the importance of the new and creative therapeutic strategies.

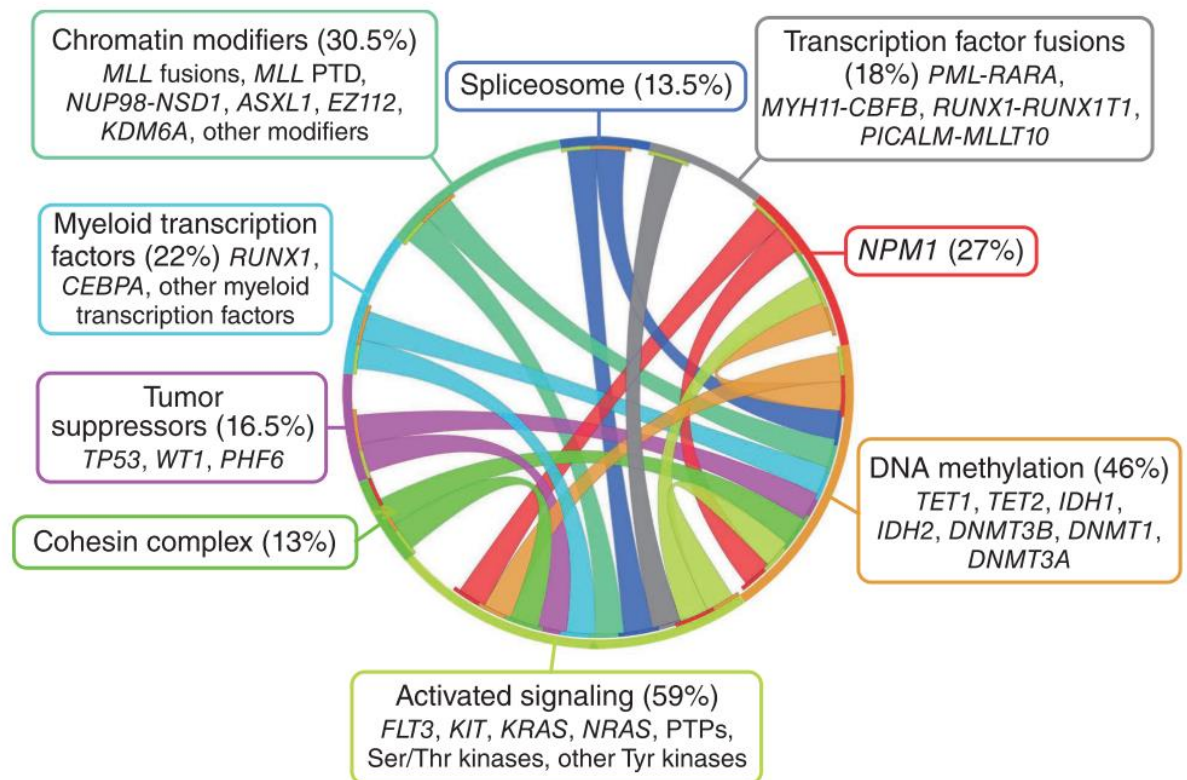


Figure 3. Panoramic view of acute myeloid leukemia.

AML patients are classified into methylation profiles with distinct biological features and distinct features, however, not all these classes could be matched with any mutation [73], resulting in an epigenetically heterogeneous population of leukemia cells. In addition to genetic diversity, this epigenetic diversity may subject the divergent cells to another level of selective pressure to compete with less adaptive cells in self-renewal and drug resistance [74].

It is now widely accepted that both genetic and epigenetic heterogeneity contribute to functional heterogeneity that can be associated with different clinical outcomes [74]. The emergence of high representation of mutations in epigenetic modifiers lead scientists to target these alterations along with genetic abnormalities by epigenome targeting drugs named as epi-drugs [47]. The reversibility of these epigenetic states makes them more attractive therapeutic targets for researchers.

³ Reprinted/adapted by permission from Springer Nature: [Nature Genetics, 2013][149]

Abstract

Acute myeloid leukemia (AML) is a complex disease, in which genetically diverse clonal events take place leading to intratumoral heterogeneity. This has been associated with the failure of durable responses in patients. Recent genomic advances revealed the landscape of AML features multiple epigenetic modifiers that are frequently mutated. Accumulating studies support the presence of distinct epigenetic cell states that prompt researchers to target these cell populations via inhibition of chromatin modifying proteins such as BRD4 in several cancer types. However, the functional impact of these epigenetic variations on chromatin architecture of cellular ecosystems during disease progression and emergence of drug resistance is still incompletely understood.

We applied single cell combinatorial indexing Assay for Transposase Accessible Chromatin sequencing (sci-ATAC-seq) to dissect the restrictive and/or permissive epigenetic cell states in 12 primary AML patient samples and 3 CD34+ primary bone-marrow derived cells from healthy controls. These patients were selected based on the mutational variations that they carry; 4 samples from each mutational group, ASXL1, DNMT3A, and TET2. In addition, each mutational group included samples with low, moderate, or high sensitivity to a BRD4 small molecule inhibitor, JQ1, detected using an ex vivo drug sensitivity assay. With this strategy, we generated 27,696 high-quality single cell libraries with a median mapped read depth of >11,000. Using a recently published cisTopic computational framework [75] along with transcription factor binding motif accessibility analysis, we attempted to define cis-regulatory regions that drive distinct cell

states that might be accountable for differential treatment responses. Both analyses revealed a vast amount of epigenetic heterogeneity across patients regardless of their mutational status. Interestingly, we noted that CD34+ healthy control cells showed a distinct cell state in their global chromatin accessibility compared to cells from AML patients. Also, we discovered multiple epigenetic sub clonal populations within each individual AML patient sample, sharing similar sub-clustering patterns and cell-type specific topics across all the samples. Overall, these findings are promising for future experimental designs to unravel the cellular ecosystem of AML tumors and improve novel strategies for precision medicine for the treatment of AML.

CHAPTER 2

Chromatin dynamics in acute myeloid leukemia subsets defined by genetic and drug response diversity

2.1. Introduction

Hematopoiesis is a tightly regulated process requiring both self-renewal and differentiation of hematopoietic stem cells (HSCs) to maintain diverse mature blood cell types. Disruption of normal hematopoiesis can lead to pathologic states, including leukemia [76]. This well-defined and organized hierarchical structure of hematopoietic cell types and easier tissue sampling provide great benefits studying blood cancers to decipher diseased vs normal cell states [77].

AML has been shown to be comprised of genetically and epigenetically heterogeneous multiple subclones [1], [78] where each can contribute to functional diversity such as therapy response and this can be modulated under the selective pressure that leads to clonal evolution over time, treatment, and acquisition of further aberrations. Deciphering this layered heterogeneity in AML can elevate our understanding of AML heterogeneity and reflect in developing better targeted therapies.

Moreover, the level of epigenomic heterogeneity makes the use of single-cell tools measuring this epigenetic diversity inevitable. Chromatin accessibility affects gene expression by altering interactions of transcriptional regulatory factors with their targets. Therefore, assessing the chromatin accessibility is a powerful tool to define cancer specific cell trajectories that are shaped by the interactions and transcriptional regulations in a complex tumor ecosystem.

To comprehensively analyze epigenetic layer of AML heterogeneity, we have compiled chromatin accessibility data on thousands of cells by sci-ATAC sequencing. To characterize the different cell states that can be associated with a specific drug response and match these states with disease markers in AML tumors I mainly used single cell chromatin accessibility data and I have integrated several datasets including exome sequencing, ex-vivo functional drug screening and RNA sequencing along with several publicly available ATAC-seq data sets.

2.2. Results

2.2.1. Single cell chromatin accessibility profiles from DNMT3A, TET2 and ASXL1 mutant primary patients

We have compiled a list of patients where we can study clonal dynamics in patients' response to a BRD4 inhibitor, JQ1. These patients are selected based on their sensitivity against JQ1 and mutational status to allow us to identify key biological features of JQ1 sensitivity that can be dissected by using genetic diversity.

We have predicted that JQ1, (BRD4 inhibitor) should have global effects on the epigenome of AML patients by replacing putative elements that are responsible for the acetylation of histones. Thus, the loss of acetylation can be matched with a particular mutation such as ASXL1, TET2 or DNMT3A. We have performed single cell combinatorial indexing assay for transposase accessible chromatin sequencing (sci-ATAC-seq) on a cohort of AML patient samples and CD34+ primary bone-marrow derived cells from healthy controls. We have included 4 samples from each mutational group, ASXL1, DNMT3A, and TET2 (n=12). Each mutational group included samples with varying levels of JQ1 response, namely, low, moderate, or high sensitivity detected using an ex vivo drug sensitivity assay (Figure 4). Each patient's clinical blast percentages have been matched with its mutational status and secondary aberrations (Figure 5).

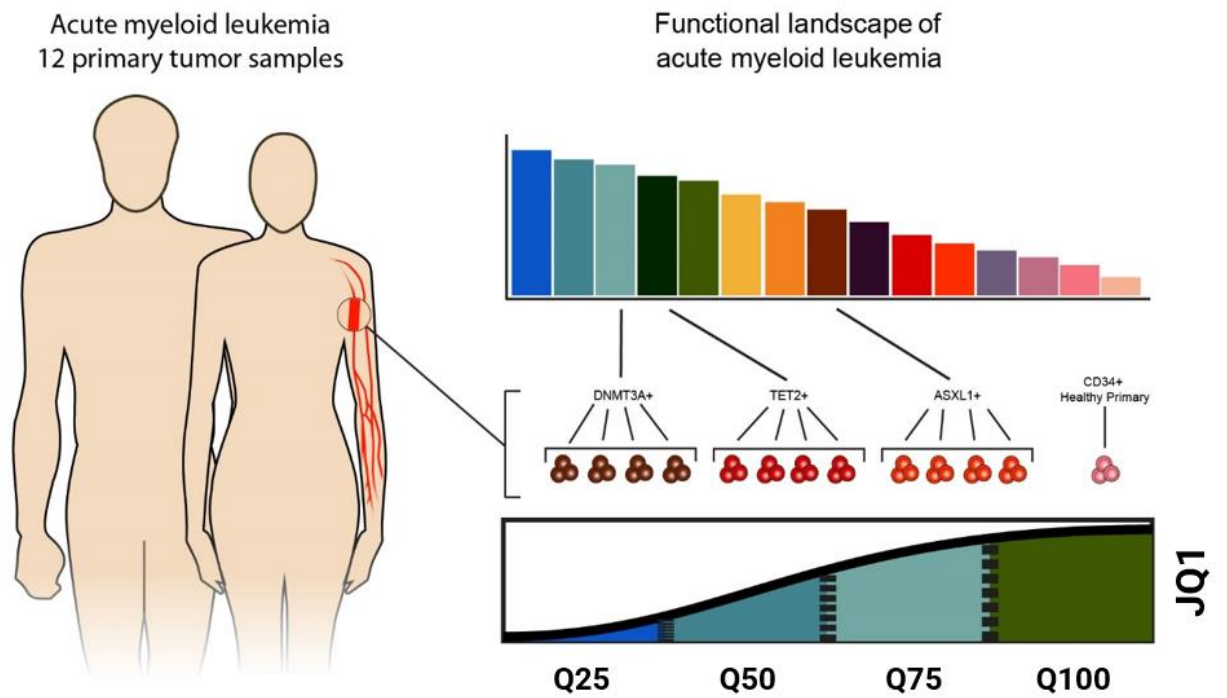


Figure 4. Cohort of AML patient samples and CD34+ primary bone-marrow derived cells from healthy controls. We have selected top recurrent mutations in AML that occurs in epigenetic modifiers (DNMT3A, TET2, ASXL1) and each mutation subgroup included 4 patient samples with varying levels of sensitivity against BRD4 inhibitor JQ1.

Patient	Gender	Cells	Blast Percentage
16-00094	F	3234	
15-00777	M	2276	
17-00195	F	366	
17-00328	M	1874	
15-00872	F	2427	
15-00979	F	1962	
17-00613	F	1467	
16-00494	F	2135	
16-00109	M	909	
16-00731	F	2776	
17-00878	M	2427	
18-00029	M	2842	
Total:	12 patients	24,695 cells	

Clinical blast count (%)

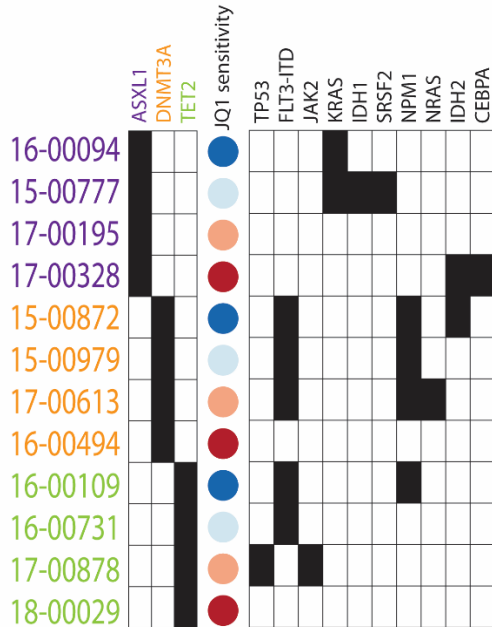
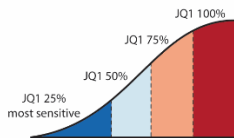


Figure 5. An overview of the AML patient cohort.

The number of cells reflect single cell ATAC-seq profiles that passed our quality thresholds (top). The pie charts next to those represent clinical blast percentages (top). The bottom panel shows genetic alterations detected in this cohort by exome-sequencing and the level of JQ1 sensitivity by ex-vivo drug screening.

2.2.2. Epigenetic heterogeneity across primary patients with selected mutations and differential drug responses

We generated a counts matrix using sci-ATAC-seq data containing cells as columns and peaks as rows via scitools matrix-counts function and filtered to exclude rows represented in less than 100 cells with -R 100 flag and columns represented in less than 1000 reads with -C 1000 flag. We then binarized this counts matrix and applied cistopic, cis-regulatory topic modelling on single cell ATAC-seq data [75].

This probabilistic tool calculates the probability of a region belonging to a topic and a topic belonging to a cell by simultaneously clustering cells and regions to define common cis-regulatory themes (topics) [75].

We have used these two probability matrices for downstream projections of cell states via uniform manifold approximation and projection (UMAP) (Figure 6) and explored several sites that can signal distinct biological cues.

Later we explored whether transcriptional signatures drive a specific subcluster which may help in resolving highly complex heterogeneity in this dataset and we have performed hierarchical clustering and heatmap generation by using RNA sequencing profiles annotated with several clinical features (e.g., major driver genes in AML; TP53, FLT3-ITD, JAK2, KRAS, IDH1, SRSF2, NPM1, NRAS, IDH2, CEBPA, gender, age, ELN 2008 classification, blast %) from these patients (Figure 7), however we did not observe any distinct transcriptional feature that could impact the cell clustering.

In summary, single cell profiling of AML tumors revealed a vast amount of epigenetic heterogeneity across patients however this was regardless of their mutational or drug resistance status.



Figure 6. Clustering of cells using t-SNE clustering and annotated and colored by Patient IDs and JQ1 sensitivity percentages next to them.

Patients were not subclustered based on their mutational or drug sensitivity status, however, the distinct regulatory programs drive healthy cells apart from other cell clusters derived from patient cells.

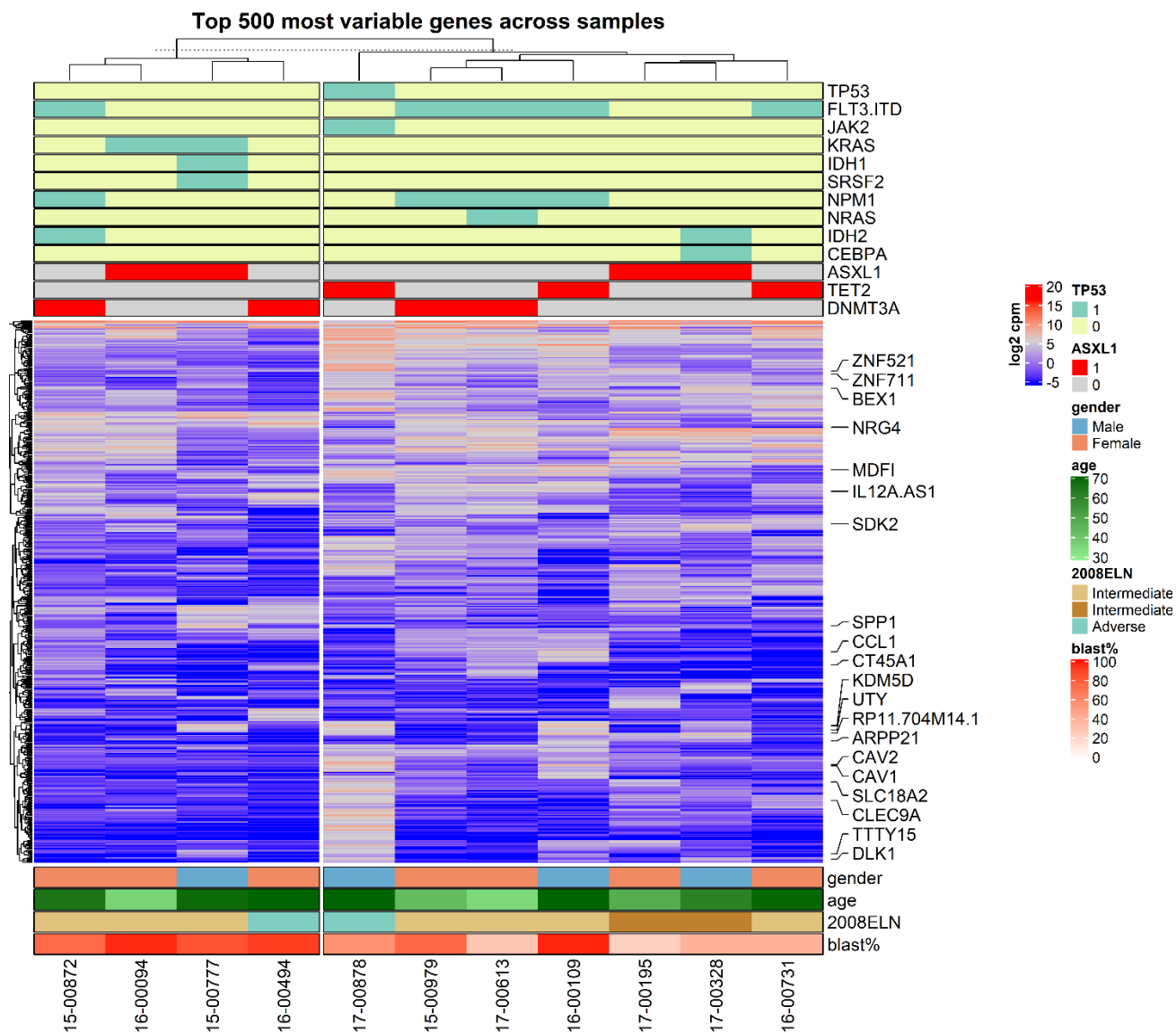


Figure 7. Transcriptomic profiles of AML patients sequenced via sci-ATAC-seq.

Key driver genes in AML and several clinical characteristics are used to annotate and cluster.

2.2.3. Identifying transcription factor associated changes in AML tumors

Human hematopoiesis is orchestrated by activities of distinct transcription factors which leads to terminal differentiation of blood cell types into mature cells (Figure 8).

We applied the chromVAR [79] to our data to infer changes in global transcription factor motif accessibility across our cell populations. Chromvar aggregates chromatin accessibility binary scores across peaks that share a common feature like a TF motif and applies bias correction to generate deviation scores for each cell and/or sample. This provides information on the putative binding of transcription-factors and consequently the possible ongoing biological processes in cell populations. The bias corrected motif deviation scores were plotted on the UMAP with the scitools plot-dims -M option for visualization (Figure 9). According to chromVAR global motif deviation z-scores, the cell clusters enriched for cell-type associated TFs are highlighted in each panel (top-bottom, CEBPA, SPI-1, GATA-1).

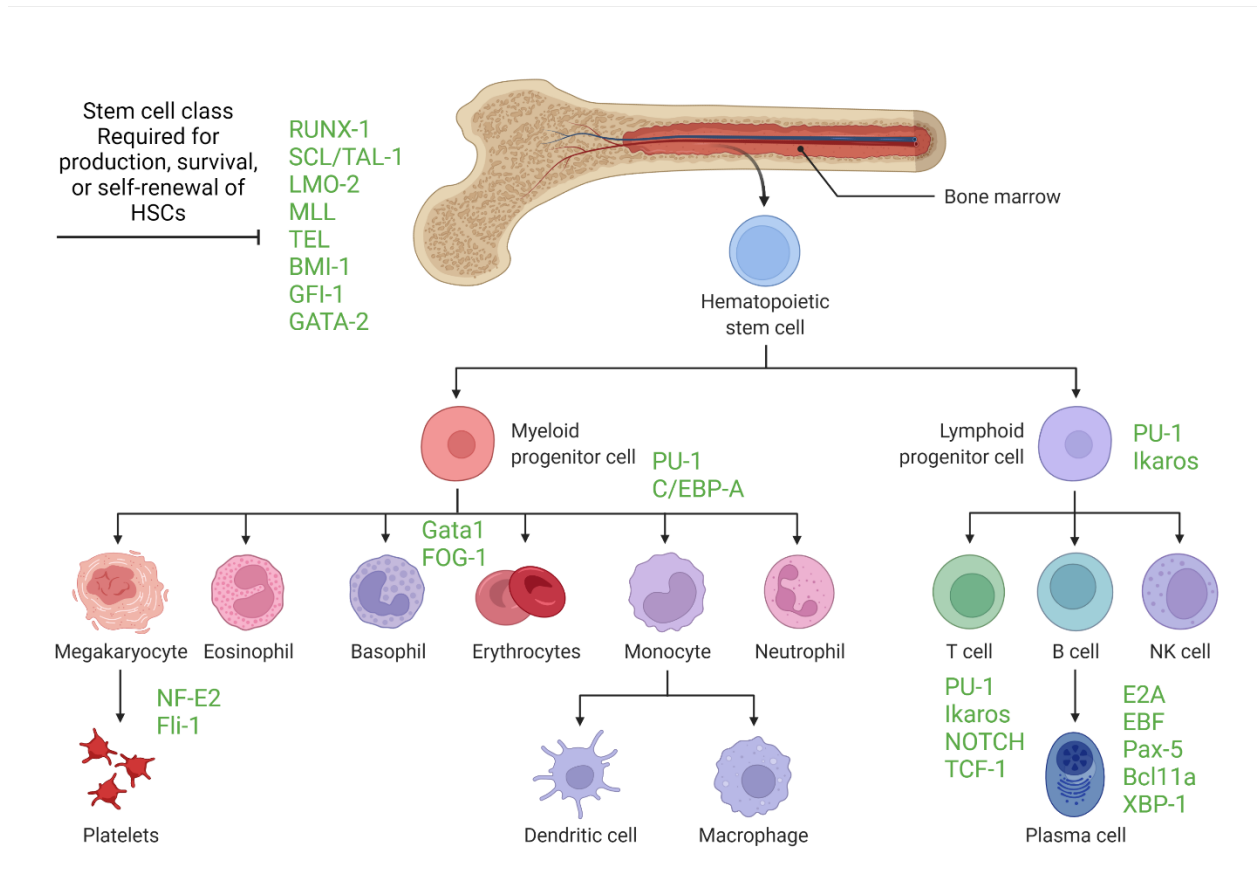


Figure 8. Orchestrate of distinct transcription factors for human hematopoiesis.

Created with BioRender.com

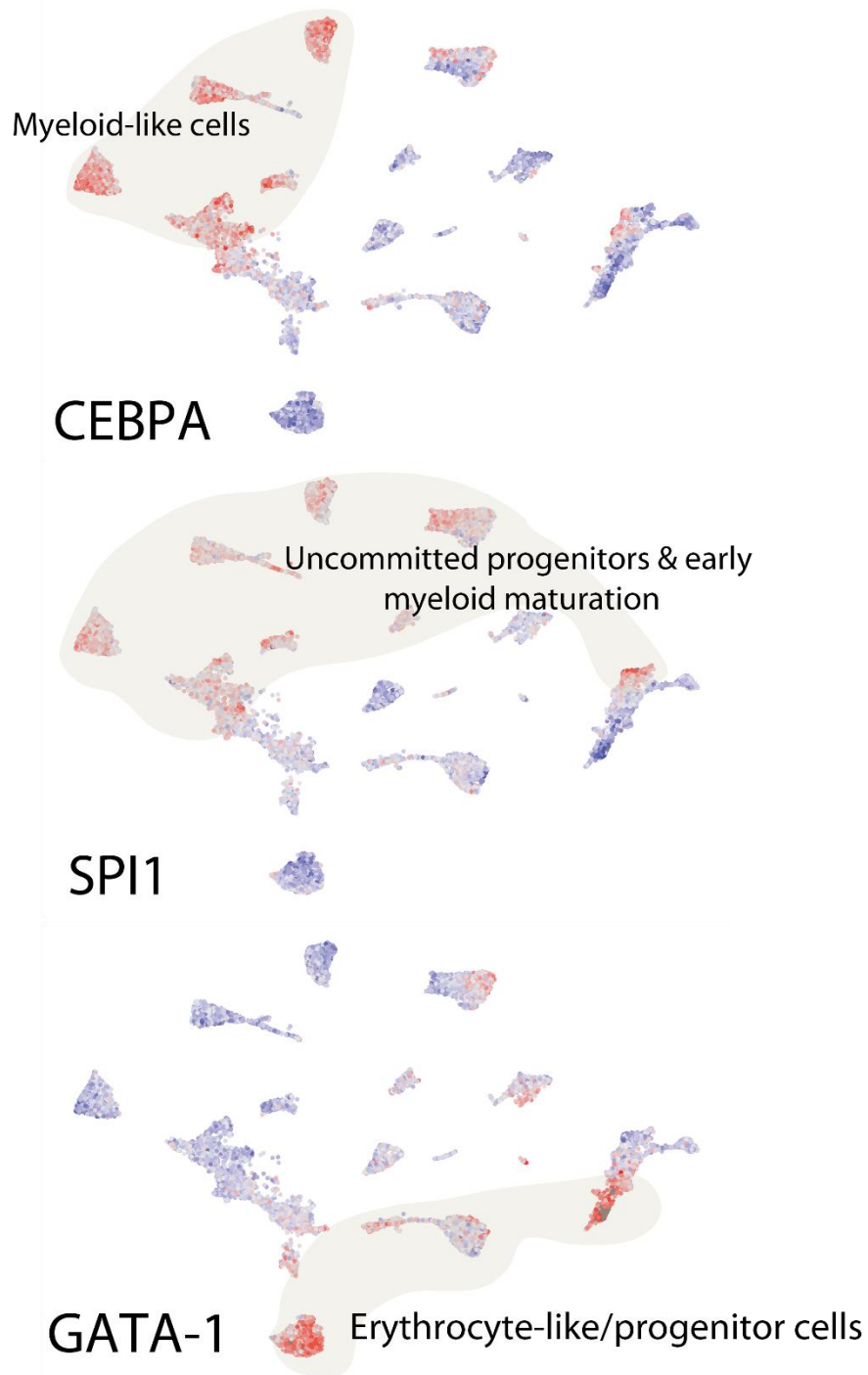


Figure 9. chromvar generated transcription factor motif deviation scores across all the cells plotted on UMAP.

Shades of red color signal for higher accessibility/activity for the annotated TF and shades of blue signal low accessibility/activity. The scale is (± 3 , red-blue)

2.2.4. Drug resistance specific chromatin changes upon JQ1 treatment in OCI-AML3 line

We have generated a second sciATAC-seq library on parental AML2 (JQ1-responsive) and resistant AML2 cell lines with and without treatment with JQ1 for 16 hours.

UMAP clustering of accessibility matrices from these samples showed that JQ1 resistance in vitro exhibits an altered epigenetic profile between parental and resistant cell lines that need to be explored further (Figure 10).

The future analysis can be expanded into: (1) the investigation of global heterogeneity in the system by Shannon Entropy, (2) detecting the loss and gain of epigenetic features, (3) use of these features in reconstructing trajectories of cell fates in the path of gaining resistance, (3) implementation of tools designed specifically for hematopoietic datasets such as ArchR suite [80] that can better deal with technical challenges of sciATAC data such as the sparsity and allowing more robust cell type identification by integrating scATAC with scRNA seq datasets for a given tissue, (4) the last but the most importantly, the overlay of the cell line data with patient samples.

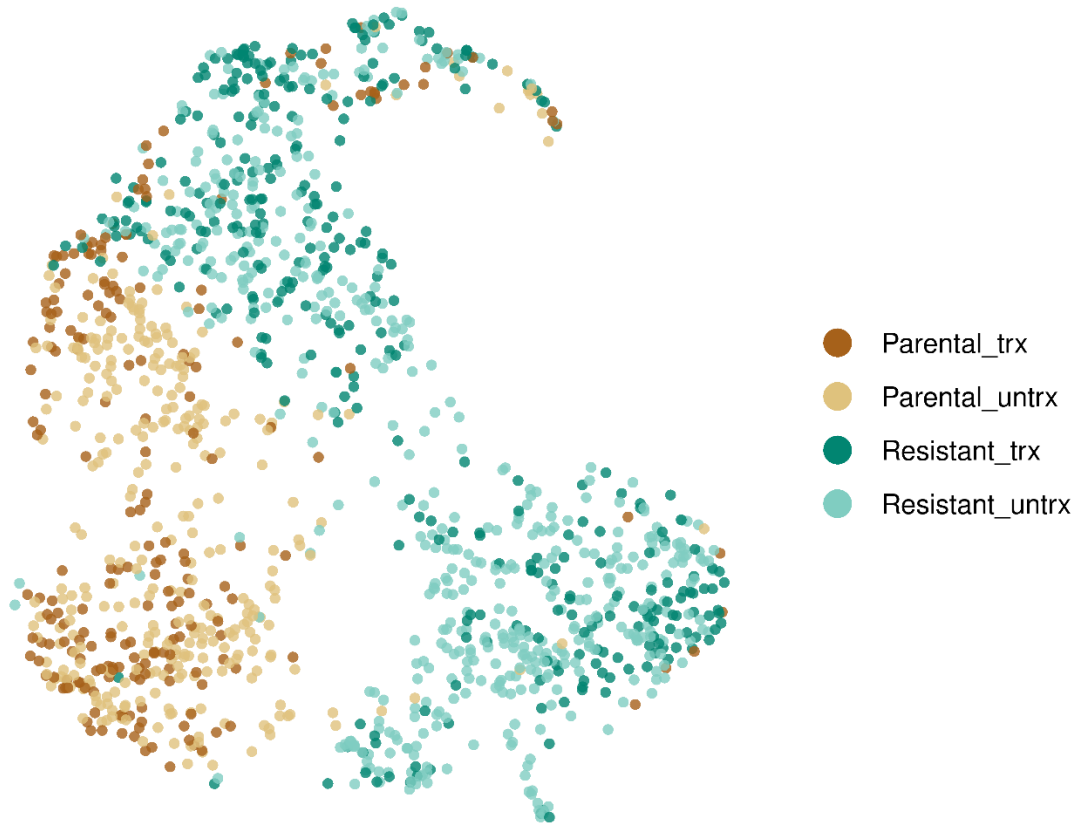
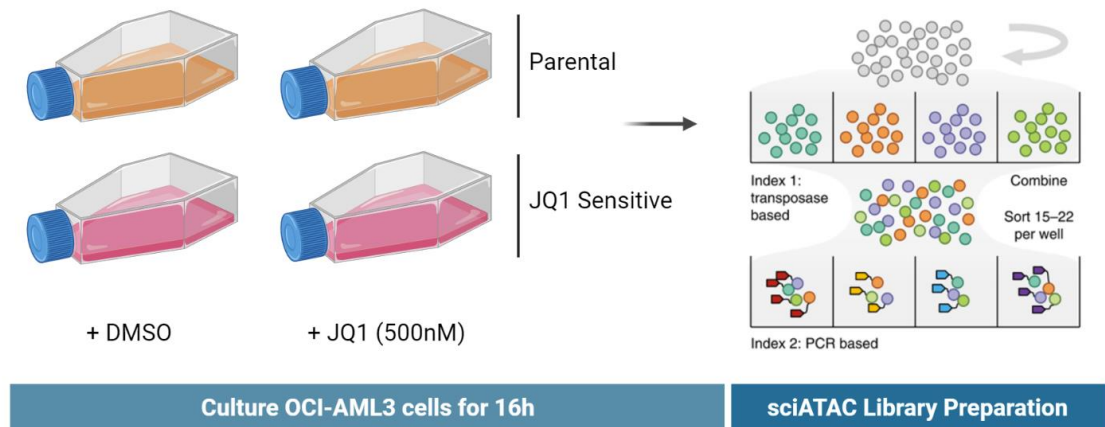


Figure 10. sciATAC on parental and resistant JQ1 cell lines.

sciATAC protocol schematic (Top).

We have generated 1394 cells with >500 accessible peaks per cell and called 206,954 peaks/features accessible in min 10 cells per peak (UMAP clustering, bottom).

2.4. Summary and Discussion

Recurrent mutations in epigenetic regulators are common events in AML that has been recently published and epigenetic therapies targeting these regulators are emerging.

A major challenge for the clinical management of AML is the heterogeneous that adds to its evolving complexity. This heterogeneity has been associated with the failure of durable drug responses.

The emergence of single cell epigenomic technologies has introduced the exciting possibility of linking genetic and transcriptional heterogeneity in the context of cancer biology. The next challenge is to leverage these tools for improved patient outcomes. We have applied single cell ATAC-seq to address the regulatory dynamics in terms of a small molecule inhibitor response that could be associated with tremendous complexity of AML. Deciphering the epigenetic layer of AML complexity can elevate our understanding of disease heterogeneity and reflection of this heterogeneity in developing better suited therapies for patients.

Chromatin accessibility at the single cell resolution is a powerful information that can define AML specific cell trajectories that are shaped by the interactions and transcriptional regulations in a complex tumor ecosystem.

To comprehensively analyze epigenetic and genetic layer of AML heterogeneity, we have compiled chromatin accessibility data on cells from a cohort of AML patient samples by sci-ATAC sequencing and assessed the relationship between drug response and epigenetic features that can be associated with it. However, no common epigenetic feature could be associated to drive or affect drug response in selected patients. Overall, we believe that the cohort of patient samples were extremely heterogeneous that could not

be stratified due to the several factors that can contribute to JQ1 response rather than mutational status alone as in predicted initially.

Further, we have implemented sci-ATAC-seq on JQ1 resistant cell lines and showed that resistance drives a distinct epigenetic state that could be further explored by (1) the investigation of global heterogeneity in the system by Shannon Entropy, (2) detecting the loss and gain of epigenetic features, (3) use of these features in reconstructing trajectories of cell fates in the path of gaining resistance, (3) implementation of tools designed specifically for hematopoietic datasets such as ArchR suite [80] that can better deal with technical challenges of sciATAC data such as the sparsity and allowing more robust cell type identification by integrating scATAC with scRNA seq datasets for a given tissue, (4) the last but the most importantly, the overlay of the cell line data with patient samples.

2.3. Materials and Methods

Patient Samples

All patient specimens were obtained after informed consent with IRB approval by Oregon Health & Science University.

Cell Cultures

OCI-AML2 cells were obtained from ATCC. All cell lines were maintained in RPMI, 20% FBS, L-Glutamine, penicillin/streptomycin. JQ1 resistant OCI-AML2 cells were generated as described here in detail [81].

Ex vivo functional drug screens

Upon receipt of a patient sample, plates were thawed at 37°C, 5% CO₂ for 1 hour and centrifuged at 800 × g. Subsequently, patient samples were suspended into culture media at a concentration of 1×10^6 cells per mL. Cells were incubated for 3 days at 37°C, 5% CO₂ and subjected to a CellTiter 96 AQueous one solution cell proliferation assay (MTS) (Promega). Each plate contained 7 wells without any drug. The average absorbance value of these 7 wells was used for data normalization and the kill curve of each drug gradient was assessed relative to this average no-drug point [82].

Sci-ATAC-seq assay and sequencing

5e6 cells in media were centrifuged in 4°C at 1100 rpm for 5 mins, media was aspirated, and cells were re-suspended and washed with ice-cold PBS twice. Nuclei Isolation buffer (NIB) was prepared according to Table 1.1. Cell pellet was suspended in 1ml NIB by pipetting and incubated on ice for 15mins for nuclei dissociation. Cells were counted using a hemacytometer and diluted with NIB to a concentration of 500 nuclei per ul.

Tagmentation of nuclei

To get to the 3,500 cells per well in a 96-well plate, 7.5ul cells in NIB mix were distributed to a 96-well tagmentation plate that has been prepared according to Table 1.2. 2.5 ul 4X fresh TAPS buffer per well was added. 96-well plate was centrifuged at 300 x g for 2 mins at 4°C, to ensure that nuclei were properly suspended in TAPS Buffer. Plates were kept on ice during the whole assay. Subsequently, 1-1.5 ul of uniquely indexed transposase complexes were added to each well of a 96-well plate. 96-well plate was sealed and incubated in 55°C at 300rpm on a thermomixer for 15 minutes. Following the completion of tagmentation the nuclei, plate was placed on ice immediately to stop reaction and over transposition.

Tagmented single nuclei sorting and PCR

96-well sort plates were prepared according to Table 1.3. To each well, a combination of 2.5 µL 10 mM i5 and 2.5 µL 10 mM i7 sci- barcoded PCR primers were added to generate a matrix where each well has a unique combination of i5 and i7 indexes.

All wells of the tagmentation plate were pooled into a 15 ml conical tube on ice. 2ul per ml DAPI (5mg/mL) was added for nuclear staining before proceeding with the flow cytometry. DAPI-stained nuclei were then sorted into 96-well sort plates, using a Sony SH800 Fluorescence-activated cell sorting (FACS). Keeping the plates and tagmented cells on

ice, gating was performed to isolate a clean population of single nuclei. 15 tagged nuclei per well were sorted into each 96-well plate by flow assisted nuclei sorting (FANS).

Table 1.1. NIB Buffer preparation.

Final Concentration	Stock Concentration	Volume of Stock
10 mM Hepes, pH 7.2	1M Hepes, pH 7.2	500 uL
10 mM NaCl	5M NaCl	100 uL of 5M NaCl
3mM MgCl ₂	1M MgCl ₂	150 uL of 1M MgCl ₂
0.1 % Igepal	10% Igepal	500 uL
0.1 % Tween	10% Tween	500 uL
ddH ₂ O		To 50 mL (add 48.25 uL)
Protease Inhibitor (Only if dealing with tissue)	2 tablets	2 tablets of Pierce Protease Inhibitor Tablets, EDTA-Free

Table 1.2. 4X TAPS Buffer preparation.

Final Concentration	Stock Concentration	Volume per ml	Volume per 96-well plate (x110)
TAPS pH 8.5	1M	132mM	132ul
Potassium Acetate	5M	264mM	52.8ul
Magnesium Acetate	1M	40mM	40ul
Dimethylformamide (DMF)	100%	64%	640ul
H ₂ O			135.2ul

Table 1.3. Transposase neutralization buffer preparation.

Final Concentration	Stock Concentration	Volume per uL
0.59mg/ml	20mg/ml BSA	0.25ul
0.059% (w/v)	1% SDS (w/v)	0.5ul
Up to 8.5ul	ddH ₂ O	7.75ul

Consequently, sort plates were spun in 4°C at 300 x g briefly to ensure that nuclei were properly suspended in neutralization buffer. 96-well neutralization buffer containing plates were incubated at 55°C for 20 minutes to denature the transposase in sorted nuclei. After denaturation, PCR reagents per well was added according to Table 1.4. and Real-time PCR (RT-PCR) was performed on a BioRad CFX machine according to Table 1.5. Libraries were amplified until majority of the wells begin to plateau around 14-18 cycles.

Table 1.4. PCR preparation.

PCR Reagents	Volume per ml
5X Kapa Hifi Buffer (GC Buffer)	5 ul
10mM Kapa dNTP Mix	0.75 ul
1U/uL Kapa HiFi DNA Polymerase	0.5 ul
100X SYBR Green I	0.25 ul
H ₂ O	5 ul

Post-PCR library column cleaning

10ul of each PCR well has been pooled into a 15 ml conical tube and the product was cleaned according to the manufacturer's protocol (QIAquick PCR Purification Kit, Qiagen).

Table 1.5. PCR steps.

Step	Temperature	Time
Initial denaturation	72°C	5 mins
Denaturation	98°C	30 sec
Primer Annealing	63°C	30 sec
Extension	72°C	30 sec
Final extension	72°C	1 min

SPRI bead size selection

Eluted DNA were further cleaned and size selected via SPRI beads to remove DNA below 200 bp. Equal volume of DNA and SPRI beads (1:1) were mixed gently by pipetting and incubated at room temperature for 15 minutes. DNA is incubated on a magnetic rack for 2 mins until all the beads have pelleted and supernatant is clear. for 15 minutes. DNA is incubated on a magnetic rack for 2 minutes to pellet bead-DNA complexes. Supernatant was discarded by being careful not to touch the pellet and beads are washed twice with 80% fresh ethanol. After second wash, the residual ethanol was discarded as much as possible being careful not to disrupt the pellet. With the tube lid open, the beads are incubated for 5-10 minutes until all the liquid evaporates. The pellet was resuspended in 30ul 10mM Tris HCl, pH 8.0 and DNA is eluted by incubating 10 minutes. Using a magnetic rack, beads are pelleted and DNA is collected as supernatant.

Quality control via TapeStation

Following this, dsDNA libraries were quantified prior to TapeStation via Qubit HS assay according to the manufacturer's protocol (Qubit dsDNA HS and BR Assay Kits, Invitrogen). Libraries were diluted to 1 ng/ul and run-on Agilent 4200 TapeStation instrument using

Agilent D5000 assay. According to the Tapestation assay results, libraries were diluted in the 120-1000bp molarity range.

Sequencing on Nexseq500 with custom chemistry

NextSeq 500/550 High Output Kit v2.5 (150 Cycles, Illumina, #20024907), the flow cell and HT1 buffer, 1N NaOH were fully thawed in RT for 1 hour. 1nM sci-ATAC libraries were combined with 0.2N NaOH and incubated at room temperature for 5 mins. 200ul 200mM Tris-HCl is added to the library mixture and 940ul HT1 solution is added to dilute libraries to 20 pM. Library is diluted to 1.1 pM concentration in HT1 buffer for a standard sciATAC sequencing assay. Custom primers used were diluted according to the Table 1.7. Reagent cartridge was loaded into NextSeq500 instrument and ran for 50 cycles of Read 1, 50 cycles of Read 2, 18 cycles of Index 1 and 18 cycles of Index 2.

Table 1.6. Custom sequencing primers and dilution.

Final Diluted Reagent	Primer Identity	Oligo Name	Loading Concentration	Volume of Stock Oligo	Volume of HT1 Dilutant	Total Volume
Custom Read 1 Primer Dilution	Read 1 Primer	AA221_CPT_R1_Primer	0.6uM	9uL	1491uL	1500uL
Custom Read 2 Primer Dilution	Read 2 Primer	AA222_CPT_R2_Primer	0.6uM	9uL	1491uL	1500uL
Custom Index Primer Dilution	Index 1 Primer	AA223_CPT_I1_Primer	0.6uM (when combined)	18uL	1482uL	1500uL
Custom Index Primer Dilution	Index 2 Primer	AA224_CPT_I2_Primer	0.6uM (when combined)	18uL	1482uL	1500uL

Raw processing of sci-ATAC-seq data and scitools platform [83]

Once sequencing is complete, BCL files were converted to FASTQ files using bcl2fastq (2.19.0). Following fastq file generation, reads were demultiplexed and formatted to accept index barcodes as unique cell identifiers (Cell ID) via scitools platform developed by Adey Lab (fastq-dump and fastq-split functions). Scitools package related scripts can be found at <https://github.com/adeylab/scitools>. Next, demultiplexed reads were aligned to the reference genome hg38 (GRCh38, NCBI) using scitools-fastq align function which runs Burrows-Wheeler (BWA-MEM) aligning algorithm. Hg38 aligned reads were filtered based on a stringent quality score 30 (Q30) and PCR duplicates (multiple amplifications of the same amplicon that can be detected by same chromosome start and end information from a single cell ID) were removed using scitools bam-rmdup. Unique reads after PCR duplicate removal were used to plot library complexity (unique reads/total reads in each unique cell defined by Cell ID, scitools plot-complexity) by using scitools bam-filter with -N flag to filter cells with more than 1000 reads. With this strategy, we generated 27,696 high-quality single cell libraries with a median mapped read depth of >11,000.

Sci-ATAC-seq peak calling

Peaks were called using scitools atac-callpeak function, which uses MACS2 to identify peaks. Peaks that extend beyond chromosome boundaries were removed and the real peaks were extended to 500 bp. With this criterion, in total, 202003 peaks were called. Peaks were intersected with a known data set of peaks associated with cell types. Peaks of cell types in hematopoiesis was used from Corces et. al [64].

Counts matrix generation

Scitools platform atac-counts function was used to generate binary matrix & values file with the fraction of reads on target for each cell. The matrix of low or noisy information is

filtered. Scitools filter-matrix is used to filter the binary matrix to exclude cells with less than 1000 reads with -C flag and reads that are present in less than 10 cells with -R flag.

Dimensionality reduction

UMAP dimensionality reduction is used by performing scitools matrix-umap and/or t-SNE by scitools matrix-tsne providing an annotation file to plot with sample ID or other variables that has been generated.

Differential motif usage to identify transcription factor–associated changes

We have used chromVAR, R package [79] to compute bias corrected deviation in accessibility from the expected accessibility based on a mean accessibility for each cell, cell group or clustering. These scores are called transcription motif deviation Z-scores. By using global motif accessibility across all the populations present, we may infer what transcription factors are active in which cell populations, therefore the active biological pathways at a given time. The human_pwmms_v2 motif set from the chromVAR motifs R package was used in this analysis. The bias-corrected motif deviation scores were used as coordinates of t-SNE by the scitools plot-dims -M flag for visualization [83].

cisTopic dimensionality reduction

We have used cistopic, cis-regulatory topic modelling on single cell ATAC-seq data [75]. This probabilistic tool calculates the probability of a region belonging to a topic and a topic belonging to a cell by simultaneously clustering cells and regions to define common cis-regulatory themes (topics) [75]. We have used these two probability matrices for downstream projections of cell states via uniform manifold approximation and projection (UMAP) and explored several sites that can signal distinct biological cues.

Abstract

Zinc finger 687 (ZNF687) encodes a transcription factor containing DNA binding C2H2 zinc fingers and has been studied in several forms of malignancies including Paget's disease of bone that is associated with giant cell tumor (PDB/GCT) and hepatocellular carcinoma. It is shown to be a part of a transcriptional network along with ZMYND8, and two additional zinc-finger proteins, ZNF592 and ZNF532, that associate with H3K4 demethylation machinery and regulate gene transcription. However, the functional significance of ZNF687 in AML has not been explored yet. We have found that ZNF687 transcription is significantly upregulated in the AML patient samples compared to control bone marrow and whole mono-nuclear bone marrow cells from healthy donors. In addition, we show for the first time a recurrent frameshift variant (p.R939Pfs*Ter36) in ZNF687 in AML patients.

We performed RNA sequencing on HEK293 cells that were transfected to express the frameshift variant, wild type, or a previously identified variant, c.2810C>G, p.Pro937Arg (P937R), that has been reported as a germline ZNF687 variant characterized to predispose to Paget's disease and giant cell tumor. Our results demonstrated the activation of oxidative phosphorylation and inactivation of histone methylation machinery by greatly downregulating the expression of lysine methyltransferase 2D (KMT2D). Overall, these suggest that ZNF687 plays an important role in AML pathogenesis and could be a promising new target for therapeutic applications in myeloid malignancies.

CHAPTER 3

The Identification of ZNF687 variants in Acute Myeloid Leukemia

Fatma Eryildiz¹, Alisa Damnernsawad^{1,2}, Mustafa Cilkiz³, Andrew C. Adey^{1,4}, Jeffrey W. Tyner¹

This chapter has been reformatted for this dissertation from the manuscript titled: “Identification and characterization of novel ZNF687 variants in acute myeloid leukemia” currently in submission.

3.1. Introduction

Acute myeloid leukemia (AML) is the most common form of acute leukemia in adults that is presented by a repertoire of epigenetic and genetic alterations. The accumulation of these events leads to uncontrolled proliferation of immature hematopoietic cells in the bone marrow and peripheral blood thereby inhibiting the normal hematopoiesis.

To date, several studies including ours[82] have identified several somatic mutations, notably, in genes important for epigenetic regulation of gene expression, transcription factors and chromatin organization. Unsurprisingly, epigenome dysregulation which is represented by the large number of mutations in these genes alters the cells' normal hematopoiesis by disrupting methylation, demethylation of CpG islands and

posttranslational modifications of histones. The crucial need to link epigenetic alterations to clinically relevant disease outcomes is clear.

ZNF687 is a C2H2 type zinc finger transcription factor that belongs to transcription regulatory multiprotein complex Z3 including ZNF592 and ZMNYD8 [84]·[85] that aids the recruitment of NuRD complex at damaged DNA sites thereby facilitating DNA repair by homologous recombination[86]. ZNF687 has been implicated in several studies; such as a founder germline mutation in ZNF687 gene (c.2810C>G, p.Pro937Arg) has been identified to be responsible for a form of bone disease associated with giant cell tumor[87], it regulates osteoclast differentiation[88]·[89], it's transcription is elevated in hepatocellular carcinoma[90] and found as a translocation partner of RUNX1, crucial for myeloid differentiation, in AML[91]. Cumulatively, all the findings suggest an oncogenic role for ZNF687.

Previously, we have reported the results of a whole-exome sequencing in a large cohort of AML patients[82] and here we identify a frameshift mutation in ZNF687 gene in these AML patients (n=8) and show its expression is significantly elevated compared to healthy control bone marrow, CD34+, and whole mono-nuclear bone marrow cells from healthy donors. Additionally, RNA-seq analysis on cell subsets that express variants showed the inactivation of histone methylation pathways that allows epigenetic gene expression regulation and the activation of ATP synthesis and oxidative phosphorylation compared to subsets that do not express variants of ZNF687. Thus, these results might establish

an oncogenic role for ZNF687 for the first time in AML through an epigenetic transcription regulatory mechanism.

3.2. Epigenetic control of transcriptional gene regulation via transcription factors

Despite recent therapeutic advances in the cure of pediatric and adult acute leukemia patients, estimated 5-year relative survival rate for adult patients remains stable ~29% according to the National Cancer Institute [92]. Thus, the identification and characterization of new therapeutic strategies and candidates are critically needed. Recent evidence reveals the abnormal activation of epigenetic regulators that leads to dysregulated oncogenic transcriptional programs [41], [45].

The regulation of gene expression by epigenetic mechanisms is such a primal concept such that it dictates cell-fate decisions from very early on in an organism's development. Target gene activation or repression is facilitated by binding of transcription factors to specific and conserved DNA motifs within enhancers and promoters, with the help of chromatin-modifying enzymes [93].

Several key events have been identified in leukemia that results in the inactivation of tumor suppressor genes or activation of oncogenes through aberrant promoter hypermethylation [94], covalent histone modifications [95], chromatin remodeling [96]. All these epigenetic events eventually become responsible for enhanced cellular proliferation or differentiation block [97].

It is fair to claim that the important workhorses of these critical jobs are transcription factors by playing a central role in regulating gene expression to coordinate numerous biological processes. Transcription factors can be categorized into classical zinc fingers [98],

homeodomains [99] and basic helix-loop-helix [100][101]. By being encoded about 2% of the human genome, classical zinc finger containing sequence-specific DNA-binding proteins constitute the largest family of all transcription factors [102], [103]. Several examples of zinc finger proteins reported so far include Cys2His2 (C2H2) like, Zinc ribbon, Gag knuckle, Treble clef, Zn2/Cys6, TAZ2 domain like, Zinc binding loops and Metallothionein [104] and these proteins are associated with plethora of distinct functions. As if these were not enough, zinc fingers contain several domains that specify their function in different biological processes, including BTB (BroadComplex, Tramtrack, and Bric-a-brac), the Krüppel-Associated Box (KRAB) domain, SET domain and SCAN (SRE-ZBP, CTfin51, AW-1 and Number 18 cDNA) domain[105]. Although, most prominent function of these proteins is DNA binding, recent studies displayed that these zinc finger motifs can interact with other macromolecules like RNA, protein and lipids under different cellular contexts or stimuli[106]–[109].

3.3. Zinc finger proteins in health and disease

Among the numerous biological functions of zinc finger proteins, transcription regulation is the most important one, relying on the binding to the promoter region of the target gene and act as a transcription enhancer by interacting with co-activators including CBP/p300 and C/EBP[110], [111] or transcription inhibitor by recruiting co-repressors[105],[112]–[114].

One study that exemplifies this function of zinc fingers is the stimulation of HDGF (hepatoma-derived growth factor) expression in endometrial carcinoma cells upon binding of ZEB1 (Zinc finger E box binding homeobox 1) to the HDGF[115]. Another interesting example comes from a study that shows a C2H2 zinc finger transcription factor can bind and recruit histone deacetylation enzyme 3 to the c-Myc promoter to inhibit its

transcriptional expression thereby maintaining stem cell-like characteristics in lung cancer[116].

C2H2-type zinc finger motif is the largest group of all zinc finger motif classes. C2H2-type zinc finger motif is composed of CX₂CX₃FX₅LX₂HX₃H which allows the protein to fold into a finger like structure upon interacting with zinc ions[117],[118]. Studies have shown that the presence of two to three successive C2H2-type zinc finger motifs provides the most efficient DNA binding[98]. The zinc ion does not directly contact the DNA that these proteins bind to. It serves as a cofactor that is essential for the stability of the tightly folded protein chain[119].

The role of zinc fingers in cancer onset and progression is like a double-edged sword that includes both tumor suppressor and oncogene function. Transcription factor function of zinc fingers is shown to be crucial to recruit chromatin modifiers or structural proteins that regulate cancer cell migration and invasion[120].

3.4. Identification of ZNF687 germline mutations

The first mutations in ZNF687 gene has been identified in a group of Paget's disease of bone patients from Italy[87]. Paget's disease of bone (PDB) is the second most common metabolic bone disorder after osteoporosis which is characterized by increased osteoclastic bone resorption followed by a compensatory increase in osteoblastic bone formation, resulting with enlarged and disorganized formation of the bone structure[121].

In less than 1% of the affected cases, the disease undergoes neoplastic transformation resulting in giant cell tumor of bone (GCT) associated with PDB (GCT/PDB) specifically at pagetic skeletal sites (Figure 11). Recently, a germline heterozygous missense mutation (c.2810C>G, p.Pro937Arg) in ZNF687 gene has been identified to be responsible in all

the cases reported in this form of secondary degeneration of PDB [88]. Later, researchers identified several other novel mutations in ZNF687 gene, commonly all affecting the sites of zinc finger motifs of the protein. These patients displayed milder disease profiles compared to the carriers of p.Pro937Arg mutation [88].

ZNF687 immunofluorescence performed on in vitro differentiated osteoclasts and GCT/PDB tumor biopsies confirmed that the greater accumulation of ZNF687 protein in the nuclei than in the cytoplasm [88]. Following analysis from this cohort of patients provided the genetic evidence that p.Pro937Arg mutation is necessary and sufficient for GCT/PDB phenotype development [87].

Collectively, the initial evidence showed that ZNF687 plays a role in bone metabolism and regulates osteoclast and osteoblast differentiation. Also, ZNF687 expression is found to be upregulated in zebrafish development and tissue regeneration suggesting a role in cell proliferation and differentiation [122].

Interestingly, ZNF687 along with the other partners in Z3 regulatory module (ZMYND8, ZNF592) have been identified as genomic targets of NFkB by multiple ChIP-seq experiments [122] and this can be hypothetically illustrated in (Figure 12).

Bioinformatic predictions suggest that ZNF687 contains a strong nuclear localization signal (NLS) between amino acid residues 938 and 954. The mutations that occur at and/or around these residues were predicted to be affecting the functional consequences of ZNF687 expression. In the case of c.2810C>G mutation, ZNF687 nuclear import is enhanced and nuclear availability of this transcription factor is increased [122].

3.5. Evidence for ZNF687 involvement in DNA damage response

A recent study identified ZNF687 as one of the putative DNA-binding zinc finger proteins interacting with nucleosome remodeling and histone deacetylase (NuRD) along with ZMYND8 in HeLa cells [85]. Recent work proves that the elimination of components of NuRD complex contributes to the reduced cell survival (increased apoptosis, reduced cell proliferation and colony forming ability) in cell lines and primary AML samples [123].

Also, this study agrees with recent data showing that ZMYND8 and ZNF687, together with ZNF532, and ZNF592, is part of the recently described transcriptional regulator complex Z3 [84] (Figure 11 and Figure 13).

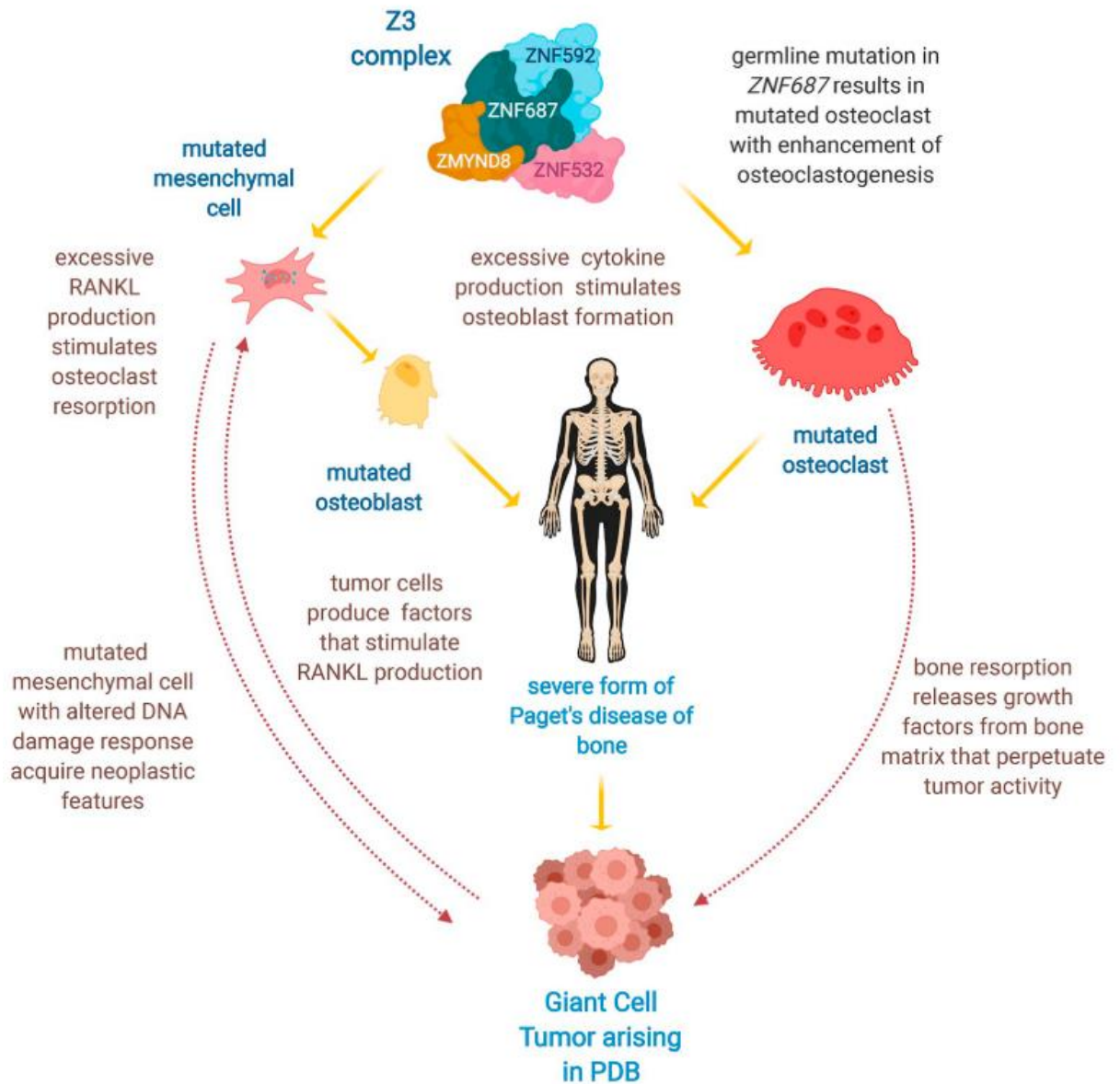


Figure 11 - ZNF687-dependent GCT/PDB formation. ZNF687 mutation increases osteoclast differentiation, in turn producing molecules that stimulate osteoblast formation and activity. Mutation-bearing mesenchymal precursors could harbor unrepaired DNA damage and undergo uncontrolled proliferation. ⁴

⁴ Adapted from [89]. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

In addition, this study [85] shows that ZNF687 occupies ZMYND8 genome-wide binding sites defined by chromatin immunoprecipitation sequencing (ChIP-seq) strengthening the connection between Z3 and NuRD complexes. Also, DNA damage generated by local irradiation rapidly recruits ZNF687 at these sites by bolstering the recruitment of an NuRD component CDH4. Right after siRNA-mediated knockdown of ZNF687 impairs the recruitment of CDH4 to these double stranded breaks (DSBs) suggesting that ZNF687 acts downstream of ZMYND8 to accumulate the NuRD complex at damaged sites [85], [86]. Also, these results points to a sequential NuRD recruitment model that depends on ZMYND8/ZNF687 axis[85]. ZNF687 is not only important for recruitment of NuRD components at the DSBs but also is required for the repair these lesions by homologous recombination (HR) [85].

Collectively, ZNF687 displays a notable role in the regulation of the epigenome, especially transcriptional silencing of genomic regions suffering DNA damage [89].

The function of ZNF687 in AML pathogenesis remains uninvestigated. However, ZNF687 mutated AML could harbor unrepaired DNA that would lead to increased mutations which could support the role of ZNF687 in uncontrolled cellular proliferation of damaged cells.

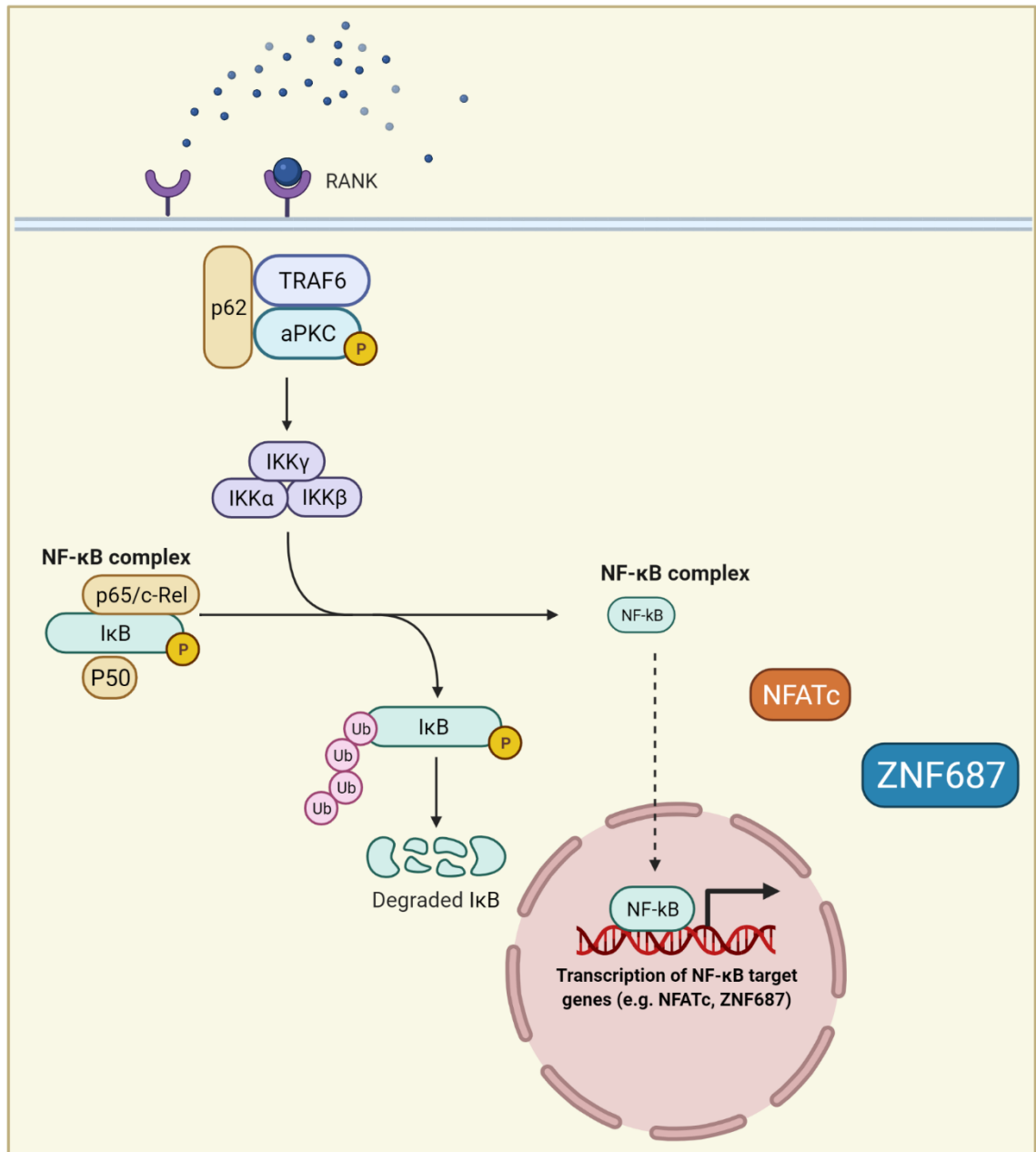


Figure 12 - Hypothetical illustration of ZNF687 involvement in bone cell proliferation as the downstream target of NFκB.

Created with [BioRender.com](https://www.biorender.com)

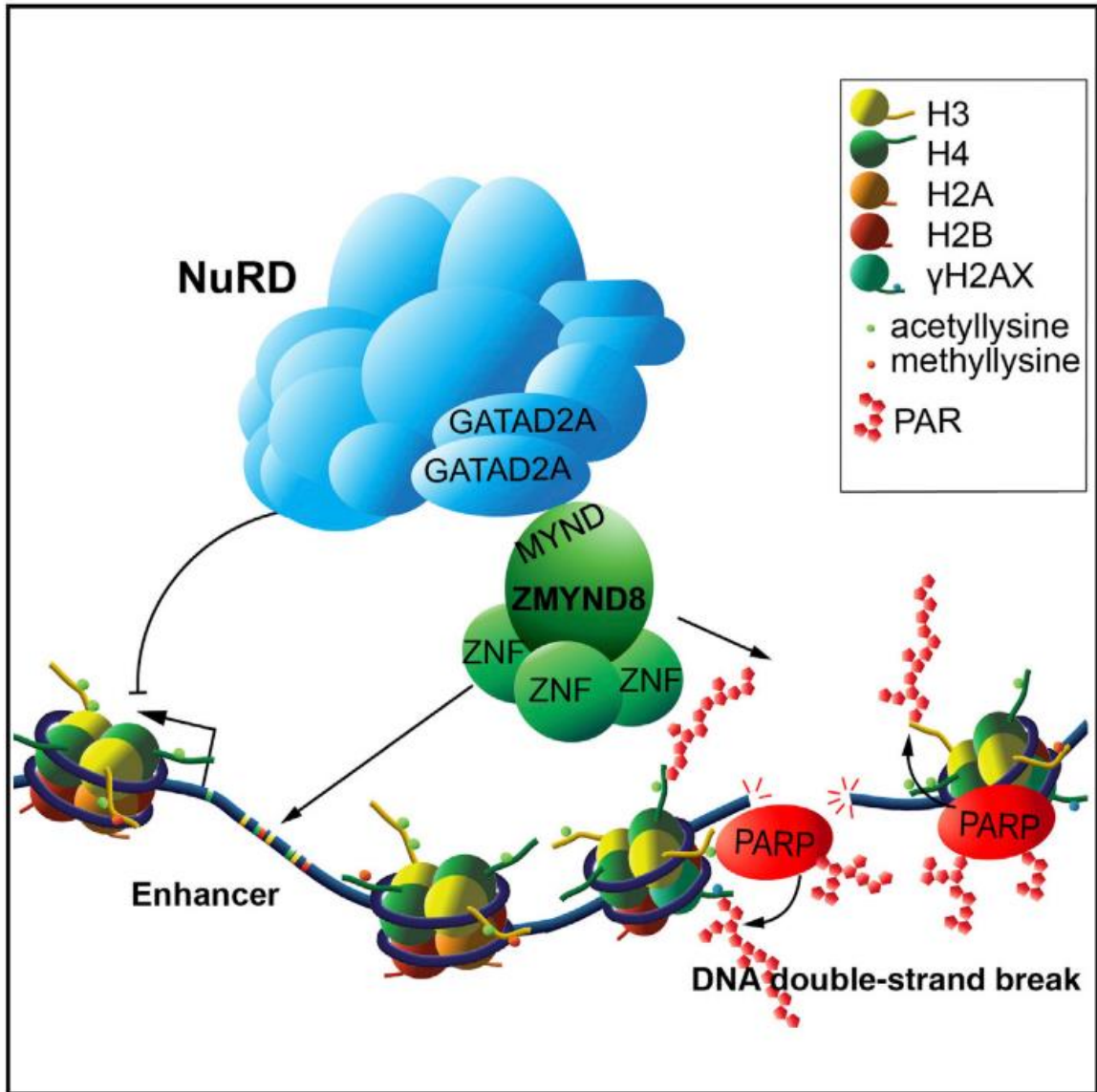


Figure 13 - Depiction of the recruitment of the NuRD complex to its genome-wide binding sites.

5

⁵ Adapted from [85]. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

3.6. Summary

ZNF687 is a C2H2 type zinc finger containing transcription factor that is known to bind DNA. A recurrent germline mutation (c.2810C>G, p.Pro937Arg) in this gene has been identified in Paget's disease of bone patients that presents with a more severe form of the disease resulting in formation of giant cell tumors specifically at the pagetic skeletal sites. This supports that *ZNF687* plays a role in bone metabolism and proper osteoclast differentiation.

The arise of several mutations at the zinc finger motifs and the increased nuclear accumulation of *ZNF687* mutant protein implies the importance of nuclear function of the gene. Beyond, its upregulation during zebrafish development supports a role for *ZNF687* in cell proliferation and differentiation.

Current literature suggests that *ZNF687* is important to facilitate the recruitment of NuRD complex at its target sites and required for the repair of DSBs through homologous recombination. Regarding the role of NuRD components in acute leukemia cell survival, *ZNF687* may be an integral interactor of this complex in AML and its involvement may lead to dysregulation of the AML epigenome. Overall, these may suggest a mechanism where *ZNF687* is functional when it's nuclear localized and facilitates the repair of lesions affected by DNA damage by recruiting several other proteins.

The core of my work outlined here is the characterization of *ZNF687* mutations and their role in AML pathogenesis. Throughout the work that will be presented here, I have used several existing bioinformatic algorithms and tools to predict the functional effects of the variants described in patients.

3.7. Results

3.7.1. A cohort of primary patients from BeatAML suggests a pathogenic somatic frameshift mutation in *ZNF687*

We have identified for the first time a *ZNF687* recurrent frameshift variant p.R939Pfs*Ter36 (C2809insC) in a group of AML patients (n=8, ~1%) from the BeatAML cohort where we have previously reported [82] (Figure 14) by whole exome sequencing (Figure 16) and confirmed by Sanger sequencing (Figure 15).

Sample ID	Diagnosis	FAB/Blast Morphology	ELN2017	Karyotype	Secondary Mutations
13-00615	AML with t(8;21)(q22;q22); RUNX1-RUNX1T1	N/A	Favorable	46,XX,del(5)(q13q33)[14]/46,XX[6]	NRAS
13-00625	AML with myelodysplasia-related changes	M1	Adverse	45,XY,-7[6]/47,XY,+13[cp5]/46,XY[12]	NRAS
14-00573	AML with myelodysplasia-related changes	M2	Adverse	47,XY,der(7)t(1;7)(q10;p10),+8[20]	CBL, EZH2, GATA2, RUNX1
15-00123	AML with mutated NPM1	N/A	Intermediate	46,XY[20]	FLT3-ITD, NPM1
16-00566	AML with mutated NPM1	N/A	Intermediate	N/A	FLT3-ITD, NPM1
16-01093	Acute myeloid leukaemia, NOS	N/A	Intermediate	47,XY,t(8;19)(q24.1;q13.1),+10[17]/46,XY[3]	N/A
18-00414	AML without maturation	M1	Adverse	46,XX,del(5)(q13q33)[14]/46,XX[6]	DNMT3A, ETV6, JAK2, KMT2A, STAT3
19-00431	AML with mutated NPM1	M2	Favorable	46,XY[20]	BLM, FLT3-D835, FLT3, NBN, NPM1, PTPN11, STAG2

Figure 14. Top Carriers of frameshift mutation in BeatAML cohort

Created with BioRender.com

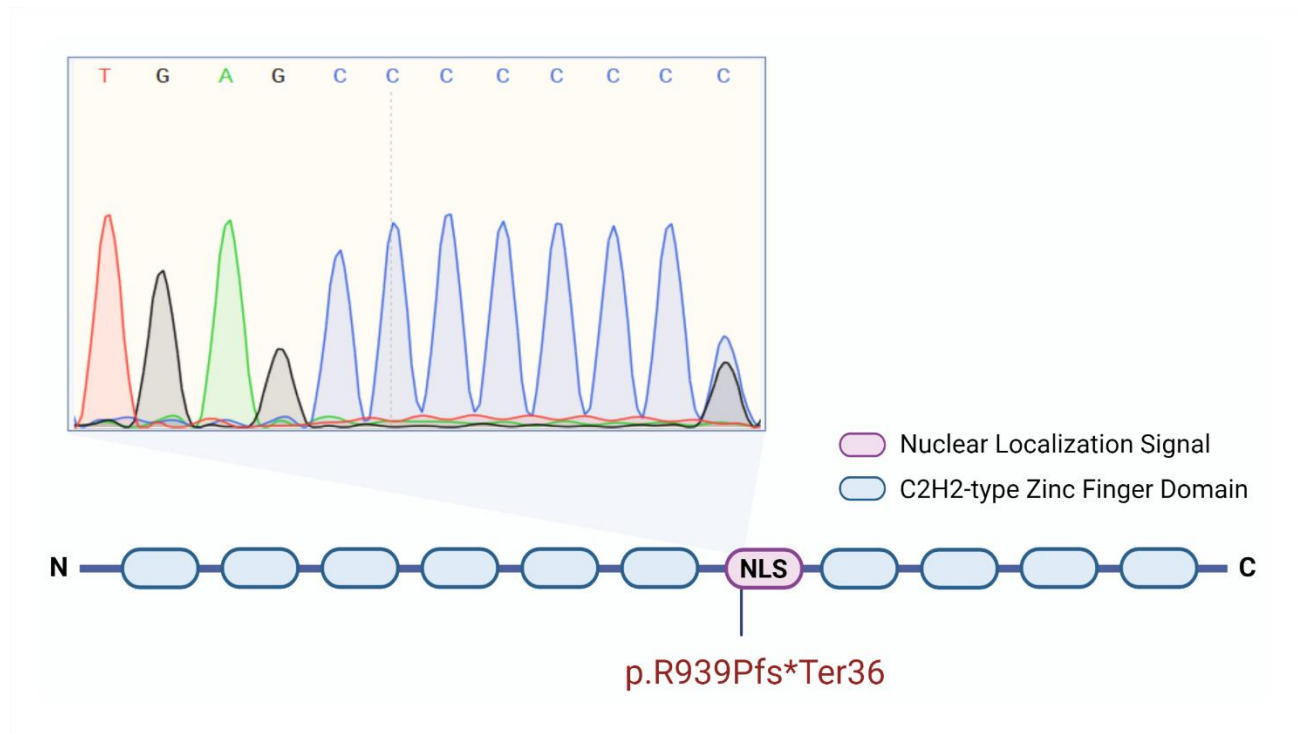


Figure 15. ZNF687 recurrent frameshift mutation R939Pfs*Ter36 (bottom) was first identified in a cohort of AML patients by whole exome sequencing and confirmed by Sanger sequencing (C2809insC, top).

Created with BioRender.com

ZNF687

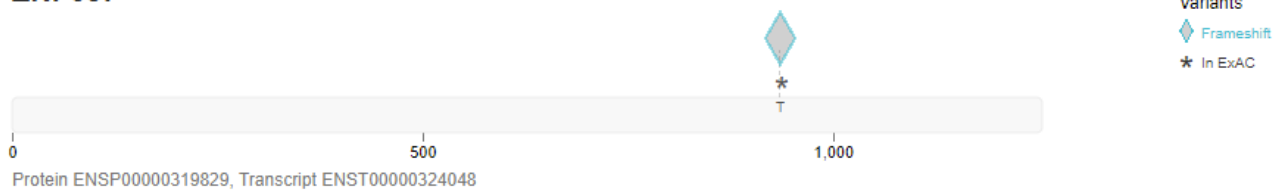


Figure 16 Frameshift mutation found in %1 of BeatAML patients. (Vizome, 2022) Asterisk indicates pathogenic P937R mutation that is recorded in ExAC/gnomAD database.

3.7.2. ZNF687 expression is significantly upregulated in primary AML patients

Previously, we have reported clinical, genomic, transcriptomic, and functional analyses of a large cohort of AML patients, a cohort of 672 tumor specimens collected from 562 patients [82]. ZNF687 expression was found to be upregulated in this large cohort of AML patient tumor samples, suggesting that ZNF687 might be involved in AML progression (Figure 18). To validate whether ZNF687 expression correlates with clinical prognosis, we looked at the correlation between ZNF687 mRNA expression and ELN 2017 genetic risk categories and found that it does not show any correlation (Figure 19). Further, we wanted to assess whether ZNF687 frameshift mutations have any relevance with its mRNA expression level, however we did not observe a significant correlation between ZNF687 frameshift mutant vs ZNF687 WT patients in ZNF687 mRNA expression, which might also be due to the low number of samples (Figure 17).

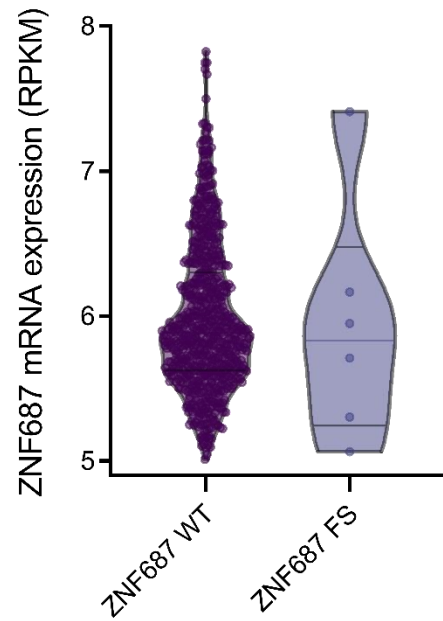


Figure 17. ZNF687 mRNA expression does not correlate with mutational status.

ZNF687 mRNA expression profiles in ZNF687 WT (n=702) and ZNF687 FS (n=6) AML patients (not significant) from BeatAML dataset.

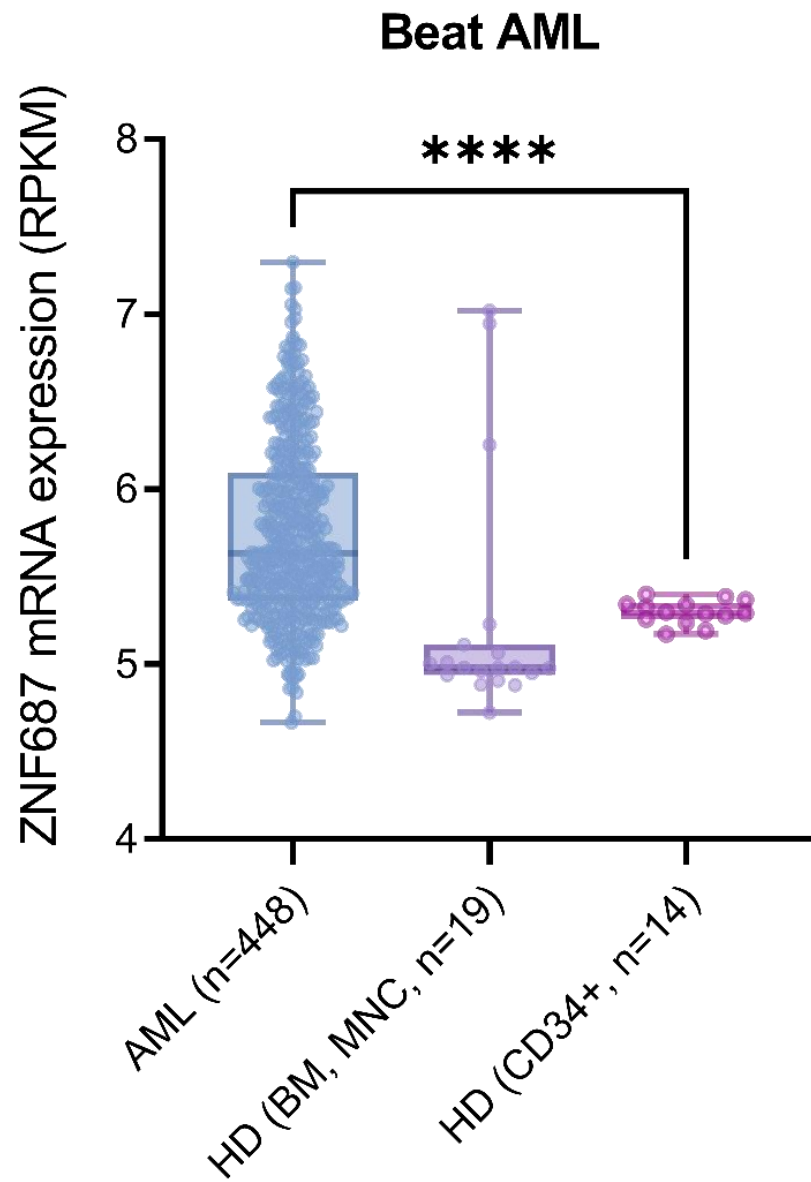


Figure 18. ZNF687 mRNA is upregulated specifically in AML patients.

ZNF687 mRNA expression profiles in AML (n=448) and bone marrow mononuclear cells of healthy donors (n=19), CD34+ healthy donors (n=14) from BeatAML dataset (****, $p < 0.001$).

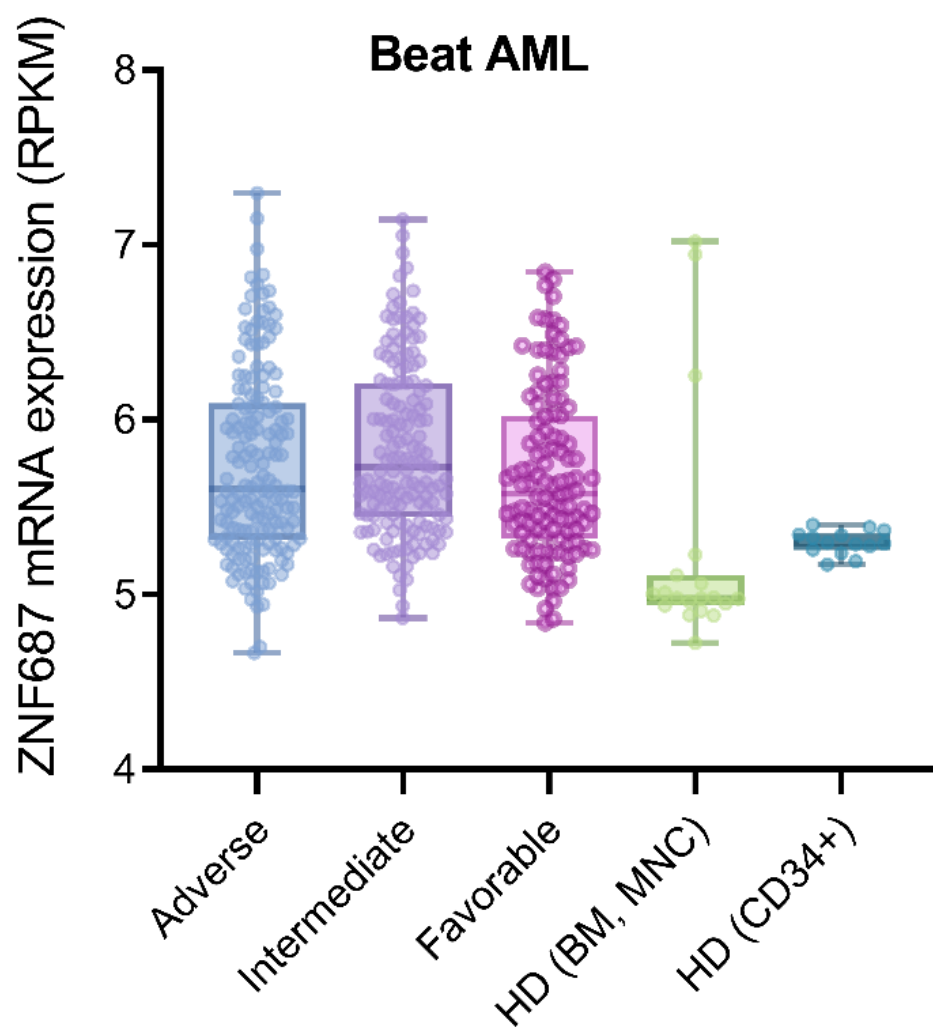


Figure 19. ZNF687 mRNA expression does not correlate with clinical severity of the disease.

ZNF687 mRNA expression profiles in Adverse (n=340), Intermediate (n=358), Favorable (n=243) risk groups stratified by ELN 2017 and bone marrow mononuclear cells of healthy donors (n=19), CD34+ healthy donors (n=14) from BeatAML dataset.

3.7.3. ZNF687 frameshift causes loss of a functional C2H2 superfamily domain and related functional features

The insertion of an extra cysteine at the 939 residue causes the synthesis of 35 novel amino acids followed by an in-frame premature stop codon (PTS) at 973 (1237 amino acid in ZNF687 native protein) by generating a C terminal truncated protein (Figure 20-21). Also, we predicted that C terminal of the native protein contains two nuclear localization signals (NLS) that leads to the expression of a primarily nuclear localized zinc finger protein. However, with the insertion of PTS, the localization signals are completely disrupted (Figure 22).

To further assess the effects of the loss of C terminal on functional domains of ZNF687, we have comparatively annotated the functional units of ZNF687 variants and WT and discovered that zinc finger C2H2_superfamily domain 11 is deleted in frameshift variant (Figure 23). This domain has been annotated as a contact point for DNA binding by recognizing the DNA sequence.

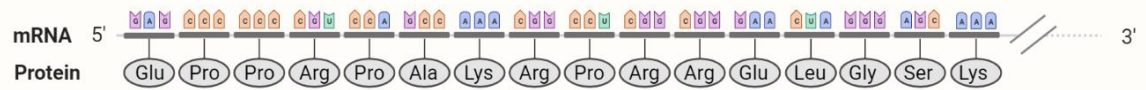
Moreover, we have predicted that the C-termini of WT ZNF687 has multiple posttranslational modifications including site-specific phosphorylation and ubiquitination (Figure 24). In addition to these serine, threonine, and tyrosine phosphorylation sites, ZNF687 has been predicted to have several ubiquitination at specific sites (Figure 25). These all modifications may impact its localization and stability.

Recent advances in protein structure prediction with deep learning methods promise to provide improved templates for molecular modeling of the proteins not yet to be elucidated [124], [125]. We have implemented Google DeepMind initiative AI assisted protein folding prediction algorithm AlphaFold to predict the effects of missense mutation, p.Pro937Arg,

and frameshift indel p.R939Pfs*Ter36 on protein folding by computing last 188 amino acids where these recurrent mutations occur. As expected, while missense mutation leads to the slight conformational change in the structure, frameshift mutation shows a more dramatic implication on protein by disrupting formation of higher-order structures like alpha helices and β sheets (Figure 27) when aligned with the native folding predicted by AlphaFold (Figure 26).

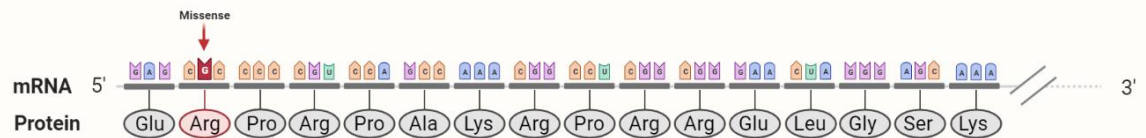
ZNF687 - Wild type

mRNA sequence without any mutation



Missense (p.P937R)

mRNA sequence with P937R mutation



Base-pair insertion (p.R939Pfs*Ter36)

Frameshift causing premature stop codon

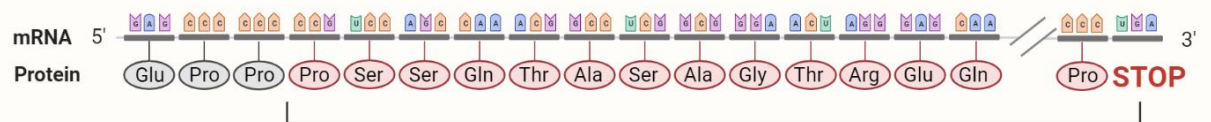


Figure 20. Illustration of ZNF687 WT, FS and 937 mutations at mRNA and protein sequence.

We have verified that a 1-base pair insertion causes a frameshift in the codon encoding arginine at amino acid position 939 (replaced by Proline), which predicts the synthesis of 35 novel amino acids (amino acids that are not encoded by non-mutant ZNF687), followed by an in-frame premature stop codon.

FS	PKTEPEELAVSQGGAAPATEESSSSSEEEVPSSPEPPSSSQTAS--AGTREQRPQGWWG	958
WT	PKTEPEELAVSQGGAAPATEESSSSSEEEVPSSPEPPRPAPKRPRRELGSKGLKGGGGGP	960
937	PKTEPEELAVSQGGAAPATEESSSSSEEEVPSSPERPRAPKRPRRELGSKGLKGGGGGP	960
	***** * :: *:: : * *	
FS	GAWRLDLWPVSL-----VP-----	973
WT	GGWTCGLCHSWFPERDEVVAHMKKEHGKSVKKFPCRLCERSFCSAPSLRRHVRVNHEGIK	1020
937	GGWTCGLCHSWFPERDEVVAHMKKEHGKSVKKFPCRLCERSFCSAPSLRRHVRVNHEGIK	1020
	..*:*	
FS	-----	973
WT	RVYPCRYCTEGKRTFSSRLILEKHVQVRHGLQLGAQSPGRGTTLARGSSARAQGPGRKRR	1080
937	RVYPCRYCTEGKRTFSSRLILEKHVQVRHGLQLGAQSPGRGTTLARGSSARAQGPGRKRR	1080
FS	-----	973
WT	QSSDSCSEEPDSTTPPAKSPRGGPGSGGHGPLRYRSSSSTEQSLMMGLRVEDGAQQCLDC	1140
937	QSSDSCSEEPDSTTPPAKSPRGGPGSGGHGPLRYRSSSSTEQSLMMGLRVEDGAQQCLDC	1140
FS	-----	973
WT	GLCFASPGSLSRHRFISHKKRRGVGKASALGLGDGEEEAAPSRSDPDGGDSPLPASGGPL	1200
937	GLCFASPGSLSRHRFISHKKRRGVGKASALGLGDGEEEAAPSRSDPDGGDSPLPASGGPL	1200
FS	-----	973
WT	TCKVCGKSCDSPLNLTHTFRTHGMAFIRARQGAVGDN	1237
937	TCKVCGKSCDSPLNLTHTFRTHGMAFIRARQGAVGDN	1237

Figure 21. Multiple sequence alignment of missense and frameshift mutations along with WT ZNF687.

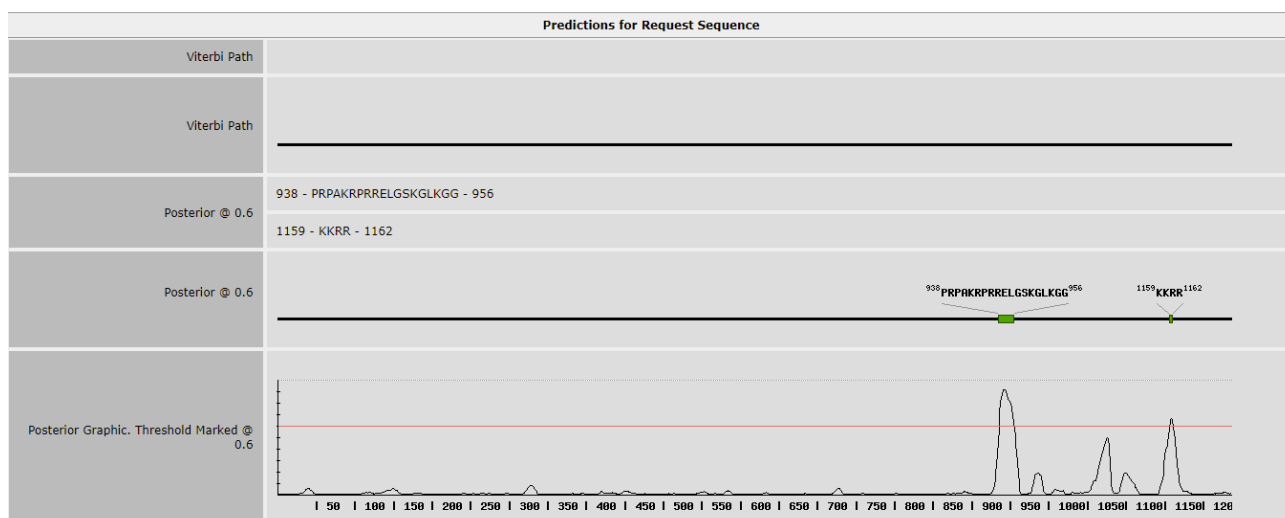


Figure 22. Nostradamus predicts nuclear localization signals.



Figure 23. In-silico analysis indicates that FS indel causes the loss of zf-C2H2_11 superfamily domain in native proteins.

Sequence data has been extracted from conserved domain database (CDD) and plotted using TBtools.

Created using <https://bio.tools/tbtools>

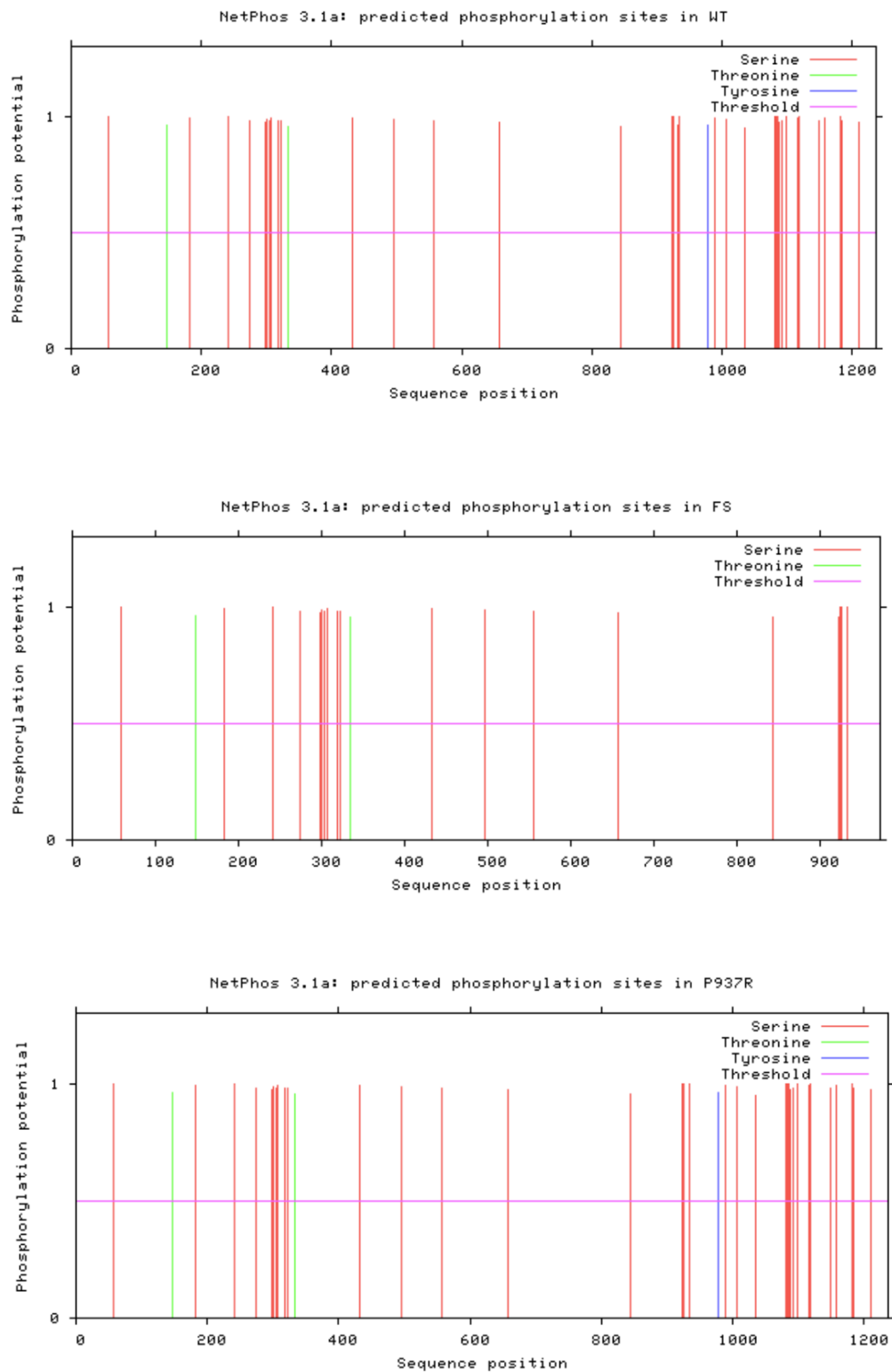


Figure 24. NetPhos3.1 is used to predict serine, threonine, and tyrosine phosphorylation sites (confidence score > 0.95).

Peptide	Position	Score	Threshold
PERPRPAKRPRRELG	34	3.58	0.3
PRRELGSKGLKGGGG	43	2.66	0.3
ELGSKGLKGGGGPG	46	3.12	0.3
DEYVAHMKKEHGKSV	75	0.62	0.3
EYVAHMKKEHGKSVK	76	0.45	0.3
HMKKEHGKSVKKFPC	80	2.31	0.3
KEHGKSVKRFPCRLC	83	1.54	0.3
EHGKSVKRFPCRLCE	84	0.90	0.3
RVNHEGIKRVYPCRY	112	1.57	0.3
CRYCTEGKRTFSSRL	124	1.17	0.3
SSRLILEKHVQVRHG	135	1.01	0.3
RAQGPGRKRRQSSDS	170	2.81	0.3

Figure 25. BDM-PUB: Prediction of Ubiquitination sites with Bayesian Discriminant Method is used to predict high confidence ubiquitination sites in native ZNF687 protein [126], [127].

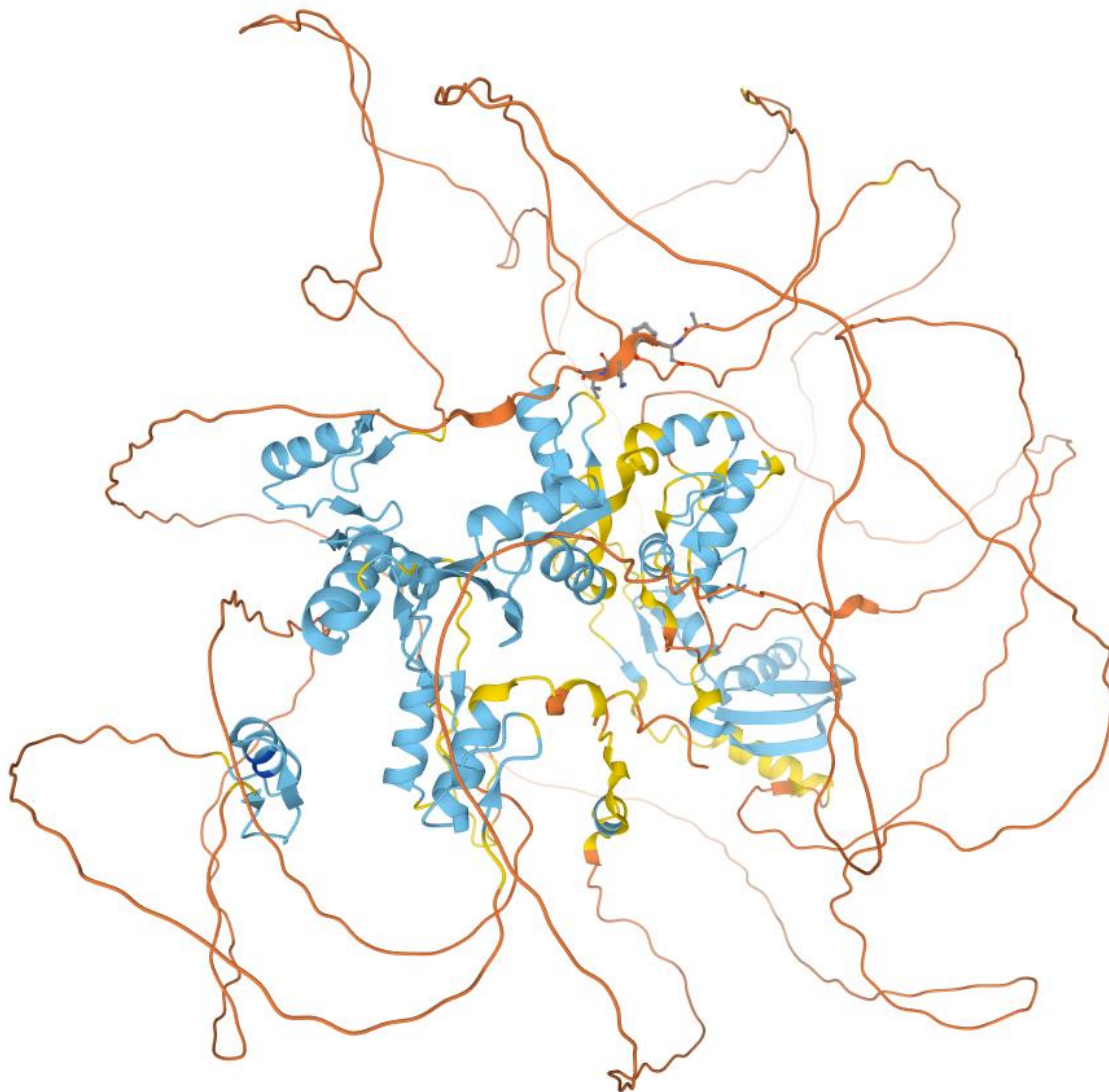


Figure 26. 3D folding of native ZNF687 protein was predicted in silico by the Alphafold2 software package with default parameters [124], [125].

3.7.4. Expression of frameshift variant leads to a change in subcellular localization of ZNF687 and frameshift protein sequesters in cytoplasm

To characterize frameshift variant which we detected in AML patients, we have expressed ZNF687 protein harboring p.R939Pfs*Ter36 and a point mutation c.2810C>G, p.Pro937Arg along with wild-type ZNF687 and empty vector in human endothelial kidney 293T17 cells, a highly transfectable derivative of the human embryonic kidney (HEK293) cell line. Immunoblots of 293T17 cells expressing parental, WT, FS (p.R939Pfs*Ter36) and P937R (p.Pro937Arg) showed, as we expected, a shift to a lower molecular weight specifically in frameshift expressing cells which proves the expression of truncated variant (Figure 28). This also shows that the truncated protein is not subject to non-sense mRNA mediated decay.

Next, to define the subcellular localization of ZNF687 variants, we have performed a detergent-based cell fractionation assay followed by western blotting. Our immunoblotting data showed the decreased abundance of ZNF687 in the nuclei of cells expressing of FS, however, this altered subcellular localization is specific to frameshift mutation and the localization of missense mutation (P937R) has not been dramatically affected (Figure 29).

To further confirm cytoplasmic localization of frameshift protein in HEK293 cells, we have performed immunocytochemical staining on transfected cells. When we express WT ZNF687 in HEK293T cells, it locates to the nucleus due to the localization signals it harbors, however clearly FS loses this capability and sequesters in cytoplasm when it is ectopically expressed (Figure 30).

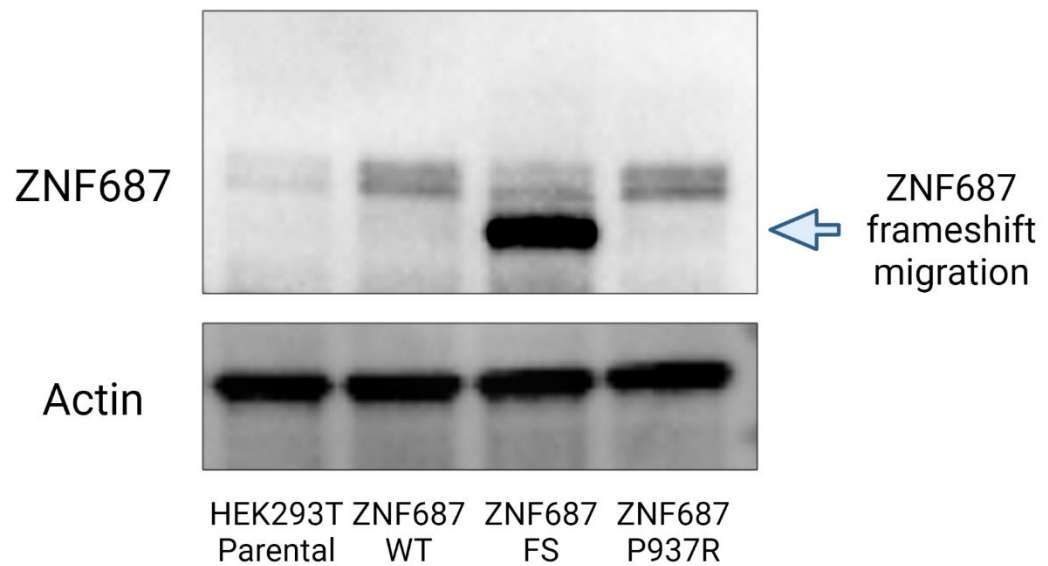


Figure 28. ZNF687 frameshift generates a truncated protein product which migrates due to decreased molecular weight in HEK293T whole cell lysates.

HEK293T cells were transfected with expression constructs containing ZNF687 WT, ZNF687 frameshift and ZNF687 P937R and lysates were subjected to immunoblot. Representative blot of 3 biological replicates.

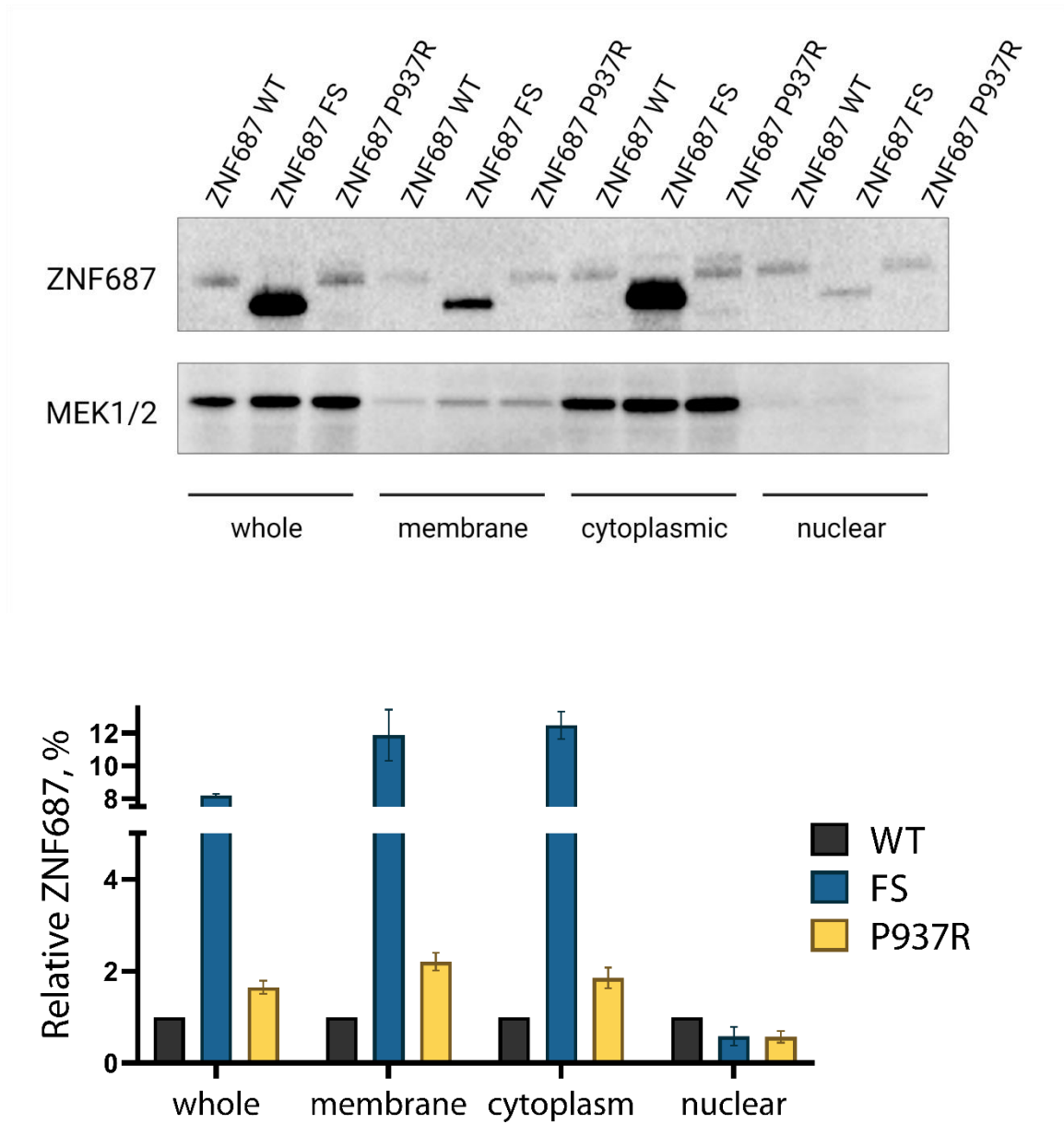


Figure 29. The expression of frameshift alters nuclear and cytoplasmic ratio of ZNF687 by sequestering almost exclusively in cytoplasm compared to WT.

P937R localization is not significantly affected. Representative blot of 3 biological replicates (top). Densitometry analysis of the top blot by ImageJ (bottom).

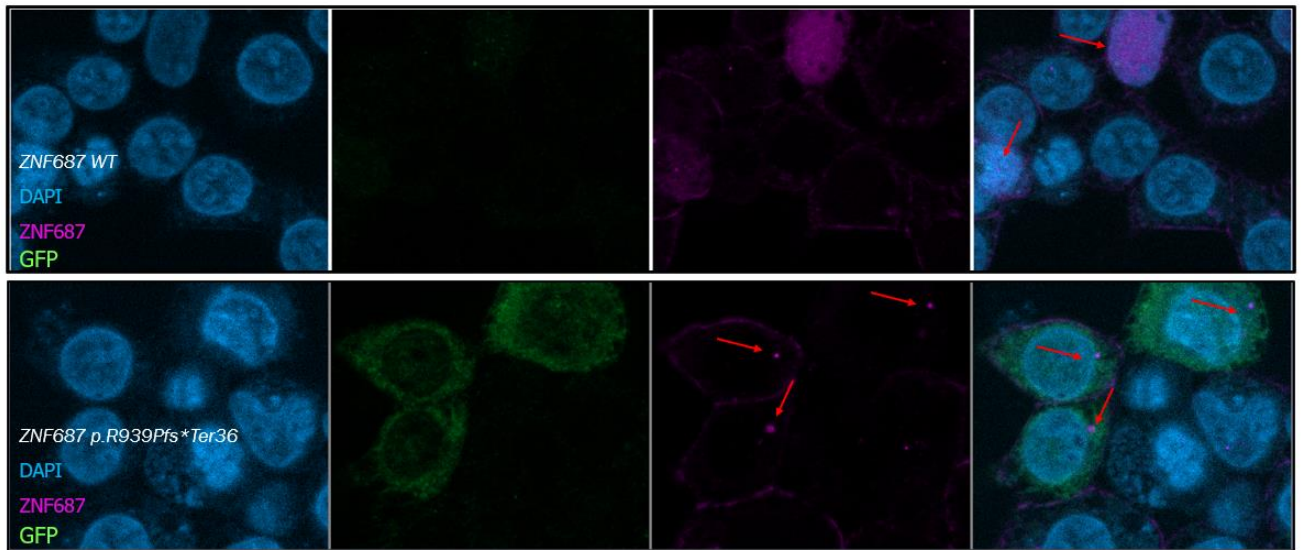


Figure 30. Immunocytochemical staining of transfected HEK293T cells.

3.7.5. Differential expression identifies transcriptional changes in frameshift mutated cell lines

Next, we performed RNA-seq analysis of HEK293T cells expressing WT, empty vector, frameshift (FS) and P937R to investigate the differentially expressed transcriptional programs. PCA clustering of replicates showed consistent transcriptional programs across the replicates in each group (Figure 31). We examined the up- and down regulated gene sets in each group by analyzing heatmap generated by the normalized expression values from RNA-seq experiment (Figure 32).

To assess the prognoses of key biological processes in ZNF687 variant expressing cell subsets, we performed pathway analyses using different databases. GO (Gene Ontology) database classifies genes according to the three categories of biological process (BP), cellular component (CC), and molecular function (MF) and annotates the function of the selected genes. In addition to GO database, we have considered Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathways, and the pathways containing at least 15 genes were evaluated (Figures 33-36).

Histone methylation, Histone H3-K4 methylation, chromatin organization programs related gene sets were found to be downregulated specifically in frameshift expressing HEK293T cell subsets compared to WT counterparts. Interestingly, regulation of hematopoiesis and regulation of myeloid cell differentiation related gene sets were also found to be downregulated in FS cluster. On the other hand, oxidative phosphorylation, ATP metabolic processes related genes were upregulated in FS compared to all the other clusters. According to some recent studies in leukemia signaling, the aberrant activation of glucose

metabolism, respiratory pathways and ATP synthesis through oxidative phosphorylation is required to support the growth of leukemia cells. Thereby, higher oxidative phosphorylation signals less differentiated and more primitive cell state to feed need for proliferation [128]–[130]. According to Reactome based enriched gene sets, “transcriptional regulation by RUNX1” related genes were downregulated in FS expressing cells. An experimental demonstration of this putative link has been previously investigated [91]. Nguyen et al. reported that a novel RUNX1-ZNF687 fusion transcription that generates a chimeric protein comprised of a truncated RUNX1 runt-homology domain (RHD), which is responsible for DNA-binding, fused to the entire ZNF687 protein [91].

We then analyzed the gene expression data to select the most differentially expressed gene sets between different variant categories, and identify specific disease linked interactions. This analysis showed the significant (FDR cut-off=0.01, min fold change=1.5) downregulation of KMT2D gene in frameshift cells vs all other groups (Figure 37).

KMT2D also known as MLL2 encodes a lysine methyltransferase 2 family protein that mono methylate lysine 4 of histone H3 tail at important regulatory regions in the genome by acting on critical chromatin structures and DNA accessibility [131]. It has been also shown to be mutated in several cancers and its deficiency has been most frequently studied in the development of follicular lymphoma and diffuse large B cell lymphoma [132]. My work here suggests a probable link between ZNF687 and KMT2D to act on important regulatory sites in genome by regulating the transcription of DNA repair response and tumor suppression genes. So that, the cytoplasmic mislocalization of ZNF687 led by frameshift mutation may impair this putative link as an alternative mechanism to KMT2D loss of function and/or deficiencies described in earlier reports.

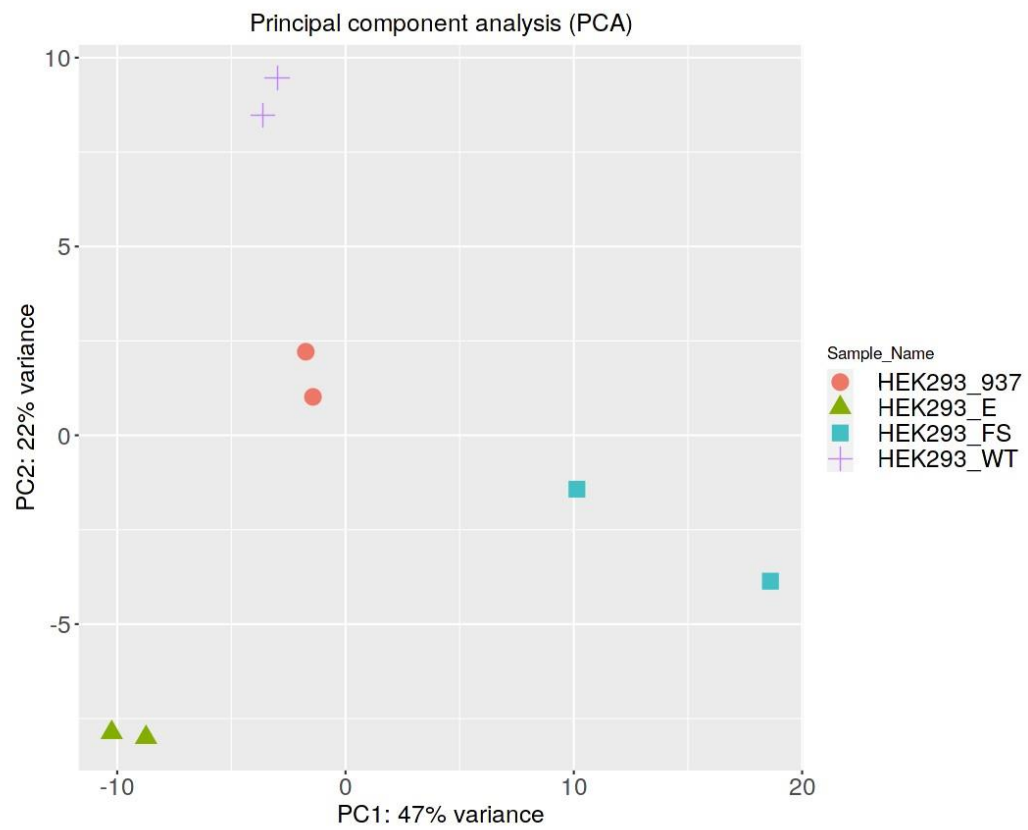


Figure 31. Scatter plot of Principal component analysis.

Each cluster was represented by two biological replicates.

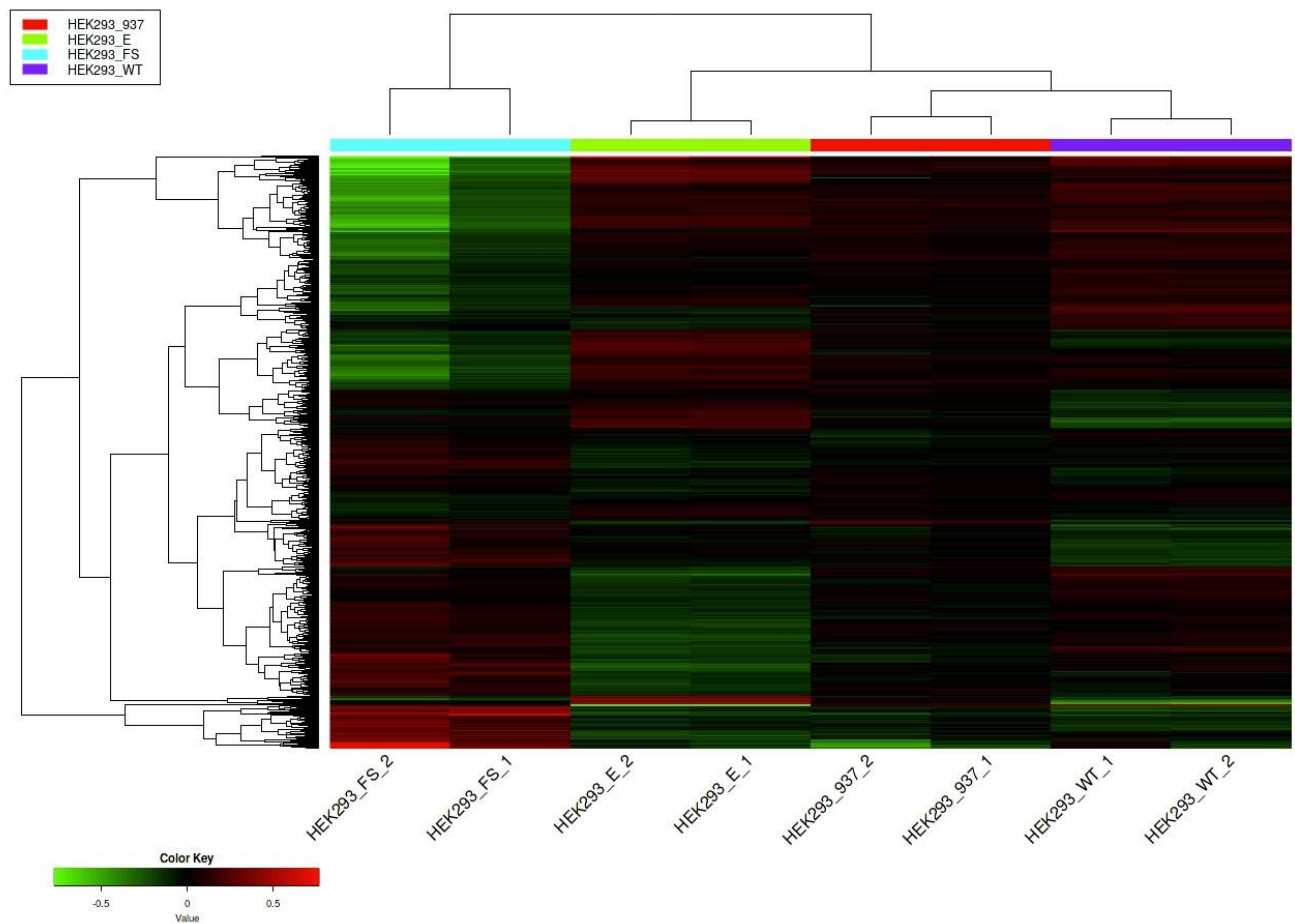
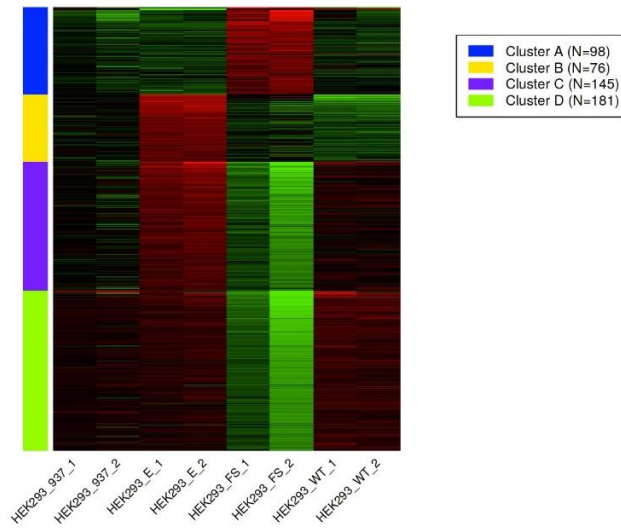


Figure 32. Heat map of the expression patterns of ZNF687 WT, Empty, FS and 937 expressing HEK293T cells.

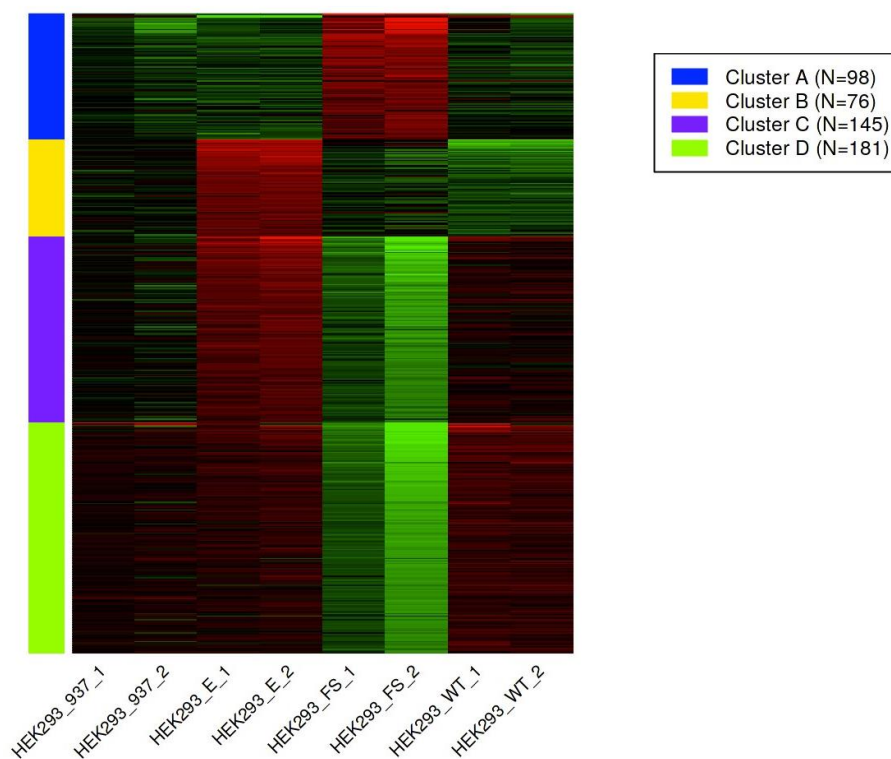
The gene expression profile of all groups was characterized by hierarchical clustering and the changes in expression levels are displayed from green (downregulated) to red (upregulated), as shown in the color key. Heatmaps were generated and hierarchical clustering was performed. Rows correspond to the top 500 most differentially expressed genes and columns correspond to the two replicates from each sample group.



Enriched pathways for each cluster

Cluster	adj.Pval	nGenes	Pathways
A	4.0e-15	16	Oxidative phosphorylation
	4.0e-15	14	ATP synthesis coupled electron transport
	4.0e-15	14	Mitochondrial ATP synthesis coupled electron transport
	2.8e-14	14	Respiratory electron transport chain
	1.1e-13	18	ATP metabolic process
	3.8e-13	15	Electron transport chain
	2.6e-11	14	Cellular respiration
	6.0e-09	8	Mitochondrial electron transport, NADH to ubiquinone
	1.2e-08	17	Generation of precursor metabolites and energy
	3.1e-07	4	Energy coupled proton transmembrane transport, against electrochemical gradient
	8.5e-07	7	NADH dehydrogenase complex assembly
	9.4e-07	8	Mitochondrial respiratory chain complex assembly
B	2.1e-05	6	L-alpha-amino acid transmembrane transport
	7.3e-05	5	Amino acid import across plasma membrane
	2.2e-04	5	L-amino acid transport
	4.5e-04	6	Amino acid transport
	8.4e-03	6	Carboxylic acid transport
	8.4e-03	6	Import into cell
D	3.8e-05	11	Histone methylation
	3.8e-05	17	Peptidyl-lysine modification
	3.8e-05	10	Histone lysine methylation
	6.1e-04	16	Histone modification
	8.4e-04	39	Regulation of developmental process
	1.3e-03	6	Histone H3-K4 methylation
	1.5e-03	20	Chromatin organization
	1.5e-03	41	Anatomical structure morphogenesis
	1.5e-03	13	Regulation of hemopoiesis
	2.2e-03	25	Regulation of multicellular organismal development
	2.4e-03	10	Regulation of myeloid cell differentiation

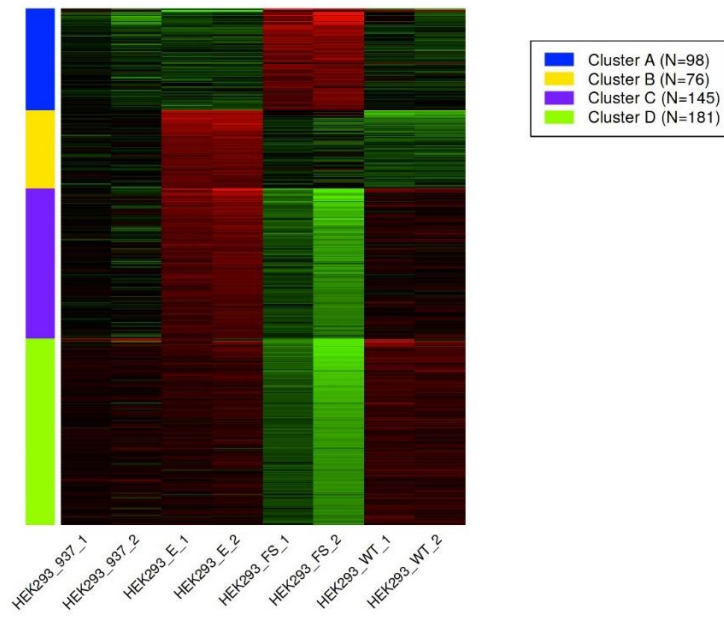
Figure 33. GO Biological Process. ENSEMBL 92 was used to analyze fold-change values of all genes in database.



Enriched pathways for each cluster

Cluster	adj.Pval	nGenes	Pathways
A	8.6e-16	15	Oxidative phosphorylation
	4.1e-15	17	Thermogenesis
	2.8e-13	17	Huntington disease
	4.0e-13	16	Prion disease
	4.1e-13	16	Parkinson disease
	5.6e-08	9	Ribosome
	2.7e-07	7	Cardiac muscle contraction
	3.4e-07	8	Retrograde endocannabinoid signaling
	1.2e-06	8	Non-alcoholic fatty liver disease
	1.2e-06	9	Coronavirus disease
D	8.3e-04	6	Lysine degradation

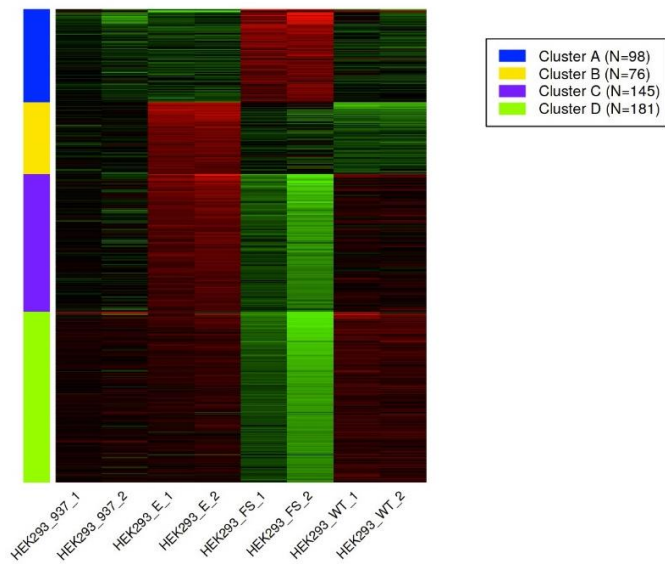
Figure 34. KEGG. KEGG Release 86.1 was used to analyze fold-change values of all genes in database.



Enriched pathways for each cluster

Cluster	adj.Pval	nGenes	Pathways
A	9.3e-17	16	Respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins.
	3.1e-15	14	Respiratory electron transport
	1.6e-07	7	Complex I biogenesis
	2.2e-06	27	Metabolism
	7.7e-06	8	Influenza Viral RNA Transcription and Replication
	8.9e-06	7	Eukaryotic Translation Elongation
	1.7e-05	10	Translation
	2.3e-05	7	SRP-dependent cotranslational protein targeting to membrane
	6.2e-05	6	Peptide chain elongation
	9.1e-03	3	Amino acid transport across the plasma membrane
D	1.2e-04	12	Chromatin modifying enzymes
	5.1e-04	6	PKMTs methylate histone lysines
	5.1e-04	5	Regulation of MECP2 expression and activity
	8.9e-04	3	Post-transcriptional silencing by small RNAs
	4.4e-03	5	Pre-NOTCH Transcription and Translation
	4.6e-03	8	Transcriptional regulation by RUNX1
	4.6e-03	5	RUNX1 regulates genes involved in megakaryocyte differentiation and platelet function
	4.6e-03	2	TET123 and TDG demethylate DNA
	5.4e-03	8	Signaling by NOTCH

Figure 35. Reactome. V64 was used to analyze fold-change values of all genes in database.



Enriched pathways for each cluster

Cluster	adj.Pval	nGenes	Pathways
A	3.0e-13	14	Electron transfer activity
	9.0e-10	8	NADH dehydrogenase activity
	9.0e-10	8	NADH dehydrogenase (ubiquinone) activity
	1.6e-07	8	Oxidoreductase activity, acting on NAD(P)H
	9.8e-07	16	Oxidoreductase activity
	1.0e-06	9	Structural constituent of ribosome
	3.9e-05	4	Cytochrome-c oxidase activity
	1.4e-04	6	Proton transmembrane transporter activity
	8.6e-04	20	RNA binding
B	2.4e-03	4	L-amino acid transmembrane transporter activity
	3.8e-03	4	Amino acid transmembrane transporter activity
	7.4e-03	3	Neutral amino acid transmembrane transporter activity
D	5.7e-04	6	Histone-lysine N-methyltransferase activity
	6.4e-04	4	Histone methyltransferase activity (H3-K4 specific)
	2.6e-03	34	Molecular function regulator
	3.5e-03	14	Transcription coregulator activity
	7.3e-03	2	Methylcytosine dioxygenase activity

Figure 36. GO Molecular Function. ENSEMBL 92 was used to analyze fold-change values of all genes in database.

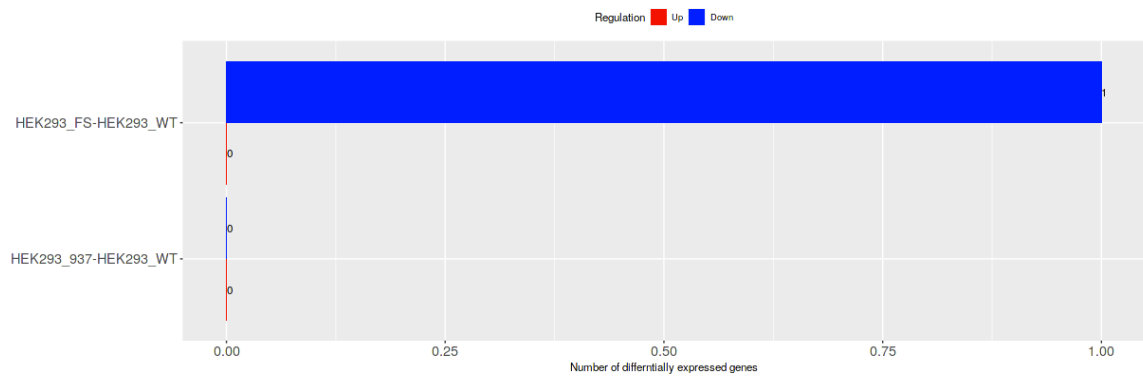


Figure 37. Differential expression analysis showed the significant downregulation of *KMT2D*, a major enhancer mono-methyltransferase.

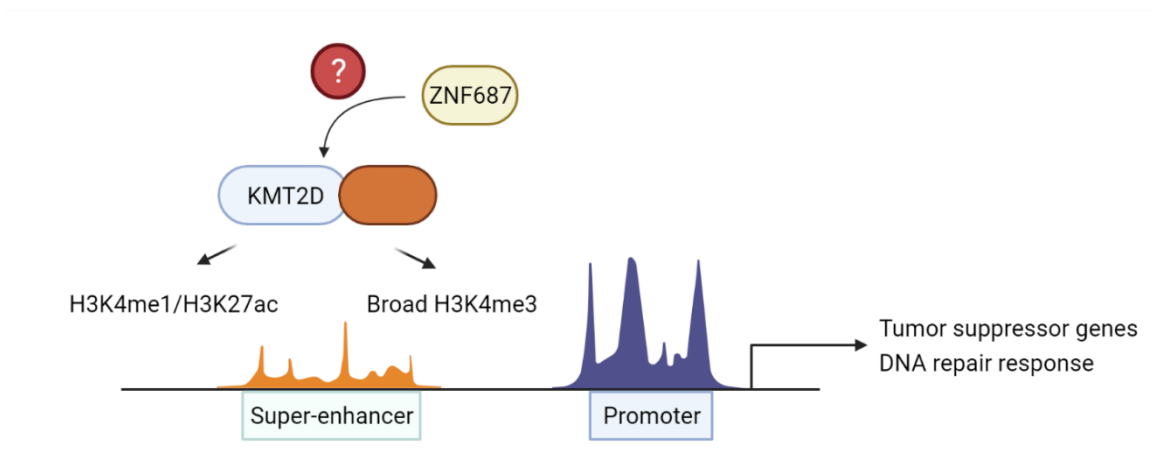


Figure 38. Hypothetical mechanistic model of a putative link between ZNF687 that upregulates *KMT2D* which acts as a chromatin remodeling machinery to monomethylate lysine-4 of histone H3 at active chromatin sites where transcription and DNA repair takes place.

3.8. Materials and Methods

Patient information

All patient specimens were obtained after informed consent with IRB approval by Oregon Health & Science University, Stanford University, University of Utah, UT-Southwestern and University of Colorado-Denver.

Cell Cultures

HEK293 cells were obtained from American Type Culture Collection (Manassas, VA) and cultured in DMEM medium (Invitrogen) supplemented with 10% FBS (Atlanta Biologicals), L-glutamine, penicillin/streptomycin (Invitrogen), and fungizone at 37°C in 5% CO₂.

Cloning and plasmid construction

Human *ZNF687* (NM_020832) in pDONR223 gateway donor was purchased from Addgene, the ORFeome Collaboration (DNASU, HsCD00399950). First, a stop codon, later *ZNF687* variants including frameshift mutation c.2809insC, p.R939Pfs*Ter36 and missense mutation c.2810C>G, p.Pro937Arg were generated using the QuikChange II XL site-directed mutagenesis kit (Agilent Technologies) on pDONR223 vector. These variants and wild type *ZNF687* were cloned into the gateway compatible MSCV-IRES-green fluorescent protein (GFP) plasmid retroviral vector (Addgene, # 9044) via Gateway Cloning System (Gateway LR Clonase/Invitrogen). Plasmids were confirmed via Sanger Sequencing (Eurofins).

HEK293 Transfection

Six-well plates were seeded with 3e5 cells overnight. HEK293 cells were transfected with *ZNF687* WT, frameshift, missense variant containing plasmids and pMIG empty using

Fugene6 (Promega) according to the manufacturer's protocol. Cells were cultured and harvested for protein lysate extraction following the observation of GFP fluorescence within 48–72 h after transfection.

NGS variant validation by Sanger sequencing

Genomic DNA from specimen 18-00414 with the highest variant allele frequency % is used for further validation of the frameshift variant detected by whole-exome sequencing[82] via Sanger sequencing. Specific primer pairs for the selected frameshift variant to be validated were designed using the Primer3 algorithm. Primer sequences were checked for the presence of single-nucleotide polymorphism (SNPs) to avoid common SNPs within the primer design (<https://genetools.org/SNPCheck/snpcheck.htm>). In-silico PCR of the UCSC Genome Browser (<https://genome.ucsc.edu/cgi-bin/hgPcr>) is used to confirm primer specificity to the region to be amplified. M13 universal sequencing tails are added to the specific primer pairs appropriately.

PCR was performed in a final volume of 25 µl using Invitrogen™ AccuPrime™ Taq DNA Polymerase, high fidelity (Invitrogen). The PCR amplicons were then purified (QIAquick PCR & Gel Cleanup Kit, Qiagen) and purified PCR products were Sanger sequenced (Eurofins) using only M13 universal primer pairs.

Cellular fractionation

2e5 HEK293 cells were pelleted at 500×g for 5 min. Nuclear and cytoplasmic fractions were isolated using the NE-PER kit according to the manufacturer's instructions (ThermoScientific).

Protein extraction and immunoblot

Transfected HEK293 whole cell protein lysates were extracted in a cell extraction buffer (Cell Lysis Buffer/Thermofisher). 50ug of total whole protein lysates were mixed with NuPAGE LDS Sample Buffer (4X) and heated at 95°C for 5 minutes. Lysates were run on 15% Criterion™ Tris-HCl Protein Gel (Biorad) and transferred to PVDF membrane and blocked for 1 hour in TBST (Tris-buffered saline with Tween-20) with 5% BSA. Blots were probed overnight at 4°C with the following antibodies: rabbit anti-ZNF687 polyclonal (#PA5-34389), rabbit mAb anti-MEK 1/2 (#9122), mouse mAb β -Actin (#3700). Membranes were blotted with appropriate anti-rabbit or anti-mouse IgG HRP conjugated antibodies following primary antibody incubation. Blot images were obtained by chemiluminescence according to the manufacturer's instructions (BioRad ECL Clarity Substrate, ChemiDocMP) and densitometry was performed using ImageJ (NIH, Bethesda MD). Bands were normalized to the loading control (Actin) before further analysis. For fractionation analysis, whole and fractionation lysates densities were calculated using ImageJ software and normalized to Actin protein levels for each membrane to ensure equal loading parameters. Then, each value is normalized to WT of each fraction to get the relative levels of the protein across different fractions.

Immunocytochemistry sample preparation and Imaging

1e5 cells are seeded and transfected in glass slides. Cells were washed with warm HBSS (Hanks Buffered Saline Solution) to prevent anomalous morphological changes due to temperature shock. Cells were fixed using pre-chilled ice-cold %100 methanol for 30 seconds, followed by rinsing with PBS and incubated in 3% BSA (Fraction V, lipid-free) in PBS solution at room temperature for blocking to minimize non-specific binding of primary and secondary antibodies in samples. In addition, all antibodies were diluted

in the blocking solution outlined above. Primary antibodies were incubated on cells for 1-4 h at RT, followed by washing with PBS 5 to 6 times. Secondary antibody incubation and wash was performed and followed by rinsing and counterstaining with DAPI (1µg/mL) to stain for nuclei for 3-5 mins. Finally, the cells were rinsed with water to prevent formation of salt crystals. All coverslips were mounted with SlowFade™ Gold Antifade Mountant agent (S36938, Invitrogen) and sealed with clear nail polish. The slides were stored in dark at 4C. For immunocytochemistry, the following primary antibodies: ZNF687 (rabbit pAb, #PA5-34389), GFP (mouse mAb, CST #GF28R), and following secondary antibodies were used: Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody Alexa Fluor 647, and Alexa Fluor 488.

RNA-Seq library preparation and Data analysis

Total RNA was extracted and purified using the QIAGEN RNeasy Mini Kit. Total RNA integrity was evaluated using the Bioanalyzer (Agilent). RNA-seq libraries were prepared from total RNA using the Illumina Stranded mRNA kit (Illumina). Libraries were then profiled on a TapeStation D1000 tape (Agilent) and quantified using an NGS Library Quantification Kit (Roche/Kapa) on a StepOnePlus Real Time PCR Workstation (Thermo/ABI). Libraries were sequenced on a NovaSeq 6000 (Illumina). Fastq files were generated from the resulting base call files using bcl2fastq (Illumina).

The paired-end reads were aligned to hg38 (NCBI build 39) using the Rsubread Package. Aligned reads were analyzed using BamTools API 1.0.2. Differential expression analysis was performed using the deSeq2 package by FPKM (fragments of kilo- base per exon per million fragments mapped) on the human RefSeq transcript database.

The top 500 genes are used in hierarchical clustering using the heatmap.2 function and heatmaps were generated. The data is centered by subtracting the average expression

level for each gene. The distance matrix is $1 - r$, where r is Pearson's correlation coefficient. The average linkage is used [133].

The R glimma, limma edgeR packages used to generate fold-change values to perform pathway analyses. 0.01 FDR cut-off and 1.5-min fold change is specified.

Statistical methods

All graphical data and statistical analyses on immunoblot were generated using GraphPad Prism software. The data were expressed as the mean $\pm$ SEM. Statistical significance was determined using two-tailed nonparametric Student's t-tests (Mann–Whitney test) or one-way analysis of variance (ANOVA) and expressed as p values (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$).

Table 3.1. ZNF687 Mutagenesis and Whole ORF sequencing Primers

Primer Name	Primer Sequence
ZNF687_end_SQ	5'-CATCAAGGTCTGACCCC-3'
ZNF687_Reverse	5'-CATGGCGTTCATCAGGGCTC-3'
M13F_SQ	5'-GTAAAACGACGGCCAGT-3'
M13R_SQ	5'-CAGGAAACAGCTATGACC-3'
ZNF687_SQ_F1	5'-CTAGCCCAACAAGGCTCAGG-3'
ZNF687_SQ_F2	5'-CTCGGTGGTGATGGTGCAAC-3'
ZNF687_SQ_F3	5'-ACCAGCAATGTGTGCCCAAC-3'
ZNF687_SQ_F4	5'-AGGACCATGCTGGAACATCTC-3'
ZNF687_ORF FWD Set 1	5'-GAGGGCATCAAGCGAGTTTA-3'
ZNF687_ORF REV Set 1	5'-GGAAGTGGGTCTTGAGGTTTAG-3'
ZNF687_ORF FWD Set 2	5'-CATCACCTCCTCTGCCATTAC-3'
ZNF687_ORF REV Set 2	5'-CTGGGTACTTCCTCCTCTTCT-3'
ZNF687_ORF FWD Set 3	5'-CCAGGTGTAGGAAGTGAATCTG-3'
ZNF687_ORF REV Set 3	5'-GACCTGAGTTACAGTCCTTGTG-3'
ZNF687_ORF FWD Set 4	5'-CAGCCCTAAGATGATTGCTAAGA-3'
ZNF687_ORF REV Set 4	5'-GTAATGGCAGAGGAGGTGATG-3'
ZNF687_ORF FWD Set 5	5'-CCAGCGTGTCAGTGTCTTTA-3'
ZNF687_ORF REV Set 5	5'-GCTTGATGCCCTCGTGATTA-3'
ZNF687_FS936_M_F	5'-CTGAGCCCCCCCCGTCCAGCCAAAC-3'
ZNF687_FS936_M_R	5'-GTTTGGCTGGACGGGGGGGGCTCAG-3'
ZNF687_P937R_M_F	5'-GCTCCCCTGAGCGCCCCCGTCCAGC-3'
ZNF687_P937R_M_R	5'-GCTGGACGGGGGCGCTCAGGGGAGC-3'
ZNF687_pat_set1 FWD Set 3	5'-CCTTGTCTCCTCTCACAGAAC-3'
ZNF687_pat_set1 REV Set 3	5'-TCAGGGAACCAGGAGTGA-3'
ZNF687_pat_M13_F1	5'-TGTA AACGACGGCCAGTCTGAGGAGTCGTCTTCATCTTC-3'
ZNF687_pat_M13_R1	5'-CAGGAAACAGCTATGACCCTGGCAGTCCCAATGCT-3'
ZNF687_pat_M13_F2	5'-TGTA AACGACGGCCAGTCCTTGTCTCCTCTCACAGAAC-3'
ZNF687_pat_M13_R2	5'-CAGGAAACAGCTATGACCTCAGGGAACCAGGAGTGA-3'

Bioinformatic predictions and multiple sequence alignment

AlphaFold 3D protein folding predictions of ZNF687 wild type and variants have been generated to include last 188 amino acids of ZNF687 by implementing publicly available Alphafold source code (<https://github.com/deepmind/alphafold/>).

The NetPhos 3.1 server was used to predict serine, threonine or tyrosine phosphorylation sites using ensembles of neural networks [134]. Both generic and kinase specific predictions are performed. Predictions are made for the following 17 kinases: ATM, CKI, CKII, CaMII, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SR, cdc2, cdk5 and p38MAPK. SIFT Web server was used to predict the functional effects of frameshift indel [135].

Clustal Omega MSA (multiple sequence alignment) ClustalW program was used to generate seeded guide trees and pairwise sequence alignment of ZNF687 variants.

CHAPTER 4

Summary and Discussion

Epigenomic dysregulation is one key feature that has been recognized in AML. This is manifested by several mutations in epigenetic modifiers, chromatin regulators and transcription factors have been characterized in leukemia that result in for enhanced cellular proliferation and differentiation block. Identifying these mutational patterns in AML progression is crucial to improve tailored drug treatments.

We have identified with high confidence a recurrent frameshift mutation (p.R939Pfs*Ter36) in ZNF687 transcriptional regulator gene for the first time in %1 of all AML cases in BeatAML [82], and my work here characterizes that ZNF687 protein is significantly overexpressed in AML patients compared to healthy controls that suggests an oncogenic role. Next, I show that ZNF687 upregulation inversely correlates with overall patient survival which imply that the expression of ZNF687 gene may represent a prognostic factor in AML patients' disease outcomes.

We have amplified target ZNF687 region and verified that a 1-base pair insertion in ZNF687 that we have detected in multiple patient samples via Sanger sequencing. This insertion causes a frameshift in the codon encoding arginine at amino acid position 939, which predicts the synthesis of 35 novel amino acids (amino acids that are not encoded by non-mutant ZNF687), followed by an in-frame premature stop codon.

The occurrence of a premature stop codon generates a C- terminal truncated polypeptide about 973 amino acid long, which is 264 amino acids shorter than the WT ZNF687. We predicted that C terminal of the native protein contains two nuclear localization signals (NLS) that leads to a primarily nuclear localized zinc finger protein. This finding is in line with previous reports that show ZNF687 encodes a large number of DNA-binding zinc fingers and associates with DNA through these functional units [85]. Interestingly, the same amino acid position has been shown to be mutated in Paget's disease of bone associated with giant cell tumor previously which supports the importance of functionality of this domain in disease progression.

ZNF687 has been associated with several functions, regulating hematopoietic cell proliferation, differentiation and most prominently mediating DNA repair together with other zinc proteins ZNF592 and ZNF532 as a part of a complex called Z3 module in Hela cells [84] [136].

We have predicted that the aberrantly untranslated portion of the C-termini of WT ZNF687 has several posttranslational modifications, phosphorylation of Y978, S989, S1007, S1036, S1057, S1068, S1069, S1082, S1083, S1085, S1087, S1092, S1099, S1116, S1118, S1119, S1149, S1157, S1182, S1184, S1211 (confidence score > 0.95, NetPhos3.1) [134]. In addition to these phosphorylation sites, ZNF687 has been predicted to have several ubiquitination at specific sites. These all modifications may impact its localization and stability.

This locus of the protein is highly intrinsically disordered, and the sequence is evolutionarily conserved. Intrinsically disordered proteins (IDPs) and domains (IDDs) have been shown to be involved in regulation of cell signaling pathways and crucial cellular processes, including the regulation of transcription, translation and the cell cycle and mutations in IDPs or changes in their cellular abundance are associated with disease

[137]–[141]. The positional importance of where the frameshift mutation occurs can suggest that this portion is structurally important in disease.

Further, we have implemented AlphaFold [124], [125] AI-based protein structure prediction tool to predict the effect of frameshift mutation on 3D ZNF687 folding structure. Indeed, while the insertion of frameshift led to the complete loss of higher-order structure elements such as alpha helices and β sheets, the missense mutation has a lower impact on the overall protein folding, however how this substitution affects protein stability was not explored.

In my study described here, the most distinguished feature of ZNF687 we have identified is that its aberrant cytoplasmic localization. The expression of frameshift leads to altered nuclear and cytoplasmic ratio of ZNF687, which is most prominently localized in the nucleus as in the native form. This altered subcellular localization is specific to frameshift mutation and the localization of missense mutation (P937R) has not been dramatically affected as measured by immunoblotting. The one model for ZNF687 dysregulation in AML could be

ZNF687 is also identified as an oncogene in hepatocellular carcinoma (HCC) and transcriptionally upregulated the expression of the pluripotency associated factors such as BMI1, OCT4 and NANOG by directly targeting promoters of these genes [90]. ZNF687 overexpression specifically in AML patients may suggest a similar function in disease progression. Its upregulation may promote the expression these genes that are important during normal hematopoiesis, however physiologically silenced in mature blood cells, suggesting a role in self-renewal [142].

The role of ZNF687 in cellular differentiation has not been further investigated, however, we can speculate that it may be related to its DNA binding function by which that it could

directly bind and replace putative factors from the nucleus that are required for proper differentiation and physiological downregulation of cell type specific architectural transcription factors which is claimed in NPM1 mutations in AML [142]–[144].

Zinc finger proteins have been shown to follow a modular recognition mechanism in DNA binding where DNA binding zinc finger domains are generously distributed across the protein to bind DNA gradually and tightly. This modular approach is even appealing to researchers, that they are designing artificial zinc fingers that can bind to a specific desired DNA sequence [145]. This modular recognition concept may support the relatively minimal effects that we have observed in P937R missense mutation, which did not lose its DNA binding activity and its subcellular localization has not practically changed, when compared to frameshift protein.

Additionally, we have performed RNA-seq analysis on cell subsets that express variants showed the inactivation of histone methylation pathways that allows epigenetic regulation of gene expression and the activation of ATP synthesis and oxidative phosphorylation which has been claimed as a therapeutic target in leukemogenesis [146] compared to subsets that do not express variants of ZNF687. Transcriptional profiling revealed significant transcriptional differences between frameshift and WT ZNF687. The results for our Gene Ontology (GO) enrichment analysis showed the differential “regulation of hematopoiesis” and “regulation of myeloid cell differentiation” specifically in frameshift expressing cell population in terms of molecular function and biological process.

Notably, “transcriptional regulation by RUNX1” is inactivated in frameshift according to Reactome pathway analysis. This is particularly interesting as one early study reported a novel RUNX1-ZNF687 fusion transcription that generates a chimeric protein comprised of a truncated RUNX1 runt-homology domain (RHD), which is responsible for DNA-binding,

fused to the entire ZNF687 protein [91]. RUNX1 is an indispensable transcription factor in differentiation of hematopoietic stem cells into mature blood cells and several chromosomal rearrangements of RUNX1 were frequently observed in myelodysplastic syndromes, acute myeloid leukemia, acute lymphoblastic leukemia and chronic myelomonocytic leukemia [147].

There are open questions remaining to be answered. The presence of frameshift mutation expression evinces that ZNF687 frameshift mRNA is not subject to nonsense-mediated mRNA decay, that can imply an evasion mechanism. Furthermore, in some cases, truncated proteins could lead to deleterious gain-of-function or dominant-negative activity of the native proteins [30] [148]. ZNF687 mutations may act similar to what was discovered with other mutations in AML that exemplify the simultaneous expression of WT and mutant protein. Both mutant protein expression can lead to the loss of native protein's function and it can further exacerbates the inhibitory effect on the biological processes where native ZNF687 is required for. The frameshift mutations in DNMT3A gave rise to the loss of the catalytic domain of DNMT3A which results in decreased methyltransferase activity associated with hypomethylation [30]. Likewise, the heterozygous expression of DNMT3A mutations affect DNMT3A WT protein by dominant-negative effects and leads to even more hypomethylation [30]. However, in ASXL1, the expression of frameshift mutant protein stabilizes BAP1 target even more than WT protein does, so that it causes a gain of function effect [41], [45]. Maybe, the most similarly, NPM1 frameshift mutations results with altered subcellular localization in NPM1 mutant cells [143].

Also, it has been shown that zinc-finger proteins may exhibit both oncogenic and tumor suppressive roles in cancer and whether to deplete or restore their expression may be a new challenge in cancer drug designs.

Overall, these results might establish an oncogenic role for ZNF687 for the first time in AML patients through an epigenetic transcription regulatory mechanism and the functional consequences of altered nuclear localization of ZNF687 mutants in AML development needs to be further investigated.

References

- [1] S. F. Cai and R. L. Levine, "Genetic and epigenetic determinants of AML pathogenesis," *Semin. Hematol.*, vol. 56, no. 2, pp. 84–89, Apr. 2019, doi: 10.1053/J.SEMINHEMATOL.2018.08.001.
- [2] J. Ladewig, P. Koch, and O. Brüstle, "Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies," *Nat. Rev. Mol. Cell Biol.*, vol. 14, no. 4, pp. 225–236, Apr. 2013, doi: 10.1038/NRM3543.
- [3] A. P. Feinberg, M. A. Koldobskiy, and A. Göndör, "Epigenetic modulators, modifiers and mediators in cancer aetiology and progression," 2016, doi: 10.1038/nrg.2016.13.
- [4] W. A. Flavahan, E. Gaskell, and B. E. Bernstein, "Epigenetic plasticity and the hallmarks of cancer," *Science (80-.)*, vol. 357, no. 6348, 2017, doi: 10.1126/science.aal2380.
- [5] J. E. Holliday, R. & Pugh, "DNA modification mechanisms and gene activity during development.," 1975.
- [6] L. T. Kagohara *et al.*, "Epigenetic regulation of gene expression in cancer: techniques, resources and analysis," *Brief. Funct. Genomics*, vol. 17, no. 1, pp. 49–63, Jan. 2018, doi: 10.1093/BFGP/ELX018.
- [7] M. Kim and J. Costello, "DNA methylation: an epigenetic mark of cellular memory," *Exp. Mol. Med.* 2017 494, vol. 49, no. 4, pp. e322–e322, Apr. 2017, doi: 10.1038/emm.2017.10.
- [8] K. D. Robertson, "DNA methylation and human disease," *Nat. Rev. Genet.*, vol. 6, no. 8, pp. 597–610, Aug. 2005, doi: 10.1038/NRG1655.

- [9] S. Gopalakrishnan, B. O. Van Emburgh, and K. D. Robertson, "DNA methylation in development and human disease," *Mutat. Res. Mol. Mech. Mutagen.*, vol. 647, no. 1–2, pp. 30–38, Dec. 2008, doi: 10.1016/J.MRFMMM.2008.08.006.
- [10] B. Jin *et al.*, "DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function," *Hum. Mol. Genet.*, vol. 17, no. 5, pp. 690–709, Mar. 2008, doi: 10.1093/HMG/DDM341.
- [11] A. P. Feinberg and B. Vogelstein, "Hypomethylation distinguishes genes of some human cancers from their normal counterparts," *Nature*, vol. 301, no. 5895, pp. 89–92, 1983, doi: 10.1038/301089A0.
- [12] L. T. Kagohara *et al.*, "Global and gene-specific DNA methylation pattern discriminates cholecystitis from gallbladder cancer patients in Chile," *Future Oncol.*, vol. 11, no. 2, pp. 233–249, Jan. 2015, doi: 10.2217/FON.14.165.
- [13] H. Liu *et al.*, "Loss of epigenetic control of synuclein- γ gene as a molecular indicator of metastasis in a wide range of human cancers," *Cancer Res.*, vol. 65, no. 17, pp. 7635–7643, Sep. 2005, doi: 10.1158/0008-5472.CAN-05-1089.
- [14] J. Paredes, A. Albergaria, J. T. Oliveira, C. Jeronimo, F. Milanezi, and F. C. Schmitt, "P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation," *Clin. Cancer Res.*, vol. 11, no. 16, pp. 5869–5877, Aug. 2005, doi: 10.1158/1078-0432.CCR-05-0059.
- [15] T. Tokizane *et al.*, "Cytochrome P450 1B1 is overexpressed and regulated by hypomethylation in prostate cancer," *Clin. Cancer Res.*, vol. 11, no. 16, pp. 5793–5801,

Aug. 2005, doi: 10.1158/1078-0432.CCR-04-2545.

- [16] A. S. Wilson, B. E. Power, and P. L. Molloy, "DNA hypomethylation and human diseases," *Biochim. Biophys. Acta - Rev. Cancer*, vol. 1775, no. 1, pp. 138–162, Jan. 2007, doi: 10.1016/J.BBCAN.2006.08.007.
- [17] S. Kono and K. Chen, "Genetic polymorphisms of methylenetetrahydrofolate reductase and colorectal cancer and adenoma," *Cancer Sci.*, vol. 96, no. 9, pp. 535–542, Sep. 2005, doi: 10.1111/j.1349-7006.2005.00090.x.
- [18] M. Pufulete *et al.*, "Folate status, genomic DNA hypomethylation, and risk of colorectal adenoma and cancer: a case control study," *Gastroenterology*, vol. 124, no. 5, pp. 1240–1248, May 2003, doi: 10.1016/S0016-5085(03)00279-8.
- [19] "A Methylenetetrahydrofolate Reductase Polymorphism and the Risk of Colorectal Cancer1 | Cancer Research | American Association for Cancer Research." <https://aacrjournals.org/cancerres/article/56/21/4862/502775/A-Methylenetetrahydrofolate-Reductase-Polymorphism> (accessed Feb. 28, 2022).
- [20] P. A. Jones, J. P. J. Issa, and S. Baylin, "Targeting the cancer epigenome for therapy," *Nat. Rev. Genet.*, vol. 17, no. 10, pp. 630–641, Oct. 2016, doi: 10.1038/NRG.2016.93.
- [21] A. D. Kelly and J. P. J. Issa, "The promise of epigenetic therapy: reprogramming the cancer epigenome," *Curr. Opin. Genet. Dev.*, vol. 42, pp. 68–77, Feb. 2017, doi: 10.1016/J.GDE.2017.03.015.
- [22] A. P. Feinberg and B. Tycko, "The history of cancer epigenetics," *Nat. Rev. Cancer*, vol. 4, no. 2, pp. 143–153, 2004, doi: 10.1038/NRC1279.
- [23] M. E. Figueroa *et al.*, "DNA Methylation Signatures Identify Biologically Distinct Subtypes

- in Acute Myeloid Leukemia,” *Cancer Cell*, vol. 17, no. 1, p. 13, Jan. 2010, doi: 10.1016/J.CCR.2009.11.020.
- [24] Y. Lu, Y. T. Chan, H. Y. Tan, S. Li, N. Wang, and Y. Feng, “Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy,” *Mol. Cancer* 2020 191, vol. 19, no. 1, pp. 1–16, Apr. 2020, doi: 10.1186/S12943-020-01197-3.
- [25] T. J. Ley *et al.*, “DNMT3A Mutations in Acute Myeloid Leukemia A b s t r a c t,” *N Engl J Med*, vol. 363, pp. 2424–2457, 2010, doi: 10.1056/NEJMoa1005143.
- [26] G. A. Challen *et al.*, “Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells,” *Cell Stem Cell*, vol. 15, no. 3, pp. 350–364, 2014, doi: 10.1016/J.STEM.2014.06.018.
- [27] G. A. Challen *et al.*, “Dnmt3a is essential for hematopoietic stem cell differentiation,” *Nat. Genet.*, vol. 44, no. 1, pp. 23–31, Jan. 2011, doi: 10.1038/NG.1009.
- [28] C. Lauber *et al.*, “Survival differences and associated molecular signatures of DNMT3A-mutant acute myeloid leukemia patients,” *Sci. Reports* 2020 101, vol. 10, no. 1, pp. 1–17, Jul. 2020, doi: 10.1038/s41598-020-69691-8.
- [29] O. Emma, S. Prideaux, and T. Chevassut, “The Epigenetic Landscape of Acute Myeloid Leukemia,” vol. 2014, p. 103175, 2014, doi: 10.1155/2014/103175.
- [30] S. J. Kim, H. Zhao, S. Hardikar, A. K. Singh, M. A. Goodell, and T. Chen, “Brief Report MYELOID NEOPLASIA A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells Key Points • Mouse Dnmt3a R878H (human R882H) mutant protein inhibits wild-type Dnmt3a/Dnmt3b in murine ES cells, suggesting dominant-nega,” 2013, doi: 10.1182/blood-2013-02.

- [31] C. Quivoron *et al.*, "TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis," *Cancer Cell*, vol. 20, no. 1, pp. 25–38, 2011, doi: 10.1016/J.CCR.2011.06.003.
- [32] O. Abdel-Wahab *et al.*, "Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies," 2009, doi: 10.1182/blood-2009-03-210039.
- [33] S. M. C. Langemeijer *et al.*, "Acquired mutations in TET2 are common in myelodysplastic syndromes," *Nat. Genet.*, vol. 41, no. 7, pp. 838–842, Jul. 2009, doi: 10.1038/NG.391.
- [34] E. Papaemmanuil *et al.*, "Clinical and biological implications of driver mutations in myelodysplastic syndromes," *Blood*, vol. 122, no. 22, pp. 3616–3627, Nov. 2013, doi: 10.1182/BLOOD-2013-08-518886.
- [35] R. Itzykson *et al.*, "Clonal architecture of chronic myelomonocytic leukemias," *Blood*, vol. 121, no. 12, pp. 2186–2198, 2013, doi: 10.1182/BLOOD-2012-06-440347.
- [36] L. Busque *et al.*, "Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis," *Nat. Genet.*, 2012, doi: 10.1038/ng.2413.
- [37] K. Moran-Crusio *et al.*, "Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation," *Cancer Cell*, vol. 20, no. 1, pp. 11–24, 2011, doi: 10.1016/J.CCR.2011.06.001.
- [38] J. Madzo *et al.*, "Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis," *Cell Rep.*, vol. 6, no. 1, pp. 231–244, 2014, doi: 10.1016/J.CELREP.2013.11.044.
- [39] O. Abdel-Wahab *et al.*, "ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression," *Cancer Cell*, vol. 22, no. 2, pp. 180–193, 2012, doi:

10.1016/j.ccr.2012.06.032.

- [40] O. Shukron, V. Vainstein, A. Kündgen, U. Germing, and Z. Agur, "Analyzing transformation of myelodysplastic syndrome to secondary acute myeloid leukemia using a large patient database," *Am. J. Hematol.*, vol. 87, no. 9, pp. 853–860, 2012, doi: 10.1002/ajh.23257.
- [41] S. Asada *et al.*, "Mutant ASXL1 cooperates with BAP1 to promote myeloid leukaemogenesis," *Nat. Commun.* 2018 91, vol. 9, no. 1, pp. 1–18, Jul. 2018, doi: 10.1038/s41467-018-05085-9.
- [42] S. Schnittger *et al.*, "ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome," *Leukemia*, vol. 27, no. 1, pp. 82–91, 2013, doi: 10.1038/leu.2012.262.
- [43] C. L. Fisher *et al.*, "Additional sex combs-like 1 belongs to the enhancer of trithorax and polycomb group and genetically interacts with Cbx2 in mice," *Dev. Biol.*, vol. 337, no. 1, pp. 9–15, Jan. 2010, doi: 10.1016/J.YDBIO.2009.10.004.
- [44] C. L. Fisher *et al.*, "Loss-of-function Additional sex combs like 1 mutations disrupt hematopoiesis but do not cause severe myelodysplasia or leukemia," *Blood*, vol. 115, no. 1, pp. 38–46, Jan. 2010, doi: 10.1182/BLOOD-2009-07-230698.
- [45] L. Wang *et al.*, "Epigenetic targeted therapy of stabilized BAP1 in ASXL1 gain-of-function mutated leukemia," *Nat. Cancer* 2021 25, vol. 2, no. 5, pp. 515–526, May 2021, doi: 10.1038/s43018-021-00199-4.
- [46] K. Ueda and U. Steidl, "Epigenetic Achilles' heel of AML Mutations in genes encoding epigenetic modifiers are frequent in acute myelogenous leukemia (AML) and have been proposed to cause AML via activation of oncogenes and repression of tumor suppressors.

Two studies now identif,” doi: 10.1038/s43018-021-00212-w.

- [47] Y. Lu, Y.-T. Chan, H.-Y. Tan, S. Li, N. Wang, and Y. Feng, “Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy,” doi: 10.1186/s12943-020-01197-3.
- [48] S. Djebali *et al.*, “Landscape of transcription in human cells,” *Nature*, vol. 489, no. 7414, pp. 101–108, Sep. 2012, doi: 10.1038/NATURE11233.
- [49] R. C. Friedman, K. K. H. Farh, C. B. Burge, and D. P. Bartel, “Most mammalian mRNAs are conserved targets of microRNAs,” *Genome Res.*, vol. 19, no. 1, pp. 92–105, Jan. 2009, doi: 10.1101/GR.082701.108.
- [50] B. Weber, C. Stresemann, B. Brueckner, and F. Lyko, “Methylation of human microRNA genes in normal and neoplastic cells,” *Cell Cycle*, vol. 6, no. 9, pp. 1001–1005, May 2007, doi: 10.4161/CC.6.9.4209.
- [51] K. Arab *et al.*, “Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A,” *Mol. Cell*, vol. 55, no. 4, pp. 604–614, Aug. 2014, doi: 10.1016/J.MOLCEL.2014.06.031.
- [52] P. Han and C. P. Chang, “Long non-coding RNA and chromatin remodeling,” *RNA Biol.*, vol. 12, no. 10, p. 1094, 2015, doi: 10.1080/15476286.2015.1063770.
- [53] K. Grønbaek *et al.*, “A critical appraisal of tools available for monitoring epigenetic changes in clinical samples from patients with myeloid malignancies,” *Haematologica*, vol. 97, no. 9, pp. 1380–1388, Sep. 2012, doi: 10.3324/HAEMATOL.2011.058305.
- [54] J. Lammerding, “Mechanics of the nucleus,” *Compr. Physiol.*, vol. 1, no. 2, pp. 783–807, Apr. 2011, doi: 10.1002/CPHY.C100038.

- [55] M. Beato and K. Eisefeld, "Transcription factor access to chromatin," *Nucleic Acids Res.*, vol. 25, no. 18, pp. 3559–3563, Sep. 1997, doi: 10.1093/NAR/25.18.3559.
- [56] G. Felsenfeld, "Chromatin as an essential part of the transcriptional mechanism," *Nature*, vol. 355, no. 6357, pp. 219–224, 1992, doi: 10.1038/355219A0.
- [57] L. L. Wallrath, Q. Lu, H. Granok, and S. C. R. Elgin, "Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures," *Bioessays*, vol. 16, no. 3, pp. 165–170, 1994, doi: 10.1002/BIES.950160306.
- [58] L. Song and G. E. Crawford, "DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells," *Cold Spring Harb. Protoc.*, vol. 2010, no. 2, Feb. 2010, doi: 10.1101/PDB.PROT5384.
- [59] R. E. Thurman *et al.*, "The accessible chromatin landscape of the human genome," *Nature*, vol. 489, no. 7414, pp. 75–82, Sep. 2012, doi: 10.1038/NATURE11232.
- [60] G. E. Crawford *et al.*, "Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS)," *Genome Res.*, vol. 16, no. 1, pp. 123–131, Jan. 2006, doi: 10.1101/GR.4074106.
- [61] D. E. Schones *et al.*, "Dynamic regulation of nucleosome positioning in the human genome," *Cell*, vol. 132, no. 5, pp. 887–898, Mar. 2008, doi: 10.1016/J.CELL.2008.02.022.
- [62] J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf, "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position," *Nat. Methods*, vol. 10, no. 12, pp. 1213–8, 2013, doi: 10.1038/nmeth.2688.
- [63] X. Chen *et al.*, "ATAC-seq reveals the accessible genome by transposase-mediated

- imaging and sequencing,” *Nat. Methods*, vol. 13, no. 12, pp. 1013–1020, 2016, doi: 10.1038/nmeth.4031.
- [64] M. R. Corces *et al.*, “Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution,” *Nat. Genet.*, vol. 48, no. 10, pp. 1193–1203, 2016, doi: 10.1038/ng.3646.
- [65] R. M. Mulqueen *et al.*, “Improved single-cell ATAC-seq reveals chromatin dynamics of in vitro corticogenesis,” *bioRxiv*, p. 637256, May 2019, doi: 10.1101/637256.
- [66] J. D. Buenrostro *et al.*, “Single-cell chromatin accessibility reveals principles of regulatory variation,” *Nat.* 2015 5237561, vol. 523, no. 7561, pp. 486–490, Jun. 2015, doi: 10.1038/nature14590.
- [67] M. R. Corces *et al.*, “The chromatin accessibility landscape of primary human cancers,” *Science*, vol. 362, no. 6413, Oct. 2018, doi: 10.1126/SCIENCE.AAV1898.
- [68] D. R. Bond, K. Uddipto, A. K. Enjeti, and H. J. Lee, “Single-cell epigenomics in cancer: charting a course to clinical impact,” doi: 10.2217/epi-2020-0046.
- [69] D. J. H. F. Knapp and C. A. Hammond, “Single-cell analysis identifies a CD33+ subset of human cord blood cells with high regenerative potential,” 2018, doi: 10.1038/s41556-018-0104-5.
- [70] C. Angermueller *et al.*, “Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity,” vol. 13, no. 3, 2016, doi: 10.1038/nmeth.3728.
- [71] H. Dohner, “Acute Myeloid Leukemia,” *N. Engl. J. Med.*, vol. 373, pp. 1136–52, 2015, doi: 10.1056/NEJMra1406184.

- [72] D. Hanahan, “Hallmarks of Cancer: New DimensionsHallmarks of Cancer: New Dimensions,” *Cancer Discov.*, vol. 12, no. 1, pp. 31–46, Jan. 2022, doi: 10.1158/2159-8290.CD-21-1059.
- [73] S. Li *et al.*, “Somatic mutations drive specific, but reversible, epigenetic heterogeneity states in AML,” *Cancer Discov.*, vol. 10, no. 12, pp. 1934–1949, Dec. 2020, doi: 10.1158/2159-8290.CD-19-0897/333426/AM/SOMATIC-MUTATIONS-DRIVE-SPECIFIC-BUT-REVERSIBLE.
- [74] A. Chen, S. Hu, and Q.-F. Wang, “Tumor heterogeneity of acute myeloid leukemia: insights from single-cell sequencing,” *Blood Sci.*, vol. 1, no. 1, pp. 73–76, Aug. 2019, doi: 10.1097/BS9.0000000000000015.
- [75] C. Bravo González-Blas *et al.*, “cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data,” *Nat. Methods*, 2019, doi: 10.1038/s41592-019-0367-1.
- [76] D. Hu and A. Shilatifard, “Epigenetics of hematopoiesis and hematological malignancies,” *Genes Dev.*, vol. 30, no. 18, pp. 2021–2041, Sep. 2016, doi: 10.1101/GAD.284109.116.
- [77] A. Chen, S. Hu, and Q.-F. Wang, “Tumor heterogeneity of acute myeloid leukemia: insights from single-cell sequencing,” *Blood Sci.*, vol. 1, no. 1, pp. 73–76, Aug. 2019, doi: 10.1097/bs9.0000000000000015.
- [78] S. Li, C. Mason, and A. Melnick, “Genetic and epigenetic heterogeneity in acute myeloid leukemia,” *Curr. Opin. Genet. Dev.*, vol. 36, pp. 100–106, Feb. 2016, doi: 10.1016/J.GDE.2016.03.011.
- [79] A. N. Schep, B. Wu, J. D. Buenrostro, and W. J. Greenleaf, “ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data,” *Nat.*

Methods, vol. 14, no. 10, pp. 975–978, 2017, doi: 10.1038/nmeth.4401.

- [80] J. M. Granja *et al.*, “Author Correction: ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis (Nature Genetics, (2021), 53, 3, (403–411), 10.1038/s41588-021-00790-6),” *Nat. Genet.*, vol. 53, no. 6, p. 935, 2021, doi: 10.1038/s41588-021-00850-x.
- [81] K. A. Romine *et al.*, “Monocytic Differentiation and AHR Signaling as Primary Nodes of BET Inhibitor Response in Acute Myeloid Leukemia,” *Blood Cancer Discov.*, vol. 2, no. 5, pp. 518–531, 2021, doi: 10.1158/2643-3230.bcd-21-0012.
- [82] J. W. Tyner *et al.*, “Functional genomic landscape of acute myeloid leukaemia,” *Nature*, vol. 562, no. 7728, pp. 526–531, 2018, doi: 10.1038/s41586-018-0623-z.
- [83] J. R. Sinnamon *et al.*, “The accessible chromatin landscape of the murine hippocampus at single-cell resolution,” *Genome Res.*, p. gr.243725.118, 2019, doi: 10.1101/gr.243725.118.
- [84] A. Malovannaya *et al.*, “Analysis of the human endogenous coregulator complexome,” *Cell*, vol. 145, no. 5, pp. 787–799, 2011, doi: 10.1016/j.cell.2011.05.006.
- [85] C. G. Spruijt *et al.*, “ZMYND8 Co-localizes with NuRD on Target Genes and Regulates Poly(ADP-Ribose)-Dependent Recruitment of GATAD2A/NuRD to Sites of DNA Damage,” *Cell Rep.*, vol. 17, no. 3, pp. 783–798, 2016, doi: 10.1016/j.celrep.2016.09.037.
- [86] F. Gong *et al.*, “Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination,” *Genes Dev.*, vol. 29, no. 2, pp. 197–211, 2015, doi: 10.1101/gad.252189.114.
- [87] G. Divisato *et al.*, “ZNF687 Mutations in Severe Paget Disease of Bone Associated with

- Giant Cell Tumor,” *Am. J. Hum. Genet.*, vol. 98, no. 2, pp. 275–286, 2016, doi: 10.1016/j.ajhg.2015.12.016.
- [88] G. Divisato, F. Scotto di Carlo, N. Petrillo, T. Esposito, and F. Gianfrancesco, “ZNF687 mutations are frequently found in pagetic patients from South Italy: implication in the pathogenesis of Paget’s disease of bone,” *Clin. Genet.*, vol. 93, no. 6, pp. 1240–1244, 2018, doi: 10.1111/cge.13247.
- [89] F. Scotto di Carlo, M. P. Whyte, and F. Gianfrancesco, “The two faces of giant cell tumor of bone,” *Cancer Lett.*, vol. 489, no. June, pp. 1–8, 2020, doi: 10.1016/j.canlet.2020.05.031.
- [90] T. Zhang *et al.*, “Overexpression of zinc finger protein 687 enhances tumorigenic capability and promotes recurrence of hepatocellular carcinoma,” *Oncogenesis*, vol. 6, no. 7, 2017, doi: 10.1038/oncsis.2017.63.
- [91] T. T. Nguyen, L. N. Ma, S. M. L., C. D. Bangs, A. M. Cherry, and D. A. Arber, “Identification of Novel RUNX1 (AML1) Translocation Partner Genes SH3D19, YTHDF2, and ZNF687 in Acute Myeloid Leukemia,” *Genes, Chromosom. Cancer*, vol. 45.10, no. January, pp. 918–932, 2006, doi: 10.1002/gcc.
- [92] R. F. Schlenk *et al.*, “Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia,” *N. Engl. J. Med.*, vol. 358, no. 18, pp. 1909–1918, 2008, doi: 10.1056/NEJMoa074306.
- [93] H. K. Long, S. L. Prescott, and J. Wysocka, “Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution,” *Cell*, vol. 167, no. 5, pp. 1170–1187, 2016, doi: 10.1016/j.cell.2016.09.018.

- [94] Y. F. Tao *et al.*, "Zinc finger protein 382 is downregulated by promoter hypermethylation in pediatric acute myeloid leukemia patients," *Int. J. Mol. Med.*, vol. 34, no. 6, pp. 1505–1515, 2014, doi: 10.3892/ijmm.2014.1966.
- [95] A. D. van Dijk *et al.*, "Histone Modification Patterns Using RPPA-Based Profiling Predict Outcome in Acute Myeloid Leukemia Patients," *Proteomics*, vol. 18, no. 8, pp. 1–10, 2018, doi: 10.1002/pmic.201700379.
- [96] Y. Heshmati *et al.*, "The chromatin-remodeling factor CHD4 is required for maintenance of childhood acute myeloid leukemia," *Haematologica*, vol. 103, no. 7, pp. 1169–1181, 2018, doi: 10.3324/haematol.2017.183970.
- [97] G. Germano *et al.*, "ZNF521 sustains the differentiation block in MLL-rearranged acute myeloid leukemia," *Oncotarget*, vol. 8, no. 16, pp. 26129–26141, 2017, doi: 10.18632/oncotarget.15387.
- [98] S. A. Wolfe, L. Nekludova, and C. O. Pabo, "DNA RECOGNITION BY Cys 2 His 2 ZINC FINGER PROTEINS," 2000. [Online]. Available: www.annualreviews.org.
- [99] M. P. Scott, J. W. Tamkun, and G. W. Hartzell, "The structure and function of the homeodomain," *BBA - Rev. Cancer*, vol. 989, no. 1, pp. 25–48, Jul. 1989, doi: 10.1016/0304-419X(89)90033-4.
- [100] S. Jones, "An overview of the basic helix-loop-helix proteins," 2004. [Online]. Available: <http://genomebiology.com/2004/5/6/226>.
- [101] J. Jen and Y. C. Wang, "Zinc finger proteins in cancer progression," *J. Biomed. Sci.*, vol. 23, no. 1, pp. 1–9, 2016, doi: 10.1186/s12929-016-0269-9.
- [102] E. S. Lander *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol.

- 409, no. 6822, pp. 860–921, Feb. 2001, doi: 10.1038/35057062.
- [103] R. Tupler, G. Perini, and M. R. Green, “Expressing the human genome,” *Nature*, vol. 409, no. 6822, pp. 832–833, Feb. 2001, doi: 10.1038/35057011.
- [104] S. S. Krishna, I. Majumdar, and N. V. Grishin, “Structural classification of zinc fingers: survey and summary,” *Nucleic Acids Res.*, vol. 31, no. 2, pp. 532–550, Jan. 2003, doi: 10.1093/NAR/GKG161.
- [105] X. Li *et al.*, “Structures and biological functions of zinc finger proteins and their roles in hepatocellular carcinoma,” doi: 10.1186/s40364-021-00345-1.
- [106] P. Konieczny, E. Stepniak-Konieczna, and K. Sobczak, “MBNL proteins and their target RNAs, interaction and splicing regulation,” *Nucleic Acids Res.*, vol. 42, no. 17, pp. 10873–10887, Jun. 2014, doi: 10.1093/NAR/GKU767.
- [107] J. Font and J. P. MacKay, “Beyond DNA: zinc finger domains as RNA-binding modules,” *Methods Mol. Biol.*, vol. 649, pp. 479–491, 2010, doi: 10.1007/978-1-60761-753-2_29.
- [108] K. J. Brayer, S. Kulshreshtha, and D. J. Segal, “The protein-binding potential of C2H2 zinc finger domains,” *Cell Biochem. Biophys.*, vol. 51, no. 1, pp. 9–19, May 2008, doi: 10.1007/S12013-008-9007-6.
- [109] J. M. Matthews and M. Sunde, “Zinc fingers--folds for many occasions,” *IUBMB Life*, vol. 54, no. 6, pp. 351–355, Dec. 2002, doi: 10.1080/15216540216035.
- [110] B. N. Jeon *et al.*, “Two ZNF509 (ZBTB49) isoforms induce cell-cycle arrest by activating transcription of p21/CDKN1A and RB upon exposure to genotoxic stress,” *Nucleic Acids Res.*, vol. 42, no. 18, pp. 11447–11461, Oct. 2014, doi: 10.1093/NAR/GKU857.

- [111] S. Meruvu, L. Hugendubler, and E. Mueller, "Regulation of adipocyte differentiation by the zinc finger protein ZNF638," *J. Biol. Chem.*, vol. 286, no. 30, pp. 26516–26523, Jul. 2011, doi: 10.1074/JBC.M110.212506.
- [112] J. J. M. Cowger, Q. Zhao, M. Iovic, and J. Torchia, "Biochemical characterization of the zinc-finger protein 217 transcriptional repressor complex: identification of a ZNF217 consensus recognition sequence," *Oncogene*, vol. 26, no. 23, pp. 3378–3386, May 2007, doi: 10.1038/SJ.ONC.1210126.
- [113] S. Fietze, H. O'Geen, K. R. Blahnik, V. X. Jin, and P. J. Farnham, "ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes," *PLoS One*, vol. 5, no. 12, 2010, doi: 10.1371/JOURNAL.PONE.0015082.
- [114] C. B. Gocke and H. Yu, "ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers," *PLoS One*, vol. 3, no. 9, Sep. 2008, doi: 10.1371/JOURNAL.PONE.0003255.
- [115] Y. Ma, X. Zheng, J. Zhou, Y. Zhang, and K. Chen, "ZEB1 promotes the progression and metastasis of cervical squamous cell carcinoma via the promotion of epithelial-mesenchymal transition," *Int. J. Clin. Exp. Pathol.*, vol. 8, no. 9, pp. 11258–11267, 2015.
- [116] J. Jen *et al.*, "Oncogenic zinc finger protein ZNF322A promotes stem cell-like properties in lung cancer through transcriptional suppression of c-Myc expression," *Cell Death Differ.* 2018 267, vol. 26, no. 7, pp. 1283–1298, Sep. 2018, doi: 10.1038/s41418-018-0204-6.
- [117] "Protein motifs 5. Zinc fingers - PubMed." <https://pubmed.ncbi.nlm.nih.gov/7768350/> (accessed Feb. 21, 2022).
- [118] A. S. McCarty, G. Kleiger, D. Eisenberg, and S. T. Smale, "Selective dimerization of a C2H2

- zinc finger subfamily,” *Mol. Cell*, vol. 11, no. 2, pp. 459–470, Feb. 2003, doi: 10.1016/S1097-2765(03)00043-1.
- [119] J. M. Berg, “Zinc Finger Domains: Hypotheses and Current Knowledge,” <https://doi.org/10.1146/annurev.bb.19.060190.002201>, vol. 19, pp. 405–421, Nov. 2003, doi: 10.1146/ANNUREV.BB.19.060190.002201.
- [120] M. Cassandri *et al.*, “Zinc-finger proteins in health and disease,” *Nat. Publ. Gr.*, vol. 3, 2017, doi: 10.1038/cddiscovery.2017.71.
- [121] L. Gennari, D. Rendina, A. Falchetti, and D. Merlotti, “Paget’s Disease of Bone,” *Calcif. Tissue Int.* 2019 1045, vol. 104, no. 5, pp. 483–500, Jan. 2019, doi: 10.1007/S00223-019-00522-3.
- [122] G. Divisato *et al.*, “ZNF687 Mutations in Severe Paget Disease of Bone Associated with Giant Cell Tumor,” *Am. J. Hum. Genet.*, vol. 98, no. 2, pp. 275–286, 2016, doi: 10.1016/j.ajhg.2015.12.016.
- [123] J. Aijaz, S. Shang, D. C. Williams, and G. D. Ginder, “Abstract 82: NuRD dependent survival of acute leukemia cells,” *Cancer Res.*, vol. 80, no. 16_Supplement, pp. 82–82, Aug. 2020, doi: 10.1158/1538-7445.AM2020-82.
- [124] M. Varadi *et al.*, “NAR Breakthrough Article AlphaFold Protein Structure Database : massively expanding the structural coverage of protein-sequence space with high-accuracy models,” vol. 50, no. November 2021, pp. 439–444, 2022.
- [125] O. Ronneberger *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. May, 2021, doi: 10.1038/s41586-021-03819-2.
- [126] Y. Xue, A. Li, L. Wang, H. Feng, and X. Yao, “PPSP: prediction of PK-specific

- phosphorylation site with Bayesian decision theory," *BMC Bioinformatics*, vol. 7, Mar. 2006, doi: 10.1186/1471-2105-7-163.
- [127] W. Qiu, C. Xu, X. Xiao, and D. Xu, "Computational Prediction of Ubiquitination Proteins Using Evolutionary Profiles and Functional Domain Annotation," *Curr. Genomics*, vol. 20, no. 5, p. 389, Nov. 2019, doi: 10.2174/1389202919666191014091250.
- [128] L. de Beauchamp, E. Himonas, and G. V. Helgason, "Mitochondrial metabolism as a potential therapeutic target in myeloid leukaemia," *Leuk. 2021 361*, vol. 36, no. 1, pp. 1–12, Sep. 2021, doi: 10.1038/s41375-021-01416-w.
- [129] J. Kreitz *et al.*, "Metabolic Plasticity of Acute Myeloid Leukemia," *Cells 2019, Vol. 8, Page 805*, vol. 8, no. 8, p. 805, Jul. 2019, doi: 10.3390/CELLS8080805.
- [130] S. B. Panina, J. Pei, and N. V. Kirienko, "Mitochondrial metabolism as a target for acute myeloid leukemia treatment," *Cancer Metab. 2021 91*, vol. 9, no. 1, pp. 1–25, Apr. 2021, doi: 10.1186/S40170-021-00253-W.
- [131] R. C. Rao and Y. Dou, "Hijacked in cancer: the MLL/KMT2 family of methyltransferases," *Nat. Rev. Cancer*, vol. 15, no. 6, p. 334, May 2015, doi: 10.1038/NRC3929.
- [132] A. Ortega-Molina *et al.*, "The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development," 2015, doi: 10.1038/nm.3943.
- [133] S. X. Ge, E. W. Son, and R. Yao, "iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data," *BMC Bioinformatics*, vol. 19, no. 1, Dec. 2018, doi: 10.1186/S12859-018-2486-6.
- [134] N. Blom, S. Gammeltoft, and S. Brunak, "Sequence and structure-based prediction of

- eukaryotic protein phosphorylation sites," *J. Mol. Biol.*, vol. 294, no. 5, pp. 1351–1362, Dec. 1999, doi: 10.1006/JMBI.1999.3310.
- [135] N.-L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, and P. C. Ng, "SIFT web server: predicting effects of amino acid substitutions on proteins," doi: 10.1093/nar/gks539.
- [136] H. I. Baymaz, C. G. Spruijt, and M. Vermeulen, "Identifying Nuclear Protein–Protein Interactions Using GFP Affinity Purification and SILAC-Based Quantitative Mass Spectrometry," *Methods Mol. Biol.*, vol. 1188, pp. 207–226, 2014, doi: 10.1007/978-1-4939-1142-4_15.
- [137] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nat. Rev. Mol. Cell Biol.* 2015 161, vol. 16, no. 1, pp. 18–29, Dec. 2014, doi: 10.1038/nrm3920.
- [138] J. Gsponer, M. E. Futschik, S. A. Teichmann, and M. M. Babu, "Tight regulation of unstructured proteins: from transcript synthesis to protein degradation," *Science (80-.)*, vol. 322, no. 5906, pp. 1365–1368, Nov. 2008, doi: 10.1126/science.1163581.
- [139] T. Vavouri, J. I. Semple, R. Garcia-Verdugo, and B. Lehner, "Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity," *Cell*, vol. 138, no. 1, pp. 198–208, Jul. 2009, doi: 10.1016/j.cell.2009.04.029.
- [140] P. Tompa, "The interplay between structure and function in intrinsically unstructured proteins," *FEBS Lett.*, vol. 579, no. 15, pp. 3346–3354, Jun. 2005, doi: 10.1016/j.febslet.2005.03.072.
- [141] M. M. Babu, R. van der Lee, N. S. de Groot, and J. Gsponer, "Intrinsically disordered proteins: regulation and disease," *Curr. Opin. Struct. Biol.*, vol. 21, no. 3, pp. 432–440,

Jun. 2011, doi: 10.1016/j.sbi.2011.03.011.

- [142] B. Falini, L. Brunetti, P. Sportoletti, and M. P. Martelli, "NPM1-mutated acute myeloid leukemia: from bench to bedside," 2020, Accessed: Mar. 06, 2022. [Online]. Available: <http://ashpublications.org/blood/article-pdf/136/15/1707/1760541/bloodbld2019004226c.pdf>.
- [143] B. Falini, L. Brunetti, and M. P. Martelli, "How I diagnose and treat NPM1-mutated AML," *Blood*, vol. 137, no. 5, pp. 589–599, Feb. 2021, doi: 10.1182/BLOOD.2020008211.
- [144] J. Krönke *et al.*, "Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia," *Blood*, vol. 122, no. 1, pp. 100–109, 2013, doi: 10.1182/blood-2013-01-479188.
- [145] M. S. Bhakta and D. J. Segal, "The Generation of Zinc Finger Proteins by Modular Assembly," doi: 10.1007/978-1-60761-753-2_1.
- [146] L. de Beauchamp, E. Himonas, and G. V. Helgason, "Mitochondrial metabolism as a potential therapeutic target in myeloid leukaemia," *Leuk. 2021 361*, vol. 36, no. 1, pp. 1–12, Sep. 2021, doi: 10.1038/s41375-021-01416-w.
- [147] R. Sood, Y. Kamikubo, and P. Liu, "Role of RUNX1 in hematological malignancies," *Blood*, vol. 129, no. 15, pp. 2070–2082, Apr. 2017, doi: 10.1182/BLOOD-2016-10-687830.
- [148] Y. F. Chang, J. S. Imam, and M. F. Wilkinson, "The Nonsense-mediated decay RNA surveillance pathway," *Annu. Rev. Biochem.*, vol. 76, pp. 51–74, 2007, doi: 10.1146/ANNUREV.BIOCHEM.76.050106.093909.
- [149] S. J. Chen, Y. Shen, and Z. Chen, "A panoramic view of acute myeloid leukemia," *Nat. Genet.*, vol. 45, no. 6, pp. 586–587, 2013, doi: 10.1038/ng.2651.

