POST TRANSLATIONAL MODIFICATIONS OF C/EBPα p30 REGULATE ITS FUNCTION IN LEUKEMOGENESIS AND DIFFERENTIATION

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

im Fach Biologie

eingereicht an der

Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von

Nguyễn Thùy Linh, M.Sc.

Präsident (komm.) der Humboldt-Universität zu Berlin

Prof. Dr. Peter Frensch

Dekan der Lebenswissenschaftlichen Fakultät

Prof. Dr. Christian Ulrichs

Gutachter/innen: 1. Prof. Dr. rer. nat Achim Leutz

2. Prof. Dr. med. Lars Bullinger

3. Priv.-Doz. Dr. med. Adrian Schreiber

Tag der mündlichen Prüfung: 08 Juli, 2022

This thesis is dedicated to my parents.

Dành tặng Bố Mẹ.

Table of Contents

Abstract

Myeloid development is regulated by the family of transcription factors CCAAT/enhancer-binding-protein (C/EBP). Aberrant expression or functioning of C/EBPs disturbs normal myeloid differentiation and is found in many types of hematopoietic malignancies. Mutations of *CEBPA* lead to imbalanced expression of the truncated isoform C/EBPα p30 and are found in approximately 15% of AML (acute myeloid leukemia) patients. Germ-line mutations of *CEBPA* mark a highly penetrant, fast progressing type of AML that associates with favorable prognosis. Mutations of *CEBPA* were also reported in many cases of therapy-related AML and in lymphoid-myeloid lineage converted AML. Yet, how C/EBPα participates in leukemic progression remains to be discovered. More specifically, the truncated isoform C/EBPα p30, although being identified as an oncogenic isoform that promotes proliferation of myeloid progenitors, still retains differentiation function. The question of how both functions of C/EBPα p30 are regulated, is of our interest.

C/EBP family also represents a group of intrinsically disordered proteins, which contain many post-translational modifications (PTMs). PTMs on C/EBPα greatly alter its functioning. Previous works have identified three arginine residues at the N-terminus of C/EBPα p30 that interact differently with others protein dependent on their methylation status. We hypothesize, that methylation of these arginine residues plays important roles in the biology of C/EBPα p30, especially in regulating hematopoietic differentiation and transformation.

In this study, we used a lymphoid-to-myeloid transdifferentiation (LMT) system to investigate the influence of arginine-methylation on C/EBPα-induced lineage switch and its pro-leukemic activity. Using amino acid substitution, we found that C/EBPα p30 mutants that resemble arginine-methylated p30 or charge-depleted p30 enhanced myeloid differentiation, while the charge-retention mutant (resembling arginine-unmethylated p30) supported renewability and proliferation of hematopoietic stem/progenitor cells. Transcriptional profiling of cells expressing C/EBPα p30 variants suggested potential targets of either methylated or unmethylated p30. The results implied that arginine methylations alter C/EBPα p30's leukemic potential and might comprise novel targets of leukemia therapy.

Zusammenfassung

Die myeloische Entwicklung wird durch die Familie der Transkriptionsfaktoren CCAAT/Enhancer-Binding-Protein (C/EBP) reguliert. Eine aberrante Expression oder Funktion von C/EBPs stört die normale myeloische Differenzierung und wird bei vielen Arten hämatopoetischer Malignome beobachtet. Mutationen von *CEBPA* führen zu einem veränderten Expressionsanteil der verkürzten Isoform C/EBPa p30 und werden bei etwa 15% der AML-Patienten (akute myeloische Leukämie) nachgewiesen. Keimbahnmutationen von *CEBPA* führen zu einer stark erhöhte Wahrscheinlichkeit an einer schnell fortschreitenden Form der AML zu erkranken, die jedoch mit einer günstigen Prognose einhergeht. Mutationen in *CEBPA* wurden auch in vielen Fällen von therapieassoziierter AML beobachtet sowie in AML bei der eine lymphoide zu myeloischer Linienumwandlung stattgefunden hat. Die genaue Rolle von C/EBPα an der leukämischen Progression ist jedoch unklar. Obwohl die verkürzte Isoform C/EBPα p30 als Onkogen identifiziert wurde da sie die Proliferation myeloischer Vorläufer fördert, behält sie dennoch eine Differenzierungsfunktion. Unser Interesse gilt der Frage, wie diese beiden Funktionen von C/EBPα p30 reguliert werden.

Die C/EBP-Familie gehört der Gruppe intrinsisch ungeordneter Proteine an, die zudem viele posttranslationale Modifikationen (PTMs) aufweisen. PTMs auf C/EBPα verändern seine biologische Funktionsweise stark. Frühere Forschungsarbeiten haben drei Argininreste am N-Terminus von C/EBPα p30 identifiziert, die aufgrund des Methylierungsstatus differentiell mit anderen Proteinen interagieren. Wir vermuten, dass diese Argininmethylierungen bei der Regulation der hämatopoetischen Differenzierung sowie der leukämischen Transformation eine wichtige Rolle spielen.

In dieser Arbeit untersuchen wir den Einfluss der C/EBPα p30 Arginin-Methylierung auf seine pro-leukämische Aktivität sowie dessen Fähigkeit zur Neuausrichtung der hämatopoietischen Differenzierungslinie unter Zuhilfenahme eines lymphoid-myeloidem-Transdifferenzierungssystem (LMT). Mit Hilfe von Aminosäuresubstitutionen fanden wir heraus, dass C/EBPα p30 Mutanten der Methylierungsmimesis oder Ladungsabschaffung die myeloische Differenzierung verstärkt, während Ladungserhalt-Mutanten die Erneuerung und Proliferation hämatopoetischer Stamm-/Vorläuferzellen unterstützt. Transkriptionelles Profiling von Zellen, die mutierte C/EBPα -p30-Varianten exprimieren, deutet auf potenzielle Ziele der methyliertem bzw. unmethyliertem C/EBPα p30 hin. Die Ergebnisse legen nahe, dass der Arginin-Methylierungsstatus das Leukämie- und Differenzierungs-Potenzial von C/EBPα p30 verändert und somit ein neues Ziel der Leukämietherapie darstellen könnten.

1. INTRODUCTION

In this section, an overview of the differentiation and malignancies of the hematopoietic system will be provided. How transcription factor C/EBPα plays its role in hematopoiesis and how its isoforms are involved in the development of leukemia will be described in detail. The concept of intrinsically disordered protein, in relation to C/EBPα function, will be introduced.

1.1. Differentiation and Leukemogenesis in the hematopoietic system

1.1.1. Acute myeloid leukemia (AML)

Acute myeloid leukemia (AML) is a molecularly diverse malignancy of the hematopoietic system. In 2019, AML was reported as the most common type of adult leukemia with the lowest survival rate in the United States (accounts 62% leukemic deaths, shortest survival rate with 5-year survival = 24%) [1]. Other epidemiologic index of AML includes increased incidence with age (median age at diagnosis of 68), as well as dependency of risks on biological gender, ethnicity, lifestyle, and environmental exposure. AML can be revealed by routine blood examinations and it also manifests medical complications such as infections, hemorrhage, intravascular coagulation [2]. Various examinations are required to classify the AML subtypes, allocate patients into the correct risk groups, and assign suitable treatments. Those include assessment of cell morphology from bone marrow and peripheral blood, marker expression analysis by flow cytometry, chromosomal abnormality by standard karyotype analysis, and molecular genetic lesions screening [3]. The classification system currently used for AML diagnostic is the 2017 edition of *World Health Organization (WHO) Classification of Tumours of Haematopoietic and Lymphoid Tissues* [4]. The classification contains six subtypes, among which, “AML with recurrent genetic aberrations” is listed as the first and one of the major subtypes. The importance of genetic aberrations was recognized after various studies involving next-generation sequencing reported that up to 95% of AML patients carried one or more driving mutations, with an average number of 3 mutations per patient [5]–[7]. This finding adds more patient-specific variable features (mutations and combinations of them) as well as requirement for appropriate personalized treatment.

Like most leukemia types, AML emerges with abnormal proliferation and survival of hematopoietic precursors of the myeloid lineage, and restriction of differentiation to more

mature myeloid cell types, which altogether leads to expansion of malignant clones. These aberrations are caused by genetic alterations, including chromosomal mutations and gene mutations, that eventually activate oncogenes and/or suppress tumor suppressors.

Next generation sequencing technology, including whole genome, RNA and micro-RNA sequencing has been widely used to study the genomic landscape of AML. An analysis by the Cancer Genomic Atlas Research Network on 200 genomes of AML patients identified genes that were significantly mutated and sorted them into various functional categories. Those includes:

i. **Transcription factors:** CEBPA, RUNX1, and GATA2. Mutations of transcription factors, especially key myelopoiesis regulators, comprise 20-25% of adult AML cases. Identified mutations includes frameshift, missense and nonsense which causes complete loss, or expression of truncated or non-coding isoforms. Impairment of key regulators results in dysregulated transcriptional program and hindered differentiation [8]–[10]. Targeted therapy for transcription factors remains challenging, with various therapeutic strategies currently being developed [11].
ii. **Signaling and kinase pathway:** FLT3, KRAS, NRAS, KIT, PTPN11, and NF1. As the most common group of mutations, they are present in nearly two-thirds (approximately 65%) of AML cases. Mutated genes affect signaling pathways confer proliferation and survival advantages [5]. Treatments for this AML category involve combination of FLT3 inhibitor and multi-kinase inhibitors that have been well established in various combinatorial regimens [12].
iii. **Epigenetic modifiers** (DNA methylation and chromatin modification): DNMT3A, IDH1, IDH2, TET2, ASXL1, EZH2, and MLL/KMT2A. They were identified in approximately 50% AML. Mutated epigenetic modifiers are considered key component of leukemogenesis [12]. Loss or impaired function of the regulators results in epigenetic silencing or activating genes that are important for hematopoietic stem cells' self-renewal, survival, differentiation, or induction of metabolic profiles that favors undifferentiated state [13]–[15]. Therapeutic approaches using small molecule inhibitors are developing rapidly with several approved products [16], [17].

iv. **Nucleophosmin:** NMP1. Nuclear chaperon protein NMP1 is one of the most common driver mutations of AML, alongside with FLT3 and DNMT3. Because of its involvement in diverse cellular processes, such as cell cycle, DNA replication, ribosome biogenesis…, mutations of NMP1 cause abnormal cell activities. Nevertheless, NMP1 mutated AML holds favorable prognosis thanks to good response to standard chemotherapy. In contrast, in the setting of DNMT3 mutated or FLT3-IDT (FLT3 internal tandem duplication) mutated AML, NMP1 mutations add up the severity and poor outcomes [18]–[20].

v. **Tumor suppressors:** TP53 mutations cause adverse effects on genomic integrity. P53 mutations identified in *de novo* AML are rare, but rather more common in therapy-related AML or AML with myelodysplasia-related changes and confer high chemo-resistant and poor outcome. Although standard treatments shows poor or moderate results, new therapy including checkpoint inhibitions, antibody-based therapy and CAR-T cell therapy is promising [21]–[23].

vi. **Spliceosome complex:** SRSF2, U2AF1, SF3B1, and ZRSR2. Mutations in components of spliceosome complex perturbs exon recognition and normal splice sites, thus leads to mis-spliced genes and dysregulated cellular processes such as epigenetic regulation, DNA-damage response [7]. These mutations account for approximately 10% of AML cases and mostly link to aged patients with poor response to treatment and poor survival rate [24].

vii. **Cohesin complex:** RAD21, STAG1, STAG2, SMC1A, and SMC3. Mutations of cohesin complex members alter chromatin accessibility, thus interfere with gene expressions; they may also hinder inflammation-induced differentiation [25]. Specific treatment for this type of AML is currently not available.

Although bearing the same entity, AML subsets manifest high heterogeneity at molecular, cytogenetic, and clinical levels. This, therefore, accentuates the need for advanced prognostic accuracy and more targeted therapies for each AML subsets. Together with the progressive development of technology, understanding of pathophysiology and revolution of AML has been growing drastically, paving the way for new treatment opportunities. Since 2017, several new agents have been approved by Food and Drug Administration for AML treatment, including: FLT3 inhibitors (Midostaurin, Gilteritinib), IDH1/2 inhibitors (Enasidenib, Ivosidenib), BCL-2 inhibitor (Venetoclax: targeting apoptosis regulator, high efficiency to

IDH1/2-mutated AML), Hedgehog pathway inhibitor (Glasdegib: targeting Hedgehog pathway receptor on leukemic stem cells to inhibit self-renewal), Anti-CD33 antibody (Gemtuzumab ozogamycin: chemotherapy-conjugated monoclonal antibody targeting myeloid leukemic blasts), Hypomethylating agents (CC-486, Oral Decitabine-cedazuridine)[12]. Depending on patient-specific physical/pathological condition, these agents are used in combinatorial regimes with/without chemotherapy or allogenic stem cell transplantation. Various new agents are currently investigated, Polo-like-1 kinase inhibitors and CAR-T cells therapy targeting CD123, CLL1 are some examples.

The establishment of new compounds is the result of translation from a large number of basic studies in molecular genetics, which have deepened our understanding of the molecular mechanisms of leukemogenesis. Along this line, this study intends to contribute to the basic knowledge of leukemia pathogenesis and aims further to prompt new treatment considerations.

1.1.2. Therapy-induced AML and leukemic lineage switching

During leukemia treatment, relapses with serious clinical complications and significant changes in lineage identity, in comparison to the point of diagnosis, is known as lineage switching, -infidelity, -ambiguity, or -promiscuity. First introduced in 1984, lineage switching was reported in relapsed acute leukemia patients with the interval time of 26-32 months after diagnosis and intensive chemotherapy [26]. Until now, many cases have been reported with phenotypic switches from lymphoblastic leukemia (T-ALL or B-ALL) to AML (most common) [27]–[30], or vice versa [31]–[33], and several switches following relapse [34][35]. The most likely explanation of this phenomenon is clonal selection upon a/few resistant clone(s) under the pressure of therapy. The therapy-resisted cell clones may exist originally in the malignant population, or may have evolved from therapy-targeted clones by acquiring additional mutations [36][34]. Although the direct evidence of mutated transcription factors-driven lineage switch is yet to be revealed, abnormal activities of PU.1, GATA-1 and C/EBPα were shown in leukemogenesis [37], or appeared in mix-lineages leukemia with signs of lineage switch [38]. Increase expression of *CEBPA* and reduction of lymphoid signature *EBF1* during a CAR-T-induced B-ALL-to-AML switch was observed in mouse [39]. This phenomenon emphasizes a challenge in clinical treatment covering leukemia plasticity and a detailed

characterization of its molecular mechanism may propose a new strategy to target the disease.

Intrinsic hematopoietic lineage plasticity underlies the malignant cell lineage switch. The mammalian blood system develops from hematopoietic stem cells (HSCs) through a series of stepwise lineage commitment events. The determination of hematopoietic lineages are under the control of external cues, lineage-instructive transcription factors, epigenetic regulators and the dynamics of their expression [40]–[42]. Changing the expression of such master regulators may alter lineage decision and may result in lineage infidelity. *In vitro*, human pro-B cells can adapt to environmental signals (in this case, cytokines) and differentiate into other lineages: T cells, macrophages and natural killer cells [43]. In response to inflammation, a fraction of bone marrow pre-B cells may give rise to functional tissue-residential and inflammatory macrophages, demonstrating lineage plasticity as a general phenomenon that remains largely undetected under non-selective conditions [44]. *In vivo*, deletion of GATA-1, an erythroid cell fate regulator, disturbs erythropoiesis while overexpression of GATA-1 directs myelomonocytic cells into megakaryocytic/erythroid fate [45][46]. Removal of B cell determinant Pax5 or ectopic expression of myeloid specific transcription factors C/EBPα and PU.1 lead to a halt of the lymphoid fate and a myeloid switch of B cells or T cells to macrophages and neutrophils [47]–[51]. Whether lineage switching is involved in the adaptation of the hematopoietic system to environmental cues remains an unresolved question, however, this type of plasticity is observed during leukemic lineage switch [5][6].

The transcription factor C/EBPα, as the main subject of this study, was implied in many reports of lineage switching leukemia. In fact, *CEBPA* gene upregulation was observed in lineage switching during leukemia progression and occurred prior to the observed switching [52]. *CEBPA* mutations were even found in ALL (Acute lymphoblastic leukemia) cases. A recent study of a cohort of Turkish pediatric acute leukemia patients reveals that *CEBPA* mutations appeared in 16/30 (53.3%) patients, among which, 6 (37.5%) patients were first diagnosed with ALL [53]. Another screening on 30 patients that carried *CEBPA* wildtype alleles at diagnosis showed an N-terminal frame-shift mutation in CEBPA at relapse in one patient (3%) [54]. Genomic profiling of B cell ALL on 172 adult and pediatric patients showed one sub-group with highly expressed myeloid genes *CEBPA*, *CEBPB*, *SPI1*, while lymphoid genes *PAX5,*

BACH2 and *EBF1* were expressed at low level, implying lineage infidelity in the genomic background of B-ALL [55]. These findings suggested that the attained mutations in *CEBPA* add survival or proliferation advantages to the existing leukemic progression.

Lineage ambiguity underscores both the complexity of the hematopoietic transcriptional network and the current challenge to precisely target certain type of leukemic mutation in cancer treatment. As the subject of this study, the lineage instructive transcription factor C/EBPα, especially its biology, regulation, function in hematopoiesis and leukemogenesis, will be discussed.

1.2. The transcription factor C/EBPα

1.2.1. C/EBPα isoforms and functions

CCAAT/enhancer binding protein alpha (C/EBPα) was the first member of the C/EBP family of transcription factors to be discovered, followed by five other members: β, γ, δ, ε and ζ, which were named in chronological order of discovery [56]. In 1986, Graves et. al reported the identification of C/EBPα as a nuclear protein fraction that showed strong affinity toward a sequence-specific motif 5'-CCAAT-3' presented at the promoter of several mRNA-coding genes [57]. C/EBPα was cloned two years after its finding by the same laboratory [58] and paved the way for the uncovering of the bZIP (basic-leucine zipper) class of DNA-binding domain transcription factors, which includes e.g. fos and jun proto-oncogenes [59]–[61]. Until 1992, other members of the C/EBP family were discovered and cloned from different species [62]–[69]. Since then, the functions of C/EBPs have been being studied intensively and revealed to play critical roles in various cellular processes, such as differentiation of adipocytes/hepatocytes/myelomonocytic cells/neuronal cells/epithelial cells, metabolic regulation, inflammation responses, proliferation and cell cycle controls [70]. Within the scope of this study, I will focus on the founding member C/EBPα and provide only brief information about other C/EBP members.

Structure

All C/EBPs have highly conserved C-terminal sequences (more than 90% similarity among members), comprising the bZIP domain; while they vary at their N-termini (20% similarity)(**Figure 1**)[70]. The bZIP domain is a sequence of 55-65 amino acids containing

b **:** a basic-amino-acid-rich region that binds the major groove of DNA,

ZIP **:** a "leucine zipper" dimerization motif, which contains 4-5 heptad repeat leucine residues allowing formation of a stable alpha helical dimer of two similar structured C/EBP polypeptides [59], [62]–[69], [71], [72].

All C/EBPs can recognize and bind to cognate binding sites on DNA and at the same time form homo- or heterodimers with intrafamilial members (except C/EBPζ, which possess a different structure at the bZIP domain due to few amino acid variations [66][56]). Besides, a bZIP-associated sequence, called bZIP "Tail" domain, which plays role in interaction with other DNA-binding factors, can be found in all C/EBPs [73], [74]. Unlike the C-terminus, the N-terminus of each C/EBP is considerably diverse. Nevertheless, there are three sequence regions that remain conserved among the members (**Figure 1**)[75]–[78]. The three conserved domains are responsible for interaction with the transcriptional machinery components to stimulate transcription (e.g. TBP/TFIIB, p300, CREB [79]–[81]), interaction with the cell cycle regulators (e.g. CDK2/CDK4, E2F [82], [83]) and chromatin remodeling machinery (SWI/SNF complex [84]); thus called transactivation domains (TADs). Some C/EBPs contain regulatory domains (RDs) that modulate the transcriptional activity of the TADs [76], [85]–[88]. These TADs and RDs vary in length and position not only among C/EBP members, but also among isoforms of some C/EBPs (**Figure 1**). The isoforms are structure variations produced from alternative translation start codons, as in the case of C/EBPα and C/EBPβ [35][54][55], or from multiple promoters usage and differential splicing, in the case of C/EBPε [91]. These variations add more complexity in structure, regulation, and functions of the C/EBPs. In the next paragraphs, details in structure variations and functions of C/EBPα will be described.

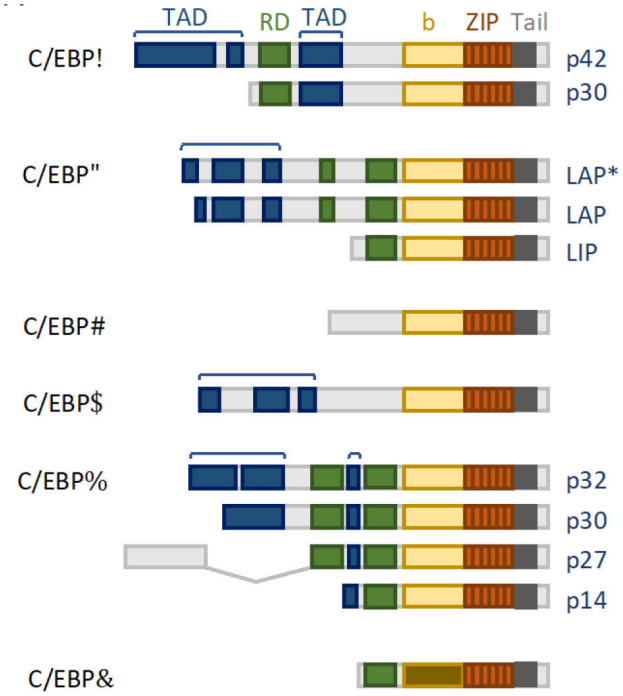


Figure 1. Schematic representation of C/EBP family.
The conserved transactivation domains (TADs) are shown as blue boxes. Regulatory domains (RDs) are shown in green. The bZIP domain consists of a basic region (b, yellow) and a tandem-repeated sequence of the leucin zipper (ZIP, brown). Lastly, the bZIP tail is also shown. Adapted from [68].

The *CEBPA/Cebpa* gene is intronless and located on 19q13.11 or 7qB1 chromosome in human or mouse, respectively. The solely transcribed mRNA consists of a protein coding sequence in between a 5'- and a 3'-UTR (untranslated region) [91]. In the first successful cloning of C/EBPα in 1988, a Western blot analysis presented **two** polypeptides species migrated with distinct molecular mass, while in Northern blot analysis, only **one** mRNA species was found [58]. This was the first sign of the present of other variations beside the fully translated form of C/EBPα. The most abundant polypeptide marked at 42 kDa, while the less abundant species, which had a mass of 30 kDa, was falsely recognized as a degradation product of the former [58]. It was later found that the 30-kDa isoform, termed p30 C/EBPα, was a product of alternative translation from a downstream in-frame start codon, thus lacking of the 117 first amino acids in comparison to the full length isoform, termed p42 C/EBPα [90]. The mechanism behind the alternative translation was brilliantly described in a study by Calkhoven et al. [92]. Among three start codons found existed on C/EBPα mRNA, the first AUG codon resides in an 18-nucleotides 5' upstream open reading frame (uORF) and is only 7

nucleotides away from the C/EBPα cistron [93]; this start codon is important for a leaky ribosomal scanning process, thus allowed skipping and alternatively initiating translation at the second AUG that gives rise to the full-length isoform p42. More importantly, the 5'uORF is indispensable for the translational re-initiation at the third AUG, which gives rise to the truncated isoform p30. This translational control of C/EBPα is dependent on the level of translation initiation complex, especially eukaryotic initiation factors eIF2a, eIF4E, which is regulated by PKR (Protein kinase R) and mTOR (Mammalian target of rapamycin) signaling pathway [92]. Using rapamycin to inhibit mTOR function results in less expression of p30 C/EBPα, while over expression of eIF4E increases the expression of the p30 isoform. This regulation allows adjustment of p30/p42 ratio in accordance with environmental cues such as nutrients and growth factors, subsequently directing cells towards proliferation or differentiation.

Function

The full-length p42 C/EBPα, but not the truncated p30, possess the transactivation domain 1 (TAD1) located within the first 117 amino acid (**Figure 1**). TAD1 activates transcription of target genes by recruiting components of the RNA polymerase II preinitiation complex, such as TFIIB (transcription factor IIB), TBP (TATA-box binding protein) [81]. TAD1 is also involved in cell cycle control function of C/EBPα by repressing E2F-dependent transcription, with the proliferation regulator c-Myc being one of the C/EBPα-E2F targets [82], [94], [95]. TAD1 is absent from the truncated isoform p30 C/EBPα, therefore the antimitotic activity mediated by TAD1 is lacking in p30. The second transactivation domain TAD2 is shared by both p42 and p30 C/EBPα. TAD2 is reported to induce and stabilize p21, a cyclin-dependent kinase inhibitor that cause growth arrest [96], and to interact with chromatin remodeling complex SWI/SNF [97].

The doctrine that 'cellular differentiation and proliferation exist in a mutually exclusive paradigm' is represented clearly in the C/EBPα structure and function, that the TADs induce differentiation on one hand and repress proliferation on the other. The differentiation of hepatocytes, adipocytes and lung are shown reliant on C/EBPα [71][72]. When the growth inhibitory activity of C/EBPα is blocked by activating PI3K/Akt pathway in hepatocyte, liver tumor cells arise [99]. While being crucial for adipogenesis and granulopoiesis, *Cebpa*$^{-/-}$ mice

manifest elevated self-renewal of hematopoietic stem cells (HSCs) and accumulation of myeloblast in bone marrow [82][97][100], [101]. In lung, C/EBPα controls the expression of bronchiolar epithelium differentiation marker CCSP (Clara cells secretory protein), and also plays an important functional role in lung epithelial cells [74][75]. Loss of C/EBPα in the respiratory epithelium not only results in defective lung function at birth due to differentiation arrest of type I/II alveolar cells, but also causes increasing number of proliferative, apoptosis resistant and expanding epithelial cells, leading to loss of airspace [104]. The most intriguing function of C/EBPα is to guard the entrance of differentiation from multipotent hematopoietic progenitor into all myeloid fate, especially important for the transition from common myeloid progenitor (CMP) to granulocyte-monocyte progenitor (GMP) [77][78]. The loss of full-length p42 C/EBPα skews hematopoiesis and with certain conditions, leads to leukemia (this topic will be described in detail in the next section). Accordingly, p42 C/EBPα is a differentiation factor and tumor suppressor with dual capacity in differentiation induction and proliferation inhibition.

Additionally, various metabolic processes are found dependent on C/EBPα. Homozygous deletion of *Cebpa* gene causes mice to die within 8 hours after birth due to hypoglycemia, a condition of failure in storing hepatic glycogen, with the root cause lies in decreased expression of gluconeogenic enzymes [107][108]. Further investigation reveals wider role of C/EBPα in hypoglycemia conditions, which is regulation of ornithine cycle enzyme genes; mutant mice lacked *Cebpa* fail to detoxify products of amino acid metabolism and have higher blood ammonia concentration than wild-type mice [109]. C/EBPα is involved in high-density lipoprotein clearance that controls plasma lipid in blood [108]. Some genes important for energy homeostasis and fall under the regulation C/EBPα are leptin (energy balance/fat storage regulator), insulin receptor substrate 1 (signal transmitter of PI3K/Akt and Erk MAP kinase pathway, which subsequently regulate glucose uptake), peroxisome proliferator-activated receptor-γ (regulator of fatty acid and glucose metabolism) [110]. The significance of C/EBPα in metabolic processes again emphasize the most prominent potential of it: governing gene expression in terminally differentiated cells. This also implies that aberrant expression or function of C/EBPα may alter normal gene regulatory networks and ultimately lead to malfunction and illness.

Dysregulated expression of *CEBPA/Cebpa* has been found in various types of solid tumor [111]. Expression of *Cebpa* in hepatocarcinoma cell lines is repressed by the oncogene YY1 (Yin Yang 1) and TNFα (Tumor necrosis factor alpha) signaling or microRNA miR-182, which subsequently releases cell cycle progression from the inhibition of C/EBPα [112]–[114]. Re-introducing C/EBPα rescues the growth inhibition and reduces hepatocarcinogenesis [115]–[118]. In breast cancer, *CEBPA* is expressed at diminished level in correlation with high c-myc, cyclin D1, cyclin E (promotes cell cycle progression), and low p21, p27, p16, Rb (cell cycle inhibitors) [119][120]. Hypoxia condition and the hypoxia-inducible factor 1a (HIF-1a) account for direct repression of *CEBPA* in advanced breast tumors; while downstream targets of C/EBPα in breast cancer cell lines are identified as miR-134, CREB (anti-apoptotic gene activator, downregulated by C/EBPα) and Bcl-2 (apoptosis regulator) [95][96]. Research on treatment of breast cancer found that C/EBPα is prerequisite for tamoxifen-induced apoptosis, it was also identified as potential target of the antiproliferation vitamin D (active form $1,25(OH)_2D_3$) [97][98]. Hypermethylation of the *CEBPA* gene leading to lowered expression can be found in lung cancer, head and neck squamous cell carcinoma and pancreatic cancer [125]–[127]. These findings again prove that C/EBPα is indispensable for a fully controlled cellular growth. Nevertheless, studies of *CEBPA* gene expression cannot discriminate between the two isoforms. The function of p30 C/EBPα is more extensively studied in the field of hematology and malignant hematopoiesis due to its involvement in acute myeloid leukemia, which is explained in 1.2.2.

To conclude, the transcription factor C/EBPα, founding member of C/EBP family, comprises two isoforms: p42 and p30, as a result of an alternative translation initiation. The full-length isoform p42 is a tumor repressor which induces gene expression, differentiation and restricts proliferation. Absence or reduction of p42 expression is correlated with elevated cell growth or even cancerous development. Moreover, the p30 isoform lacks the antimitotic features of p42, yet maintains distinct regulatory functions.

1.2.2. C/EBPα in hematopoiesis

C/EBPα can be found abundantly in various cells of the hematopoietic system, including neutrophils, splenic basophils, monocytes, macrophages; a minimal amount expressed in bone marrow short-term HSC, MMP-expressing bone marrow cells, dendritic cells and

eosinophils [128][129]. Nearly no expression of C/EBPα is found in megakaryocytes, erythrocytes, mast cells and lymphoid cells, as C/EBPα is known for inhibition of differentiation into those lineages [130]–[132]. This lineage specific expression represents, in part, the inducing function of C/EBPα exclusively in myelopoiesis and more specifically granulocytic and monocytic development, rather than other myeloid fates.

C/EBPα plays a role as a myeloid priming factor in hematopoietic stem/progenitor cells (HSPCs). The first study of germ-line deletion of *Cebpa* in 1997 shows that differentiation into neutrophils, but not monocyte, is halted in the absence of *Cebpa*. Further transplantation of *Cebpa*$^{-/-}$ fetal liver reconstitutes lymphoid but not neutrophilic cells in irradiated recipients, suggesting that loss of *Cebpa* affects the differentiation of multipotential myeloid precursors, not only the granulocytic precursors [133]. In a similar study, hematopoietic progenitors in *Cebpa*$^{-/-}$ fetal liver are hyperproliferative and show reduced differentiation [134]. While *Cebpa*-deficient animals die shortly after birth due to lung and liver failures, thus cause difficulty to study adult hematopoiesis, the establishment of the Mx1-Cre conditional knock out mice gave further insights into C/EBPα's involvement in this process. Conditional *Cebpa*$^{-/-}$ mice presents accumulation of myeloblast in bone marrow, enhanced HSPCs repopulation capacity, increased expression of self-renewal regulatory factor *Bmi-1* (B lymphoma Mo-MLV insertion region 1 homolog) [134]. C/EBPα is indeed expressed in LT-HSCs (Long-term hematopoietic stem cell), as reported in several studies, and these *CEBPA*-expressing LT-HSCs possess robust myeloid potential [135]–[137]. Knocking out *Cebpa* results not only in block of myeloid differentiation but also relieve quiescence of LT-HSCs, with the latter showing active cell cycle progression, expansion of the HSC population and subsequent exhaustion of LT-HSC [134][138]. More recent study utilizing CHIP-sequencing technology provides a detailed observation of C/EBPα activity alongside of enhancer establishment during G-M (granulocyte and monocyte/macrophage) differentiation [139]. By profiling the CMP/preGMP (Common myeloid progenitor/pre- granulocyte macrophage progenitor) cells from either *Cebpa*-KO or *Cebpa*-WT mice (termed preGM-KO or preGM-WT, respectively), the authors found exceeded expression of stem cells and megakaryocyte-erythrocyte precursors' signatures in preGM-KO, as well as depleted expression of the preGM signatures. Based on features of enhancer usage (marked by histone methylation status), chromatin binding dynamic of PU.1 (C/EBPα regulatory partner), and global gene expression patterns, the preGM-KO cells are recognized

as more similar to $Lin^{-}c\text{-}Kit^{+}Sca\text{-}1^{+}$ hematopoietic stem cells. As such, C/EBPα is suggested to be important for the downregulation of Sca-1. Moreover, C/EBPα can access closed chromatin regions of myeloid-specific enhancers before they are modified with active marks. These results demonstrate a concept which was first proposed by the same group (B.T. Porse) in 2014: C/EBPα is a pioneer transcription factor that primes myeloid commitment at the early transitional stage from HSPCs to more differentiated progenitors [139], [140].

Myelopoiesis proceeds further with the transition from CMP to GMP and subsequently to monocyte/granulocyte, which is critically dependent on C/EBPα. Conditional disruption of C/EBPα results in loss of the GMP population in bone marrow, while other populations including CMPs, CLPs (Common lymphoid progenitor), MEPs (Megakaryocyte–erythroid progenitor) remain unaffected [134]. In fact, ectopic expression of C/EBPα stimulates the myeloid program and gives rise to GMP in various cell types. Transduction of *CEBPA* into human $CD34^{+}$ HSPCs/AML cells leads to growth arrest of the cells and expression of typical myeloid markers CD14, CD15, CD11b, neutrophilic elastase and granulocyte colony-stimulating factor receptor (G-CSFR) [141], [142]. *In vitro*, lymphoid T-cell and B-cell progenitors can be redirected into myeloid progenitors and further develop to granulocytes/monocytes when C/EBPα, in collaboration with PU.1 or C/EBPβ, is overexpressed [143], [144]. C/EBPα-primed reprogramming by the Yamanaka factors (Oct4, cMyc, Klf4 and Sox2/3 [145]) converts B cells into induced pluripotent stem cells through an intermediate GMP fate [146]–[148]. Cells from nonhematopoietic background, such as fibroblast from mouse embryo or even adult skin, can transdifferentiate into functioning macrophage by overexpressing C/EBPα and PU.1 [51]. These findings depict the important role of C/EBPα as a major differentiation factor of myeloid lineage, which is indispensable to activate myeloid program from various cell types.

Beyond the GMP stage, the role of C/EBPα in lineage choice between granulocytic and monocytic fate remains unsettled. Nevertheless, various studies show that C/EBPα is required for both specifications, possibly in a dosage dependent manner.

- Apparently, more C/EBPα is required for granulopoiesis than monopoiesis [149]. Immature granulocyte precursors, identified by $Lin^{-}Sca\text{-}1^{-}cKit^{+}GCSFR^{+}MCSFR^{-}$ present higher expression of *Cebpa, Cebpe, Gfi1* and *Klf5*; while the immature

monocyte precursors $Lin^{-}Sca\text{-}1^{-}cKit^{+}GCSFR^{-}MCSFR^{+}$ show high *Klf4* and *Irf8*. Knocking down of C/EBPα by shRNA results in reduction of *Cebpe, Gfi1* and *Klf5.* These factors are known for their importance in granulopoiesis: *Cebpe* deficiency leads to functional defects in neutrophil progenitors [150], *Gfi*-deficient mice suffer from severe neutropenia and fail to induce neutrophil differentiation upon G-CSF stimulation [151], Klf5 expression is required for granulocyte differentiation in 32Dcl3 cells [152]. C/EBPα is found inducing various miRNAs to promote granulopoiesis, including miR-223 (miR-223 represses the translation of nuclear factor NFI-A, which normally acts as granulopoiesis inhibitor) [153], [154], and miR-30c (miR-30c downregulates Notch1, which is also a granulocytic differentiation inhibitor)[155]–[157]. Other genes important for early and late neutrophilic development are regulated by C/EBPα, including *Mpo* (myeloperoxidase), *Elane* (neutrophil elastane), *Lyz* (lysozyme) and *Ltf* (lactoferrin) [158]. The strongest evidence for the prominence of C/EBPα in granulopoiesis is that $Cebpa^{-/-}$ mice lack of peripheral eosinophils and neutrophils but not monocytes [133], and ectopic expression of *Cebpa* triggers granulocytic differentiation *in vitro* [159], [160].

- C/EBPα directly binds and/or regulates expression of both G and M lineages instructive growth factors G-CSF [161], GM-CSF[162] and M-CSF [163], [164] in a synergistic manners with others transcription factors PU.1 and RUNX1 (RUNX family transcription factor 1, alias AML1). A recent study using CRISP-sequencing with sgRNA against *Cebpa* reveals a depletion of both monocytic and granulocytic differentiation [106] (this study, however, used $Lin^{-}c\text{-}Kit^{+}$ hematopoietic stem cells for sgRNA infection; therefore, a valid argument can be that knockdown of *Cebpa* already blocked G-M differentiation at the precursor GMP stage). The monocyte/macrophage differentiation marker CD14 shows promoter bound heterodimerized C/EBPα and C/EBPβ [165]. Conditional expression of *Cebpa*-estradiol receptor (ER) chimera in immature myeloid cells in GM-CSF-containing medium yields up to 1.9 fold more monocytes than granulocytes, even with low concentration of estradiol (E2) [165]. Interestingly, this observation varies upon (1) growth factor addition: addition of G-CSF gives rise to only granulocytic CFU-G, while addition of M-CSF produces only monocytic CFU-M; and (2) C/EBPα level: in

GM-CSF-supplemented medium, low/physiologic level of C/EBPα results in more CFU-M while increased level of C/EBPα brings almost equal CFU-G and CFU-M numbers. These findings not only point out the role of C/EBPα in monopoiesis, and underscore the cooperating lineage instruction between C/EBPα and the CSFs, but also reveal how C/EBPα expression levels variegate the fate choice outcome. Transdifferentiation of lymphoid T- or B-cells using C/EBPα overexpression generates mainly monocyte-derived cells [143]. A study by Cirovic et al*.,* in which transdifferentiation capacity of C/EBP family members was tested side by side, also reported conversion of B-cells to monocytes, but not granulocytes, by C/EBPα [166]. Concrete explanation for this failure of granulocytic differentiation is still missing, nevertheless, one suggested possibility is growth inhibition effect by the overexpressed C/EBPα.

To sum up, in hematopoiesis, the full-length C/EBPα functions as a predominant lineage regulator which promotes the development of myeloid monocytic-granulocytic lineages. C/EBPα function is suggested to harmonize differentiation induction and proliferation inhibition, since loss of C/EBPα leads to loss of differentiated cells and accumulation of hyperproliferative progenitors. The following paragraph addresses how C/EBPα expressions and activities are regulated, as well as the mechanism behind such regulation.

The regulation of C/EBPα will be discussed briefly in this paragraph, mainly focusing in two aspects: regulation of *Cebpa* gene expression and regulation of C/EBPα function.

- **Regulation of *Cebpa* gene expression**. The promoter of *Cebpa* gene is auto-activated by C/EBPα protein, as proven by the presence of C/EBPα-binding site in the promoter region, and the transactivation of C/EBPα promoter-luciferase construct by transient C/EBPα expression [167][168]. The proximal promoter is also shown activated to a lesser extent by other family member C/EBPβ and C/EBPδ. In human, *CEBPA* promoter is thought to be auto-activated via stimulation and binding of USF (Upstream stimulation factor) to USF consensus element within *CEBPA* promoter [168]. Zinc finger protein 143 (ZNF143) is recently found to bind a conserved regulatory sequence in *CEBPA* promoter and activate gene expression distinctively in myeloid cells [169]. An autonomous enhancer at +42 kb (conserved

homologous enhancer at +37 kb in mice) is found critical for *CEBPA* expression in myeloid cells, which contain binding sites of various transcription factors including C/EBP, Ets factors, SCL (TAL BHLH transcription factor 1), GATA2 (GATA binding protein 2), MYB (Avian myeloblastosis viral oncogene homolog) [170], [171]. RUNX1 and the Ets-family member PU.1, two transcription factors essential for hematopoietic differentiation, also bind this enhancer [170]. Runx1 directly activates both *Spi1* and *Cebpa* (gene encodes PU.1 and C/EBPα, respectively) transcription; *Runx1* deletion/inhibition strongly reduced C/EBPα level, thus leads to impairment of myeloid differentiation and granulopoiesis [172], [173]. Interestingly, deletion of the *Cebpa* enhancer reduces expression of *Spi1*, pointing out the regulation of C/EBPα on *Spi1* [172]. Other direct regulators of *Cebpa* expression include LEF-1 (Lymphoid enhancer binding factor 1, reduced expression of which results in impaired granulopoiesis and congenital neutropenia) [174], HIF-1a (Hypoxia-inducible factor 1 alpha, mediates differentiation of AML cell lines via C/EBPα induction) [175], and ecCEBPA (extra-coding CEBPA, a 4.5 kb RNA transcribed from the *Cebpa* locus, interacts with DNMT1 to prevent methylation at *Cebpa* promoter, thus increases gene expression) [176].

To inhibit the myeloid differentiation program during cell fate decision, non-myeloid factors suppress *Cebpa* expression. Hes1, a Notch signal target important for T-cell development, directly binds *Cebpa* promoter and represses gene expression in T- cell progenitor [177]. However, T-cell lymphopoiesis is restored without Hes1 upon *Cebpa* deletion, providing evidence that T-cell development requires restriction of C/EBPα that, in turns, acts as a repressor of T-cell-fate. Increasing of repressive histone marks at *Cebpa* locus, leading to lowered *Cebpa* expression, is also found during T-cell commitment in thymus [178]. A recent study, using sequence-base thermodynamic model and follow-up experiments found dominant repression of *Cebpa* promoter by GATA and Myb bindings in erythroid cell lineage [179].

- **C/EBPα function regulation.** During myelopoiesis, activity of C/EBPα is shown varied following amino acid modifications, which can be directed by lineage

instructive growth factors. Both macrophage-colony stimulating factor M-CSF and granulocyte-colony stimulating factor G-CSF can direct lineage differentiation of sorted GMP [180]. M-CSF activates ERK (extracellular signal-regulated kinase), which sequentially stimulates C/EBPα serine 21 (S21) phosphorylation in Lin^- marrow cells [181]. The phosphorylation of C/EBPα S21 via ERK1/2 was reported previously to favor monopoiesis by inhibiting granulopoiesis [182]. M-CSF also stabilizes c-Fos, thus indirectly supporting C/EBPα:c-Fos hetero-dimerization that stimulate more monopoiesis [181], [183]. G-CSF, on the other hand, activates SHP2 (Src-homology-region-2 domain-containing phosphatase 2) which reduces ERK activities and thus, reduces C/EBPα S21 phosphorylation, enabling more granulocyte development [181]. Phosphorylation of C/EBPα S21 is also found regulated by p38 MAPK (Mitogen activated protein kinase) and leads to similar outcome: inhibition of neutrophil differentiation from $CD34^+$ progenitor [184]. SUMOylation of C/EBPα is shown important to maintain myelopoiesis/erythropoiesis balance in zebra fish myeloid-erythroid progenitors; hypo-SUMOylation promotes the myelopoiesis process at the expense of the erythropoiesis [185]. More on C/EBPα function variations upon protein modifications will be described in 1.3.

- **Joint action with other transcription factors.** C/EBPα functions are furthermore fine-tuned by crosstalk with other transcription factors during hematopoiesis. PU.1 is a well-studied lineage determining transcription factor that has a dynamic interplay with C/EBPα. The expression of their encoding genes, *Spi1* and *Cebpa*, is dependent on each other: *Cebpa* is induced by PU.1 binding to the +37 kb enhancer and *Spi1* induced by C/EBPα binding to -14 kb enhancer [170], [186]. Recent study by Pundhir and Bratt Lauridsen et. al, which investigates the functional crosstalk between C/EBPα and PU.1 during myelopoiesis with regards to chronology of enhancer establishment, reveals pattern of enhancer binding by both factors: PU.1. binds majorly to early-stage ($Lin^-cKit^+Sca\text{-}1^+$ progenitor stage) established enhancers while C/EBPα predominantly binds to GMP-stage established enhancers [187]. The authors describe pioneering activity of C/EBPα during the transition into GMP stage and moreover finds CEBPA-dependent binding of PU.1 onto many GM-

lineage enhancers. Interferon regulatory factor 8 (IRF8) is proven by co-immunoprecipitation assay to physically bind C/EBPα and prevent chromatin binding of C/EBPα, subsequently blocking expression of C/EBPα target genes in monocyte-dendritic cells progenitors (MDPs) and common monocyte progenitors (cMoPs) [187]. Max, a heterodimeric partner of Myc oncogene, directly interacts and enhances transactivation activity of C/EBPα, thus promotes granulopoiesis; as a result, loss of Max function reduces differentiation potential of C/EBPα in myeloid progenitors [188].

The above-mentioned points altogether demonstrate the tight control of C/EBPα expression and function. While *Cebpa* gene sustains induction or repression by lineage priming transcription factors, the functioning of C/EBPα protein differs upon **its own active structures** – determined by post-translational modification – and **collaborative partners**. This allows an accuracy as well as a wide variety of roles that C/EBPα plays during hemato-/myelo-poiesis. However, a complex regulatory network also accommodates risks of dysregulation, which are indeed found in C/EBPα in leukemic context. In the next section, the aberrations of C/EBPα reported in leukemia will be described.

1.2.3. C/EBPα in leukemia

As curated in 1.2.1, C/EBPα is a tumor suppressor which is found dysregulated in many types of solid tumor, this is also true in hematopoietic malignancies. Different impairments of C/EBPα function are found rooted in deregulation of *CEBPA* gene and/or protein expression.

- **Transcriptional deregulation**. Hypermethylation of *CEBPA* promoter, majorly in distal promoter (-1422 to -1121 upstream of transcription start site, TSS), is observed in 37% AML patients [189]. Similar observation were made by Fasan et al., in which, 38.2% AML cases show high distal promoter methylation and 2.5% show core promoter (-141 to +103 upstream of TSS) methylation [190]. Stratified AML patients with recurrent cytogenetic aberrations (inv(16), t(8;21), t(15;17), t(9;11)) or complex karyotypes raised incidents of dense *CEBPA* promoter methylation to 51% [190]. Interestingly, this study discovers the C/EBPα mRNA-

targeted microRNA miR-124a as a post-transcriptional regulator, which recognizes the 3' untranslated regions and strongly reduces C/EBPα protein level [190]. These finding underscore the importance of epigenetic regulation of C/EBPα and signify *CEBPA* DNA hyper-methylations as a deregulated mechanism supporting myeloid malignancy. Another mechanism to reduce *CEBPA* expression is enhancer hijacking by oncogenes. A zinc finger oncoprotein EVI1 (Ecotropic viral integration site 1) is found binding to the conserved *Cepba* +35, +37 kb enhancer elements with strong affinity and significantly suppress gene expression [191], [192]. Of note, this enhancer region holds binding sites of lineage specific regulators and is important for myeloid differentiation function of C/EBPα [170]. The suppressed *Cebpa* expression by EVI1 is observed in AML-transformed cell lines, AML patient samples and mouse bone-marrow stem cells, but not in committed progenitors, indicating its interference with lineage differentiation [192]. Recent analysis of EVI1-deregulated high-risk AML subtype points at C/EBPα as associated protein of EVI1, suggesting the enhancer hijacking mechanism that blocks downstream myeloid genes expression [193].

- **Post-transcriptional deregulation.** Translation of C/EBPα mRNA is inhibited by several factors. Oncoproteins BCR-ABL induce the production of poly(rC)-binding protein hnRNP-E2, which in turn directly binds to a specific motif in C/EBPα mRNA and impede the translation [194]. As a consequence, C/EBPα-induced expression of G-CSF receptor is blocked and myeloid precursor cells fail to differentiate upon G-CSF induction. BCR-ABL also downregulates miR-328, a microRNA that releases C/EBPα mRNA from the hnRNP-E2; and loss of miR-328 is often found in blast crisis chronic myeloid leukemia [195]. Oncogenic miR-182, which is highly expressed in AML with C-terminal mutated C/EBPα, and C/EBPα itself hold a regulatory loop: they directly obstruct the expression of each other. Enforced expression of miR-182 inhibits G-CSF mediated granulopoiesis and enhances replating capacity of bone marrow cells [196].

- **Post-translational deregulation.** Serine 21 of C/EBPα protein can be phosphorylated by ERK1/2 and leads to granulocytic differentiation block (see

1.2.2). In nearly 30% of AML patients, constitutive activation of FLT3 (Fms-like tyrosine kinase 3) is present and constantly induces downstream ERK1/2 pathway, thus affecting C/EBPα S21 phosphorylation, leading to leukemic blasts [197]. Differentiation of granulocyte can be rescued by either pharmacological inhibiting FLT3, or mutating S21 into a null version Alanine; in contrast, mutation into phosphorylation-mimicking Aspartate blocks differentiation. Lysine acetyltransferase GCN5 (General control non-derepressible 5) mediates acetylation of C/EBPα lysine residues K298 and K302 to diminish DNA binding activity, leading to loss of gene activation function of C/EBPα [198]. Trib2 (Tribble homolog 2) is an oncogene found in a subset of AML patient and inhibits C/EBPα [199]. Trib2 mediates ubiquitination at lysine residues K48 and K63 of C/EBPα, as marks for ubiquitin-dependent proteasomal degradation and indirectly degrades the p42 isoform, but not p30 [200]. Absence of C/EBPα disables Trib2-induced leukemogenesis; and interestingly, proteasomal inhibitors can target Trib2 function and rescue C/EBPα p42 [200].

Altogether, the described findings show that dysregulation of C/EBPα is involved in multiple pathways that lead to myeloid malignancy. Although the initiation of leukemia in the listed studies relies on aberration of other factors, involvement of C/EBPα is still strongly required as a differentiation mediator. Indeed, C/EBPα-regulated differentiation was shown to be prerequisite for the acquisition of leukemic stem cell and AML initiation [201].

In a separate development, **AML including mutated C/EBPα manifests distinct onset and mechanism**. In 2001, Pabst et al. published the first report of *CEBPA* mutations in AML patients [202]. Among 137 investigated patients, 10 (7.3%) carried at least one mutation of *CEBPA*, 7 in these 10 patients show no karyotypic abnormalities, meaning that the *CEBPA* mutation(s) is(are) the driver of malignancy. Five patients carry N-terminal mutations that truncate the full-length p42 isoform, while expression of p30 isoform elevates to 4-8 folds. The p30-expressing cells fail to induce promoters of C/EBPα target genes (in this case, *GCSFR*), and are deficient in transactivation function and block granulopoiesis [202]. Many studies in the following years, especially recently with the application of high resolution sequencing technology, revealed more insights into genetic traits and prognosis of *CEBPA*-mutated AML [6], [10], [14], [148], [188], [203]–[207], including:

- Frequency of *CEBPA*-mutated AML is 5-15% varied between cohorts of studies. Among which:
- nearly 1/3 of cases exhibit mutations on only one allele, which is termed *CEBPA* monoallelic mutants or single mutants, ***CEBPA*-sm**. Among these *CEBPA*-sm mutants, majority contains N-terminal mutations while C-terminal mutations are rare. *CEBPA*-sm frequently co-occurs with mutations on other major transcription factors, such as NMP1 (nucleophosmin 1), FLT3.
- the rest 2/3 of cases exhibit biallelic or double mutants, termed ***CEBPA*-dm**. Among the *CEBPA*-dm mutants, almost 95% show mutations in both terminals: 1 out-of-frame mutation at N-terminal, 1 in-frame mutation at C-terminal; in short, N/C mutant. The remaining 5% is very rare cases that carry biallelic mutations at either N- or C-terminal (N/N or C/C mutant).
- *CEBPA*-mutated AML can be familial, since 5-15% *CEBPA*-dm patients carry germline mutation [208]–[211]. The identified germline mutations are mainly N-terminal out-of-frame mutations, which are proven to have high penetrance because most patients develop AML at young age (median age at diagnose 24.5) when an additional C-terminal mutation is acquired. C-terminal germline mutations are often not shown in family history, patients acquire addition N- or C-terminal mutations at the time of diagnosis.

The high frequency of N-terminal mutations observed in both *CEBPA*-sm and *CEBPA*-dm, with or without germline disposition, suggests a critical role of them in the development of AML. This will be discussed in the coming paragraphs. In 2017, the European Leukemia Net (ELN) officially recognized *CEBPA*-dm as a full entity of AML, on account of distinct molecular profile and association with more favorable prognosis [10], [211]–[214]. Routine screening for *CEBPA* mutations at first diagnosis of AML was also recommended by ELN from 2010, long before assigning it to an independent category [215]. On the other hand, treatment responses and outcomes cannot be predicted in *CEBPA*-sm AML cases; treatment consideration and outcome anticipation are based on accompanied mutations, such as *FLT3-IDT, NMP1, IDH1/2, WT-1* [216]–[218].

The N-terminal out-of-frame mutations disturb normal translation of the full-length isoform p42 by introduce a stop codon after the usual AUG, leading to a premature stop of

translation and a resumed initiation at the next in-frame start codon (see 1.2.1). This disturbance extinguishes p42 and, instead, produces more dominant-negative p30 than normal, which tips the p42:p30 balance, causing failure in granulocytic differentiation and enhanced proliferation of myeloid progenitors (**Figure 2**) [202], [219], [220]. As described in 1.2.1, p30 isoform lacks of the first TAD (TAD1), which interacts with E2F proteins to inhibit cell cycle progression, therefore also lacks of this growth inhibitory function [221], [222]. Isoform p30 derived from *CEBPA*-dm AML patients also show decreased DNA binding activity, despite the DNA binding domain (bZIP) is still intact, preceding the disease onset [202]. It was also shown that even with the intact bZIP domain, C/EBPα p30 binds its specific site on *Spi1* and *Gr* genes, with 3-6 times lower affinity, thus less capable of directing granulopoiesis upon G-CSF signal [223]. Nevertheless, the lower DNA binding activity of C/EBPα p30 is possibly due to reduced interactions with other proteins, which is normally executed by TAD1. On the other hand, C-terminal mutations disrupt the bZIP domain. Defective basic DNA-binding domain impairs the DNA-binding activity, thus abolish transcription activation; while defective leucine zipper domain impairs homo- and hetero-dimerization [220], [224], [225]. As such, the *CEBPA*-dm AML patients possess only a truncated isoform and a defective full-length C/EBPα, with a complete lack of wild-type C/EBPα.

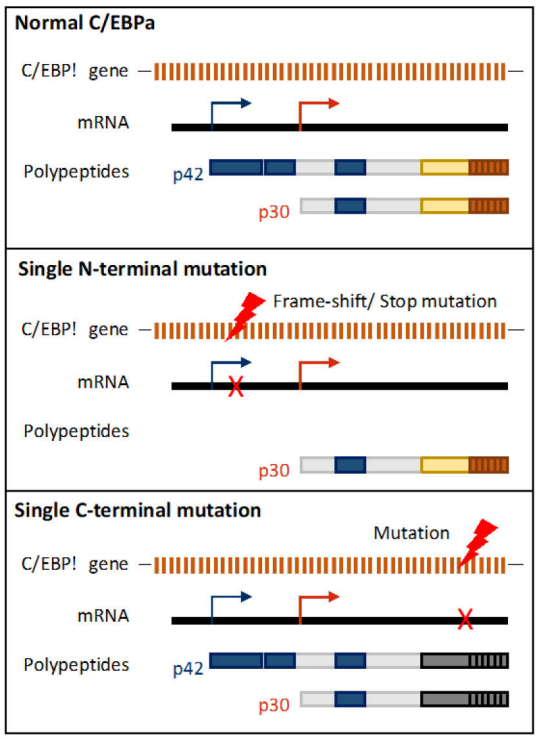


Figure 2. Schematic representation of the C/EBPα translation in normal or mutated conditions.

(Top) Normal C/EBPα exists in both isoforms and in a regulated balance.

(Middle) N-terminal mutations cause loss of the p42 isoform and increased production of p30 isoform, thus overturn the normal p42:p30 ratio. DNA binding function of the produced p30 remains unchanged.

(Bottom) C-terminal mutations do not affect the p42:p30 ratio, instead, disrupt the bZIP domain. Both products, p42 and p30, loss DNA binding function. In most CEBPA-dm AML, C-terminal mutations are found on one allele while on the other allele, mutations occur at N-terminal, leaving the bZIP domain intact.

This raises the question of which *CEBPA* mutations trigger the AML leukemogenesis in human: N-terminal mutations or C-terminal mutations. First of all, *Cebpa*$^{-/-}$ mice show enhanced proliferation of hematopoietic progenitors and presence of myeloblasts in bone marrow, which is similar to human AML. But more importantly, these mice resemble some characteristics of AML, such as accumulation of immature cells, yet do not develop AML with other leukemic traits (anemia, thrombocytopenia, disease succumbence, etc.) [134]. Experimental mouse models of both type of mutations were developed by Nerlov's lab. Cre-mediated deletion of a LoxP-flanked cassette in between p42 and p30 initiation codons allows formation of a stop codons, thus abrogates p42 translation and expresses only p30 (termed in publication L/L for homozygous, +/L for heterozygous with one wild-type allele) [226][221].

Mimesis of C-terminal mutation is lysine duplication K313KK (termed K/K), generated by targeting knock-in of additional lysine [227]; other C-terminal mutation includes point mutation at bZIP domain causes loss of E2F interaction, termed BRM2 [224]. And by combining the two models, various patient mutations, including *CEBPA*-sm (+/L and +/K) and *CEBPA*-dm (L/L, K/K and L/K - representing N/N, C/C and N/C mutant AML, respectively) could be generated. Heterozygous genotype +/L and +/K both show no transformation of hematopoietic progenitors, meanwhile homozygous mice L/L expressing only p30 show myeloid transformation and AML development within 9-18 weeks. Bone marrow analysis of the L/L reveals accumulation of an immature Mac-1^{+}Gr-1^{lo} granulocytic population, high myeloid blast count, enlarged spleen, the animals die of liver and bone marrow failure. Timed monitoring of pre-leukemic L/L cells showed enrichment of cKit^{+} progenitors and L/L GMPs exhibit enhanced serial replating efficiency, indicating that p30 expression, without p42, is enough to drive lineage differentiation to GMP stage [226]. The biallelic BRM2 mice possess transplantable granulocytic myeloproliferative disorder and accumulation of blasts, nevertheless, many pathological features of AML were missing [224]. The K/K mutant shows a strongly enhanced proliferation of Lin^{-}Sca-1^{+}cKit^{+} stem cells, however, very few transformed cells show myeloid identity and the model slowly develops erythroid leukemia, indicating that mutations at C-terminal disrupt myeloid differentiation due to failed DNA-binding of C/EBPα, leading to another leukemia type but not AML [227]. Interestingly, heterozygous mutant L/K accelerates leukemic transformation with even stronger kinetics and more malignant phenotype than the L/L combination. Altogether, these findings meet at an agreement that the *CEBPA*-driven leukemogenesis requires a formation of myeloid committed progenitor GMP, possibly by p30's residual functions and escalates expansion of those premalignant precursors upon C-terminal mutation (**Table 1**). The concept, that transformation-susceptible GMP is required for leukemic development, is further supported by studies on transformation by other factors. BCR-ABL, MLL-ENL or HoxA9/Meis1 cannot induce myeloid leukemia without C/EBPα, as this transcription factor is needed for CMP to GMP commitment [105], [228], [229].

Table 1. Models of mutated C/EBPα in normal and malignant hematopoiesis

Model	Represented AML type	C/EBPa function	Phenotype	Ref.
C/EBPα ***-/-***	-	Null	No GMP and granulocyte differentiation, enhanced HSC self-renewal, no AML	[134]
C/EBPα ***+/Lp30***	N-terminal *CEBPA*-sm	Only p30 expressed on 1 allele	Normal hematopoiesis, no AML	[226]
C/EBPα ***Δ/Lp30***	-	Only p30 expressed	Reduced GMP production, myeloid differentiation block, GMP wit partial lymphoid profile, transformation to AML with myeloid blast and pathologic traits	[221]
C/EBPα ***Lp30/Lp30***	N/N *CEBPA*-dm	Only p30 expressed	Normal GMP formation. Transformation to AML with granulocytic blast, pathologic and cytogenetic traits	[226] [227]
C/EBPα ***Δ/BRM2***	-	Loss of E2F repression	Reduced GMP production, myeloid differentiation block, enhanced self-renewal of GMP but not immortalized, no AML.	[221]
C/EBPα ***BRM2/BRM2***	-	Loss of E2F repression	Transplantable granulocytic myelo-proliferative disorder, no AML	[224]
C/EBPα ***+/K313KK***	C-terminal *CEBPA*-sm	Loss of DNA binding on 1 allele	No expansion of HSC, no AML	[227]
C/EBPα ***K313KK/K313KK***	C/C *CEBPA*-dm	Loss of DNA binding	Increased HSCs number, no GMP formation, slow transformation of erythroid progenitor, minimal myeloid differentiation, no AML	[227]
C/EBPα ***K313KK/Lp30***	N/C *CEBPA*-dm	Loss of DNA binding on 1 allele, only p30 expressed on 1 allele	Increased HSCs number, low GMP formation, rapid and lethal transformation, accumulation of granulocytic blast, AML confirmed by pathologic and cytogenetic traits	[227]

Models with AML transformation are high-lighted.

N: N-terminal; C: C-terminal; E2F: E2 Factor; GMP: granulocyte-macrophage progenitor; HSCs: hematopoietic stem cells; AML: acute myeloid leukemia.

Considering the role of N-terminal mutations, or more specifically, the C/EBPα p30 isoform, in AML, the entitlement of it as a negative regulator of p42 is no longer correct, as it clearly shown lineage commitment capacity. However, the loss of growth inhibitory function and elevated proliferation induction are also proven features of p30. It is, however, unclear **how p30 navigates between the two functions: differentiation versus proliferation**. C/EBPα p30 is shown to bind stronger than p42 to certain target genes (MPP11, p84N5, SMYD2) and induce their expression in hepatocyte, which could not be seen with p42 [230]. A recent study by Jacobsen et al., using combinatorial approach on both human AML and mouse model p30 AML, revealed a novel addition to mechanism of *CEBPA*-driven leukemogenesis [231]. Using CHIP-seq for genomic occupancy analysis on GMP cells from WT or L/L mice (see above), the author revealed that 87.7% of the CEBPA-bound genomic regions are common between two isoforms, while 7.9% are p30 specific and only 4.5% are bound solely by p42. While p42 evinces pioneer potential allowing access to closed chromatin regions, p30 binds only open (H3K4me1 marked) or active (H3K27ac marked) enhancers. P42 was associated with more down-regulated genes than up-regulated genes, while p30 binding regions are preferentially in up-regulated genes; this means that p42 functions as a suppressor and this function is lessened when switching to p30. Transcription profiling also reveals genes specifically regulated by p30, among which *NT5E/Nt5e* (encoding CD73) is a proven functionally target by p30 in both species. *NT5E* is known as a cancer-associated genes in various tumor types. An enhancer at -40 kb upstream of TSS (-48 kb in human) of *Nt5e* is activated by p30 and manifests tumor-promoting effects. In addition to discovering positive regulatory functions of p30 this study also adds one more mechanism of AML transformation by CEBPA p30, which includes

- loss of p42-specific gene activation functions (differentiation-related genes);
- loss of p42-specific gene suppressing functions (proliferation-related genes);
- at the shared binding regions, reduced transcription due to low gene activation function of p30; and finally,
- activation of p30-specific transcription at established enhancer.

Finding of distinct activity of p30 has already shown benefit for targeting therapy: using of anti-CD73 antibody in combination with inhibitory small molecules shows reduced tumor growth, elevated apoptosis of AML cells and higher survivor rates, proving therapeutic benefit

for *CEBPA*-mutated AML [231]. Therefore, the topic of C/EBPα p30 function is becoming an open field for more studies and promising options for better treatment of *CEBPA*-mutated AML.

1.3. Intrinsically disordered regions, post-translational modifications and C/EBPα functions

1.3.1. Intrinsically disordered regions

Protein function has been traditionally believed to be governed by the structure – function paradigm: function of a protein is critically defined by a stable folded three-dimensional structure of polypeptides determined by its genetic sequence. This paradigm was first challenged in 1950 by Karush, who noticed that albumin presented more than one configuration of binding activity, allowed it to bind differently shaped molecules, unlike the classic lock-and-key logic of enzymatic bindings [232]. After decades and numerous supporting studies, Dunker et al. [233], and Wright and Dyson [234], suggested a reassessment of the structure – function paradigm and proposed the concept of intrinsically disordered protein (IDP) in 2001, which opened the door for a new and rapidly developing field of protein and proteomic research.

Unlike the ordered proteins, which have defined structure as conventionally believed, intrinsically disordered proteins comprise polypeptide segments that lack of well-defined tertiary structure and can be specified by

- compositional amino acid bias,
- low number of bulky hydrophobic amino acids,
- high number of charged hydrophilic amino acids [235].

Those segments are called intrinsically disorder regions (IDRs), that cannot form a stable coherent hydrophobic core, thus dynamically vary their conformations. Among the human protein-coding genes, nearly 44% are predicted to contain IDRs of more than 30 amino acids [236]. It is now agreed that IDRs are also strongly involved in specifying various functions of proteins, in addition to the functions mediated by ordered regions [237]. Those includes:

- Serving as predisposed sites for post-translational modifications (PTMs), which subsequently code or decode certain interactions to other proteins, leading to specific functions in variable cellular context.
- Uncovering short linear motifs (SLiMs) or molecular recognition features (MoRFs), which serve as docking sites for large number of binding partners or vary the types of binding partners. This function is also known as “protein interaction networks hub”.
- Switching its own structure upon binding to different interaction partners. IDRs can become “ordered” upon binding to certain partners, a process known as coupled folding and binding, thus directly alter downstream function, or even change other parts of their own structures.
- Fine-tuning protein half-life by changing efficiency of proteasomal degradation.

Of note, one of the biophysical characteristics of IDRs includes high specificity but low affinity bindings to partners, allowing transient effects that can be rapidly and spontaneously dissociated [235]. Other biophysical features are heat stability and capability to enable phase transitions by forming aggregations to maintain functional assembles [238]. As such, IDRs containing proteins are often controlled tightly to maintain balance level of expression and abundance, thus maintain also solubility and interaction accuracy at all time in cells [237].

Impaired functions of IDRs-containing proteins are found in many diseases, for example, cancer, neurodegeneration, diabetes, cardiovascular diseases [239]. The well-studied p53 tumor suppressor contains a structured DNA-binding domain and intrinsically disordered N- and C-terminals [240], [241]. As the central regulator of many cellular processes (namely, cell cycle progression, DNA-damage response, cellular stress response and apoptosis induction…), p53 interacts with a large amount of transcription factors, activators and inhibitors to carry out its signal induction [242], [243]. Among those, 70% of interactions are found taking place at the IDRs of p53 molecules, with 86-90% of which are biased toward PTM sites [244]. These studies underline the role of p53 IDRs in modulating its wide range of functions; loss of p53 function through mutations or other mechanism, especially at IDRs, gravely contribute to malignant transformation [245]. Similar to p53, BRCA1 (Breast cancer type 1 susceptibility protein 1) contains a large central IDR that harbor both DNA binding sites and protein interaction sites, mediating its anti-tumor activities [246]. Majority of functioning interactors

of BRCA1 are at this IDR including DNA-damage response sensors/proteins and signal transducers; impairment of which directly lead to cancer [247], [248]. Tau is a protein family associated with neurodegenerative Alzheimer's disease and is a disordered protein [249][250]. Hyperphosphorylation of Tau is found in many pathological conditions. *In vitro* hyperphosphorylation of Tau changes its structure from normally unfolded to a semi-folded form, causes aggregation of Tau in neuronal cells [251], [252]. Since many of major transcription factors, signal transducers, cellular sensors/detectors contain IDRs or be IDPs, it is expectable that disruption at IDRs, which strongly abolish or redirect protein functions, are associated with diseases [239]. PTMs, as important functionality modulator of the IDRs, will be focused on in the next paragraphs.

To conclude, IDRs multiply the plethora of binding partners of a protein, hence diversify its functions in accordance with various signals, by triggering different pathways and ultimately, lead to versatile cellular responses. All different cascades of events take steps by only a single sequence of peptides. In plain texts, as described by Vladimir N. Uversky, IDRs/IDPs are "interaction professionals", who "like to move it, move it" [253]. Understanding the functional spectrums of proteins in various contexts and in compliance with IDRs will be profoundly beneficial for biological understanding as well as diseases targeted treatments.

1.3.2. Post-translational modifications contribute to intrinsically disordered structures

The diversity of mammalian proteome is vastly expanded by post-translational modifications, in addition to mRNA splicing [254]. PTMs provide structural, biophysical and functional diversity to one or more positions on a protein by many ways: addition of chemical groups (e.g. methylation, phosphorylation, glycosylation), attachment of peptides/proteins (e.g. ubiquitination, SUMOylation), chemically modification and spatial distribution of amino acids (e.g. oxidation, deamidation), proteolytic cleavage [255]. As a result, a protein functions according not only to its amino acid sequence but rather flexibly changes its effect by PTMs identity. Based on site of their effects, PTMs can also be grouped into two classes: PTMs at structured regions or proteins, and PTMs at IDRs or IDPs [256]. PTMs at structured regions/proteins are often crucial for structure stabilization, catalytic function, or enzymatic activities; those include oxidation, formylation, protein splicing. On the other hand, many

PTM sites are found at overlapping regions with IDRs [257]. In comparison to structured regions, the IDRs shows advantages, those are: ease of transient interacting (low affinity, high specificity) with catalytic sockets of the modifiers/enzymes; and ease of access to and recognition of the post-translationally modified IDRs by effectors that trigger downstream response [238]. PTMs at IDRs/IDPs include phosphorylation, methylation, acetylation, carboxylation, glycosylation, and many others [256].

Once integrated, PTMs diversify structural and functional properties of the proteins by: (i) altering primary structure by changing hydrophobicity, steric nature, and electrostatics, (ii) stabilizing or destabilizing, (iii) inducing local and long-range structural changes by enhancing/reducing spatial distance between motifs or binding partners, leading to assembly/disassembly of protein complexes and may also lead to order-to-disorder/disorder-to-order transitions [258], [259]. Although phosphorylation has been regarded as the most common type of PTMs in eukaryote, we focus on methylation as the topic of interest.

Methylation of proteins got the first insight in 1971, however, due to lack of evidence of its biological function, the topic was idle for decades. Only until the 1990s, the discovery and functional study of PRMT1 (protein arginine N-methyltransferase 1) ignited an explosion of new findings on this topic [260]. Methylation is a PTM which adds methyl group(s) $-CH_3$ to side chains or C-termini of amino acids at the expense of one hydrogen atom for each methyl group [261]. Taking away a hydrogen equals to a loss of capacity to form hydrogen bonds, which lead to changes of structural formation, increased hydrophobicity and subsequent gain or loss of interaction preferences [261]. The universal methyl donor in cells is S-adenosyl-L-Methionine (SAM or AdoMet), which is added directly to amino acids via the catalyst by methyltransferase enzymes [262]. Methylation can take place on arginine, lysine, histidine, proline and carboxyl groups, however, lysine and arginine methylation are more extensively investigated [263][264]:

- Lysine (K) methylation: is catalyzed by lysine methyltransferases. Almost all of the lysine methyltransferases contain a conserved SET domain (comprises three *Drosophila* proteins Su(var)3-9, Enhancer-of-zeste and Trithorax). Up to three methyl groups can be added to a lysine residue to for mono-, di- or tri-methylated. The classic example of lysine methylation (Kme) is its major involvement in “histone

code" (patterns of PTMs on histone tails that define gene expression by regulating protein recognitions). Methylation on the histone 3 (H3) tail can take place at K4, 9, 14, 27, 36, 79, while methylation on histone 4 (H4) occurs at K20, 59. At each residue, methylation status causes specific outcome, for example: H3K4me marks activation of euchromatic gene at the site, meanwhile, H3K9me or H3K27me marks repression of specific euchromatic genes and formation of heterochromatin.

- Arginine (R) methylation: is a common modification in mammals proteome, as common as phosphorylation and ubiquitination [265]. R-methylation catalyzed by the family of protein arginine methyltransferase (PRMTs). Methylation of arginine falls into each of these three categories: (i) monomethylarginine, MMA, catalyzed by PRMT7 and is intermediate product of other PRMTs; (ii) asymmetric dimethylarginine, ADMA, catalyzed by PRMT1-4, PRMT6 and PRMT8 (grouped into type-I PRMTs); and (iii) symmetric dimethylarginine, SDMA, catalyzed by PRMT5 and PRMT9 (type-II PRMTs). Similar to lysin methylation, arginine methylation on histone regulates specific gene expression: gene activation marks include H4R3me2a/s (histone H4 arginine R3 asymmetric- or symmetric dimethylation), H3R17me2a, H3R26me2a; gene repressive marks include H3R2me2a, H3R8me2a/s, H4R3me2s. Beside transcription regulation, arginine methylation plays central role in many biological processes, including mRNA-splicing, cytoplasmic shuttling, growth factor-mediated signal transduction, and DNA repair [261][266].

Within the scope of this study, we will focus on arginine methylation beside other types of PTMs and beside methylation of other amino acids. More specifically, important function of R-methylation in hematopoietic lineage maturation in normal and pathological contexts will be briefly described below.

Establishment and maintenance of hematopoietic lineages are strongly influenced by arginine methylation and PRMTs activities, therefore, it is also expected that malfunction of PRMTs leads to hematopoietic maladies [267]. Conditional knockout of PRMT5 in hematopoietic stem cells (HSCs) or inhibition of PRMT5 results in short-term promotion of HSC commitments but also disrupts cell cycle progression of these committed progenitors, suggesting the role of PRMT5 in quiescence of HSCs [268]. Enhanced expression of PRMT5 is

observed in various cancer cell lines and patient samples, including mantle cell lymphoma (MCL) and chronic B-lymphocytic leukemia, which causes hypermethylation at histones H3R8 and H4R3. Increased H3R8me2s and H4R3me2s suppress expression of PRMT5 target tumor suppressor genes, thus promote tumor growth [269], [270]. PRMT1 regulates R-methylation on RUNX1, a major transcription factor known to be involved in myeloid differentiation and development of lymphocytes. R-methylated RUNX1 is found binding to the promoter of CD41 (expressed in primitive multipotent progenitors) and PU.1 (myeloid differentiation regulator), and activate their expression, while knockdown of PRMT1 using shRNA greatly reduces level of these target genes [271]. Abrogating methylation at two arginine residues of RUNX1 by lysine substitutions shows decreased and defective population of $CD3^+$ and $CD4^+$ T-cell in peripheral lymphoid organs [272]. Elevated expression of PRMT1 is often found in hematopoietic malignancies, such as AML, ALL and lymphoma [273]–[276]. AML-ETO fusion protein, found in 10% of AML cases, is R-methylated by PRMT1 and also recruits PRMT1 to co-bind target genes' promoters to enrich H4R3me, thus activates transcriptions. Similarly, in mixed-lineage leukemia MLL, MLL-EEN fusion protein and PRMT1 increase H4R3me marks at the promoter of HoxA9 (Homeobox protein A9), a leukemic transformation maintaining transcription factor which is associated with poor prognosis [277]. In AML, PRMT1 catalyzed R-methylation of FLT3-ITD tandem duplicated protein, prompts leukemic maintaining in a methylation dependent fashion; while R-to-K (arginine to lysine) mutation, which disrupts R-methylation, strongly reduced $cKit^+$ leukemic cells and enhanced survival in mouse model [278].

Various PRMTs inhibitor have been developed with therapeutic efficacy. PRMT1 inhibitor AMI-408 efficiently suppresses the di-methylation of H4R3 in MLL-GAS7 cell lines, thus reduces self-renewability of leukemic cells, reduces disease penetrance and enhance survival of tumor-transplanted mice [279], [280]. PRMT4, also known as CARM1, can be selectively inhibited by compound EPZ025654 [281]. Treatment of primary AML cells from patients shows reduced cell growth and reduced colony size, possibly via inhibition of mRNA stability, induction of p53 responses and downregulated E2F family [282]. Several inhibitors of PRMT5 are currently in clinical trials phase I for treatment of AML, non-Hogdkin's lymphoma and advanced solid tumors [283]. Comprehensive lists of PRMTs inhibitors and their progression in clinical trials are available in cited reviews [261], [283].

To sum up, the human proteomes retains an impressive complexity and diversity, a considerable part of which are contributed by hundred types of PTMs and nearly a million sites of modified peptides motifs [259]. Computationally calculation reveals approximately 5% of disease-causing mutations affect PTM sites and this is, indeed, observed in many types of malignancies beside leukemia [284]–[287]. Among that, arginine methylation has proved itself to be a crucial regulator of biological development, cancerous transformation and fortunately, a druggable entity that promises great medicinal value.

1.3.3. Post-translational modifications on C/EBPα

C/EBP family manifests features of IDPs by containing various IDRs, PTM sites, SLiMs and MoRF [288][289]. It was suggested that during transdifferentiation, the cell type specification of B cells progenitor was altered by not only structure but also the post-translational modification of C/EBPβ. In detailed, granulocytic-transdifferentiation ability of C/EBPβ varies based on the methylation status of R3, K39, K156, E158 [290]. Recently, a novel method, PRISMA (Protein Interaction Screen On Peptide Matrices), was developed to screen protein interactions on peptide matrices, which reveals hundreds of C/EBPβ interactors, including PTM-specific ones [289]. Subsequently, an interaction between C/EBPβ and TLE3 (Transducin-like enhancer of Split 3, involved in Notch signaling pathway to regulate cell fate determination) was experimentally proven to be regulated by Arginine methylation.

Similarly, C/EBPα shows dependency on several PTMs (**Table 2**). Function of C/EBPα in granulopoiesis depends of acetylation status of K298, K302 [198]. C/EBPα normally recruits HDAC3 (Histone deacetylate 3) to promoter region of the cell cycle regulator *CyclinD1* and suppresses its expression. When being methylated at R35, R156, R165 by PRMT1, C/EBPα-HDAC3 complex is dissociated, thus releases transcription activity of *CyclinD1* and promotes growth of breast cancer cells [291]. A recent study using combination of PRISMA and BioID (Biotin proximity labeling identification) revealed important insights into PTM-dependent interactome of both C/EBPα p42 and p30 [292]. R-methylation dependent interactions are identified between C/EBPα and many binding partners, including SWI-SNF component SMARCE1 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1), SWI-SNF subunits ARID1A, ARIB1B, ARID2 (AT-rich interactive domain-containing protein 1A/1B/2), ubiquitin ligase TRIM33 (Tripartite motif-containing 33),

DREAM complex component LIN9, LIN37, among others. Substitution of R142 on C/EBPα p30 by methylation-mimicking leucine enhanced interaction with the SWI-SNF complex members. Furthermore, this study also indicates distinct binding signatures between p42 and p30, with erythroid transcription factor GATA1 preferentially interacts with p30. Some p30-specific interactors are found important to predict AML survival, including TFAP4, BCL11A, GATA1.

Altogether, these findings confirm the intrinsically disordered identity of C/EBPs, especially C/EBPα, and further discover interaction hotspots on structural regions on C/EBPα. As the interaction pattern of C/EBPα p30 is distinctive and R-methylation dependent, we anticipated that p30 activity in both normal and leukemic context is strongly directed by this type of PTM.

Table 2. Post-translational modifications discovered on C/EBPα

Residues	Type	Consequences of decorated residues	Ref.
R35, R156, R165	Methylation	Loss of CyclinD1 suppression, promoted proliferation of breast cancer cell lines	[291]
K298, K302	Acetylation	Inhibited granulopoiesis	[198]
S21	Phosphorylation	Promoted monopoiesis, inhibited granulopoiesis	[182]
K159	SUMOylation	Loss of cell cycle arrest, altered protein-protein interactive partners.	[293]
S193	Phosphorylation	Blocked proliferation of hepatocytes.	[294]
S248	Phosphorylation	Enhanced expression of G-CSF, promoted granulopoiesis	[295]

1.4. C/EBP-induced lymphoid-to-myeloid transdifferentiation system

The lineage redirecting function of C/EBPα was introduced early on by the group of Thomas Graf. By overexpressing C/EBPα, in combination with PU.1, B, T cells and even fibroblasts can be redirected into macrophages [50], [296], [297]. The C/EBPα-induced transdifferentiation from B cells to inflammatory macrophages is a robust process with 100% efficiency within 2 cell-generation (3-5 days) [144]. Interestingly, the authors found no apparent expression of HSC markers, indicating that the lineage conversion is direct without de-differentiation into an intermediate multipotent progenitor. Later, an estrogen inducible *CEBPA* expressing system was developed and used in pre-B cell lines to study various aspects of hematopoiesis systemically and functionally, for instance:

- *Study of histone modification* using pre-B cell line HAFTL1 (Ha-ras transformed fetal liver cell line) found histone deacetylase HDAC7, which normally suppresses non-B cell genes, was downregulated during lymphoid-myeloid lineage conversion [298].
- *Study of gene expression and transcription factors activity* using similar approach found a C/EBPα target: methylcytosine hydroxylase Tet2 as a de-repressor of myeloid genes, which hydroxyl-methylates target genes' promoter [299].
- *Study of enhancer establishment* using *CEBPA*-ER fusion in pre-B cell lines revealed coordination of C/EBPα and PU.1 during myeloid enhancer binding and pioneering activity of C/EBPα on a subset of myeloid enhancers [300].
- Other findings of genome topology, cell cycle progression and non-coding RNA expression were uncovered using the transdifferentiation system [148][301], [302].

One of the advantages of the C/EBPα induced transdifferentiation system is that it can be a powerful tool to examine hematopoiesis in larger scale, especially in proteomics and biochemistry studies that usually require large amount of material, in comparison to using human- or animal-derived primary materials. The system also allows more flexibility when using genetically modified B cells as starting materials, besides using genetically engineered animal models, which are more laborious to generate. Above all, we appreciate that the transdifferentiation system is appropriate to investigate the biology of C/EBPα itself and indeed, employed it for this study.

1.5. Aims of the thesis

We envisage dysregulation of C/EBPα isoforms function, in particular, post-translational modification of its p30 isoform, is a driving force of AML and of lineage conversion.

Despite tremendous amount of information has been established on C/EBPα p42, its truncated isoform p30 is still not receiving adequate attention. More extensive studies on the nature of C/EBPα p30 functions promises great benefit because:

- Unlike the conventional viewing of truncated isoform of a protein as a dominant-negative form and often antagonizes the full-length form, p30 proves itself as an active form which regulates its own transcriptional program; thus, investigation of

p30 function brings more insights into how C/EBPα isoforms cooperatively regulate developmental processes.

- Although carrying the intermediate risk and favorable outcome value, *CEBPA*-mutated AML treatment remains untargeted and needs more strategic options coming from basic research. Besides, the risk of developing AML as a secondary leukemia, although still lack of evidence of C/EBPα involvement, is a valid concern. Study of p30 function complements our understanding of AML etiology and may potentially disclose novel therapeutic targets.
- Arginine methylation evidently directs activity of C/EBPα isoforms, nevertheless, biological implications and supporting mechanism are still missing. Even though it is now possible to target PRMTs to alter R-methylation, a mechanistic description of R-methylated/unmethylated p30 functioning may allow us to redirect p30 function more accurately.

We therefore considered exploring the effect of C/EBPα p30 arginine methylation on its instructive function during hematopoietic lineage commitment and transformation. The recently developed C/EBP-induced LMT system was used in this study as an amenable tool to study many aspects of cell-fate decisions, from mechanism to specification of cell type outcome. These matters should be assigned:

(i) Evaluation of R-methylation dependent p30 differentiation potential using lymphoid to myeloid transdifferentiation system.

(ii) Evaluation of R-methylation dependent p30 proliferation potential in bone-marrow derived hematopoietic progenitors.

(iii) Investigation of downstream mechanism directed by p30 R-methylation.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Mice

Mice used in this study were maintained under pathogen-free condition at the Animal Core Facility of the Max Delbrück Center for Molecular Medicine (MDC), Berlin, Germany. All mice were treated according to recommendations of good animal handling practice. The experiments followed institutional guidelines and German Federal Animal Protection Act.

2.1.2. Cell culture

Reagent	Supplier
Cryostor CS10	STEMCELL Technology
DMEM GlutaMAX-I (high Glucose)	Gibco
Fetal Bovine Serum	Gibco
Gentamycin	Sigma
HEPES Buffer Solution (1 M)	Gibco
IMDM GlutaMAX-I	Gibco
Methocult M3434	STEMCELL Technology
mIL-3	STEMCELL Technology
mIL-6	STEMCELL Technology
mSCF	STEMCELL Technology
Penicillin/Streptomycin (100X)	PAA
Polyethylenimine hrdrochroride	Polysciences, Inc.
Trypan Blue	Sigma
β-Mercaptoethanol (50 mM)	Gibco

2.1.3. Buffer

FACs Buffer	
FCS	2%
EDTA	2 mM
in PBS, filtered	

LB medium	
Tryptone	1 % (w/v)
Yeast extract	0.5 % (w/v)
NaCl	0.5 % (v/v)
in dH_2O	

Cell Lysis Buffer	
NaCl	150 mM
Tris-HCl	50 mM
Nonidet P-40	1 % (v/v)
Glycerol	10 % (v/v)
EDTA	1 mM
in dH_2O	

Completed protease inhibitor cocktail is added freshly

4x protein loading buffer	
Tris-HCl	200 mM
SDS	1 % (w/v)
Glycerol	40 % (v/v)
DTT	400 mM
Bromophenol blue	0.4 % (w/v)

Ponceau S Solution	
Ponceau S	0.1 % (w/v)
Acetic acid	5 % (v/v)
in dH_2O	

TBS-T	
Tris-HCl	50 mM
NaCl	150 mM
Tween 20	0.1 % (v/v)
in dH_2O	

TAE Buffer	
Tris base	80 mM
EDTA	1 mM
Acetic acid	0.11 % (v/v)
in dH_2O	

Western Blot running buffer	
Tris base	25 mM
Glycine	200 mM
SDS	0.1 % (w/v)
in dH_2O	

PBS 10X	
NaCl	1.37 mM
KCl	25 mM
$Na_2HPO_4.H_2O$	100 mM
KH_2PO_4	18 mM
in dH_2O, pH 7.4	

Red blood cell lysis buffer	
NH4Cl	150 mM
KHCO3	10 mM
EDTA	0.1 mM
in dH2O (adjust pH 7.4, filtered)	

2.1.4. Flow cytometry antibodies

TruStain FcX (anti-mouse CD16/32, BioLegend) was used for Fc block in all panels, unless other fluorophore conjugated CD16/32 antibodies were used.

Lymphoid-myeloid transdifferentiation panel			
Specificity	**Conjugation**	**Clone**	**Manufacturer**
CD115 (M-CSFR)	APC	AFS98	eBioscience
CD11b (Mac-1)	PE	M1/70	BD Pharmingen
CD19	PE-Cy7	1D3	eBioscience
Ly6-G	Brilliant Violet 421	1A8	BioLegend

cKit enrichment			
Specificity	**Conjugation**	**Clone**	**Manufacturer**
CD117 (cKit)	Biotin	2B8	BioLegend
Anti-biotin microbeads			Miltenyi Biotec

$Lin^{neg}cKit^{+}Sca\text{-}1^{+}$ (LSK) sorting panel			
Specificity	**Conjugation**	**Clone**	**Manufacturer**
CD117 (cKit)	PE-Cy7	2B8	eBioscience
Streptavidin	PE-Cy7	-	BioLegend
Ly-6A/E (Sca-1)	PE	E13-161.7	BD Pharmingen
CD3e	APC	145-2C11	BD Pharmingen
Ly-6C	APC	RB6-8C5	BD Pharmingen
CD11b	APC	M1/70	BD Pharmingen
Ter119	APC	TER-119	BioLegend
B220	APC	RA3-6B2	BD Pharmingen
CD11c	APC	HL3	BD Pharmingen
CD5	APC	53-7.3	eBioscience
CD115	APC	AFS98	BioLegend

Colony characterization panel			
Specificity	**Conjugation**	**Clone**	**Manufacturer**
CD16/32	PE	2.4G2	BD Pharmingen
CD117 (cKit)	PE-Cy7	2B8	eBioscience
CD34	AlexaFluor 647	SA376A4	BioLegend
Ly-6G	AlexaFluor 700	1A8	BioLegend
Ly-6C	APC-Cy7	HK1.4	BioLegend
F4/80	Pacific Blue	BM8	BioLegend
CD115	Brilliant Violet 605	AFS98	BioLegend
CD11b	Brilliant Violet 711	M1/70	BioLegend
CD11c	PerCP-Cy5.5	N418	BioLegend

2.1.5. Reagent and consumables

Name	**Supplier**
2.0 ml Polyprolylene Microcentrifuge Tube	BrandTech Scientific
3 cc Syringes (use with methylcellulose-based medium)	STEMCELL Technology
50 ml Conical Centrifuge Tube	BD
Agarose pure	Thermo Fisher Scientific
Blunt-end Needles 16 Gauge	STEMCELL Technology
Bovine Serum Albumin Standards	Thermo Fisher Scientific
Bradford Reagent	Sigma
Bromophenol blue	Roth
Cell Proliferation Reagent WST-1	Roche Diagnostics
Cell Strainer 100 µm Nylon	BD

Complete Protease Inhibitor Cocktail	Roche Diagnostics
Complete protease inhibitor mix	Roche Diagnostics
Criterion TGX Precast Gels 4-15%	BioRad
Cutfix Stainless Steel Scalpel #21	B.Braun
DNA Gel Loading Dye (6X)	Thermo Fisher Scientific
DNA LoBind Tubes 1.5ml	Eppendorf
DNase I	Roche Diagnostics
dNTP Set	GE Healthcare
DTT	Sigma
Dynabeads™ mRNA DIRECT™ Purification Kit	Thermo Fisher Scientific
Ethidium bromide solution 1 %	Roth
Falcon 1.5 ml Polypropylene Microcentrifuge Tube	BrandTech Scientific
Falcon 15 ml Conical Centrifuge Tube	BD
Falcon Round-Bottom Polystyrene Tubes	BD
Falcon Round-Bottom Polystyrene Tubes with cell strainer cap	BD
Falcon Round-Bottom Polystyrene Tubes with snap cap	BD
GeneRuler 1 kb DNA Ladder	Thermo Fisher Scientific
GeneRuler 100 bp DNA Ladder	Thermo Fisher Scientific
Glass pasteur pipettes	VWR
Glycerol	Roth
Hydrochloric acid	Merck
Immobilon Chemiluminescence HRP Substrate	Millipore
Invisorb Spin DNA Extraction Kit	Stratec
Invisorb Spin Plasmid Mini Two Kit	Stratec
LS Columns	Miltenyi Biotec
May-Grünwald solution	Merck
Microscope Slides Menzel-Gläser Superfrost	Thermo Fisher Scientific
MidiMACS™ Separator	Miltenyi Biotec
Millex-GV Filter Unit 0.22µm PVDF	Milipore
NucleoBond Xtra Midi Kit	Macherey-Nagel
PageRuler Plus Prestained Protein Ladder	Thermo Fisher Scientific
Petri Dish 94/16 mm, PS	Greiner Bio-One
Pfu DNA-polymerase	Thermo Fisher Scientific
Pierce Reagent	Thermo Fisher Scientific
Ponceau S	Serva
Protein LoBind *Tubes* 1.5ml	Eppendorf
QuadroMACS™ Separator	Miltenyi Biotec
Qubit RNA HS Assay Kit	Thermo Fisher Scientific
Quick-RNA kit	Zymo Research
Rescriction enzymes and buffers	New England Biolabs

RevertAid First Strand Synthesis Kit	Thermo Fisher Scientific
RNeasy Mini Kit	Qiagen
Roti®-Histokitt II	Roth
Shrimp Alkalin Phosphatase	Roche
Skim Milk Powder	Fluka
SmartDish	STEMCELL Technology
Sterican Hypodermic Needle, 24 G	B.Braun
Super RX-N X-ray film	Fuji
Syringe, 1 ml luer	B.Braun
Syringe, 10 ml luer	B.Braun
Syringe, 5 ml luer	B.Braun
T4 DNA Ligase	New England Biolabs
Taq DNA Polymerase	Thermo Fisher Scientific
TC Dish 100, Standard	Sarstedt
TC Dish 150, Standard	Sarstedt
TC Flask T25, Standard, Ventilated Cap	Sarstedt
TC Flask T75, Standard, Ventilated Cap TC	Sarstedt
TC Plate 12 Well, Standard	Sarstedt
TC Plate 24 Well, Standard	Sarstedt
TC Plate 6 Well, Standard	Sarstedt
TC Plate 96 Well, Standard	Sarstedt
TipOne Pipette Tips 10/20µl XL, Graduated	StarLab
TipOne Pipette Tips 1000µl, Graduated	StarLab
TipOne Pipette Tips 200µl, Graduated	StarLab
Trans-Blot Turbo Midi-size 0.2 µm Nitrocellulose	BioRad
UltraPure Agarose	Thermo Fisher Scientific
WesternBright ECL HRP substrate	Advansta
Whatman Syringe Filter FP, 0.2 µm	GE Healthcare
Whatman Syringe Filter FP, 0.45 µm	GE Healthcare
Wright-Giemsa staining solution	Merck

2.1.6. Equipment

Name	Supplier
510 Precision Balance	Kern
7120 Hematology Aerospray Slide Stainer Centrifuge	Wescor
BD FACSAria Fusion Cell Sorter	BD Biosciences
BD LSR II Flow Cytometer	BD Biosciences
BDAdigital gel documentation system	Biometra
Bioanalyzer 2100	Agilent
BP210D Analytical Balance	Satorius

CanoScan 9000F Mark II	Canon
Citizen CL-S631 label printer	Citizen
CKX41 Inverted Phase Contrast Microscope	Olympus
Criterion Vertical Electrophoresis Cell	BioRad
Cryotherm Biosafe 420 SC β	Cryotherm
Duomax 1030 platform shaker	Heidolph
Easypet pipette	Eppendorf
EVOS FL Auto Imaging System	Thermo Fisher Scientific
GFL 3017 orbital shaker	GFL
GFL Water Bath 1008	GFL
Heraeus Multifuge X3FR	Thermo Fisher Scientific
HU10 Mini-Plus Horizontal gel electrophoresis unit	Scie-Plas
Ice maker	Ziegra
iMark Microplate Absorbance Reader	BioRad
Incubator CB150, CB210c	Binder
Innova40 incubated shaker	New Brunswick Scientific
Laboratory pH Meter 766	Knick
Laminar hood BDK-SK 1500	BDK
LGex 3410 MediLine freezer (-20 °C)	Liebherr
Mastercycler nexus GX2 PCR system	Eppendorf
Microcentrifuge 5417R	Eppendorf
Milli-Q Water Purification System	Milipore
Mr. Frosty Cryo 1°C Freezing Containers	Nalgene
Multipette Plus	Eppendorf
Multitron Standard incubated shaker	Infors HT
Nanodrop 2000c	Thermo Fisher Scientific
NextSeq 500 Sequencing System	Illumina
Optimax 2010 film developer	Protec
QuantStudio 6 Flex Real-Time PCR system	Thermo Fisher Scientific
Qubit 3 Fluorometer	Thermo Fisher Scientific
Reax top vortex mixer	Heidolph
Resarch plus pipettes	Eppendorf
Rotator mixer	Bachofer
Rotilabo Mini-Centrifuge	Roth
Scissors and forceps	KLS Martin
Sonoplus HD70 ultrasonic homogeniser	Bachofer
Thermomixer compact	Eppendorf
Trans-Blot Turbo system	BioRad
Variomag Power Direct magnetic stirrer	Thermo Fisher Scientific
VIP Ultra-low Temperature Freezer (-80 °C)	Sanyo
Wescor Aerospray slide stainer cytocentrifuge	Wescor

2.1.7. Software

Name	Producer
Canon IJ Scan Utility	Canon
etiLabel Professional 3.0.0.68	Etisoft
EVOS FL Cell Imaging System Software V1.4	Thermo Fisher Scientific
ExpressionSuite Software v1.1	Applied Biosystems
FACSDiva™ Software	BD Biosciences
FlowJo V10	FlowJo, LLC
GraphPad Prism	GraphPad Software
Illustrator CC 2017	Adobe
ImageJ	NIH
Mendeley Desktop	Mendeley Ltd
Microsoft Office 365 ProPlus	Microsoft
NanoDrop 2000 Software	Thermo Fisher Scientific
QuantStudio Real-Time PCR Software	Applied Biosystems
SnapGene 4.2.4	GSL Biotech

2.2. Cell biology methods

2.2.1. Cell lines

All cell lines used in this study were culture at 30^{o}C, 5% CO_2 in water-jacketed incubators. Completed medium contained basal cell culture medium (DMEM, IMDM) with 10% fetal bovine serum (FBS), 1% Penicillin/Streptomycin and 10 mM HEPES.

The V-Abl-transformed B cell lines, including the *Cebpa*$^{fl/fl}$*Cebpb*$^{fl/fl}$ cell line (termed B-WT) and *Cebpa*$^{-/-}$*Cebpb*$^{-/-}$ (termed B-dKO), were generated as described in previous study [166]. They were culture in completed DMEM medium, supplemented with 50 µM β-Mercaptoethanol. Bone marrow derived cells were culture in complete IMDM, supplemented with 10 ng/mL IL-3, 10 ng/mL IL-6 and 20 ng/mL mSCF. The Platinum-E (PlatE) cells, a derivation of 293T cell line, stably expressed retroviral *gag*, *pol* and *env* and was used as packaging cell line.

2.2.2. Retroviral transduction and infection

To generate supernatant containing retrovirus particles, PlatE cells were used as packaging cells line. 5x10^6 PlatE cells were seeded in complete DMEM in 10 cm plates and cultured for 3-5 hours until cells became attached. Then, the medium was refreshed before transfection

to assure optimal cell growth. The transfection was performed with 5 µg of desired pMSCV-based constructs in 500 µL of un-supplemented DMEM; in separate tubes, 30 µL of PEI was mixed with 470 µL of un-supplemented DMEM. PEI solution was added to the construct-containing mixture, pipetted thoroughly, and incubated for 15 minutes at room temperature. Transfection mixture was added dropwise to PlatE cells. Medium was refreshed with completed DMEM at 24 hours. The infectious supernatant was harvested at 48 hours and 72 hours, filtered with 45 µM syringe filters, and used freshly. In case of prolonged use, infectious supernatant could be stored in -80^{o}C after snap-freezing in liquid nitrogen. Frozen supernatant was thawed and used one time.

Retroviral infection was performed in well-plate format. Cells were harvested, diluted into 4x10^{5} cells/ml dilution in adequate medium with 2X concentration of supplements (cytokines or β -Mercaptoethanol) and 2X Polybrene (16 µg/mL). For each well of a 24-well plate, 500 µL of cell suspension and 500 µL of retroviral supernatant were added. Plate was centrifuge at 2200 rpm, 37^{o}C for 60 minutes. Medium was change after 24 hours (for cells grew in DMEM) or 5 hours (for cells grew in medium other than DMEM, such as bone marrow derived cells).

2.2.3. Flow cytometric analysis and sorting

Cells were harvested into centrifuge tubes and washed with cold FACs buffer. Harvested cells were incubated with Fc Block reagent (TruStain FcX anti-mouse CD16/32, BioLegend) at 1:200 vol/vol dilution for 10 minutes at room temperature. Cells were then washed twice and proceed to labelling with fluorescence-conjugated antibodies for 30 minutes at 4^{o}C in the dark. After washing, cells were resuspended in FACs buffer containing PI (7-AAD) for discrimination of dead cells. Samples were measured using LSRFortessa analyzer or sorted using FACSAria, FACSAriaII and FACSAriaIII sorters. Measurement was recorded by FACSDiva software and analyzed by Flowjo software.

For sorting, catching tubes were coated with FBS for 30 minutes at 4^{o}C to reduce stickiness of tube wall. Gentamicin was added to buffer and post-sort cell culture medium at final concentration 10 µg /mL. Sorted cells were spun down in gentle cycle (700 rpm for 7 minutes) and resuspend in fresh medium containing Gentamycin.

2.2.4. Isolation of bone marrow cells and LSK cells

Bone marrow derived cells were isolated from femur, tibia, and part of hip joints of 8–12-week-old mice. Two isolation methods were used in this study:

- For experiments in 3.4, prepared bones were cut at the epiphysis. Bone marrow was flushed out with cold PBS using a syringe and a 24-gauge needle.
- To enhance the yield and reduce processing time, bones were crushed in cold PBS under sterile condition.

Cell suspension was filtered through a 70 µM cell strainer before being incubated with red blood cell lysis buffer for 8-10 minutes on ice. Cells were then washed and cultured in cytokine supplemented-IMDM or resuspend in PBS for further processing.

2.2.5. Enrichment of cKit$^+$ cells

cKit$^+$ cells were enriched by MACs separation (magnetic-activated cell sorting). Isolated bone marrow cells were incubated with biotin-conjugated cKit$^+$ antibody for 20 minutes at 4^{o}C. After washing, anti-biotin magnetic beads were added to cell suspension (20 µL beads per 10^7 cells) and incubated for 30 minutes at 4^{o}C. Magnetically labelled cells were washed and resuspend in at most 10^9 cells/mL and passed through equilibrated selecting columns (LS Column, Miltenyi) mounted on MidiMACs or QuadroMACs separator. For each mouse, 2 columns were used to avoid oversaturation on columns, which might lead to loss of targeted cells. Cell-loaded columns were washed 3 times with MACs buffer, then put into a 15 mL centrifuge tube and eluted with 1 mL MACs buffer. Cells were then washed and cultured in cytokine supplemented-IMDM or prepared for further processing.

2.2.6. Colony serial replating assay

Colony serial replating assay was performed on C/EBPα p30 (WT or mutants) infected bone marrow cells. At day 1 post-infection, Lin$^-$cKit$^+$GFP$^+$ cells were sorted, spun down and resuspended 5x10^4 cells/mL in cytokine-supplemented IMDM. For colony formation, semi-solid cytokine supplemented Methocult medium (MC3434, STEMCELL Technology) was used; stock medium was first aliquoted 3.6 mL into tubes. 360 µL of cell suspension was added to 3.6 mL MC3434 aliquot, which made final concentration 5000 cells/mL, mixed well, and let rest for 15 minutes to avoid cells being trapped in air bubbles. In a 6-well meniscus-free dish (SmartDish™, STEMCELL Technology), 1.1 mL of MC3434 cell suspension was added to each well using a 3cc syringe and a blunt-end 16-gauge needle, 3 replicates were seeded for each experiment group. Sterile water was added into the empty spaces between wells to keep humidity and cells were culture under normal conditions. Whole colony forming dishes were

scanned using EVOS™ FL™ auto imaging system (Thermo Fisher Scientific) every 7 days for 4 rounds (day 7, day 14, day 21 and day 28) at 4x magnification. Colonies were counted and classified manually based on morphology (size, shape, tightness).

To replate, after scanning, all 3 replicates were harvested and pooled together. Colonies were well-suspended to obtain single cells suspension. Total cell number was counted, and cells were diluted $5x10^4$ cells/ mL. Cell suspension was added to MC3434 aliquots and seeded as described above. Leftover cells of day-7 plates were subjected to further analysis, including flow cytometric analysis, growth curve analysis, WST-1 assay, and cytospin.

2.2.7. Growth curves

Growth curves were determined by accumulative cell counting. From cells obtained as described in 2.2.6, 10^5 cells were seeded in 1mL cytokine supplemented IMDM into a 24-well plate, 3 replicates were set up for each group. Every 24 hours, cells in each well were well-suspended, 10 µL of cells suspension was drawn and mixed 1:1 with Trypan blue. 10 µL of the mixture were then transferred to a Neubauer chamber and counted manually. Cell concentration was calculated as:

$$\text{Concentration} = \text{number of cells counted} \times 2 \times 10^4 / \text{number of squares}$$

2.2.8. WST-1 assay

WST-1 assay was performed using cells obtained as described in 2.2.6. Cells were diluted 10^5 cells/mL in cytokine supplemented IMDM, 100 µL of cell suspension containing 10^4 cells were added to a microplate (96-well plate, flat bottom). Three replicates were seeded for each group, five plates were set up, with Plate 0 measured on the same day when cells were seeded. To quantify metabolic activity of the cells, which only viable cells had, 10 µL of WST-1 reagent was added to each well (1:10 dilution) and incubated in cell culture incubator. At 30 and 60 minutes after incubating, absorbance of 450 nm wavelength was measured using an iMark microplate absorbance reader (Bio-Rad). Measured OD was normalized against OD of blank control wells, which contained equivalent cell culture medium. Measurement was obtained every 24 hours for 4 rounds.

2.2.9. Cytospin and May-Grünwald/Giemsa staining

2×10^5 cells were centrifuged using a Wescor Aerospray slide stainer cytocentrifuge at 500 rpm for 5 minutes, which spun cells into glass slides. Cells on slide were let dry completely

before fixing with Methanol for 5 minutes and air dried. Slides were immersed into May-Grünwald solution for 5 minutes, then washed by dipping 2-3 times in ddH_2O. Slides were then immediately immersed in freshly prepared Giemsa solution and incubate for 35-45 minutes until desired vividity was obtained. Stained slides were rinsed under running water for 3 minutes and air dried. One drop of mounting solution (Roti-Histokitt II) was added onto stained area and slides were covered by coverslips. Stained cells were observed under the microscope and captured at multicolor mode (EVOS, Invitrogen).

2.3. Molecular biology methods

2.3.1. Expression constructs

All C/EBPs in this study were expressed from the same construct backbone: the pMSCV retroviral vector system. The original vector was used as empty-vector control, which contained IRES_EGFP as reporter for successfully transduced cells, thus bear the term MIEG (pMSCV_IRES_EGFP). The 3xFLAG tagged C/EBPα variants were cloned by Dr. Elisabeth Kowen-Leutz (AG Achim Leutz, Max Delbrück Center for Molecular Medicine, Berlin) (triple-mutants C/EBPα p30 3A, 3L, 3K and single-mutants R154A/L/K) and the author (single-mutants R140A/L/K and R147A/L/K).

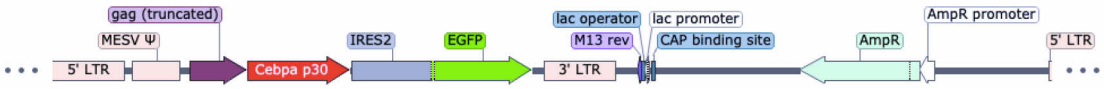


Figure 3. Schematic representation of expression constructs used in this study.
*pMIEG-based vector construct used in this study for the ectopic expression of C/EBPs (including C/EBP**α** p42, p30, p30 mutants and C/EBP**β** LAP1).*

Point-mutated rat C/EBPα p42 construct was commercially synthesized in pUC57 backbone (MWG-Biotech), which was then subcloned into pBluescript vector using HindIII-XbaI ligation. A 3xFLAG tag sequence was extracted from a pCDNA3-based C/EBPα-p42-WT_3xFLAG tag construct (provided by Dr. Elisabeth Kowen-Leutz) using MluI-XbaI digestion, and subsequently ligated into pBluescript vector containing mutated p42 to obtain the final construct pBluescript_p42-mutant_3xFLAG. This plasmid was used as template for polymerase chain reaction (PCR) with primers containing restriction sites of EcoR1 and XhoI to generate EcoRI-XhoI-ended p30 and p42 fragments. These fragments were digested by EcoRI-XhoI to create sticky ends before being ligated into an EcoRI-XhoI-digested MIEG

vector. In the end, the process produced pMSCV_p42-mutant_3xFLAG_IRES_EGFP and pMSCV_p30-mutant_3xFLAG_IRES_EGFP vectors. Expression of C/EBPα p42 and p30 from these constructs were validated by transduction into HEK-293 cells and following Western Blot analysis.

2.3.2. General plasmid cloning methods

Major steps of subcloning a vector includes digestion and ligation reactions. Digestion reactions were set up with 1-3 µg DNA, 1 µL of each restriction enzyme, 1X of equivalent buffer (NEB) in a total 50 µL reaction volume. Incubation was at 37°C in 1 hour. Additional 30 minutes incubation with shrimp alkaline phosphatase (SAP) at 37°C was required to dephosphorylate, thus avoid self-ligation of backbone fragment. Digested fragments were separated on agarose gel by electrophoresis; fragments of interest were excised and purified using Invisorb Spin DNA Extraction Kit (Stratec). Ligation reactions were set up with 3:1 ratio of insert:backbone fragments, 1 µL of T4 DNA ligase, equivalent buffer in a total 20 µL reaction volume. Incubation was at room temperature in 2 hours. For transformation, 3 µL of ligation reaction were mixed with 30 µL competent E.Coli strain TOP10. The mixture was incubated on ice for 20 minutes before being heat-shocked at 42°C for 70 seconds, then cooled down on ice for 2 minutes. Heat-shocked bacteria were recovered by adding 900 µL LB medium to the mixture and incubated in a bacteria shaker at 37°C in 1 hour. Afterward, 100 µL of mixture were spread on LB agar plate containing Ampicillin for selection of successfully transformed cells and incubated at 37°C overnight. Colonies were picked next day and expanded in 2 mL LB medium containing Ampicillin in a bacterial shaker at 37 °C for 16-18 hours. Extraction of plasmid DNA was performed next day using Invisorb Spin Plasmid Mini Two Kit (Stratec). Test-digestion was used to screen positive clones, 3-5 positive clones were later verified by Sanger sequencing. For up-scaling, 100 µL of a pre-cultured bacteria were inoculated in 100 mL of LB medium containing Ampicillin at 37°C overnight. Plasmid DNA was extracted using NucleoBond Xtra Midi Kit (MACHEREY-NAGEL).

2.3.3. Polymerase chain reaction (PCR)

PCR was performed to amplify single mutated p42 or p30 fragments. A reaction was set up with 400 ng DNA template, 1 µL of each primer, 1 µL dNTP, 0.5 µL Pfu DNA-polymerase with equivalent buffer, 2.5 µL DMSO in the total 50 µL reaction volume. PCR was performed using a Mastercycler Nexus GX2 (Eppendorf) with the following program: 98°C initiation (10

minutes); 15-20 cycles of 95°C denaturation (30 seconds), 60°C annealing (30 seconds), 72°C extension (30 seconds); a final extension step at 72°C (5 minutes). Final products were separated on agarose gel by electrophoresis, purified and digested before subjected to ligation.

To amplify mutated p42 fragment with EcoRI/XhoI sites at ends, these primers were used:

p42 FW (628) 5'-gcgaagcttgaattcgccatggagtcggccgacttctac-3'

p42 RV (626) 5'-ccgctcgagctagagcttgtcatcgtcatccttgtaatc-3'

To amplify mutated p30 fragment with EcoRI/XhoI sites at ends, these primers were used:

p30 FW (627) 5'-gcgaagcttgaattcgccatggcggcgggggcgcacgga-3'

p30 FW (626) 5'-ccgctcgagctagagcttgtcatcgtcatccttgtaatc-3'

2.3.4. Western Blot

Western Blot was performed to ensure stable expression of the C/EBP retroviral constructs. HEK-293 cells were harvested at 48h after transduction with C/EBP constructs. To prepare total protein lysates, snap-frozen cell pellets were incubated with cell lysis buffer for 20 minutes on ice before being fragmented by sonicating for 10 pulses at 20 kHz sonic waves. Total lysates were centrifuged 11000 rpm at 4°C in 20 minutes to remove insoluble components and debris; pellets were discarded afterward. Protein concentration was measured using Pierce Reagent (Thermo Scientific). 10-30 µg of protein were mixed with 4x protein loading buffer, boiled at 96°C for 2 minutes and loaded into pre-cast 4-15% SDS-polyacrylamide gel. Electrophoresis was run in running buffer at 110 V until bromophenol blue dye reached the end of the gel. Proteins were transferred onto a nitrocellulose membrane using Trans-Blot Turbo Midi System (BioRad). The membrane was stained with Ponceau S to confirm successful transfer and proceed to blocking in 5 % skim milk (in TBS-T) for 1 hour. Detection of antigen of interest was done by incubating the membrane with specific antibodies against C/EBPα (14AA, Santa Cruz Biotechnology) at 4°C overnight. The membrane was washed three times with TBS-T for 5 minutes before incubated with 5 % skim milk (in TBS-T) containing appropriate horseradish peroxidase-tagged secondary antibodies against mouse or rabbit IgG (GE healthcare) for 1 hour. After washing three times with TBS-T for 5 minutes, freshly prepared chemiluminescence reaction solution (Milipore) was applied onto the surface of the membrane and gently shook for 2 minutes. The membrane was exposed to a Super RX-N X-ray film (Fuji) and developed using the OPTIMAX 2010 (Protec).

2.3.5. Bulk RNA-sequencing

Cells at day 4 post-infection was subjected to GFP$^+$ sorting (**Figure 9**). Cell pellet was immediately resuspended in RNA lysis buffer (RNeasy Kit, Qiagen) and store in -80^{o}C condition. Samples from 4 independent biological replicates were processed at once using RNAeasy Kit (Qiagen). The concentration of extracted total RNA was measured using Qubit 3 Fluorometer (Thermo Fisher Scientific). For each sample, 1 ug of total RNA was subjected to further RNA quality and integrity measurement using the Eukaryote Total RNA Nano assay on Bioanalyzer 2100 (Agilent Technologies). Further steps were performed at the EMBL Genomic Facilities as a collaboration with Dr. Vladimír Beneš, including preparation of barcoded mRNA-seq library, sequencing of the prepared library (Next Generation Sequencing, NGS) using NextSeq500 platform (Illumina) with paired-end reading at 75 bps read-length.

For experiment on cells harvested at day 2 post-infection (**Figure 12**), 20 000 GFP$^+$ cells were sorted directly into RNA lysis buffer (Dynabeads™ mRNA DIRECT™ Purification Kit, Thermo Fisher Scientific). Subsequent processes were performed by Dorothea Dörr (AG Achim Leutz, Max Delbrück Center for Molecular Medicine, Berlin), including RNA extraction, preparation of barcoded mRNA-seq library and sequencing. NGS was performed at the Next Generation Sequencing Unit (Scientific Genomics Platforms, Max Delbrück Center for Molecular Medicine, Berlin) using NextSeq500 platform (Illumina) with paired-end reading at 75 bps read-length.

All subsequent demultiplexing and bioinformatic analysis was performed by Dr. Karin Zimmermann (AG Achim Leutz, Max Delbrück Center for Molecular Medicine, Berlin).

2.3.6. Single-cell RNA Sequencing

Single-cell RNA sequencing (Figure 4, 5) was performed on C/EBPβ LAP1-infected cells harvested by GFP$^+$ sorting at day 6 post-infection. Cells were sorted into PBS and processed at the Single Cell Technologies Unit (Scientific Genomics Platforms, Max Delbrück Center for Molecular Medicine, Berlin) by Dr. Cornelius Fischer and Caroline Braeuning, following standard workflow by 10X Genomics technology. Briefly, cells were partitioned and barcoded using a Chromium Automation (10X Genomics) before being sequenced using a HiSeq4000 (Illumina). The experiment was performed together with and supported by Dr. Alexander Mildner (AG Achim Leutz, Max Delbrück Center for Molecular Medicine, Berlin), subsequent

bioinformatic analysis was performed by Dr. Karin Zimmermann (AG Achim Leutz, Max Delbrück Center for Molecular Medicine, Berlin).

2.4. Others

2.4.1. Statistical methods

Data was analyzed and visualized using GraphPad Prism (versions varied from 7.0.0 to 9.2.0, GraphPad Software, San Diego, California USA, www.graphpad.com). Used statistical methods were stated at each figure. Error bars represent standard error of the mean (SEM). Threshold level for assuming significance was $p < 0.05$. Significance levels were represented as follow: ns > 0.05, *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.005, ****p ≤ 0.001.

2.4.2. Bioinformatic analysis of mRNA sequencing data

Analysis was performed by Dr. Karin Zimmermann (AG Achim Leutz, Max Delbrück Center for Molecular Medicine, Berlin).

3. RESULTS

3.1. System validation

3.1.1. Characterization of C/EBP-induced lymphoid-myeloid transdifferentiation (LMT) system

The full length C/EBPα p42 isoform is a cell fate instructive factor and can trans-differentiate B cells into functional myeloid cells [50], [144], [296]. Transdifferentiation of B cells can also be induced by C/EBPβ and, interestingly, the outcome markedly varies upon PTMs of C/EBPβ [290]. In detail, methylation at arginine residue R3 abrogates neutrophil differentiation, enhances residential macrophage and dendritic cell differentiation, while unmethylated R3 favors toward inflammatory macrophage differentiation. Elimination of SUMOylation at K156 and E158 results in enhanced neutrophil-like differentiation. These findings prompted us to first explore the full potential of the C/EBP-induced LMT system.

As enforced expression of C/EBPα p42 is associated with cell cycle arrest and finally with cell death, we therefore started with C/EBPβ LAP1 for the abovementioned purpose. The open reading frame of rat C/EBPβ LAP1, fused with 3xFLAG tag, was constructed in the pMSCV retroviral vector, which also contains EGFP (enhanced green fluorescent protein) as a reporter (see 2.3.1). For transdifferentiation assay, we used an established v-Abl transformed pre-B cell line (termed from now on as B cell), which was generated as described before [166]. Two v-Abl-transformed B cell lines: B-WT (contained flox sites flanking *Cebpa* and *Cebpb*, *Cebpa*$^{fl/fl}$*Cebpb*$^{fl/fl}$) and B-dKO (double knocked-out *Cebpa* and *Cebpb, Cebpa*$^{-/-}$*Cebpb*$^{-/-}$) were retrovirally transduced with C/EBPβ LAP1. Construct-harboring GFP^{+} cells were sorted at day 6 pi (post-infection) and underwent single-cell RNA sequencing.

Transdifferentiated cells, including 3297 cells originated from B-WT cells and 1423 originated from B-dKO cells, were pooled in an integrated mapping shown in **Figure 4A**. Despite the different cell numbers, 8 clusters defined by unsupervised clustering could be found in both samples, with similar distribution patterns. Comparison of clusters distribution showed that cluster 2 and 6 were more prominent in B-dKO-originated cells than B-WT (35% and 20% in B-dKO, in comparison to 10% and 2% in B-WT) (**Figure 4B**). Gene expression analysis revealed that B cells specific genes (*Ebf1, Vpreb2, Vpreb3*) were exclusively expressed

in both clusters (**Figure 4C**), identifying them as residual B cells that possibly failed to transdifferentiate. In contrary, the remaining clusters 1, 3, 4, 5, 7 showed higher occupancy in B-WT-originated cells in comparison to B-dKO, while cluster 0 and 7 were quite similar between both samples. Myeloid markers (e.g., *Lyz2, S100a8, Cd74,* among others) were highly expressed in these clusters (0, 1, 3-5, 7), thus marking them with myeloid identity. These observations indicated that transdifferentiation in both B-WT and B-dKO cells progressed in a similar fashion and produced similar outcomes, although transdifferentiation in B-WT cells were more efficient (more myeloid cells and less undifferentiated B cells) than in B-dKO cells.

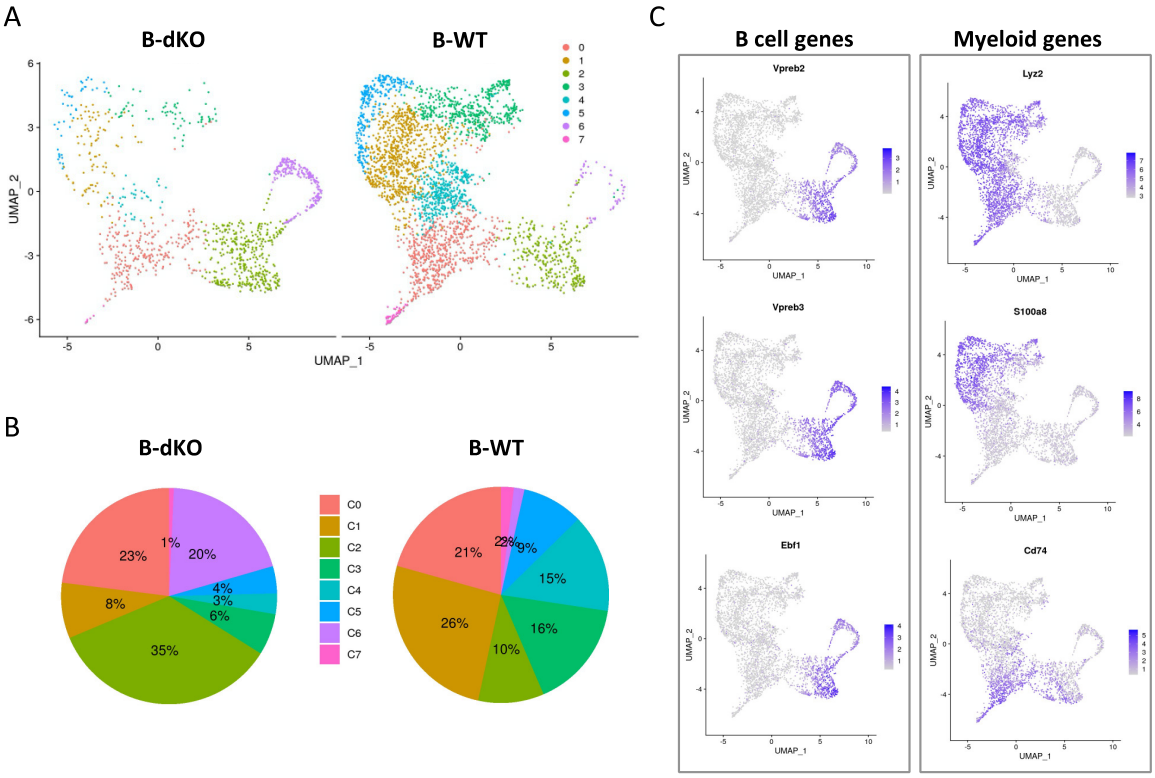


Figure 4. LMT system characterization: C/EBPβ LAP1 converted B cells to myeloid cells regardless of endogenous C/EBPα and C/EBPβ.

(A) Cebpa$^{fl/fl}$Cebpb$^{fl/fl}$ B cells (WT-B cells) or Cebpa$^{-/-}$Cebpb$^{-/-}$ B cells (DKO-B cells) were infected with C/EBPβ LAP1 using retrovirus. At day 6-pi, GFP^{+} cells were sorted and subjected to single cell RNA-sequencing. Results of two groups were pooled and sub-sampled, followed by dimensional reduction, UMAP projection and unsupervised clustering.

(B) Distribution of clusters in each cell type.

(C) Feature blots showing expression of representative B cell genes and myeloid genes.

Cluster annotation was further processed using combined events from both B-WT and B-dKO originated cells. Eight clusters defined by unsupervised clustering (**Figure 5A**) were

annotated based on gene expression profile and literature references. Specifically expressed genes in clusters 2 and 6 were of B cell identity (*Cd79a, Vpreb1, Igll1*) (**Figure 5B**). Still, only cluster 6 exhibited distinguishable expression of a group of genes related to cell division and cell cycle process, including *Top2a* (DNA topoisomerase 2 alpha), *Tubb5/Tuba1b* (Tubulin chain beta-5/alpha-1b), *Ube2s/Ube2c* (Ubiquitin-conjugating enzyme E2 S/C). These expressions indicated higher a proliferation rate in cluster 6 in comparison to cluster 2 and all other clusters, suggesting that these cells were the originating v-Abl-transformed B cells. On the other hand, the B cells in cluster 2 appeared to have low/no proliferation and were projected in a closer proximity to the myeloid counterparts. We presumed that cells in cluster 2 were in an early transdifferentiation phase where B cell identity was not yet lost, but cells were primed for myeloid conversion by ceasing proliferation.

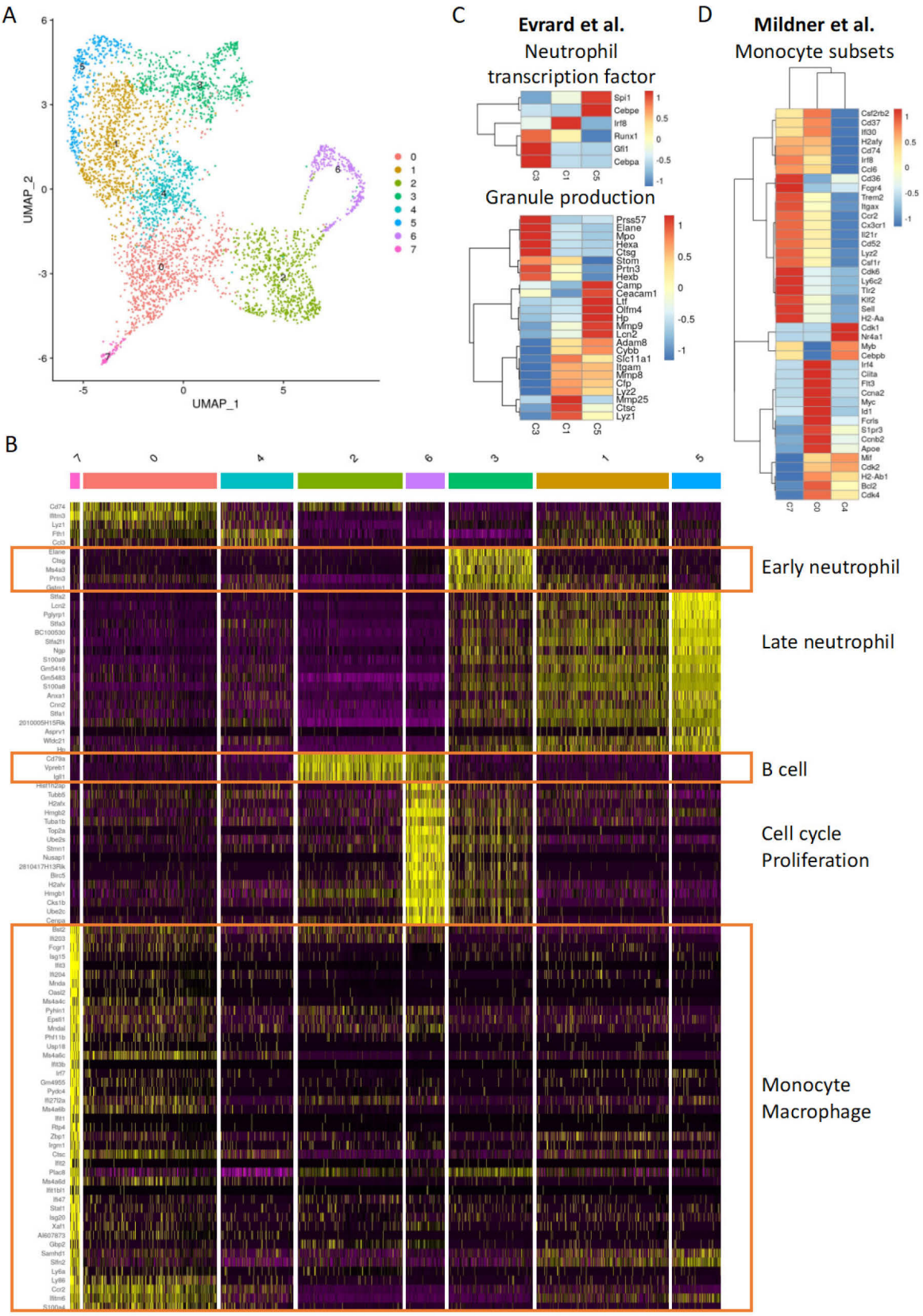


Figure 5. LMT system characterization: bipotentiality of the transdifferentiated cells.

(A) UMAP of combined events from WT-B cells and DKO-B cells. Unsupervised clustering was applied. Clusters were annotated based on expression of cluster-specific markers, representatively shown in (B).
(B) Heatmap of differentially expressed genes covering all clusters (in columns). Genes with p-value < 0.05 are shown (in rows). Additional texts annotated common features of genes in each grouping, based on literature.
(C) Expression of neutrophil genes selected from Evrard et al. [303] *in cluster 1, 3, 5.*
(D) Expression of monocyte genes selected from Mildner et al. [304] *in cluster 0, 4, 7.*

Neutrophil specific genes were found expressed in cluster 1, 3 and 5 (**Figure 5B**).

- Distinctively expressed in cluster 3 were *Elane* (Neutrophil elastase), *Prtn3* (Proteinase 3), *Ms4a3* (Membrane spanning 4-domains A3) and *Gstm1* (Glutathione S-transferase Mu 1), which have been determined as early expressed gene during GMP-Neutrophil differentiation [106].
- Other neutrophilic genes expressed faintly in cluster 3, slightly raised in cluster 1 and peaked in cluster 5, including *S100a8/9* (S100 calcium binding protein a8/9), *Stfa1/2/3* (Stefin-1/2/3), *Anxa1* (Annexin-1). These genes were known to be expressed in mature neutrophils [106].

A recent study by Evrard et. al revealed a subset of proliferative neutrophils emerged during GMP-Neutrophil transition and was termed pre-neutrophils [303]. We projected the profiling of three clusters 1, 3, 5 onto the signature gene list of Evrard's pre- or differentiated neutrophils and found strong similarity. In details, neutrophil transcription factors highly expressed in pre-neutrophils, *Gfi1, Runx1* and *Cebpa* were shown in clusters 3, while differentiated neutrophils genes *Cebpe, Spi1* were stronger in cluster 5 (**Figure 5C**). Genes important for primary granules formation expressed highly in pre-neutrophils and clusters 3, while secondary and tertiary granules genes were more accumulated in cluster 1 and 5. Considering also moderate expression of cell cycle genes (**Figure 5B**), we endorsed cluster 3 as pre-neutrophils, while cluster 5 was identified as differentiated neutrophils. In all analysis, cluster 1 showed partial similarity to either cluster 3 or 5, thus, was annotated as intermediated neutrophils.

Cluster 0, 4 and 7 presented monocytic traits due to expression of *Cd74* (Major histocompatibility complex class II), *Ccl3* (inflammatory chemokine ligand 3), *Ctsc* (Cathepsin C, lysosomal protease), *Irf7* (Interferon regulatory factor 7) and many of *Ifit* family (interferon

induced proteins) (**Figure 5B**). We referred to a Ly6C-based study by Mildner el al. to further identify monocyte subsets in these clusters [304].

- Classical monocyte traits $Ly6C^{hi}Sell^{+}Ccr2^{+}CD74^{+}$ was noticeably shown in gene expressions of cluster 7 (**Figure 5D**).
- In cluster 4, high expression of *Nr4a1, Cebpb* and minimal *Ly6c2* pointed it toward $Ly6C^{-}$ monocyte, according to the finding that C/EBPβ regulates the transition from $Ly6C^{+}$ to $Ly6C^{-}$ *Nr4a1*-high monocytes [304].
- Cluster 0 moderately expressed the $Ly6C^{+}$ monocyte genes, namely, *Ly6c2* (Lymphocyte antigen 6C2), *Sell* (Selectin L), *Ccr2* (Monocyte Chemotactic Protein 1 Receptor), *Lyz2* (Lysozyme 2), etc. Genes related to other lineages could also be found in this cluster, including dendritic cell (DC)-related genes (*Ciita, Flt3, H2-Ab1*), microglia genes (*Apoe, Fcrls*). These observations were in accordance with $Ly6C^{int}$ monocytes categorization, and furthermore, hinted a DC-biased subset in this cluster 0.

Precise identification of all transdifferentiated clusters required further functional analysis. Therefore, we decided to assign this task to a separated study (manuscript in preparation). Nevertheless, the abovementioned profiling results led us to two conclusions: (i) the C/EBPβ LAP1 based LMT system has bi-lineage (granulocytic and monocytic) potential; (ii) transdifferentiated neutrophils and monocytes appear in various developmental stages, which might resemble normal granulo-/monopoiesis. This reaffirmed our viewpoint that the LMT system is an applicable model to study myelopoiesis and, on the technical side, is a comprehensive tool to examine differentiation capacity of C/EBP constructs/mutants (**Figure 6**).

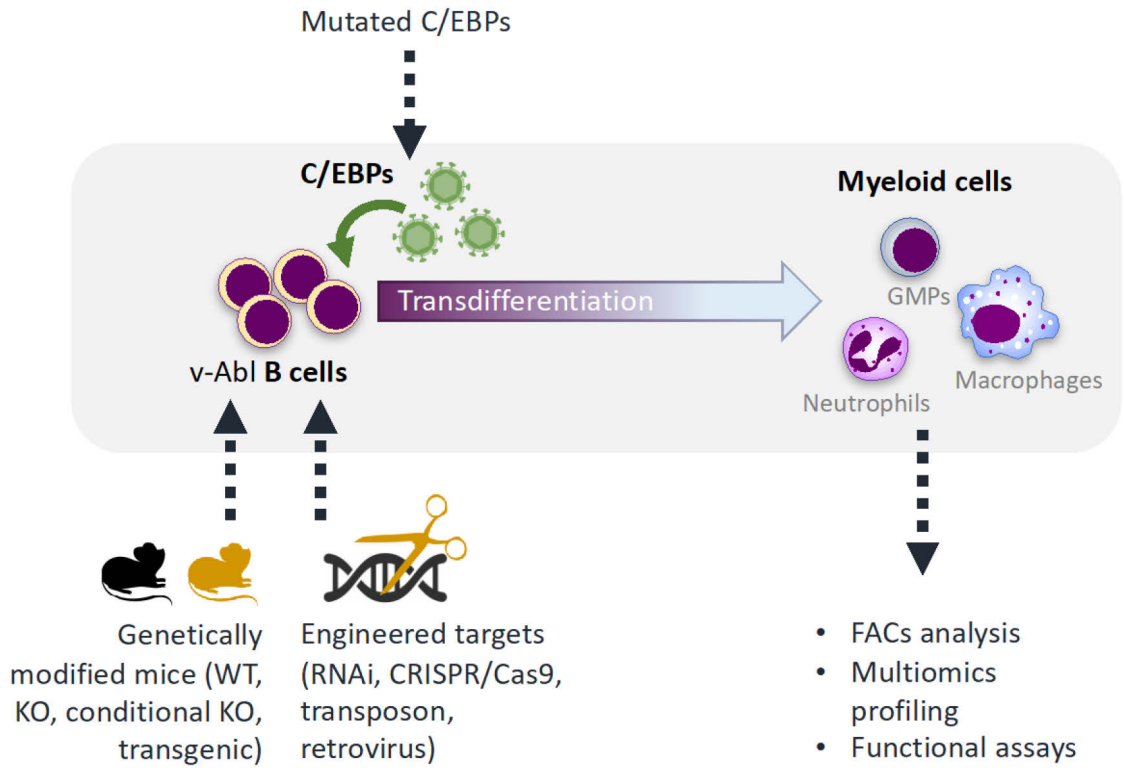


Figure 6. The C/EBP-induced LMT system.

The C/EBP-induced LMT system is a versatile experimental tool for various study purposes (grey area). Input includes primary B cells immortalized by v-Abl and C/EBP constructs, which could be modified (dashed inward arrows):

- *To study function of certain genes, B cells from genetically modified animals or in vitro engineered B cells could be used.*
- *To study function of C/EBPs, mutated construct could be used.*

The transdifferentiation process generates cells of myeloid lineages, including the bi-potential GMP and its differentiated lineages. The process could be traced or timestamped to study myelopoiesis. Output cells could be subjected to further analysis as read-outs (outward arrow). This figure contained icons designed by PNGTree (https://pngtree.com).

3.1.2. C/EBPα p30 can transdifferentiate B cells

Major members of the C/EBP family, including C/EBPα, β, δ, ε were previously examined their lineage conversion capacity using the LMT system [166]. However, only full-length isoform of each member was tested, among which, C/EBPα p42 showed robust transdifferentiation potential toward monocytes but not granulocytes/neutrophils. In contrast to p42 , the short isoform p30 was known to impair myeloid differentiation from Granulocyte-Macrophage Progenitor (GMP) [224] and failed to induce transdifferentiation in normal primary pre-B-WT cells [290]. Here, we examined the myeloid transdifferentiation

capacity of p42 and p30 in v-Abl transformed B cells (B-WT and B-dKO, see 3.1) in more detail, using the LMT system.

Both B cell lines were retrovirally transduced with C/EBP constructs shown in **Figure 7A**. C/EBPβ LAP1 and C/EBPα p42 served as positive controls, while negative control was the vector backbone without C/EBP inserts (MIEG). Cells were harvested in bulk for flow cytometric analysis at various time points and construct-harboring cells were further selected by $EGFP^+$ gating (**Figure 7B**). Successful transdifferentiation was represented by reduced expression of B cell marker CD19 and emerging expression of myeloid marker CD11b, according to flow cytometric measurement. Percentages of $CD19^{dim/neg}CD11b^+$ events were used to evaluate transdifferentiation efficiency of each construct (**Figure 7B**, red number).

Time-course measurements showed the capability of C/EBPα p30 to induced transdifferentiation in both cell types (**Figure 7C**), albeit with lower efficiency in comparison to C/EBPα p42 and C/EBPβ LAP1. Cell expressing p30 could be maintained and expanded in culture, while p42 induced cell death (observation, data not shown). Similar to our observation in 3.1.1, B-WT cells yielded more $CD11b^+$ myeloid cells than B-dKO cells when induced by the same C/EBP. Although myeloid cell output was low, $CD11b^+$ cell yield across time points were significant with p30 in comparison to the MIEG control. This result validated our LMT system as suitable for functional p30 studies.

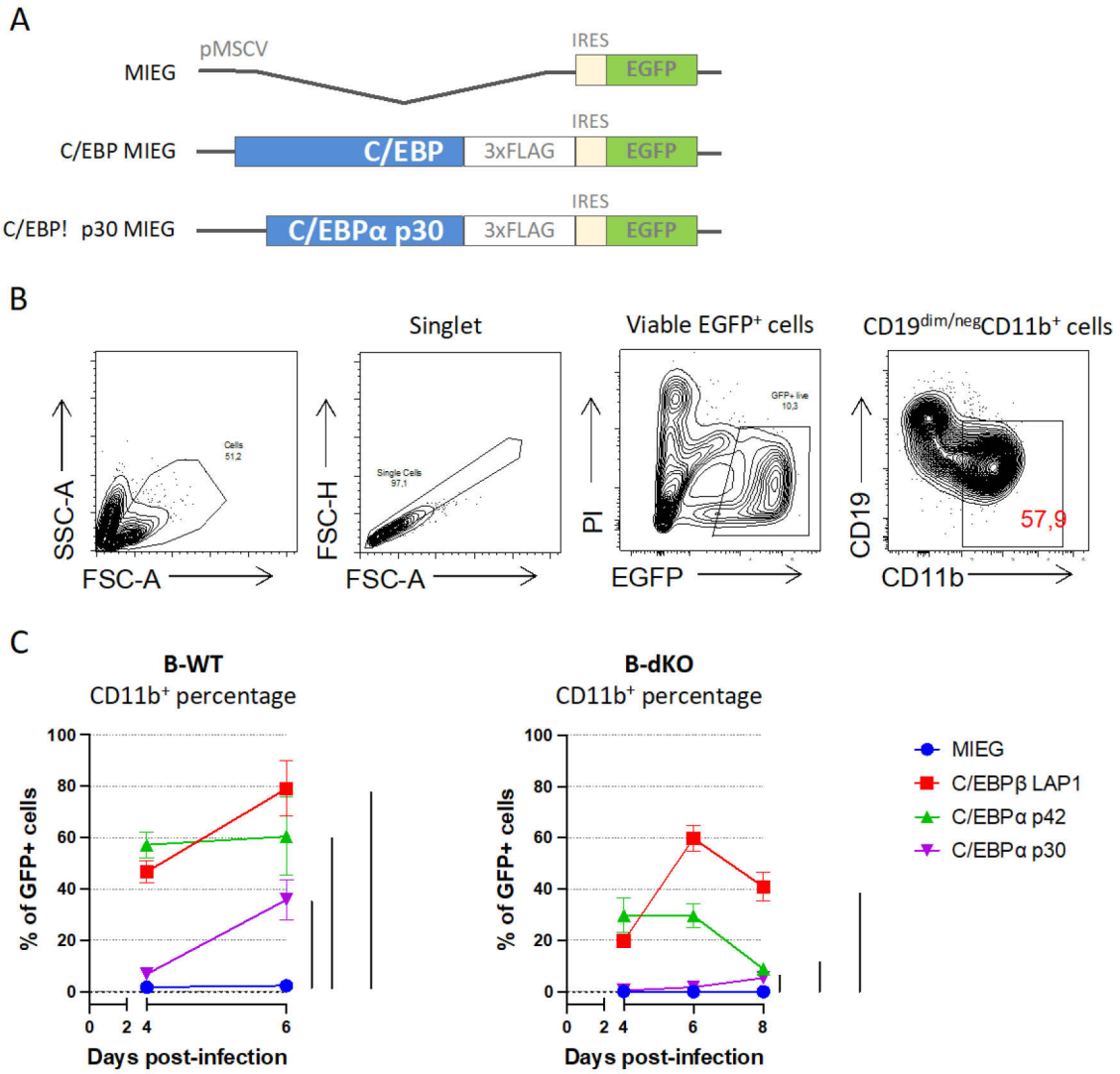


Figure 7. Transdifferentiation potential of C/EBPα p30.

(A) Simplified structure of the expressing constructs used in this study. MIEG construct was used as vector control. MIEG-based C/EBP constructs (either C/EBPα p42 or C/EBPβ LAP1) were used as positive control of transdifferentiation, as studied before [166].

(B) Successful transdifferentiation was measured by CD19 down-regulation and CD11b expression as determined by FAC analysis. General gating strategy is shown as follow: after excluding debris, dead cells, and doublets, GFP⁺ living events were subjected to analysis of CD19 and CD11b. Percentage of gated CD19$^{dim/neg}$CD11b⁺ (shown in red) was used for graphs in (C).

(C) Percentage of CD19$^{dim/neg}$CD11b⁺ cells transdifferentiated from either B-WT cells or from B-dKO cells. Data are shown as mean ± SEM. Statistical significance was determined by mixed-model analysis, followed by Dunnett's test. Multiple comparisons were between C/EBP and MIEG control. Mean values of all time points were used to calculate overall significance (main column effect).

3.2. Arginine methylations alter p30-induced transdifferentiation

Based on the hypothesis that N-terminal Arginine methylation affects the biology of C/EBPα p30, we first modified arginine residues to mimic methylated or unmethylated status. The arginine residues of interest are R140, R147 and R154, which were chosen based on (i) preliminary experiments showing methylation-dependent differential binding of p30 to chromatin remodeling complexes (examined by Dr. Elisabeth Kowen-Leutz, data not shown), and (ii) C/EBPα interactome analysis by Ramberger et al. showing p30-specific binding function dependent on R142 PTMs (human homolog of mouse R140) [292]. These arginine residues reside in a cross-species conserved transactivation domain (TAD2) and are separated to each other by 6 amino acids. Amino acid substitution was used to exchange Arginine by either Lysin, Leucin or Alanine (**Figure 8A**).

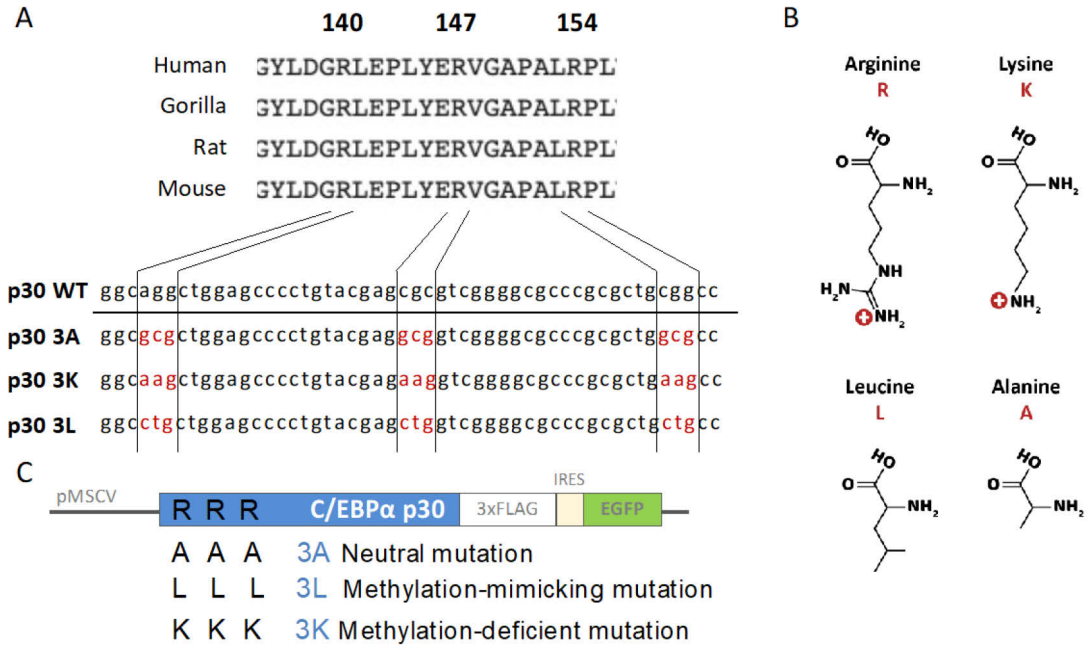


Figure 8. Schematic illustration of the used amino acid substituted constructs.
(A) Upper: amino acid alignment showing cross-species conserved arginine residues on C/EBPα p30. Arginine of-interest are shown in highlight. Peptide sequence alignment was generated on https://www.ncbi.nlm.nih.gov/ using MUSCLE tool. Lower: nucleotide alignment of substituted Cebpa sequence showing replacement of Arginine codons by Alanine/Lysine/Leucine codons. Nucleotide alignment was performed on Sanger-sequenced data of the plasmids using SnapGene software.
(B) Illustrated structural formular of Arginine and the substitutions.

(C) Simplified structure of the retrovirus constructs used in this study to overexpress C/EBPα. Relative sites of arginine residues and the substitutions, as well as roles of the substitutions are shown.

Arginine is a basic, positively charged amino acid. Methylation of arginine adds up to two methyl groups in exchange of hydrogen atoms that can form hydrogen bonds with certain residues on binding partners [305], [306]. Arginine shares physicochemical properties with lysin (K) [307] (**Figure 8B**). Substitution of arginine by lysine is a conservative substitution, which preserves the positive charge while abolishing the required substrate for methylation by the PRMTs and thus may acts as unmethylated arginine mimesis. On the other hand, both leucine (L) and alanine (A) have little similarity to arginine and are more interfering substitutions for R. Alanine has been widely used as neutral mutation because of its small size, non-polarity, and non-active residue function. Leucine may serve as a hydrophobic substitution with neither capacity to donor hydrogen nor form hydrogen bonds, which is similar to methylated arginine.

On the retroviral C/EBPα p30 construct, point mutations were introduced at three consecutive arginine codons, termed triple mutation or 3A, 3K, 3L, for each type of substitution (**Figure 8C**). Both B cell lines were retrovirally transduced at day 0 and harvested in bulk at different time points. As described in **Figure 7C**, GFP^+ transdifferentiated cells were determined as $CD19^{dim/neg}CD11b^+$ and blotted. As shown in **Figure 9A** emergence of $CD11b^+$ myeloid cells by C/EBP p30 could be found (**Figure 9B, C**). This capacity of p30 was strongly elevated by mutations 3A and 3L, showing by the steep rise of $CD11b^+$ cells percentage over time, while 3K mutations produced significantly less $CD11b^+$ cells. Next, single Arginine mutations were examined to determine the most critical Arginine residue for alteration of p30 functions (**Figure 9D, E**). Measurement at day 4- or 6-days post-infection presented similar patterns in both cell types and in all arginine residues: A and L mutations yielded more myeloid $CD11b^+$ cells in comparison to p30 WT, while K mutations showed the opposite effect. These differences were smaller with single mutations (1-4x higher/lower than p30 WT in B-WT cells, 1-7x in B-dKO cells), and much larger with triple mutations (6-9x higher/lower than p30 WT in B-WT cells, 10-18x higher/lower in B-dKO cells). We proposed that the changes caused by single mutation were possibly compensated by the other wild-type arginine

residues. These observations suggested that all three Arginine R140/147/154 were involved in coordination with each other to regulate C/EBPα p30 functions.

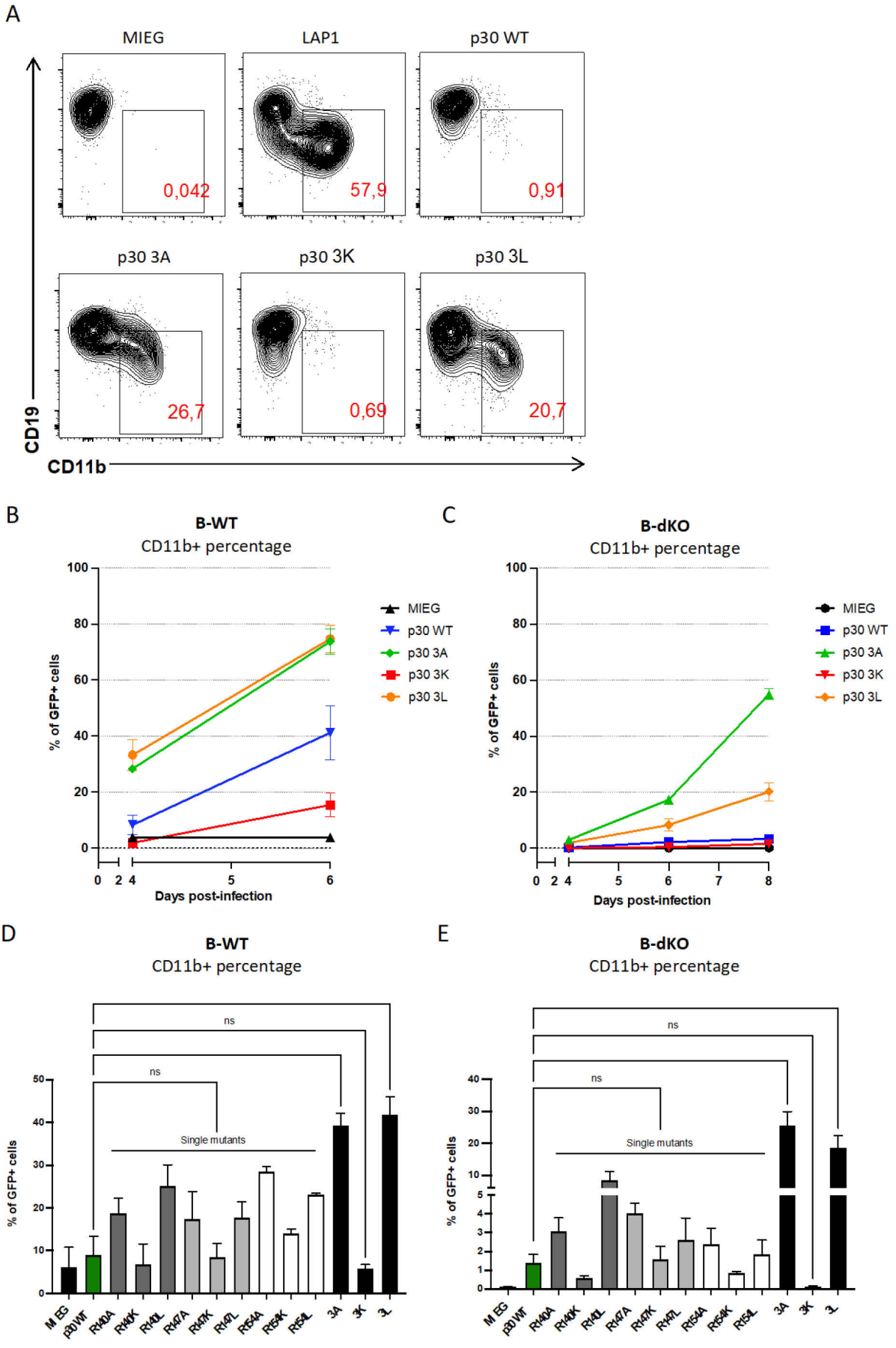


Figure 9. Arginine methylations altered p30-induced transdifferentiation.

Figure 9. Arginine methylations altered p30-induced transdifferentiation.
(A) Representative FACs analysis of transdifferentiation analyzed on B-dKO expressing Cebpa constructs at day 6 post-infection. Gating was shown in ***Figure 7****, in which, all living cells were gated for GFP$^+$ before further CD19 CD11b gating. Percentage of gated CD19^{neg}CD11b$^+$ (shown in red) was used for graphs in (B)-(E).*
(B) Percentage of CD11b$^+$ cells at various time-points, transdifferentiated from either B-WT cells or (C) from B-dKO cells. Data are shown as mean ± SEM, significance was determined by two-way ANOVA analysis followed by Dunnett's multiple comparisons test, n=4. Only significant comparisons between p30 WT and other constructs are shown by asterisks in matching color with lines, insignificant comparisons are not shown.
(D) Percentage of CD11b$^+$ cells induced by single mutations or triple mutations, transdifferentiated from B-WT cells or (E) from B-dKO cells. Data are shown as mean ± SEM, significance was determined by one-way ANOVA analysis followed by Dunnett's multiple comparisons test, n=3, significance between p30 WT and mutants at day 4-pi (B-WT) or day 6-pi (B-dKO) are shown.

3.3. Transcriptional profile of methylation dependent transdifferentiation

3.3.1. C/EBPα p30 mutants differentially induced myeloid transcriptional profile

Bulk RNA-sequencing using EGFP$^+$ sorted cells at day 4 post-infection was performed to evaluate the effect of C/EBPα p30 mutations at transcriptional level (**Figure 10A**). Pair-wise comparisons identified 2639 differentially expressed genes (DEGs) (**Figure 10B**). Highly expressed lineage specific genes in the control group (MIEG) were mostly B cell markers, including *Vpreb3, Cd22, Bach2, or Bcl2*. These genes expressed in a uniform pattern: slightly diminished by 3K or p30 WT, while strongly reduced by 3L and 3A p30 mutants (**Figure 10C**). In contrast, a large set of genes was regulated in an opposing, yet uniform pattern: highest expression in 3A, gradually reduced level in the order of 3L, p30WT and 3K, and low/no expression in control group. Genes in this category marked myeloid identity, such as transcription factors (*Spi1, Fosl2*), growth factors (*Csf1, Csf1r, Csf2ra/b*), cytokines receptors (*Flt3*, *Il12rb1, Ifngr1*) and surface markers (*Itgam, Cd33*). Principle component analysis using DEGs confirmed the observations (**Figure 10D**), that although all p30 constructs could induce transcriptional changes in B cells (p30 variants were clustered separately from MIEG), the extremity of changes varied among them. When compared with MIEG, the 3A cluster was at the farthermost variance distance, followed by 3L, p30WT and 3K, respectively.

Recently, Jakobsen et al. have identified C/EBPα p30-specific targets in AML patients and mouse models, which could also be found in our data set (**Figure 10E**) [231]. Accordingly, specific binding of C/EBPα p30, but not p42, was found at a -40 kb enhancer region of *Nt5e* (5'-Nucleotidase Ecto); and *NT5E* was found highly expressed in *CEBPA*-mutated AML samples from patients and a mouse model. In our transdifferentiation system, *Nt5e* was upregulated nearly equally by all p30 constructs, suggesting normal gene activation of the p30 mutants, although not showing methylation dependency. On the other hand, other p30-upregulated genes (in [231]) differed in their expression in accordance with methylation: 3K failed to or reduced expression of *Tyrobp* (TYRO Protein Tyrosine Kinase Binding Protein), *Hck* (Hemopoietic Cell Kinase) and *Arrb2* (Arrestin Beta 2), while 3A, 3L enhanced their expression. Gene down regulated by p30 (in [231]) also exhibited methylation dependent expression (**Figure 10E,** lower row)**.**

C/EBPα showed isoform-specific co-binding activity with various proteins, including the ETS transcription factor family, for instance, ETS member ETS1, ELK4, ETV4 were found more likely to co-bind with p30 to target genes, rather than p42 [231]. In our dataset, we found that gene expression of ETS members were regulated by p30, such as *Etv3/4* (Ets variant 3/4), *Elk4* (ETS Transcription Factor ELK4), *Spi1* (Spleen focus forming virus (SFFV) pro-viral integration oncogene) (**Supplementary Figure 1**). Expression of these genes suggested that besides having protein-protein interactions, C/EBPα p30 also regulates the expression of its co-factors; thus initiates or mediates its unique transcriptional program.

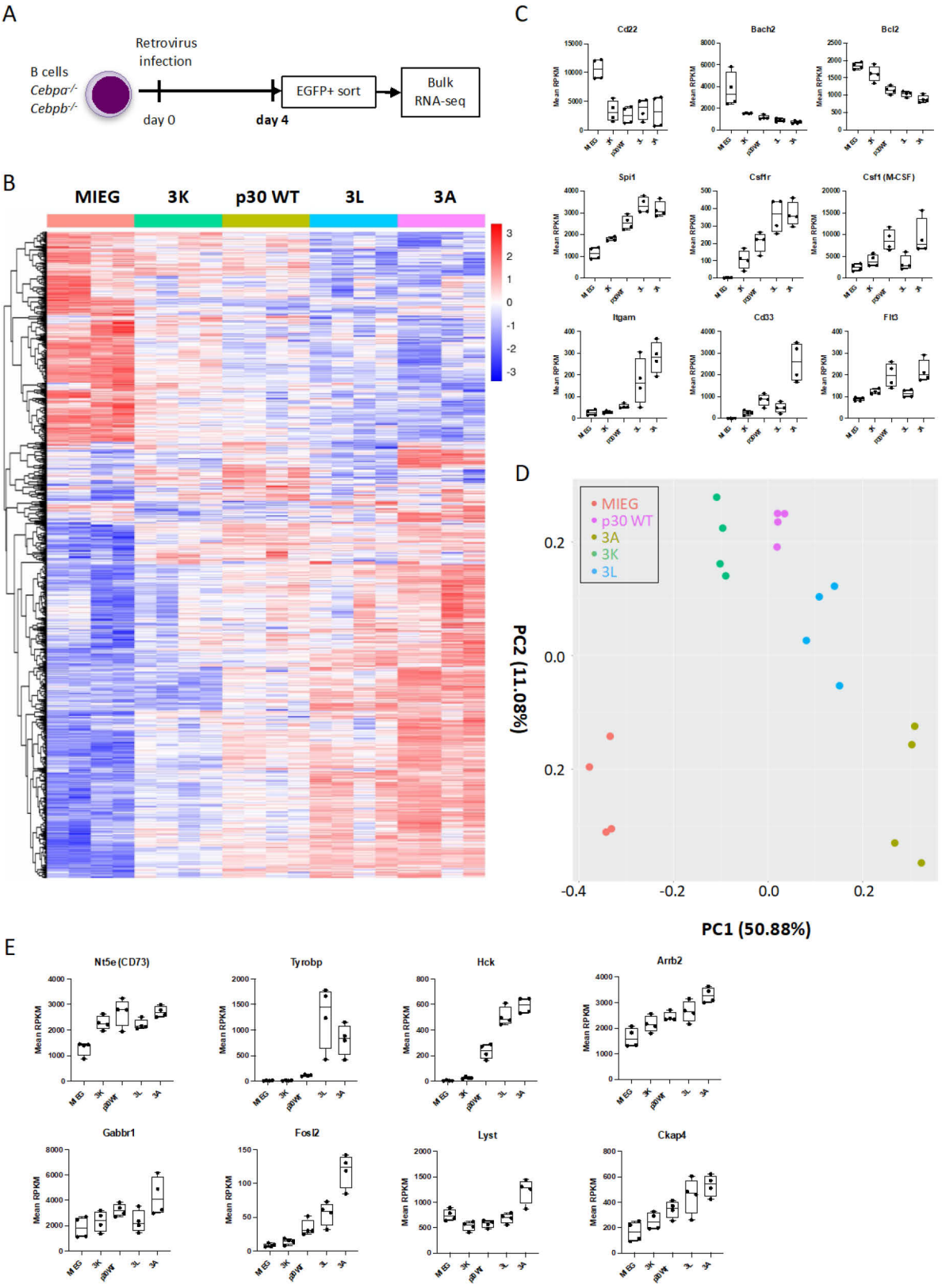


Figure 10. Transcriptional profile of methylation dependent transdifferentiation.

(A) Schematic illustration of sample preparation. B-dKO cells were retrovirally infected with p30 constructs or MIEG control. GFP⁺ cells were sorted directly into RNA lysis buffer.

Samples were in quadruplicates; all samples were processed together. 1 μg of total RNA was used for library preparation and sequencing.

(B) Heatmap presenting pairwise differential gene expression analysis. Gene expressed differentially in at least 1 comparison with p-value <0.05 was listed. Experimental groups (in columns) were manually arranged based on similarity of gene expression pattern.

(C) Relative expression of representative B cell specific genes (top row) and myeloid specific genes (second and third row).

(D) Principle component analysis using DEGs in (B).

(E) Relative expression of p30 target genes, as identified in Jakobsen et al. [231].

Gene ontology (GO) enrichment analysis was performed to identify biological processes enriched by each mutant, using the DEGs as listed above. Transdifferentiation was verified by terms related to monocytic activity including chemotaxis, inflammatory response, production of Interleukin IL-1, IL-1b, IL-6, which could be found enriched in p30, 3A and to a lesser extend in 3L and 3K (**Figure 11A**). Interestingly, biological process term "Superoxide anion generation", which represented the capacity of producing free radical oxygen important for killing of phagocyted bacterial, was enriched by every p30 variants and distinct from B cell control (MIEG). This implied that p30 could induce a cell fate with phagocytosis characteristic, regardless of the methylation status. To focus on the effect of p30 mutants, we used pairwise comparisons of DEGs of each mutant against p30 WT to analyze GO biological process enrichment (**Figure 11B**). In comparison to p30 WT, the 3A mutant showed stronger indication of myeloid fate by terms related to inflammatory responses and cell mobility (leukocyte chemotaxis, heterotypic cell-cell adhesion, purigenic receptor signaling pathway [308]). On the other hand, processes enriched by 3L were linked to cellular events of protein translation which required rRNA, ribosome formation and amino acid synthesis. The term "Mitochondrial translation" could be found as a part of intracellular anatomical structure [309], and suggested an enhanced activity in cell morphology or mobility. These enrichments indicated that cells expressing 3L might be in a transitional phase that required a massive synthesis of new materials, as a result of a change in transcriptome. Non-coding RNAs (ncRNAs) have been recently focused in various studies of hematopoietic system, which revealed ncRNA-fingerprints specifying distinct lineages [310], [311]. We noted two terms related to ncRNA enriched by 3L mutants, which could be inferred as a particular ncRNA-mediated change in cell fate caused by this mutant. Lastly, the 3K mutant presented terms related to cell replication or proliferation. We considered this enrichment as the identity of

the original un-transdifferentiated v-Abl B-dKO cells, which also dominated in the MIEG control group (**Figure 11C**). Nevertheless, “Monocyte chemotaxis” shown by 3K (**Figure 11B**) marked a potential of 3K to induce transdifferentiation, although at a much lower extend, as comparing with other p30 variants.

In general, C/EBPα p30 variants demonstrated diverse transdifferentiation potential. At 4 days post-infection, both null-mutant and R-methylated mimesis of p30 (3A and 3L mutant, respectively) successfully induced myeloid transcriptional program and suppressed B cell program to a significantly greater extent than the wild-type p30 and R-unmethylated mimesis p30 (3K). It was possible that 3K mutant presented delayed transdifferentiation, which might take longer than our observation time window (8 days). This possibility was ruled out by monitoring 3K-expressing B cells for a prolonged period of 30 days, during which, myeloid population emerged at very low rate (less than 1%, data not shown).

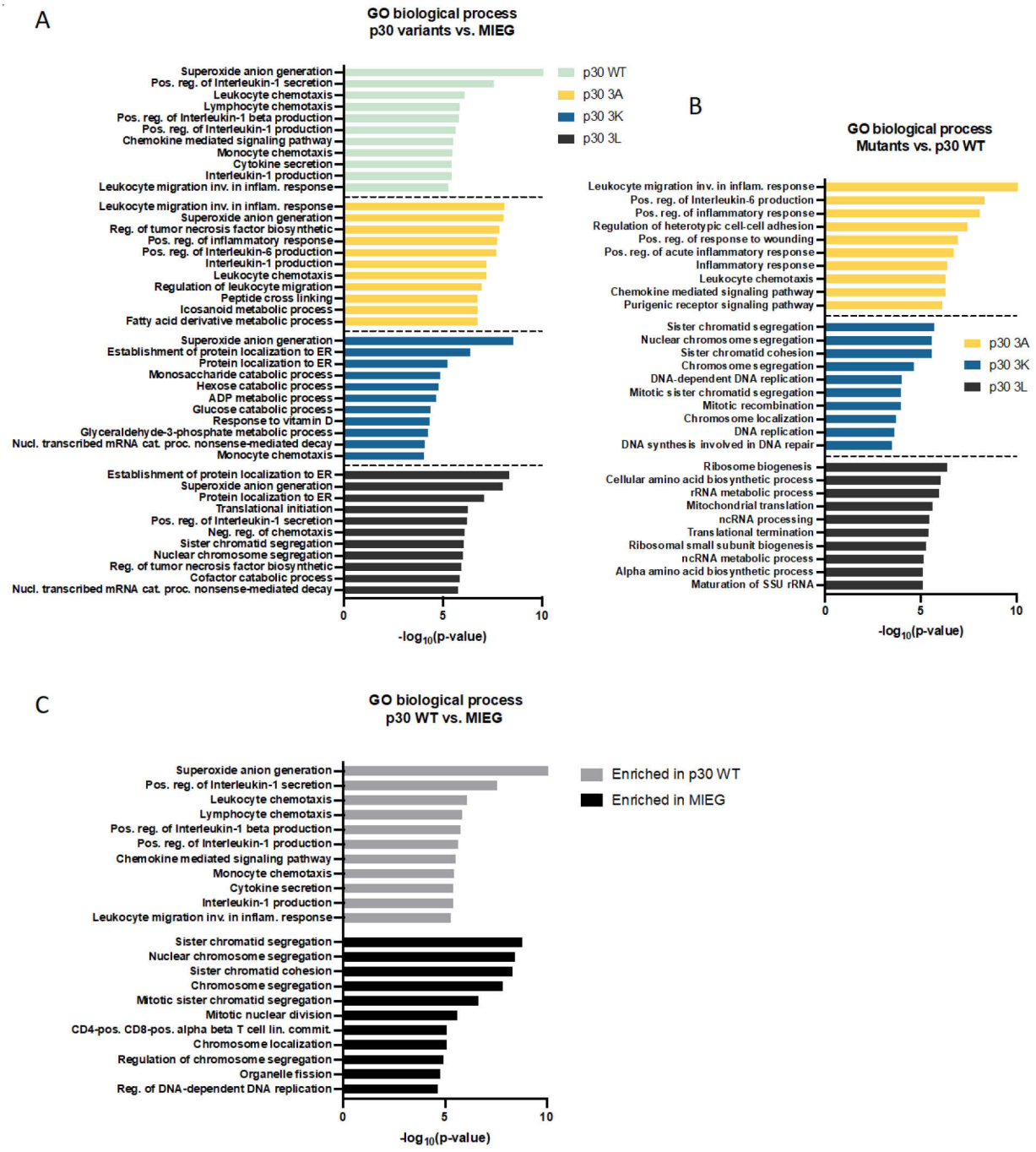


Figure 11. Pair-wise GO-terms analysis of differentially expressed genes.

Top enriched GO-Biological processes (GO-BP) were shown in order of increasing -log10(p-value). Comparison included:

(A) p30 variants vs. MIEG control. Only terms upregulated in p30 variants were shown.

(B) p30 mutants vs. p30 WT. Only terms upregulated in p30 mutants were shown.

(C) MIEG control vs. p30 WT.

Abbreviations: pos. (positive), neg. (negative), reg. (regulation), nucl. (nuclear), cat. (catabolic), proc. (process), inflam. (inflammatory), inv. (involved).

3.3.2. Identifying distinct targets of p30 mutants

To discover the underlying mechanism of p30 mutants' function, we aimed to identify genes that were exclusively induced by each p30 mutant. However, current data set of **day 4**-transdifferentiated cells was found biased by the myeloid proportion. In detail, bulk RNA-sequencing was performed on GFP$^+$ cells, which contains both transdifferentiated (GFP$^+$CD11b$^+$) and un-transdifferentiated (GFP$^+$CD11b^{neg}) cells; proportion of CD11b$^+$ transdifferentiated cells varied in each group as shown in **Figure 8, 12A**. These transdifferentiated portions might determine level of expression of myeloid genes, for example, *Itgam* (encoding CD11b) and its downstream targets *Il1b* and *Vegfa* [312] (**Figure 12B)**. We speculated that genes expressed in the similar pattern to *Itgam* (inclination following the order of MIEG – 3K – p30 WT – 3A – 3L) were secondary events after achieving the myeloid fate. It is therefore difficult to conclude the direct relationship between p30 mutants and *Il1b* or *Vegfa*. Accordingly, the CD11b-bias might hinder the finding of exclusive target of each p30 constructs.

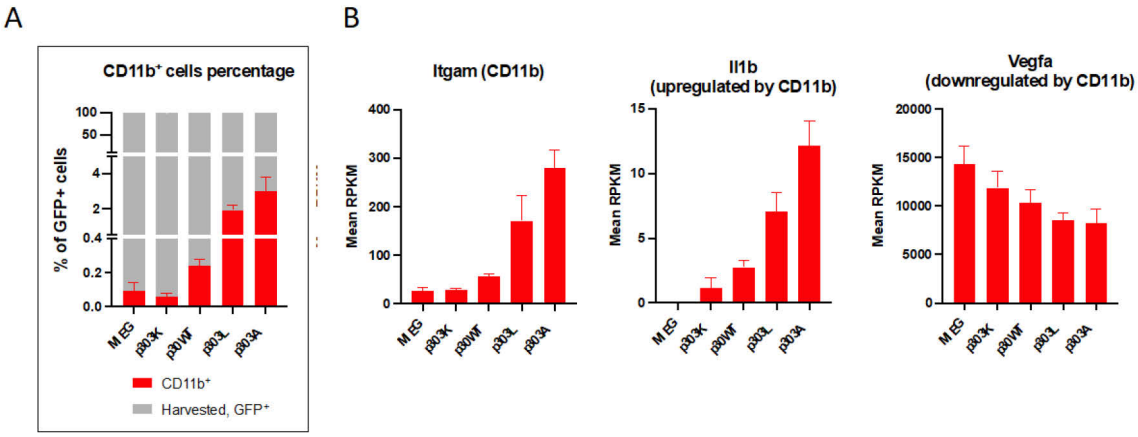


Figure 12. The CD11b-bias.
(A) For RNA-seq experiments, whole population of GFP$^+$ cells were harvested (100%, gray bars). Within this population, size of CD11b$^+$ subpopulation varied among groups (red bars) and accounted for Itgam gene expression.
(B) Relative expression of Itgam and its representative downstream targets Il1b, Vegfa, as determined by bulk RNA-seq.

To rule out the bias caused by transdifferentiation, we performed bulk RNA-sequencing using GFP$^+$ cells harvested at **day 2** post-infection (**Figure 13A**). At the point of harvesting, all GFP$^+$ cells were CD11b^{neg} (data not shown), thus the profiling might capture early events of transdifferentiation before the cells show the myeloid phenotype. Pairwise comparisons

between p30 variants and MIEG control resulted in 503 differentially expressed genes (DEGs) and they were intersected with day 4 DEGs (**Supplementary Figure 2A**). Joint DEGs in both day 2 and day 4 datasets were considered specific targets of p30 constructs at early transdifferentiation, and shown in **Table 3**. Similar to previous observations, cells expressing 3A showed the largest differences when compared to MIEG (165 DEGs), followed by p30 WT (60 DEGs), 3L (50 DEGs) and 3K (25 DEGs).

GO-Biological process (GO-BP) term analysis using lists of DEGs also showed enrichment of myeloid related terms, such as inflammatory response, leukocyte cell-cell adhesion, leukocyte/neutrophil migration (**Supplementary Figure 2B**), confirming upregulation of myeloid program by p30 WT/3A/3L. Mutant 3K did not yield any significant enrichment, possibly due to small number of DEGs (25 genes), nevertheless, various myeloid genes were induced by 3K, such as *Ncf4* (Neutrophil cytosol factor 4), *Trem1* (Triggering receptor expressed on myeloid cells 1) (**Table 3**). Curiously, 3L showed enrichment of terms related to development of mesodermal derivatives, including lateral (heart morphogenesis) and intermediate mesoderm (prostate gland, renal, urogenital and kidney development). Those terms presented two genes in common: *Ahr* (Aryl hydrocarbon receptor) and *Notch1* (Neurogenic locus notch homolog protein 1), both of which were found important for definitive hematopoiesis and differentiation of hematopoietic stem cells [313]. GO term Respiratory burst appeared enriched by p30 WT and 3A and was considered signature of neutrophilic development, as previously seen in C/EBPα p42-induced transdifferentiation [182], [197].

Table 3. Differentially expressed genes in comparison to MIEG, intersect of day 2 and day 4 datasets.

p30 WT **60 genes**	**3A** **165 genes**	**3L** **50 genes**	**3K** **25 genes**
Abcc5, Arel1, Arhgap31, Arhgef18, Atf5, B430306N03Rik, Ccl6, Cd33, Cd44, Cd47, Cpm, Cxcr5, Cybb, Evi2a, Ffar2, Frmd5, Gbp8,	*Acp6, Acpp, Ahr, Aldh3b1, Amotl1, Anxa3, Arel1, Arhgap31, Arhgef18, Arrb2, Atf3, Atf5, B430306N03Rik, Bach2, Bcar3, Bcl7a, Blk, C3, Camk2d, Carmil2, Cbr3, Ccl6, Ccn3, Cd2, Cd24a, Cd33, Cd44, Cd47, Cdh17, Cdkn2a, Cebpa, Clec2i, Cnn2, Csf2rb2, Ctsz, Cxcr5, Cyba, Dab2ip, Ddc, Dedd2, Dgat1, Dgat2,*	*Acadsb, Ahr, Aldh3b1, Amotl1, Angptl6, Arhgef18, Atf3, Atf5, Camk4, Ccn3, Cd44, Cdkn1a, Cox6a2, Cxcr5,*	*Abcc5, Aldh3b1, Arhgef18, Atf5, Cebpa, Cybb, Ell2, Ffar2, Hcst, Id3, Jchain, Lrrc32, Milr1, Ncf4,*

p30 WT 60 genes	3A 165 genes	3L 50 genes	3K 25 genes
Gcnt1, Gpat3, Gpr84, Hepacam2, Hmgn3, Id3, Ifngr1, Inppl1, Jchain, Lat, Lax1, Lrrc32, Lta4h, Ltf, Ly6d, Mgll, Milr1, Mindy1, Ncf2, Ncf4, Neurl3, Notch1, Nupr1, Plaur, Rab27a, Rab7b, Rflnb, Sell, Slamf6, Slc1a4, Slc8b1, Slpi, Smim41, Stom, Syndig1l, Tgm2, Tmem86a, Trem1, Trem3, Umod, Vsir, Xbp1, Xrcc5	*Dhrs1, Dhrs3, Dhrs7, Dstn, E130215H24Rik, E2f8, Epsti1, Erc1, Ero1l, Fam117a, Fam234a, Ffar2, Flot2, Frmd5, Fxyd5, G6pdx, Gadd45b, Gas7, Gcnt1, Gem, Gfi1, Glipr1, Gm2788, Gm8369, Gnb4, Gpat3, Gpr84, Grap2, Gsto1, Hmgn3, Hpgds, Hsd11b1, Id3, Ier3, Ifngr1, Inppl1, Irf7, Jchain, Kbtbd11, Kynu, Lat, Lax1, Lck, Lcmt1, Lgals3, Litaf, Lta, Lta4h, Ltb4r1, Ltf, Lxn, Ly6c2, Ly6d, Lyz2, Map1lc3a, Mc1r, Megf10, Mgll, Milr1, Mindy1, Mrgpre, Msrb1, Mycn, N4bp2, Ncf1, Ncf2, Neurl3, Nfil3, Ninj1, Nupr1, Papolg, Plaur, Plek, Pnkp, Prg3, Rab3d, Rab7b, Rasgrp1, Rflnb, Rgs8, Rnf130, Rnf150, Rragd, S1pr3, Sbk1, Scn4b, Sell, Serpinb1a, Sertad4, Sit1, Slamf6, Slc1a4, Slc31a2, Slc8b1, Sln, Slpi, Smarca4, Smim41, Snx20, Sp100, Spi1, St3gal6, Stom, Syndig1l, Tbc1d16, Tcn2, Tgm2, Tmem255a, Tnfsf11, Trem1, Trem3, Trp53inp2, Ttll9, Tuba8, Tubb3, Twsg1, Txnrd3, Tyrobp, Ubtd1, Umod, Vsir, Zbp1, Zfp961*	*Cyba, Cybb, Dpp4, Ell2, Ffar2, Gcnt1, Gnb4, Gng4, Gpr84, Grap2, Gsto1, Hepacam2, Hmgn3, Id3, Jchain, Lax1, Lck, Lrrc32, Lta, Mindy1, Neurl3, Notch1, Plaur, Plek, Rflnb, Sell, Slamf6, Slc8b1, Sln, Slpi, Smim41, Stom, Syndig1l, Tnfsf11, Trib3, Umod*	*Nupr1, Pim1, Plaur, Slamf6, Slc1a4, Slc8b1, Slpi, Stom, Syndig1l, Tgm2, Trem1*

To find unique induction by each p30 variant, DEGs lists in **Table 3** were categorized by up/downregulated genes and intersected (**Figure 13B, C**). Early changes in gene expression were noted as follow:

- All p30 variants induced expression of essential genes for myelopoiesis, including *Atf5* (Cyclic AMP-dependent transcription factor ATF-5, interacts with C/EBPα, C/EBPε and C/EBPγ) *Slpi* (Secretory leukocyte protease inhibitor, required for granulopoiesis), *Plaur* (Urokinase plasminogen activator surface receptor, regulates proliferation, apoptosis of hematopoietic cells), *Ffar2* (Free fatty acid receptor 2, expressed in neutrophils and monocytes). They also suppressed

lymphoid specific genes, e.g., *Id3* (DNA-binding protein inhibitor ID-3), *Jchain* (Immunoglobulin J chain).

- Unique upregulated gene by 3K mutant included *Pim1* (Proto-Oncogene Serine/Threonine-Protein Kinase Pim-1), which was reported to regulate hematopoiesis by expanding population of $Lin^{-}cKit^{+}Sca\text{-}1^{+}$ HSCs [314]; and *Hcst* (Hematopoietic Cell Signal Transducer), a trans membrane adapter which is not well-characterized in hematopoiesis. Downregulated genes by 3K included lymphoid-specific genes *Ddc* (Dopa Decarboxylase), *Gimap4* (GTPase of Immunity-Associated protein 4) and interestingly, myeloid specific *Dlx1* (Distal-Less Homeobox 1).
- Noteworthy upregulated DEGs in 3L were the C/EBPα targets and negative regulator *Trib3* (Tribbles Pseudokinase 3), and cellular proliferation inhibitor *Cdkn1a* (Cyclin Dependent Kinase Inhibitor 1A, or p21).
- Wildtype p30 induced known C/EBPα target genes (*Rab27a, Evi2a, Cpm*). Interestingly, *Xbp1* (X-box binding protein 1), which were previously found upregulated in a group of AML patients and might inhibit myeloid differentiation by restricting *CEBPA* translation [315]. The expression of *Xbp1* may suggest mechanism of p30-driven differentiation block.
- Mutant 3A showed pronounced effect on gene expression, with 11 down- (including B-/lymphoid-specific genes) and 61 upregulated genes (majorly myeloid-/neutrophil-specific genes).

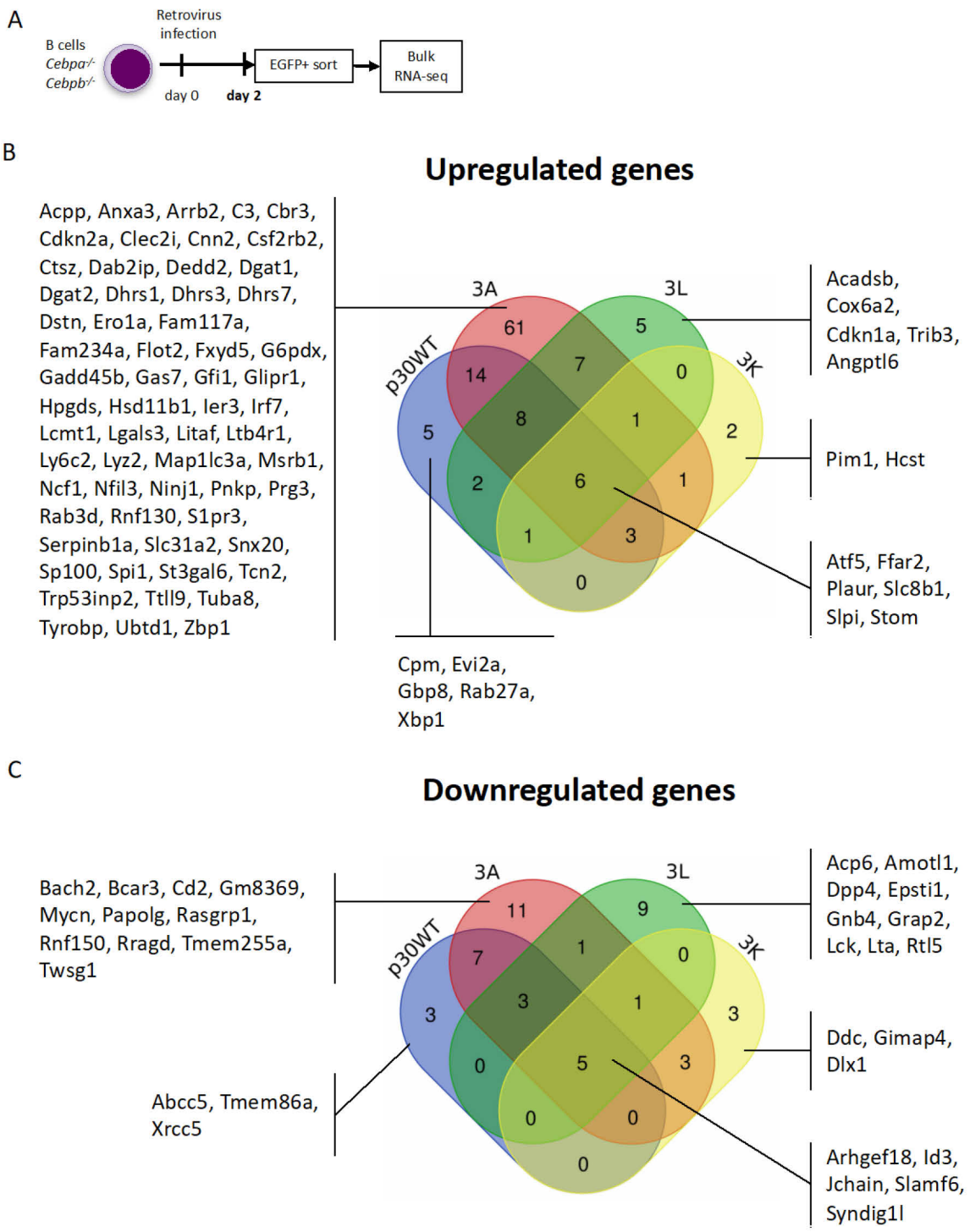


Figure 13. Transcriptional profile of transdifferentiation at day 2 post-infection.
*(A) Schematic illustration of sample preparation. B-dKO cells were retrovirally infected with p30 constructs or MIEG control. At day 2-pi, 20 000 construct-harboring GFP⁺ cells were sorted directly into RNA lysis buffer and subjected to library preparation and bulk RNA-sequencing. (B) and (C) Pairwise DEGs analysis was performed similar to day 4 analysis (**Figure 10B**). DEGs in comparison to MIEG control from two datasets were intersected. Overlapped upregulated or downregulated genes were then subjected to composing of Venn diagrams to identify*

unique affected genes in each group. Intersection and Venn diagram was composed using webtool http://bioinformatics.psb.ugent.be/webtools/Venn/.

A complete list of genes comprising the Venn diagrams are shown in **Supplementary table 1, 2**. Unique targets of each C/EBPα p30 variants hinted at the mechanism of their functions, which required further validation and investigation. For many genes, experimental evidence of regulation involving C/EBPα, especially p30, is still missing. Hypothetical connection of C/EBPα and the interesting target will be further discussed in Discussion session.

3.4. Increased proliferative potential of Arginine-to-Lysin mutated C/EBPα p30 (3K)

C/EBPα p30 isoform is known as an oncogene for granulocyte-macrophage progenitors, in the absence of cell-cycle restriction function of the full length p42 isoform [226]. Pro-proliferative activity of p30 was observed as accumulation of undifferentiated cells, increased proliferation, and enhanced replating efficiency. To examine whether Arginine methylation may be involved in proliferation function of p30, we performed serial replating assay using bone marrow-derived c-Kit-enriched cells from *Cebpa*$^{fl/fl}$*Cebpb*$^{fl/fl}$ mice. These cells were retrovirally infected with C/EBPα p30 constructs and GFP-sorted 2 days post-infection. Sorted cells were seeded in cytokine-supplemented semi-solid medium; they formed visible colonies after approximately 7 days. Colony plates were scanned every 7 days and colonies were classified into CFU-G, CFU-M, CFU-GM, CFU-GEMM and quantified manually by colony counts. For replating, colonies were resuspended to form a homogeneous single-cell suspension; 5000 cells were extracted and replated in fresh semi-solid medium. Replating was repeated every 7 days.

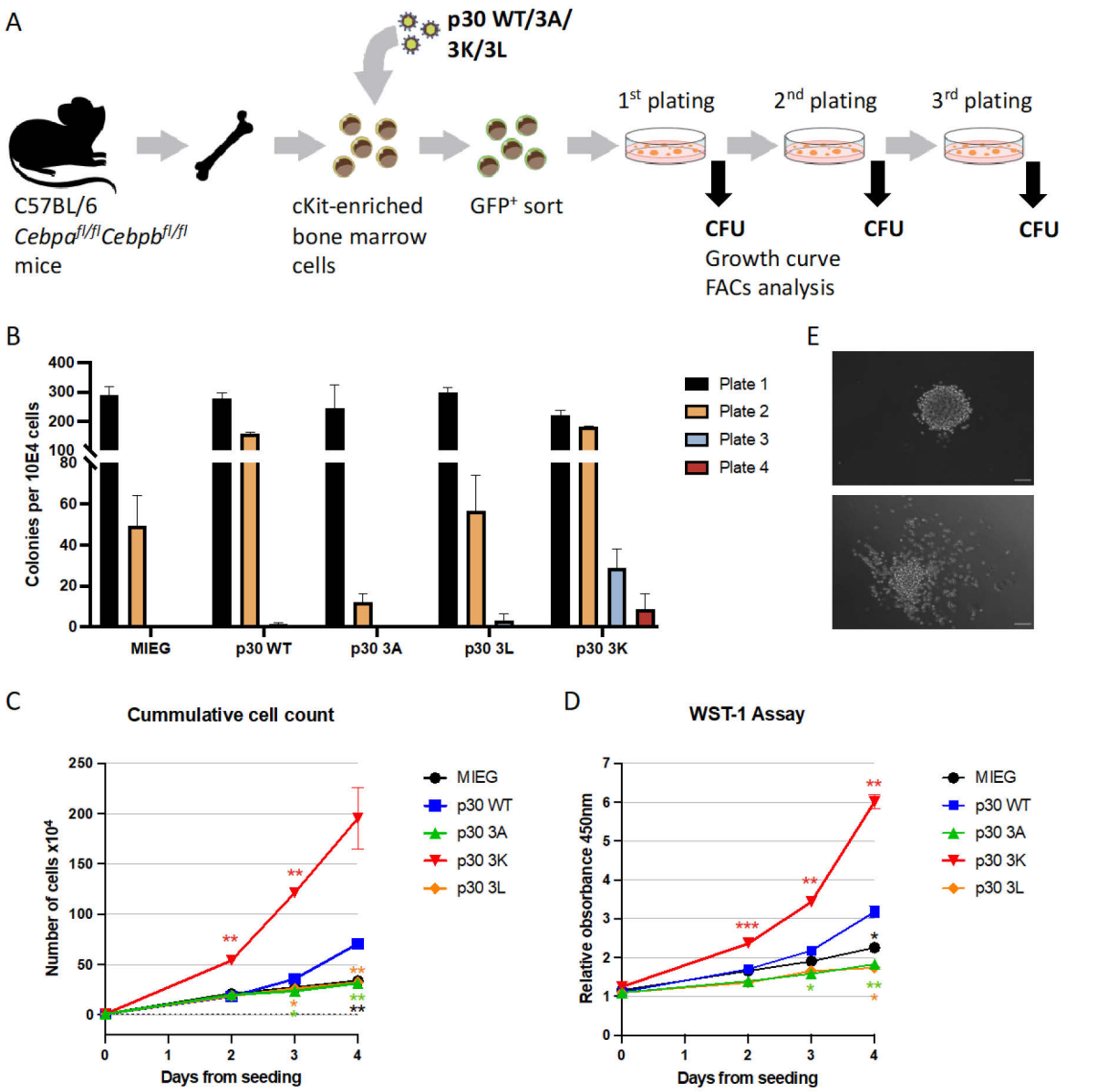


Figure 14. Increased proliferative potential of Arginine-to-Lysin mutated C/EBPα p30 (3K).

(A) Schematic illustration of experimental process. Hematopoietic progenitors Lin^{neg}cKit^{+}Sca-1^{+} were harvested from mouse bone marrow. Retroviral infections were performed at 24 hours after cell harvesting, date of infection was counted as day 0. GFP^{+} cells were sorted at day 2-pi and seeded as described. Colony was counted every 7 days from seeding/replating.

(B) At 7 days after seeding, whole wells containing colonies were scanned. Manual counting was performed on scanned images. Representative result of colony formation rate from one replating experiment with three technical replicate is shown.

(C) and (D) Growth curve measured by cumulative cell count and WST-1 assay using cells from Plate 1 (as shown in (A)). Data are shown as mean ± SEM, significance was determined by two-way ANOVA analysis followed by Dunnett's multiple comparisons test, only significance between p30 WT and other constructs is shown by asterisks in matching color with lines. Results of one independent experiment with three technical replicates are shown.

(E) Representative colony formed by 3K-expressing cells at day 52-pi. Scale bar 100 μm.

Despite forming relatively equal numbers of colonies in the 1st plate, p30 constructs later varied distinctively in replating efficiency (**Figure 14B**). All constructs could form colonies after replating into the 2nd plate, with the highest count by p30 WT and 3K mutant. After the second replating, only 3K could maintain the colony formation through 4 passages, indicating increased self-renewal in the plated progenitors. The pro-proliferative function of 3K cells was confirmed by higher growth rate, as shown by cell counting (**Figure 14C**) and WST-1 assay (**Figure 14D**) when cultured in growth factor supplemented medium, in comparison to the p30 WT. Following long-term cultivation, bone marrow cells gradually ceased growth due to differentiation and subsequent apoptosis (data not shown). Nevertheless, we observed prolonged survival of 3K cells for up to 7 months post-infection, which could continue to form colonies when cultured in semi-solid medium (**Figure 14E**). In contrast, cells expressing p30 3A or 3L showed poor replating efficiency and proliferation, and could not be propagated.

We, again, observed nearly similar effect of 3A and 3L mutant, which lacked the positively charged side chain that the p30 WT and 3K mutant retained. These observations suggested that the enhanced self-renewal and cell growth were induced by the positive charge, accelerated by the unmethylation mimesis (K substituted for R), and inhibited by removing the charge (3A and 3L). This hinted toward enhanced oncogenicity of p30 in its un-/de-methylated form.

3.5. C/EBPα p30 Arginine methylation determined granulocytic/monocytic differentiation

Classification of colonies, as described in 3.3, permitted to evaluate how differentiation was affected by p30 WT or p30 mutants. Types of colony were determined based on morphology, including shape, tightness, size of cells and spreading of cells at colony border (**Figure 15A**) [316]. Among the groups, 3L and 3K mutants presented two opposing effects: growing of CFU-G was more favorable by 3L, while 3K comprised higher numbers of CFU-M than other colony types (**Figure 15B, C, Supplementary Figure 3**).

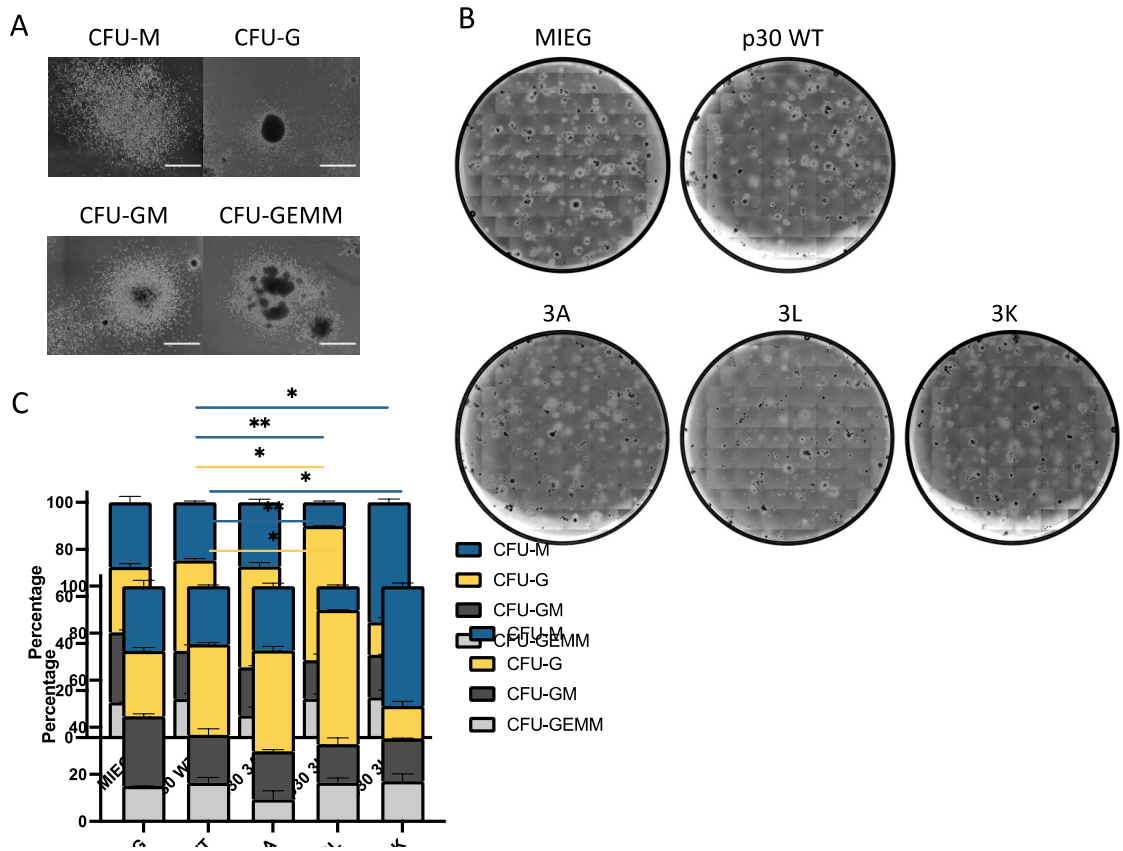


Figure 15. C/EBPα p30 Arginine methylation determined granulocytic/monocytic differentiation.

(A) Morphologic representation of colony categories. Pictures of colonies on Plate 1 of replating assay were taken at day 7 after seeding, before scanning and replating. Scale bar 100 µm. Determination of colony types was based on recommendation of Thermo Fisher Scientific (provider of methylcellulose based medium M3434 for colony formation assay used in this experiment).

(B) Distribution of colonies in replating wells.

(C) Percentages of primary colonies in each category. Manual counting and categorizing of colony type were performed on scanned images of the first colony plates (Plate 1, as in Figure 14). Representative results from one experiment with three technical replicates are shown. Data are shown as mean ± SEM, significance was determined by two-way ANOVA analysis followed by Dunnett's multiple comparisons test. Only significant comparison between p30 WT and other constructs is shown, color represented colony type in the comparison.

Flow cytometric analysis of cells emerged in the first plate (Plate 1) revealed a larger population of $Ly6G^+$ neutrophils by 3A and 3L, in comparison to p30 WT and 3K (**Figure 16A, B**). On the other hand, 3K and p30 WT gave rise to more $Ly6C^+CD115^+$ monocytes, in line with the colony classification results. Interestingly, we also noted a strong accumulation of $Ly6C^{neg}$

cells by the p30 3K mutant, which appeared to express CD16/32 (FcγRIII/II) in later replicates (**Figure 16C, Supplementary Figure 3**). The Ly6C^{neg} populations minimally overlapped with the CD11b^{+}CD115^{+} population (less than 10%) (**Figure 16D, Supplementary Figure 3**), excluding the possibility that these cells were Ly6C^{-} monocytes [304]. This phenotype hinted that these cells remained undifferentiated and had GMP identity, however, expression of c-Kit and CD34, two important progenitor markers, were not properly recorded in these experiments, thus, necessitating further examination of the identity of the p30 3K Ly6C^{neg} cells.

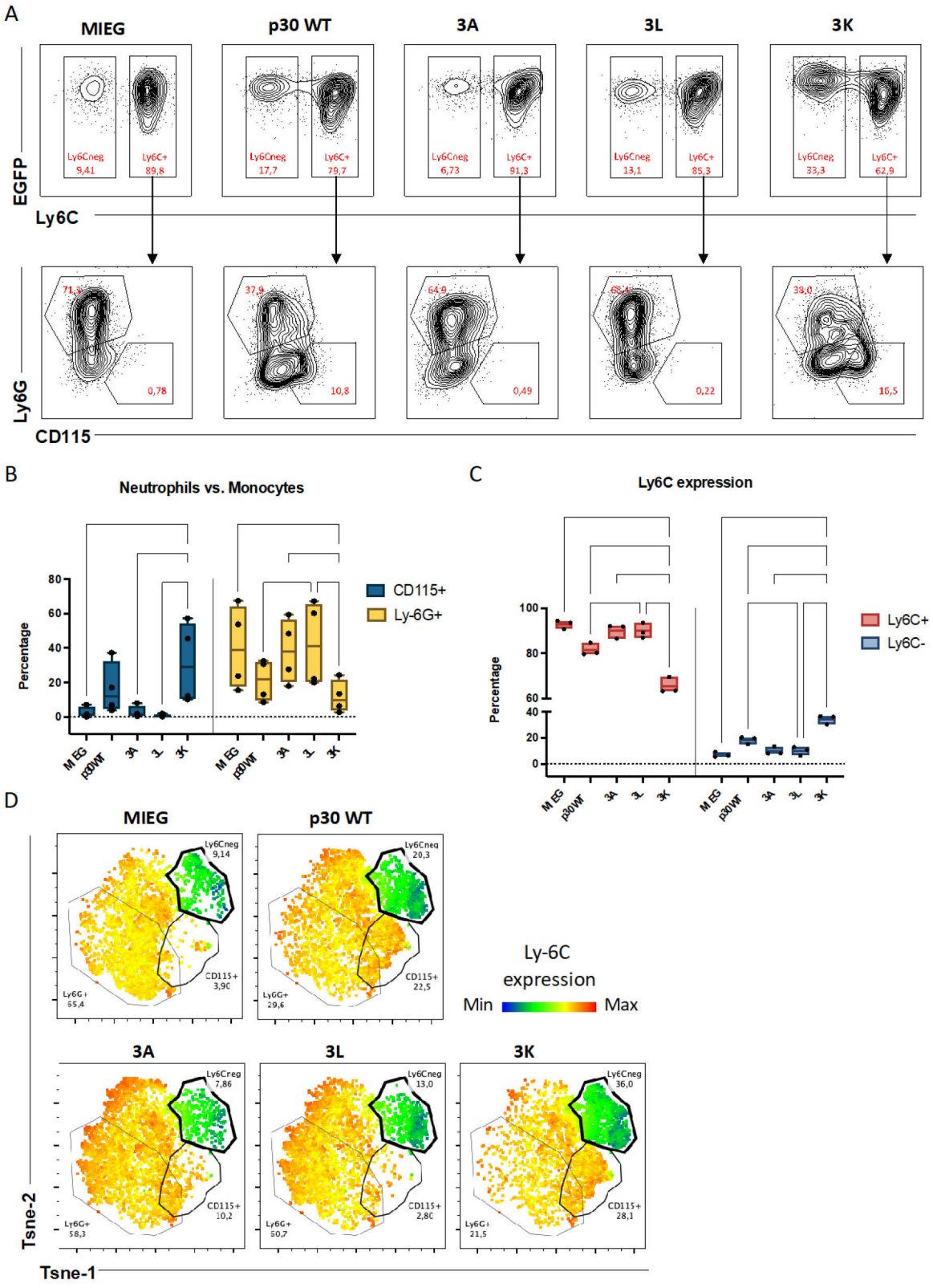


Figure 16. Cells expressing p30 3K enhanced monocytic differentiation and Ly6C^{neg} accumulation.

(A) FACs analysis of cells pooled from colony-assay Plate 1. CD11b^{+} cells were gated before subjected to Ly6C expression. Ly6C^{+} cells were further subjected to neutrophil (Ly6G^{+}) or

monocyte (CD115$^+$) determination. Percentage of each population was used to compose graphs in (B) and (C). Representative result of one experiment is shown.

(B) Percentage of CD115$^+$ monocytes and Ly6G$^+$ neutrophils in 3 independent experiments. Data are shown as mean ± SEM, significance was determined by two-way ANOVA analysis followed by Dunnett's multiple comparisons test. Only significant comparisons are shown.

(C) Percentage of Ly6C^{neg} and Ly6C$^+$ cells 3 independent experiments. Data are shown as mean ± SEM, significance was determined by two-way ANOVA analysis followed by Dunnett's multiple comparisons test. Only significant comparisons are shown.

(D) t-SNE blots of cells analyzed in (A). Heatmap showed expression level of Ly6C marker, Ly6C-negative cells were marked in thick line gating. Ly6G$^+$ and CD115$^+$ population were also shown. From all FACs recorded events, 5000 events of each sample were randomly sampled using DownSample plugin of Flowjo. Further sample-concatenation and dimensional reduction was applied by t-SNE tool (Flowjo). Gating was determined based on unstained and fluorescence-minus-one controls. Representative result of one experiment is shown.

Nonetheless, the myeloid differentiation toward granulocyte/neutrophil or monocyte/macrophage was markedly influenced by C/EBPα p30 Arginine mutations (which resemble various methylation states). Neutrophil fate was more favorable following the overexpression of the p30 3A null mutant or methylation mimesis p30 3L, which both lacked the positively charged group. Monocyte fate, instead, was promoted by p30 WT and even to a larger extend by the unmethylated p30 3K, which maintained the charged side chain. The charged property of an amino acid plays important role in structure and function of protein, since it defines interaction among differently charged amino acids or between amino acid and non-protein molecules [307]. Therefore, we hypothesized that due to the charge property of arginine residues on C/EBPα p30,

(i) the p30 mutants differentially formed interactions **with binding partners**, thus leading to distinct fate choice reflected by different transcriptional profiles;

(ii) the p30 mutants might interact differently **with other parts of p30 molecule**, thus forming unique structures that lead to distinct functions.

Although methylation of arginine residues does not alter the charge properties of arginine sidechains, methyl groups still affect bulkiness and hydrophobicity of the amino acid, thus vary the charge-based interactions. These possibilities will be discussed further in the Discussion section.

4. DISCUSSION

4.1. Transcription activation by the truncated C/EBPα p30

We investigated how the C/EBPs structure alters hematopoietic lineage instruction, using and refining a previously established lymphoid to myeloid transdifferentiation system (LMT). This system is suitable to study the myeloid differentiation potential of C/EBP members [166], among which, C/EBPα is a robust (trans)differentiation factor that impairs tumorigenicity of leukemia/lymphoma cell lines [317]; and by acquiring mutations, C/EBPα can become an oncogene [202]. The oncogenic C/EBPα truncated isoform p30 was initially believed to be dominant-negative. However, recent studies proved otherwise: C/EBPα p30 was shown to contain residual transdifferentiation capacity.

The research groups of Nerlov and Porse, who have studied various p30 mouse models, reported the capability of *Cebpa*$^{p30/p30}$ or *Cebpa*$^{\Delta/p30}$ hematopoietic precursors to commit to the GMP stage [221], [226]. Our observations are in line with these finding, and furthermore, show differentiation capacity of p30 not only during development but also in switching of the B cell fate to new identities. Phenotypic and transcriptional suppression of the B cells program, as well as activation of myeloid program, are both induced by ectopic expression of *Cebpa p30*, without requirement of endogenous *Cebpa p42* or *Cebpb* (**Figure 7C**).

How does p30 activate myeloid programs? The transactivation domain TAD2, which is contained in both C/EBPα p42 and p30, was shown early on to be critical for the interaction between C/EBPα and the SWI/SNF remodeling complex [318]. In immunoprecipitation assays, protein constructs lacking TAD2 (termed TE-III in the cited publication) failed to bind SWI/SNF core subunits (Brm and BAF155), while the intact C/EBPα p42 and p30 showed co-precipitation. This SWI/SNF recruitment by TAD2 was also found to be critical for fibroblast-to-adipocyte conversion, as well as the activation of myeloid transcription in non-myeloid tissue [318]. Interestingly, growth-arrest function of p42 is strongly disrupted upon either removal of TAD2 or deficiency of SWI/SNF [97]. Cells expressing p30, on the other hand, showed undisturbed cell growth, indicating that the proliferation control requires further conditions provided by the TAD1 domain, which is not contained in p30. These studies suggest that the interaction with SWI/SNF via TAD2 is necessary for both differentiation and proliferation control by C/EBPα.

C/EBPα p30 function has been recently studied and concluded to retain distinctive transcriptomic and interaction profile [231], [292]. Although most C/EBPα interactors defined by PRISMA and BioID interact with both p42 and p30, 7 proteins were found to exclusively bind to the p30 isoform. Four out of the 7 proteins proved influential to survival in AML cell lines, including the SWI/SNF subunit BCL11A (BAF Chromatin Remodeling Complex Subunit BCL11A). C/EBPα p30-specific transcriptomes was also identified and appeared to be highly conserved between human and mice [231]. Tumor-promoting genes are selectively induced by p30, for example, *SOX4*, *TYROBP* and *HCK*, which may imply a more complex mechanism of p30-induced leukemogenesis, than the widely assumed lack of E2F-interaction via TAD1. Our analyses, on one hand, confirm the transcription activation function of C/EBPα p30 (**Figure 10E**), and on the other hand, add more possible p30-targets for further considerations.

Studies of enhancer binding activity of C/EBPα isoforms revealed pioneer binding of p42 to "closed" chromatin regions [187], meanwhile, p30 can access only open (H3K4me1) and active (H3K27ac) chromatin regions [231]. These finding may explain the lower transdifferentiation activity by p30 in B-dKO cells in comparison to B-WT cells in our study (when comparing B-dKO cells to B-WT cells, LAP1-induced and p42-induced transdifferentiation was 2x less efficient, while p30-induced transdifferentiation was nearly 10x less efficient (**Figure 6C, day 6 points**)). Our data therefore show that endogenous C/EBPα and C/EBPβ accelerated transdifferentiation; and that p30, without the pioneer activity, only weakly induced transdifferentiation. C/EBPβ may replace C/EBPα in the liver, when *Cebpb* gene was expressed from *Cebpa* locus [319]. SWI/SNF recruiting function that resides in the N-terminus of C/EBPβ has been shown to functionally compensate loss of TAD2 in C/EBPα in the activation of myeloid gene expression [318]. The redundancy between the two C/EBPs may be account for leukemogenicity in *CEBPA*-dm AML context, that the loss of C/EBP p42 could be compensated by C/EBPβ to prime myeloid commitment while p30, which could only access active chromatin regions, follows the lead of C/EBPβ and largely blocks terminal differentiation.

4.2. C/EBPα p30-induced differentiation/proliferation: arginine methylation may tip the balance.

We discovered a PTM dependent regulated activity of p30, which has not been recognized before. As such, methylation of arginine residues on the N-terminus of p30 may facilitate transdifferentiation from B- to myeloid cells, while un-methylated arginine residues may favor proliferation and an undifferentiated state.

Our findings are in agreement with the current interaction mapping of C/EBPα (PRISMA and BioID) [292], in which, p30 R142/149/156 (equivalent to rat R140/147/154) showed methylation dependent binding to SMARCE1, TRIM33, BAF/SWI/SNF subunits (ARID1A/1B/2) and Myb-Muvb/DREAM complex members (LIN9, LIN37, MYBL2). These factors were previously shown to be involved in hematopoietic differentiation: TRIM33-deficient mice show accumulated undifferentiated granulocytic-monocytic progenitors [320]; deficiency of ARID1A blocks granulocytic differentiation in human leukemic cells [321]; Arid2 knock-out cells enhance signatures of myeloid multipotent progenitors [322]; loss of B-Myb (MYBL2) triggers death of myeloid progenitors [323]. These details point toward regulative processes of cell differentiation, which is induced or enhanced by p30 R-methylation. Especially, the occurrence of the SWI/SNF remodeling complex subunits supports our concept that C/EBPα transdifferentiation, which required genes activation by SWI/SNF, is R-methylation dependent.

Interaction between C/EBPα and members of SWI/SNF remodeling complex has been reported previously. On both C/EBPA p42 and p30, the transactivation domain 2 (TAD2, termed TE-III in cited publication, harbors R140/147/154) directly interacts with SWI/SNF components (Brm, (Brahma/SMARCA2) and BAF155) to activated SWI/SNF-dependent genes involved in myelopoiesis and adipogenesis [318]. Brm, together with SNF5 (or SMARCB1), was found interacting with C/EBPα p42 and p30 in another study [97]. These studies strongly support the role of C/EBPα-SWI/SNF interplay, how these interactions are regulated by C/EBPα PTMs has not yet been examined. The TAD2 SWI/SNF recruiting domain can be replaced by a similar function SWI/SNF recruiting domain CR1 on C/EBPβ [318], and this C/EBPβ-SWI/SNF (Brm, BAF155, BAF47) interaction was also shown to be R3-methylation sensitive [324]. Accordingly, we speculate that methylation of the targeted arginine residues

residing in TAD2 alters the interaction with SWI/SNF complexes, yielding diverse outcomes. In fact, arginine methylation has strong implication on SWI/SNF-mediated gene expression. Recruitment of chromatin remodelers are usually associated with histone modifying enzymes, including the PRMTs (protein arginine N-methyltransferases). PRMT5 catalyzes H3R8me2, thus instigate binding of SWI/SNF subunit Brg1 to target genes' promoters; this was found important for gene activation during adipogenesis and myogenesis [325], [326] or repression of tumor suppressors [327]. PRMT7 mediates H3R2me2s, thus induces binding of Brg1 and BAF subunits to promoter of DNA repair genes [328]. Arginine methylation on non-histone proteins also affects their SWI/SNF interaction. PRMT7 methylates R70 on p38MAPK, allowing p38MAPK to recruit SWI/SNF subunits Brg1, BAF160a to target promoters important for myoblast differentiation [329]. C/EBPβ R3 methylation regulates its interaction with SWI/SNF subunits, as described in previous paragraph [324].

The unmethylated arginine mimesis (K mutant), on the other hand, fails to efficiently induce the myeloid program, yet accelerate growth rate and self-renewal of hematopoietic progenitors (**Figure 12**). The role of arginine methylation in cell cycle control, with relation to PRMTs activities, has been widely studied. Although most of the PRMTs are found upregulated in various types of malignancy, suggesting that methylated arginine is associated with enhanced proliferation, various studies also describe otherwise [330]. For instance, methylation at R55, R73, R82, R163 of CDK4 (Cyclin-dependent kinase 4) by PRMT1 destabilizes CDK4-cyclinD3 binding, thus preventing proliferation and prompting differentiation of pre-B cells [331]. Asymmetric methylation at R109 of E2F-1 by PRMT1 strongly induces apoptosis and inhibits cell growth; symmetric methylation at R111, R113 by PRMT5 favors proliferation [332]. Methylation at R754 of p300 by CARM1 (Coactivator-associated arginine methyltransferase 1) modulates the recruitment of BRCA1 to p21 promoter, which subsequently inhibits cell cycle progression [333]. Expression of p21, which is dependent on p53, was also attenuated when methylation at R213 of p53 was blocked by R-to-K mutation [334]. In these studies, arginine methylation either directly affects the major components of the cell cycle progression machinery (e.g., CDK4, E2F-1) or alters the activation of genes important for proliferation (e.g., p21 expression by p300). Arginine methylation of C/EBPα p30 may also follow those mechanisms. C/EBPα p42 has been known for its growth inhibition function, which is orchestrated by

(i) interaction with SWI/SNF complex via TAD2 [97],

(ii) direct repression of E2F complexes via TAD1 [82], [94],

(iii) inhibition of Cdk2 and Cdk4 via TAD2 (amino acid 175-188) [335],

(iv) stabilization of cyclin-dependent kinase inhibitor p21 via TAD2 [96].

In case of C/EBPα p30, the lack of TAD1 represents a fundamental condition for leukemic transformation and will not be further discussed [point (ii)]. Interpretation of p30-SWI/SNF complex [point (i)] was described in 4.1. Regarding point (iii), our targeted arginine residues are located out of the Cdk2 and Cdk2 interaction site; therefore, the mutants and their effects on cell differentiation/proliferation presumably do not directly interfere with p30-Cdk2/Cdk4 interaction. Considering the enhanced cell growth by p30 3K mutant, we surmise that interaction with p21 [point (iv)] at arginine residues may account for this enhanced cell growth. Although C/EBPα binds promoter and induces expression of p21, when comparing *Cebpa*$^{+/+}$ and *Cebpa*$^{-/-}$ newborn mice livers, mRNA level of p21 remained unchanged while protein level varied greatly [96], [336], indicating that C/EBPα translationally/post-translationally regulates p21. The importance of stabilizing p21 to C/EBPα function was challenged by a study by Müller et al., who showed that C/EBPα inhibits cell cycle progression even in *p21*$^{-/-}$ embryonic fibroblasts [337]. Stabilizing p21 may therefore not be a crucial pathway of C/EBPα growth arrest, yet loss of C/EBPα associates with degradation of p21 and promoted cell growth. In case of C/EBPα p30, a role in cell cycle regulation may rely on the weak transactivation domain TAD2, which interacts with p21. Discovering interaction pattern of C/EBPα p30 variants with p21, or SWI/SNF complex as described above, will therefore be part of our future work.

The mechanism behind diverse functions of C/EBPα p30 upon Arginine methylation will require further examinations. Our transcriptomic analyses revealed early and late target genes of each p30 variants; however, one of the difficulties in interpretation arose in p30 3K group, which showed minimal changes in gene expression. Nevertheless, we identified two interesting genes, which might hint at a mechanism of p30 variant activities.

- Among few upregulated genes, we found early induction of *Pim1* (Serine/Threonine-Protein Kinase Pim-1) by 3K mutant (**Figure 12B**). *Pim1* is a proto-oncogene; Pim1 kinase binds and phosphorylates cyclin-dependent kinase inhibitors, including p21 and p27 [338], [339]. Phosphorylation of p27 by Pim1

increase nuclear export and degradation of p27, promoting cell cycle progression. Similarly, phosphorylation of p21 by Pim1 re-localized p21 to cytoplasm, instead of the nucleus. Interestingly, when being re-localized into the cytoplasm after monocytic differentiation, p21 plays role as an anti-apoptosis factor by inhibiting stress-activated responses [340]. These finding support our idea of p30-(Pim1)-p21 regulation: speculatively, 3K mutant may induce *Pim1* expression, leading to p21 re-localization into cytoplasm; this, on one hand, release cell cycle progression in the nucleus, on the other hand, enhance survival of HSC. Further protein-protein interaction analysis will be helpful to test this speculation.

- *Phf10* (PHD Finger Protein 10) is found exclusively expressed in the 3K group at day 4 post-infection (**Supplementary Figure 2**). BAF45a, encoded by *Phf10*, is a non-core subunit of the SWI/SNF-like polybromo-based PBAF chromatin remodeling complex. BAF45 is crucial for the maintenance of long-term HSC and exist at lower level in myeloid progenitors (CMP and GMP) but not in mature myeloid cells, as well as cells of other lineages [341]. Expression of *Phf10* was highest in 3K (**Supplementary figure 1**) and did not follow the trend of *Itgam*/CD11b expression (the CD11b bias, **Figure 12**), suggesting that this is an exclusive target of 3K – the mimesis of unmethylated p30 C/EBPα. PHF10 was found to bind to MYC and to cooperatively activate proliferation [342]; meanwhile, Myc represses differentiation-induced expression of p21, hence, promoting proliferation [343].

Further investigations will include more thorough analysis of the early (day 2) transcriptional profiles of p30 mutant expressing cells, besides studies of the interaction pattern of p30 variants using massspectometry-based proteomics analysis.

In experiments using p30 WT and 3K expressing bone marrow cells, we observed a larger $CD115^{+}$ population and, in particular with the p30 3K mutant, accumulated $Ly6C^{neg}$ population. Because bone marrow derived cells used in these experiments were isolated from $Cebpa^{fl/fl}Cebpb^{fl/fl}$ mice, we considered the activity of endogenous C/EBPα p42 when explaining our observations. In a study of granulocytic-monocytic differentiation under the governing of CEBPA and PU.1, Pundhir and Lauridsen et al. described that after GMP stage, granulocytic enhancers establishment were dependent more on CEBPA, while monocytic enhancers showed less CEBPA-binding and more PU.1-binding [187]. Granulocytic vs.

Monocytic fate choice was previously shown dependent on expression level of *Cebpa*: higher C/EBPα level gives rise to more neutrophils while lower level favors monocytic differentiation [166]. Not only level of expression, the isoforms balance is essential for C/EBPα function, since many chromatin regions are found bound specifically by either p42, p30, or both [231]. The concept of context-dependent C/EBPα isoform balance, leading to diverse outcome in both normal and malignant scenarios, was reviewed recently [344]. Along these lines, we anticipate that granulocytic differentiation requires a C/EBPα-with-differentiation-function, which means p42 or modified p30, such as with 3A and p30 3L. On the other hand, p30 WT or 3K overexpression may form abundance of wild type p30/p30 or 3K/3K homodimers that may displace p42 dimers and thus lead to monocytic differentiation. In such case, the question of "whether p30 3A and p30 3L reserve the differentiation potential similar to that of p42 or not" will be addressed in future works.

4.3. Hypothetical model of R-methylation dependent action

One of the big questions is why 3A, which served as a null mutation with no similarity to methylated-R, behaved in quite similar fashion to the R-methylated mimicry 3L. We summarized major characteristic of each p30 mutants, based on explanation in 3.2 and our finding throughout this study, as below (**Table 4**).

Table 4. Properties of p30 mutants

	Positive charge	Methylation resemblance	Amino acid characteristics	Proliferation	Differentiation
p30 WT (R)	+		Charged through polar guanidinium group, interacts with negatively charged groups, also hydrophobic via long side chain	+/-	+/-
Mutant K	+	-	Charged, polar, interacts with negatively charged groups	+	-
Mutant L	-	+	Hydrophobic via long side chain	-	+
Mutant A	-	-	Non-hydrophobic, non-polar, small	-	+

From our observations, it could be hypothesized that the positive charge on arginine was critical for p30 functions related to the cell cycle and undifferentiated state, as only the non-charged substitutions (A and L) could induce myeloid differentiation. We hypothesized that the positively charged side chain of arginine was involved in certain interactions that could be lost upon methylation.

One of the common interactions found in protein structures are cation-π interactions, a noncovalent binding between a cationic group and an aromatic group [345]. The majority of the cation-π interactions are found between adjacent residues in one sequence, and 70% of arginine residues identified from Protein Data Bank had their sidechains situated near an aromatic sidechain [346]. From the fact that interaction between arginine (a cationic amino acid) and tyrosine (an aromatic amino acid) is the most abundant cation-π interaction, we were prompted to look for aromatic amino acids, especially tyrosine (Y), in the C/EBPα p30 protein sequence. Indeed, three tyrosine residues (Y129, Y136, Y145) were found in proximity with the interrogated arginine residues (R140, R147, R154) (**Figure 17A**). Among various possibilities of how these R and Y residues interact, we consider a folding back structure that may involve three cation-π interactions (**Figure 17B**).

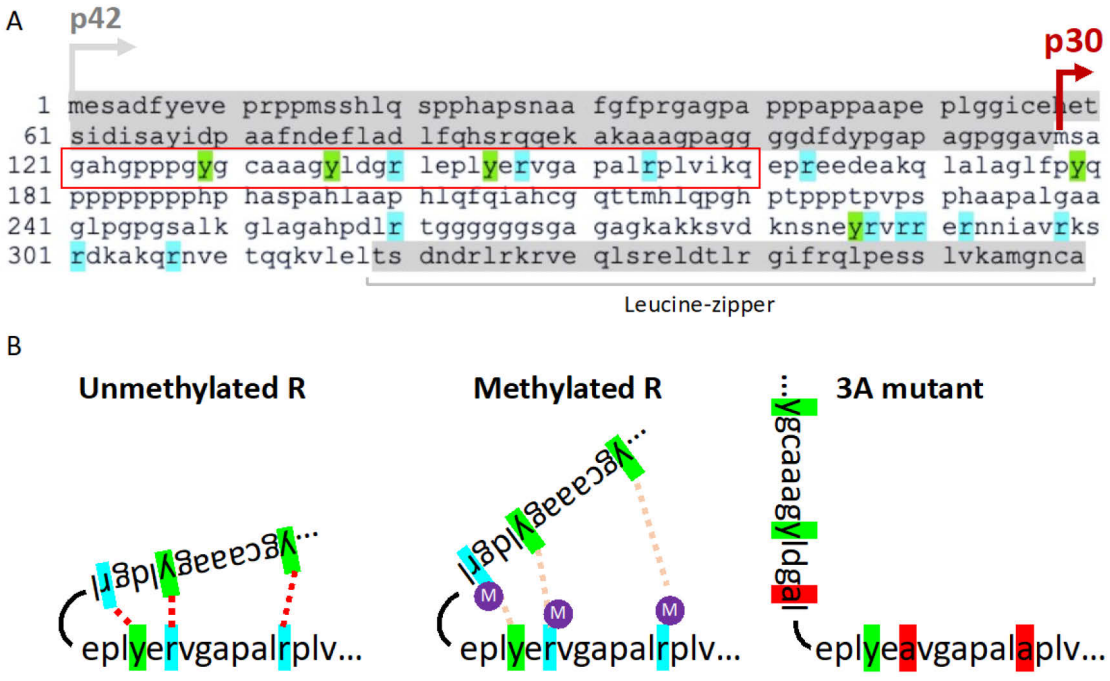


Figure 17. Hypothetical cation-π-based folding structure of C/EBPα p30.

(A) Peptide sequence of C/EBPα. N-terminus of p30 starts at methionine marked by red arrow. Arginine and tyrosine residues are highlighted in cyan and green, respectively. The interrogated arginine residues (R140, R147, R152) are in red box, which also covers three tyrosine residues.

(B) Cation-π interaction could be formed between arginine and tyrosine, allowing the formation of an intramolecular "fold back on itself" structure. Cation-π interaction could be weakened by methylation of arginine, which subsequently released the folding structure and resulted in a low-affinity folded p30 peptide. In case of mutant 3A, alanine is lack of a charged property, cation-π interaction and the folded structure could not be formed. The proposed interactions and structural considerations are based on a hypothesis developed and currently examined by Prof. Achim Leutz.

Many studies have described the chemical nature and biological relevance of the arginine:tyrosine cation-π interactions, and furthermore, in connection with R-methylation. The best example is a recent study by Qamar et al. on FUS (Fused in Sarcoma), an RNA-binding protein often found mutated in familial amyotrophic lateral sclerosis and frontotemporal lobar degeneration [347]. FUS was reported to behave dependently on the cation-π interactions between an arginine-rich structural domain and a tyrosine-rich low-complexity domain. By "behave", the authors indicated the function of FUS to form phase separation (discussed in the next paragraph). Disruption of cation-π interactions by replacing either arginine or tyrosine by alanine (R to A, or Y to A mutations) abrogated phase separation by FUS. Contrastingly, FUS phase separation was maintained when substituting arginine by a cationic lysine (R to K) or substituting tyrosine by an aromatic phenylalanine (Y to F) [347]. Moreover, R-methylation strongly affects phase separation: methylation weakens cation-π interactions and reduces phase separation, while hypomethylation strengthen this interaction and promotes phase separation.

4.3.1. Phase separation

The study by Qamar et al. also provided us a hint toward the question how cation-π interaction may affect binding partners of a protein. A newly found characteristic of intrinsically disordered proteins (IDPs) or proteins with intrinsically disordered regions (IDRs) is that many of them promote liquid-liquid phase separation (LLPS) in cyto-/nucleoplasm [235]. LLPS property of a protein depends on cation-π interaction, among others, and

subsequently alters its interactome. Since the topic of phase separation was not touched in the Introduction, some relevant details are listed below.

- Liquid-liquid phase separation (LLPS) is a de-mixing process of an initially homogeneous solution (liquid de-mixing is a state when two types of liquid do not mix but coexist, think oil and water). Certain protein can transit to and from various material states (from solute to solid state, with intermediate states in between), thus changes the de-mixing state of it surrounding (e.g., cytoplasm). As a result, a phase separated protein may rearrange itself in a certain membrane-less subcellular region; these regions contain certain concentration of said protein. This process changes a homogeneous solution into a solution containing droplets formed by certain proteins. An example of a membrane-less compartment is *C.Elegans*' P granules, which are liquid-like bodies separated from the cytoplasm and contains many proteins and RNAs [348].
- What are the driving forces of LLPS? LLPS is driven by multivalency, which comes from (i) folded proteins contain defined modular interaction domains, which form multivalency by interacting with their associates; (ii) folded domains of proteins connected by linkers, thus form an oligomer of multivalent proteins; and (iii) IDRs that scaffold multiple short linear motifs [349].
- How do IDRs encode LLPS properties? IDRs often contain abundantly amino acids with polar (e.g., glutamine, glycine, serine), charged (e.g., arginine, lysine) or aromatic side chains (tyrosine, phenylalanine). They are usually distributed in short linear interaction motifs that facilitate formation of various inter- or intramolecular interactions, such as electrostatic, π–π, cation–π, or hydrogen bonding interactions. Disruption of such interactions were experimentally shown to perturb phase separation [349], with the abovementioned FUS as an example.
- PTMs appear to be important for LLPS thresholds by changing valency and solubility of proteins. In case of FUS and also Ddx4 (DEAD-Box Helicase 4), arginine methylations were found reducing LLPS by weakening cation–π interaction with tyrosine [347], [350].
- The functional effects of LLPS can result in concentrated amount of resident chemical species and macromolecules, thus affecting movement, binding affinity,

enzymatic activity and even specificity of biochemical processes. For instance, highly concentrated polymer inside a compartment forms a “net” that allows movement of small molecules through spaces between polymer components, but blocks movement of larger molecules, or even immobilizes polymer binding molecules. A phase separated compartment may also concentrate a protein with certain interactors, favoring certain biological pathway while blocking unselective or alternative pathways. At the same time, molecules that are not inside the phase separated compartment may be blocked from accessing activities inside. As an example, T-cell receptor phosphorylation is shown forming liquid-like clusters that selectively concentrate kinases but not phosphatases [351].

Coming back to the raised question, we speculate that cation–π interactions, which are strongly influenced by PTM may play a significant role in regulating proteins interaction and function by alternating LLPS.

4.3.2. C/EBPα and phase separation

LLPS is an emerging topic with plenty of unresolved questions, among which, how to predict phase separation tendency based on protein sequence. In our case, a prediction is needed to find out **whether C/EBPα can phase-separate and whether phase separation properties are altered by PTMs**. Primary structure of C/EBPα was predicted to have high tendency to initiate LLPS at the N-terminus [292]. As described in previous sections, C/EBPα is identified as an IDP which can be heavily decorated with PTMs, and indeed, arginine methylation showed pronounced effect on its function. Together with the developing of LLPS concept, it is reasonable to connect the dots and propose that arginine methylation is critical for a specific structure of C/EBPα p30, which allows distinct functioning via phase separating. This idea is supported by a study on gene activations by transcription factors by Boija et al. [352]. Accordingly, several transcription factors (OCT4, ER, GCN4) were shown forming phase separated condensates with coactivators (Mediator complexes) via IDRs in the activation domains. Those IDRs in activation domains were found crucial for both processes: phase separating and mediators recruiting. Hence, the authors suggest a general model as follow: to activate target enhancers, transcription factors interact with the Mediator; such interaction is fueled by the capacity of activation domains to form phase-separated

condensate. Indeed, p30 R-methylation at R142 (equivalent to rat R140) displayed strong differential binding to many components of the Mediator complex [292], again pointing toward phase separation capacity of C/EBPα.

Assuming that R-methylation regulates C/EBPα p30 function via phase separation, how does p30 phase separation facilitate downstream biological processes? The study by Boija et al. suggested that the LLPS-mediated recruitment of Mediator is fundamental for enhancer activation. Another newly developed model, which fits well with the LLPS model of gene activation, is nucleosome detergent model, suggested by Erkine et al. [353]. The new model is different from the classical "direct recruitment model", which describes that TADs on transcription factors directly form complexes with coactivators (such as Mediators, SWI/SNF) and bring them to target gene's promoter. The nucleosome detergent model proposes that TADs distort the nucleosome structure by disintegrating histones and peeling-off nucleosomal DNA from the histones. In turn, chromatin remodeling complex recruitment is triggered, followed by the recruitment of transcription machinery. Distortion of nucleosomes is caused by "fuzzy" interactions, which form transients electrostatic and aromatic (π–π) bonds between TAD-histone tails or TAD-nucleosomal DNA. The nucleosome detergent model is based on the facts that most TADs-coactivators interactions have low affinity, low specificity, and the TADs do not contain fixed or well-determined sequences, structures and targeting (characteristics of IDRs). A recent machine learning analysis by the same group revealed that most functional TADs contain enriched aromatic and acidic residues, while basic amino acids are detrimental for functionality of TADs [354]. This prediction may explain the lack of transcription activation by the C/EBPα p30 WT and 3K, which contain the basic (positively charged) residues R and K, and the gaining of transcription activation by replacing R by A or L. Moreover, post translational modifications could either facilitate or lessen the interaction with nucleosome components, hence, alter the nucleosome distortion and the subsequent gene activation.

By this logic, the activity of C/EBPα p30 could be summarized as in **Figure 18**. Hypothetically, methylation at critical arginine residues abolished the arginine:tyrosine cation-π interactions, thus giving C/EBP p30 a distinctively folded structure. This structure may permit (i) phase separation leading to Mediator recruitment, and/or (ii) nucleosome

distortion leading to recruitment of chromatin remodeling complexes. These interactions activate transcription of p30's target genes, resulting in myeloid differentiation. Evidences supporting this supposition include stronger binding to SWI/SNF subunits and Mediator to the mutated R142L on C/EBPα p30 [292]. Contrastingly, in the absence of p30 arginine methylation, arginine:tyrosine cation-π interactions may adopt a folded structure that may not have the capacity to interact with nucleosome or other complexes; nevertheless, other interactions could be favored and subsequently induce proliferation. Our speculation of unmethylated p30 seems opposite to Qamar's finding, that arginine hypo-methylation strengthens cation-π interactions, thus promote FUS phase separation [347]. Yet, from their results, it was interpreted that phase separation was initiated by intermolecular cation-π interactions, which form β-sheets by accumulating the low-complexity domains of many FUS molecules together in a restricted space. The authors suggested that alternative intramolecular cation-π interactions (the theoretical "tightly folded structure of p30" falls into this category) might interfere with β-sheets formation and lower phase separation. Our model requires further experimental approaches, which includes analysis of histone or chromatin association, protein-protein interaction, phase separation capacity of p30 variants. In a broader picture, many questions await further investigation, for instance, how is p30 methylated? If only certain amount of p30 is methylated in one cell, how does the cell harmonize the governing of both unmethylated and methylated p30? What is the threshold of methylation that p30 must overcome to induce leukemogenesis? At the upstream of the methylation process, which PRMT methylates arginine residues of p30? How is methylation regulated, in normal myelopoiesis and in leukemia? And finally, is methylation reversible? Answers to these questions may provide insights to the progression of CEBPA-mutated AML, as well as treatment possibility, in addition to mechanistic insights of molecular function of C/EBPs.

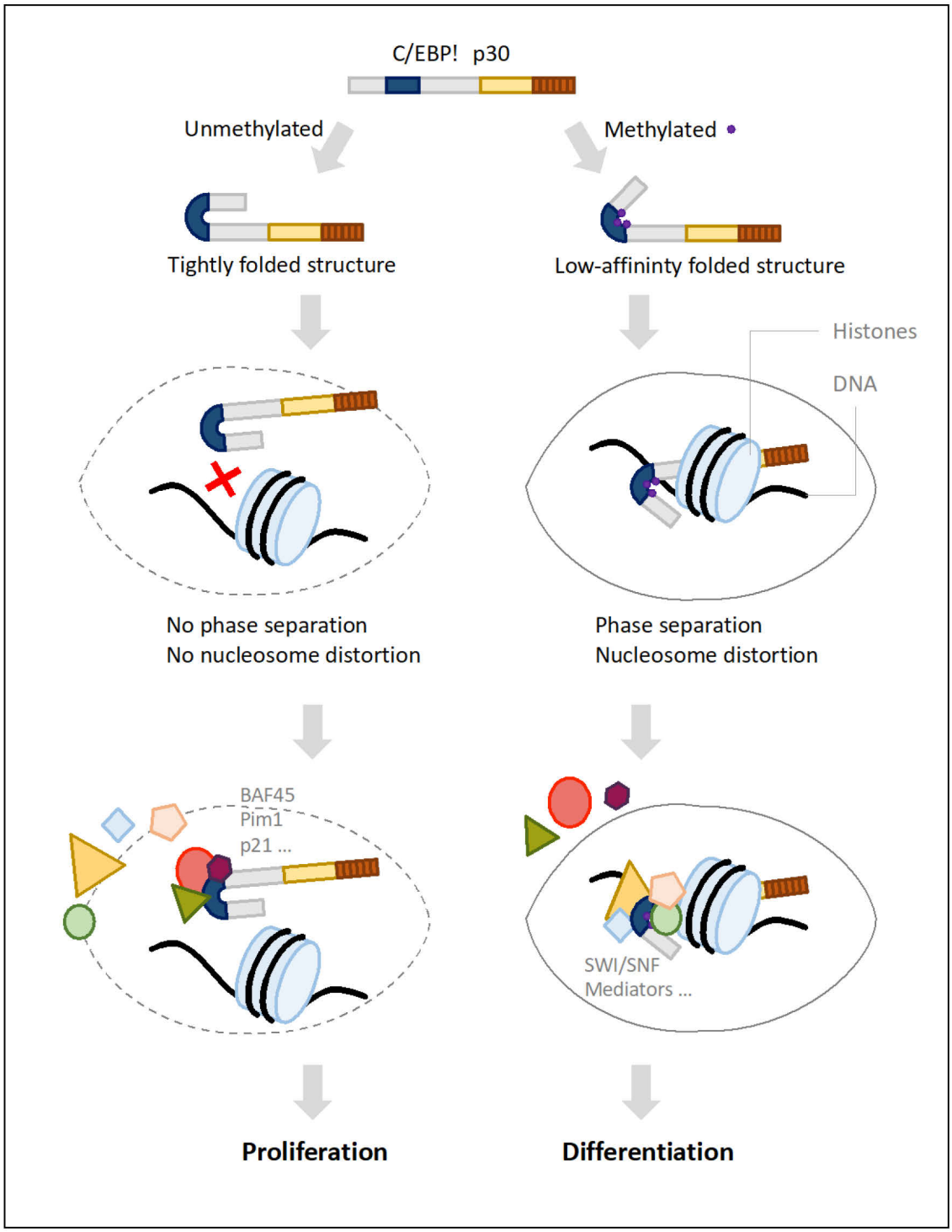


Figure 18. Model of methylation-dependent functioning of C/EBPα p30.
Methylation status of arginine residues on C/EBPα p30 determined the folding structure of the p30 peptide. Low affinity folded p30 allowed gene activation by nucleosome distortion and/or by phase-separation (which favored interactions with SWI/SNF remodeling complexes and mediators, thus supporting gene activation); ultimately induced myeloid differentiation. Tightly folded p30 might not be able to phase-separate or distort nucleosomes, however, could maintain interaction with structure-specific interactors like BAF45, p21 to direct proliferation, or possibly induce expressions of oncogene Pim1.

4.4. Relevance to C/EBPα p30-dependent leukemogenesis

The pro-proliferation effect of p30 3K mutant was only observed when overexpressed in bone marrow cells from *Cebpa*$^{fl/fl}$*Cebpb*$^{fl/fl}$ mice, but not in LMT system which uses B cells. This suggests that proliferation function of C/EBPα p30 is not effective on non-myeloid cells, and a myeloid fated cells is required for p30 transformation. This concept was raised and suggested by prior studies (see 1.2.3). Impaired myeloid differentiation, either by complete loss of C/EBPα *(Cebpa*$^{-/-}$*)* or by C-terminal mutations on both allele (K/K or BRM2) does not lead to AML (**Table 1**) [134], [224], [227]. Only models of N-terminal mutations, namely Δ/Lp30, Lp30/Lp30 (L/L, bi-allelic p30) and Lp30/K (L/K) retain the leukemic transformation capability (**Table 1**) [226]. Serial replating assay using Lin$^-$cKit$^+$Sca-1$^+$ pluripotent HSPCs from L/L mice showed enhanced efficiency only from the 4th replating round, while using GMPs from similar mice showed superior efficiency from the first two rounds [226], suggesting that L/L HSPCs require additional time to reach committed myeloid progenitor stage (GMP) before showing enhanced self-renewability. In our replating assays, the required myeloid fate was possibly induced by endogenous C/EBPα with or without the ally of endogenous C/EBPβ, before the 3K mutant manifested its pro-proliferative effect.

C/EBPα p30 is no longer identified as a negative isoform of p42, but rather a functional isoform with oncogenic features. The oncogenicity of p30 does not only come from the lack of TAD1, but also from p30's unique promoters/enhancers binding distribution and differential interactions with others protein, including co-activators and epigenetic modifiers. We described regulated functions of C/EBPα p30, which involve pro-proliferative effect when arginine residues R140/147/154 remain unmethylated. Self-renewability of HSCs was shown enhanced *in vitro* by unmethylated mimesis of p30 (3K) (**Figure 12**); however, *in vivo* hematologic malignancy testing method is still in the planning state. AML mouse models, including L/L, L/K and K/K were subjected to competitive transplantation experiments into wildtype recipients and together with wildtype competitor cells [226], [227]. Lethal transformation occurs in committed myeloid compartment from all the above genotypes, with accelerated transformation by L/K, delayed and mixed lineage transformation by K/K. However, future transplantation of the p30 unmethylated mimesis (presumably *Cebpa*$^{3K/3K}$) may be incapable of leukemic transformation, due to the lack of differentiation capacity by

p30 3K. Alternatively, using heterozygous *Cebpa*$^{3K/p30}$ might be useful since expression of p30 WT from one allele is adequate to differentiation and may, thus, provide a myeloid platform for leukemic transformation.

It could be anticipated that p30 R-methylation regulates leukemic transformation in a dynamic fashion, which first utilizes the differentiation function (as shown in 3L) to reach GMP stage, and subsequently promotes proliferation. In that event, monitoring activity of protein methylation "writers" (the PRMTs) and "erasers" (demethylases) during p30-driven transformation would provide helpful insights. Prior studies have reported cyclic manner of protein methylation, for instance, H3R17me2a at promoter of *SP2* gene peaks every 20 minutes [355], methylation of ERα (Estrogen receptor alpha) peaks within 5 minutes after ligand binding and lost in 10 minutes afterward [356]. It is widely believed that beside the large number of methyltransferases have been identified, a comparable number of demethylases do exist and remain to be discovered. Regulated methylation and demethylation are mostly studied on histone modification. Up until now, two classes of enzymes were identified as catalyzers of histone lysine demethylation or: the LSDs (Lysine-specific demethylases) and the JmjC family (Jumonji C- terminal domain); both of which could turn methylated histone lysine to unmethylated residues, thus strongly affect gene expression [357]. The appearance and disappearance of methylation marks in the above examples suggest that a process of demethylating histone arginine and non-histone protein does exist, yet genuine arginine-demethylases have not yet been discovered. Methylated arginine residues could be converted to citrulline by a family of peptidyl arginine deiminases (PADs or PADIs). However, this process is considered to have low activity in physiological conditions; although does not directly convert arginine to citrulline, the PADIs compete with PRMTs to antagonize arginine methylation and subsequently prevent gene activation [358][359]. Interaction of C/EBPα and PRMT1 is reported to promote proliferation of breast cancer cells [291], however, interaction of C/EBPα and the PADIs, LSDs and JmjC family remains to be elucidated. A prior study from our group has identified binding of PADI4 (peptidyl arginine deiminase 4) to several sites on C/EBPα, including R297 in the basic domain; disrupting the citrullination by knocking-down PADI4 resulted in granulocytic differentiation [360]. This result encourages our implication that functioning of C/EBPα is regulated by a dynamic methylation-demethylation interplay and requires further examinations.

In the light of clinical the relevance of our findings toward AML research and treatment, we consider some aspects: how methylation marks on C/EBPα arginine changes during transformation, how protein methylation profiles of AML patients differ from healthy individuals, and how to interfere with methylation/demethylation as a treatment option.

- The first question is experimentally approachable, thanks to the available models of CEBPA-mutated AML discussed above, and could be also approached with our lymphoid-myeloid transdifferentiation system. Methylation marks could be traced using specific antibodies on various cell lineages and stages from AML animal models. Alternatively, utilizing inducible expression of p30 in LMT system might allow timely control of the study. However, the requirement of a myeloid platform for leukemic transformation must be met, which demands additional improvement of the system.
- Second question requires global protein methylome analysis methods which are thoroughly reviewed [361]. Generally, global analysis of protein methylome includes combination of enrichment methods (using antibodies against methylated arginine or heavy isotopic labeling) and peptide identification using LC-MS/MS. Deep coverage can be achieved in arginine methylation profiling; however, when applying to analyzing patients' samples, difficulties lie in sample handlings, quality controls, standardized and bioinformatic pipelines, which requires more improvements in the future.
- Regarding the third question, various therapeutic products have been developed targeting the upstream of methylation/demethylation process. Considering that mutant resembling unmethylated C/EBPα p30 showed stronger oncogenicity in our study, inhibition of arginine demethylation should be aimed for. However, both arginine demethylases and inhibitor for the PADIs (arginine deiminase) are yet to be discovered and require more extensive research. Inhibition of lysine demethylase LSD1 was shown to decrease H3K4me2 and induce myeloid differentiation of acute promyelocytic leukemic blasts [362]. Inhibition of JmjC member KDM4A (Lysine Demethylase 4A) also shown anti-cancer benefit [363], [364]. Those studies are encouraging confirmation that protein methylation/demethylation is a promising target in cancer, although they aim at

lysine instead of arginine de-methylation. Alternatively, downstream of methylation/demethylation process may also embody druggable target. In our case, analysis of binding partners or downstream signaling pathways that associated exclusively with different C/EBPα p30 variants is set to be the most recent task. Lastly, together with the emerging topic of phase separation (LLPS), excitement is highly raised over interventions targeting condensate formation [365], [366]. Our implication of C/EBPα p30 arginine methylation – LLPS connection may suggest unconventional approach to control this process (targeting, for example, cation-π disruption, or enzymes modulating LLPS).

4.5. Conclusion and future perspectives

Using the LMT transdifferentiation system and standard hematological analysis, we described regulated functions of the oncogenic C/EBPα isoform p30, which are dependent on arginine methylation. Highlights from this study include:

- C/EBPα p30 is capable of myeloid (trans)differentiation by activating myeloid transcriptional program and suppressing non-myeloid program.
- Arginine residues R140, R147 and R154 are critical for the function of C/EBPα p30. Single mutation at each arginine mildly affects p30's function, while triple mutations strongly altered p30's function.
- Transdifferentiation: alteration of R140, R147, R154 significantly induce transdifferentiation from B cells to myeloid cells. Transdifferentiation capacity is abrogated when replacing residues maintain the charge but remain refractile to arginine methylation.
- Proliferation: in *Cebpa*$^{fl/fl}$*Cebpb*$^{fl/fl}$ bone marrow derived cells, methylated mimesis of p30 (3L mutant) directed myeloid differentiation into granulocytic lineage, rather than monocytic lineage. The p30 3K mutant directed myeloid differentiation into monocytic lineage and at the same time, maintain a population of undifferentiated cells. Unmethylated mimesis of p30 (3K mutant) enhanced replating efficiency of hematopoietic stem/progenitor cells.
- Transcriptional profiling suggested *Pim1* and *Phf10* as potential target gene of 3K mutant.

Mechanisms of methylation-dependent p30 functioning were suggested based on transcriptional profiling and requires further validations. Nevertheless, the findings imply that oncogenicity of C/EBPα p30 can be regulated.

Outlook of this project includes finding of upstream and downstream pathways of C/EBPα p30 methylation. Particularly, interaction of p30 with the PRMTs, as well as p30 functional assays in the present or absents of PRMTs, will be of interest. Downstream events, including exclusive target genes of each p30 mutants and the signaling that they are involve in, will be identified and validated based on our gene expression profiling data at two time points. A strong supportive analysis is being conducted as a parallel project, in which, methylation-dependent binding partners of p30 will be identified using BioID labeling and MS-based analysis.

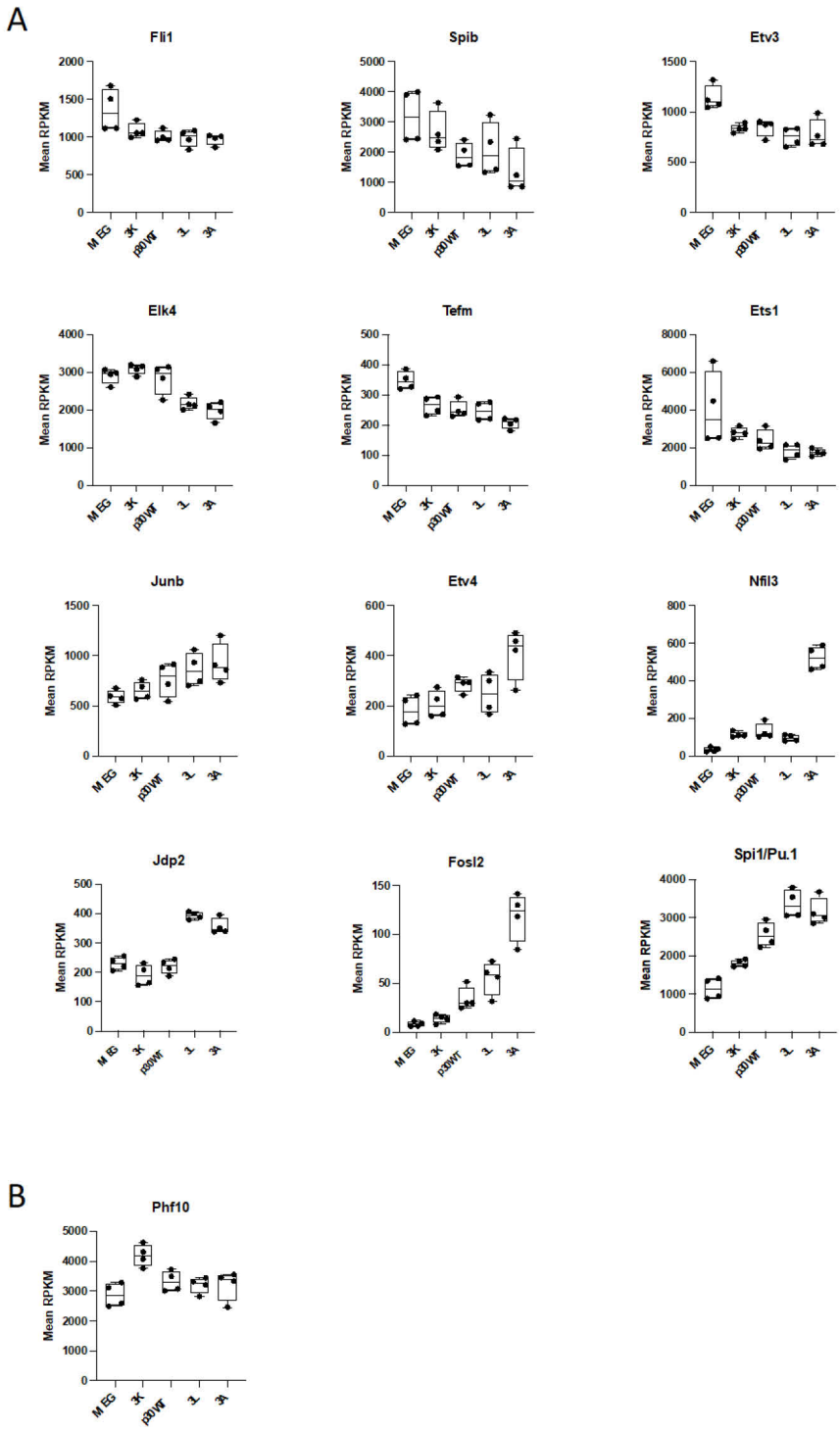
We supposed that the p30 3K mutant, although showing pro-proliferation function, still requires a myeloid platform to incite leukemic transformation. We plan to construct a dual inducible system, in which, C/EBPβ LAP1 is transiently induced to prime B cells into a myeloid fate before p30 3K induction. As such, several issues need to be addressed, such as determination of time between inductions and preventing leakiness of the inducible constructs. One of our main concerns is validation of oncogenicity of the p30 3K mutant, which requires a strategy for adoptive transplantation of 3K expressing cells, yet still provide adequate myeloid differentiation for leukemic transformation.

We are also curious to test our hypothesis of LLPS as described above. Upon purification of C/EBP p30 mutants, the proteins can be subjected to several assay including droplet assay, co-operative mixing experiments or turbidity assay. A nature folding structure of p30 mutant is also interesting to explore.

In closing, our study provided strong evidence of methylation-regulated functioning of the transcription factors C/EBPα p30, which is valuable for characterization of p30-driven AML and may promise future clinical benefit. Further investigations may also connect biological function of p30 to phase separation, a new concept that is re-defining how molecules functionally interact in the cell.

Supplementary Data

Supplementary Figure 1



Supplementary Figure 1. Expression of ETS family members and other noted genes

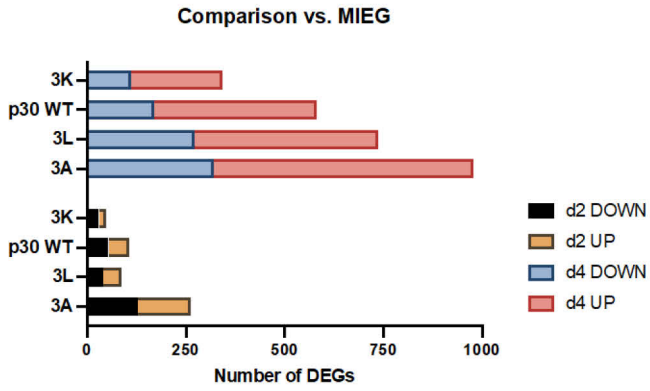
*(A) Relative expression of ETS family members was extracted from RNA-sequencing data (**Figure 10**). The members include:*

- *Down regulated by p30: Etv3 (ets variant 3), Elk4 (ELK4, member of ETS oncogene family), Ets1 (E26 avian leukemia oncogene 1, 5' domain), Fli1 (Friend leukemia integration 1), Spib (Spi-B transcription factor (Spi-1/PU.1 related)), Tefm (transcription elongation factor, mitochondrial).*
- *Upregulated by p30: Etv4 (ets variant 4), Fosl2 (fos-like antigen 2), Jdp2 (Jun dimerization protein 2), Junb (jun B proto-oncogene), Nfil3 (nuclear factor, interleukin 3, regulated), Spi1/PU.1 (spleen focus forming virus (SFFV) proviral integration oncogene Spi1)*

(B) Induced expression of Phf10 (PHD Finger Protein 10) by p30 3K mutant

Supplementary Figure 2

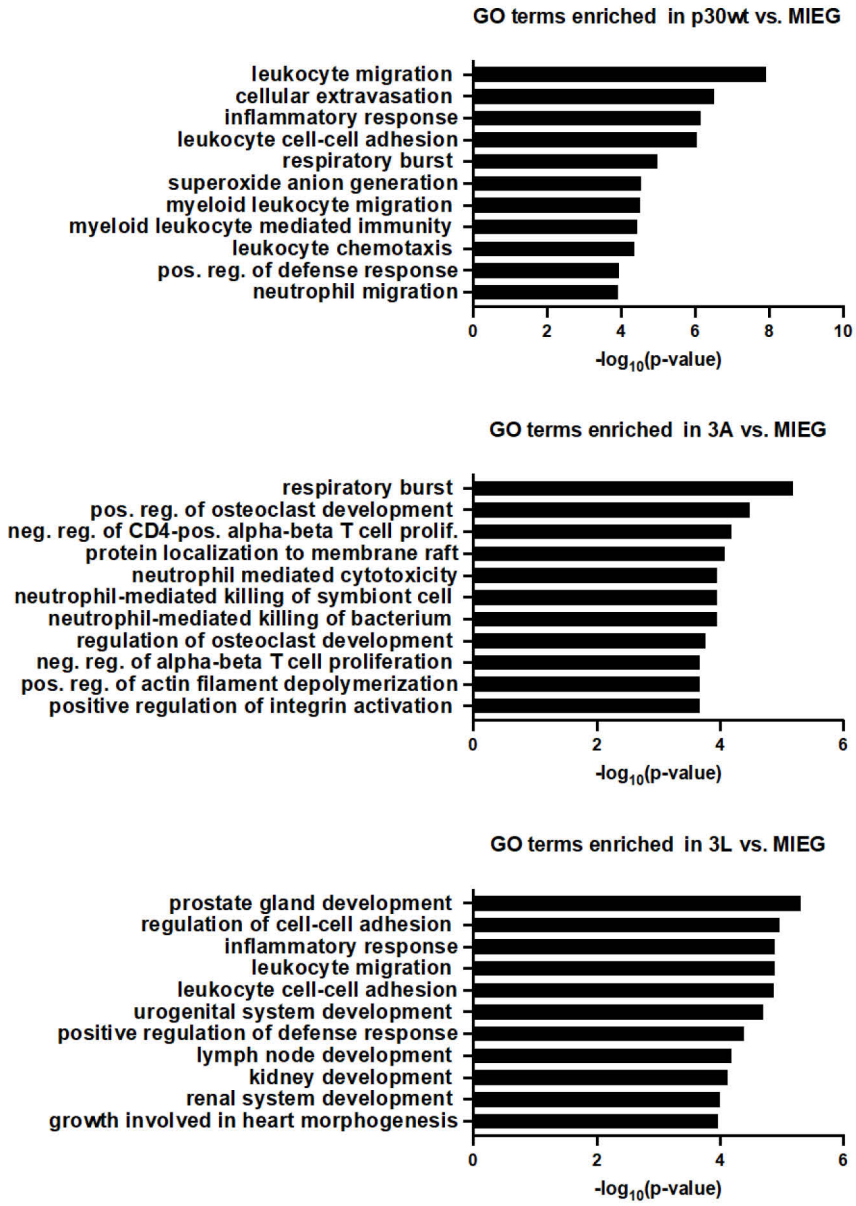
A



Comparison vs. MIEG

	Day 2		Day 4	
	UP	DOWN	UP	DOWN
3A	134	128	657	321
3L	48	40	464	272
3K	17	30	232	112
p30 WT	52	54	412	169

B

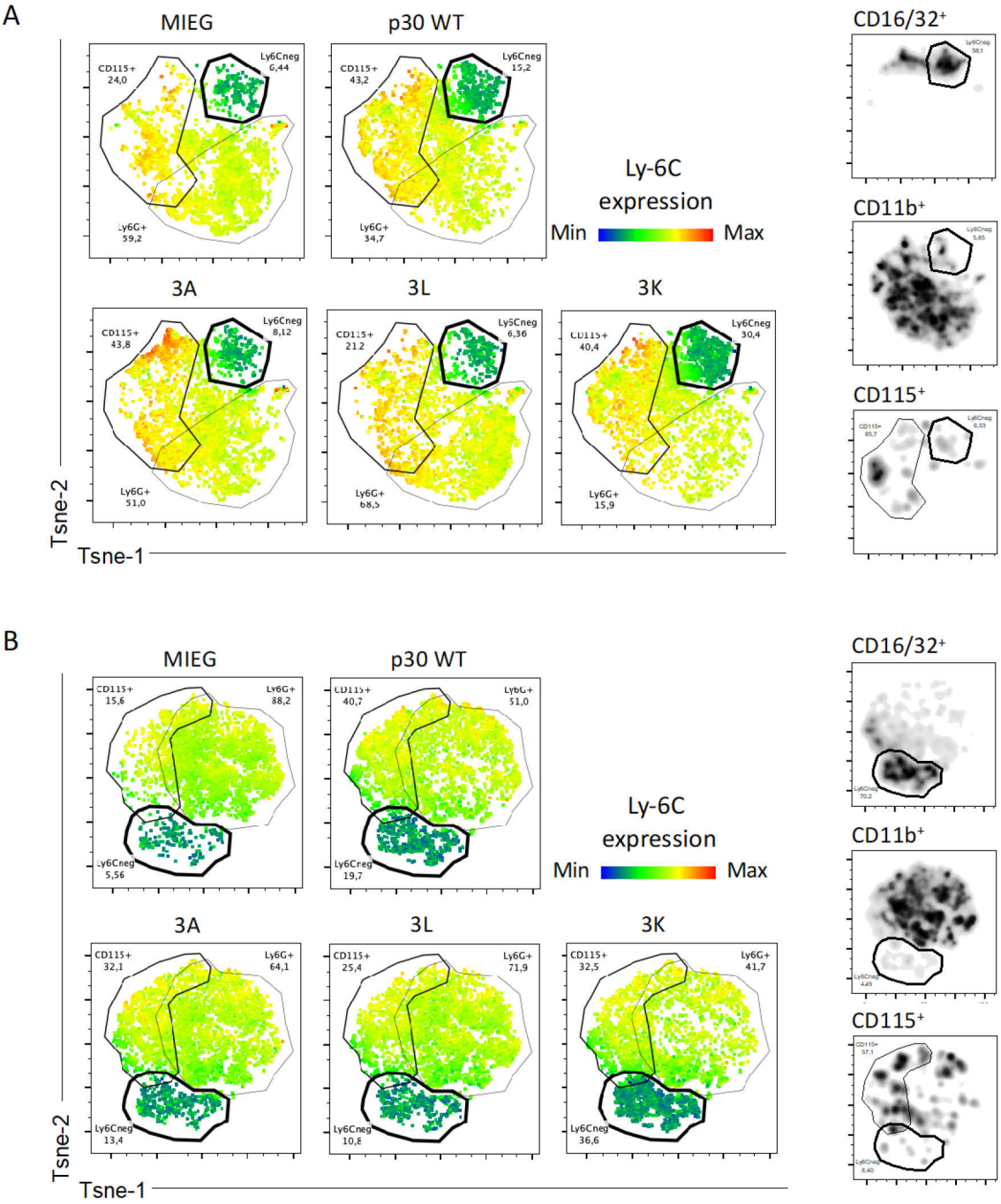


Supplementary Figure 2. Transcriptional profile of methylation dependent transdifferentiation

(A) DEGs between MIEG and p30 variants from both datasets, day 4 (d4) and day 2 (d2) were summed up as shown in the bar chart and table.

(B) Pair-wise GO-terms analysis using DEGs between MIEG and p30 variants in day 2 dataset. Top enriched GO – Biological processes (GO-BP) were shown in order of increasing -log10(p-value). Abbreviations: pos. (positive), neg. (negative), reg. (regulation), prolif. (proliferation).

Supplementary Figure 3



Supplementary Figure 3. Marker expression analysis of bone-marrow derived p30-expressing cells.

Results of other two independent replicates were subjected to t-SNE plots calculation as described in ***Figure 15****. Additional markers (CD16/32, CD11b) and CD115 were gated based on unstained control; positive events were shown in density plots (right columns). For instance, CD16/32⁺ plots represented only CD16/32⁺ gated events, CD16/32^{neg} cells were not shown.*

Supplementary Table 1. Genes upregulated by p30 variants, in comparison to MIEG
*Related to Venn's diagram (**Figure 13A**).*

Intersection	Number	Genes
3A 3K 3L p30WT	6	*Slc8b1, Slpi, Stom, Atf5, Plaur, Ffar2*
3A 3L p30WT	8	*Cd44, Gcnt1, Rflnb, Smim41, Neurl3, Mindy1, Gpr84, Sell*
3A 3K p30WT	3	*Tgm2, Trem1, Milr1*
3K 3L p30WT	1	*Lrrc32*
3A 3K 3L	1	*Aldh3b1*
3A p30WT	14	*Ltf, Ccl6, B430306N03Rik, Lta4h, Mgll, Cd47, Trem3, Arel1, Vsir, Cd33, Rab7b, Ifngr1, Arhgap31, Ncf2*
3L p30WT	2	*Hepacam2, Notch1*
3A 3L	7	*Cyba, Sln, Atf3, Gsto1, Plek, Ahr, Ccn3*
3A 3K	1	*Cebpa*
p30WT	5	*Cpm, Xbp1, Rab27a, Evi2a, Gbp8*
3A	61	*Serpinb1a, Dstn, Cdkn2a, Lcmt1, Dhrs1, Trp53inp2, Ninj1, Cnn2, Clec2i, S1pr3, Prg3, Ubtd1, Dgat2, Ctsz, Sp100, Fam234a, Arrb2, G6pdx, Ly6c2, Cbr3, Hsd11b1, Slc31a2, Acpp, Ltb4r1, Rab3d, Fam117a, Map1lc3a, Rnf130, Litaf, Tuba8, Dedd2, Tyrobp, Snx20, Lyz2, Irf7, Gadd45b, Dgat1, Msrb1, Anxa3, Gas7, Dab2ip, Pnkp, Ero1a, St3gal6, Dhrs7, Tcn2, Hpgds, Csf2rb2, Nfil3, Glipr1, Zbp1, Ier3, Dhrs3, Flot2, Ncf1, C3, Spi1, Ttll9, Gfi1, Lgals3, Fxyd5*
3L	5	*Acadsb, Cox6a2, Cdkn1a, Trib3, Angptl6*
3K	2	*Pim1, Hcst*

Supplementary Table 2. Genes upregulated by p30 variants, in comparison to MIEG.
*Related to Venn's diagram (**Figure 13B**).*

Intersection	Number	Genes
3A 3K 3L p30WT	5	*Arhgef18, Id3, Jchain, Slamf6, Syndig1l*
3A 3K 3L p30WT	3	*Lax1, Cxcr5, Umod*
3A 3K 3L	1	*Hmgn3*
3A 3K 3L	7	*Nupr1, Ncf4, Slc1a4, Frmd5, Lat, Ly6d, Inppl1*
3A 3L	1	*Tnfsf11*
3A 3L	3	*Cd24a, Gpat3, Ell2*
3A 3L	3	*Abcc5, Xrcc5, Tmem86a*
3A	11	*Papolg, Gm8369, Cd2, Mycn, Twsg1, Rragd, Tmem255a, Rnf150, Bcar3, Rasgrp1, Bach2*
3L	9	*Gnb4, Rtl5, Acp6, Lta, Dpp4, Amotl1, Epsti1, Lck, Grap2*
3K	3	*Ddc, Gimap4, Dlx1*

List of Figures

List of Tables

List of Abbreviations

A	alanine
ADMA	asymmetric dimethylarginine
AdoMet	S-adenosyl-L-Methionine
AML	acute myeloid leukemia
ANOVA	analysis of variance
APC	allophycocyanin
B-ALL	B-cell acute lymphoblastic leukemia
B-dKO	*Cebpa*$^{-/-}$ *Cebpb*$^{-/-}$ (double knock-out) B-cell
B-WT	*Cebpa*$^{fl/fl}$*Cebpb*$^{fl/fl}$ (double knock-out) B-cell
BCR-ABL	fusion of BCR (breakpoint cluster region protein) and ABL (Tyrosine-protein kinase ABL1)
Bio-ID	proximity-dependent biotin identification)
BR	basic region
BRM	basic region mutant
bZIP	basic leucine-zipper domain
C-terminus	carboxyl-terminus
C/EBP	CCAAT/enhancer binding protein
CAR-T	chimeric antigen receptor T cells
Cdk/CDK	cyclin-dependent kinase
CFU	colony forming unit
CFU-G	colony forming unit, granulocyte
CFU-GEMM	colony forming unit, granulocyte, erythrocyte, monocyte, megakaryocyte
CFU-GM	colony forming unit, granulocyte, monocyte
CFU-M	colony forming unit, monocyte
ChIP-sequencing	chromatin immunoprecipitation-sequencing
CLP	common lymphoid progenitor
CMP	common myeloid progenitor
CRISP-seq	CRISP-sequencing, perturbed sequencing
DEG	differentially expressed gene
DMEM	Dulbecco's modified Eagle medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
E2F	E2 Factor
EDTA	ethylenediamine-tetra acetic acid

EGFP	enhanced green fluorescent protein
ELK	ETS-like factor
ER	estrogen receptor
ETS	avian erythroblastosis virus E26 (V-Ets) oncogene
ETV	ETS Variant Transcription Factor
FACs	fluorescence activated cell sorting
FBS	fetal bovine serum
FCS	fetal calf serum
FLAG	DYKDDDDK octapeptide
FSC-A	forward scatter area
FSC-H	forward scatter height
FW	forward (PCR primer)
G-CSF	granulocyte colony-stimulating factor
GFP	green fluorescent protein
GM	granulocyte monocyte
GM-CSF	granulocyte-macrophage colony-stimulating factor
GMP	granulocyte-macrophage progenitor
GO	gene ontology
GO-BP	gene ontology - biological process
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSC	hematopoietic stem cell
HSPC	hematopoietic stem progenitor cell
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IL	interleukin
IMDM	Iscove's Modified Dulbecco's Medium
IRES	internal ribosomal entry site
K	lysine
kDa	kilodaltons
L	leucine
LAP1 (C/EBPβ)	liver-enriched transcription-activating protein
LB (medium)	lysogeny broth (or Luria broth or Luria–Bertani medium)
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LLPS	liquid-liquid phase separation
LMT	lymphoid-myeloid transdifferentiation
LT-HSC	long-term hematopoietic stem cell
M	molar
M-CSF	macrophage colony-stimulating factor
m/v	mass/volume

MACs	magnetic-activated cell sorting
MED	Mediator complex
MEP	megakaryocyte–erythroid progenitor
mg	milligram
MIEG	plasmid pMSCV_IRES_EGFP
mL	milliliter
MLL	mixed lineage leukemia
mM	millimolar
MMA	monomethylarginine
MMP	matrix metalloproteinase
MoRF	molecular recognition features
N-terminus	amino terminus
n.s	not significant
ng	nanogram
NGS	next generation sequencing
nM	nanomolar
^{o}C	degree Celsius
OD	optical density
PADI	peptidyl arginine deiminase
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PE	phycoerythrin
PE-Cy7	phycoerythrin-cyanine7
PEI	polyethyleneimine
PerCP-Cy5.5	peridinin chlorophyll protein-cyanine5.5
pH	potentia hydrogenii
pi	post-infection
PI	propidium iodide
PlatE	platinum-E (packaging cell line)
pMSCV	Murine Stem Cell virus plasmid
PRISMA	Protein Interaction Screen on peptide Matrices
PRMT	protein arginine methyltransferases
PTM	post-translational modification
R	arginine
RD	regulatory domain
RNA	ribonucleic acid
RNA-seq	RNA-sequencing
rpm	rotations per minute
RV	reverse (PCR primer)

SAM	S-adenosyl-L-Methionine
SCF	stem cell factor
SDMA	symmetric dimethylarginine
SDS	sodium dodecyl sulfate
SEM	standard error of mean
shRNA	short/small hairpin RNA
SLiM	short, linear motif
SSC-A	side scatter - area
SWI/SNF	SWItch/Sucrose Non-Fermentable complex
T-ALL	T-cell acute lymphoblastic leukemia
t-SNE	t-distributed stochastic neighbor embedding
TAD	transactivation domain
TBS-T	tris-buffered saline-tween
TSS	transcriptional start site
uORF	upstream open reading frames
V	Volt
v-Abl	Abelson murine leukemia viral oncogene
v/v	volume/volume
WST-1	tetrazolium salt WST-1 (use in colorimetric assays)
WT	wildtype
Y	tyrosine
µg	microgram
µL	microliter
µm	micrometer
µM	micromolar

References

[1] R. M. Shallis, R. Wang, A. Davidoff, X. Ma, and A. M. Zeidan, “Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges,” *Blood Rev.*, vol. 36, pp. 70–87, Jul. 2019.

[2] A. Pelcovits and R. Niroula, “Acute Myeloid Leukemia: A Review.,” 2020.

[3] B. Johansson and C. J. Harrison, “Acute Myeloid Leukemia,” *Cancer Cytogenet.*, vol. 373, no. 12, pp. 45–139, Sep. 2010.

[4] D. Bell and F. Gaillard, *WHO classification of tumours of haematopoietic and lymphoid tissues*. International Agency for Research on Cancer, 2010.

[5] E. Papaemmanuil *et al.*, “Genomic Classification and Prognosis in Acute Myeloid Leukemia,” *N. Engl. J. Med.*, vol. 374, no. 23, pp. 2209–2221, Jun. 2016.

[6] “Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia,” *N. Engl. J. Med.*, vol. 368, no. 22, pp. 2059–2074, May 2013.

[7] L. Bullinger, K. Döhner, and H. Dohner, “Genomics of acute myeloid leukemia diagnosis and pathways,” *Journal of Clinical Oncology*, vol. 35, no. 9. American Society of Clinical Oncology, pp. 934–946, 20-Mar-2017.

[8] M. W. Wlodarski *et al.*, “Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents,” *Blood*, vol. 127, no. 11, pp. 1387–1397, Mar. 2016.

[9] B. Jacob *et al.*, “Stem cell exhaustion due to Runx1 deficiency is prevented by Evi5 activation in leukemogenesis,” *Blood*, vol. 115, no. 8, pp. 1610–1620, Feb. 2010.

[10] B. J. Wouters, B. Löwenberg, C. A. J. Erpelinck-Verschueren, W. L. J. Van Putten, P. J. M. Valk, and R. Delwel, “Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome,” *Blood*, vol. 113, no. 13, pp. 3088–3091, Mar. 2009.

[11] H. Takei and S. S. Kobayashi, “Targeting transcription factors in acute myeloid leukemia,” *Int. J. Hematol.*, vol. 109, no. 1, pp. 28–34, Jan. 2019.

[12] H. Kantarjian *et al.*, “Acute myeloid leukemia: current progress and future directions,” *Blood Cancer Journal*, vol. 11, no. 2. Springer Nature, p. 41, 01-Feb-2021.

[13] C. D. Dinardo *et al.*, “Characteristics, clinical outcome, and prognostic significance of IDH mutations in AML,” *Am. J. Hematol.*, vol. 90, no. 8, pp. 732–736, Aug. 2015.

[14] 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, Aug. 2012.

[15] 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, Jul. 2011.

[16] H. Sharma, “Development of Novel Therapeutics Targeting Isocitrate Dehydrogenase Mutations in Cancer,” *Curr. Top. Med. Chem.*, vol. 18, no. 6, pp. 505–524, May 2018.

[17] J. Yu, Y. Li, D. Zhang, D. Wan, and Z. Jiang, “Clinical implications of recurrent gene mutations in acute myeloid leukemia,” *Exp. Hematol. Oncol.*, vol. 9, no. 1, pp. 1–11, Mar. 2020.

[18] T. Alpermann *et al.*, “Molecular subtypes of npm1 mutations have different clinical profiles, specific patterns of accompanying molecular mutations and varying outcomes in intermediate risk acute myeloid leukemia,” *Haematologica*, vol. 101, no. 2. Ferrata

Storti Foundation, pp. e55–e58, 01-Feb-2016.
[19] B. Falini *et al.*, “Cytoplasmic Nucleophosmin in Acute Myelogenous Leukemia with a Normal Karyotype,” *N. Engl. J. Med.*, vol. 352, no. 3, pp. 254–266, Jan. 2005.
[20] H. Becker *et al.*, “Favorable Prognostic Impact of NPM1 Mutations in Older Patients with Cytogenetically Normal de Novo Acute Myeloid Leukemia and Associated Gene- And MicroRNA-Expression Signatures: A Cancer and Leukemia Group B Study,” *J. Clin. Oncol.*, vol. 28, no. 4, pp. 596–604, Feb. 2010.
[21] F. G. Rücker *et al.*, “TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome,” *Blood*, vol. 119, no. 9, pp. 2114–2121, Mar. 2012.
[22] R. Devillier *et al.*, “Role of ASXL1 and TP53 mutations in the molecular classification and prognosis of acute myeloid leukemias with myelodysplasiarelated changes,” *Oncotarget*, vol. 6, no. 10, pp. 8388–8396, 2015.
[23] A. M. Hunter and D. A. Sallman, “Current status and new treatment approaches in TP53 mutated AML,” *Best Pract. Res. Clin. Haematol.*, vol. 32, no. 2, pp. 134–144, Jun. 2019.
[24] C. D. Di Nardo and J. E. Cortes, “Mutations in AML: Prognostic and therapeutic implications,” *Hematology*, vol. 2016, no. 1, pp. 348–355, 2016.
[25] S. Cuartero, A. J. Innes, and M. Merkenschlager, “Towards a Better Understanding of Cohesin Mutations in AML,” *Frontiers in Oncology*, vol. 9. Frontiers Media S.A., p. 867, 09-Sep-2019.
[26] S. Stass, J. Mirro, S. Melvin, C. H. Pui, S. B. Murphy, and D. Williams, “Lineage switch in acute leukemia,” *Blood*, vol. 64, no. 3, pp. 701–706, 1984.
[27] O. Imataki *et al.*, “Lineage switch from precursor B cell acute lymphoblastic leukemia to acute monocytic leukemia at relapse,” *Int. J. Clin. Oncol.*, vol. 15, no. 1, pp. 112–115, 2010.
[28] A. Zoghbi, U. zur Stadt, B. Winkler, I. Müller, and G. Escherich, “Lineage switch under blinatumomab treatment of relapsed common acute lymphoblastic leukemia without MLL rearrangement,” *Pediatr. Blood Cancer*, vol. 64, no. 11, 2017.
[29] F. Perna and M. Sadelain, “Myeloid leukemia switch as immune escape from CD19 chimeric antigen receptor (CAR) therapy,” *Translational Cancer Research*, vol. 5, pp. S221–S225, 2016.
[30] R. Gardner *et al.*, “Acquisition of a CD19-negative myeloid phenotype allows immune escape of MLL-rearranged B-ALL from CD19 CAR-T-cell therapy,” *Blood*, vol. 127, no. 20, pp. 2406–2410, 2016.
[31] R. Ihle, H. Matthes, and H. Ihle, “Phenotype switch in acute leukemia patients after intensive chemotherapy.,” *Haematol. Blood Transfus.*, vol. 32, pp. 101–103, 1989.
[32] E. Dorantes-Acosta, F. Arreguin-Gonzalez, C. A. Rodriguez-Osorio, S. Sadowinski, R. Pelayo, and A. Medina-Sanson, “Acute myelogenous leukemia switch lineage upon relapse to acute lymphoblastic leukemia: A case report,” *Cases J.*, vol. 2, no. 10, 2009.
[33] W. van den Ancker *et al.*, “Uncommon lineage switch warrants immunophenotyping even in relapsing leukemia,” *Leukemia Research*, vol. 33, no. 7. 2009.
[34] B. Wu *et al.*, “Lineage Switch between B-Lymphoblastic Leukemia and Acute Myeloid Leukemia Intermediated by ‘occult’ Myelodysplastic Neoplasm,” in *American Journal of Clinical Pathology*, 2017, vol. 148, no. 2, pp. 136–147.
[35] A. Aujla, M. Hanmantgad, H. Islam, F. Shakil, D. Liu, and K. Seiter, “Lineage switch from T-cell lymphoblastic leukemia/lymphoma to acute myeloid leukemia and back to T-cell lymphoblastic leukemia/lymphoma in a patient diagnosed during pregnancy,” *Stem*

Cell Investig., vol. 6, no. May, pp. 1–6, 2019.

[36] T. Hu, R. Murdaugh, and D. Nakada, "Transcriptional and microenvironmental regulation of lineage ambiguity in Leukemia," *Front. Oncol.*, vol. 7, no. NOV, pp. 1–7, 2017.

[37] Y. Aikawa *et al.*, "Essential role of PU.1 in maintenance of mixed lineage leukemia-associated leukemic stem cells," *Cancer Sci.*, vol. 106, no. 3, pp. 227–236, 2015.

[38] F. Monma *et al.*, "Molecular and phenotypic analysis of Philadelphia chromosome-positive bilineage leukemia: Possibility of a lineage switch from T-lymphoid leukemic progenitor to myeloid cells," *Cancer Genet. Cytogenet.*, vol. 164, no. 2, pp. 118–121, 2006.

[39] E. Jacoby *et al.*, "CD19 CAR immune pressure induces B-precursor acute lymphoblastic leukaemia lineage switch exposing inherent leukaemic plasticity," *Nat. Commun.*, vol. 7, no. May, 2016.

[40] E. H. Davidson, "Emerging properties of animal gene regulatory networks," *Nature*, vol. 468, no. 7326, pp. 911–920, 2010.

[41] H. Iwasaki *et al.*, "The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages," *Genes Dev.*, vol. 20, no. 21, pp. 3010–3021, 2006.

[42] R. Stadhouders, G. J. Filion, and T. Graf, "Transcription factors and 3D genome conformation in cell-fate decisions," *Nature*, vol. 569, no. 7756. pp. 345–354, 2019.

[43] D. Reynaud, N. Lefort, E. Manie, L. Coulombel, and Y. Levy, "In vitro identification of human pro-B cells that give rise to macrophages, natural killer cells, and T cells," *Blood*, vol. 101, no. 11, pp. 4313–4321, 2003.

[44] T. Audzevich *et al.*, "Pre/pro-B cells generate macrophage populations during homeostasis and inflammation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 20, pp. E3953–E3954, 2017.

[45] H. Kulessa, J. Frampton, and T. Graf, "GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts," *Genes Dev.*, vol. 9, no. 10, pp. 1250–1262, 1995.

[46] J. E. Visvader, A. G. Elefanty, A. Strasser, and J. M. Adams, "GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line," *EMBO J.*, vol. 11, no. 12, pp. 4557–4564, 1992.

[47] S. L. Nutt, B. Heavey, A. G. Rolink, and M. Busslinger, "Commitment to the B-lymphoid lineage depends on the transcription factor Pax5," *J. Immunol.*, vol. 195, no. 3, pp. 766–772, 2015.

[48] G. J. Lindeman, J. M. Adams, S. Cory, and A. W. Harris, "B-lymphoid to granulocytic switch during hematopoiesis in a transgenic mouse strain," *Immunity*, vol. 1, no. 6, pp. 517–527, 1994.

[49] C. V. Laiosa, M. Stadtfeld, H. Xie, L. de Andres-Aguayo, and T. Graf, "Reprogramming of Committed T Cell Progenitors to Macrophages and Dendritic Cells by C/EBPα and PU.1 Transcription Factors," *Immunity*, vol. 25, no. 5, pp. 731–744, 2006.

[50] H. Xie, M. Ye, R. Feng, and T. Graf, "Stepwise reprogramming of B cells into macrophages," *Cell*, vol. 117, no. 5, pp. 663–676, 2004.

[51] R. Feng *et al.*, "PU.1 and C/EBPα/β convert fibroblasts into macrophage-like cells," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 16, pp. 6057–6062, Apr. 2008.

[52] L. Slamova *et al.*, "CD2-positive B-cell precursor acute lymphoblastic leukemia with an early switch to the monocytic lineage," *Leukemia*, vol. 28, no. 3, pp. 609–620, 2014.

[53] D. F. Akin, D. A. Oner, E. Kurekci, and N. Akar, *Determination of CEBPA mutations by next generation sequencing in pediatric acute leukemia*, vol. 119, no. 6. Bratisl Lek Listy, 2018, pp. 366–372.

[54] I. H. I. M. Hollink *et al.*, "Characterization of CEBPA mutations and promoter hypermethylation in pediatric acute myeloid leukemia," *Haematologica*, vol. 96, no. 3, pp. 384–392, 2011.

[55] Y. F. Liu *et al.*, "Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia," *EBioMedicine*, vol. 8, pp. 173–183, 2016.

[56] J. Tsukada, Y. Yoshida, Y. Kominato, and P. E. Auron, "The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation," *Cytokine*, vol. 54, no. 1. pp. 6–19, 01-Apr-2011.

[57] B. J. Graves, P. F. Johnson, and S. L. McKnight, "Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene," *Cell*, vol. 44, no. 4, pp. 565–576, Feb. 1986.

[58] W. H. Landschulz, P. F. Johnson, E. Y. Adashi, B. J. Graves, and S. L. McKnight, "Isolation of a recombinant copy of the gene encoding C/EBP.," *Genes Dev.*, vol. 2, no. 7, pp. 786–800, 1988.

[59] W. H. Landschulz, P. F. Johnson, and S. L. McKnight, "The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins," *Science (80-.).*, vol. 240, no. 4860, pp. 1759–1764, 1988.

[60] P. Agre, P. F. Johnson, and S. L. McKnight, "Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP," *Science (80-.).*, vol. 246, no. 4932, pp. 922–926, Nov. 1989.

[61] W. H. Landschulz, P. F. Johnson, and S. L. McKnight, "The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite," *Science (80-.).*, vol. 243, no. 4899, pp. 1681–1688, Mar. 1989.

[62] S. Akira *et al.*, "A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family," *EMBO J.*, vol. 9, no. 6, pp. 1897–1906, Jun. 1990.

[63] S. C. Williams, C. A. Cantwell, and P. F. Johnson, "A family of C/EBP-related proteins capable of forming covalently linked leucine zipper dimers in vitro," *Genes Dev.*, vol. 5, no. 9, pp. 1553–1567, Sep. 1991.

[64] C. Roman, J. S. Platero, J. Shuman, and K. Calame, "Ig/EBP-1: A ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP," *Genes Dev.*, vol. 4, no. 8, pp. 1404–1415, Aug. 1990.

[65] P. Descombes, M. Chojkier, S. Lichtsteiner, E. Falvey, and U. Schibler, "LAP, a novel member of the C/EBP gene family, encodes a liver-enriched transcriptional activator protein," *Genes Dev.*, vol. 4, no. 9, pp. 1541–1551, Sep. 1990.

[66] D. Ron and J. F. Habener, "CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription," *Genes Dev.*, vol. 6, no. 3, pp. 439–453, 1992.

[67] Z. Cao, R. M. Umek, and S. L. McKnight, "Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells," *Genes Dev.*, vol. 5, no. 9, pp. 1538–1552, 1991.

[68] C. J. Chang, T. T. Chen, H. Y. Lei, D. S. Chen, and S. C. Lee, "Molecular cloning of a transcription factor, AGP/EBP, that belongs to members of the C/EBP family," *Mol. Cell. Biol.*, vol. 10, no. 12, pp. 6642–6653, Dec. 1990.

[69] V. Poli, F. P. Mancini, and R. Cortese, "IL-6DBP, a nuclear protein involved in interleukin-

6 signal transduction, defines a new family of leucine zipper proteins related to C EBP," *Cell*, vol. 63, no. 3, pp. 643–653, Nov. 1990.

[70] D. P. Ramji and P. Foka, "CCAAT/enhancer-binding proteins: Structure, function and regulation," *Biochem. J.*, vol. 365, no. 3, pp. 561–575, 2002.

[71] P. Agre, P. F. Johnson, and S. L. McKnight, "Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP," *Science (80-.).*, vol. 246, no. 4932, pp. 922–926, 1989.

[72] W. H. Landschulz, P. F. Johnson, and S. L. McKnight, "The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite," *Science (80-.).*, vol. 243, no. 4899, pp. 1681–1688, 1989.

[73] J. A. Listman *et al.*, "Conserved ETS domain arginines mediate DNA binding, nuclear localization, and a novel mode of bZIP interaction," *J. Biol. Chem.*, vol. 280, no. 50, pp. 41421–41428, Dec. 2005.

[74] T. H. Tahirov *et al.*, "Mechanism of c-Myb-C/EBPβ cooperation from separated sites on a promoter," *Cell*, vol. 108, no. 1, pp. 57–70, Jan. 2002.

[75] C. Cooper, A. Henderson, S. Artandi, N. Avitahl, and K. Calame, "Ig/EBP (C/EBPγ) is a transdominant negative inhibitor of C/EBP family transcriptional activators," *Nucleic Acids Res.*, vol. 23, no. 21, pp. 4371–4377, 1995.

[76] J. G. Tang and H. P. Koeffler, "Structural and Functional Studies of CCAAT/Enhancer-binding Protein ε," *J. Biol. Chem.*, vol. 276, no. 21, pp. 17739–17746, Jan. 2001.

[77] C. Nerlov and E. B. Ziff, "Three levels of functional interaction determine the activity of CCAAT/enhancer binding protein-α on the serum albumin promoter," *Genes Dev.*, vol. 8, no. 3, pp. 350–362, 1994.

[78] S. C. Williams, M. Baer, A. J. Dillner, and P. F. Johnson, "CRP2 (C/EBPβ) contains a bipartite regulatory domain that controls transcriptional activation, DNA binding and cell specificity," *EMBO J.*, vol. 14, no. 13, pp. 3170–3183, 1995.

[79] K. A. Kovács, M. Steinmann, P. J. Magistretti, O. Halfon, and J. R. Cardinaux, "CCAAT/enhancer-binding protein family members recruit the coactivator CREB-binding protein and trigger its phosphorylation," *J. Biol. Chem.*, vol. 278, no. 38, pp. 36959–36965, Sep. 2003.

[80] C. Schwartz *et al.*, "Recruitment of p300 by C/EBPβ triggers phosphorylation of p300 and modulates coactivator activity," *EMBO J.*, vol. 22, no. 4, pp. 882–892, Feb. 2003.

[81] C. Nerlov and E. B. Ziff, "CCAAT/enhancer binding protein-α amino acid motifs with dual TBP and TFIIB binding ability co-operate to activate transcription in both yeast and mammalian cells," *EMBO J.*, vol. 14, no. 17, pp. 4318–4328, Sep. 1995.

[82] B. T. Porse *et al.*, "E2F repression by C/EBPα is required for adipogenesis and granulopoiesis in vivo," *Cell*, vol. 107, no. 2, pp. 247–258, Oct. 2001.

[83] B. T. Porse *et al.*, "The Proline-Histidine-Rich CDK2/CDK4 Interaction Region of C/EBPα Is Dispensable for C/EBPα-Mediated Growth Regulation In Vivo," *Mol. Cell. Biol.*, vol. 26, no. 3, pp. 1028–1037, Feb. 2006.

[84] T. Å. Pedersen, E. Kowenz-Leutz, A. Leutz, and C. Nerlov, "Cooperation between C/EBPα TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation," *Genes Dev.*, vol. 15, no. 23, pp. 3208–3216, Dec. 2001.

[85] E. Kowenz-Leutz, G. Twamley, S. Ansieau, and A. Leutz, "Novel mechanism of C/EBPβ (NF-M) transcriptional control: Activation through derepression," *Genes Dev.*, vol. 8, no. 22, pp. 2781–2791, Nov. 1994.

[86] E. A. Williamson *et al.*, "Identification of transcriptional activation and repression

domains in human CCAAT/enhancer-binding protein ε," *J. Biol. Chem.*, vol. 273, no. 24, pp. 14796–14804, Jun. 1998.

[87] H. S. Abdou, E. Atlas, and R. J. G. Haché, "A positive regulatory domain in CCAAT/enhancer binding protein β (C/EBPβ) is required for the glucocorticoid-mediated displacement of histone deacetylase 1 (HDAC1) from the C/ebpα promoter and maximum adipogenesis," *Endocrinology*, vol. 154, no. 4, pp. 1454–1464, Apr. 2013.

[88] D. Q. Pei and C. H. Shih, "An 'attenuator domain' is sandwiched by two distinct transactivation domains in the transcription factor C/EBP," *Mol. Cell. Biol.*, vol. 11, no. 3, pp. 1480–1487, Mar. 1991.

[89] V. Ossipow, P. Descombes, and U. Schibler, "CCAAT/enhancer-binding protein mRNA is translated into multiple proteins with different transcription activation potentials," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 17, pp. 8219–8223, 1993.

[90] F. T. Lin, O. A. MacDougald, A. M. Diehl, and M. D. Lane, "A 30-kDa alternative translation product of the CCAAT/enhancer binding protein α message: Transcriptional activator lacking antimitotic activity," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 20, pp. 9606–9610, 1993.

[91] R. Yamanaka *et al.*, "CCAAT/enhancer binding protein ε is preferentially up-regulated during granulocytic differentiation and its functional versatility is determined by alternative use of promoters and differential splicing," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 12, pp. 6462–6467, Jun. 1997.

[92] C. F. Calkhoven, C. Muller, and A. Leutz, "Translational control of C/EBPα and C/EBPβ isoform expression," *Genes Dev.*, vol. 14, no. 15, pp. 1920–1932, 2000.

[93] C. F. Calkhoven, P. R. J. Bouwman, L. Snippe, and G. Ab, "Translation start site multiplicity of the CCAAT/enhancer binding protein α mRNA is dictated by a small 5′ open reading frame," *Nucleic Acids Res.*, vol. 22, no. 25, pp. 5540–5547, Dec. 1994.

[94] B. A. Slomiany, K. L. D'Arigo, M. M. Kelly, and D. T. Kurtz, "C/EBPα Inhibits Cell Growth via Direct Repression of E2F-DP-Mediated Transcription," *Mol. Cell. Biol.*, vol. 20, no. 16, pp. 5986–5997, Aug. 2000.

[95] L. M. Johansen *et al.*, "c-Myc Is a Critical Target for C/EBPα in Granulopoiesis," *Mol. Cell. Biol.*, vol. 21, no. 11, pp. 3789–3806, Jun. 2001.

[96] N. A. Timchenko, M. Wilde, M. Nakanishi, J. R. Smith, and G. J. Darlington, "CCAAT/enhancer-binding protein α (C/EBPα) inhibits cell proliferation through the p21 (WAF-1/CIP-1/SDI-1) protein," *Genes Dev.*, vol. 10, no. 7, pp. 804–815, Apr. 1996.

[97] C. Müller, C. F. Calkhoven, X. Sha, and A. Leutz, "The CCAAT Enhancer-binding Protein α (C/EBPα) Requires a SWI/SNF Complex for Proliferation Arrest," *J. Biol. Chem.*, vol. 279, no. 8, pp. 7353–7358, Feb. 2004.

[98] R. M. Umek, A. D. Friedman, and S. L. Mcknight, "CCAAT-enhancer binding protein: A component of a differentiation switch," *Science (80-.).*, vol. 251, no. 4991, pp. 288–292, 1991.

[99] G. L. Wang, P. Iakova, M. Wilde, S. Awad, and N. A. Timchenko, "Liver tumors escape negative control of proliferation via PI3K/ Akt-mediated block of C/EBPα growth inhibitory activity," *Genes Dev.*, vol. 18, no. 8, pp. 912–925, Apr. 2004.

[100] P. Zhang, A. Iwama, M. W. Datta, G. J. Darlington, D. C. Link, and D. G. Tenen, "Upregulation of interleukin 6 and granulocyte colony-stimulating factor receptors by transcription factor CCAAT enhancer binding protein α (C/EBPα) is critical for granulopoiesis," *J. Exp. Med.*, vol. 188, no. 6, pp. 1173–1184, Sep. 1998.

[101] D. E. Zhang, P. Zhang, N. D. Wang, C. J. Hetherington, G. J. Darlington, and D. G. Tenen,

"Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 2, pp. 569–574, Jan. 1997.

[102] T. N. Cassel, T. Berg, G. Suske, and M. Nord, "Synergistic transactivation of the differentiation-dependent lung gene Clara cell secretory protein (secretoglobin 1a1) by the basic region leucine zipper factor CCAAT/enhancer-binding protein α and the homeodomain factor Nkx2.1/Thyroid transcription factor-1," *J. Biol. Chem.*, vol. 277, no. 40, pp. 36970–36977, Oct. 2002.

[103] P. C. Martis, J. A. Whitsett, Y. Xu, A. K. T. Perl, H. Wan, and M. Ikegami, "C/EBPα is required for lung maturation at birth," *Development*, vol. 133, no. 6, pp. 1155–1164, Mar. 2006.

[104] D. S. Bassères *et al.*, "Respiratory Failure Due to Differentiation Arrest and Expansion of Alveolar Cells following Lung-Specific Loss of the Transcription Factor C/EBPα in Mice," *Mol. Cell. Biol.*, vol. 26, no. 3, pp. 1109–1123, Feb. 2006.

[105] E. Ohlsson *et al.*, "Initiation of MLL-rearranged AML is dependent on C/EBPα," *J. Exp. Med.*, vol. 211, no. 1, pp. 5–13, Jan. 2014.

[106] A. Giladi *et al.*, "Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis," *Nat. Cell Biol.*, vol. 20, no. 7, pp. 836–846, Jul. 2018.

[107] N. D. Wang *et al.*, "Impaired energy homeostasis in C/EBPα knockout mice," *Science (80-.).*, vol. 269, no. 5227, pp. 1108–1112, 1995.

[108] T. Å. Pedersen *et al.*, "Distinct C/EBPα motifs regulate lipogenic and gluconeogenic gene expression in vivo," *EMBO J.*, vol. 26, no. 4, pp. 1081–1093, Feb. 2007.

[109] T. Kimura *et al.*, "Hypoglycemia-associated hyperammonemia caused by impaired expression of ornithine cycle enzyme genes in C/EBPα knockout mice," *J. Biol. Chem.*, vol. 273, no. 42, pp. 27505–27510, Oct. 1998.

[110] W. J. Roesler, "The role of C/EBP in nutrient and hormonal regulation of gene expression," *Annual Review of Nutrition*, vol. 21. Annu Rev Nutr, pp. 141–165, 2001.

[111] A. R. Lourenço and P. J. Coffer, "A tumor suppressor role for C/EBPα in solid tumors: More than fat and blood," *Oncogene*, vol. 36, no. 37, pp. 5221–5230, Sep. 2017.

[112] C. Wang *et al.*, "MiR-182 is up-regulated and targeting Cebpa in hepatocellular carcinoma," *Chinese J. Cancer Res.*, vol. 26, no. 1, pp. 17–29, 2014.

[113] P. Foka, N. N. Singh, R. C. Salter, and D. P. Ramji, "The tumour necrosis factor-α-mediated suppression of the CCAAT/enhancer binding protein-α gene transcription in hepatocytes involves inhibition of autoregulation," *Int. J. Biochem. Cell Biol.*, vol. 41, no. 5, pp. 1189–1197, May 2009.

[114] S. Zhang *et al.*, "Yin Yang-1 suppresses differentiation of hepatocellular carcinoma cells through the downregulation of CCAAT/enhancer-binding protein alpha," *J. Mol. Med.*, vol. 90, no. 9, pp. 1069–1077, Sep. 2012.

[115] Y. C. Shi *et al.*, "C/EBPα inhibits hepatocellular carcinoma by reducing Notch3/Hes1/p27 cascades," *Dig. Liver Dis.*, vol. 45, no. 10, pp. 844–851, Oct. 2013.

[116] P. J. Watkins, J. P. Condreay, B. E. Huber, S. J. Jacobs, and D. J. Adams, "Impaired proliferation and tumorigenicity induced by CCAAT/enhancer-binding protein," *Cancer Res.*, vol. 56, no. 5, pp. 1063–1067, 1996.

[117] C. Zeng *et al.*, "A novel GSK-3 beta-C/EBP alpha-miR-122-insulin-like growth factor 1 receptor regulatory circuitry in human hepatocellular carcinoma," *Hepatology*, vol. 52, no. 5, pp. 1702–1712, Nov. 2010.

[118] H. Huan *et al.*, "C/EBPα short-activating RNA suppresses metastasis of hepatocellular carcinoma through inhibiting EGFR/β-catenin signaling mediated EMT," *PLoS One*, vol. 11, no. 4, p. e0153117, Apr. 2016.
[119] S. Gery, S. Tanosaki, S. Bose, N. Bose, J. Vadgama, and H. P. Koeffler, "Down-regulation and growth inhibitory role of C/EBPα in breast cancer," *Clin. Cancer Res.*, vol. 11, no. 9, pp. 3184–3190, May 2005.
[120] K. Milde-Langosch, T. Löning, and A. M. Bamberger, "Expression of the CCAAT/enhancer-binding proteins C/EBPα, C/EBPβ and C/EBPδ in breast cancer: Correlations with clinicopathologic parameters and cell-cycle regulatory proteins," *Breast Cancer Res. Treat.*, vol. 79, no. 2, pp. 175–185, May 2003.
[121] J. Zhang, Y. Ma, S. Wang, F. Chen, and Y. Gu, "C/EBPα inhibits proliferation of breast cancer cells via a novel pathway of miR-134/CREB," *Int. J. Clin. Exp. Pathol.*, vol. 8, no. 11, pp. 14472–14478, 2015.
[122] R. Seifeddine *et al.*, "Hypoxia down-regulates CCAAT/enhancer binding protein-α expression in breast cancer cells," *Cancer Res.*, vol. 68, no. 7, pp. 2158–2165, Apr. 2008.
[123] J. Cheng, D. V. Yu, J. H. Zhou, and D. J. Shapiro, "Tamoxifen induction of CCAAT enhancer-binding protein α is required for tamoxifen-induced apoptosis," *J. Biol. Chem.*, vol. 282, no. 42, pp. 30535–30543, Oct. 2007.
[124] P. Dhawan, R. Weider, and S. Christakos, "CCAAT enhancer-binding protein α is a molecular target of 1,25-dihydroxyvitamin D3 in MCF-7 breast cancer cells," *J. Biol. Chem.*, vol. 284, no. 5, pp. 3086–3095, Jan. 2009.
[125] T. Kumagai *et al.*, "Epigenetic regulation and molecular characterization of C/EBPα in pancreatic cancer cells," *Int. J. Cancer*, vol. 124, no. 4, pp. 827–833, Feb. 2009.
[126] Y. Tada, R. M. Brena, B. Hackanson, C. Morrison, G. A. Otterson, and C. Plass, "Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein α activity in lung cancer," *J. Natl. Cancer Inst.*, vol. 98, no. 6, pp. 396–406, Mar. 2006.
[127] K. L. Bennett *et al.*, "Tumor suppressor activity of CCAAT/enhancer binding protein a is epigenetically down-regulated in head and neck squamous cell carcinoma," *Cancer Res.*, vol. 67, no. 10, pp. 4657–4664, May 2007.
[128] T. S. P. Heng *et al.*, "The immunological genome project: Networks of gene expression in immune cells," *Nature Immunology*, vol. 9, no. 10. pp. 1091–1094, 2008.
[129] C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, and A. M. Klein, "Lineage tracing on transcriptional landscapes links state to fate during differentiation," *Science (80-.).*, vol. 367, no. 6479, Feb. 2020.
[130] C. S. Hyung, J. Gooya, K. Renn, A. D. Friedman, P. F. Johnson, and J. R. Keller, "C/EBPα determines hematopoietic cell fate in multipotential progenitor cells by inhibiting erythroid differentiation and inducing myeloid differentiation," *Blood*, vol. 107, no. 11, pp. 4308–4316, Jun. 2006.
[131] L. M. Scott, C. I. Civin, P. Rorth, and A. D. Friedman, "A novel temporal expression pattern of three C/EBP family members in differentiating myelomonocytic cells," *Blood*, vol. 80, no. 7, pp. 1725–1735, Oct. 1992.
[132] H. Iwasaki *et al.*, "The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages," *Genes Dev.*, vol. 20, no. 21, pp. 3010–3021, Nov. 2006.
[133] D. E. Zhang, P. Zhang, N. D. Wang, C. J. Hetherington, G. J. Darlington, and D. G. Tenen, "Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice," *Proc. Natl. Acad.*

Sci. U. S. A., vol. 94, no. 2, pp. 569–574, Jan. 1997.
[134] P. Zhang *et al.*, "Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBPα," *Immunity*, vol. 21, no. 6, pp. 853–863, Dec. 2004.
[135] F. Paul *et al.*, "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors," *Cell*, vol. 163, no. 7, pp. 1663–1677, Dec. 2015.
[136] A. Olsson *et al.*, "Single-cell analysis of mixed-lineage states leading to a binary cell fate choice," *Nature*, vol. 537, no. 7622, pp. 698–702, Aug. 2016.
[137] A. Wölfler *et al.*, "Lineage-instructive function of C/EBPα in multipotent hematopoietic cells and early thymic progenitors," *Blood*, vol. 116, no. 20, pp. 4116–4125, Nov. 2010.
[138] M. Ye *et al.*, "C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence," *Nat. Cell Biol.*, vol. 15, no. 4, pp. 385–394, Apr. 2013.
[139] S. Pundhir *et al.*, "Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors," *Cell Rep.*, vol. 23, no. 9, pp. 2744–2757, May 2018.
[140] M. S. Hasemann *et al.*, "C/EBPα Is Required for Long-Term Self-Renewal and Lineage Priming of Hematopoietic Stem Cells and for the Maintenance of Epigenetic Configurations in Multipotent Progenitors," *PLoS Genet.*, vol. 10, no. 1, p. 1004079, Jan. 2014.
[141] H. Schepers, A. T. J. Wierenga, D. Van Gosliga, B. J. L. Eggen, E. Vellenga, and J. J. Schuringa, "Reintroduction of C/EBPα in leukemic CD34+ stem/progenitor cells impairs self-renewal and partially restores myelopoiesis," *Blood*, vol. 110, no. 4, pp. 1317–1325, Aug. 2007.
[142] J. Cammenga, J. C. Mulloy, F. J. Berguido, D. MacGrogan, A. Viale, and S. D. Nimer, "Induction of C/EBPα activity alters gene expression and differentiation of human CD34+ cells," *Blood*, vol. 101, no. 6, pp. 2206–2214, Mar. 2003.
[143] C. V. Laiosa, M. Stadtfeld, H. Xie, L. de Andres-Aguayo, and T. Graf, "Reprogramming of Committed T Cell Progenitors to Macrophages and Dendritic Cells by C/EBPα and PU.1 Transcription Factors," *Immunity*, vol. 25, no. 5, pp. 731–744, Nov. 2006.
[144] A. Di Tullio, T. P. Vu Manh, A. Schubert, R. Månsson, and T. Graf, "CCAAT/enhancer binding protein α (C/EBPα)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 41, pp. 17016–17021, Oct. 2011.
[145] K. Takahashi and S. Yamanaka, "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," *Cell*, vol. 126, no. 4, pp. 663–676, Aug. 2006.
[146] B. Di Stefano *et al.*, "C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells," *Nature*, vol. 506, no. 7487, pp. 235–239, 2014.
[147] C. Bueno *et al.*, "Reprogramming human B cells into induced pluripotent stem cells and its enhancement by C/EBPα," *Leukemia*, vol. 30, no. 3, pp. 674–682, Mar. 2016.
[148] A. Di Tullio and T. Graf, "C/EBPα bypasses cell cycle-dependency during immune cell transdifferentiation," *Cell Cycle*, vol. 11, no. 14, pp. 2739–2746, Jul. 2012.
[149] O. Ma, S. H. Hong, H. Guo, G. Ghiaur, and A. D. Friedman, "Granulopoiesis requires increased C/EBPα compared to monopoiesis, correlated with elevated Cebpa in immature G-CSF receptor versus M-CSF receptor expressing cells," *PLoS One*, vol. 9, no. 4, pp. 1–14, Apr. 2014.
[150] J. A. Lekstrom-Himes, S. E. Dorman, P. Kopar, S. M. Holland, and J. I. Gallin, "Neutrophil-

specific granule deficiency results from a novel mutation with loss of function of the transcription factor CCAAT/enhancer binding protein ε," *J. Exp. Med.*, vol. 189, no. 11, pp. 1847–1852, Jun. 1999.

[151] H. Karsunky *et al.*, "Inflammatory reactions and severe neutropenia in mice lacking the transcriptional repressor GFi1," *Nat. Genet.*, vol. 30, no. 3, pp. 295–300, 2002.

[152] S. M. Diakiw, C. H. Kok, L. B. To, I. D. Lewis, A. L. Brown, and R. J. D'Andrea, "The granulocyte-associated transcription factor Krüppel-like factor 5 is silenced by hypermethylation in acute myeloid leukemia," *Leuk. Res.*, vol. 36, no. 1, pp. 110–116, Jan. 2012.

[153] F. Fazi *et al.*, "A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPα regulates human granulopoiesis," *Cell*, vol. 123, no. 5, pp. 819–831, Dec. 2005.

[154] L. M. Starnes *et al.*, "NFI-A directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls β-globin and G-CSF receptor expression," *Blood*, vol. 114, no. 9, pp. 1753–1763, 2009.

[155] C. Katzerke *et al.*, "Transcription factor C/EBPα-induced microRNA-30c inactivates Notch1 during granulopoiesis and is downregulated in acute myeloid leukemia," *Blood*, vol. 122, no. 14, pp. 2433–2442, 2013.

[156] M. Nakamura, L. Wu, S. Kojika, J. D. Griffin, and K. Sugita, "Activation of Notch1 Signaling Suppresses Granulocytic Differentiation and Maintains a Part of Myeloid Progenitor Cells At the Immature Stage," *Blood*, vol. 118, no. 21, pp. 2375–2375, Nov. 2011.

[157] L. A. Milner, A. Bigas, R. Kopan, C. Brashem-Stein, I. D. Bernstein, and D. I. K. Martin, "Inhibition of granulocytic differentiation by mNotch1," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 23, pp. 13014–13019, Nov. 1996.

[158] A. D. Friedman, "Transcriptional regulation of granulocyte and monocyte development," *Oncogene*, vol. 21, no. 21 REV. ISS. 2, pp. 3377–3390, May 2002.

[159] H. S. Radomska, C. S. Huettner, P. Zhang, T. Cheng, D. T. Scadden, and D. G. Tenen, "CCAAT/Enhancer Binding Protein α Is a Regulatory Switch Sufficient for Induction of Granulocytic Development from Bipotential Myeloid Progenitors," *Mol. Cell. Biol.*, vol. 18, no. 7, pp. 4301–4314, Jul. 1998.

[160] X. Wang, E. Scott, C. L. Sawyers, and A. D. Friedman, "C/EBPα bypasses granulocyte colony-stimulating factor signals to rapidly induce PU.1 gene expression, stimulate granulocytic differentiation, and limit proliferation in 32D cl3 myeloblasts," *Blood*, vol. 94, no. 2, pp. 560–571, Jul. 1999.

[161] L. T. Smith, S. Hohaus, D. A. Gonzalez, S. E. Dziennis, and D. G. Tenen, "PU.1 (Spi-1) and C/EBPα regulate the granulocyte colony-stimulating factor receptor promoter in myeloid cells," *Blood*, vol. 88, no. 4, pp. 1234–1247, Aug. 1996.

[162] S. Hohaus, M. S. Petrovick, M. T. Voso, Z. Sun, D. E. Zhang, and D. G. Tenen, "PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene," *Mol. Cell. Biol.*, vol. 15, no. 10, pp. 5830–5845, Oct. 1995.

[163] D. E. Zhang, C. J. Hetherington, H. M. Chen, and D. G. Tenen, "The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor," *Mol. Cell. Biol.*, vol. 14, no. 1, pp. 373–381, Jan. 1994.

[164] M. S. Petrovick, S. W. Hiebert, A. D. Friedman, C. J. Hetherington, D. G. Tenen, and D.-E. Zhang, "Multiple Functional Domains of AML1: PU.1 and C/EBPα Synergize with

Different Regions of AML1," *Mol. Cell. Biol.*, vol. 18, no. 7, pp. 3915–3925, Jul. 1998.
[165] Z. Pan, C. J. Hetherington, and D. E. Zhang, "CCAAT/enhancer-binding protein activates the CD14 promoter and mediates transforming growth factor β signaling in monocyte development," *J. Biol. Chem.*, vol. 274, no. 33, pp. 23242–23248, Aug. 1999.
[166] B. Cirovic *et al.*, "C/EBP-Induced Transdifferentiation Reveals Granulocyte-Macrophage Precursor-like Plasticity of B Cells," *Stem Cell Reports*, vol. 8, no. 2, pp. 346–359, 2017.
[167] R. J. Christy, K. H. Kaestner, D. E. Geiman, and M. Daniel Lane, "CCAAT/enhancer binding protein gene promoter: Binding of nuclear factors during differentiation of 3T3-L1 preadipocytes," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, no. 6, pp. 2593–2597, 1991.
[168] C. Legraverend, P. Antonson, P. Flodby, and G. Xanthopoulos, "High level activity of the mouse CCAAT/enhancer binding protein (C/EBPα) gene promoter involves autoregulation and several ubiquitous transcription factors," *Nucleic Acids Res.*, vol. 21, no. 8, pp. 1735–1742, Apr. 1993.
[169] D. Gonzalez *et al.*, "ZNF143 protein is an important regulator of the myeloid transcription factor C/EBPα," *J. Biol. Chem.*, vol. 292, no. 46, pp. 18924–18936, Nov. 2017.
[170] S. Cooper, H. Guo, and A. D. Friedman, "The +37 kb Cebpa enhancer is critical for Cebpa myeloid gene expression and contains functional sites that bind SCL, GATA2, C/EBPα, PU.1, and additional Ets factors," *PLoS One*, vol. 10, no. 5, p. e0126385, May 2015.
[171] R. Avellino *et al.*, "An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation," *Blood*, vol. 127, no. 24, pp. 2991–3003, 2016.
[172] H. Guo, O. Ma, N. A. Speck, and A. D. Friedman, "Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis," *Blood*, vol. 119, no. 19, pp. 4408–4418, May 2012.
[173] G. Huang *et al.*, "PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis," *Nat. Genet.*, vol. 40, no. 1, pp. 51–60, Jan. 2008.
[174] J. Skokowa *et al.*, "LEF-1 is crucial for neutrophil granulocytopoiesis and its expression is severely reduced in congenital neutropenia," *Nat. Med.*, vol. 12, no. 10, pp. 1191–1197, 2006.
[175] Y. Jiang *et al.*, "Desferrioxamine induces leukemic cell differentiation potentially by hypoxia-inducible factor-1α that augments transcriptional activity of CCAAT/enhancer-binding protein-α," *Leukemia*, vol. 19, no. 7, pp. 1239–1247, May 2005.
[176] A. Di Ruscio *et al.*, "DNMT1-interacting RNAs block gene-specific DNA methylation," *Nature*, vol. 503, no. 7476, pp. 371–376, Oct. 2013.
[177] M. E. De Obaldia *et al.*, "T cell development requires constraint of the myeloid regulator C/EBP-α by the Notch target and transcriptional repressor Hes1," *Nat. Immunol.*, vol. 14, no. 12, pp. 1277–1284, Nov. 2013.
[178] J. A. Zhang, A. Mortazavi, B. A. Williams, B. J. Wold, and E. V. Rothenberg, "Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity," *Cell*, vol. 149, no. 2, pp. 467–482, Apr. 2012.
[179] A. Repele, S. Krueger, T. Bhattacharyya, M. Y. Tuineau, and Manu, "The regulatory control of Cebpa enhancers and silencers in the myeloid and red-blood cell lineages," *PLoS One*, vol. 14, no. 6, p. e0217580, Jun. 2019.
[180] M. A. Rieger, P. S. Hoppe, B. M. Smejkal, A. C. Eitelhuber, and T. Schroeder, "Hematopoietic cytokines can instruct lineage choice," *Science (80-.).*, vol. 325, no. 5937, pp. 217–218, Jul. 2009.

[181] G. D. Jack, L. Zhang, and A. D. Friedman, “M-CSF elevates c-Fos and phospho-C/EBPα(S21) via ERK whereas G-CSF stimulates SHP2 phosphorylation in marrow progenitors to contribute to myeloid lineage specification,” *Blood*, vol. 114, no. 10, pp. 2172–2180, 2009.

[182] S. E. Ross *et al.*, “Phosphorylation of C/EBPα Inhibits Granulopoiesis,” *Mol. Cell. Biol.*, vol. 24, no. 2, pp. 675–686, Jan. 2004.

[183] D. H. Cai, D. Wang, J. Keefer, C. Yeamans, K. Hensley, and A. D. Friedman, “C/EBPα:AP-1 leucine zipper heterodimers bind novel DNA elements, activate the PU.1 promoter and direct monocyte lineage commitment more potently than C/EBPα homodimers or AP-1,” *Oncogene*, vol. 27, no. 19, pp. 2772–2779, Apr. 2008.

[184] C. R. Geest *et al.*, “p38 MAP kinase inhibits neutrophil development through phosphorylation of C/EBPα on serine 21,” *Stem Cells*, vol. 27, no. 9, pp. 2271–2282, Sep. 2009.

[185] H. Yuan *et al.*, “Sumoylation of CCAAT/enhancer-binding protein α promotes the biased primitive hematopoiesis of zebrafish,” *Blood*, vol. 117, no. 26, pp. 7014–7020, Jun. 2011.

[186] C. Yeamans, D. Wang, I. Paz-Priel, B. E. Torbett, D. G. Tenen, and A. D. Friedman, “C/EBPα binds and activates the PU.1 distal enhancer to induce monocyte lineage commitment,” *Blood*, vol. 110, no. 9, pp. 3136–3142, Nov. 2007.

[187] S. Pundhir *et al.*, “Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors,” *Cell Rep.*, vol. 23, no. 9, pp. 2744–2757, May 2018.

[188] A. A. Zada *et al.*, “Proteomic discovery of Max as a novel interacting partner of C/EBPα: A Myc/Max/Mad link,” *Leukemia*, vol. 20, no. 12, pp. 2137–2146, 2006.

[189] E. Musialik *et al.*, “Comparison of promoter DNA methylation and expression levels of genes encoding CCAAT/enhancer binding proteins in AML patients,” *Leuk. Res.*, vol. 38, no. 7, pp. 850–856, Jul. 2014.

[190] A. Fasan *et al.*, “Frequency and Prognostic Impact of CEBPA Proximal, Distal and Core Promoter Methylation in Normal Karyotype AML: A Study on 623 Cases,” *PLoS One*, vol. 8, no. 2, Feb. 2013.

[191] M. Wilson, Y. Zhang, X. Cui, C. A. Wuertzer, and A. S. Perkins, “EVI1, A Potential Regulator Of CEBPA; A Contributory Mechanism In Leukemogenesis?,” *Blood*, vol. 122, no. 21, pp. 2495–2495, Nov. 2013.

[192] M. Wilson, V. Tsakraklides, M. Tran, Y. Y. Xiao, Y. Zhang, and A. S. Perkins, “EVI1 interferes with myeloid maturation via transcriptional repression of Cebpa, via binding to two far downstream regulatory elements,” *J. Biol. Chem.*, vol. 291, no. 26, pp. 13591–13607, Jun. 2016.

[193] S. Kiehlmeier *et al.*, “Identification of therapeutic targets of the hijacked super-enhancer complex in EVI1-rearranged leukemia,” *Leukemia*, vol. 35, no. 11, pp. 3127–3138, Apr. 2021.

[194] D. Perrotti *et al.*, “BCR-ABL suppresses C/EBPα expression through inhibitory action of hnRNP E2,” *Nat. Genet.*, vol. 30, no. 1, pp. 48–58, Dec. 2002.

[195] A. M. Eiring *et al.*, “miR-328 Functions as an RNA Decoy to Modulate hnRNP E2 Regulation of mRNA Translation in Leukemic Blasts,” *Cell*, vol. 140, no. 5, pp. 652–665, Mar. 2010.

[196] A. A. Wurm *et al.*, “Disruption of the C/EBPα - MiR-182 balance impairs granulocytic differentiation,” *Nat. Commun.*, vol. 8, no. 1, Dec. 2017.

[197] H. S. Radomska *et al.*, “Block of C/EBPα function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations,” *J. Exp. Med.*, vol. 203, no. 2, pp. 371–381, 2006.
[198] D. Bararia *et al.*, “Acetylation of C/EBPα inhibits its granulopoietic function,” *Nat. Commun.*, vol. 7, p. 10968, 2016.
[199] K. Keeshan *et al.*, “Tribbles homolog 2 inactivates C/EBPα and causes acute myelogenous leukemia,” *Cancer Cell*, vol. 10, no. 5, pp. 401–411, 2006.
[200] C. O’Connor *et al.*, “The presence of C/EBPα and its degradation are both required for TRIB2-mediated leukaemia,” *Oncogene*, vol. 35, no. 40, pp. 5272–5281, Oct. 2016.
[201] M. Ye *et al.*, “Hematopoietic Differentiation Is Required for Initiation of Acute Myeloid Leukemia,” *Cell Stem Cell*, vol. 17, no. 5, pp. 611–623, Nov. 2015.
[202] T. Pabst *et al.*, “Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-α (C/EBPα), in acute myeloid leukemia,” *Nat. Genet.*, vol. 27, no. 3, pp. 263–270, 2001.
[203] J. Snaddon *et al.*, “Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2,” *Genes Chromosom. Cancer*, vol. 37, no. 1, pp. 72–78, May 2003.
[204] S. B. van Waalwijk van Doorn-Khosrovani *et al.*, “Biallelic mutations in the CEBPA gene low CEBPA expression levels as prognostic markers in intermediate-risk AML,” *Hematol. J.*, vol. 4, no. 1, pp. 31–40, 2003.
[205] C. Preudhomme *et al.*, “Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: A study from the Acute Leukemia French Association (ALFA),” *Blood*, vol. 100, no. 8, pp. 2717–2723, Oct. 2002.
[206] A. F. Gombart *et al.*, “Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein α in myelodysplastic syndromes and acute myeloid leukemias,” *Blood*, vol. 99, no. 4, pp. 1332–1340, Feb. 2002.
[207] “scholar.” .
[208] T. Pabst, M. Eyholzer, S. Haefliger, J. Schardt, and B. U. Mueller, “Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia,” *J. Clin. Oncol.*, vol. 26, no. 31, pp. 5088–5093, Sep. 2008.
[209] K. Tawana *et al.*, “Disease evolution and outcomes in familial AML with germline CEBPA mutations,” *Blood*, vol. 126, no. 10, pp. 1214–1223, Sep. 2015.
[210] V. Grossmann *et al.*, “CEBPA double-mutated acute myeloid leukaemia harbours concomitant molecular mutations in 76·8% of cases with TET2 and GATA2 alterations impacting prognosis,” *Br. J. Haematol.*, vol. 161, no. 5, pp. 649–658, Jun. 2013.
[211] E. Taskesen *et al.*, “Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: Further evidence for CEBPA double mutant AML as a distinctive disease entity,” *Blood*, vol. 117, no. 8, pp. 2469–2475, Feb. 2011.
[212] A. Dufour *et al.*, “Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome,” *J. Clin. Oncol.*, vol. 28, no. 4, pp. 570–577, Feb. 2010.
[213] H. Döhner *et al.*, “Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel,” *Blood*, vol. 129, no. 4, pp. 424–447, Jan. 2017.
[214] T. Pabst, M. Eyholzer, J. Fos, and B. U. Mueller, “Heterogeneity within AML with CEBPA mutations; Only CEBPA double mutations, but not single CEBPA mutations are

associated with favourable prognosis," *Br. J. Cancer*, vol. 100, no. 8, pp. 1343–1346, Apr. 2009.

[215] H. Döhner *et al.*, "Diagnosis and management of acute myeloid leukemia in adults: Recommendations from an international expert panel, on behalf of the European LeukemiaNet," *Blood*, vol. 115, no. 3. Blood, pp. 453–474, 21-Jan-2010.

[216] S. Sarojam, S. Raveendran, S. Vijay, J. Sreedharan, G. Narayanan, and H. Sreedharan, "Characterization of CEBPA mutations and polymorphisms and their prognostic relevance in de novo acute myeloid leukemia patients," *Asian Pacific J. Cancer Prev.*, vol. 16, no. 9, pp. 3785–3792, 2015.

[217] A. Fasan *et al.*, "The role of different genetic subtypes of CEBPA mutated AML," *Leukemia*, vol. 28, no. 4, pp. 794–803, 2014.

[218] Y. Zhang *et al.*, "Companion gene mutations and their clinical significance in AML with double mutant CEBPA," *Cancer Gene Ther.*, vol. 27, no. 7–8, pp. 599–606, Sep. 2020.

[219] K. Keeshan, G. Santilli, F. Corradini, D. Perrotti, and B. Calabretta, "Transcription activation function of C/EBPα is required for induction of granulocytic differentiation," *Blood*, vol. 102, no. 4, pp. 1267–1275, Aug. 2003.

[220] H. Leroy, C. Roumier, P. Huyghe, V. Biggio, P. Fenaux, and C. Preudhomme, "CEBPA point mutations in hematological malignancies," *Leukemia*, vol. 19, no. 3, pp. 329–334, Jan. 2005.

[221] M. B. Schuster, A. K. Frank, F. O. Bagger, N. Rapin, J. Vikesaa, and B. T. Porse, "Lack of the p42 form of C/EBPα leads to spontaneous immortalization and lineage infidelity of committed myeloid progenitors," *Exp. Hematol.*, vol. 41, no. 10, pp. 882-893.e16, Oct. 2013.

[222] F. D'Alo' *et al.*, "The amino terminal and E2F interaction domains are critical for C/EBPα-mediated induction of granulopoietic development of hematopoietic cells," *Blood*, vol. 102, no. 9, pp. 3163–3171, Nov. 2003.

[223] R. Cleaves, Q. F. Wang, and A. D. Friedman, "C/EBPαp30, a myeloid leukemia oncoprotein, limits G-CSF receptor expression but not terminal granulopoiesis via site-selective inhibition of C/EBP DNa binding," *Oncogene*, vol. 23, no. 3, pp. 716–725, Jan. 2004.

[224] B. T. Porse *et al.*, "Loss of C/EBPα cell cycle control increases myeloid progenitor proliferation and transforms the neutrophil granulocyte lineage," *J. Exp. Med.*, vol. 202, no. 1, pp. 85–96, 2005.

[225] C. Nerlov, "C/EBPα mutations in acute myeloid leukaemias," *Nature Reviews Cancer*, vol. 4, no. 5. Nat Rev Cancer, pp. 394–400, 2004.

[226] P. Kirstetter *et al.*, "Modeling of C/EBPα Mutant Acute Myeloid Leukemia Reveals a Common Expression Signature of Committed Myeloid Leukemia-Initiating Cells," *Cancer Cell*, vol. 13, no. 4, pp. 299–310, Apr. 2008.

[227] O. Bereshchenko *et al.*, "Hematopoietic Stem Cell Expansion Precedes the Generation of Committed Myeloid Leukemia-Initiating Cells in C/EBPα Mutant AML," *Cancer Cell*, vol. 16, no. 5, pp. 390–400, Nov. 2009.

[228] K. Wagner *et al.*, "Absence of the transcription factor CCAAT enhancer binding protein α results in loss of myeloid identity in bcr/abl-induced malignancy," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 16, pp. 6338–6343, Apr. 2006.

[229] C. Collins *et al.*, "C/EBPα is an essential collaborator in Hoxa9/Meis1-mediated leukemogenesis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 27, pp. 9899–9904, Jul. 2014.

[230] C. Wang, X. Chen, Y. Wang, J. Gong, and G. Hu, “C/EBPαp30 plays transcriptional regulatory roles distinct from C/EBPαp42,” *Cell Res.*, vol. 17, no. 4, pp. 374–383, Apr. 2007.
[231] J. S. Jakobsen *et al.*, “Mutant CEBPA directly drives the expression of the targetable tumor-promoting factor CD73 in AML,” *Sci. Adv.*, vol. 5, no. 7, 2019.
[232] F. Karush, “Heterogeneity of the Binding Sites of Bovine Serum Albumin,” *J. Am. Chem. Soc.*, vol. 72, no. 6, pp. 2705–2713, Jun. 1950.
[233] A. K. Dunker *et al.*, “Intrinsically disordered protein,” *J. Mol. Graph. Model.*, vol. 19, no. 1, pp. 26–59, 2001.
[234] P. E. Wright and H. J. Dyson, “Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm,” *J. Mol. Biol.*, vol. 293, no. 2, pp. 321–331, Oct. 1999.
[235] P. E. Wright and H. J. Dyson, “Intrinsically disordered proteins in cellular signalling and regulation,” *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 1, pp. 18–29, Dec. 2015.
[236] M. E. Oates *et al.*, “D2P2: Database of disordered protein predictions,” *Nucleic Acids Res.*, vol. 41, no. D1, p. D508, Jan. 2013.
[237] M. M. Babu, “The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease,” *Biochem. Soc. Trans.*, vol. 44, no. 5, pp. 1185–1200, Oct. 2016.
[238] R. Van Der Lee *et al.*, “Classification of intrinsically disordered regions and proteins,” *Chem. Rev.*, vol. 114, no. 13, pp. 6589–6631, 2014.
[239] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, “Intrinsically disordered proteins in human diseases: Introducing the D 2 concept,” *Annu. Rev. Biophys.*, vol. 37, pp. 215–246, May 2008.
[240] H. Lee *et al.*, “Local structural elements in the mostly unstructured transcriptional activation domain of human p53,” *J. Biol. Chem.*, vol. 275, no. 38, pp. 29426–29432, Sep. 2000.
[241] R. Dawson, L. Müller, A. Dehner, C. Klein, H. Kessler, and J. Buchner, “The N-terminal domain of p53 is natively unfolded,” *J. Mol. Biol.*, vol. 332, no. 5, pp. 1131–1141, Oct. 2003.
[242] C. W. Anderson and E. Appella, “Signaling to the p53 tumor suppressor through pathways activated by genotoxic and non-genotoxic stresses,” *Handb. Cell Signaling, 2/e*, vol. 3, pp. 2185–2204, Jan. 2010.
[243] K. H. Vousden and X. Lu, “Live or let die: The cell’s response to p53,” *Nature Reviews Cancer*, vol. 2, no. 8. Nat Rev Cancer, pp. 594–604, Aug-2002.
[244] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Qu, V. N. Uversky, and A. K. Dunker, “Flexible nets: Disorder and induced fit in the associations of p53 and 14-3-3 with their partners,” *BMC Genomics*, vol. 9, no. SUPPL. 1, Mar. 2008.
[245] V. N. Uversky, “p53 proteoforms and intrinsic disorder: An illustration of the protein structure–function continuum concept,” *Int. J. Mol. Sci.*, vol. 17, no. 11, Nov. 2016.
[246] W. Y. Mark *et al.*, “Characterization of segments from the central region of BRCA1: An intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions?,” *J. Mol. Biol.*, vol. 345, no. 2, pp. 275–287, Jan. 2005.
[247] A. R. Venkitaraman, “Cancer susceptibility and the functions of BRCA1 and BRCA2,” *Cell*, vol. 108, no. 2. Cell, pp. 171–182, 25-Jan-2002.
[248] C. X. Deng, “BRCA1: Cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution,” *Nucleic Acids Res.*, vol. 34, no. 5, pp. 1416–1426, 2006.

[249] R. A. Crowther and M. Goedert, “Abnormal tau-containing filaments in neurodegenerative diseases,” *J. Struct. Biol.*, vol. 130, no. 2–3, pp. 271–279, 2000.
[250] M. Von Bergen, S. Barghorn, S. Jeganathan, E. M. Mandelkow, and E. Mandelkow, “Spectroscopic approaches to the conformation of tau protein in solution and in paired helical filaments,” *Neurodegener. Dis.*, vol. 3, no. 4–5, pp. 197–206, Oct. 2006.
[251] V. N. Uversky, S. Winter, O. V. Galzitskaya, L. Kittler, and G. Lober, “Hyperphosphorylation induces structural modification of tau-protein,” *FEBS Lett.*, vol. 439, no. 1–2, pp. 21–25, Nov. 1998.
[252] T. Hagestedt, B. Lichtenberg, H. Wille, E. M. Mandelkow, and E. Mankelkow, “Tau protein becomes long and stiff upon phosphorylation: Correlation between paracrystalline structure and degree of dephosphorylation,” *J. Cell Biol.*, vol. 109, no. 4 I, pp. 1643–1651, 1989.
[253] V. N. Uversky, “Intrinsically disordered proteins and their ‘Mysterious’ (meta)physics,” *Frontiers in Physics*, vol. 7, no. FEB. 2019.
[254] C. T. Walsh, S. Garneau-Tsodikova, and G. J. Gatto, “Protein posttranslational modifications: The chemistry of proteome diversifications,” *Angewandte Chemie - International Edition*, vol. 44, no. 45. pp. 7342–7372, 2005.
[255] D. Vuzman, Y. Hoffman, and Y. Levy, “Modulating protein-DNA interactions by post-translational modifications at disordered regions,” in *Pacific Symposium on Biocomputing*, 2012, pp. 188–199.
[256] H. Xie *et al.*, “Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions,” *J. Proteome Res.*, vol. 6, no. 5, pp. 1882–1898, May 2007.
[257] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradović, “Intrinsic disorder and protein function,” *Biochemistry*, vol. 41, no. 21, pp. 6573–6582, 2002.
[258] A. L. Darling and V. N. Uversky, “Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter,” *Front. Genet.*, vol. 9, no. MAY, pp. 1–18, 2018.
[259] A. Bah and J. D. Forman-Kay, “Modulation of intrinsically disordered protein function by post-translational modifications,” *J. Biol. Chem.*, vol. 291, no. 13, pp. 6696–6705, 2016.
[260] J. Murn and Y. Shi, “The winding path of protein methylation research: Milestones and new frontiers,” *Nat. Rev. Mol. Cell Biol.*, vol. 18, no. 8, pp. 517–527, 2017.
[261] E. Guccione and S. Richard, “The regulation, functions and clinical relevance of arginine methylation,” *Nat. Rev. Mol. Cell Biol.*, vol. 20, no. 10, pp. 642–657, 2019.
[262] M. Fontecave, M. Atta, and E. Mulliez, “S-adenosylmethionine: Nothing goes to waste,” *Trends Biochem. Sci.*, vol. 29, no. 5, pp. 243–249, May 2004.
[263] D. Y. Lee, C. Teyssier, B. D. Strahl, and M. R. Stallcup, “Role of protein methylation in regulation of transcription,” *Endocr. Rev.*, vol. 26, no. 2, pp. 147–170, 2005.
[264] M. T. Bedford, “Methylation of Proteins,” *Encycl. Ref. Genomics Proteomics Mol. Med.*, pp. 1070–1075, Sep. 2006.
[265] S. C. Larsen *et al.*, “Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells,” *Sci. Signal.*, vol. 9, no. 443, Aug. 2016.
[266] M. T. Bedford and S. G. Clarke, “Protein Arginine Methylation in Mammals: Who, What, and Why,” *Mol. Cell*, vol. 33, no. 1, pp. 1–13, 2009.
[267] S. M. Greenblatt, F. Liu, and S. D. Nimer, “Arginine methyltransferases in normal and malignant hematopoiesis,” *Exp. Hematol.*, vol. 44, no. 6, pp. 435–441, 2016.

[268] R. S. Blanc and S. Richard, “Arginine Methylation: The Coming of Age,” *Mol. Cell*, vol. 65, no. 1, pp. 8–24, 2017.

[269] L. Wang, S. Pal, and S. Sif, “Protein Arginine Methyltransferase 5 Suppresses the Transcription of the RB Family of Tumor Suppressors in Leukemia and Lymphoma Cells,” *Mol. Cell. Biol.*, vol. 28, no. 20, pp. 6262–6277, Oct. 2008.

[270] S. Pal, R. A. Baiocchi, J. C. Byrd, M. R. Grever, S. T. Jacob, and S. Sif, “Low levels of miR-92b/96 induce PRMT5 translation and H3R8/H4R3 methylation in mantle cell lymphoma,” *EMBO J.*, vol. 26, no. 15, pp. 3558–3569, Aug. 2007.

[271] X. Zhao *et al.*, “Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity,” *Genes Dev.*, vol. 22, no. 5, pp. 640–653, Mar. 2008.

[272] S. Mizutani, T. Yoshida, X. Zhao, S. D. Nimer, M. Taniwaki, and T. Okuda, “Loss of RUNX1/AML1 arginine-methylation impairs peripheral T cell homeostasis,” *Br. J. Haematol.*, vol. 170, no. 6, pp. 859–873, Sep. 2015.

[273] M. Yoshimatsu *et al.*, “Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers,” *Int. J. Cancer*, vol. 128, no. 3, pp. 562–573, Feb. 2011.

[274] S. Leonard, N. Gordon, N. Smith, M. Rowe, P. G. Murray, and C. B. Woodman, “Arginine methyltransferases are regulated by Epstein-Barr virus in B cells and are differentially expressed in Hodgkin’s lymphoma,” *Pathogens*, vol. 1, no. 1, pp. 52–64, Sep. 2012.

[275] W. J. Shia *et al.*, “PRMT1 interacts with AML1-ETO to promote its transcriptional activation and progenitor cell proliferative potential,” *Blood*, vol. 119, no. 21, pp. 4953–4962, May 2012.

[276] L. Zou *et al.*, “Correlation of SRSF1 and PRMT1 expression with clinical status of pediatric acute lymphoblastic leukemia,” *J. Hematol. Oncol.*, vol. 5, no. 1, pp. 1–11, Jul. 2012.

[277] N. Cheung, L. C. Chan, A. Thompson, M. L. Cleary, and C. W. E. So, “Protein arginine-methyltransferase-dependent oncogenesis,” *Nat. Cell Biol.*, vol. 9, no. 10, pp. 1208–1215, Sep. 2007.

[278] X. He *et al.*, “PRMT1-mediated FLT3 arginine methylation promotes maintenance of FLT3-ITD1 acute myeloid leukemia,” *Blood*, vol. 134, no. 6, pp. 548–560, 2019.

[279] K. Bonham, S. Hemmers, Y. H. Lim, D. M. Hill, M. G. Finn, and K. A. Mowen, “Effects of a novel arginine methyltransferase inhibitor on T-helper cell cytokine production,” *FEBS J.*, vol. 277, no. 9, pp. 2096–2108, May 2010.

[280] N. Cheung *et al.*, “Targeting Aberrant Epigenetic Networks Mediated by PRMT1 and KDM4C in Acute Myeloid Leukemia,” *Cancer Cell*, vol. 29, no. 1, pp. 32–48, Jan. 2016.

[281] A. E. Drew *et al.*, “Identification of a CARM1 Inhibitor with Potent in Vitro and in Vivo Activity in Preclinical Models of Multiple Myeloma,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–13, Dec. 2017.

[282] S. M. Greenblatt *et al.*, “CARM1 Is Essential for Myeloid Leukemogenesis but Dispensable for Normal Hematopoiesis,” *Cancer Cell*, vol. 33, no. 6, pp. 1111-1127.e5, 2018.

[283] J. Jarrold and C. C. Davies, “PRMTs and Arginine Methylation: Cancer’s Best-Kept Secret?,” *Trends Mol. Med.*, vol. 25, no. 11, pp. 993–1009, 2019.

[284] J. J. Park and M. Lee, “Increasing the α 2, 6 sialylation of glycoproteins may contribute to metastatic spread and therapeutic resistance in colorectal cancer,” *Gut and Liver*, vol. 7, no. 6. pp. 629–641, 2013.

[285] P. Radivojac, P. H. Baenziger, M. G. Kann, M. E. Mort, M. W. Hahn, and S. D. Mooney, “Gain and loss of phosphorylation sites in human cancer,” in *Bioinformatics*, 2008, vol. 24, no. 16.
[286] S. Chatterjee, P. Senapati, and T. K. Kundu, “Post-translational modifications of lysine in DNA-damage repair,” *Essays in Biochemistry*, vol. 52, no. 1. pp. 93–111, 2012.
[287] Z. Huang and S. Bao, “Ubiquitination and deubiquitination of REST and its roles in cancers,” *FEBS Letters*, vol. 586, no. 11. pp. 1602–1605, 2012.
[288] O. Pless, E. Kowenz-Leutz, G. Dittmar, and A. Leutz, “A differential proteome screening system for post-translational modificationg-dependent transcription factor interactions,” *Nat. Protoc.*, vol. 6, no. 3, pp. 359–364, Feb. 2011.
[289] G. Dittmar *et al.*, “PRISMA: Protein Interaction Screen on Peptide Matrix Reveals Interaction Footprints and Modifications- Dependent Interactome of Intrinsically Disordered C/EBPβ,” *iScience*, vol. 13, pp. 351–370, 2019.
[290] B. Stoilova, E. Kowenz-Leutz, M. Scheller, and A. Leutz, “Lymphoid to Myeloid Cell Trans-Differentiation Is Determined by C/EBPβ Structure and Post-Translational Modifications,” *PLoS One*, vol. 8, no. 6, 2013.
[291] L. M. Liu, W. Z. Sun, X. Z. Fan, Y. L. Xu, M. Bin Cheng, and Y. Zhang, “Methylation of C/EBPa by PRMT1 inhibits its tumor-suppressive function in breast cancer,” *Cancer Res.*, vol. 79, no. 11, pp. 2865–2877, 2019.
[292] E. Ramberger *et al.*, “PRISMA and BioID disclose a motifs-based interactome of the intrinsically disordered transcription factor C/EBPα,” *iScience*, vol. 24, no. 6, p. 102686, 2021.
[293] Y. Sato, K. Miyake, H. Kaneoka, and S. Iijima, “Sumoylation of CCAAT/enhancer-binding protein α and its functional roles in hepatocyte differentiation,” *J. Biol. Chem.*, vol. 281, no. 31, pp. 21629–21639, Aug. 2006.
[294] G.-L. Wang and N. A. Timchenko, “Dephosphorylated C/EBPα Accelerates Cell Proliferation through Sequestering Retinoblastoma Protein,” *Mol. Cell. Biol.*, vol. 25, no. 4, pp. 1325–1338, Feb. 2005.
[295] G. Behre *et al.*, “Ras signaling enhances the activity of C/EBPα to induce granulocytic differentiation by phosphorylation of serine 248,” *J. Biol. Chem.*, vol. 277, no. 29, pp. 26293–26299, Jul. 2002.
[296] L. H. Bussmann *et al.*, “A Robust and Highly Efficient Immune Cell Reprogramming System,” *Cell Stem Cell*, vol. 5, no. 5, pp. 554–566, 2009.
[297] R. Feng *et al.*, “PU.1 and C/EBPα/β convert fibroblasts into macrophage-like cells,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 16, pp. 6057–6062, 2008.
[298] B. Barneda-Zahonero *et al.*, “HDAC7 Is a Repressor of Myeloid Genes Whose Downregulation Is Required for Transdifferentiation of Pre-B Cells into Macrophages,” *PLoS Genet.*, vol. 9, no. 5, p. e1003503, May 2013.
[299] E. M. Kallin *et al.*, “Tet2 facilitates the derepression of myeloid target genes during CEBPα-Induced transdifferentiation of Pre-B cells,” *Mol. Cell*, vol. 48, no. 2, pp. 266–276, Oct. 2012.
[300] C. Van Oevelen *et al.*, “C/EBPα Activates Pre-existing and de Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis,” *Stem Cell Reports*, vol. 5, no. 2, pp. 232–247, Aug. 2015.
[301] J. Rodriguez-Ubreva, L. Ciudad, C. van Oevelen, M. Parra, T. Graf, and E. Ballestar, “C/EBPa-Mediated Activation of MicroRNAs 34a and 223 Inhibits Lef1 Expression To Achieve Efficient Reprogramming into Macrophages,” *Mol. Cell. Biol.*, vol. 34, no. 6, pp.

1145–1157, Mar. 2014.
[302] P. H. L. Krijger *et al.*, “Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming,” *Cell Stem Cell*, vol. 18, no. 5, pp. 597–610, May 2016.
[303] M. Evrard *et al.*, “Developmental Analysis of Bone Marrow Neutrophils Reveals Populations Specialized in Expansion, Trafficking, and Effector Functions,” *Immunity*, vol. 48, no. 2, pp. 364-379.e8, Feb. 2018.
[304] A. Mildner *et al.*, “Genomic Characterization of Murine Monocytes Reveals C/EBPβ Transcription Factor Dependence of Ly6C− Cells,” *Immunity*, vol. 46, no. 5, pp. 849-862.e7, May 2017.
[305] M. Evich, E. Stroeva, Y. G. Zheng, and M. W. Germann, “Effect of methylation on the side-chain pKa value of arginine,” *Protein Sci.*, vol. 25, no. 2, pp. 479–486, 2016.
[306] N. M. Luscombe, R. A. Laskowski, and J. M. Thornton, “Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level,” *Nucleic Acids Res.*, vol. 29, no. 13, pp. 2860–2874, Jul. 2001.
[307] M. J. Betts and R. B. Russell, “Amino Acid Properties and Consequences of Substitutions,” in *Bioinformatics for Geneticists*, 1st ed., M. R. Barnes and I. C. Gray, Eds. Wiley, 2003, pp. 289–316.
[308] M. Adamiak, A. Abdel-Latif, and M. Z. Ratajczak, “Purinergic signaling regulates mobilization of hematopoietic stem cells,” *Oncotarget*, vol. 9, no. 90. Impact Journals LLC, pp. 36052–36054, 01-Nov-2018.
[309] D. Binns, E. Dimmer, R. Huntley, D. Barrell, C. O’Donovan, and R. Apweiler, “QuickGO: A web-based tool for Gene Ontology searching,” *Bioinformatics*, vol. 25, no. 22, pp. 3045–3046, 2009.
[310] M. J. Delás *et al.*, “lncRNA Spehd Regulates Hematopoietic Stem and Progenitor Cells and Is Required for Multilineage Differentiation,” *Cell Rep.*, vol. 27, no. 3, pp. 719-729.e6, 2019.
[311] A. Schwarzer *et al.*, “The non-coding RNA landscape of human hematopoiesis and leukemia,” *Nat. Commun.*, vol. 8, no. 1, pp. 1–17, Dec. 2017.
[312] M. C. Schmid *et al.*, “Integrin CD11b activation drives anti-tumor innate immunity,” *Nat. Commun.*, vol. 9, no. 1, pp. 1–14, Dec. 2018.
[313] A. Leung *et al.*, “Notch and Aryl Hydrocarbon Receptor Signaling Impact Definitive Hematopoiesis from Human Pluripotent Stem Cells,” *Stem Cells*, vol. 36, no. 7, pp. 1004–1019, Jul. 2018.
[314] N. An *et al.*, “Pim1 serine/threonine kinase regulates the number and functions of murine hematopoietic stem cells,” *Stem Cells*, vol. 31, no. 6, pp. 1202–1212, Jun. 2013.
[315] J. A. Schardt, M. Eyholzer, N. A. Timchenko, B. U. Mueller, and T. Pabst, “Unfolded protein response suppresses CEBPA by induction of calreticulin in acute myeloid leukaemia,” *J. Cell. Mol. Med.*, vol. 14, no. 6 B, pp. 1509–1519, Jun. 2010.
[316] K. Chotinantakul, P. Prasajak, and W. Leeanansaksiri, “Wnt1 accelerates an ex vivo expansion of human cord blood CD34 +CD38- cells,” *Stem Cells Int.*, vol. 2013, 2013.
[317] F. Rapino, E. F. Robles, J. A. Richter-Larrea, E. M. Kallin, J. A. Martinez-Climent, and T. Graf, “C/EBPα Induces Highly Efficient Macrophage Transdifferentiation of B Lymphoma and Leukemia Cell Lines and Impairs Their Tumorigenicity,” *Cell Rep.*, vol. 3, no. 4, pp. 1153–1163, Apr. 2013.
[318] T. Å. Pedersen, E. Kowenz-Leutz, A. Leutz, and C. Nerlov, “Cooperation between C/EBPα TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation,” *Genes Dev.*, vol. 15, no. 23, pp. 3208–3216, 2001.

[319] S.-S. Chen, J.-F. Chen, P. F. Johnson, V. Muppala, and Y.-H. Lee, " C/EBPβ, When Expressed from the C/ebp α Gene Locus, Can Functionally Replace C/EBPα in Liver but Not in Adipose Tissue ," *Mol. Cell. Biol.*, vol. 20, no. 19, pp. 7292–7299, Oct. 2000.
[320] M. L. Chrétien *et al.*, "Trim33/Tif1γ is involved in late stages of granulomonopoiesis in mice," *Exp. Hematol.*, vol. 44, no. 8, pp. 727-739.e6, 2016.
[321] L. Han *et al.*, "Chromatin remodeling mediated by ARID1A is indispensable for normal hematopoiesis in mice," *Leukemia*, vol. 33, no. 9, pp. 2291–2305, Mar. 2019.
[322] T. Bluemn *et al.*, "Arid2 regulates hematopoietic stem cell differentiation in normal hematopoiesis," *Exp. Hematol.*, vol. 94, pp. 37–46, Feb. 2021.
[323] S. J. Baker *et al.*, "B-myb is an essential regulator of hematopoietic stem cell and myeloid progenitor cell development," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 8, pp. 3122–3127, Feb. 2014.
[324] E. Kowenz-Leutz, O. Pless, G. Dittmar, M. Knoblich, and A. Leutz, "Crosstalk between C/EBPB phosphorylation, arginine methylation, and SWI/SNF/Mediator implies an indexing transcription factor code," *EMBO J.*, vol. 29, no. 6, pp. 1105–1115, 2010.
[325] S. Albini and P. L. Puri, "SWI/SNF complexes, chromatin remodeling and skeletal myogenesis: It's time to exchange!," *Exp. Cell Res.*, vol. 316, no. 18, pp. 3073–3080, Nov. 2010.
[326] S. E. LeBlanc *et al.*, "Protein arginine methyltransferase 5 (Prmt5) promotes gene expression of peroxisome proliferator-activated receptor γ2 (PPARγ2) and its target genes during adipogenesis," *Mol. Endocrinol.*, vol. 26, no. 4, pp. 583–597, Apr. 2012.
[327] S. Pal, S. N. Vishwanath, H. Erdjument-Bromage, P. Tempst, and S. Sif, " Human SWI/SNF-Associated PRMT5 Methylates Histone H3 Arginine 8 and Negatively Regulates Expression of ST7 and NM23 Tumor Suppressor Genes ," *Mol. Cell. Biol.*, vol. 24, no. 21, pp. 9630–9645, Nov. 2004.
[328] V. Karkhanis, L. Wang, S. Tae, Y. J. Hu, A. N. Imbalzano, and S. Sif, "Protein arginine methyltransferase 7 regulates cellular response to DNA damage by methylating promoter histones H2A and H4 of the polymerase δ catalytic subunit gene, POLD1," *J. Biol. Chem.*, vol. 287, no. 35, pp. 29801–29814, Aug. 2012.
[329] H. J. Jeong *et al.*, "Prmt7 promotes myoblast differentiation via methylation of p38MAPK on arginine residue 70," *Cell Death Differ.*, vol. 27, no. 2, pp. 573–586, Jun. 2020.
[330] A. E. Raposo and S. C. Piller, "Protein arginine methylation: An emerging regulator of the cell cycle," *Cell Div.*, vol. 13, no. 1, pp. 1–16, 2018.
[331] E. Dolezal *et al.*, "The BTG2-PRMT1 module limits pre-B cell expansion by regulating the CDK4-Cyclin-D3 complex," *Nat. Immunol.*, vol. 18, no. 8, pp. 911–920, Jun. 2017.
[332] S. Zheng *et al.*, "Arginine methylation-dependent reader-writer interplay governs growth control by E2F-1," *Mol. Cell*, vol. 52, no. 1, pp. 37–51, Oct. 2013.
[333] Y. H. Lee, M. T. Bedford, and M. R. Stallcup, "Regulated recruitment of tumor suppressor BRCA1 to the p21 gene by coactivator methylation," *Genes Dev.*, vol. 25, no. 2, pp. 176–188, Jan. 2011.
[334] Y. Zhang, Y. Zhang, Y. J. Zhang, H. Y. Zhao, Q. L. Zhai, and Y. F. Shen, "The impact of R213 mutation on p53-mediated p21 activity," *Biochimie*, vol. 99, no. 1, pp. 215–218, Apr. 2014.
[335] H. Wang *et al.*, "C/EBPα arrests cell proliferation through direct inhibition of Cdk2 and Cdk4," *Mol. Cell*, vol. 8, no. 4, pp. 817–828, Oct. 2001.
[336] N. A. Timchenko *et al.*, "CCAAT/enhancer binding protein alpha regulates p21 protein

and hepatocyte proliferation in newborn mice," *Mol. Cell. Biol.*, vol. 17, no. 12, pp. 7353–7361, 1997.

[337] C. Müller, M. Alunni-Fabbroni, E. Kowenz-Leutz, X. Mo, M. Tommasino, and A. Leutz, "Separation of C/EBPα-mediated proliferation arrest and differentiation pathways," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 13, pp. 7276–7281, Jun. 1999.

[338] Z. Wang, N. Bhattacharya, P. F. Mixter, W. Wei, J. Sedivy, and N. S. Magnuson, "Phosphorylation of the cell cycle inhibitor p21Cip1/WAF1 by Pim-1 kinase," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1593, no. 1, pp. 45–55, Dec. 2002.

[339] D. Morishita, R. Katayama, K. Sekimizu, T. Tsuruo, and N. Fujita, "Pim kinases promote cell cycle progression by phosphorylating and down-regulating p27kip1 at the transcriptional and posttranscriptional levels," *Cancer Res.*, vol. 68, no. 13, pp. 5076–5085, Jul. 2008.

[340] M. Asada *et al.*, "Apoptosis inhibitory activity of cytoplasmic p21(Cip1/WAF1) in monocytic differentiation," *EMBO J.*, vol. 18, no. 5, pp. 1223–1234, 1999.

[341] V. Krasteva, G. R. Crabtree, and J. A. Lessard, "The BAF45a/PHF10 subunit of SWI/SNF-like chromatin remodeling complexes is essential for hematopoietic stem cell maintenance," *Exp. Hematol.*, vol. 48, pp. 58-71.e15, 2017.

[342] N. V. Soshnikova *et al.*, "PHF10 subunit of PBAF complex mediates transcriptional activation by MYC," *Oncogene*, vol. 40, no. 42, pp. 6071–6080, Aug. 2021.

[343] S. Wu *et al.*, "Myc represses differentiation-induced p21CIP1 expression via Miz-1-dependent interaction with the p21 core promoter," *Oncogene*, vol. 22, no. 3, pp. 351–360, Jan. 2003.

[344] L. Schmidt *et al.*, "CEBPA-mutated leukemia is sensitive to genetic and pharmacological targeting of the MLL1 complex," *Leukemia*, vol. 33, no. 7, pp. 1608–1619, 2019.

[345] P. B. Crowley and A. Golovin, "Cation-π interactions in protein-protein interfaces," *Proteins Struct. Funct. Genet.*, vol. 59, no. 2, pp. 231–239, May 2005.

[346] J. P. Gallivan and D. A. Dougherty, "Cation-π interactions in structural biology," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 17, pp. 9459–9464, Aug. 1999.

[347] S. Qamar *et al.*, "FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions," *Cell*, vol. 173, no. 3, pp. 720-734.e15, Apr. 2018.

[348] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, "Biomolecular condensates: Organizers of cellular biochemistry," *Nat. Rev. Mol. Cell Biol.*, vol. 18, no. 5, pp. 285–298, May 2017.

[349] S. Boeynaems *et al.*, "Protein Phase Separation: A New Phase in Cell Biology," *Trends Cell Biol.*, vol. 28, no. 6, pp. 420–435, 2018.

[350] T. J. Nott *et al.*, "Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles," *Mol. Cell*, vol. 57, no. 5, pp. 936–947, Mar. 2015.

[351] X. Su *et al.*, "Phase separation of signaling molecules promotes T cell receptor signal transduction," *Science (80-.).*, vol. 352, no. 6285, pp. 595–599, Apr. 2016.

[352] A. Boija *et al.*, "Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains," *Cell*, vol. 175, no. 7, pp. 1842-1855.e16, 2018.

[353] A. M. Erkine, "'Nonlinear' Biochemistry of Nucleosome Detergents," *Trends Biochem. Sci.*, vol. 43, no. 12, pp. 951–959, 2018.

[354] B. K. Broyles *et al.*, "Activation of gene expression by detergent-like protein domains," *iScience*, vol. 24, no. 9, p. 103017, 2021.

[355] R. Métivier *et al.*, “Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter,” *Cell*, vol. 115, no. 6, pp. 751–763, Dec. 2003.
[356] M. Le Romancer *et al.*, “Regulation of Estrogen Rapid Signaling through Arginine Methylation by PRMT1,” *Mol. Cell*, vol. 31, no. 2, pp. 212–221, Jul. 2008.
[357] Y. Shi and J. R. Whetstine, “Dynamic Regulation of Histone Lysine Methylation by Demethylases,” *Mol. Cell*, vol. 25, no. 1, pp. 1–14, 2007.
[358] J. Fuhrmann, K. W. Clancy, and P. R. Thompson, “Chemical Biology of Protein Arginine Modifications in Epigenetic Regulation,” *Chem. Rev.*, vol. 115, no. 11, pp. 5413–5461, 2015.
[359] G. L. Cuthbert *et al.*, “Histone deimination antagonizes arginine methylation,” *Cell*, vol. 118, no. 5, pp. 545–553, Sep. 2004.
[360] Q. Liu, “Post-translational modification on arginine and function of CCAAT / enhancer binding protein alpha,” Humboldt University, 2012.
[361] S. M. Carlson and O. Gozani, “Emerging technologies to map the protein methylome,” *J. Mol. Biol.*, vol. 426, no. 20, pp. 3350–3362, Oct. 2014.
[362] T. Schenk *et al.*, “Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia,” *Nat. Med.*, vol. 18, no. 4, pp. 605–611, Mar. 2012.
[363] H. Lin *et al.*, “Small molecule KDM4s inhibitors as anti-cancer agents,” *J. Enzyme Inhib. Med. Chem.*, vol. 33, no. 1, pp. 777–793, Jan. 2018.
[364] D. H. Lee, G. W. Kim, Y. H. Jeon, J. Yoo, S. W. Lee, and S. H. Kwon, “Advances in histone demethylase KDM4 as cancer therapeutic targets,” *FASEB J.*, vol. 34, no. 3, pp. 3461–3484, Mar. 2020.
[365] R. J. Wheeler, “Therapeutics-how to treat phase separation-associated diseases,” *Emerg. Top. Life Sci.*, vol. 4, no. 3, pp. 331–342, Dec. 2020.
[366] A. Mullard, “Biomolecular condensates pique drug discovery curiosity,” *Nat. Rev. Drug Discov.*, Apr. 2019.

Acknowledgement

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Achim Leutz, for his invaluable guidance, support and encouragement throughout this study.

To the following people, whom I have met and learned from, I thank:
Karin, for the great contribution to this project with her skills, and also for her creativity and positivity.
Akki, who was always willing to share his knowledge, bear with my questions and sing me a birthday song.
Thea, for her great help with the sequencing experiments and everything else inside and outside the lab, for being the best fellow I could ever wish for.
Elisabeth, for providing me with the materials and extremely helpful advice.
Maria H., Jelena and Alexandra, who were my first teachers in the lab, and still are.
Lab members and friends, Radek, Valeriia, Milda, Karina, Vicky, Ruth and Evelyn, for all the advice, conversations and support that helped me through the up and down of my journey. I sincerely treasure the time we spent together.

I wish to thank my Advisory Committee: Dr. Baris Tursun and Dr. Martin Janz, for the thoughtful comments on this project. I would like to thank my Doctoral Degree Committee for taking time reviewing my work and contributing their expert insights.

I would like to acknowledge the Berlin School of Integrative Oncology (BSIO), Max Delbrück Center for Molecular Medicine (MDC) and Berliner Krebsgesellschaft e.V. for funding; the Flow Cytometry technology platform (MDC), Genomics platform (MDC) and the EMBL Genomics Core Facility for technical support.

A thank to my friends: Kiều, Mám, Tình, Tố, for the calls across time zones. Finally, my most special thanks to my family and my girlfriend, Nhi, for always be the source of my strength. Cảm ơn bố mẹ vì luôn tin tưởng và động viên con.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe. Die Stellen der Arbeit, die anderen Werken in Wortlaut oder Sinn nach entnommen wurden, sind in jedem Einzelfall kenntlich gemacht.

Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad. Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde. Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt Universität zu Berlin vom 5. März 2015.

Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsbearbeiterinnen/Promotionsberatern stattgefunden hat und dass die Grundsätze der Humboldt Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Berlin, den ..... Nguyen Thuy Linh .....

Publications

Le, P. H. A., Nishimura, K., Kuno, A., **Nguyen, L. T.**, Tetsuo, K., Fukuda, A., & Hisatake, K. (2021). Downregulation of Odd-skipped Related 2, a novel regulator of epithelial- mesenchymal transition, enables efficient somatic cell reprogramming. *Stem Cell Reports*. (Manuscript in revision).

Do, H. Q., Truong, N. H., Lam, T. T., **Nguyen, L. T.**, Le, D. M., Dinh, N. H.-T., … Pham, P. Van. (2019). The effects of the *Panax Vietnamensis* ethanol fraction on proliferation and differentiation of mouse neural stem cells. *Biomedical Research and Therapy*, *6*(10), 3422–3432. https://doi.org/10.15419/bmrat.v6i10.571.

Truong N.H., Dinh, N., Le, D., **Nguyen, L. T.**, Lam, T., Phan, N., & Pham, P. (2014). Isolation and culture of neural stem cells from murine foetal brain. *Research Opinions in Animal and Veterinary Sciences*, *4*(1), 24–29.