The transcriptional regulation of normal and malignant blood cell development

Edginton-White, Benjamin; Bonifer, Constanze

Link to publication on Research at Birmingham portal

STATE-OF-THE-ART REVIEW

The transcriptional regulation of normal and malignant blood cell development

Benjamin Edginton-White and Constanze Bonifer [iD]

Institute of Cancer and Genomic Sciences, College of Medicine and Dentistry, University of Birmingham, UK

Development of multicellular organisms requires the differential usage of our genetic information to change one cell fate into another. This process drives the appearance of different cell types that come together to form specialized tissues sustaining a healthy organism. In the last decade, by moving away from studying single genes toward a global view of gene expression control, a revolution has taken place in our understanding of how genes work together and how cells communicate to translate the information encoded in the genome into a body plan. The development of hematopoietic cells has long served as a paradigm of development in general. In this review, we highlight how transcription factors and chromatin components work together to shape the gene regulatory networks controlling gene expression in the hematopoietic system and to drive blood cell differentiation. In addition, we outline how this process goes astray in blood cancers. We also touch upon emerging concepts that place these processes firmly into their associated subnuclear structures adding another layer of the control of differential gene expression.

Introduction

Hematopoiesis is one of the best understood developmental pathways [1,2] and has extensively been studied in mice. The origin of blood cell development in the embryo is the mesodermal germ layer in the mammalian embryo, and hematopoietic specification occurs in two waves: The first wave takes place in the extraembryonic blood islands of the yolk sac and gives rise to primitive progenitor cells with mostly erythroid and myeloid potential [3,4]; the second wave gives rise to definitive hematopoietic stem cells (HSCs) and takes place at the ventral part of the dorsal aorta in the aorta–gonad–mesonephros (AGM) region of the embryo [5]. Cells emerging during the second wave migrate first to the fetal liver and later to the bone marrow. Here, they are maintained in a specialized niche and are largely quiescent, and if growing, either self-renew or enter differentiation to sustain mature blood cell production throughout lifetime. All HSCs are born from a specialist endothelial cell layer, the hemogenic endothelium (HE), which communicates with the dorsal mesenchyme. In response to signals, HE cells undergo a cellular shape transition, the endothelial–hematopoietic transition (EHT), forming intra-aortic clusters, which undergo several maturation steps before floating off into the bloodstream. Blood cell development therefore involves a carefully regulated cascade of gene expression changes that are regulated by molecular mechanisms linking genomic responses to a multitude of signals coming from the outside. It matters, where a cell has been and who it has talked to. In turn, it matters whether a cell is responsive to an outside signal, making development

and differentiation an intricate, but highly robust balancing act that occurs in multiple cells at the same time. This review will summarize seminal studies, which uncovered the players involved in this balancing act and highlight the notion that signaling-responsive transcription regulation and chromatin dynamics are at the heart of the mechanisms maintaining and changing cellular identity. We will also highlight that perturbing any of these mechanisms leads to a disturbance of differentiation and, in some cases, to the development of malignant cells that have opted out of normal growth and differentiation control.

Transcription factors control blood cell development and differentiation

Blood cell lineage-specific gene expression is under the control of specific transcription factors. A large number of studies employing genetically modified mice showed that the absence of lineage specifically expressed factors leads to a perturbation of differentiation or a complete absence of the respective lineage. One of the first examples of a knockout removing an important TF was that of the erythroid-specifically expressed TF GATA1, which led to a complete absence of erythroid (and megakaryocytic) differentiation, while other lineages appeared to be unperturbed [6]. The underlying molecular mechanism of differentiation defects in the absence of a lineage-determining TF is the deregulation of genes carrying binding sites for this TF. Knockout experiments also highlighted the fact that TFs act in a hierarchical fashion. The elimination of earlier acting TFs such as TAL1 or RUNX1 affects HSC formation and thus the development of the entire hematopoietic system, whereas elimination of others such as PU.1 largely affects the myeloid and B-cell lineages. Recent studies using single-cell RNA-sequencing approaches have visualized the successive activation of specific developmental trajectories by performing 'nearest neighbor' analyses, which determine changes in the gene expression patterns of single cells and order them according to the direction of increased maturation. Such analyses highlight the different branches of the hematopoietic system and show that they deviate from each other earlier than previously thought [7,8]. Performing such studies with hematopoietic cells lacking a lineage-determining factor clearly showed the absence of specific branches in the trajectory [9]. Another hallmark of the hierarchical action of TFs is the finding that they are often only critically required at specific stages of development even if still expressed at other stages. Examples for this notion are again RUNX1 and TAL1, and

the removal of their genes from the germ line strongly blocks HSC development, but when removing it conditionally after HSCs have formed, their maintenance is not affected [10–12].

However, such a clear-cut result is not seen with all TF knockouts and the reason for this behavior is the fact that TFs operate within large interacting protein assemblies as explained in further detail below. TFs have a modular structure with different domains that fulfill different functions and interact with different proteins. Recent studies showed that crippling TFs by removing individual domains can have unexpected effects that shed light on their actual function in gene regulation and point to an amazing robustness of protein complex formation driving gene expression. An example is again the transcription factor TAL1. Deletion of the whole factor abolishes HSC emergence completely, but deletion of the DNA-binding domain alone has a much milder phenotype and factor binding can be detected at a subset of genomic targets [13]. A different result was observed after removing the DNA-binding domain from the ubiquitously expressed TF SP1. This mutation affects all developmental pathways, and a germ-line mutation is an embryonic lethal [14]. However, in contrast to lineage-determining factors, removing it conditionally later in development had very little effect [15]. The explanation for this finding came with the analysis of the differentiation of mouse embryonic stem cells into blood precursors *in vitro*, which showed that the knockout still expressed a truncated protein and that the effect of the mutation on gene expression was cumulative. The full knockout of the *Sp1* gene was incompatible with differentiation and so was the full deletion of the SP1 orthologue SP3 in a SP1 hypomorphic genetic background, indicating that the truncated version of SP1 needed SP3 to function. During the differentiation of cells expressing a truncated Sp1, bulk gene expression patterns of purified cells became more and more diverse. Single-cell RNA-Seq experiments and the analysis of differentiation trajectories of such cells showed why this was the case: Cells entered the correct gene expression trajectory, but seemed to do this at different time points, forming transcriptionally diverse cell populations. In essence, cells do not execute cell fate decisions as a cohort, meaning that robustness of differentiation was lost [15]. However, the system could only tolerate a certain level of deregulation: Once past the progenitor stage, differentiation crashed, and mutant cells were unable to form terminally differentiated blood cells [16]. A break-down of robustness of differentiation can also be seen when another crucial level of control of differentiation is disturbed: the expression of correct TF levels. Many crucial TFs show haploinsufficiency

phenotypes when one genetic copy is deleted or expression levels are reduced by the mutation of important cis-regulatory elements, with the kinetics of development being perturbed. This notion is true for GATA2 whose downregulation by the mutation of an essential enhancer [17] or by haploinsufficiency [18] causes various hematopoietic defects and predisposes to leukemia. The latter is also true for a cis-regulatory mutation of the gene encoding PU.1, *SPI1* [19]. Last, but not least, RUNX1 needs to be expressed at carefully controlled levels to drive hematopoietic differentiation and specification [20,21].

Taken together, these studies show that the effects of the crippling of an essential TF or its gene on gene expression control have to be seen within the context of development being a dynamic and highly robust process. The system is composed of large interacting protein assemblies and partly redundant components, which compensate until they fall apart or are malfunctioning, meaning that the defect occurs way before phenotypic alterations can be seen. The current challenge is to identify the point when this occurs. This notion will become important when trying to interpret how mutant transcription factors set differentiating cells on the path to cancer.

Transcription factors collaborate and respond to signals

TFs come in families that bind to specific DNA-binding motifs within cis-regulatory elements such as enhancers and promoters, which are responsible for determining how a gene is regulated and when and in which cell type it is expressed. Each regulatory region contains multiple TF-binding motifs, which often are highly conserved depending on the nature of a gene and whether its function is conserved in evolution. A good example is the 'Heptad', a consortium of co-localizing transcription factors such as GATA2, TAL1, RUNX1, and FLI1 and the bridging factors LDB1/LMO2 that specify the cis-regulatory elements of genes expressed in hematopoietic stem and progenitor cells [22,23]. The spatial arrangement of TF-binding sites is often not conserved [24], but there are exceptions with TFs that directly interact on DNA and whose binding is interdependent. Here, the spacing of binding motifs can be very precise, as, for example, seen with the pair AP-1/TEAD in the hemogenic endothelium [25], the pair RUNX1/ETS1 [26], or the GATA/E-Box motifs within the heptad [27].

As exemplified by SP1, ubiquitously expressed TFs cooperate with tissue-specific factors to set up differential gene expression patterns, and thus, the binding patterns of TFs are highly specific for each cell type [28].

Importantly, binding patterns are highly dynamic and can be maintained in self-renewing cells [29] or change during development [30,31]. In this context, it is noteworthy that transcription is not a uniform process but occurs in bursts that are regulated by the burst frequency, indicating that genes are in intricate contact with their environment [32,33]. The cell receives signals from various sources, which trigger developmental changes and are integrated within the genome by the action of inducible and signaling-responsive TFs. Many of these factors can be activated in all cells and include the AP-1 (JUN/FOS) factor families, which respond to MAP kinase signaling [34], STATs responding to cytokine receptor signaling [35], SMADs mediating TGF-β signaling [36,37], TEAD/YAP mediating Hippo signaling [38], or NFAT family members (linked to Ca++-signaling) [39]. As a result of the activation of such factors, enhancer elements can be activated de novo, or become more active, driving increased levels of gene expression. However, note that also noninducible TFs present can be regulated in their activity by signaling-dependent post-translational modifications such as phosphorylation with RUNX1 being a prominent example [40]. While we have a fairly good idea about what regulates the activity of inducible transcription factors, we know very little about how the different modes of signal transmission interplay with each other across the genome. This notion becomes important when different signals are being integrated at the genome level by regulating TF binding. For example, the abolition of AP-1 binding during hematopoietic specification by using a dominant-negative version of FOS led to a loss of binding of the Hippo signaling-responsive factor TEAD at the composite genomic sites described above [25]. Hippo signaling is activated by the onset of blood flow, which creates biomechanical forces stimulating Rho-GTPase signaling [41]. During T-cell activation, MAP kinase and Ca++ signaling are integrated by a cooperation of AP-1 and NFAT and at a specific subset of sites with composite-binding motifs one factor cannot bind without the other [42,43], thus ensuring that genes respond only when both signals are present. These few examples show that we are only now starting to obtain a glimpse of the principles and staggering complexity of how the multitude of signaling inputs that a cell encounters are integrated within the genome and shape a genomic response.

TFs can both activate and repress gene expression and form dynamic gene regulatory networks

A large number of Zn++ finger TFs are bona fide repressors with REST, a factor that is required to

repress neuronal genes in other tissues, being a prominent example [44]. Other factors can activate or repress depending on the genomic context, which determines whether they recruit co-activator or co-repressors (see below) or interfere with the activity of lineage-determining TFs that set up alternate gene expression patterns. Examples for the latter are the B-cell commitment factors PAX5 and GATA2. PAX5 is required to activate the expression of B-cell-specific genes, but at the same time represses the expression of myeloid genes [45,46] and thus finalizes commitment. GATA2 activates multiple hematopoietic genes but is required for the repression of cardiac genes [47]. PU.1 and GATA-1 form a similar antagonistic pair during erythropoiesis [48,49]. A very interesting example of how to turn an activator into a repressor during dynamic gene activation is provided by Mylona *et al.* [50], who showed that the type of response of the serum-responsive TF ELK1 depends on the timing of its post-translational modification. The protein contains multiple ERK kinase-dependent phosphorylation sites that are modified with different kinetics, fast, intermediate, and slow. After the fast sites are modified, co-activators and mediator are recruited and genes are activated, once the slower sites are modified, co-repressors are recruited and gene expression is switched off. Many other TFs contain multiple phosphorylation sites as well, making it highly likely that such dynamic behavior of factors responding to signaling is widespread and ensures that gene expression does not overshoot.

An important feature of TF function is the fact that they can bind to genes encoding other TFs and form gene regulatory networks (GRNs), and this is also true for blood cells [31,51,52]. In order to be able to construct such a network, it is necessary to identify regulatory relationships between TFs and their target genes. This aim can be achieved by inference, a strategy by which the expression of individual putative regulators is perturbed followed by determining which genes respond and whether they are upregulated or repressed [53]. A more direct way is to identify the actual TF-binding events using *in vivo* footprinting [54] or chromatin immunoprecipitation (ChIP) assays and then link binding sites to their associated genes. GRNs consist of nodes (TFs or TF families) that are interconnected (edges), all of which bind to non-TF genes that actually specify a cell type (Fig. 1). It is now clear that highly interconnected nodes are important for the maintenance of a specific cell type. Moreover, it matters how the network is structured and how the different components are wired. Factors binding and regulating their own and other TF-encoding genes can form recursively wired circuits, thus carefully controlling their expression levels and binding patterns, and are a hallmark of self-renewing hematopoietic stem and progenitor cells [29]. However, when differentiation is kicked off by a signal or by the upregulation of expression of a specific TF, connections are altered, meaning that TF-binding patterns are altered as well, and factors move to different locations [31]. A striking example of how the expression of one factor can drive the rewiring of an entire GRN is provided by experiments that expressed an inducible version of RUNX1 in mouse embryonic stem cells with a RUNX1 null genetic background. Differentiation of such cells to blood cells *in vitro* is blocked at the hemogenic endothelium stage. Induction of RUNX1 allows the differentiation of the hematopoietic progenitor stage and, importantly, leads to a genome-wide relocation of other factors, such as TAL1, LDB1, and FLI1 to different locations close to RUNX1-binding sites. Importantly, at least at the early stages of induction, relocation was reversible. Rewiring and RUNX1-dependent differentiation into blood progenitors require the chromatin reader BRD4 with the final complex recruiting mediator and CDk9 kinase to activate transcription [30,55]. The analogy of GRNs with differentially wired circuits and the availability of global binding and gene expression data have attracted the attention of computational biologists and mathematical modelers who strive to create models that could predict the behavior of GRNs in response to perturbation [56]. However, these efforts face formidable challenges, both experimentally and bioinformatically. For example, due to the signaling responsiveness of many TFs and their cofactors, their binding does not necessarily mean that a gene is expressed. So far, these models are therefore only capable of predicting simple subaspects of gene expression control, such as whether a gene is likely to be expressed or not [57]. Due to the multiple parameters feeding into the system, predicting gene expression patterns in a dynamic or even a developmental context is so far out of reach. However, such methodology is essential, if we want to predict the response of a GRN to changes in transcription factor binding as a result of DNA sequence changes, changes in the signaling environment, or in perturbation experiments such as drug treatment.

Transcription factors interact with a specific chromatin landscape

The most important feature of TFs is that they recognize specific DNA sequences and therefore are able to read the genetic code. However, within the eukaryotic
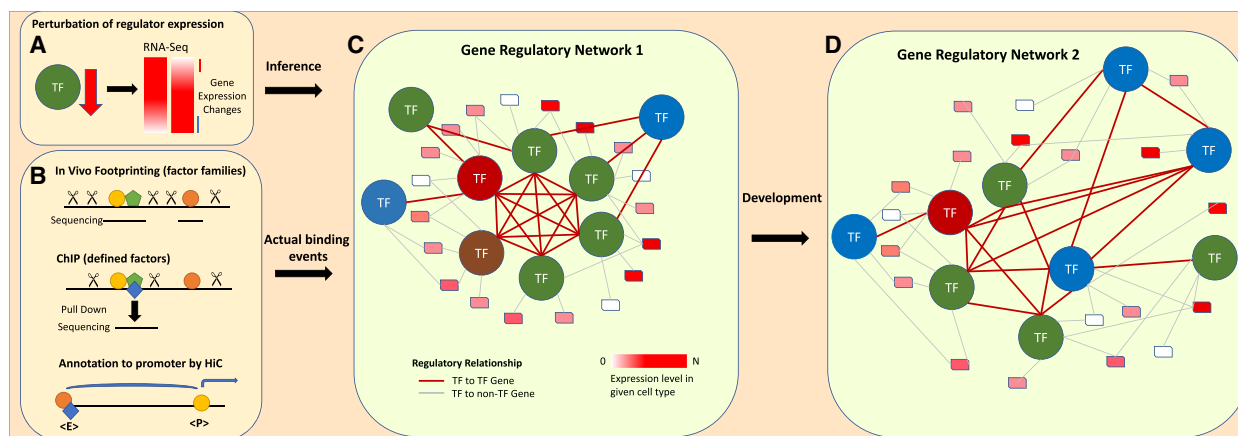
Fig. 1. Development involves the alterations of gene regulatory networks. (A). Inferring regulatory relationships between transcription factor genes by perturbation of the expression of one factor and measuring the gene expression response of all genes. Note that such experiments require multiple measurements to identify statistically significant correlations. (B) Identifying regulatory relationships by direct binding experiments as described in Ref. [119]. This includes digital footprinting, which identifies factor families binding to the same motif or identifying the precise factor by using chromatin immunoprecipitation assays. Identified binding sites are then annotated to their rightful promoter using promoter capture HiC (chromosome conformation capture). Compared with gene expression analysis, this strategy allows to identify the differential wiring of GRNs in different cell types as depicted in C and D. Genes bound by the different TFs or TF families are depicted as rectangles with the color highlighting their expression level as indicated in the figure.

nucleus they encounter a formidable obstacle to this process in the form of chromatin. Here, DNA is wrapped around nucleosomes, which are then packaged into higher-order structures of differential compaction, depending on whether the genes within these structures are active, potentially active, or stashed away in heterochromatin. In order for the genetic code to be accessed by TFs, chromatin needs to be remodeled and modified, which is achieved by different mechanisms. One mechanism is the opening of chromatin by pioneer factors, which are capable of binding to nucleosomal DNA and then cooperate with other factors to nucleate a transcription factor complex [58]. Other TFs interact with nucleosomes in different ways with most factors binding the nucleosomal linker regions [59]. All binding modes have in common that after a stable TF assembly is established, TFs recruit chromatin remodelers such as SWI/SNF complexes that use ATP to 'peel' DNA off the nucleosome and free up sequences for further binding [60]. A variant of the second mechanism is 'assisted loading' whereby an inducible factor binds, recruits chromatin remodelers that enable the binding of a second factor which cannot normally bind, and then leaves again, leaving stably remodeled and TF-bound chromatin behind [61,62]. During the assembly process, TF complexes recruit further cofactors such as histone acetyltransferases (HATs) that facilitate transcription by modifying the N-terminal tails of the surrounding

nucleosomes and stabilize an open chromatin structure that is devoid of nucleosomes and exists as a nuclease hypersensitive site [63]. Given the importance of chromatin remodelers and modifiers in gene activation, it does not come as a surprise that these proteins are essential components of the regulatory machinery driving hematopoiesis [64–66].

The establishment of stable TF complexes and modified chromatin is not the only mechanism that is required to activate transcription. TF and cofactor complexes at the different enhancers and the promoter of a gene contact each other within nuclear space [67] and form large protein–DNA complexes on cis-regulatory elements that contain all the factors necessary to activate mRNA synthesis by RNA polymerase, and form a regulatory unit or chromatin hub [68]. The architecture of such units can be simple or complex—depending on the complexity of gene regulation during development and in different tissues [69,70]. The reason for such complexity is that during development, genes can be regulated by a relay of differentially/tissue-specifically active cis elements. A good example for this notion is the chicken lysozyme locus, which is expressed in the oviduct or in macrophages and uses different and shared tissue-specific elements and factors to drive different regulatory modes of gene expression [71]. Moreover, even genes that are expressed in every cell, that is, 'housekeeping genes', are regulated by a relay of different factors thus keeping chromatin

open and ensuring their sustained activity [31]. The latter mechanism highlights several important concepts in gene regulation: An active, transcriptionally permissive chromatin structure has to be actively maintained. In the absence of activators, an inactive chromatin structure is established by repressing factors such as DNA methyltransferase and histone deacetylases, which methylate DNA and remove the acetylation mark from histones. Secondly, an active chromatin pattern that is nuclease accessible and carries active histone marks is cell-type specific. Finally, it is not the promoters, but the nonpromoter elements that contain the information of tissue-specific gene expression and mirrors tissue-specific gene expression patterns [72,73].

Each transcription cycle is regulated by the balance of activating and repressing factors responding to outside signals [74]. A large number of genes maintain their transcriptionally active structure throughout cell division. However, during mitosis, TF complexes are largely stripped off chromatin and the question arises how they reform. It is now clear that the parent set of modified histones are distributed to the two daughter strands. Modification patterns are therefore retained during mitosis and mark genes that are activated after mitosis [75]. It is also clear that certain TFs, such as FOXA1, are capable of binding to mitotic chromatin and form the basis of re-assembled TF complexes creating an active chromatin structure once the nuclear environment has been reformed [74,76]. This transcriptional memory is often dependent on signaling processes, as shown during the formation of T-cell memory: Stable TF binding allowing rapid reactivation of genes by a second stimulus is dependent on the constant reinforcement of factor binding by cytokine signaling, employing inducible TFs. The absence of cytokine signaling leads to a loss of an active and transcriptionally permissive chromatin structure [77]. A similar transcriptional memory is also established in macrophages after a first inflammatory stimulus [78]. In a developmental context, this interplay of inducible and constitutive factors establishing an early memory of a previously received signal, also referred to as priming [79], plays a decisive role in changing or maintaining cell identities, as exemplified by neuronal development of C. elegans. The developmental timing of regulation of the Lsy-6 miRNA locus is dictated by a NOTCH responsive an early enhancer. Those neuronal precursor cells receiving the signal upregulate the gene earlier as compared to those who did not with a strong impact on gene expression patterns. The result is a functional left–right asymmetry in otherwise morphologically symmetric neurons [80]. Developing blood cells are embedded in a sea of signals that have a profound impact on gene expression. One of the challenges in the next years will be to unravel the order of events of how genes are activated in hematopoietic development and how external signals such as soluble factors, mechanical forces, and spatial context regulate the ordered formation of HSCs, cells of the different hematopoietic lineages, and hematopoietic tissues such as the thymus and lymph nodes. Single-cell analyses of chromatin changes and expression patterns in developing cells together with spatial information will be crucial to answer these questions [9,81]. Such studies need to be combined with studies of surface molecule mapping [82] and the analysis of intracellular signaling processes using advanced imaging—a formidable task.

Gene regulatory processes take place in different parts of the nucleus

Gene regulation cannot be viewed without taking into account where it takes place—in the nucleus (Fig. 2). In recent years, it has become clear that this organelle displays a highly organized structure, with genes occupying different compartments depending on their activity state, the nature of their neighbors, and whether they are transiently or permanently silenced [83–85]. The latter distinction is important, because transcription can be rapidly switched off with genes remaining in a poised state ready for further activation of repression, which is mediated by polycomb-repressive complexes (PRCs). PRC complexes come as two general types, PRC1 and PRC2. PRC2 contains the EZH1/2 methyltransferase, which deposits methyl groups on histone H3K27. H3K27me3 binds the PRC1 complex, which then ubiquitinates histone H2A at target promoters resulting in a block of transcriptional elongation by RNA polymerase II, with the nonelongating form of RNA polymerase still being associated at these sites [86,87]. PRC complexes at promoters interact with each other in nuclear space and form a long-range network of transcriptionally silent genes [88,89] that can intermingle with active genes [90] to rapidly switch from one state to another. In contrast, true heterochromatic regions such as centromeres, telomeres, repeat elements, and genes that are stably silenced display a highly compact chromatin structure, are not bound by RNA PolII, and are associated with the nuclear periphery and the nuclear lamina [91].

Another level of chromosomal organization of higher eukaryotes, which is associated with differential gene expression, are topologically associated domains (TADs) [92,93]. TADs partition chromosomes into regulatory domains inhibiting interactions between neighboring chromosomal regions. TADs are in average

Fig. 2. Gene expression is controlled by transcription factors, chromatin components, and signaling and takes place in specialized compartments within the nucleus. Round shapes highlight different factor assemblies and regulatory components as well as their interactions linked by arrows. Rounded rectangular shapes highlight intranuclear/intracellular compartments. TAD, topology-associated domain.

100–200 kb in size, and their borders are bound by the CCCTC-binding factor (CTCF) [94]. TADs can contain both active and inactive genes displaying active and inactive chromatin features, whereby the cis-regulatory elements of active genes interact with each other inside, but not outside the TAD boundaries, forming distinct subcompartments. The presence of CTCF is essential for forming the TAD structure, and the presence of the boundary is important for the insulation of genes from neighboring TADs. However, while CTCF depletion abolishes TAD boundaries and insulation, the organization into active and inactive genes and their interactions are largely unaffected [95], which is in line with the observation that TAD structures are not tissue-specific and gene expression patterns are programmed by transcriptional and epigenetic regulators. The vast majority of all TF and cofactor interactions within gene regulatory elements take place within the TAD boundaries with both TFs and cofactors participating in mediating these contacts [96] bringing together large regions of DNA that are highly tissue-specific [97]. Such extruded DNA loops can be encircled and thus stabilized by the structural maintenance of chromosome (SMC) complex, which contains cohesin and condensin and uses ATP to reel in DNA [98]. It is likely that once formed, such structures are required for genes to be able to respond to outside signals with a burst of transcription without having to build up the entire 3D structure from scratch.

In the last few years, another feature within the nucleus has caught attention—that of nuclear speckles or membrane-less organelles (MLOs). Such structures can be formed by liquid–liquid phase separation (LLPS), which in biological systems is essentially a process that is based on an interaction of molecules that excludes water [99]. The nucleus contains a multitude of such structures [100]. The best known are nuclear speckles, which are the sites of splicing, and the nucleolus, which is the site of rRNA synthesis that originates from multiple repeats of rDNA genes. RNA itself is sufficient to nucleate the formation of a nucleolus, which is faithfully reformed after cell division [101]. Transcriptionally inactive heterochromatin containing the heterochromatin protein HP1 consists of another nuclear compartment at the nuclear periphery, which protects the genome from mechanical stress [102]. Proteins, such as HP1, are capable to form condensates by themselves [103] and a tell-tale sign of their ability to do so are domains of intrinsically disordered regions that appear to be devoid of structure but are essential for phase separation [104]. A large number of TFs, including those important for hematopoietic differentiation processes [105,106], contain such regions and are able to form large assemblies without having to be too selective and sprout-specific domains for every possible interaction [107]. Under physiological salt condition, unmodified chromatin undergoes phase separation *in vitro* or when injected

into the nucleus, and this feature is modified by histone acetylation and protein binding [108]. A large number of factors contribute to a specific speckle type indicating that such structures play a global role in organizing nuclear processes [109]. Transcription is no exception. Microscopic analysis had shown many years ago that RNA polymerase II is organized in foci within the nucleus and appears to occur at fixed sites called 'transcription factories' [110]. More recently, the partition of transcriptional processes into separate assemblies was revived with the advent of global chromatin immunoprecipitation assays that uncovered that genes with complex regulatory regions (also termed 'superenhancers') form large, DNA-dependent molecular assemblies. It was suggested that these assemblies are able to undergo phase transition, thus forming regulatory entities with their own rules [111,112]. Moreover, it was also suggested that RNA polymerase II can shuttle between a transcription and a splicing compartment depending on its phosphorylation status [113]. However, while it is clear that such protein–DNA assemblies containing TFs and their cofactors form condensates *in vitro* and speckles *in vivo*, there is still some controversy whether DNA-dependent factor assembly represents true LLPS in living cells [114,115]. Nevertheless, it is now clear that compartmentalization is an essential part of regulatory processes within the nucleus, which drives the behavior of proteins in terms of their assembly kinetics and activity of enzymes. The challenge in the next years will be to precisely define the role of each compartment and which factors are involved in deciding how genes choose where to go to and are involved in driving compartmentalization.

The malignant state—differentiation going sideways

It is now clear that all of the mechanisms described in this review so far are important for normal development. Decades of research using knockout mice have shown that the machinery regulating differential gene expression is highly robust with a high inbuilt level of redundancy. However, they also showed that defects do not always manifest themselves immediately but can appear later in the life of an organism in the form of cancer, which is exemplified by certain types of blood cancers, occurring in families with inherited mutations in TF genes [116] that predispose patients to acute myeloid leukemia (AML). However, most cancer-causing mutations occur as somatic mutations in early hematopoietic precursor and stem cells. Recurrent mutations are seen in genes controlling gene regulation and epigenetic processes impacting cell fate

decisions [117]. This involves genes encoding TFs (i.e., RUNX1 or C/EBPα), chromatin remodelers and modifiers (i.e., CHD4, CBP), polycomb family members (i.e., EZH2), DNA methyltransferases (i.e., DNMT3A) but also demethylases such as TET1/2. Moreover, we also find mutations in genes encoding signaling molecules controlling gene expression driving growth such as RAS, genes encoding architectural proteins such as CTCF, and genes encoding splicing factors. AML is mostly a disease of the elderly, with mutations in genes encoding transcriptional and epigenetic regulators occurring first, which are then followed by additional mutations in growth-promoting genes [118,119]. Such successive acquisition of mutations first generates progenitor cells with slightly impaired differentiation capacity, which manifests itself as clonal hematopoiesis where the normally tightly regulated balance of differentiation is disturbed. One particular progenitor clone expands and contributes excessively to blood cell development without causing any overt disease phenotype. However, the seed is then laid for secondary mutations, which then lead to a complete impediment of differentiation and excessive malignant growth. It is now clear that different driver mutations have a different impact on the differentiation trajectory and the epigenetic landscape. As a result of a defect in an important regulator of cell fate driving a normal developmental trajectory, malignant cells adopt new identities distinct from normal cells and differentiation goes 'sideways' [120,121] (Fig. 3). The question now arises, what is the nature of these new cellular identities and how are they maintained as compared to normal cells. Normal cellular differentiation processes have been shaped and perfected by evolution over millions of years, whereas malignant cell differentiation is a product of patient-specific clonal selection that occurs in a much smaller time frame: Being 'imperfect', the question arises of why are malignant cells so difficult to eradicate?

The answer to this question lays in the robustness and plasticity of the differentiation process, that is, life itself. Similar to normal cells, malignant cells are maintained by distinct GRNs that drive common and AML subtype-specific signaling and metabolic pathways. It should be noted that while cancers come in different forms and can arise from many tissues, the rewiring of normal GRNs into one that sustains a malignant phenotype is a hallmark of all of them. In AML, each mutation shapes the aberrant differentiation process in a different way, and even different mutations in a single TF-encoding gene such as *RUNX1,* which give rise to different aberrant version of RUNX1, can lead to completely different disease outcomes and cellular
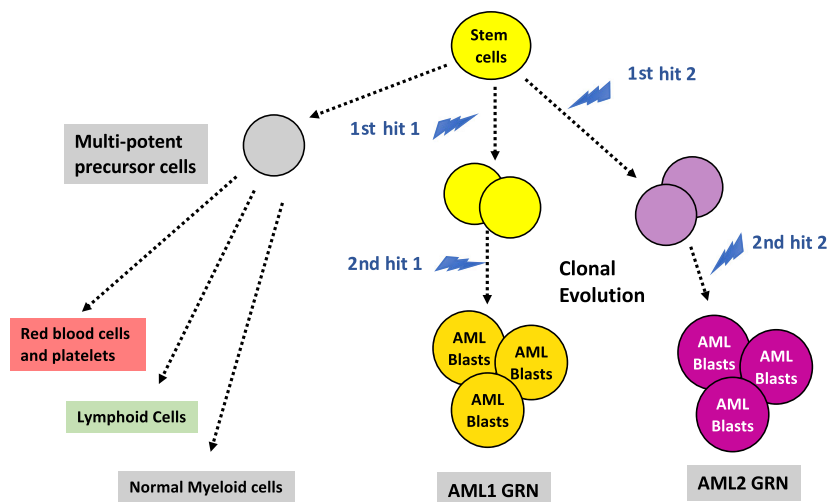
Fig. 3. Different types of mutations shape the development of alternate gene regulatory networks (GRNs) and drive subclonal evolution in acute myeloid leukemia. The figure depicts the alternate differentiation pathway of normal and aberrant blood cell development, starting from normal hematopoietic stem cells, which are capable of forming all blood cell types (left panel). In AML (right panel), the first mutational hit creates a cell that may still be capable of some level of differentiation, but differentiation is impaired due to the skewing of their GRNs. The nature of this impediment is different depending on the type of mutation. Once a second hit occurs, the GRNs of these cells are skewed further during clonal evolution, differentiation is further impaired, and proliferating AML cells develop a new identity with GRNs that are distinct from that of normal cells.

identities with distinct chromatin landscapes [122,123]. Moreover, the inducible expression of different RUNX1 oncoproteins causes an immediate reprogramming of their chromatin and TF-binding landscape, which is specific for each aberrant protein [123,124]. These data suggest that once different epigenetic landscapes have been set up after the first oncogenic hit, cells on their way to malignancy tweak their GRNs to compensate for the weakness of one differentiation process to activate another to maintain a stable state that is compatible with growth. We find the aberrant activation of genes encoding lineage-inappropriate TFs, which then become essential part of the network of abnormal but not normal cells [121,125,126]. We also find compensatory mechanisms whereby the mutation of one allele encoding a TF leads to shift in the GRN so that it now is dependent on the function of the wild-type allele [127]. Compensatory mechanisms and rewiring of signaling pathways are also common and tend to appear during therapy with the development of different subclonal populations. Examples for this phenomenon are the eradication of cells carrying a mutant FLT3 growth factor receptor after FLT3 inhibitor therapy and the appearance of RAS mutant cells either from preleukemic cells carrying the original driver mutation or from mutated leukemic cell escaping therapy [118,128]. A glimmer of hope comes from studies that profiled the chromatin landscape and gene expression of prospectively isolated subclonal

population pairs from different patients carrying different founder mutations. Each subclonal population displayed a different chromatin accessibility pattern indicating that the acquisition of additional genetic changes led to the formation of different chromatin landscapes [129]. However, when different subclonal pairs from different patients were compared, the chromatin accessibility patterns of each pair still clustered in a patient-specific way, demonstrating that epigenetic landscapes cannot drift apart in a disorderly way, that is, the cells have still much in common. Identifying the nature of these commonalities together with the differences will be crucial for the identification of patient-specific therapies.

Perspectives

In this review, we have only been able to show a glimpse of the complexity of the gene regulatory mechanisms that are encoded in our genome and that drive cell differentiation and we face significant challenges in our understanding of the molecular basis of developmental processes. We have deliberately left out the RNA world, and we have not mentioned how protein–protein interactions and metabolic processes impact on genome function and many other regulatory processes, many of which also play part in multiple pathologies. For our understanding of cancer as described above, it becomes clear that (a) each type of cancer has to be

seen as a different entity with an entirely unique underlying biology, (b) that we need to understand this biology if we want to get away from therapeutic approaches that target unregulated growth only (chemotherapy) which in itself is genotoxic, and (c) that we need to start thinking how we can reprogram GRNs without touching normal cells. Note, that these statements are valid for a number of pathological processes. We need to directly target the gene regulatory machinery in a disease-specific way, and we need to block the compensatory escape routes that are used by cancer cells, be it the rewiring of signaling pathways or increasing genomic instability thus jumbling GRNs and speeding up evolution. With the development of drugs targeting TFs such as MYC [130], RUNX [131,132]), chromatin regulators (BET [133,134], MLL [135]), and repair mechanisms (ATMi [135–137], PARPi [138]), we are starting to develop the right tool box. However, what is clear is that neither our normal environment nor pathological processes can be understood without knowing the rules of gene regulation and cellular biology and the players dictating these rules. This review is a passionate appeal to keep studying how life operates in all its amazing complexity.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author contributions

BE-W and CB both contributed to writing the paper.

References

1 Elsaid R, Soares-da-Silva F, Peixoto M, Amiri D, Mackowski N, Pereira P, Bandeira A & Cumano A (2020) Hematopoiesis: a layered organization across chordate species. *Front Cell Dev Biol* **8**, 606642.

2 Ivanovs A, Rybtsov S, Ng ES, Stanley EG, Elefanty AG & Medvinsky A (2017) Human haematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323–2337.

3 Palis J, McGrath KE & Kingsley PD (1995) Initiation of hematopoiesis and vasculogenesis in murine yolk sac explants. *Blood* **86**, 156–163.

4 Moore MA & Metcalf D (1970) Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. *Br J Haematol* **18**, 279–296.

5 Medvinsky A & Dzierzak E (1996) Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897–906.

6 Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, Orkin SH & Costantini F (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257–260.

7 Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D *et al.* (2017) Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**, 271–281.

8 Liggett LA & Sankaran VG (2020) Unraveling hematopoiesis through the lens of genomics. *Cell* **182**, 1384–1400.

9 Ton MN, Guibentif C & Gottgens B (2020) Single cell genomics and developmental biology: moving beyond the generation of cell type catalogues. *Curr Opin Genet Dev* **64**, 66–71.

10 Bonifer C, Levantini E, Kouskoff V & Lacaud G (2017) Runx1 structure and function in blood cell development. In RUNX Proteins in Development and Cancer (Groner Y, Ito Y, Liu P, Neil JC, Speck NA & van Wijnen A, eds), pp. 65–81.Springer Singapore, Singapore.

11 Growney JD, Shigematsu H, Li Z, Lee BH, Adelsperger J, Rowan R, Curley DP, Kutok JL, Akashi K, Williams IR *et al.* (2005) Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood* **106**, 494–504.

12 Mikkola HK, Klintman J, Yang H, Hock H, Schlaeger TM, Fujiwara Y & Orkin SH (2003) Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature* **421**, 547–551.

13 Kassouf MT, Hughes JR, Taylor S, McGowan SJ, Soneji S, Green AL, Vyas P & Porcher C (2010) Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. *Genome Res* **20**, 1064–1083.

14 Marin M, Karis A, Visser P, Grosveld F & Philipsen S (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* **89**, 619–628.

15 Gilmour J, O'Connor L, Middleton CP, Keane P, Gillemans N, Cazier JB, Philipsen S & Bonifer C (2019) Robust hematopoietic specification requires the ubiquitous Sp1 and Sp3 transcription factors. *Epigenetics Chromatin* **12**, 33.

16 Gilmour J, Assi SA, Jaegle U, Kulu D, van de Werken H, Clarke D, Westhead DR, Philipsen S & Bonifer C (2014) A crucial role for the ubiquitously expressed transcription factor Sp1 at early stages of hematopoietic specification. *Development* **141**, 2391–2401.

17 Johnson KD, Conn DJ, Shishkova E, Katsumura KR, Liu P, Shen S, Ranheim EA, Kraus SG, Wang W, Calvo KR *et al.* (2020) Constructing and deconstructing GATA2-regulated cell fate programs to establish developmental trajectories. *J Exp Med* **217**, e20191526.

18 Ostergaard P, Simpson MA, Connell FC, Steward CG, Brice G, Woollard WJ, Dafou D, Kilo T, Smithson S, Lunt P *et al.* (2011) Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* **43**, 929–931.

19 Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y, Akashi K, Fiering S & Tenen DG (2004) Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet* **36**, 624–630.

20 Lie ALM, Marinopoulou E, Lilly AJ, Challinor M, Patel R, Lancrin C, Kouskoff V & Lacaud G (2018) Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development* **145**, dev149419.

21 Lie ALM, Mevel R, Patel R, Blyth K, Baena E, Kouskoff V & Lacaud G (2020) RUNX1 dosage in development and cancer. *Mol Cells* **43**, 126–138.

22 Beck D, Thoms JA, Perera D, Schutte J, Unnikrishnan A, Knezevic K, Kinston SJ, Wilson NK, O'Brien TA, Gottgens B *et al.* (2013) Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood* **122**, e12–e22.

23 Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P, Chilarska PM, Kinston S, Ouwehand WH, Dzierzak E *et al.* (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544.

24 Khoueiry P, Girardot C, Ciglar L, Peng PC, Gustafson EH, Sinha S & Furlong EE (2017) Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. *Elife* **6**, e28440.

25 Obier N, Cauchy P, Assi SA, Gilmour J, Lie ALM, Lichtinger M, Hoogenkamp M, Noailles L, Cockerill PN, Lacaud G *et al.* (2016) Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. *Development* **143**, 4324–4340.

26 Kim WY, Sieweke M, Ogawa E, Wee HJ, Englmeier U, Graf T & Ito Y (1999) Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. *EMBO J* **18**, 1609–1620.

27 Wadman IA, Osada H, Grutz GG, Agulnick AD, Westphal H, Forster A & Rabbitts TH (1997) The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J* **16**, 3145–3157.

28 Zhou Q, Liu M, Xia X, Gong T, Feng J, Liu W, Liu Y, Zhen B, Wang Y, Ding C *et al.* (2017) A mouse tissue transcription factor atlas. *Nat Commun* **8**, 15089.

29 Pimanda JE, Ottersbach K, Knezevic K, Kinston S, Chan WY, Wilson NK, Landry JR, Wood AD, Kolb-Kokocinski A, Green AR *et al.* (2007) Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc Natl Acad Sci USA* **104**, 17692–17697.

30 Lichtinger M, Ingram R, Hannah R, Muller D, Clarke D, Assi SA, Lie ALM, Noailles L, Vijayabaskar MS, Wu M *et al.* (2012) RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J* **31**, 4318–4333.

31 Goode DK, Obier N, Vijayabaskar MS, Lie ALM, Lilly AJ, Hannah R, Lichtinger M, Batta K, Florkowska M, Patel R *et al.* (2016) Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev Cell* **36**, 572–587.

32 Suter DM, Molina N, Gatfield D, Schneider K, Schibler U & Naef F (2011) Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–474.

33 Bartman CR, Hamagami N, Keller CA, Giardine B, Hardison RC, Blobel GA & Raj A (2019) Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol Cell* **73**, 519–532 e4.

34 Shaulian E & Karin M (2002) AP-1 as a regulator of cell life and death. *Nat Cell Biol* **4**, E131–E136.

35 Lin JX & Leonard WJ (2019) Fine-tuning cytokine signals. *Annu Rev Immunol* **37**, 295–324.

36 Malhotra N & Kang J (2013) SMAD regulatory networks construct a balanced immune system. *Immunology* **139**, 1–10.

37 Blank U & Karlsson S (2011) The role of Smad signaling in hematopoiesis and translational hematology. *Leukemia* **25**, 1379–1388.

38 Wu Z & Guan KL (2020) Hippo signaling in embryogenesis and development. *Trends Biochem Sci* **46**, 51–63.

39 Clipstone NA & Crabtree GR (1992) Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357**, 695–697.

40 Yoshimi M, Goyama S, Kawazu M, Nakagawa M, Ichikawa M, Imai Y, Kumano K, Asai T, Mulloy JC, Kraft AS et al. (2012) Multiple phosphorylation sites are important for RUNX1 activity in early hematopoiesis and T-cell differentiation. Eur J Immunol 42, 1044–1050.

41 Lundin V, Sugden WW, Theodore LN, Sousa PM, Han A, Chou S, Wrighton PJ, Cox AG, Ingber DE, Goessling W et al. (2020) YAP regulates hematopoietic stem cell formation in response to the biomechanical forces of blood flow. Dev Cell 52, 446–460 e5.

42 Johnson BV, Bert AG, Ryan GR, Condina A & Cockerill PN (2004) Granulocyte-macrophage colony-stimulating factor enhancer activation requires cooperation between NFAT and AP-1 elements and is associated with extensive nucleosome reorganization. Mol Cell Biol 24, 7914–7930.

43 Brignall R, Cauchy P, Bevington SL, Gorman B, Pisco AO, Bagnall J, Boddington C, Rowe W, England H, Rich K et al. (2017) Integration of kinase and calcium signaling at the level of chromatin underlies inducible gene activation in T cells. J Immunol 199, 2652–2667.

44 Schoenherr CJ & Anderson DJ (1995) The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. Science 267, 1360–1363.

45 Nutt SL, Heavey B, Rolink AG & Busslinger M (1999) Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature 401, 556–562.

46 Tagoh H, Ingram R, Wilson N, Salvagiotto G, Warren AJ, Clarke D, Busslinger M & Bonifer C (2006) The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction. EMBO J 25, 1070–1080.

47 Castano J, Aranda S, Bueno C, Calero-Nieto FJ, Mejia-Ramirez E, Mosquera JL, Blanco E, Wang X, Prieto C, Zabaleta L et al. (2019) GATA2 promotes hematopoietic development and represses cardiac differentiation of human mesoderm. Stem Cell Reports 13, 515–529.

48 Stopka T, Amanatullah DF, Papetti M & Skoultchi AI (2005) PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. EMBO J 24, 3712–3723.

49 Palii CG, Cheng Q, Gillespie MA, Shannon P, Mazurczyk M, Napolitani G, Price ND, Ranish JA, Morrissey E, Higgs DR et al. (2019) Single-cell proteomics reveal that quantitative changes in co-expressed lineage-specific transcription factors determine cell fate. Cell Stem Cell 24, 812–820 e5.

50 Mylona A, Theillet FX, Foster C, Cheng TM, Miralles F, Bates PA, Selenko P & Treisman R (2016) Opposing effects of Elk-1 multisite phosphorylation

shape its response to ERK activation. Science 354, 233–237.

51 Gottgens B (2015) Regulatory network control of blood stem cells. Blood 125, 2614–2620.

52 Rothenberg EV & Gottgens B (2021) How haematopoiesis research became a fertile ground for regulatory network biology as pioneered by Eric Davidson. Curr Opin Hematol 28, 1–10.

53 Kucinski I, Wilson NK, Hannah R, Kinston SJ, Cauchy P, Lenaerts A, Grosschedl R & Gottgens B (2020) Interactions between lineage-associated transcription factors govern haematopoietic progenitor states. EMBO J 39, e104983.

54 Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS et al. (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods 6, 283–289.

55 Gilmour J, Assi SA, Noailles L, Lichtinger M, Obier N & Bonifer C (2018) The co-operation of RUNX1 with LDB1, CDK9 and BRD4 drives transcription factor complex relocation during haematopoietic specification. Sci Rep 8, 10410.

56 Davidson EH (2010) Emerging properties of animal gene regulatory networks. Nature 468, 911–920.

57 Agarwal V & Shendure J (2020) Predicting mRNA abundance directly from genomic sequence using deep convolutional neural networks. Cell Rep 31, 107663.

58 Zaret KS (2020) Pioneer transcription factors initiating gene network changes. Annu Rev Genet 54, 367–385.

59 Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, Dodonova SO, Nitta KR, Morgunova E, Taipale M et al. (2018) The interaction landscape between transcription factors and the nucleosome. Nature 562, 76–81.

60 Kassabov SR, Zhang B, Persinger J & Bartholomew B (2003) SWI/SNF unwraps, slides, and rewraps the nucleosome. Mol Cell 11, 391–403.

61 Voss TC, Schiltz RL, Sung MH, Yen PM, Stamatoyannopoulos JA, Biddie SC, Johnson TA, Miranda TB, John S & Hager GL (2011) Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. Cell 146, 544–554.

62 Goldstein I, Paakinaho V, Baek S, Sung MH & Hager GL (2017) Synergistic gene expression during the acute phase response is characterized by transcription factor assisted loading. Nat Commun 8, 1849.

63 Cockerill PN (2011) Structure and function of active chromatin and DNase I hypersensitive sites. FEBS J 278, 2182–2210.

64 Xu B, Cai L, Butler JM, Chen D, Lu X, Allison DF, Lu R, Rafii S, Parker JS, Zheng D et al. (2018) The chromatin remodeler BPTF activates a stemness gene-expression program essential for the maintenance of

adult hematopoietic stem cells. *Stem Cell Reports* **10**, 675–683.

65 Hsu J, Huang HT, Lee CT, Choudhuri A, Wilson NK, Abraham BJ, Moignard V, Kucinski I, Yu S, Hyde RK *et al.* (2020) CHD7 and Runx1 interaction provides a braking mechanism for hematopoietic differentiation. *Proc Natl Acad Sci USA* **117**, 23626–23635.

66 Tu J, Liu X, Jia H, Reilly J, Yu S, Cai C, Liu F, Lv Y, Huang Y, Lu Z *et al.* (2020) The chromatin remodeler Brg1 is required for formation and maintenance of hematopoietic stem cells. *FASEB J* **34**, 11997–12008.

67 Dekker J, Rippe K, Dekker M & Kleckner N (2002) Capturing chromosome conformation. *Science* **295**, 1306–1311.

68 de Laat W & Grosveld F (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* **11**, 447–459.

69 Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA & Young RA (2013) Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947.

70 Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI & Young RA (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319.

71 Bonifer C, Jagle U & Huber MC (1997) The chicken lysozyme locus as a paradigm for the complex developmental regulation of eukaryotic gene loci. *J Biol Chem* **272**, 26075–26078.

72 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H & Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576–589.

73 Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ *et al.* (2016) Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193–1203.

74 Fuda NJ, Ardehali MB & Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186–192.

75 Behera V, Stonestrom AJ, Hamagami N, Hsiung CC, Keller CA, Giardine B, Sidoli S, Yuan ZF, Bhanu NV, Werner MT *et al.* (2019) Interrogating histone acetylation and BRD4 as mitotic bookmarks of transcription. *Cell Rep* **27**, 400–415 e5.

76 Palozola KC, Lerner J & Zaret KS (2019) A changing paradigm of transcriptional memory propagation through mitosis. *Nat Rev Mol Cell Biol* **20**, 55–64.

77 Bevington SL, Keane P, Soley JK, Tauch S, Gajdasik DW, Fiancette R, Matei-Rascu V, Willis CM, Withers DR & Cockerill PN (2020) IL-2/IL-7-inducible factors pioneer the path to T cell differentiation in advance of lineage-defining factors. *EMBO J* **39**, e105220.

78 Kamada R, Yang W, Zhang Y, Patel MC, Yang Y, Ouda R, Dey A, Wakabayashi Y, Sakaguchi K, Fujita T *et al.* (2018) Interferon stimulation creates chromatin marks and establishes transcriptional memory. *Proc Natl Acad Sci USA* **115**, E9162–E9171.

79 Bonifer C & Cockerill PN (2017) Chromatin priming of genes in development: concepts, mechanisms and consequences. *Exp Hematol* **49**, 1–8.

80 Cochella L & Hobert O (2012) Embryonic priming of a miRNA locus predetermines postmitotic neuronal left/right asymmetry in *C. elegans*. *Cell* **151**, 1229–1242.

81 Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, Mulas C, Ibarra-Soria X, Tyser RCV, Ho DLL *et al.* (2019) A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495.

82 Digre A & Lindskog C (2020) The Human Protein Atlas- spatial localization of the human proteome in health and disease. *Protein Sci* **30**, 218–233.

83 Szabo Q, Donjon A, Jerkovic I, Papadopoulos GL, Cheutin T, Bonev B, Nora EP, Bruneau BG, Bantignies F & Cavalli G (2020) Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet* **52**, 1151–1157.

84 Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293.

85 Hildebrand EM & Dekker J (2020) Mechanisms and functions of chromosome compartmentalization. *Trends Biochem Sci* **45**, 385–396.

86 Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS & Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043.

87 de Napoles M, Mermoud JE, Wakao R, Tang YA, Endoh M, Appanah R, Nesterova TB, Silva J, Otte AP, Vidal M *et al.* (2004) Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* **7**, 663–676.

88 Denholtz M, Bonora G, Chronis C, Splinter E, de Laat W, Ernst J, Pellegrini M & Plath K (2013) Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602–616.

89 Schoenfelder S, Sugar R, Dimond A, Javierre BM, Armstrong H, Mifsud B, Dimitrova E, Matheson L,

Tavares-Cadete F, Furlan-Magaril M *et al.* (2015) Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet* **47**, 1179–1186.

90 Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu CT & Zhuang X (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422.

91 Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951.

92 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS & Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380.

93 Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385.

94 Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295.

95 Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA & Bruneau BG (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944, e22.

96 Kim S & Shendure J (2019) Mechanisms of interplay between transcription factors and the 3D genome. *Mol Cell* **76**, 306–319.

97 Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ *et al.* (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 e19.

98 Ganji M, Shaltiel IA, Bisht S, Kim E, Kalichava A, Haering CH & Dekker C (2018) Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105.

99 Shin Y & Brangwynne CP (2017) Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382.

100 Strom AR & Brangwynne CP (2019) The liquid nucleome – phase transitions in the nucleus at a glance. *J Cell Sci* **132**, jcs235093.

101 Shevtsov SP & Dundr M (2011) Nucleation of nuclear bodies by RNA. *Nat Cell Biol* **13**, 167–173.

102 Nava MM, Miroshnikova YA, Biggs LC, Whitefield DB, Metge F, Boucas J, Vihinen H, Jokitalo E, Li X, Garcia Arcos JM *et al.* (2020) Heterochromatin-driven nuclear softening protects the genome against mechanical stress-induced damage. *Cell* **181**, 800–817 e22.

103 Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S & Narlikar GJ (2017) Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240.

104 Uversky VN (2017) Intrinsically disordered proteins in overcrowded milieu: membrane-less organelles, phase separation, and intrinsic disorder. *Curr Opin Struct Biol* **44**, 18–30.

105 Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM *et al.* (2018) Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 e16.

106 Wang Y, Zolotarev N, Yang CY, Rambold A, Mittler G & Grosschedl R (2020) A prion-like domain in transcription factor EBF1 promotes phase separation and enables B cell programming of progenitor chromatin. *Immunity* **53**, 1151–1167 e6.

107 Staby L, O'Shea C, Willemoes M, Theisen F, Kragelund BB & Skriver K (2017) Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem J* **474**, 2509–2532.

108 Gibson BA, Doolittle LK, Schneider MWG, Jensen LE, Gamarra N, Henry L, Gerlich DW, Redding S & Rosen MK (2019) Organization of chromatin by intrinsic and regulated phase separation. *Cell* **179**, 470–484 e21.

109 Berchtold D, Battich N & Pelkmans L (2018) A systems-level study reveals regulators of membrane-less organelles in human cells. *Mol Cell* **72**, 1035–1049 e5.

110 Jackson DA, Hassan AB, Errington RJ & Cook PR (1993) Visualization of focal sites of transcription within human nuclei. *EMBO J* **12**, 1059–1065.

111 Hnisz D, Shrinivas K, Young RA, Chakraborty AK & Sharp PA (2017) A phase separation model for transcriptional control. *Cell* **169**, 13–23.

112 Shrinivas K, Sabari BR, Coffey EL, Klein IA, Boija A, Zamudio AV, Schuijers J, Hannett NM, Sharp PA, Young RA *et al.* (2019) Enhancer features that drive formation of transcriptional condensates. *Mol Cell* **75**, 549–561 e7.

113 Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, Spille JH, Afeyan LK, Zamudio AV, Shrinivas K *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548.

114 McSwiggen DT, Hansen AS, Teves SS, Marie-Nelly H, Hao Y, Heckert AB, Umemoto KK, Dugast-

Darzacq C, Tjian R & Darzacq X (2019) Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *Elife* **8**, e47098.

115 McSwiggen DT, Mir M, Darzacq X & Tjian R (2019) Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev* **33**, 1619–1634.

116 Churpek JE & Bresnick EH (2019) Transcription factor mutations as a cause of familial myeloid neoplasms. *J Clin Invest* **129**, 476–488.

117 Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ Jr, Laird PW *et al.* (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059–2074.

118 Miles LA, Bowman RL, Merlinsky TR, Csete IS, Ooi AT, Durruthy-Durruthy R, Bowman M, Famulare C, Patel MA, Mendez P *et al.* (2020) Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature* **587**, 477–482.

119 Morita K, Wang F, Jahn K, Hu T, Tanaka T, Sasaki Y, Kuipers J, Loghavi S, Wang SA, Yan Y *et al.* (2020) Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun* **11**, 5327.

120 Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, van Waalwijk B, van Doorn-Khosrovani S, Boer JM, Beverloo HB, Moorhouse MJ, van der Spek PJ *et al.* (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* **350**, 1617–1628.

121 Assi SA, Imperato MR, Coleman DJL, Pickin A, Potluri S, Ptasinska A, Chin PS, Blair H, Cauchy P, James SR *et al.* (2019) Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet* **51**, 151–162.

122 Matheny CJ, Speck ME, Cushing PR, Zhou Y, Corpora T, Regan M, Newman M, Roudaia L, Speck CL, Gu TL *et al.* (2007) Disease mutations in RUNX1 and RUNX2 create nonfunctional, dominant-negative, or hypomorphic alleles. *EMBO J* **26**, 1163–1175.

123 Kellaway SG, Keane P, Edginton-White B, Regha K, Kennett E & Bonifer C (2021) Different mutant RUNX1 oncoproteins program alternate haematopoietic differentiation trajectories. *Life Sci Alliance* **4**, e202000864. ePub ahead of print.

124 Nafria M, Keane P, Ng ES, Stanley EG, Elefanty AG & Bonifer C (2020) Expression of RUNX1-ETO rapidly alters the chromatin landscape and growth of early human myeloid precursor cells. *Cell Rep* **31**, 107691.

125 Ray D, Kwon SY, Tagoh H, Heidenreich O, Ptasinska A & Bonifer C (2013) Lineage-inappropriate PAX5 expression in t(8;21) acute myeloid leukemia requires signaling-mediated abrogation of polycomb repression. *Blood* **122**, 759–769.

126 Somerville TD, Wiseman DH, Spencer GJ, Huang X, Lynch JT, Leong HS, Williams EL, Cheesman E & Somervaille TC (2015) Frequent derepression of the mesenchymal transcription factor gene FOXC1 in acute myeloid leukemia. *Cancer Cell* **28**, 329–342.

127 Ben-Ami O, Friedman D, Leshkowitz D, Goldenberg D, Orlovsky K, Pencovich N, Lotem J, Tanay A & Groner Y (2013) Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Rep* **4**, 1131–1143.

128 McMahon CM, Ferng T, Canaani J, Wang ES, Morrissette JJD, Eastburn DJ, Pellegrino M, Durruthy-Durruthy R, Watt CD, Asthana S *et al.* (2019) Clonal selection with RAS pathway activation mediates secondary clinical resistance to selective FLT3 inhibition in acute myeloid leukemia. *Cancer Discov* **9**, 1050–1063.

129 de Boer B, Prick J, Pruis MG, Keane P, Imperato MR, Jaques J, Brouwers-Vos AZ, Hogeling SM, Woolthuis CM, Nijk MT *et al.* (2018) Prospective isolation and characterization of genetically and functionally distinct AML subclones. *Cancer Cell* **34**, 674–689 e8.

130 Soucek L, Whitfield J, Martins CP, Finch AJ, Murphy DJ, Sodir NM, Karnezis AN, Swigart LB, Nasi S & Evan GI (2008) Modelling Myc inhibition as a cancer therapy. *Nature* **455**, 679–683.

131 Illendula A, Pulikkan JA, Zong H, Grembecka J, Xue L, Sen S, Zhou Y, Boulton A, Kuntimaddi A, Gao Y *et al.* (2015) Chemical biology. A small-molecule inhibitor of the aberrant transcription factor CBFbeta-SMMHC delays leukemia in mice. *Science* **347**, 779–784.

132 Cunningham L, Finckbeiner S, Hyde RK, Southall N, Marugan J, Yedavalli VR, Dehdashti SJ, Reinhold WC, Alemu L, Zhao L *et al.* (2012) Identification of benzodiazepine Ro5-3335 as an inhibitor of CBF leukemia through quantitative high throughput screen against RUNX1-CBFbeta interaction. *Proc Natl Acad Sci USA* **109**, 14592–14597.

133 Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T, Hickman TT, Felletar I *et al.* (2010) Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073.

134 Krivtsov AV, Evans K, Gadrey JY, Eschle BK, Hatton C, Uckelmann HJ, Ross KN, Perner F, Olsen SN, Pritchard T *et al.* (2019) A Menin-MLL inhibitor induces specific chromatin changes and eradicates disease in models of MLL-rearranged leukemia. *Cancer Cell* **36**, 660–673 e11.

135 Hickson I, Zhao Y, Richardson CJ, Green SJ, Martin NM, Orr AI, Reaper PM, Jackson SP, Curtin NJ & Smith GC (2004) Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res* **64**, 9152–9159.

136 Pike KG, Barlaam B, Cadogan E, Campbell A, Chen Y, Colclough N, Davies NL, de-Almeida C, Degorce SL, Didelot M et al. (2018) The identification of potent, selective, and orally available inhibitors of ataxia telangiectasia mutated (ATM) kinase: the discovery of AZD0156 (8-{6-[3-(Dimethylamino) propoxy]pyridin-3-yl}-3-methyl-1-(tetrahydro-2 H-pyran-4-yl)-1,3-dihydro-2 H-imidazo[4,5- c]quinolin-2-one). *J Med Chem* **61**, 3823–3841.

137 Karlin J, Allen J, Ahmad SF, Hughes G, Sheridan V, Odedra R, Farrington P, Cadogan EB, Riches LC, Garcia-Trinidad A et al. (2018) Orally bioavailable and blood-brain barrier-penetrating ATM inhibitor (AZ32) radiosensitizes intracranial gliomas in mice. *Mol Cancer Ther* **17**, 1637–1647.

138 Murai J, Huang SY, Das BB, Renaud A, Zhang Y, Doroshow JH, Ji J, Takeda S & Pommier Y (2012) Trapping of PARP1 and PARP2 by clinical PARP inhibitors. *Cancer Res* **72**, 5588–5599.