The role of noncoding mutations in blood cancers

Sunniyat Rahman and Marc R. Mansour*

ABSTRACT

The search for oncogenic mutations in haematological malignancies has largely focused on coding sequence variants. These variants have been critical in understanding these complex cancers in greater detail, ultimately leading to better disease monitoring, subtyping and prognostication. In contrast, the search for oncogenic variants in the noncoding genome has proven to be challenging given the vastness of the search space, the intrinsic difficulty in assessing the impact of variants that do not code for functional proteins, and our still primitive understanding of the function harboured by large parts of the noncoding genome. Recent studies have broken ground on this quest, identifying somatically acquired and recurrent mutations in the noncoding genome that activate the expression of proto-oncogenes. In this Review, we explore some of the best-characterised examples of noncoding mutations in haematological malignancies, and highlight how a significant majority of these variants impinge on gene regulation through the formation of aberrant enhancers and promoters. We delve into the challenges faced by those that embark on a search for noncoding driver mutations, and provide a framework distilled from studies that have successfully identified such variants to overcome some of the most salient hurdles. Finally, we discuss the current therapeutic strategies being explored to target the oncogenic mechanism supported by recurrent noncoding variants. We postulate that the continued discovery and functional characterisation of somatic variants in the noncoding genome will not only advance our understanding of haematological malignancies, but offer novel therapeutic avenues and provide important insights into transcriptional regulation on a broader scale.

KEY WORDS: Enhancers and promoters, Gene regulation, Haematological malignancy, Noncoding genome

Introduction

Haematological malignancies cover a broad spectrum of cancers of bone marrow, lymphatic or thymic origin. These malignancies arise through the acquisition of genetic aberrations that drive proliferation, inhibit differentiation, enable cellular survival and evade immune surveillance. Haematological malignancies are the fourth most common cause of cancer mortality, surpassed only by lung, colorectal and breast cancer (www.cancerresearchuk.org). Huge strides have been made in treatment, with notable improvements in acute lymphoblastic leukaemia (ALL), which now has a survival rate of more than 90% for those aged 14 or younger (www.cancerresearchuk.org). Unfortunately, 5-year

Department of Haematology, UCL Cancer Institute, University College London, London WC1E 6DD, UK.

*Author for correspondence (m.mansour@ucl.ac.uk)

S.R., 0000-0001-5374-7376; M.R.M., 0000-0002-7449-845X

survival in adult acute myeloid leukaemia (AML) and ALL are still less than 50%, with significant treatment challenges including treatment-resistant disease, clonal heterogeneity of the underlying disease, treatment-associated toxicities and poor tolerance for intensive treatment regimens, particularly in older patients with comorbidities (Kansagra et al., 2018). Because of this, there is a significant impetus to discover new genetic aberrations, including driver mutations (see Box 1 for a glossary of terms), that may provide insight into the mechanisms of malignant haematopoiesis, as well as offering novel therapeutic opportunities.

Detailed genetic characterisation of haematological malignancies has already identified alterations that are now being used for better diagnosis, prognostication, subtype identification and to inform therapeutic decisions (Taylor et al., 2017). The vast majority of these genetic alterations have been identified by studies focused on the coding sequences, which represent just 2% of the human genome, leaving the noncoding genome largely unexplored (ENCODE Project Consortium, 2012). Here, we discuss examples of noncoding mutations that have been identified in haematological malignancies so far, and explore how these examples have shaped our understanding about what constitutes a functional or driver noncoding mutation. Furthermore, we describe the challenges in identifying noncoding mutations that are drivers, rather than passengers, within the trajectory of cellular transformation, and begin to outline a framework through which one can potentially address some of these challenges to identify novel noncoding mutations that have functional significance. Finally, we provide some insight into therapeutic strategies that are currently being explored to disrupt the oncogenic mechanisms that arise from noncoding oncogenic mutations.

Rationale for the identification and characterisation of mutations in the noncoding genome

There is a strong rationale for exploring the noncoding genome for biomarkers, therapeutic targets and somatically acquired driver mutations (Box 1). First, it has become clear that the noncoding genome itself is rich with *cis*-regulatory DNA elements such as promoters, enhancers and insulators (Box 1) (ENCODE Project Consortium, 2012). Whereas promoters bind the core transcriptional machinery, including RNA polymerase II, enhancers are able to bind regulatory proteins called transcription factors (TFs; Box 1) that can dramatically alter the activity of promoters, even from distal positions within the genome and without orientation constraints (Haberle and Stark, 2018). Ultimately, these regulatory mechanisms ensure that genes are expressed at an appropriate time and magnitude during the trajectory of cellular differentiation. This is especially true in haematopoiesis, where master TFs such as GATA1, SPI1, RUNX1 and MYB intimately regulate the expression of genes critical for the development of cells from erythroid, myeloid and lymphoid lineages (Jagannathan-Bogdan and Zon, 2013). Therefore, thorough interrogation of the noncoding genome is warranted in the context of haematological malignancies in which mutations disrupt nominal haematopoietic developmental programmes. Under these

Box 1. Glossary

APOBEC: 'Apolipoprotein B mRNA editing enzyme, catalytic peptide-like' are a family of catalytically active proteins that can insert mutations in both DNA and RNA through the deamination of cytidine to uridine, which is believed to be a significant endogenous mutagen in cancers.

ATAC-seq: 'Assay for transposase-accessible chromatin using sequencing' is a technique used to identify nucleosome-free regions of the genome. Such open regions of chromatin are susceptible to digestion by Tn5 transposase, which simultaneously digests and ligates adapter sequences for high throughput sequencing.

Chimeric fusions: A class of mutations that create a new protein by fusing two or more different coding sequences.

Chromatin immunoprecipitation (ChIP): An antibody-based experimental technique that identifies the location of protein binding events to DNA at nucleotide resolution. The immunoprecipitated DNA fragments can be sequenced (a technique called ChIP-seq) and mapped to the reference genome sequence.

Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET): A technique used to detect interactions between disparate DNA sequences via a protein of interest. This is achieved by combining chromatin immunoprecipitation of a protein of interest, proximity ligation of DNA fragments and high throughput sequencing.

Class switch recombination: Allows B cells to rearrange the constant region genes in the immunoglobulin heavy chain locus to switch expression from one class of immunoglobulin to another.

Cohesin: A ring-shaped protein complex that is required for sister chromatid cohesion and to make contacts between distal chromatin segments for gene regulation.

DNase hypersensitivity sequencing (DNase-seq): A technique that allows for the sequencing of accessible genomic regions as determined by sensitivity to cleavage by DNase. These sites may have regulatory potential and are often bound by transcription factors and coactivators.

Driver mutations: A class of mutations that confer a growth or survival advantage in cancerous cells and thus promote cancer development.

Enhancer: A regulatory sequence element that can alter the transcriptional output of a gene.

Hi-C: A technique used to capture the conformational organisation of genomes. Briefly, it involves chromatin fixation, digestion and ligation, which also introduces a biotin-labelled nucleotide at ligation junctions. Following biotin-selective purification, the DNA fragments can be subjected to high-throughput sequencing to map many-to-many chromatin contacts and interactions.

Indels: Small nucleotide insertions or deletions or a combination of both at specific genomic loci.

Insulator: A regulatory sequence that creates a boundary between enhancers and promoters to minimise the activity of these sequences on gene regulatory processes.

Kataegis: A form of localised hypermutation that is characterised by clusters of C>T and/or C>G mutations that are substantially enriched at TpCpN trinucleotides and on the same DNA strand.

Monoallelic expression: Where expression of a gene is identified as being from a single allele only, either maternal or paternal.

MonoMAC syndrome: 'Monocytopaenia and mycobacterial infection' syndrome arises from heterozygous mutations in *GATA2* leading to loss of function. This gene is critical for functional haematopoiesis and lymphatic formation, so loss of function leads to significantly reduced numbers of circulating monocytes, dendritic cells, natural killer and B cells, as well as an increased likelihood of opportunistic infections and haematological malignancies.

Promoter: A regulatory sequence element nearest to the transcriptional start site of a gene that is bound by the core transcriptional machinery, including RNA pol II, and capable of activating gene expression.

Transcription factor (TF): A protein that binds specific DNA sequences through a DNA binding domain, and that can activate or repress gene expression.

Transcription start site (TSS): The nucleotide position of transcriptional initiation, which usually corresponds to the 5′ cap of an mRNA transcript.

V(D)J recombination: An endogenous mutagenic process that facilitates the recombination of V, D and J gene segments of developing T and B cells that results in diverse T cell receptor and immunoglobulin repertoires, respectively.

circumstances, one can hypothesise that novel uncharacterised variants may reside in the noncoding genome, particularly in regions with regulatory potential.

Secondly, the continued advancement of next generation sequencing (NGS) technologies has allowed for better exploration of cancer genomes. There are now several detailed and integrated datasets of human malignancies available thanks to large-scale collaborative efforts. These include The Cancer Genome Atlas (TCGA), which has generated data on 27 cancer types (including AML) through exome sequencing, copy number variation analysis using single nucleotide polymorphism (SNP) arrays, DNA methylation and RNA sequencing (Cancer Genome Atlas Research Network et al., 2013a,b). Additional data has been collated and generated by the Catalogue of Somatic Mutations In Cancer (COSMIC) project, which has reported approximately six million coding mutations and explored other genetic mechanisms that can spearhead cancer progression, including gene fusions, drug resistance mutations and, more recently, noncoding mutations (Tate et al., 2019). In the context of haematological malignancies, focused studies on specific diseases, such as T cell ALL (T-ALL) and AML, have identified coding variants, chromosomal translocations, chimeric fusions (Box 1) and mutations, which have then been coupled with clinical outcome data for improved prognostication (Liu et al., 2017; Papaemmanuil et al., 2016; Zhang et al., 2012). These studies have been essential in distilling important details about the pathogenesis of these malignancies and the mechanisms

by which specific mutations disrupt normal cellular function, yet the vast majority of these studies have reported mutations only in the exome, leaving the noncoding genome relatively unexplored.

Thirdly, there are strong endogenous mutagenic processes required for the development of mature lymphocytes. This includes V(D)J recombination (Box 1) by RAG1/2 for antigen receptor diversity, AID (AICDA)-mediated class switch recombination (Box 1) and somatic hypermutation. Off-target activity of these cellular processes has been attributed to chromosomal translocations, such as translocation of the T cell receptor locus with its strong endogenous enhancers into close regulatory proximity to the oncogenes *LMO2*, *TAL1* and *TAL2* in T-ALL, and the AID-dependent *MYC/IGH* translocations in Burkitt's lymphoma (Marculescu et al., 2002; Robbiani et al., 2008). These endogenous mutagenic processes are a source of double-strand DNA breaks in developing lymphocytes, where off-target events are subjected to imperfect repair processes such as non-homologous end joining and homology directed repair (Helleday et al., 2014). Together, these processes can create lesions, including indels (Box 1), tandem duplications and translocations across the genome. Given RAG1/2 is allosterically activated upon binding to H3K4me3 (trimethylated lysine 4 of histone 3), a marker of active promoters, it is reasonable to postulate that genes that are co-expressed with RAG during cell development are at greater risk of off-target RAG endonuclease activity (Bettridge et al., 2017). There are also more generalised

mutagenic processes at work in cancer genomes. Systematic analysis of somatic mutations in cancer genomes, including those of ALLs and AMLs, demonstrate that many mutations conform to mutational signatures associated with ageing and APOBEC-like cytidine deaminase activity (Box 1) (Alexandrov et al., 2013). This study also identified localised hypermutation termed kataegis (Box 1) in chronic lymphocytic leukaemia (CLL), B cell lymphoma and ALL genomes. Furthermore, the noncoding genome is subjected to higher mutation rates compared to its coding counterpart, although these mutations are under relatively weak selection pressures unless the mutations themselves confer a survival advantage (Weinhold et al., 2014). Given the many mutagenic processes at play, we should aim to open the search space for driver mutations in haematological malignancies beyond the exome to intergenic, intronic and promoter regions, in which such mutagenic events also occur.

Finally, there is merit in searching for noncoding mutations. A landmark discovery in 2013 showed that the promoter of *TERT*, which encodes the reverse transcriptase subunit of telomerase, can somatically acquire mutations leading to its overexpression in human melanoma (Huang et al., 2013). This finding was important in demonstrating that the noncoding genome itself can acquire driver mutations. Further work confirmed that these *TERT* promoter mutations recur in other malignancies, meaning similar variants were selected for during the development of multiple neoplasms (Horn et al., 2013; Vinagre et al., 2013). The mutations at the *TERT* promoter are excellent examples of noncoding driver mutations, as these lesions generate *de novo* consensus binding sites for ETS family TFs leading to consequential overexpression of *TERT*, telomere length maintenance through cell divisions, and thus continued cellular survival. Similar mutational hotspots have also been identified in noncoding regions of multiple cancer genomes. Examples include recurrent *FOXA1* promoter mutations, a known driver of hormone-receptor positive breast cancer, recurrent mutations in *cis*-regulatory elements that interact with the *ETV1* promoter in colorectal cancer and affect patient survival, and recurrent noncoding mutations in liver cancer (Fujimoto et al., 2016; Orlando et al., 2018; Rheinbay et al., 2017). These studies help to further rationalise the importance in searching for driver mutations in the noncoding genome.

Architecture of a classical enhancer-promoter interaction

Many of the noncoding mutations identified thus far in haematological malignancies appear to impinge on the pre-transcriptional regulation of genes. A number of regulatory noncoding elements interact before transcriptional initiation at a gene locus (Fig. 1B). This includes the promoter, which is located <1 kb upstream from the transcriptional start site (TSS; Box 1) of a gene and may include nearby proximal regulatory elements, and distal enhancers (Vernimmen and Bickmore, 2015). Whereas promoters effectively recruit the core transcriptional machinery, such as elongation factors and RNA pol II, it is enhancers that control spatiotemporal gene expression and tissue-specific gene expression programmes, often over vast genomic distances (Schoenfelder and Fraser, 2019). Importantly, enhancers can activate distal promoters independently of location and orientation relative to the target gene. Furthermore, enhancers are enriched with histone marks such as H3K27ac (histone 3 lysine 27 acetylation) and H3K4me1 (histone 3 lysine 4 monomethylation), and are bound by TFs and coactivators (Lambert et al., 2018; Medina-Rivera et al., 2018). Transmission of enhancer activity to RNA pol II at a gene promoter is achieved by the large multimeric protein complex called

Mediator (Allen and Taatjes, 2015). Together, interactions with TFs and individual subunits of Mediator allow for stable association of these proteins with RNA pol II and the consequential formation of enhancer-promoter DNA loops. To ensure that the activity of enhancers is contained to specific target gene promoters, these enhancer-promoter loops are often contained within boundary elements, referred to as topologically associating domains (TADs), that are anchored by CCCTC-binding factor (CTCF) dimers and the cohesin (Box 1) complex (Hadjur et al., 2009; Hnisz et al., 2016a).

Enhancer-promoter interactions can be assembled *de novo* through somatic mutation of noncoding sequences (Fig. 1B). This can arise from activating promoter or enhancer mutations, which can result in aberrant activation of proto-oncogenes, and changes in histone modifications and/or DNA methylation. Furthermore, mutations of CTCF boundary elements, hypermethylation of CTCF-binding motifs, or cohesin-complex mutations, may disengage CTCF binding, leading to loss of enhancer insulation at TAD boundaries, which results in aberrant gene expression of neighbouring genes.

Somatically acquired noncoding mutations and their mechanisms of action in haematological malignancies
Noncoding mutations that nucleate *de novo* enhancers and promoters of oncogenes

Well-characterised examples of noncoding enhancer and promoter mutations have recently been identified in T-ALL genomes. In this cancer, translocations activating transcriptional regulators such as *LMO1/2*, *TAL1/2* and *TLX1/3*, together with the acquisition of activating mutations in the *NOTCH1* gene, are considered hallmarks of the disease (Liu et al., 2017; Pear et al., 1996; Weng et al., 2004). Both TAL and LMO TFs are expressed physiologically during early T cell differentiation and are then progressively silenced, unlike TLX TFs, which are not expressed in the T lineage (Dadi et al., 2012; Dik et al., 2005; Herblot et al., 2000). Together, dysregulation of the aforementioned genes leads to T cell differentiation arrest and rapid proliferation of malignant progenitors. Transcriptional analysis of primary patient samples showed that several oncogenic TFs, including *LMO2*, *TAL1* and *TLX1*, exhibited upregulated expression without an underlying cytogenetic lesion, leaving a subset of cases as genetically 'unresolved' (Ferrando and Look, 2003). In the case of TAL1-positive T-ALL, over half of the patient samples examined overexpress *TAL1*, many with monoallelic expression (Box 1) of *TAL1*, with no detectable cytogenetic aberration at the locus (Ferrando et al., 2004). This finding suggested that another mechanism (undiscovered at the time of this study) was activating the expression of *TAL1* in a subset of patients with T-ALL.

Focusing on this subset of 'unresolved' TAL1-positive cases, our group identified a novel mechanism of oncogene activation, whereby a somatic mutation in a noncoding sequence ~7 kb upstream from the *TAL1* TSS nucleated an aberrant transcriptional enhancer capable of driving TAL1 expression (Table 1) (Fig. 2A) (Mansour et al., 2014). In the study, somatically acquired heterozygous indels 2-18 bp in length introduced binding motifs for the TF MYB at this precise noncoding site. These noncoding mutations were found in 5.5% of unselected paediatric primary T-ALL patient samples, making this region a mutation hotspot.

The fact that these mutations created MYB binding sites was a significant observation. MYB is a master TF in T-ALL and capable of recruiting the TAL/LMO complex assembled from multiple transcriptional activators, including GATA3, HEB (TCF12), RUNX1, the histone acetyltransferase CBP (CREBBP), and TAL1 itself. Biochemical assays have shown that direct binding
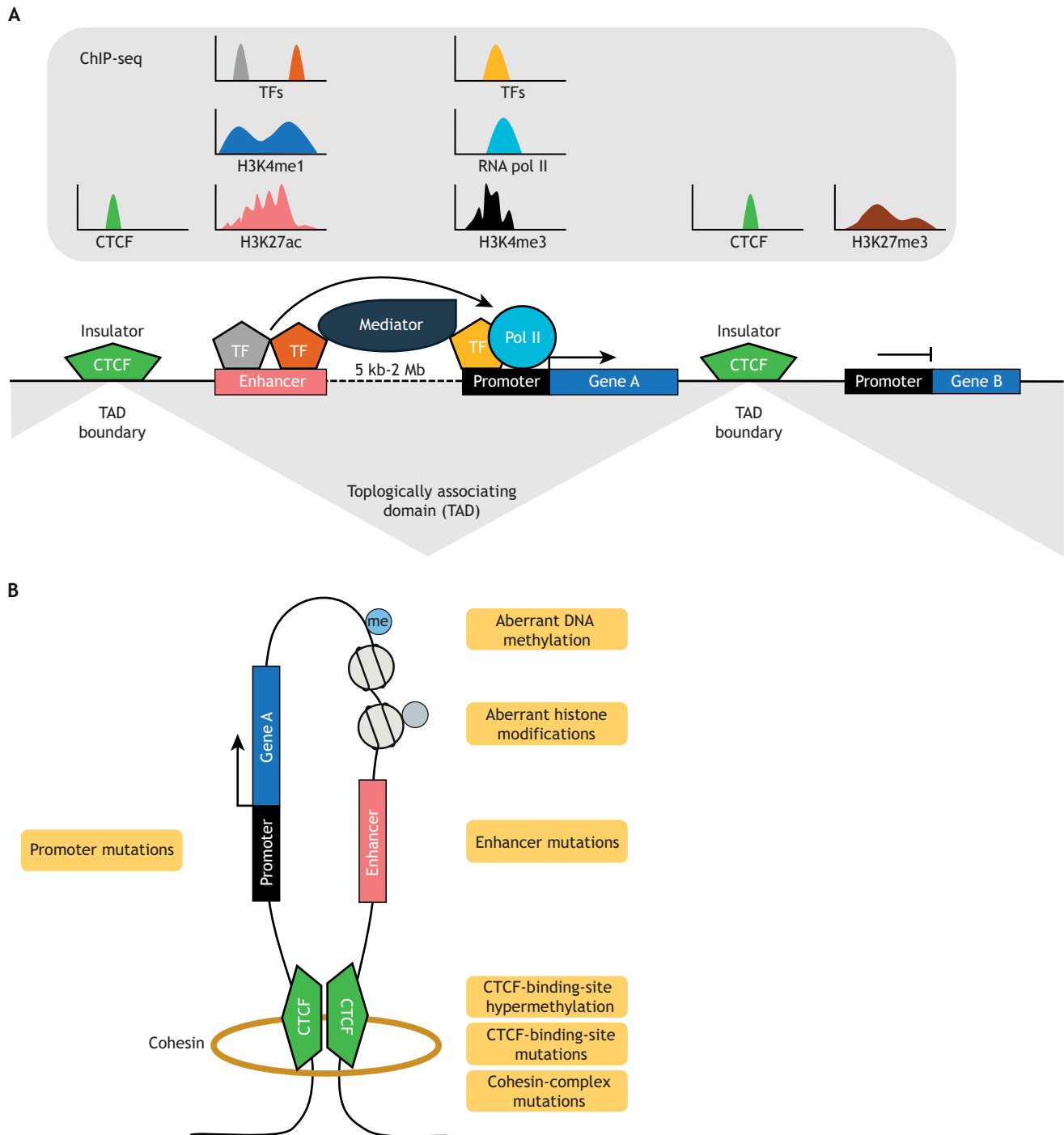
Fig. 1. Canonical enhancer to proximal promoter interaction. (A) Transcription factors bind cognate sequences in the enhancer, which recruits cofactors, histone acetyltransferases and larger protein complexes such as Mediator. Loop formation between the enhancer and proximal promoter of the gene allows enhancer elements to interact with RNA pol II, followed by active transcription of the target genes. Enhancer-promoter interactions can be kept insulated from other genomic loci by the CTCF/cohesin complex, which closes this interaction into an insulated genomic neighbourhood, also known as a topologically associating domain. (B) An enhancer-promoter loop can be formed through somatically acquired mutations leading to aberrant gene regulation. These can arise from mutations that nucleate *de novo* regulatory elements such as promoters or enhancers, by mutations in noncoding sequences bound by CTCF or by mutations of the cohesin-complex members. Such mutation may yield detectable aberrations in DNA methylation and histone modifications.

between MYB and CBP through its KIX domain is required for MYB-dependent transcriptional activation (Zor et al., 2004). Indeed, recruitment of the TAL/LMO complex to this noncoding site upstream of *TAL1* was validated using ChIP-seq (Box 1) in the Jurkat T-ALL cell line and the locus was enriched for H3K27 acetylation, a marker of an active enhancer facilitated by intrinsic

CBP histone acetyltransferase activity. This led to robust expression of *TAL1* from the mutant allele, and no expression from the wild-type allele, explaining the monoallelic expression of *TAL1* described in the Jurkat cell line many years before (Leroy-Viard et al., 1994). It is also congruent that a TF such as MYB, which relies on explicit recognition and binding of a DNA sequence, would be able to exploit

Table 1. Known examples of noncoding mutations in haematological malignancies and associated affects observed

Gene	Mutation	Effect of mutation	Noncoding location	Reference
Chronic lymphocytic leukaemia				
PAX5	Somatic – indel	Reduced expression	Intergenic	(Puente et al., 2015)
NOTCH1	Somatic – indel	Generation of novel splice variant	3′ UTR	(Puente et al., 2015)
T cell acute lymphoblastic leukaemia				
TAL1	Somatic – indel	Increased expression through activating enhancer	Intergenic	(Mansour et al., 2014) (Navarro et al., 2015)
TAL1	Somatic – RAG1/2-mediated episomal reintegration	Increased expression through translocation	Intergenic	(Navarro et al., 2015)
TAL1	Somatic – indel	Increased expression through loss of insulator	Intergenic	(Hnisz et al., 2016b)
LMO2	Somatic – indel	Increased expression through activating enhancer	Intergenic	(Abraham et al., 2017)
LMO2	Somatic – indel	Increased expression through activating promoter	Intronic	(Rahman et al., 2017) (Hu et al., 2017)
LMO2	Somatic – cryptic chromosomal deletion	Increased expression through loss of a negative regulatory element	Intergenic	(Van Vlierberghe et al., 2006)
LMO2	Somatic – indel	Increased expression through loss of insulator	Intergenic	(Hnisz et al., 2016b)
LMO1	Somatic – indel	Increased expression through activating promoter	Promoter	(Li et al., 2017) (Hu et al., 2017)
MYC	Somatic – focal amplification	Increased expression through NOTCH1-dependent enhancer	Intergenic	(Herranz et al., 2014)
Acute myeloid leukaemia				
EVI1	Somatic – chromosomal inversion	Increased expression through translocation of GATA2 enhancer	Intergenic	(Gröschel et al., 2014) (Yamazaki et al., 2014)
GATA2	Somatic – chromosomal inversion	Decreased expression due to the loss of an endogenous enhancer	Intergenic	(Gröschel et al., 2014) (Yamazaki et al., 2014)
NPM1	Somatic – indel	Decreased expression via aberrant microRNA binding to 3′ UTR	3′ UTR	(Cheng et al., 2013)
B cell lymphoma				
TPRG1	Somatic – indel	Increased expression	Intergenic	(Cornish et al., 2019)
CD69	Somatic – copy number variation (loss)	Increased expression	Intergenic	(Cornish et al., 2019)
MMP14	Somatic – copy number variation	Increased expression observed with copy number loss, decreased expression observed with copy number gain	Intergenic	(Cornish et al., 2019)
IGLL5	Somatic – copy number variation (loss)	Decreased expression	Intergenic	(Cornish et al., 2019)
Burkitt lymphoma				
PVT1	Somatic – indel	Potential impact on MYC regulation by the noncoding RNA PVT1	Intronic	(Grande et al., 2019)
PAX5	Somatic – indel	Mutations within distal enhancer of PAX5, an important regulator of B cell differentiation	Intergenic	(Grande et al., 2019)

a somatic mutation in the noncoding genome. All the mutations identified in primary patient samples either matched or were similar to the cognate MYB binding sequence, allowing for MYB-dependent enhancer formation. Furthermore, CRISPR/Cas9-mediated excision of the mutation in the Jurkat T-ALL cell line led to full abrogation of *TAL1* expression (Mansour et al., 2014).

Navarro et al. also independently reported *cis*-acting lesions in the same noncoding locus upstream of *TAL1* in T-ALL, leading to diminished H3K27 trimethylation (Table 1) (Navarro et al., 2015). This study also reported the first case of RAG1/2-mediated episomal reintegration in the same noncoding locus. This event integrated a ~370 kb section of chromosome 7 flanked by recombination signal sequences leading to aberrant *TAL1* expression, demonstrating that this specific noncoding site is able to transcriptionally activate *TAL1* through differing genetic lesions.

Similar to the TAL1-positive cases, a proportion of T-ALL samples also overexpress *LMO2* without a known cytogenetic lesion. Using aberrant ChIP-seq peaks as a discovery tool, our group identified a mutation hotspot within intron 1 of *LMO2*, leading to the formation of an aberrant promoter (Table 1) (Rahman et al., 2017). These somatically acquired *cis*-acting heterozygous mutations create high-confidence *de novo* TF binding motifs, not only for MYB, but also for ETS1 and RUNX1. These somatic *LMO2* mutations were identified in 3.7% of paediatric and 5.5% of adult T-ALL patient samples and in the PF-382 and DU.528 T-ALL cell lines, suggesting that these mutations were indeed driver events.

Analysis of 5′ capped RNA transcripts in the mutant cell lines showed that *LMO2* transcripts skipped exon 1 and instead started transcription from exon 2, confirming that these intronic mutations generate an aberrant *cis*-acting promoter. Interestingly, a nascent MYB motif already exists in this site, with several of the mutations introducing a second MYB site 10 or 20 bp away, suggesting MYB may bind to this mutant locus as a dimer, separated by one or two helical coils of DNA. Indeed, CRISPR/Cas9-mediated deletion that disrupts this spacing led to an almost total loss of *LMO2* expression (Rahman et al., 2017).

Again using ChIP-seq as the discovery tool, somatically acquired noncoding mutations in the *LMO1* oncogene were also identified in 2% of paediatric T-ALL samples and in the Jurkat T-ALL cell line (Table 1) (Li et al., 2017). Li et al. described a single C-to-T nucleotide transition 4 kb upstream of *LMO1* that created a *de novo* MYB binding site, leading to high levels of *LMO1* expression, although it is unclear whether the mutations create an enhancer or a promoter. Further analysis of this mutation by the authors demonstrated that it conforms to an APOBEC-like cytidine deaminase mutation signature, characterised by TCN-to-TTN changes. This mutational signature has been observed across multiple cancer types, beyond those of lymphoid origin, suggesting that somatically acquired oncogenic promoter or enhancer mutations may be found in other cellular lineages (Alexandrov et al., 2013).

In an independent study, Hu et al. used whole-genome sequencing (WGS) of 31 paediatric T-ALL patients and matched

A

2. Recruitment of MYB and other cofactors to create an aberrant enhancer

H3K27ac enrichment

3. Enhancer looping to proximal promoter

| GATA3 | MYB | RUNX1 |

ACCGTTA
ACCGTTAGG
AAACGGTAA

1. Somatically acquired *de novo* MYB motif recurrent in T-ALL patient samples

RNA pol II — *TAL1*

B

3. Long-range enhancer interaction with promoter (~1.4 Mb)

RBP-J
NOTCH-1

2. Binding of NOTCH1 and binding partner RBP-J induces enhancer formation

RNA pol II — *MYC*

4. Activation of *MYC* proto-oncogene

N-Me

1. Focal amplification in T-ALL patients of a NOTCH1 dependent distal enhancer

C

2. Aberrant splice variant is created from a cryptic donor site in exon 34

Aberrant splicing

1. Recurrent mutations identifie in CLL patient samples create cryptic splice acceptor site

NOTCH1

AGGT | PEST |

530 bp deletion

3. Truncated NOTCH1 lacks the negative regulatory PEST domain
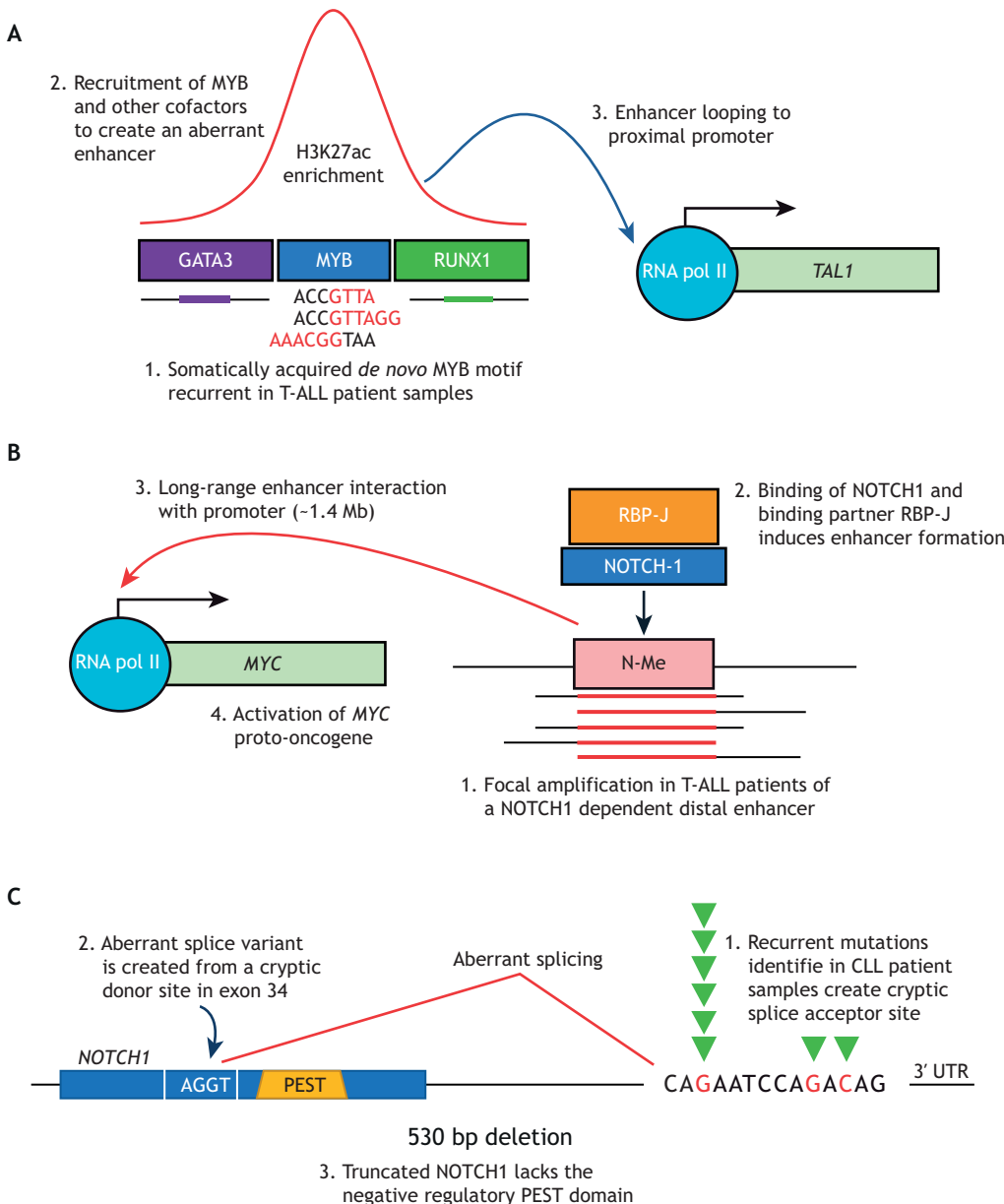
CAGAATCCAGACAG   3′ UTR

Fig. 2. Examples of noncoding mutations in haematological malignancies and mechanisms of action. (A) Recurrent heterozygous somatic mutations of a noncoding element 5′ to *TAL1* creates a *de novo* binding site for the transcription factor MYB in T-ALL. Recruitment of MYB and additional cofactors leads to the formation of an aberrant enhancer, as seen by an enrichment of H3K27 acetylation. This enhancer is then able to interact with the proximal promoter of *TAL1* causing its monoallelic expression. (B) An endogenous NOTCH1-dependent enhancer is located 1.4 Mb away from the proximal promoter of *MYC* in T-ALL. Examination of primary patient samples demonstrated that this element is frequently and focally amplified in T-ALL, a cancer that frequently presents with mutations that constitutively activate NOTCH1. (C) Recurrent mutations in the 3′ UTR of *NOTCH1* were identified in CLL. These mutations lead to the formation of an aberrant splice acceptor site. A truncated splice variant is created between a cryptic splice donor site in the preceding exon and this mutant splice acceptor site. This allows for the formation of transcript that excises the negative regulatory PEST domain of NOTCH1, resulting in a more stable NOTCH1 protein.

germline samples to identify the previously discussed somatically acquired noncoding variants at the *TAL1* and *LMO2* loci, but also the *LMO1* promoter mutation (Table 1) (Hu et al., 2017). The authors also transduced the Jurkat T-ALL cell line with a doxycycline-inducible CRISPR/Cas9 system targeting the enhancer mutation and injected these engineered cells into immunocompromised mice to evaluate the effects of the *LMO1* enhancer mutation *in vivo*. Importantly, mice on a doxycycline-supplemented diet, in which the activated CRISPR/Cas9 reagents excised the *LMO1* enhancer mutation and thus suppressed LMO1 overexpression, exhibited lower leukaemic engraftment at week 4 and improved survival compared to mice on a regular diet, providing clear evidence that this enhancer variant is important for the pathogenesis of this leukaemia.

Genetic lesions that alter the function of endogenous enhancers

Cis-acting enhancers can regulate genes across vast genomic distances. An excellent example of this is the NOTCH1-bound *MYC* enhancer (N-Me) in T-ALL, which is situated +1.47 Mb

relative to the *MYC* TSS (Table 1) (Fig. 2B) (Herranz et al., 2014). Unlike the previous examples, this *cis*-acting enhancer is not acquired through mutation, but is a highly conserved endogenous regulatory sequence with a critical role in T cell development. The authors demonstrated that NOTCH1 binds to this *cis*-acting element along with its binding partner RBPJ, leading to transcriptional activation of *MYC*. Analysis of primary T-ALL patient samples identified focal duplications of chromosome 8q24, which includes the N-Me locus, in 5% of cases, suggesting these leukaemias amplify this locus to enhance MYC expression. The authors also determined that this enhancer is essential for normal thymic development. Deletion of N-Me reduced thymic cellularity and mature T cell counts, but made knockout mice resistant to NOTCH1-induced T-ALL. Further analysis of this regulatory element has identified the role for GATA3-mediated nucleosomal eviction in its activation (Belver et al., 2019)

The examples described thus far illustrate the gain-of-function capability of noncoding aberrations in driving oncogene expression. In a study of CLL, the converse was shown to be true – noncoding

lesions leading to the inactivation of a tumour suppressor gene. Here, recurrent indels inactivate the B cell developmental enhancer of *PAX5*, leading to PAX5 haplosufficiency (Table 1) (Puente et al., 2015). Importantly, the authors functionally validated the enhancer using CRISPR/Cas9 genome editing in which deletions of the putative enhancer in a lymphoblastoid B cell line led to a 40% reduction in *PAX5* expression. A similar mechanism has also been described in the inherited MonoMAC syndrome (Box 1), in which mutations inactivate a *GATA2* intronic enhancer leading to GATA2 haplosufficiency (Hsu et al., 2013).

Genetic lesions that alter the function of endogenous enhancers have also been characterised in AML. Gröschel et al. showed that recurrent inv(3)(q21q26.2) or t(3;3)(q21;q26.2) [inv(3)/t(3;3)] rearrangements, which lead to aberrant expression of the proto-oncogene *EVI1* in AML, are caused by the translocation of an endogenous enhancer which usually regulates *GATA2* (Table 1) (Gröschel et al., 2014). These genetic lesions translocate an 18 kb noncoding sequence to the *EVI1* (*MECOM*) gene locus. ChIP-seq analysis of the inv(3) MOLM-1 AML cell line defined a putative enhancer element within this translocated 18 kb noncoding sequence owing to its significant enrichment for the p300 transcriptional coactivator (EP300) and for the H3K27ac, H3K4me3 and H3K4me1 histone marks. Interestingly, the authors also show that loss of the enhancer from its usual resident position in the genome leads to *GATA2* haploinsufficiency and to its monoallelic expression from the unaffected allele. Cloning of this enhancer sequence into a reporter assay resulted in strong reporter gene induction within the inv(3) myeloid cell lines MUTZ-3 and MOLM-1, but not in non-myeloid cell lines. This suggested that this enhancer required myeloid-specific transcriptional components to activate. Excision of this p300-bound translocated enhancer with TALE nucleases silenced *EVI1* expression and stalled cellular growth in the inv(3) AML cell line, further demonstrating the driver nature of this genetic aberration. This finding was confirmed in a complimentary *in vivo* study by Yamazaki et al., in which a bacterial artificial chromosome encoding the inversion generated a transplantable leukaemia in mice (Table 1) (Yamazaki et al., 2014). The inv(3)/t(3;3) offers a good example of enhancer hijacking, and shows how a single genetic lesion can lead to activation of an oncogene and inactivation of a tumour suppressor in a single 'hit'.

Noncoding variants that lead to aberrant splicing

Alternative splicing facilitates greater transcriptional diversity from a single precursor messenger RNA under differing cellular contexts. Most human genes (94%) can generate different mRNA isoforms, which can ultimately be translated into different proteins (Gerstein et al., 2014; Pan et al., 2008). Data from TCGA has shown that cancer genomes can acquire somatic mutations that alter normal splicing mechanisms, including intron retention, which is thought to inactivate tumour suppressors (Jung et al., 2015). By analysing genome-wide patterns of RNA splicing in AML genomes, Dvinge and Bradley identified that the extent of intron retention correlated with mutations in *RUNX1*, *IDH1* and *IDH2* relative to wild-type AML samples. The authors postulate that differential methylation arising from *IDH1/2* mutations may have an impact on RNA splicing through a crosstalk mechanism, although they did not experimentally investigate this (Dvinge and Bradley, 2015). TCGA data have also shown that AMLs can acquire somatic mutations within the noncoding sequences that demarcate the two positions that should be spliced together, also known as the 5′ splice donor and the 3′ splice acceptor sites (Kandoth et al., 2013).

By interrogating CLL genomes by WGS, Puente et al. identified recurrent mutations in the noncoding 3′ untranslated region (UTR) of *NOTCH1* (Table 1; Fig. 2C) (Puente et al., 2015). RNA-seq examination of the mutated cancers confirmed the presence of aberrant splicing of *NOTCH1*, where the final exon was spliced into a newly formed splice acceptor site in the 3′ UTR of *NOTCH1*. As a result of this, 158 coding bases were deleted, including those encoding the negative regulatory PEST domain of the mature NOTCH1 protein. Loss of this domain interferes with the ubiquitination-mediated degradation of NOTCH1, giving rise to increased levels of activated NOTCH1. Indeed, immunohistochemistry confirmed the presence NOTCH1 in the nuclei of mutant CLL cells, and data from patients with the 3′ UTR *NOTCH1* mutations suggested that this mutation was concomitant with adverse prognosis.

Genetic lesions that disrupt negative regulatory noncoding sequences

In addition to enhancers, the noncoding genome is also rich with negative regulatory sequences, or insulators. These block the activity of enhancers, and are usually positioned between the enhancer of a gene and its proximal promoter. Negative regulatory elements have been identified upstream of the *LMO2* gene. In a study by Hammond et al., the authors set out to understand the contribution of upstream noncoding sequences on the regulation of *LMO2* expression (Hammond et al., 2005). First, they cloned 3190 bp 5′ to the TSS into luciferase reporters and tested reporter activity in different cell lines. In erythroid cells, where *LMO2* is constitutively expressed, expression of the reporter gene was maintained, however the Jurkat T-ALL cell line, where *LMO2* is not active, had significantly reduced reporter activity. This suggested the presence of a tissue-specific negative regulatory element within the noncoding sequence. Building on this research, Van Vlierberghe et al. identified a recurrent cryptic deletion del(11)(p12p13) in T-ALL using the array-comparative genome hybridisation (array-CGH) technique (Table 1) (Van Vlierberghe et al., 2006). This cryptic deletion led to the loss of the aforementioned negative regulatory region of *LMO2*, leading to aberrant activation of the proximal promoter and *LMO2* expression. A follow up study successfully improved on the detection method by using a multiplex ligations probe assay to identify additional monoallelic LMO2 expressors with deletions of the negative regulatory element (Van Vlierberghe et al., 2008).

Higher-order chromatin conformation can regulate genes by dynamically altering the interactions between gene promoters and regulatory elements. These are often held together in DNA loops, where the loop itself is closed by the binding of two CTCFs, creating an insulated neighbourhood. These loops can further be organised in larger TADs (Dixon et al., 2012; Dowen et al., 2014). This led Hnisz et al. to postulate that mutations could affect the boundary elements between insulated neighbourhoods, which in turn may lead to the deregulated expression of proto-oncogenes (Table 1) (Hnisz et al., 2016b). The authors mapped the Jurkat T-ALL cell line genome with chromatin interaction analysis by paired-end tag sequencing (ChIA-PET; Box 1). This technique allowed for the identification of CTCF-CTCF looping interactions across the genome. Notably, the authors showed that *TAL1* is located in its own insulated neighbourhood in Jurkat cells, with one CTCF site found in the middle of a neighbouring gene called *STIL*, and the other located in a noncoding region 3′ of *TAL1*. Interestingly, a subset of primary T-ALL patient samples appeared to have genomic deletions of the CTCF site within *STIL*, which in effect destroys the insulation boundary of the neighbourhood. This now meant that

TAL1 could be regulated by distal enhancers, which it would usually be insulated from, leading to its overexpression. Similar mutations were also found in patient samples at the *LMO2* locus, although it is unclear whether these also deleted the previously discussed negative regulatory element of *LMO2* (Van Vlierberghe et al., 2008). CRISPR/Cas9-mediated deletion of CTCF-associated boundary sites in HEK293T cells led to an almost 2-fold increase in both *TAL1* and *LMO2* expression, demonstrating that loss of the insulated neighbourhood in an unrelated cell line can upregulate the expression of these proto-oncogenes.

Although recurrent mutations disrupting CTCF motifs themselves have been described in solid tumours (Katainen et al., 2015), to our knowledge, such mutations have not been described in haematological malignancies. However, it is notable that CTCF binding to DNA is sensitive to DNA methylation (Bell and Felsenfeld, 2000; Hark et al., 2000). In gliomas, Flavahan et al. describe how gain-of-function IDH family gene mutations leads to the production of an onco-metabolite called 2-hydroxygluterate, which interferes with the function of TET family 5′-methylcytosine hydroxylases (Flavahan et al., 2016). This ultimately leads to increased DNA methylation, which, when occurring at CTCF sites, impairs CTCF binding. The authors demonstrate that methylation of a CTCF site downstream of *FIP1L1* leads to loss of insulation and upregulation of the neighbouring oncogene *PDGFRA*. Given the prevalence of *IDH1/2* and *TET2* mutations in AML, it is tempting to speculate that similar mechanisms are at play in this disease (Cancer Genome Atlas Research Network et al., 2013a; Marcucci et al., 2010; Papaemmanuil et al., 2016). Furthermore, many TFs, including MYB, have recently been shown to have altered affinity for methylated DNA, and it is therefore likely that DNA methylation plays an important role in the transcriptional dysregulation seen in many haematopoietic malignancies. (Yin et al., 2017).

Challenges and potential methods to identify novel functional noncoding mutations
Challenges
The identification and characterisation of noncoding mutations in haematological malignancies is fraught with challenges. The principal problem is to separate driver mutations from passenger mutations. Intrinsic to this issue is that variants in the noncoding genome do not alter amino acids, which means that immediate functional characterisation of a variant is not possible. Furthermore, passenger mutations are acquired over the course of an average human lifespan, which could occur before or after the acquisition of a driver mutation. Martincorena and Campbell discuss this in their review and also state the importance of sequencing normal tissues alongside malignant ones to distinguish passenger mutations from rare germline variants (Martincorena and Campbell, 2015). An important challenge is to isolate mutations that are positively selected for during the development of the malignancy, and under this circumstance, mutational recurrence and allele frequency across primary patient samples may provide a clue.

Given the vastness of the noncoding genome, another challenge is to know where to look for driver mutations. In much of the noncoding genome, variations will have no observable phenotypic affect, therefore it is important to restrict search space to sequences that may have regulatory potential on other genes. To this end, acquisition of datasets that provide insights into accessible regions of chromatin, and the position of active enhancers or promoters may be valuable. Without further experimental interrogation, it can still be difficult to connect specific enhancers to the target genes that these elements regulate. It is therefore also important to establish causality with a

possible phenotypic outcome. Genome engineering techniques such as CRISPR/Cas9 have been crucial in this regard, as they allow for the reversion of genetic lesions back to wild type, or to knock in mutations where they were not present before and then assess the phenotype.

Proposed methods to identify functional noncoding variants
A number of differing methods have been employed to identify functional noncoding variants. Some studies, including those from our own research group, have successfully examined biological samples that exhibit monoallelic expression of proto-oncogenes without underlying cytogenetic lesions at the locus (Li et al., 2017; Mansour et al., 2014; Rahman et al., 2017). In these studies, integrated datasets have been helpful where ChIP-seq for H3K27 acetylation and the TF MYB has limited search space for noncoding sequences with regulatory potential, which have then ultimately harboured recurrent somatically acquired mutations in primary patient samples. Monoallelic expression can be identified by RNA-seq, albeit with a number of challenges. Some notable examples being that genes imprinted as part of mammalian development will be identified, and such genes may not harbour *cis*-acting lesions or be relevant to the disease state. Furthermore, accurate calling of monoallelic expression relies on an awareness of heterozygous SNP positions in exons or UTRs to appropriately distinguish expression between two alleles. Genes that have no heterozygous SNPs cannot be assessed for allelic expression using this method, limiting its utility as a truly genome-wide approach.

Chromatin accessibility can indicate the presence of active enhancers or promoters. These can be determined through multiple techniques, including ChIP-seq, assay for transposase-accessible chromatin sequencing (ATAC-seq), and DNase-seq (Box 1) (Boyle et al., 2008; Buenrostro et al., 2013, 2015). Abraham et al. developed a novel bioinformatic pipeline to examine H3K27 acetylation ChIP-seq reads for indels (Table 1) (Abraham et al., 2017). H3K27ac limits the search space to only ∼2% of the genome, and is particularly suited for discovery of gain-of-function mutations. This pipeline effectively uncovered a recurrent mutation hotspot 40 kb upstream from *LMO2* that forms an aberrant enhancer to drive *LMO2* expression in T-ALL.

Integration of WGS with ATAC-seq can also identify cancer-relevant mutations in the noncoding genome (Corces et al., 2018). In this study, mutations were identified in ATAC-seq reads with variant allele frequencies (VAF) >80%, even though the same variant was identified as heterozygous with a 50% VAF by WGS reads. This discrepancy suggested that the mutant allele was more accessible to transposase and the site of a novel *cis*-regulatory element. This approach successfully characterised a noncoding enhancer mutation creating a predicted *de novo* binding site for the TF NKX2-8, leading to upregulation of *FGD4*, an important regulator of the actin cytoskeleton in bladder cancer. In a recent study, WGS from 117 patients with B cell lymphoma was coupled with promoter capture Hi-C (Box 1) data from naïve B cells to define *cis*-regulatory elements (Table 1) (Cornish et al., 2019). Using this method, the authors identified copy number variations and single nucleotide variants at *cis*-regulatory elements that affect the expression of *MMP14* and *IGLL5*, where low expression is believed to be associated with poor prognosis. The authors also identified aberrations that affected the expression of *CD69* and *TPRG1*, where increased expression has been previously implicated in lymphoma oncogenesis.

By examining the existing literature, it is possible to generate a framework to identify noncoding mutations (Fig. 3). In almost all studies, the search space for novel mutations was restricted by

exploring a gene of interest previously implicated in disease pathology, or by identifying regulatory elements in the noncoding genome. Studies also had access to transcriptomic data to assess the impact of newly identified regulatory regions on candidate genes.

Given the complexity of the data from sequencing runs, robust bioinformatic analysis is crucial to identify novel indels or to subject candidate noncoding regions to motif analysis (Hume et al., 2015; Khan et al., 2018; Matys et al., 2003). This may help identify
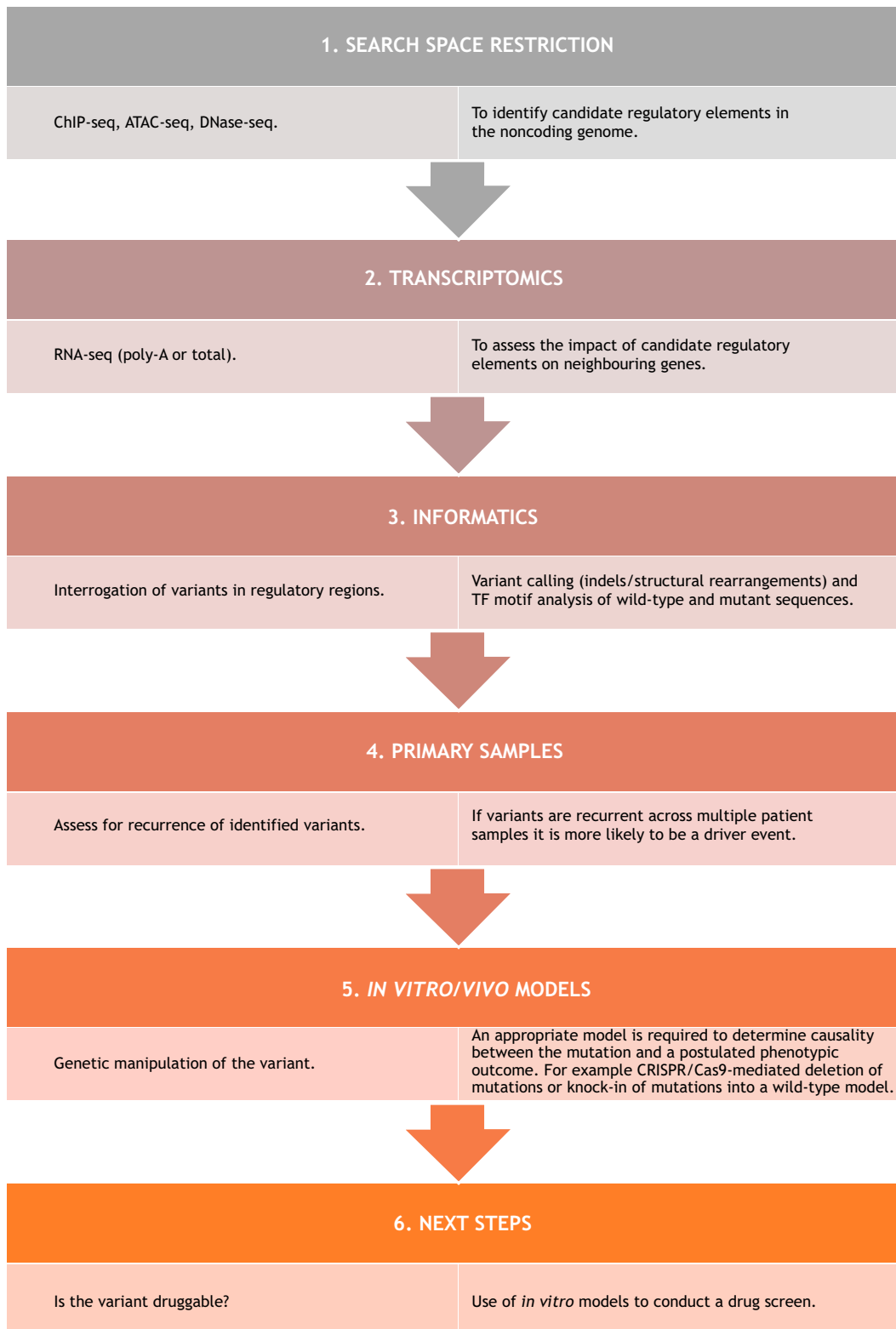
1. SEARCH SPACE RESTRICTION

ChIP-seq, ATAC-seq, DNase-seq.

To identify candidate regulatory elements in the noncoding genome.

2. TRANSCRIPTOMICS

RNA-seq (poly-A or total).

To assess the impact of candidate regulatory elements on neighbouring genes.

3. INFORMATICS

Interrogation of variants in regulatory regions.

Variant calling (indels/structural rearrangements) and TF motif analysis of wild-type and mutant sequences.

4. PRIMARY SAMPLES

Assess for recurrence of identified variants.

If variants are recurrent across multiple patient samples it is more likely to be a driver event.

5. *IN VITRO/VIVO* MODELS

Genetic manipulation of the variant.

An appropriate model is required to determine causality between the mutation and a postulated phenotypic outcome. For example CRISPR/Cas9-mediated deletion of mutations or knock-in of mutations into a wild-type model.

6. NEXT STEPS

Is the variant druggable?

Use of *in vitro* models to conduct a drug screen.

Fig. 3. Framework for identifying novel noncoding mutations. Proposed research framework for the successful identification and functional characterisation of noncoding variants in human malignancies.

Disease Models & Mechanisms

de novo TF binding motifs, which itself may provide insight into the mechanism of action. If a variant is a true oncogenic noncoding mutation, it should be recurrent, and so access to primary patient-derived samples to validate the presence of variants is important, and such variants should be absent from germline controls, or absent from remission samples if indeed they are somatically acquired. Finally, detailed mechanistic insights have been achieved by examining noncoding mutations in suitable cellular or organismal models. These variants are best explored within their endogenous loci, hence detailed phenotypic assessment should be attempted following CRISPR/Cas9-mediated excision of the mutations in models which already harbour the mutation, or knock-in of the mutations into a wild-type locus.

Potential therapeutic interventions

Somatically acquired noncoding mutations identified thus far in haematological malignancies have common molecular features. As described in this Review, these variants create stable and potent enhancers with massive enrichment of H3K27 acetylation, orders of magnitude above that of other endogenous enhancers. Some researchers have postulated that these enhancers are in effect 'super-enhancers', locus-control regions or just 'strong' enhancers (Pott and Lieb, 2015). Although there is continued debate regarding semantics, there is growing consensus that these enhancer complexes can be disassembled through therapeutic intervention. Stable enhancers are usually occupied by RNA pol II, Mediator, cohesin, p300, CBP and the bromodomain and extraterminal (BET) protein BRD4 (Hnisz et al., 2013). BET proteins such as BRD4 detect histone acetylation at specific genomic loci and TF binding events, which then further recruits the Mediator complex and RNA pol II elongation factors (Shi and Vakoc, 2014).

Lovén et al. examined the effects of the BET inhibitor JQ1 in multiple myeloma (Lovén et al., 2013). In the study, the authors observed that treatment with JQ1 depletes BRD4 specifically at 'super-enhancers' that are massively enriched with MED1 and BRD4. This included the selective inhibition of the enhancers that regulate the *MYC* oncogene. This was a particularly significant finding as it suggested there could be a therapeutic window to selectively inhibit the strongest enhancers, which in turn regulate the strongest oncogenes, in this malignancy. Further success with JQ1 was also described by Gröschel et al., where treatment with the compound led to significantly reduced *EVI1* expression and growth arrest of inv(3) AML cells, but no sensitivity in cells that have no rearrangement at that locus and yet overexpress *EVI1* (Gröschel et al., 2014). The authors postulate that this is because there is an active 'super-enhancer' driving *EVI1* expression within inv(3) AML cells. Recent evidence suggests that BET degraders may also be an effective way to stop aberrant enhancer activity (Winter et al., 2017). Unlike BET inhibitors such as JQ1, which may only be effective in a subset of massively enriched enhancer regions, the authors state that a BET degrader called dBET6 leads to a total collapse of global transcriptional elongation, at a similar magnitude of CDK9 inhibition. By treating T-ALL cells with dBET6, the authors demonstrate efficacy of the degrader at lower concentrations compared to JQ1, and increased survival of T-ALL-engrafted mice treated with the degrader, compared to mice treated with the inhibitor. Furthermore, inhibitors of the proteins that regulate C-terminal domain phosphorylation of RNA pol II have been shown to effectively downregulate the expression of genes that are regulated by enhancers. This includes the CDK7 inhibitor THZ1, which leads to potent suppression of the *TAL1* enhancer in Jurkat T-ALL cells, raising the possibility that targeting noncoding drivers

of oncogenes may offer an exploitable therapeutic window in cases where oncogenes are aberrantly expressed but do not carry an obvious cytogenetic change (Kwiatkowski et al., 2014).

Future outlook

Recent advances have made it clear that the noncoding genome is not only acquiring passenger mutations, but also driver mutations that cause the dysregulated expression of proto-oncogenes in haematological malignancies. It is probable that such mutations exist in other malignancies, but have remained under-reported owing to the challenges faced with initial discovery and subsequent functional characterisation. As we continue into the genomic age, researchers are generating vast integrated datasets that include WGS, which can readily identify mutations, RNA-seq for the quantification of gene expression, and ChIP/ATAC/DNase-seq, all of which are highly effective at mapping gene regulatory elements. This provides an opportunity to not only discover novel noncoding mutations, but to assess these mutations for phenotypic outcomes, and to establish cause and effect with the assistance of genetic perturbation and gene editing technologies that have become ubiquitous in translational research. Identifying such mutations has several potential implications; they may have prognostic relevance, large indels may be amenable to minimal residual disease analysis, they can highlight novel oncogenes that are potentially druggable, and they can provide important insights into the optimal DNA syntax required for nucleation of multiprotein TF complexes. The factors that interplay to orchestrate transcriptional regulation in normal development, and transcriptional dysregulation in cancer, are highly complex. It may be that noncoding mutations selected for during tumorigenesis will prove to be 'experiments of nature', the Rosetta stone that helps us understand the language of the noncoding genome.

References

Abraham, B. J., Hnisz, D., Weintraub, A. S., Kwiatkowski, N., Li, C. H., Li, Z., Weichert-Leahey, N., Rahman, S., Liu, Y., Etchin, J. et al. (2017). Small genomic insertions form enhancers that misregulate oncogenes. *Nat. Commun.* **8**, 14385. doi:10.1038/ncomms14385

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L. et al. (2013). Signatures of mutational processes in human cancer. *Nature* **500**, 415-421. doi:10.1038/nature12477

Allen, B. L. and Taatjes, D. J. (2015). The mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155-166. doi:10.1038/nrm3951

Bell, A. C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-485. doi:10.1038/35013100

Belver, L., Yang, A. Y., Albero, R., Herranz, D., Brundu, F. G., Quinn, S. A., Perez-Duran, P., Alvarez, S., Gianni, F., Rashkovan, M. et al. (2019). Gata3-controlled nucleosome eviction drives Myc enhancer activity in T-cell development and leukemia. *Cancer Discov.* doi:10.1158/2159-8290.CD-19-0471

Bettridge, J., Na, C. H., Pandey, A. and Desiderio, S. (2017). H3K4me3 induces allosteric conformational changes in the DNA-binding and catalytic regions of the V(D)J recombinase. *Proc. Natl. Acad. Sci. USA* **114**, 1904-1909. doi:10.1073/pnas.1615727114

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S. and Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-322. doi:10.1016/j.cell.2007.12.014

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic

profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213-1218. doi:10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Chang, H. Y. and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–9. doi:10.1002/0471142727.mb2129s109

Cancer Genome Atlas Research Network, **Ley, T. J., Miller, C., Ding, L., Raphael, B. J., Mungall, A. J., Robertson, A. G., Hoadley, K., Triche, T. J., Laird, P. W. et al.** (2013a). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059-2074. doi:10.1056/NEJMoa1301689

Cancer Genome Atlas Research Network, **Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J. M.** (2013b). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113-1120. doi:10.1038/ng.2764

Cheng, C. K., Kwan, T. K., Cheung, C. Y., Ng, K., Liang, P., Cheng, S. H., Chan, N. P. H., Ip, R. K. L., Wong, R. S. M., Lee, V. et al. (2013). A polymorphism in the 3′-untranslated region of the NPM1 gene causes illegitimate regulation by microRNA-337-5p and correlates with adverse outcome in acute myeloid leukemia. *Haematologica* **98**, 913-917. doi:10.3324/haematol.2012.073015

Corces, M. R., Granja, J. M., Shams, S., Louie, B. H., Seoane, J. A., Zhou, W., Silva, T. C., Groeneveld, C., Wong, C. K., Cho, S. W. et al. (2018). The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898. doi:10.1126/science.aav1898

Cornish, A. J., Hoang, P. H., Dobbins, S. E., Law, P. J., Chubb, D., Orlando, G. and Houlston, R. S. (2019). Identification of recurrent noncoding mutations in B-cell lymphoma using capture Hi-C. *Blood Adv.* **3**, 21-32. doi:10.1182/bloodadvances.2018026419

Dadi, S., Le Noir, S., Payet-Bornet, D., Lhermitte, L., Zacarias-Cabeza, J., Bergeron, J., Villarèse, P., Vachez, E., Dik, W. A., Millien, C. et al. (2012). TLX homeodomain oncogenes mediate T cell maturation arrest in T-ALL via interaction with ETS1 and suppression of TCRα gene expression. *Cancer Cell* **21**, 563-576. doi:10.1016/j.ccr.2012.02.013

Dik, W. A., Pike-Overzet, K., Weerkamp, F., de Ridder, D., de Haas, E. F. E., Baert, M. R. M., van der Spek, P., Koster, E. E. L., Reinders, M. J. T., van Dongen, J. J. M. et al. (2005). New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201**, 1715-1723. doi:10.1084/jem.20042524

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380. doi:10.1038/nature11082

Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., Weintraub, A. S., Schuijers, J., Lee, T. I., Zhao, K. et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374-387. doi:10.1016/j.cell.2014.09.030

Dvinge, H. and Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45. doi:10.1186/s13073-015-0168-9

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74. doi:10.1038/nature11247

Ferrando, A. A. and Look, A. T. (2003). Gene expression profiling in T-cell acute lymphoblastic leukemia. *Semin. Hematol.* **40**, 274-280. doi:10.1016/S0037-1963(03)00195-1

Ferrando, A. A., Herblot, S., Palomero, T., Hansen, M., Hoang, T., Fox, E. A. and Look, A. T. (2004). Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. *Blood* **103**, 1909-1911. doi:10.1182/blood-2003-07-2577

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suvà, M. L. and Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114. doi:10.1038/nature16490

Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y., Tanaka, H., Taniguchi, H., Kawakami, Y., Ueno, M. et al. (2016). Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500-509. doi:10.1038/ng.3547

Gerstein, M. B., Rozowsky, J., Yan, K.-K., Wang, D., Cheng, C., Brown, J. B., Davis, C. A., Hillier, L., Sisu, C., Li, J. J. et al. (2014). Comparative analysis of the transcriptome across distant species. *Nature* **512**, 445-448. doi:10.1038/nature13424

Grande, B. M., Gerhard, D. S., Jiang, A., Griner, N. B., Abramson, J. S., Alexander, T. B., Allen, H., Ayers, L. W., Bethony, J. M., Bhatia, K. et al. (2019). Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood* **133**, 1313-1324. doi:10.1182/blood-2018-09-871418

Gröschel, S., Sanders, M. A., Hoogenboezem, R., de Wit, E., Bouwman, B. A. M., Erpelinck, C., van der Velden, V. H. J., Havermans, M., Avellino, R., van Lom, K. et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369-381. doi:10.1016/j.cell.2014.02.019

Haberle, V. and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621-637. doi:10.1038/s41580-018-0028-8

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G. and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**, 410-413. doi:10.1038/nature08079

Hammond, S. M., Crable, S. C. and Anderson, K. P. (2005). Negative regulatory elements are present in the human LMO2 oncogene and may contribute to its expression in leukemia. *Leuk. Res.* **29**, 89-97. doi:10.1016/j.leukres.2004.05.013

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M. and Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486-489. doi:10.1038/35013106

Helleday, T., Eshtad, S. and Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585-598. doi:10.1038/nrg3729

Herblot, S., Steff, A.-M., Hugo, P., Aplan, P. D. and Hoang, T. (2000). SCL and LMO1 alter thymocyte differentiation: inhibition of E2A-HEB function and pre-T alpha chain expression. *Nat. Immunol.* **1**, 138-144. doi:10.1038/77819

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S. A., Belver, L., Wendorff, A. A., Xu, L., Castillo-Martin, M., Llobet-Navás, D., Cordon-Cardo, C. et al. (2014). A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130-1137. doi:10.1038/nm.3665

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A. and Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947. doi:10.1016/j.cell.2013.09.053

Hnisz, D., Day, D. S. and Young, R. A. (2016a). Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* **167**, 1188-1200. doi:10.1016/j.cell.2016.10.024

Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., Goldmann, J., Lajoie, B. R., Fan, Z. P., Sigova, A. A. et al. (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458. doi:10.1126/science.aad9024

Horn, S., Figl, A., Rachakonda, P. S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K. et al. (2013). TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-961. doi:10.1126/science.1230062

Hsu, A. P., Johnson, K. D., Falcone, E. L., Sanalkumar, R., Sanchez, L., Hickstein, D. D., Cuellar-Rodriguez, J., Lemieux, J. E., Zerbe, C. S., Bresnick, E. H. et al. (2013). GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* **121**, 3830-3837. doi:10.1182/blood-2012-08-452763

Hu, S., Qian, M., Zhang, H., Guo, Y., Yang, J., Zhao, X., He, H., Lu, J., Pan, J., Chang, M. et al. (2017). Whole-genome noncoding sequence analysis in T-cell acute lymphoblastic leukemia identifies oncogene enhancer mutations. *Blood* **129**, 3264-3268. doi:10.1182/blood-2017-03-771162

Huang, F. W., Hodis, E., Xu, M. J., Kryukov, G. V., Chin, L. and Garraway, L. A. (2013). Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-959. doi:10.1126/science.1229259

Hume, M. A., Barrera, L. A., Gisselbrecht, S. S. and Bulyk, M. L. (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **43**, D117-D122. doi:10.1093/nar/gku1045

Jagannathan-Bogdan, M. and Zon, L. I. (2013). Hematopoiesis. *Development* **140**, 2463-2467. doi:10.1242/dev.083147

Jung, H., Lee, D., Lee, J., Park, D., Kim, Y. J., Park, W.-Y., Hong, D., Park, P. J. and Lee, E. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **47**, 1242-1248. doi:10.1038/ng.3414

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A. et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-339. doi:10.1038/nature12634

Kansagra, A., Dahiya, S. and Litzow, M. (2018). Continuing challenges and current issues in acute lymphoblastic leukemia. *Leuk. Lymphoma* **59**, 526-541. doi:10.1080/10428194.2017.1335397

Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., Gylfe, A. E., Ristolainen, H., Hänninen, U. A., Cajuso, T. et al. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* **47**, 818-821. doi:10.1038/ng.3335

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S. R., Tan, G. et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260-D266. doi:10.1093/nar/gkx1126

Kwiatkowski, N., Zhang, T., Rahl, P. B., Abraham, B. J., Reddy, J., Ficarro, S. B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B. et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616-620. doi:10.1038/nature13393

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R. and Weirauch, M. T. (2018). The human transcription factors. *Cell* 172, 650-665. doi:10.1016/j.cell.2018.01.029

Leroy-Viard, K., Vinit, M. A., Lecointe, N., Mathieu-Mahul, D. and Romeo, P. H. (1994). Distinct DNase-I hypersensitive sites are associated with TAL-1 transcription in erythroid and T-cell lines. *Blood* 84, 3819-3827. doi:10.1182/blood.V84.11.3819.bloodjournal84113819

Li, Z., Abraham, B. J., Berezovskaya, A., Farah, N., Liu, Y., Leon, T., Fielding, A., Tan, S. H., Sanda, T., Weintraub, A. S. et al. (2017). APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL. *Leukemia* 31, 2057-2064. doi:10.1038/leu.2017.75

Liu, Y., Easton, J., Shao, Y., Maciaszek, J., Wang, Z., Wilkinson, M. R., McCastlain, K., Edmonson, M., Pounds, S. B., Shi, L. et al. (2017). The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* 373, 1541. doi:10.1038/ng.3909

Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I. and Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320-334. doi:10.1016/j.cell.2013.03.036

Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., Etchin, J., Lawton, L., Sallan, S. E., Silverman, L. B. et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* 346, 1373-1377. doi:10.1126/science.1259037

Marcucci, G., Maharry, K., Wu, Y.-Z., Radmacher, M. D., Mrózek, K., Margeson, D., Holland, K. B., Whitman, S. P., Becker, H., Schwind, S. et al. (2010). IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J. Clin. Oncol.* 28, 2348-2355. doi:10.1200/JCO.2009.27.3730

Marculescu, R., Le, T., Simon, P., Jaeger, U. and Nadel, B. (2002). V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J. Exp. Med.* 195, 85-98. doi:10.1084/jem.20011578

Martincorena, I. and Campbell, P. J. (2015). Somatic mutation in cancer and normal cells. *Science* 349, 1483-1489. doi:10.1126/science.aab4082

Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A. E., Kel-Margoulis, O. V. et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31, 374-378. doi:10.1093/nar/gkg108

Medina-Rivera, A., Santiago-Algarra, D., Puthier, D. and Spicuglia, S. (2018). Widespread enhancer activity from core promoters. *Trends Biochem. Sci.* 43, 452-468. doi:10.1016/j.tibs.2018.03.004

Navarro, J.-M., Touzart, A., Pradel, L. C., Loosveld, M., Koubi, M., Fenouil, R., Le Noir, S., Maqbool, M. A., Morgado, E., Gregoire, C. et al. (2015). Site- and allele-specific polycomb dysregulation in T-cell leukaemia. *Nat. Commun.* 6, 6094. doi:10.1038/ncomms7094

Orlando, G., Law, P. J., Cornish, A. J., Dobbins, S. E., Chubb, D., Broderick, P., Litchfield, K., Hariri, F., Pastinen, T., Osborne, C. S. et al. (2018). Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nat. Genet.* 50, 1375-1380. doi:10.1038/s41588-018-0211-z

Pan, Q., Shai, O., Lee, L. J., Frey, B. J. and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413-1415. doi:10.1038/ng.259

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., Potter, N. E., Heuser, M., Thol, F., Bolli, N. et al. (2016). Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* 374, 2209-2221. doi:10.1056/NEJMoa1516192

Pear, W. S., Aster, J. C., Scott, M. L., Hasserjian, R. P., Soffer, B., Sklar, J. and Baltimore, D. (1996). Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. *J. Exp. Med.* 183, 2283-2291. doi:10.1084/jem.183.5.2283

Pott, S. and Lieb, J. D. (2015). What are super-enhancers? *Nat. Genet.* 47, 8-12. doi:10.1038/ng.3167

Puente, X. S., Beà, S., Valdés-Mas, R., Villamor, N., Gutiérrez-Abril, J., Martín-Subero, J. I., Munar, M., Rubio-Pérez, C., Jares, P., Aymerich, M. et al. (2015). Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 526, 519-524. doi:10.1038/nature14666

Rahman, S., Magnussen, M., León, T. E., Farah, N., Li, Z., Abraham, B. J., Alapi, K. Z., Mitchell, R. J., Naughton, T., Fielding, A. K. et al. (2017). Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia. *Blood* 129, 3221-3226. doi:10.1182/blood-2016-09-742148

Rheinbay, E., Parasuraman, P., Grimsby, J., Tiao, G., Engreitz, J. M., Kim, J., Lawrence, M. S., Taylor-Weiner, A., Rodriguez-Cuevas, S., Rosenberg, M. et al. (2017). Recurrent and functional regulatory mutations in breast cancer. *Nature* 547, 55-60. doi:10.1038/nature22992

Robbiani, D. F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D. J., Chen, H. T., Corcoran, A. E., Nussenzweig, A. et al. (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* 135, 1028-1038. doi:10.1016/j.cell.2008.09.062

Schoenfelder, S. and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437-455. doi:10.1038/s41576-019-0128-0

Shi, J. and Vakoc, C. R. (2014). The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol. Cell* 54, 728-736. doi:10.1016/j.molcel.2014.05.016

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E. et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941-D947. doi:10.1093/nar/gky1015

Taylor, J., Xiao, W. and Abdel-Wahab, O. (2017). Diagnosis and classification of hematologic malignancies on the basis of genetics. *Blood* 130, 410-423. doi:10.1182/blood-2017-02-734541

Van Vlierberghe, P., van Grotel, M., Beverloo, H. B., Lee, C., Helgason, T., Buijs-Gladdines, J., Passier, M., van Wering, E. R., Veerman, A. J. P., Kamps, W. A. et al. (2006). The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. *Blood* 108, 3520-3529. doi:10.1182/blood-2006-04-019927

Van Vlierberghe, P., Beverloo, H. B., Buijs-Gladdines, J., van Wering, E. R., Horstmann, M., Pieters, R. and Meijerink, J. P. P. (2008). Monoallelic or biallelic LMO2 expression in relation to the LMO2 rearrangement status in pediatric T-cell acute lymphoblastic leukemia. *Leukemia* 22, 1434-1437. doi:10.1038/sj.leu.2405063

Vernimmen, D. and Bickmore, W. A. (2015). The hierarchy of transcriptional activation: from enhancer to promoter. *Trends Genet.* 31, 696-708. doi:10.1016/j.tig.2015.10.004

Vinagre, J., Almeida, A., Pópulo, H., Batista, R., Lyra, J., Pinto, V., Coelho, R., Celestino, R., Prazeres, H., Lima, L. et al. (2013). Frequency of TERT promoter mutations in human cancers. *Nat. Commun.* 4, 2185. doi:10.1038/ncomms3185

Weinhold, N., Jacobsen, A., Schultz, N., Sander, C. and Lee, W. (2014). Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat. Genet.* 46, 1160-1165. doi:10.1038/ng.3101

Weng, A. P., Ferrando, A. A., Lee, W., Morris, J. P., Silverman, L. B., Sanchez-Irizarry, C., Blacklow, S. C., Look, A. T. and Aster, J. C. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* 306, 269-271. doi:10.1126/science.1102160

Winter, G. E., Mayer, A., Buckley, D. L., Erb, M. A., Roderick, J. E., Vittori, S., Reyes, J. M., di Iulio, J., Souza, A., Ott, C. J. et al. (2017). BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. *Mol. Cell* 67, 5-18.e19. doi:10.1016/j.molcel.2017.06.004

Yamazaki, H., Suzuki, M., Otsuki, A., Shimizu, R., Bresnick, E. H., Engel, J. D. and Yamamoto, M. (2014). A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell* 25, 415-427. doi:10.1016/j.ccr.2014.02.008

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F. et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356, eaaj2239. doi:10.1126/science.aaj2239

Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S. L., Payne-Turner, D., Easton, J., Chen, X., Wang, J., Rusch, M. et al. (2012). The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 481, 157-163. doi:10.1038/nature10725

Zor, T., De Guzman, R. N., Dyson, H. J. and Wright, P. E. (2004). Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.* 337, 521-534. doi:10.1016/j.jmb.2004.01.038