bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Epigenomic translocation of H3K4me3 broad domains following super-enhancer hijacking

Authors

Aneta Mikulasova[1], Marco Trevisan-Herraz[1], Kent Fung[2], Cody Ashby[3], Nefeli Karataraki[2], Gareth J. Morgan[4], Anne Corcoran[5], Sophie Hambleton[2,6], Brian A. Walker[7], Daniel Rico[1,8,*], Lisa J. Russell[2,8,*]

Affiliations

[1]Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[2]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[3]Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR USA

[4]NYU Langone Medical Center, Perlmutter Cancer Center, NYU Langone Health, New York City, NY, USA

[5]Babraham Institute, Cambridge, United Kingdom

[6]Great North Children's Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, UK

[7]Division of Hematology Oncology, Indiana University, Indianapolis, IN, USA

[8]Co-senior authors

[*]Correspondence should be addressed to lisa.russell@newcastle.ac.uk or daniel.rico@newcastle.ac.uk

Key words

epigenome, translocation, H3K4me3 broad domain, immunoglobulin, enhancer, oncogene activation

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Chromosomal translocations are important drivers of haematological malignancies whereby proto-oncogenes are activated by juxtaposition with super-enhancers, often called *super-enhancer hijacking*. We analysed the epigenomic consequences of rearrangements between the enhancers of the *immunoglobulin heavy chain locus* (*IGH)* and proto-oncogene *CCND1* that are common in B-cell malignancies. By integrating BLUEPRINT epigenomic data with DNA breakpoint detection, we characterised the normal chromatin landscape of the human *IGH* locus and its dynamics after pathological genomic rearrangement. We detected an H3K4me3 broad domain (BD) within the *IGH* locus of healthy B-cells that was absent in samples with *IGH-CCND1* translocations. The appearance of H3K4me3-BD over *CCND1* in the latter was associated with overexpression and extensive chromatin accessibility of this locus. We observed similar cancer-specific H3K4me3-BDs associated with super-enhancer hijacking of other common oncogenes in B-cell (*MAF*, *MYC* and *FGFR3*) and in T-cell malignancies (*LMO2, TLX3* and *TAL1*). Our analysis suggests that H3K4me3-BDs are created by super-enhancers and supports the new concept of *epigenomic translocation*, where the relocation of H3K4me3-BDs accompanies the translocation of super-enhancers.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

In healthy cells, many proto-oncogenes and tumour-suppressor genes coordinate to control cell proliferation. However, these essential genes can be activated or inactivated by different genomic alterations to their coding sequence, including nucleotide substitutions, gene amplification or loss, and gene fusion. Proto-oncogenes can also be converted to oncogenes without alterations to the protein-coding sequence; structural alterations can result in juxtaposition of proto-oncogenes and super-enhancers promoting their over-expression, a situation that is referred to as *enhancer adoption* (Lettice et al. 2011) or, more often, *enhancer hijacking* (Northcott et al. 2014; Beroukhim, Zhang, and Meyerson 2016). Other structural alterations that disrupt local chromatin architecture can produce epigenomic activation of proto-oncogenes to resemble adjacent chromosomal neighbourhoods (Hnisz et al. 2016). However, little is known about the epigenomic consequences of enhancer hi-jacking. To begin to understand how structural alterations activate proto-oncogenes, we have analysed the changes in chromatin states associated with super-enhancer translocation events in cancer.

Tri-methylation of lysine 4 on histone H3 (H3K4me3) is a chromatin modification classically associated with the promoters of transcriptionally active genes (Howe et al. 2017) and also present at some active enhancers (Pekowska et al. 2011; Shen et al. 2016; Hu et al. 2017; Russ et al. 2017; Q.-L. Li et al. 2019). Although H3K4me3 marks tend to show sharp 1–2 kb peaks around promoters, some genes have broader regions of H3K4me3, also known as H3K4me3 broad domains (H3K4me3-BDs) that expand over part or all of the coding sequence of the gene (up to 20 kb) (Pekowska et al. 2010, 2011; K. Chen et al. 2015; Benayoun et al. 2014). These H3K4me3-BDs are associated with cell identity genes (Pekowska et al. 2010; Benayoun et al. 2014) and cell-specific tumour-suppressor genes (K. Chen et al. 2015), where they favour transcriptional consistency and increased expression (Benayoun et al. 2014). Interestingly, it has been observed that some tumour-suppressor genes show a narrower breadth of these domains in cancer, associated with their downregulation in malignant cells (K. Chen et al. 2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Super-enhancers were originally defined as enhancers with unusually high levels of certain transcriptional co-activators (Whyte et al. 2013; Lovén et al. 2013) and a median size larger than 8 kb (Pott and Lieb 2015). Recent studies of 3D genome networks suggest that the presence of H3K4me3-BDs in genes is associated with increased interactions with active super-enhancers (Cao et al. 2017; Thibodeau et al. 2017). More recently, Dhar and co-workers have shown that the histone methyl-transferase, MLL4, is essential for the methylation of both super-enhancers and H3K4me3-BDs, and proposed a hypothetical model whereby MLL4 would be essential to maintain the interaction of super-enhancers and tumour-suppressor genes with BDs (Dhar et al. 2018). However, the data from these previous studies could not address whether H3K4me3-BDs are formed as a consequence of super-enhancer activity.

The immunoglobulin (Ig) loci contain powerful super-enhancers to drive antibody formation and expression. The *immunoglobulin heavy chain locus* (*IGH*) has four regions encompassing: constant (C_H), joining (J_H), diversity (D_H) and variable (V_H) gene segments. The human C_H region encodes nine different Ig isotypes: IGHA2, IGHE, IGHG4, IGHG2, IGHA1, IGHG1, IGHG3, IGHD and IGHM (**Supplementary Figure 1**). Complex regulatory and genomic rearrangements in the *IGH* locus are required to ensure that immunoglobulin transcripts containing only one of each of these gene segments are expressed in each B-cell at the correct stage of B-cell differentiation. These natural processes predispose human Ig loci to inappropriate translocation events. As a result, these super-enhancer rich regions are commonly hijacked in B-cell malignancies providing the optimal model in which to address whether H3K4me3-BDs are formed as a consequence of super-enhancer activity.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

A common translocation involves the proto-oncogene *cyclin D1* (*CCND1*) at 11q13 and *IGH* at 14q32 (*IGH-CCND1*). To begin to understand how this translocation event leads to *CCND1* activation, we characterised the epigenomic landscape in these two regions from healthy and malignant B-cells. To precisely define the location and activity of the human *IGH* super-enhancers and promoters and the chromatin dynamics of the *CCND1* locus in B-cells, we used 103 chromatin state maps for lymphoid and myeloid cells built with over 700 ChIP-seq datasets from BLUEPRINT (Carrillo-de-Santa-Pau et al. 2017; Stunnenberg et al. 2016) (**Supplementary Table 1–3**). These data included samples from four stages of healthy B-cell differentiation, plasma cells and both primary tumour cells and cell lines derived from four different B-cell haematological malignancies. We used seven healthy non-B-cell cell types (T-cells, neutrophils, eosinophils, monocytes, macrophages, erythroblasts and megakaryocytes) as negative controls in which chromatin activity signals in the *IGH* locus are not predicted.

Epigenomic cartography of the human *IGH* locus

Three super-enhancers have been previously described in the non-variable region of *IGH*: Eα2, Eα1 and Eμ (Mills et al. 1983; C. Chen and Birshtein 1997; Mills et al. 1997). However, the precise definition, location and activity of these regulatory regions throughout different stages of healthy human B-cell differentiation and human B-cell derived malignancies have not been previously investigated.

The three known B-cell specific active enhancers (high signal of both histone marks H3K4me1 and H3K27ac) (~40 kb) were found within C_H region at the following genomic locations: chr14:106,025,200–106,056,800 (Eα2), chr14:106,144,200–106,179,400 (Eα1) and chr14:106,281,800–106,326,200 (Eμ). The Eμ enhancer showed dynamic changes associated with B-cell development stage. A larger region appears active in naïve B-cells before entering the germinal centre (GC), whereas in post-GC B-cells the region of activity is reduced (class-switched memory B-cells and plasma cells, **Figure 1**). A similar area of

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reduced activity was also observed in malignant samples derived from cells post- class-switch recombination, including diffuse large B-cell lymphoma (DLBCL), Burkitt lymphoma (BL) and multiple myeloma (MM) in comparison to pre-GC-like malignancies such as mantle cell lymphoma (MCL). We propose that a second, independent enhancer should be distinguished from Eμ based on the observation of a ~10 kb enhancer chromatin state gap between these two epigenomic regions that is supported by high quality mapping. We refer to this as Eδ (Eδ at chr14:106,281,800–106,289,800 and Eμ at chr14:106,299,800–106,326,200) (**Figure 1**). None of these four enhancer regions were detected in healthy myeloid and T-cells.

Epigenomic cartography of the human *CCND1* gene

The proto-oncogene *CCND1* is one of the most common *IGH* translocation partners in haematological malignancies, a hallmark of MCL (Vose 2017) and frequently observed in MM (Walker et al. 2013). The chromatin states in the *CCND1* promoter and immediate upstream region (**Figure 2**) suggest that the Polycomb repressive complex may be implicated in the regulation of this cell cycle gene in healthy human B-cells. When the promoter is inactive in non-proliferating GC B-cells and terminally differentiated plasma cells, the Polycomb state covers the whole promoter and upstream region (**Figure 2**). Interestingly, the Polycomb state is maintained upstream of the promoter in healthy proliferating haematopoietic cell types including, pre-GC and post-GC B-cells, some T-cells, macrophages and eosinophils (Polycomb short).

We observed a different chromatin landscape for *CCND1* across four haematological malignancy subtypes (**Figure 2**). In the majority of cases, the *CCND1* promoter shows an active state. Chronic lymphocytic leukaemia (CLL) samples show a similar pattern to healthy B-cells, with the Polycomb state domain upstream of the active promoter. DLBCL and BL samples show a narrow H3K4me3 active promoter but lack the upstream Polycomb domain. Intriguingly, all MCL and some MM samples have a larger active promoter/enhancer region extending upstream and downstream of the transcription start site, with most of the gene body containing H3K4me3 marks – a pattern that looks comparable to an H3K4me3-BD (Pekowska et al. 2010, 2011; Benayoun et al. 2014; K. Chen et al. 2015). Cytogenetic analysis

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for BLUEPRINT primary patient samples confirmed the presence of an *IGH-CCND1* rearrangement in all MCL patient samples (personal communication with J.I. Martin-Subero, provider of those samples for BLUEPRINT). Although we were unable to confirm this for the MM samples, it is reported that 20% of MM patients will have an *IGH-CCND1* rearrangement (Avet-Loiseau et al. 2007). We therefore hypothesised that it is the juxtaposition of an *IGH* super-enhancer close to the coding region of *CCND1* results in the presence of an H3K4me3-BD over the coding region of the gene.

Re-location of an H3K4me3-BD from *IGH* to *CCND1* as a consequence of super-enhancer hijacking

We discovered a B-cell specific H3K4me3-BD at genomic location, chr14:106,346,800-106,387,800 (41 kb), overlapping the D_H region (**Figure 1**). This element was characterised by high signal of both H3K4me3 and H3K27ac (typical for promoters), combined with occasional H3K4me1 high signal. Interestingly, this H3K4me3-BD is absent or significantly reduced in malignant B-cells. These data suggest that the disappearance of the H3K4me3-BD from the *IGH* locus may be the consequence of *IGH* super-enhancer hijacking via genomic translocation events that are known to be present in the malignancy subtypes represented.

To further investigate the presence or absence of H3K4me3-BD as a result of *IGH* translocations, we studied the MM cell line U266 in which the *IGH* super-enhancer, Eα1, is inserted next to the *CCND1* proto-oncogene (Gabrea et al. 1999). This cell line provides the opportunity to analyse the consequence of the relocation of an isolated *IGH* super-enhancer. Using paired-end read targeted DNA sequencing, we precisely mapped the chromosomal changes at the *IGH* locus (**Supplementary Figure 2**). We confirmed two chromosomal breakpoints that occur in *IGHE* and *IGHA* switch regions and result in the cut and paste of the *IGH* Eα1 super-enhancer into chromosome 11, approximately 12 kb upstream of *CCND1* (**Figure 3**). The vast majority of the H3K4me3-BD observed over the D_H region in healthy B-cells is absent from this translocated cell line. A small region of H3K4me3 is still observed over the D_H region, perhaps due to the presence of *IGH* Eα2 super-enhancer

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that enables the expression of the IgE monoclonal immunoglobulin produced by U266 cells (Hellman et al. 1988). Importantly, an aberrant H3K4me3-BD is observed covering most of the *CCND1* gene body (**Figure 3**). This observation suggests the presence of an *epigenomic translocation* of the H3K4me3-BD from *IGH* to the *CCND1* locus, present exclusively in MCL and MM samples (**Figure 1–2**). Interestingly, this epigenomic translocation can result from both errors in VDJ recombination (Stamatopoulos et al. 1999) as observed in MCL, and class-switch recombination (Walker et al. 2013) as observed in MM.

Relocation of H3K4me3-BD is associated with increased chromatin accessibility and transcription

To understand the effect of the epigenomic translocation on chromatin accessibility we used BLUEPRINT DNaseI hypersensitivity data available for five cell lines derived from B-cell malignancies (**Supplementary Table 3**). Two of the cell lines had an *IGH-CCND1* rearrangement including U266 (already described above) and MCL cell line, Z-138. We observed the H3K4me3-BD and strong DNaseI hypersensitivity signals encompassing the entire coding region of the *CCND1* gene in both cell lines (**Figure 3**). The increased chromatin accessibility extended beyond the coding region of *CCND1*. This pattern is associated with an increase in *CCND1* expression in U266 and Z-138 (**Figure 3**). The presence of the H3K4me3-BD, the strong DNaseI hypersensitivity signal and increased transcript levels for *CCND1* were not observed in the remaining three cell lines without the *IGH-CCND1* rearrangement (**Figure 3**).

Interestingly, an aberrant H3K4me3-BD, increased chromatin accessibility and expression of the adjacent gene *MYEOV* was observed in the U266 cell line (**Supplementary Figure 3**). This suggests that the inserted Eα1 super-enhancer has a bi-directional effect, generating an H3K4me3-BD over another adjacent gene. This was not observed for Z-138, since the reciprocal nature of this translocation results in *MYEOV* being retained on chromosome 11, with the relocation of *CCND1* to chromosome 14 juxtaposing it to all three *IGH* super-enhancers (Guikema et al. 2005).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cancer-specific H3K4me3-BDs are associated with immunoglobulin translocations in multiple myeloma

To investigate whether the epigenomic translocation of the *IGH* H3K4me3-BD can occur over additional proto-oncogenes in B-cell malignancies, we generated ChIP-seq data for the same six histone marks for an additional two MM cell lines (KMS11 and MM1S) and seven MM patients whose diagnostic samples were engrafted in murine models (patient-derived xenografts, or PDXs). Each cell line had a complex rearrangement identified by targeted sequencing of the *IGH* locus involving multiple proto-oncogenes (**Supplementary Table 3**). 10X genome sequencing detected the second most common *IGH* translocation in MM, t(14;16)(q32;q23), involving the transcription factor *MAF* in two patients (P1, P2). A further two patients (P3, P4) had *IGH* translocations t(4;14)(p16;q32) involving *FGFR3* and *NSD2* with the remaining three patients (P5–7) having no detectable translocation involving any of the Ig loci (**Supplementary Table 4**). No cell lines or patients had an *IGH-CCND1* rearrangement detected by targeted or genome sequencing therefore, as expected, we did not observe an aberrant H3K4me3-BD or an increase in mRNA transcript levels for *CCND1* in these samples (**Figure 4a**). This suggests that epigenomic translocations of H3K4me3-BDs over *CCND1* are a direct result of a specific genomic translocation event.

In the two cell lines and four of the patient samples (P1–4) we observed an H3K4me3-BD over the coding region of *MAF* that was coupled with higher *MAF* expression observed by RNA-seq (**Figure 4b**). This was associated with the presence of the *IGH-MAF* rearrangement in the two cell lines and patients P1 and P2. Our targeted sequencing identified the involvement of *MYC* as well in the complex translocations observed in KMS11 and MM1S cell lines (**Supplementary Table 3**). Intriguingly, we observed a Polycomb chromatin state over *MYC*, but the analysis of each histone modification separately identified an H3K4me3-BD over *MYC* in the two cell lines (**Supplementary Figures 4a** and **5a**), possibly reflecting repression of one allele of *MYC* and activation of the other one as a result of the *IGH* enhancer hijacking (Affer et al. 2014).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In the remaining two patients (P3, P4), we were unable to detect a genomic rearrangement of a super-enhancer in close proximity *to MAF* however, we see overexpression of *MAF* in association with an H3K4me3-BD (**Figure 4b**). These two patients have the translocation t(4;14)(p16;q32). Interestingly, we observed an H3K4me3-BD and increased transcript levels for *FGFR3* and *NSD2,* genes that are involved in this rearrangement event (**Supplementary Figure 4b and 5b**). Interestingly, a Polycomb chromatin state without H3K4me3-BD was observed over *MYC* coupled with low *MYC* expression in both patients.

In summary, we have observed the appearance of H3K4me3-BDs over a variety of proto-oncogenes when they are involved in *IGH* super-enhancer hijacking. Whilst the chromatin state model (Carrillo-de-Santa-Pau et al. 2017) identified BDs over *CCND1* and *MAF*, it preferentially selected a Polycomb chromatin state for both *MYC* and *FGFR3* in some samples. The presence of both H3K27me3 and H3K4me3 domains over these genes potentially highlights the repression of the second allele (Nag et al. 2015), a mechanism the cell can use to control the level of overexpression. With the H3K4me3-BD signature observed over *MYC* in both healthy and malignant B-cells, future work will be required to unravel the epigenetic impact upon this locus after different super-enhancer hijacking scenarios.

Super-enhancer-driven H3K4me3-BDs as a wider phenomenon associated with oncogene deregulation in haematological malignancies

T-cell receptor loci (*TRA, TRD, TRB, TRG*) are also regulated by super-enhancers, undergo similar inherent somatic rearrangement events and are involved in translocation events in T-cell acute lymphoblastic leukaemia (T-ALL) (Larmonie et al. 2013). To investigate whether these loci also have the ability to generate H3K4me3-BDs over proto-oncogenes involved in translocation events, we studied the KOPT-K1 T-ALL cell line that results in the juxtaposition of *LMO2* and the super-enhancer of the *TRA/D* locus as a result of the translocation, t(11;14)(q13;q11). The Polycomb chromatin state was observed in healthy CD4 and CD8 expressing T-cells (**Figure 5a**). We indeed observed an aberrant H3K4me3-BD

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

over the coding region of *LMO2* in KOPT-K1 suggesting that T-cell receptor super-enhancers can also generate H3K4me3-BD that results in proto-oncogene activation and overexpression (**Figure 5a**) (Sanda et al. 2012). An H3K4me3-BD was not observed over two additional cell lines that do not have rearrangements of the *LMO2* locus.

To investigate the wider involvement of H3K4m3-BDs in T-ALL we considered the T-cell identity (Ha et al. 2017) and tumour suppressor gene, *BCL11B* that is rearranged in T-ALL (Liu et al. 2017). Since this gene is known to be regulated by an upstream super-enhancer element (Nagel et al. 2007; L. Li et al. 2013), we questioned whether the relocation of this region from chromosome 14 band q32 next to *TLX3* on chromosome 5 band q35 would result in the appearance of an aberrant H3K4me3-BD over the coding region of the gene. Using publicly available ChIP-seq data for T-ALL cell line DND-41 (Knoechel et al. 2014) we observed high signal for H3K27ac, H3K4me3 and H3K4me1 encompassing the entire coding region of *TLX3* and non-coding regions either side (**Figure 5b**). This correlated with a high transcript level of when compared to healthy cells (Petryszak et al. 2016). We did not observe the presence of these histone marks in normal healthy CD4 and CD8 T-cells and cell lines without rearrangements of these loci (**Figure 5b**).

Finally we wanted to assess whether *de novo* super-enhancers, generated by somatic nucleotide insertions, could also generate aberrant H3K4me3-BD. An aberrant H3K4me3-BD was observed across the *TAL1* gene in the T-ALL cell line Jurkat (**Figure 5c**) where a 12 bp insertion generates a super-enhancer element upstream in the gene (Mansour et al. 2014) .

In summary we found that the appearance of H3K4me3-BDs is associated with cancer-specific super-enhancer activation in both B-cell and T-cell derived haematological malignancies. Our data suggest that H3K4me3-BDs could be a signature of super-enhancer targets in general and a useful marker to identify deregulated genes affected by super-enhancer hijacking in cancer.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

We aimed to understand how the juxtaposition of super-enhancers and proto-oncogenes results in exceptionally high, but restricted levels of oncogene expression. For this, we first dissected the epigenomic landscape of two loci that are commonly translocated in haematological malignancies, *CCND1* at 11q13 and the *IGH* locus at 14q32. We have shown that genomic relocation of the *IGH* E1 super-enhancer alters the location of an H3K4me3-BD in rearranged malignant B-cells. We also observed H3K4me3-BDs over additional proto-oncogenes rearranged in B-cell and T-cell malignancies including the involvement of gene specific and de novo super-enhancers. H3K4me3-BDs are associated with high levels of stable transcription of cell identity and tumour suppressor genes (Pekowska et al. 2010, 2011; Benayoun et al. 2014; K. Chen et al. 2015) and we show here, for the first time that they can be a feature of oncogene activation by enhancer hijacking (as a result of a genomic rearrangement).

Based on these data, we present a model of *epigenomic translocation*, a consequence of super-enhancer hijacking. In this model a super-enhancer driven wild-type H3K4me3-BD "re-locates" to a target proto-oncogene as the result of a genomic rearrangement leading to oncogene activation (**Figure 6**).

Our model implies that H3K4me3-BDs are generated by super-enhancers in their wildtype neighbourhood. Interestingly, the H3K4me3 writer MLL4 is essential for the maintenance of both H3K4me3-BDs and super-enhancers, but not for narrow-peak promoters and standard enhancers (Dhar et al. 2018). Initially, a strong emphasis was placed on the distinction between H3K4me3-BDs and super-enhancers (Benayoun et al. 2014) even if they were later shown to be close or sometimes overlapping in the linear genome (Cao et al. 2017) and in 3D (Thibodeau et al. 2017). Whilst super-enhancers may be responsible for generating broad domains very few studies have addressed the length and maintenance of these regions.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Recent simulations and experimental data suggest that very active enhancers may induce high levels of gene expression by global chromatin decompaction that may actually increase the distance between promoter and enhancer regions (Buckle et al. 2018). We have observed that the presence of the H3K4me3-BD over *CCND1* coincides with high chromatin accessibility of the whole gene body. If super-enhancers induce unusual chromatin decompaction and access to DNA, this could explain how the target genes of oncogenic translocations can show such high levels of gene expression. The effects of the super-enhancer could be restricted within the boundaries of the topologically associated domains (TADs) where the hijacked oncogene and super-enhancer reside. Indeed, it has been suggested that one reason why TAD boundaries are conserved across species (Lazar et al. 2018; Krefting, Andrade-Navarro, and Ibn-Salem 2018) could be that they prevent undesirable enhancer hijacking (Fudenberg and Pollard 2019).

The epigenomic translocation scenario is a simple yet powerful conceptual model that extends the original enhancer adoption/hijacking model (Northcott et al. 2014; Beroukhim, Zhang, and Meyerson 2016; Lettice et al. 2011) and allows inference to translocation events without the analysis of structural alterations in the genome. Although we propose that broad H3K4me3 domains are the consequence of active super-enhancers, future studies will be needed to elucidate the possible cooperation between genetic features, other epigenetic marks, the broad domain and the super-enhancer chromatin structures and how they maintain each other.

This work will hopefully pave the way for new therapeutic approaches based on chromatin remodelling to revert the local epigenome of super-enhancer activated proto-oncogenes, returning their expression back to wild-type levels. We anticipate the model proposed here will focus attention on the regulatory effects of different genomic rearrangements including translocations, identifying key oncogenes in each patient, and open exciting new avenues for novel diagnostic and therapeutic approaches.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

ChIP-seq, DNaseI hypersensitivity and RNA-seq of BLUEPRINT dataset

Previously processed ChIP-seq chromatin state BLUEPRINT samples (n=103) were used (**Supplementary Table 2–3**). This dataset comprised 80 samples from healthy donors (15 B-cell lineage, 15 T-cell lineage and 50 myeloid lineage) and 23 samples from patients (16 primary and seven cell lines) with different B-cell haematological malignancies. Five of these cell lines had DNAseI hypersensitivity and RNA-seq data available (**Supplementary Table 3**). Detailed laboratory methods and data processing of BLUEPRINT experiments are described elsewhere (Carrillo-de-Santa-Pau et al. 2017), (Fernández et al. 2016; Stunnenberg et al. 2016). Briefly, sequencing reads were mapped using BWA (ChIP-seq and DNase-seq, v0.5.9) or GEM mapper (Marco-Sola et al. 2012) (RNA-seq) to GRCh37 human genome assembly. ChromHMM (Ernst and Kellis 2012) (v1.10) was used to determine chromatin states, based on a combination of six histone modifications as follows: H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K27Ac and H3K36me3 (Carrillo-de-Santa-Pau et al. 2017). Selected chromatin states used in this study are described in **Supplementary Table 1.** Data used in this study were downloaded from ftp://ftp.ebi.ac.uk/pub/databases/blueprint/paper_data_sets/chromatin_states_carrillo_build37 and ftp://ftp.ebi.ac.uk/pub/databases/blueprint/data/homo_sapiens/GRCh37.

Epigenomic consensus of the *IGH* and *CCND1* loci determined by BLUEPRINT ChIP-seq data

Epigenomic annotation of *IGH* (super-enhancers, promoters and H3K4me3-BD) and *CCND1* (promoter and polycomb) loci was built using 15 BLUEPRINT B-cells/plasma cell samples derived from healthy donors (**Supplementary Table 2**). MCL (JVM-2 and Z-138) and MM (U266) cell lines with known *IGH-CCND1* rearrangements (**Supplementary Table 3**) were used to determine the relocated *CCND1* H3K4me3-BD.

Segmentation files of relevant samples were filtered for the subset of chromatin states of interest as follows: active canonical enhancer (state 9) for super-enhancers, repressed Polycomb regulatory region (state 7) for Polycomb, and a combination of active promoter (state 11) and active non-canonical H3K4me3 enhancer (state 10) for promoters

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and H3K4me3-BDs. For each of these epigenomic elements, subsets were intersected and merged (both bedtools v2.27.1) into continuous regions with skipping gaps smaller than 1 kb. *IGH* super-enhancers were called as regions bigger than 2.5 kb overlapping the C_H and *IGH* intronic region. Promoters were determined as a segment bigger than 2.5 kb within the expected promoter positions of *IGH* (C_H and intronic region) and *CCND1* genes. H3K4me3-BDs were scanned at *IGH* C_H, J_H and D_H and *CCND1* loci, with a minimum size of 15 kb. Polycomb at the *CCND1* locus was obtained as regions upstream and overlapping with the promoter.

Targeted DNA sequencing of myeloma cell lines

DNA of myeloma cell lines U266, KMS11 and MM1S (**Supplementary Table 3**) were sequenced using targeted next generation sequencing covering the following genomic regions: 1) 4.2 Mb; extended coverage for chromosomal abnormality detection within *IGH*, *IGK*, *IGL* and *MYC* (4.2 Mb); and 2) 0.6 Mb, exonic regions of 127 myeloma specific genes for mutation analysis and 27 additional regions for efficient data normalisation. GRCh37 human genome assembly was used for sequence mapping using BWA-MEM (v0.7.12). Chromosomal rearrangements were called using Manta v.0.29.6 (Mikulasova et al. 2019; X. Chen et al. 2016). Detailed methods are described previously (Mikulasova et al. 2019).

Myeloma patient derived xenografts

Patient derived xenografts were generated by passaging primary patient CD138+ selected cells through the previously described SCID-rab myeloma mouse model (Yata and Yaccoby 2004; Mikulasova et al. 2019). Detailed methods are described previously (Mikulasova et al. 2019). Seven patient derived xenografts were used in this study (**Supplementary Table 4**).

Whole genome sequencing of patient derived xenograft samples

DNA from seven patient derived xenograft samples were sequenced using phased whole genome sequencing (10X Genomics, Pleasanton, CA, USA) at Hudson Alpha (Huntsville, AL, USA). Long Ranger (10X Genomics) pipelines were used for data processing,

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

including alignment to GRCh38 genome assembly and structural variant calling. Germline controls were used to distinguish somatic abnormalities and chromosomal breakpoints of detected translocations (**Supplementary Table 4**) were manually inspected.

ChIP-seq of myeloma cell lines and patient derived xenograft samples

ChIP-seq was performed as previously reported (Mikulasova et al. 2019), for the myeloma cell lines KMS11 and MM1S as well as seven myeloma patient derived xenograft samples. ChIP-seq for the histone marks H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K27Ac, and H3K36me3 (Active Motif, Carlsbad, CA, USA) were included in this study. Controls without antibody input were performed to ensure data quality. GRCh38 human assembly was used for alignment.

ChIP-seq chromatin states were determined by ChromHMM (v1.20), the model used in the BLUEPRINT dataset (Carrillo-de-Santa-Pau et al. 2017). Additionally, the peaks of individual histone marks were detected using MACS2 (v2.2.5) by a pipeline available at BLUEPRINT Data Coordination Center Portal (Fernández et al. 2016). For direct comparison of this data and BLUEPRINT data, The liftOver (University of California, Santa Cruz, CA, USA) tool was used to convert chromatin state segments and histone peaks within the genomic regions included in this study from GRCh38 to GRCh37 genome assembly. No segment was lost during this conversion.

RNA-seq of patient derived xenograft samples

RNA-seq was performed using 100 ng total RNA with genomic DNA removal using the TURBO DNA-free kit (Ambion, Austin, TX, USA). RNA was prepared using the TruSeq stranded total RNA Ribo-zero gold kit (Illumina, San Diego, CA, USA) and libraries were sequenced using 75 bp paired end reads on a NextSeq500 (Illumina). RNA-seq data was analysed as previously reported (Mikulasova et al. 2019). Briefly, raw data were aligned to the human genome assembly GRCh38 with gene transcript quantification being processed by Star (v2.5.1b) and Salmon (v0.6.0) algorithms. Read counts per gene were read into R and using the DESeq2 (v1.20.0) R library, normalised across samples and the log2 expression calculated.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ChIP-seq of T-cell acute lymphoblastic leukemia cell lines

Publicly available ChIP-seq data were used for histone modification analysis of three T-ALL cell lines KOPT-K1, DND-41 and Jurkat (**Supplementary Table 3**). ChiP-seq data are available at Gene Expression Omnibus (National Center for Biotechnology Information (NCBI), Bethesda, MD, USA) under the accession numbers GSE54379 (KOPT-K1 and DND-41) and GSE65687 (Jurkat). Raw files were downloaded from NCBI Sequence Read Archive and converted to fastq files using NCBI fasterq-dump (SRA Toolkit v2.9.6-1). Appropriate sequencing runs were merged and sequences were mapped to GRCh37 genome assembly using BWA (v0.7.17), followed by processing to MACS2 (v2.2.5) peaks, using a pipeline available at BLUEPRINT Data Coordination Centre Portal (http://dcc.blueprint-epigenome.eu/). Histone marks H3K4me1, H3K4me3, H3K27me3, H3K27Ac, and H3K36me3 in all three cell lines were used in this study, with the exception of H3K27me3 in Jurkat as this was not available.

Data availability

All publicly available datasets that were used are referred to in the sections above. The new data generated will be available in ENA and EGA repositories (data transfer in progress, accession numbers will be available in the next version of this manuscript).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Affer, Maurizio, Marta Chesi, Wei-Dong G. Chen, Jonathan J. Keats, Yulia N. Demchenko, Anna V. Roschke, Scott Van Wier, Rafael Fonseca, P. Leif Bergsagel, and W. Michael Kuehl. 2014. "Promiscuous MYC Locus Rearrangements Hijack Enhancers but Mostly Super-Enhancers to Dysregulate MYC Expression in Multiple Myeloma." *Leukemia* 28 (8): 1725–35.

Avet-Loiseau, H., M. Attal, P. Moreau, C. Charbonnel, F. Garban, C. Hulin, S. Leyvraz, et al. 2007. "Genetic Abnormalities and Survival in Multiple Myeloma: The Experience of the Intergroupe Francophone Du Myelome." *Blood*. https://doi.org/10.1182/blood-2006-08-040410.

Benayoun, Bérénice A., Elizabeth A. Pollina, Duygu Ucar, Salah Mahmoudi, Kalpana Karra, Edith D. Wong, Keerthana Devarajan, et al. 2014. "H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency." *Cell* 163 (5): 1281–86.

Beroukhim, Rameen, Xiaoyang Zhang, and Matthew Meyerson. 2016. "Copy Number Alterations Unmasked as Enhancer Hijackers." *Nature Genetics*.

Buckle, Adam, Chris A. Brackley, Shelagh Boyle, Davide Marenduzzo, and Nick Gilbert. 2018. "Polymer Simulations of Heteromorphic Chromatin Predict the 3D Folding of Complex Genomic Loci." *Molecular Cell* 72 (4): 786–97.e11.

Cao, Fan, Yiwen Fang, Hong Kee Tan, Yufen Goh, Jocelyn Yeen Hui Choy, Bryan Thean Howe Koh, Jiong Hao Tan, et al. 2017. "Super-Enhancers and Broad H3K4me3 Domains Form Complex Gene Regulatory Circuits Involving Chromatin Interactions." *Scientific Reports* 7 (1): 2186.

Carrillo-de-Santa-Pau, Enrique, David Juan, Vera Pancaldi, Felipe Were, Ignacio Martin-Subero, Daniel Rico, Alfonso Valencia, and on behalf of The BLUEPRINT Consortium. 2017. "Automatic Identification of Informative Regions with Epigenomic Changes Associated to Hematopoiesis." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx618.

Chen, C., and B. K. Birshtein. 1997. "Virtually Identical Enhancers Containing a Segment of Homology to Murine 3'IgH-E(hs1,2) Lie Downstream of Human Ig C Alpha 1 and C Alpha 2 Genes." *Journal of Immunology* 159 (3): 1310–18.

Chen, Kaifu, Zhong Chen, Dayong Wu, Lili Zhang, Xueqiu Lin, Jianzhong Su, Benjamin Rodriguez, et al. 2015. "Broad H3K4me3 Is Associated with Increased Transcription Elongation and Enhancer Activity at Tumor-Suppressor Genes." *Nature Genetics*. https://doi.org/10.1038/ng.3385.

Chen, Xiaoyu, Ole Schulz-Trieglaff, Richard Shaw, Bret Barnes, Felix Schlesinger, Morten Källberg, Anthony J. Cox, Semyon Kruglyak, and Christopher T. Saunders. 2016. "Manta: Rapid Detection of Structural Variants and Indels for Germline and Cancer Sequencing Applications." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv710.

Dhar, Shilpa S., Dongyu Zhao, Tao Lin, Bingnan Gu, Khusboo Pal, Sarah J. Wu, Hunain Alam, et al. 2018. "MLL4 Is Required to Maintain Broad H3K4me3 Peaks and Super-Enhancers at Tumor Suppressor Genes." *Molecular Cell*. https://doi.org/10.1016/j.molcel.2018.04.028.

Ernst, Jason, and Manolis Kellis. 2012. "ChromHMM: Automating Chromatin-State Discovery and Characterization." *Nature Methods*. https://doi.org/10.1038/nmeth.1906.

Fernández, José María, Victor de la Torre, David Richardson, Romina Royo, Montserrat Puiggròs, Valentí Moncunill, Stamatina Fragkogianni, et al. 2016. "The BLUEPRINT Data Analysis Portal." *Cell Systems*. https://doi.org/10.1016/j.cels.2016.10.021.

Fudenberg, Geoff, and Katherine S. Pollard. 2019. "Chromatin Features Constrain Structural Variation across Evolutionary Timescales." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1808631116.

Gabrea, Ana, P. Leif Bergsagel, Marta Chesi, Yaping Shou, and W. Michael Kuehl. 1999. "Insertion of Excised IgH Switch Sequences Causes Overexpression of Cyclin D1 in a Myeloma Tumor Cell." *Molecular Cell*. https://doi.org/10.1016/s1097-2765(00)80180-x.

Guikema, Jeroen E. J., James A. L. Fenton, Conny de Boer, Karin Kleiverda, Antoinette A. T. P. Brink, Anton K. Raap, Zeev Estrov, Ed Schuuring, and Philip M. Kluin. 2005. "Complex Biallelic IGH Rearrangements in IgM-Expressing Z-138 Cell Line: Involvement of Downstream

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Immunoglobulin Class Switch Recombination." *Genes, Chromosomes & Cancer* 42 (2): 164–69.

Ha, V. L., A. Luong, F. Li, D. Casero, J. Malvar, Y. M. Kim, R. Bhatia, G. M. Crooks, and C. Parekh. 2017. "The T-ALL Related Gene BCL11B Regulates the Initial Stages of Human T-Cell Differentiation." *Leukemia* 31 (11): 2503–14.

Hellman, Lars, Staffan Josephson, Helena Jernberg, Kenneth Nilsson, and Uif Pettersson. 1988. "Immunoglobulin Synthesis in the Human Myeloma Cell Line U-266; Expression of Two Immunoglobulin Heavy Chain Isotypes (ε and α) after Long-Term Cultivationin Vitro." *European Journal of Immunology*. https://doi.org/10.1002/eji.1830180611.

Hnisz, D., A. S. Weintraub, D. S. Day, A. -L. Valton, R. O. Bak, C. H. Li, J. Goldmann, et al. 2016. "Activation of Proto-Oncogenes by Disruption of Chromosome Neighborhoods." *Science*. https://doi.org/10.1126/science.aad9024.

Howe, Françoise S., Harry Fischl, Struan C. Murray, and Jane Mellor. 2017. "Is H3K4me3 Instructive for Transcription Activation?" *BioEssays*. https://doi.org/10.1002/bies.201600095.

Hu, Deqing, Xin Gao, Kaixiang Cao, Marc A. Morgan, Gloria Mas, Edwin R. Smith, Andrew G. Volk, et al. 2017. "Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification." *Molecular Cell* 65 (3): 460–75.e6.

Knoechel, Birgit, Justine E. Roderick, Kaylyn E. Williamson, Jiang Zhu, Jens G. Lohr, Matthew J. Cotton, Shawn M. Gillespie, et al. 2014. "An Epigenetic Mechanism of Resistance to Targeted Therapy in T Cell Acute Lymphoblastic Leukemia." *Nature Genetics* 46 (4): 364–70.

Krefting, Jan, Miguel A. Andrade-Navarro, and Jonas Ibn-Salem. 2018. "Evolutionary Stability of Topologically Associating Domains Is Associated with Conserved Gene Regulation." *BMC Biology*. https://doi.org/10.1186/s12915-018-0556-x.

Larmonie, Nicole S. D., Willem A. Dik, Jules P. P. Meijerink, Irene Homminga, Jacques J. M. van Dongen, and Anton W. Langerak. 2013. "Breakpoint Sites Disclose the Role of the V(D)J Recombination Machinery in the Formation of T-Cell Receptor (TCR) and Non-TCR Associated Aberrations in T-Cell Acute Lymphoblastic Leukemia." *Haematologica* 98 (8): 1173–84.

Lazar, Nathan H., Kimberly A. Nevonen, Brendan O'Connell, Christine McCann, Rachel J. O'Neill, Richard E. Green, Thomas J. Meyer, Mariam Okhovat, and Lucia Carbone. 2018. "Epigenetic Maintenance of Topological Domains in the Highly Rearranged Gibbon Genome." *Genome Research* 28 (7): 983–97.

Lettice, Laura A., Sarah Daniels, Elizabeth Sweeney, Shanmugasundaram Venkataraman, Paul S. Devenney, Philippe Gautier, Harris Morrison, Judy Fantes, Robert E. Hill, and David R. FitzPatrick. 2011. "Enhancer-Adoption as a Mechanism of Human Developmental Disease." *Human Mutation* 32 (12): 1492–99.

Li, Long, Jingli A. Zhang, Marei Dose, Hao Yuan Kueh, Ruzbeh Mosadeghi, Fotini Gounari, and Ellen V. Rothenberg. 2013. "A Far Downstream Enhancer for Murine Bcl11b Controls Its T-Cell Specific Expression." *Blood* 122 (6): 902–11.

Li, Qing-Lan, Dan-Ya Wang, Lin-Gao Ju, Jie Yao, Chuan Gao, Pin-Ji Lei, Lian-Yun Li, Xiao-Lu Zhao, and Min Wu. 2019. "The Hyper-Activation of Transcriptional Enhancers in Breast Cancer." *Clinical Epigenetics* 11 (1): 48.

Liu, Yu, John Easton, Ying Shao, Jamie Maciaszek, Zhaoming Wang, Mark R. Wilkinson, Kelly McCastlain, et al. 2017. "The Genomic Landscape of Pediatric and Young Adult T-Lineage Acute Lymphoblastic Leukemia." *Nature Genetics* 49 (8): 1211–18.

Lovén, Jakob, Heather A. Hoke, Charles Y. Lin, Ashley Lau, David A. Orlando, Christopher R. Vakoc, James E. Bradner, Tong Ihn Lee, and Richard A. Young. 2013. "Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers." *Cell* 153 (2): 320–34.

Mansour, Marc R., Brian J. Abraham, Lars Anders, Alla Berezovskaya, Alejandro Gutierrez, Adam D. Durbin, Julia Etchin, et al. 2014. "Oncogene Regulation. An Oncogenic Super-Enhancer Formed through Somatic Mutation of a Noncoding Intergenic Element." *Science* 346 (6215): 1373–77.

Marco-Sola, Santiago, Michael Sammeth, Roderic Guigó, and Paolo Ribeca. 2012. "The GEM Mapper: Fast, Accurate and Versatile Alignment by Filtration." *Nature Methods* 9 (12): 1185–88.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mikulasova, Aneta, Cody Ashby, Ruslana G. Tytarenko, Pingping Qu, Adam Rosenthal, Judith A. Dent, Katie R. Ryan, et al. 2019. "Microhomology-Mediated End Joining Drives Complex Rearrangements and over Expression of MYC and PVT1 in Multiple Myeloma." *Haematologica*. https://doi.org/10.3324/haematol.2019.217927.

Mills, F. C., L. M. Fisher, R. Kuroda, A. M. Ford, and H. J. Gould. 1983. "DNase I Hypersensitive Sites in the Chromatin of Human Mu Immunoglobulin Heavy-Chain Genes." *Nature* 306 (5945): 809–12.

Mills, F. C., N. Harindranath, M. Mitchell, and E. E. Max. 1997. "Enhancer Complexes Located Downstream of Both Human Immunoglobulin Calpha Genes." *The Journal of Experimental Medicine* 186 (6): 845–58.

Nag, Anwesha, Sébastien Vigneau, Virginia Savova, Lillian M. Zwemer, and Alexander A. Gimelbrant. 2015. "Chromatin Signature Identifies Monoallelic Gene Expression Across Mammalian Cell Types." *G3* 5 (8): 1713–20.

Nagel, Stefan, Michaela Scherr, Alexander Kel, Klaus Hornischer, Gregory E. Crawford, Maren Kaufmann, Corinna Meyer, Hans G. Drexler, and Roderick A. F. MacLeod. 2007. "Activation of TLX3 and NKX2-5 in t(5;14)(q35;q32) T-Cell Acute Lymphoblastic Leukemia by Remote 3'-BCL11B Enhancers and Coregulation by PU.1 and HMGA1." *Cancer Research* 67 (4): 1461–71.

Northcott, Paul A., Catherine Lee, Thomas Zichner, Adrian M. Stütz, Serap Erkek, Daisuke Kawauchi, David J. H. Shih, et al. 2014. "Enhancer Hijacking Activates GFI1 Family Oncogenes in Medulloblastoma." *Nature* 511 (7510): 428–34.

Pekowska, Aleksandra, Touati Benoukraf, Pierre Ferrier, and Salvatore Spicuglia. 2010. "A Unique H3K4me2 Profile Marks Tissue-Specific Gene Regulation." *Genome Research* 20 (11): 1493–1502.

Pekowska, Aleksandra, Touati Benoukraf, Joaquin Zacarias-Cabeza, Mohamed Belhocine, Frederic Koch, Hélène Holota, Jean Imbert, Jean-Christophe Andrau, Pierre Ferrier, and Salvatore Spicuglia. 2011. "H3K4 Tri-Methylation Provides an Epigenetic Signature of Active Enhancers." *The EMBO Journal* 30 (20): 4198–4210.

Petryszak, Robert, Maria Keays, Y. Amy Tang, Nuno A. Fonseca, Elisabet Barrera, Tony Burdett, Anja Füllgrabe, et al. 2016. "Expression Atlas Update--an Integrated Database of Gene and Protein Expression in Humans, Animals and Plants." *Nucleic Acids Research* 44 (D1): D746–52.

Pott, Sebastian, and Jason D. Lieb. 2015. "What Are Super-Enhancers?" *Nature Genetics* 47 (1): 8–12.

Russ, Brendan E., Moshe Olshansky, Jasmine Li, Michelle L. T. Nguyen, Linden J. Gearing, Thi H. O. Nguyen, Matthew R. Olson, et al. 2017. "Regulation of H3K4me3 at Transcriptional Enhancers Characterizes Acquisition of Virus-Specific CD8 T Cell-Lineage-Specific Function." *Cell Reports* 21 (12): 3624–36.

Sanda, Takaomi, Lee N. Lawton, M. Inmaculada Barrasa, Zi Peng Fan, Holger Kohlhammer, Alejandro Gutierrez, Wenxue Ma, et al. 2012. "Core Transcriptional Regulatory Circuit Controlled by the TAL1 Complex in Human T Cell Acute Lymphoblastic Leukemia." *Cancer Cell* 22 (2): 209–21.

Shen, Hongjie, Wenqi Xu, Rui Guo, Bowen Rong, Lei Gu, Zhentian Wang, Chenxi He, et al. 2016. "Suppression of Enhancer Overactivation by a RACK7-Histone Demethylase Complex." *Cell*. https://doi.org/10.1016/j.cell.2016.02.064.

Stamatopoulos, Kostas, Christos Kosmas, Chrysoula Belessi, Panagiotis Kyriazopoulos, Theodora Papadaki, Dimitra Anagnostou, and Dimitris Loukopoulos. 1999. "Molecular Analysis of Bcl-1/IgH Junctional Sequences in Mantle Cell Lymphoma: Potential Mechanism of the t(11;14) Chromosomal Translocation." *British Journal of Haematology*. https://doi.org/10.1111/j.1365-2141.1999.01314.x.

Stunnenberg, Hendrik G., International Human Epigenome Consortium, and Martin Hirst. 2016. "The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery." *Cell* 167 (5): 1145–49.

Thibodeau, Asa, Eladio J. Márquez, Dong-Guk Shin, Paola Vera-Licona, and Duygu Ucar. 2017. "Chromatin Interaction Networks Revealed Unique Connectivity Patterns of Broad H3K4me3 Domains and Super Enhancers in 3D Chromatin." *Scientific Reports* 7 (1): 14466.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Vose, Julie M. 2017. "Mantle Cell Lymphoma: 2017 Update on Diagnosis, Risk-Stratification, and Clinical Management." *American Journal of Hematology* 92 (8): 806–13.

Walker, B. A., C. P. Wardell, D. C. Johnson, M. F. Kaiser, D. B. Begum, N. B. Dahir, F. M. Ross, F. E. Davies, D. Gonzalez, and G. J. Morgan. 2013. "Characterization of IGH Locus Breakpoints in Multiple Myeloma Indicates a Subset of Translocations Appear to Occur in Pregerminal Center B Cells." *Blood*. https://doi.org/10.1182/blood-2012-12-471888.

Whyte, Warren A., David A. Orlando, Denes Hnisz, Brian J. Abraham, Charles Y. Lin, Michael H. Kagey, Peter B. Rahl, Tong Ihn Lee, and Richard A. Young. 2013. "Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes." *Cell* 153 (2): 307–19.

Yata, K., and S. Yaccoby. 2004. "The SCID-Rab Model: A Novel in Vivo System for Primary Human Myeloma Demonstrating Growth of CD138-Expressing Malignant Cells." *Leukemia* 18 (11): 1891–97.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author contributions

L.J.R. and D.R. conceived and designed the study; A.C., S.H., B.A.W., D.R. and L.J.R supervised the research; A.M. led the data analysis and interpreted results with L.J.R and D.R.; A.M., M.T., C.A., K.F. and N.K. processed and analysed genomic, epigenomic and chromatin accessibility data; C.A., G.J.M. and B.A.W. were responsible for collection and generation of myeloma data; A.M., D.R. and L.J.R took the lead in writing the manuscript; all authors discussed the results, wrote, reviewed and approved the manuscript.

Competing interests

Authors declare no competing interests.

Acknowledgements

This work was supported by the Wellcome Trust (206103/Z/17/Z to D.R. and 207556/Z/17/Z to S.H.), the Leukemia & Lymphoma Society. (to G.J.M. and B.A.W.) and Kidscan Children's Cancer Research (L.J.R). S.H. would also like to thank the Newcastle upon Tyne Hospitals NHS Charities, the NIHR Newcastle Biomedical Research Centre and the Sir Jules Thorn Charitable Trust (12/JTA). L.J.R and D.R. have Newcastle University Research Fellowships and A.M. a Newcastle University Faculty Fellowship.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures Legends

Figure 1: Genomic and epigenomic architecture of the *IGH* locus (14q32) in healthy and malignant human haematopoietic cells. Each panel represents collapsed cell-type specific signal of ChIP-seq chromatin states included in this study (**Supplementary Table 1**). Epigenomic consensus of individual elements was determined using healthy B-cell lineage samples, detailed described in **Methods**. AID motif clusters were detected as high enrichment of AID motifs (>200 of RGYW motifs per 2.5 kb). Abbreviations: GC – Germinal centre, CLL – Chronic lymphocytic leukemia, MCL – Mantle cell lymphoma, DLBCL – Diffuse large B-cell lymphoma, BL – Burkitt lymphoma, MM – Multiple myeloma.

Figure 2: Genomic and epigenomic architecture of the *CCND1* locus (11p13) in healthy and malignant human haematopoietic cells. Each panel represents collapsed cell-type specific signal of ChIP-seq chromatin states included in this study (**Supplementary Table 1**). Epigenomic consensus of individual elements was determined using healthy B-cell lineage samples, detailed described in Methods. Abbreviations: GC – Germinal centre, CLL – Chronic lymphocytic leukemia, MCL – Mantle cell lymphoma, DLBCL – Diffuse large B-cell lymphoma, BL – Burkitt lymphoma, MM – Multiple myeloma.

Figure 3: Chromatin landscape of the *CCND1* locus in five cell lines derived from B-cell haematological malignancies. Upper line of each cell line (**Supplementary Table 3**) represents the selected ChIP-seq chromatin states (**Supplementary Table 1**) and the lower line shows DNaseI hypersensitivity sites for the *CCND1* locus. Vertical black line in U266 mark position of the inserted Eα1 super-enhancer. The breakpoints within the *IGH* locus are characterised in **Supplementary Figure 2**. Shades of purple arrows symbolise translocation orientation. Numbers in coloured squares (red denotes high expression and blue low expression) show *CCND1* expression detected using RNA-seq in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Abbreviations: PC – Polycomb, P – Promoter.

Figure 4: Chromatin landscape of the *CCND1* (11q13) and *MAF* (16q23) loci in healthy human B-cells and multiple myeloma samples. Upper panels show selected ChIP-seq chromatin states (**Supplementary Table 1**) of **(a)** *CCND1* and **(b)** *MAF* loci in BLUEPRINT healthy B-cells. Each line of lower panel represents the ChIP-seq chromatin states for myeloma cell lines KMS11 and MM1S and seven multiple myeloma patients (P1-P7, patient derived xenograft material) for **(a)** *CCND1* and **(b)** *MAF* loci. Numbers in coloured squares (red denotes high expression and blue low expression) show gene expression detected by RNA-seq and displayed as Log2 normalised counts (Log2 NC). GC – Germinal centre, *sample contains chromosomal translocation involving the displayed region, described in detail in **Supplementary Table 3–4**.

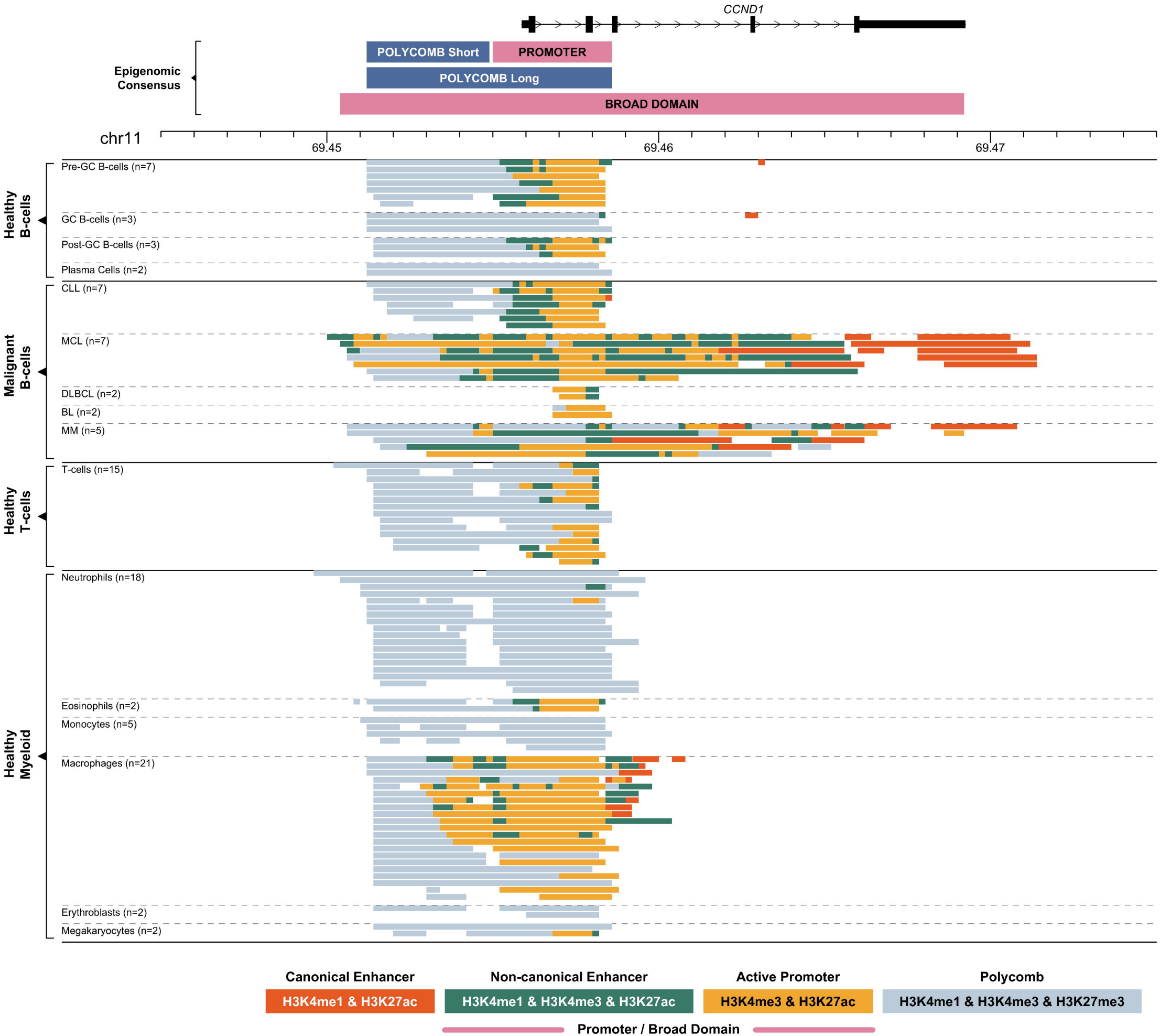
bioRxiv preprint doi: https://doi.org/10.1101/2020.02.12.938563; this version posted February 12, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

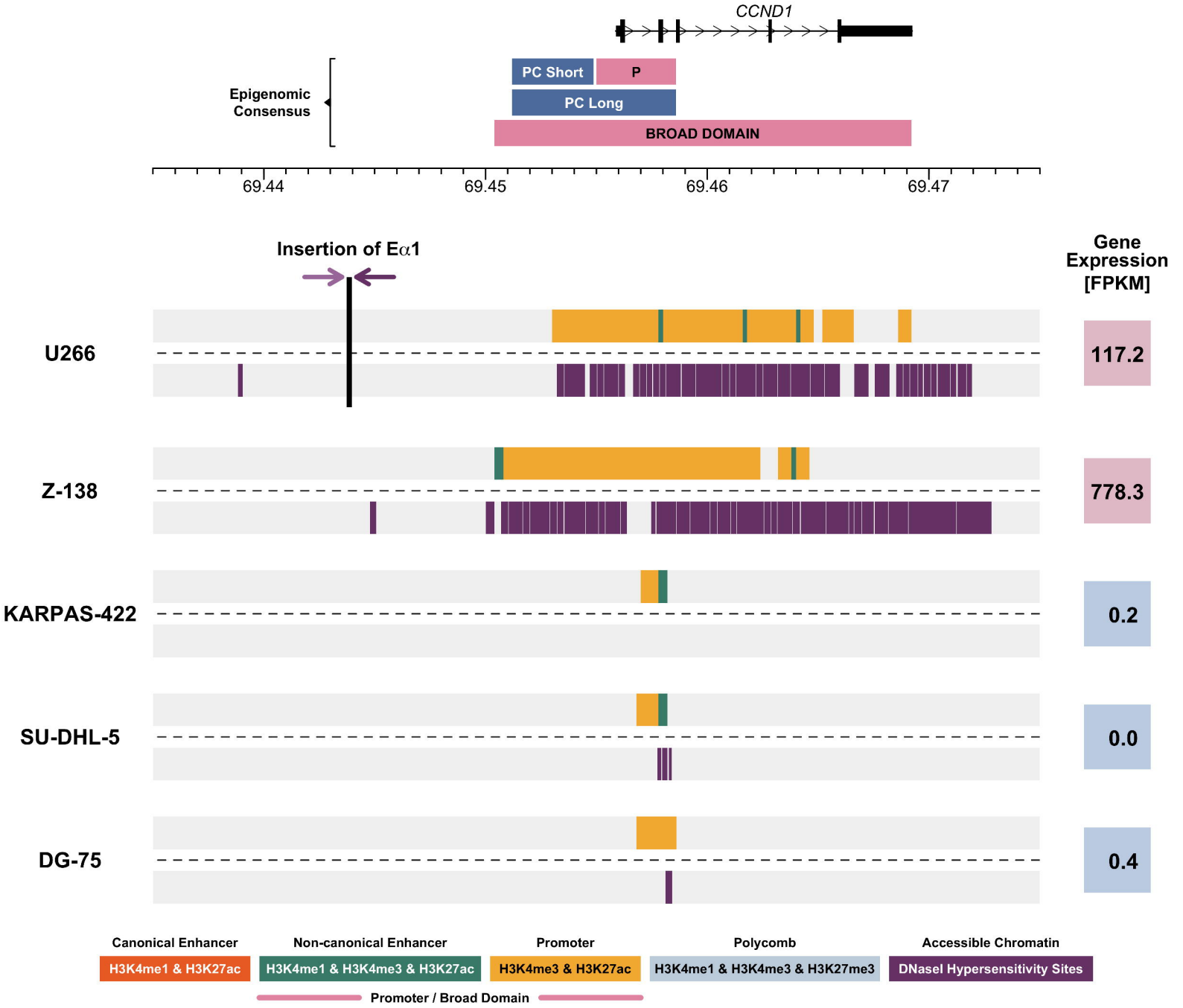
Figure 5. Chromatin landscape of the *LMO2* (11p13), *TLX3* (5q35) and *TAL1* (1p33) loci in healthy mature T-cells and three cell lines derived from T-cell haematological malignancies. Upper panels show selected ChIP-seq chromatin states (**Supplementary Table 1**) of **(a)** *LMO2*, **(b)** *TLX3* and **(c)** *TAL1* loci in BLUEPRINT healthy mature T-cells. Each line of lower panels represents individual histone marks peaks, separately for each cell line. *cell line contains chromosomal aberration involving the displayed region as follows: KOPT-K1 – *LMO2-TRA/TRD*, DND-41 – *TLX3-BCL11B*, and Jurkat – 12 bp insertion upstream of the *TAL1* gene (for details see **Supplementary Table 3**). #missing data for histone mark H3K27me3 in Jurkat cell line.

Figure 6: Scheme of epigenomic translocation model. In healthy cells, a cell identity or tumour suppressor gene is expressed by regulation of a super-enhancer that generates an H3K4me3 broad domain (H3K4me3-BD), allowing rapid and consistent activation of the transcriptional machinery (upper left). Proto-oncogenes are strictly regulated to control important physiological processes including the cell cycle (upper right). During a genomic translocation event, a super- enhancer is juxtaposed close to a proto-oncogene resulting in oncogenic activation (bottom right). Re-location of the super-enhancer brings the transcriptional machinery including the H3K4me3-BD. This epigenomic signature disappears at the original locus (bottom left).

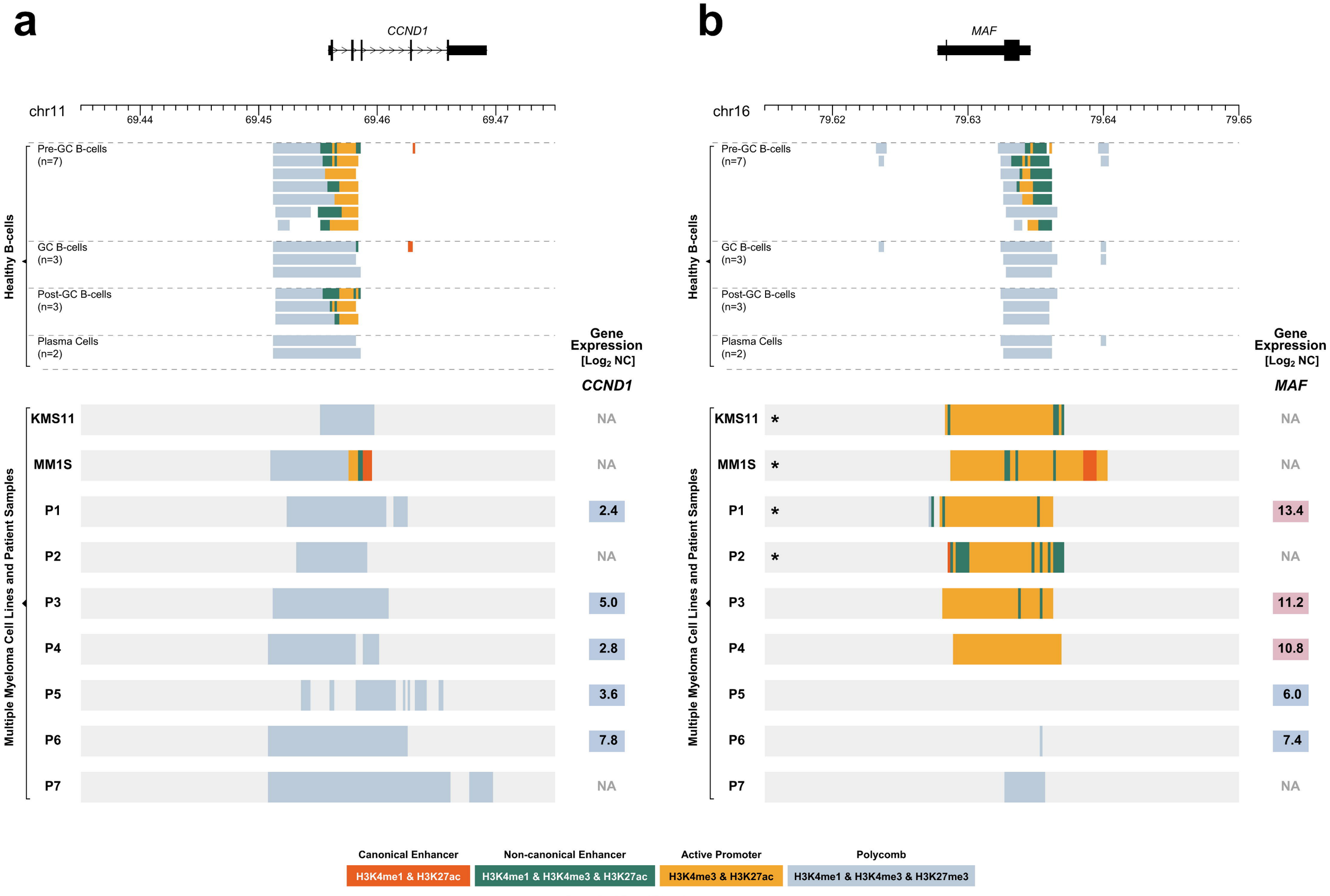
available under aCC-BY-NC-ND 4.0 International license.

available under aCC-BY-NC-ND 4.0 International license.

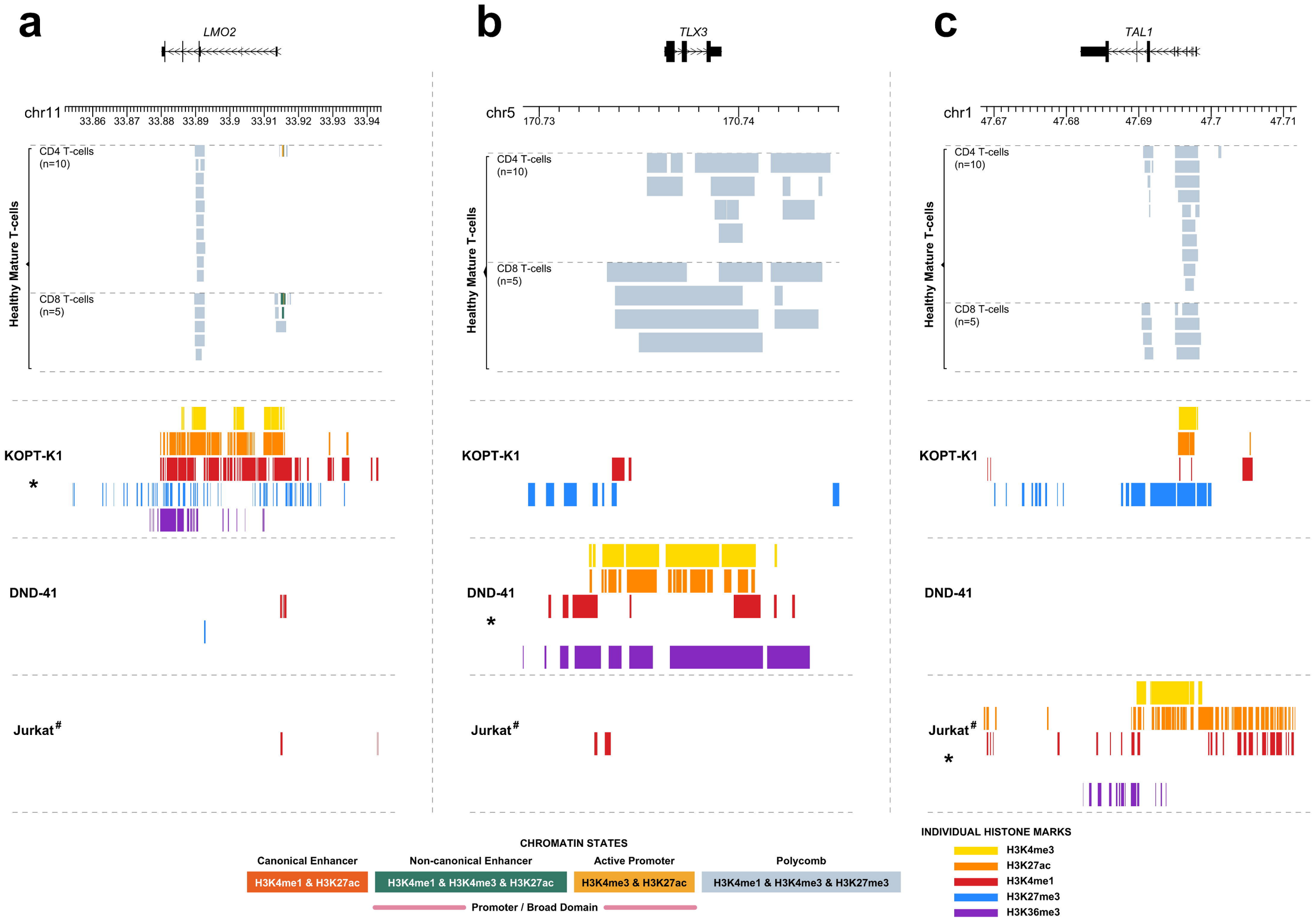


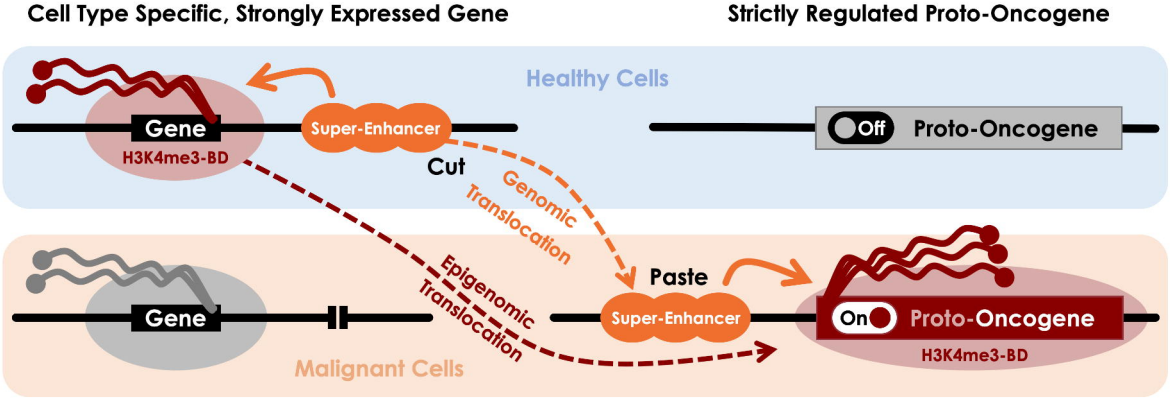

CCND1
Epigenomic Consensus
PC Short
P
PC Long
BROAD DOMAIN
69.44
69.45
69.46
69.47
Insertion of Eα1
Gene Expression [FPKM]
U266
117.2
Z-138
778.3
KARPAS-422
0.2
SU-DHL-5
0.0
DG-75
0.4
Canonical Enhancer
H3K4me1 & H3K27ac
Non-canonical Enhancer
H3K4me1 & H3K4me3 & H3K27ac
Promoter
H3K4me3 & H3K27ac
Polycomb
H3K4me1 & H3K4me3 & H3K27me3
Accessible Chromatin
DNaseI Hypersensitivity Sites
Promoter / Broad Domain

available under aCC-BY-NC-ND 4.0 International license.



available under aCC-BY-NC-ND 4.0 International license.




Cell Type Specific, Strongly Expressed Gene
Strictly Regulated Proto-Oncogene
Healthy Cells
Gene
H3K4me3-BD
Super-Enhancer
Cut
Genomic Translocation
Off
Proto-Oncogene
Epigenomic Translocation
Paste
Gene
Super-Enhancer
On
Proto-Oncogene
H3K4me3-BD
Malignant Cells