High-resolution simulations of chromatin folding at genomic rearrangements in malignant B-cells provide mechanistic insights on proto-oncogene deregulation

Daniel Rico*[1], Dan Kent[2], Aneta Mikulasova[1], Nefeli Karataraki[2], Brian A. Walker[3], Biola M. Javierre[4], Lisa J. Russell*[2], and Chris A. Brackley*[5]

[1] *Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK*
[2] *Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK*
[3] *Melvin and Bren Simon Comprehensive Cancer Center, Division of Hematology Oncology, Indiana University, Indianapolis, IN, USA*
[4] *Josep Carreras Leukaemia Research Institute (IJC), IJC Building, Campus ICO-Germans Trias i Pujol, Ctra de Can Ruti, Camí de les Escoles s/n, 08916 Badalona, Barcelona, Spain*
[5] *SUPA, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK*
* *Senior authors. Address corresponence to: C.Brackley@ed.ac.uk, lisa.russell@newcastle.ac.uk, or daniel.rico@newcastle.ac.uk*

Abstract Genomic rearrangements are known to result in proto-oncogene deregulation in many cancers, but the link to genome 3D structure remains poorly understood. Here we used the highly-predictive heteromorphic polymer (HiP-HoP) model to predict chromatin conformations at the proto-oncogene *CCND1* in healthy and malignant B-cells. After confirming that the model gives good predictions of Hi-C data for the non-malignant human B-cell derived cell line GM12878, we generated predictions for two cancer cell lines, U266 and Z-138. These possess genome rearrangements involving *CCND1* and the immunoglobulin heavy chain locus (*IGH*), which we mapped using targeted genome sequencing. Our simulations showed that a rearrangement in U266 cells where a single *IGH* super-enhancer is inserted next to *CCND1* leaves the local topologically associated domain (TAD) structure intact. They also revealed extensive changes in enhancer-promoter interactions within the TAD, and suggested that it is the downstream "epigenetic rearrangement" or chromatin remodelling which gives rise to this, rather than the presence of the inserted super-enhancer *per se*. Simulations of the *IGH-CCND1* reciprocal translocation in Z-138 cells suggested that an oncogenic fusion TAD encompassing *CCND1* and the *IGH* super-enhancers is created. We used the model to predict how the structure of *CCND1* changes in these different cell lines, and demonstrated that the simulations can be used to predict differences in chromatin interactions and gene expression for different translocation break-points.

Keywords
polymer model | chromatin conformation | epigenome | translocation | immunoglobulin | oncogene activation

Introduction

Chromatin structure, nuclear organisation and the epigenome are intimately linked to gene regulation and cell function, and are tightly controlled during differentiation. Genome rearrangements leading to structural variants (including deletions, insertions, duplication's and translocations) can disturb these, leading to gene mis-regulation and cancer [1]. An important example occurs during B-cell differentiation, where programmed breakage and recombination of the genome takes place in order to generate the broad heterogeneity of immunoglobulins (Igs) required for immune system function [2]. Errors in this process can lead to, for example, repositioning of Ig regulatory elements which then drives proto-oncogene activation [3]. The diversity in translocations and the difficulties in accessing and handling patient samples means that detection, accurate mapping, and characterisation of their functional consequences are inherently problematic.

Advances in molecular probes of genomic and epigenomic structure, such as ChIP-seq and chromosome-conformation-capture methods like Hi-C, have transformed our understanding of the regulatory link between structure and function. For example, Hi-C, which probes chromosome interactions genome-wide, has revealed that the genome can be partitioned into topologically associated domains (TADs) [4]. These are contiguous chromosome regions which show enriched self-interactions, and are thought to be associated with cis-regulatory mechanisms; they are often bounded by binding sites for the CCCTC-binding factor CTCF [5]. ChIP-seq profiling of histone modifications and protein binding has identified super-enhancers, clusters of enhancer elements thought to drive expression patterns responsible for cell identity [6]. Super-enhancers have been found to preferentially sit within TADs which are highly insulated from their surroundings [7], and have been identified as drivers of oncogene expression in many tumours [8, 9].

A full epigenetic and three-dimensional (3D) structural

characterisation of genome rearrangements is crucial for our understanding of proto-oncogene activation. However, the small size of patient samples and their inherent variability means that routinely applying methods like Hi-C and ChIP-seq to primary cancer samples remains challenging. This is why much work to date has focused on the biogenesis of chromosome translocations and the factors which control where they occur [10, 11], but comparatively less attention has been paid to the structural *consequences* of genome rearrangements [12, 13]. In this paper we show how computer simulations can help to understand the effects of common genome rearrangements in B-cell malignancies *in silico*, and how these lead to the changes in the 3D structure of gene loci which in turn leads to deregulation.

We used our "highly-predictive heteromorphic polymer" (HiP-HoP) model [14, 15] to study genomic rearrangements involving the Ig heavy chain locus (*IGH*) [16] and the *CCND1* gene encoding the cell cycle protein Cyclin D1. *IGH-CCND1* is a common translocation observed in mantle cell lymphoma (MCL) and multiple myeloma (MM); when translocated together the *IGH* super-enhancers drive *CCND1* over-expression. In previous work [9] we analysed the changes in chromatin states associated with super-enhancer translocation events, finding that relocation of the *IGH* super-enhancer results in local chromatin remodelling and the emergence of patterns of histone modifications including H3K4me3 broad domains [9, 17, 18]. Here we study how these epigenetic changes alter 3D chromosome structure.

HiP-HoP uses data on DNA accessibility, chromatin states, and protein binding to generate an ensemble of 3D chromosome structures. We previously showed that the method gives good predictions of both population level data (from methods like Hi-C or 4C) and single cell measurements (*e.g.*, from fluorescence microscopy) [14]. Importantly, HiP-HoP does not use Hi-C (or any other prior 3D genome) data as an input (unlike other popular models which use fitting-based methods [19]). This means that we can make predictions about the 3D structure of cell types or tissues where Hi-C data is not available.

Here we found that HiP-HoP method gives good predictions of Hi-C data at the *CCND1* locus in a human cell line derived from healthy B-cells. Simulations of an MM cell line where a single *IGH* super-enhancer is inserted into the *CCND1*-TAD predicted that while the TAD remains intact, chromatin interactions among gene promoters and emergent regulatory regions are radically altered. Predictions for an MCL cell line possessing a translocation which repositions *CCND1* next to the *IGH* joining region [16] showed an oncogenic fusion TAD encompassing *CCND1* and four previously identified *IGH* super-enhancer regions [9, 20–22]. Further simulations suggest that it is the downstream epigenetic changes which drive the 3D changes, rather than the genomic rearrangements itself. This points to a scenario where the translocated *IGH* super-enhancers first recruit chromatin remodelers to the region; this then drives epigenomic changes which in turn disrupts the wider 3D structure leading to dysregulation of *CCND1*. Importantly, the simulations allowed us to investigate different possible rearrangements, to infer the pathway through which the *CCND1* proto-oncogene becomes activated, and to generate testable hypotheses which will direct new experiments suggesting targets for therapeutic intervention. This points to a future role for the method in a clinical setting, *e.g.* in personal medicine, where structural predictions based on available or easily obtainable data could be extremely valuable.

Results

HiP-HoP predicts the domain structure of the *CCND1* locus in a B-cell derived cell line

We previously showed that in malignant B-cells, genome rearrangements involving *CCND1* lead to local changes in histone modifications [9]. We hypothesised that this in turn leads to changes in 3D structure and chromatin interactions which drive *CCND1* expression. To understand these malignant structural changes, it is first necessary to study the healthy case. To this end we considered the human B-cell derived lymphoblastoid cell line GM12878, adapting our HiP-HoP method to study chromatin structure around *CCND1*. This is an ideal cell line due to the abundance of publicly available data.

In the model a 3 Mbp chromosome region was represented by a chain of beads (a common approach in polymer physics [15]). The region was chosen because it is large enough to encompass approximately seven TADs surrounding *CCND1* such that its local chromatin context is captured, but small enough to give feasible simulation run times. The model combines three mechanisms which drive locus structure: (i) diffusing spheres representing complexes of chromatin binding proteins which can bind to multiple points at the same time to form molecular bridges (*e.g.*, between promoters and enhancers) [24–26]; (ii) a heteromorphic polymer structure, where different sections of the bead chain have different properties (thickness/compaction and flexibility); and (iii) the loop extrusion mechanism [27, 28]. Loop extrusion is a molecular mechanism where chromatin is pushed onto loops by factors (probably the cohesin complex [29]), which are halted by CTCF proteins bound in a convergent orientation [5].

A way to identify different types of chromatin is via high-throughput analysis of histone modification data and hidden Markov modelling (HMM) [30, 31]; this approach classifies chromosome regions into a number of states (*e.g.* promoter/enhancer associated, polycomb repressed, and heterochromatin states, see Suppl. Table 1). In the model we had two different types of chromatin structure (the heteromorphic polymer): more open regions (thinner, more flexible polymer) and more compact regions (thicker, "crumpled" polymer, see Suppl. Fig. 1 for a schematic). We used the chromatin states to identify these regions: H3K27ac associated states have the more open structure [32]. The model included three different species of bridge forming protein: a general "active binder", and two species of repressive binders (this is an extension of the scheme in

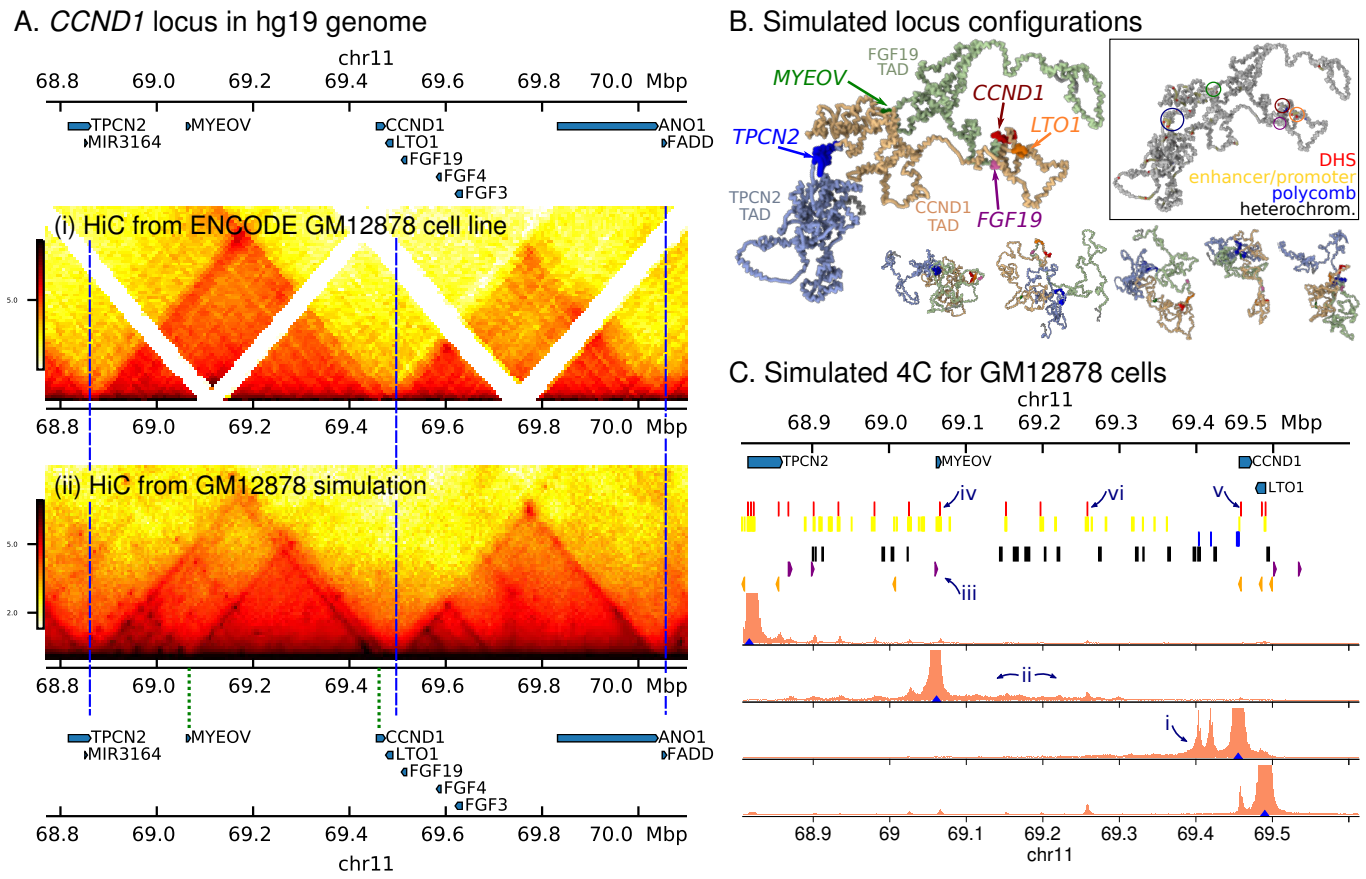

Figure 1. **A.** A map of the genes in the vicinity of *CCND1* is shown above heat maps for: (i) Hi-C data for GM12878 cells obtained from Ref. [5]; and (ii) simulated Hi-C from the same region. TAD boundary positions are indicated with blue dashed lines; green dashed lines indicate a "subTAD" region of increased self-interaction (see text). Some unmappable regions are visible as gaps in the data (white stripes). Simulated Hi-C maps are obtained from 4400 independent configurations. **B.** Simulation snapshots of the 3D structure of the *CCND1* region, including the *CCND1*-TAD (pale orange region) and the two neighbouring TADs (pale green and blue regions) (a 1.3 Mbp section of the 3 Mbp region simulated is shown). One representative snapshot is shown in large, with five other smaller examples. Several genes within the region are shown in bright colours as indicated by labels. Pale colours show the TADs. For clarity proteins are not shown. Box: the same configuration is shown but coloured according to the input data used in the simulation, as indicated by label colours. Gene positions are circled. **C.** Simulated 4C interaction profiles for GM12878 cells are shown from four viewpoints (blue triangles; these are positioned at promoters of *TPCN2*, *MYEOV*, *CCND1* and *LTO1*). Data used as simulation input (obtained from ENCODE [23], see Supplementary Methods and Suppl. Table 2) is shown as coloured blocks above the plots. Red blocks indicate DNase hypersensitive sites (DHS), used to infer binding sites for active proteins. Blue and black blocks indicate regions with chromatin states corresponding to polycomb association and heterochromatin respectively, used to infer binding for the corresponding proteins. Yellow blocks indicate regions with chromatin states associated with H3K27ac, and indicate regions which have a more open chromatin structure in the model. Orange and purple arrowheads indicate the position and directionality of CTCF sites, used to infer loop extruder anchors. Some features are labelled with numbered arrows as referred to in the text.

Ref. [14] which only included active proteins). To identify active protein binding sites we used DNA accessibility measured via DNase-seq experiments, assuming that DNase-hypersensitive sites (DHS) are binding sites (see Methods and Supplementary Methods). Repressive protein binding sites were identified using the chromatin states, with one species binding to states rich in H3K9me3, and one species representing polycomb repressive complexes and binding to states rich in H3K27me3. Finally, ChIP-seq for CTCF was used to identify "anchor" sites where loop extrusion is halted. Full details of the simulations are given in the Supplementary Methods, and a schematic is shown in Suppl. Fig. 1.

To verify that HiP-HoP gives good predictions for the *CCND1* locus, we generated simulated Hi-C interaction maps and compared with publicly available data [5] (Fig. 1B; Suppl. Fig. 1 shows a similar plot but for the entire 3 Mbp region which was simulated, along with the experimental data used as an input to the model). The simulations gave a good prediction of the TAD pattern across the region. *CCND1* sits at the far right of a TAD, which is bounded by convergent CTCF sites and starts just to the left of (but not encompassing) the gene *TPCN2*, and ends between *LTO1* (immediately downstream of *CCND1*) and *FGF19*. We shall refer to these three neighbouring domains as the *TPCN2*, *CCND1*, and *FGF19*-TADs. The gene *MYEOV* sits just left of the centre of the *CCND1*-TAD.

Simulated Hi-C maps tend to show more structure within the domains than experiments (visible as dark spots and

stripes); typically these features arise due to model protein mediated enhancer-promoter interactions. This was also observed in previous HiP-HoP studies, and could be due either to an overestimation of local interactions in the simulations, or because the experimental data lacks sufficient resolution to reveal these features (*e.g.*, promoter-enhancer interactions are more apparent in targeted 4C [33], nucleosome resolution MicroC [34], and recent Hi-C data which has been treated using a method for removing biases [35]). In the simulated Hi-C it is clear that the *CCND1*-TAD is split into two sub-TADs, with a weaker (internal) domain boundary near *MYEOV*; this is also evident, albeit weakly, in the experimental data, though it is obscured by an unmappable DNA region (white stripe in Fig. 1B). A typical simulated structure, with the three TADs and gene positions highlighted, is shown in Fig. 1A; also shown are 5 other representative structures (from a population of 4400 generated in the simulations).

A more focused view of chromatin interactions at gene promoters is shown in simulated 4C interaction profiles in Fig. 1C. These reveal that *CCND1* often interacts with polycomb associated regions (arrow 'i' in the figure); this is because there is enrichment of H3K27me3 at the promoter. There are also H3K4me1/3 modifications, which could indicate variation between alleles or across the population, with the gene sometimes expressed and sometimes repressed; RNA-seq data obtained from ENCODE [23, 36] show weak expression of the gene (Fig. 2F).

MYEOV (not expressed in these cells, see Fig. 2F) shows H3K4me1 and H3K27ac marks (an enhancer-associated chromatin state). While there are no DHS at *MYEOV* to drive strong interactions, there is a general enrichment of interactions within a broad region downstream of the gene (arrow 'ii' in Fig. 1C); this is probably driven by a CTCF site immediately upstream of the gene, and a DHS immediately downstream of the gene (arrows 'iii' and 'iv' respectively). *MYEOV* does not show strong interactions with other nearby genes. *TPCN2* and *LTO1* are both expressed in these cells, both have a DHS at their promoter, and in simulations they interact with several other nearby DHS. Despite *TPCN2* being in the adjacent TAD, it shows peaks of interaction with a number of DHS between it and *MYEOV*, though the *MYEOV* proximal CTCF site (arrow 'iii') seems to isolate it from further downstream interactions. The *LTO1* promoter interacts with a DHS within the *CCND1* gene body (which has chromatin state indicative of weak transcription, arrow 'v'), and with a DHS between *MYEOV* and *CCND1* which has an enhancer chromatin state (arrow 'vi' in Fig. 1C).

Super-enhancer insertion near *CCND1* drives changes in 3D structure within the TAD

We next used HiP-HoP to study *CCND1* in the MM cell line U266. These cells possess a genomic rearrangement where an enhancer region from *IGH* is inserted upstream of *CCND1*, within its TAD boundaries. A number of super-enhancers have previously been described in the non-variable *IGH* region, including the long recognised Eα1, Eα2 and Eμ enhancers [20–22]. More recently we proposed that a fourth element Eδ should be identified as distinct from Eμ, based on high quality chromatin state mapping [9]. The ∼100 kbp inserted region in U266 includes Eα1 (accurately mapped in Ref. [9]). U266 is a highly useful model because it allows study of the effect of relocating a single *IGH* enhancer within an existing TAD; strikingly, it leads to significant changes in expression which are similar to those resulting from a reciprocal translocation.

In order to simulate U266 cells we generated an alternative genome build which includes the insert, denoting this hg19_u266 (compared to the unaltered hg19 genome used for GM12878 above). We obtained U266 DNA accessibility and chromatin states from the BLUEPRINT project [37], and used CTCF binding site data for B-cells (from ENCODE [23]). Figure 2A shows simulated Hi-C for U266 along with the input data; the insert region is marked in green. For comparison, Fig. 2B shows the intact locus from the GM12878 simulations (no insert). Example configurations are shown in Fig. 2C.

The TAD structure is unchanged by the presence of the insert, however, interactions within the domains show significant differences between the two cell lines. Particularly, in U266 there are several strongly interacting regions (dark spots in the Hi-C map, some highlighted with green circles), including close to *CCND1* and within the *IGH* insert itself. Many of these are between DHS, between regions with heterochromatin states, or between regions with polycomb states. DHS often coincide with regulatory elements such as promoters; particularly in U266 there is a cluster of DHS near and within *CCND1* which is not present in GM12878. Together these observations suggest that loop extruders and CTCF sites drive TAD formation, while bridging proteins drive interactions within TADs. Extrusion is unchanged by the presence of the insert (which does not contain any CTCF sites), while significant remodelling of chromatin in U266 compared to GM12878 leads to changes within the TAD.

Figure 2D shows simulated 4C with viewpoints positions at gene promoters (top four rows). We found that there were new (compared to GM12878) strong interactions between the *CCND1* promoter and several DHS across the TAD, including three prominent peaks within the insert (arrow 'i' in Fig. 2D), as well as several peaks at DHS between *MYEOV* and the insert (arrows 'ii'; some of these DHS were not present in GM12878). Reciprocal interactions were observed as expected (Fig. 2D, bottom four rows). Many of the regions around the new DHS also gained enhancer or promoter chromatin states compared to GM12878. In U266 *CCND1* also interacts with nearby *LTO1* (its promoter and gene body, arrow 'iii' in Fig. 2D), as well as showing weaker interactions with the *MYEOV* and *TPCN2* promoters (arrows 'iv' and 'v'). These changes result from the fact that in U266, *CCND1* gains DHS and active promoter and enhancer chromatin states (H3K27ac and H3K4me3) at the promoter and across its body (while losing H3K4me1 and H3K27me3 at the promoter). These features are consistent with H3K4me3 broad domains). These have previously been associated with super-enhancers, and their appearance has been implicated in cancer-specific super-enhancer hijacking at a

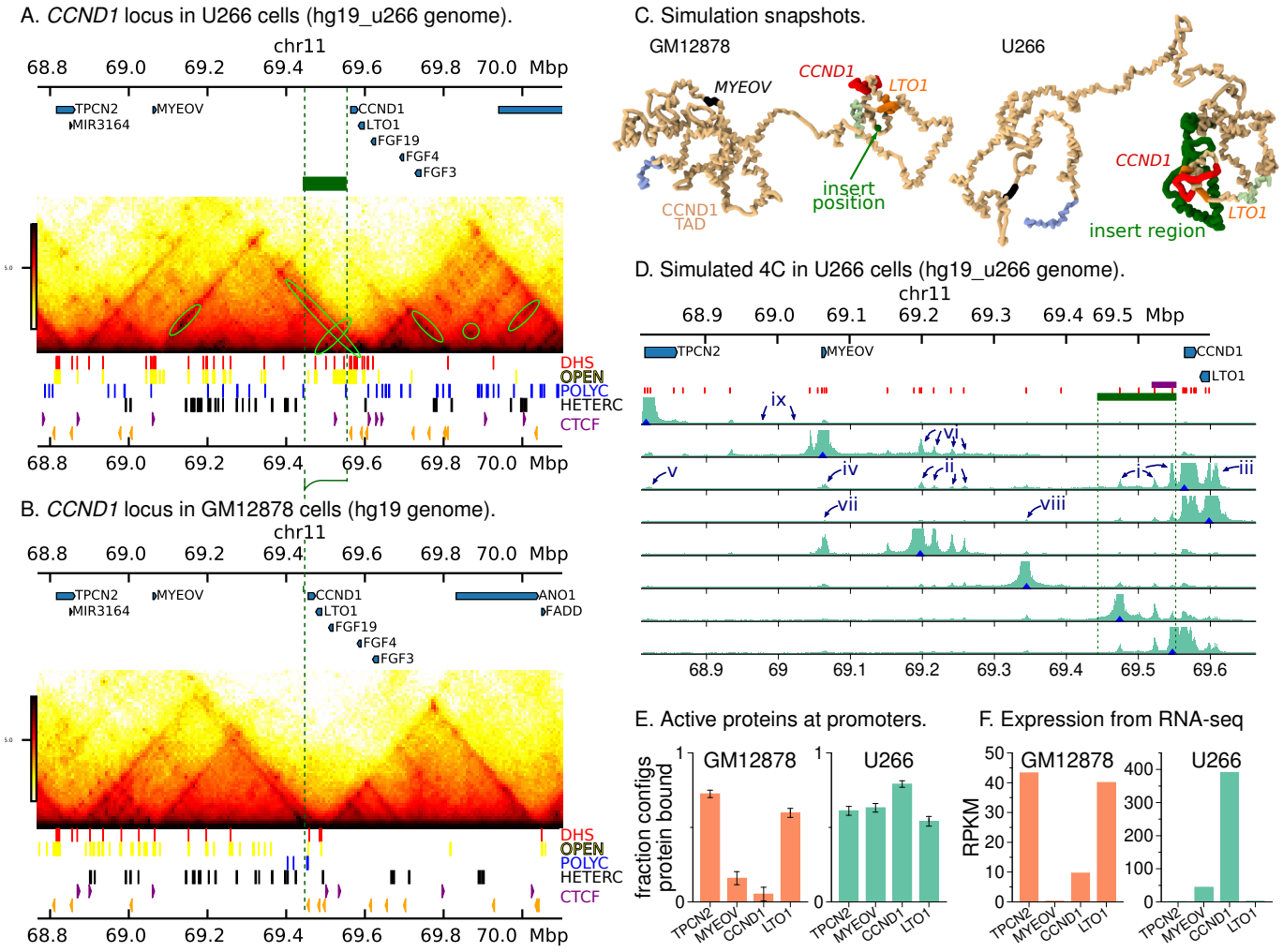

Figure 2. **A.** Simulated Hi-C data is shown below a map of the genes within the rearranged *CCND1* locus as found in the U266 cell line (hg19_u266 genome; a region encompassing the *CCND1* and *FGF19*-TADs is shown). Green rings highlight additional interactions not present in GM12878 cells. The green block and dashed lines show the position of the inserted region. Data used as simulation input (obtained from the BLUEPRINT and ENCODE projects [23, 37]) is shown as coloured blocks under the Hi-C map using the same scheme as in Fig. 1. **B.** Simulated Hi-C data is shown for the intact *CCND1* locus for GM12878 cells (hg19 genome). The green dashed line indicates the insertion point for the U266 rearrangement. **C.** Example snapshots from simulations showing the *CCND1*-TAD in GM12878 (left) and U266 (right). In the latter the *IGH* insert containing Eα1 is shown in green; for comparison in the GM12878 snapshot the position of the insert is shown as a single green bead. **D.** Simulated 4C plots are shown for U266 at various viewpoints (blue triangles). In the top four tracks viewpoints are at promoters of *TPCN2*, *MYEOV*, *CCND1* and *LTO1*. The bottom four tracks show reciprocal viewpoints from interacting regions. The green block and dashed lines show the position of the insert, red blocks show DHS, and the purple block shows Eα1. Some features are labelled with numbered arrows as detailed in the text. **E.** Bar plots showing the proportion of simulated configurations in which the different promoters are bound by an active protein. This is a crude predictor of expression. **F.** Plots showing expression level of genes as determined from RNA-seq data (obtained from BLUEPRINT and ENCODE [23, 37]).

number of oncogenes (including *CCND1* [9, 17, 18]).

The promoter of *MYEOV* also gains a DHS in U266 compared to GM12878, and the chromatin state changes from enhancer to promoter. It also shows a number of new interaction peaks, particularly in a region downstream where there is a cluster of new DHS with active enhancer and promoter chromatin state (arrows 'vi' in Fig. 2D). Surprisingly, *MYEOV* shows very little interaction with the *IGH* insert itself. The promoter of *LTO1* interacts strongly with *CCND1*, but more weakly with the *IGH* insert. Interestingly, there are two DHS with which *LTO1* showed strong interactions in GM12878, but only weak interactions in U266 (arrows 'vii' and 'viii'); the chromatin at these positions also shifts from enhancer to promoter state.

Unlike the other genes, the promoter of *TPCN2* seems to show fewer interaction peaks in U266 compared to GM12878; this can be explained as there is also a loss of a DHS (arrows 'ix' in Fig. 2D point to positions where there were interaction peaks in GM12878 cells). However, the level of interactions with some sites upstream of *TPCN2* (in the *TPCN2*-TAD) is also reduced. The reasons for this are unclear, but one possibility is that the abundance of new DHS in and around the insert sequesters active proteins, meaning fewer are available to stabilise loops involving *TPCN2*.

In our simulations, the active proteins represent complexes

of transcription factors and polymerases. A simple, though perhaps crude, way to infer gene activity from the simulated structures, is to ask how often a promoter is bound by one of these proteins within the population of structures (see Supplementary Methods). We previously showed in simpler whole-chromosome simulations that this does correlate with gene expression level [38]. Figure 2E shows this measure for several genes in the two cell lines, while Fig. 2F shows expression levels inferred from RNA-seq data (obtained from BLUEPRINT [37] and ENCODE [23, 36]). The simulations correctly predicted much higher *CCND1* expression in U266 than in GM12878, and a smaller increase in expression of *MYEOV*. On the other hand, the severe decrease in expression of *TPCN2* and *LTO1* observed in RNA-seq is not correctly predicted by the simulations. This could be because the interaction partners of these promoters change in U266, *i.e.* they no longer interact with their canonical enhancers. For example, recent high-resolution promoter capture data showed that super-enhancer interaction is not sufficient for high expression [39], suggesting only some genes are amenable to super-enhancer activation.

Since HiP-HoP generates a population of many structures, each representing a single cell, we can also obtain single cell like measurements (Suppl. Fig. 2). For example we can obtain separations between specific points, such as would be measured experimentally in 3D DNA fluorescence *in situ* hybridisation (3D DNA FISH) microscopy measuring the separation of fosmid clone probes. We can also measure the overall 3D size of a given region. We find that the mean separation of probes on either side of the insert position, and the overall size of the *CCND1*-TAD, tends to be larger in U266 where the insert is present. This is expected because not only is the genomic (linear) separation of the probes larger, but in the model more of the region has the open (H3K27ac associated) structure. The latter effect is highlighted by examining the 3D size of a region of fixed genomic length around *CCND1* in the two cell lines: the 3D size is again larger in U266 cells (Suppl. Fig 2D, and see also Fig. 4C below).

The reciprocal translocation t(11;14)(q13;q32) preserves boundaries to generate an oncogenic TAD fusion

The MCL cell line Z-138 harbours a translocation that engenders a dramatic change to the local environment of *CCND1*: it relocates to chromosome 14 where it becomes juxtaposed with the joining region of the *IGH* locus. While the t(11;14) rearrangement is common, the precise breakpoint has not previously been mapped in this cell line. Using paired-end read targeted DNA sequencing, we precisely mapped the chromosomal changes (Suppl. Fig. 3). With this information we could then perform simulations of the rearranged genome, which we denote hg19_z138.

Figure 3 shows results from simulations using Z-138 data (again obtained from BLUEPRINT [37] and ENCODE [23]). The simulated Hi-C map (Fig. 3A) predicts that a new TAD will form which encompasses *CCND1*, *LTO1* and the *IGH* super-enhancers (Eα2, Eα1, Eμ, and Eδ). Internal to this

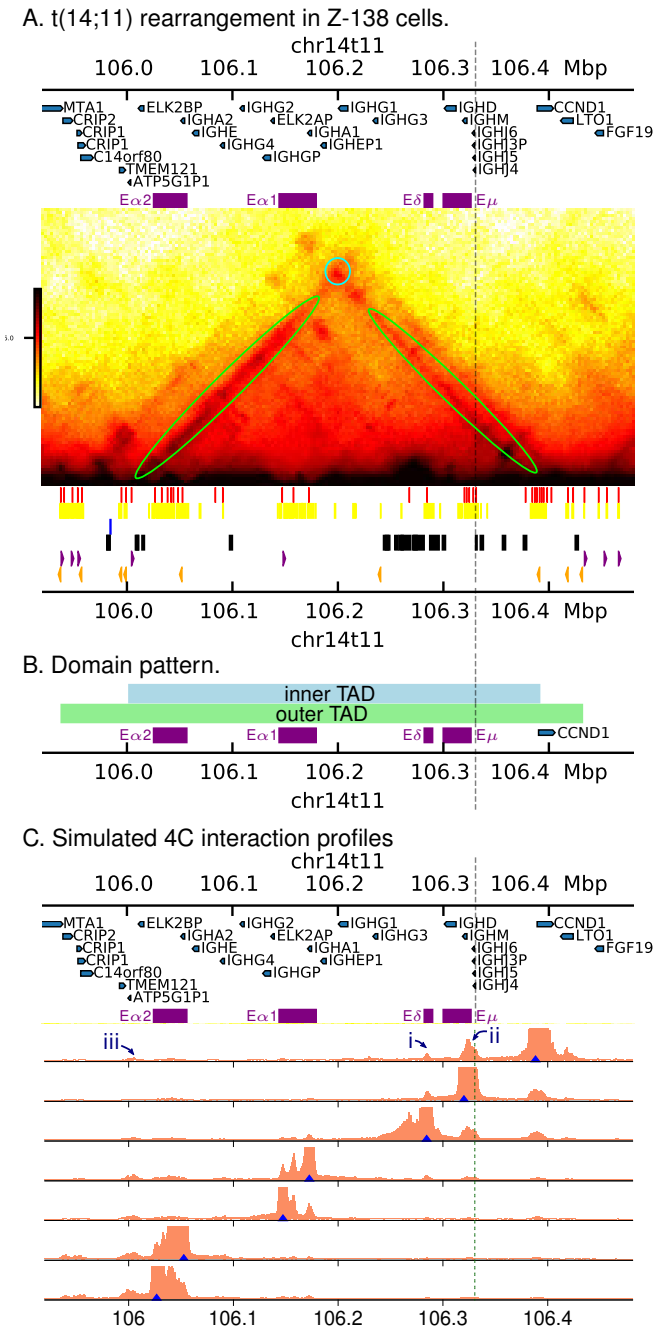

Figure 3. **A.** Simulated Hi-C map from simulations of Z-138 cells which possess a t(11;14) translocation (hg19_z138 genome build). Positions of the previously identified *IGH* super-enhancers are shown as purple blocks above the map [9]. Simulation input data (DHS, chromatin states and CTCF peaks) are shown under the map as in previous figures. Blue and green circles and ellipses on the map highlight some interactions discussed in the main text. **B.** Positions of TADs obtained from inspection of the simulated Hi-C map shown in A. **C.** Simulated 4C from viewpoints positioned at DHS across the region. Blue triangles show viewpoint positions. Some features are highlighted with numbered arrows as discussed in the text.

there is also a stronger subTAD (or 'inner TAD'). These domains are shown in Fig. 3B. The boundary of the outer TAD on the right is formed by a pair of divergently oriented CTCF sites, and is conserved from the chromosome 11 boundary

observed in GM12878 and other cell types. The left-hand outer TAD boundary is formed by three inward pointing CTCF sites; a boundary-like pattern is also seen at this position in chromosome 14 in some other cell types (which do not have a translocation). In summary, the simulations predict that the boundaries are conserved after the translocation, and this leads to formation of a new *IGH-CCND1* TAD. The inner TAD is also bounded by CTCF sites, which show a strong interaction peak (blue circle in Fig. 3A), and encompasses the Eα2 enhancer and the *CCND1* promoter (both of which show a "stripe" of interaction in the Hi-C, green ellipses in the figure).

Figure 3C shows simulated 4C interactions with viewpoints at the *CCND1* promoter and several positions within the *IGH* super-enhancers. *CCND1* shows prominent interaction peaks in both Eδ and Eμ; these coincide with DHS within H3K27ac regions (arrows 'i' and 'ii' in the figure; the interaction is strongest where the DHS are clustered together). There are also smaller peaks of interactions within Eα1 and Eα2, and with a number of other DHS (notably a cluster of DHS to the left of Eα2, arrow 'iii'). Similar to U266, there are a number of DHS at the *CCND1* promoter and within the gene body which are not present in the GM12878 cell line. In the simulations the interactions of the promoter are driven by binding (and molecular bridging) of active proteins; this bridging is likely further promoted by CTCF/cohesin driven loop extrusion (particularly the longer ranged interactions). Viewpoints positioned at DHS within the enhancers show reciprocal interactions with *CCND1*. There are also strong interactions between DHS within the same enhancer, and weaker interactions between DHS in different enhancers, as expected due to the general dependence on genomic separation.

Here we have simulated the break-point and rearrangement as mapped for the Z-138 cell line. It is also possible to simulate alternative rearrangements to better understand the effect of break-point variation between patients. Up to 50% of t(11;14) MCL break-points are within the so-called major translocation cluster (MTC) about 120 kbp upstream of *CCND1* [40], with the rest scattered throughout a broader 380 kbp region around this (the Z-138 break-point is about 60 kbp upstream of *CCND1*). Supplementary Figure 4 shows simulated 4C from two different examples where the chromosome 14 break-point is kept the same but we move the break-point on chromosome 11 to different positions: one 11 kbp upstream of *CCND1*, and one about 100 kbp upstream. We found that in general interactions between the gene and the Eμ enhancer are weaker if their genomic separation is larger.

Chromatin remodelling and *CCND1* 3D structures differ in U266 and Z-138

In Fig. 4A we show simulated 4C plots with a viewpoint positioned at the *CCND1* promoter, but the data are mapped back to the hg19 reference genome, and interactions with the *IGH* locus are shown. For U266 cells only the small insert is in the vicinity of *CCND1*.

First, we note that there are differences in the pattern of chromatin states and DHS between the U266 and Z-138 cell lines. Particularly, within Eα1, there are three DHS in Z-138 cells, while there are only two in U266. In the latter these show strong interactions with *CCND1* due to their proximity to the gene, whereas the interactions are much weaker in Z-138. We also note that a DHS at around chr11:106,100,000 in U266 which interacts with *CCND1* is not present in Z-138 (arrow 'i' in the figure); the chromatin state here is also different in the two cell lines. This suggests that the rearrangement not only leads to a remodelling of chromatin around *CCND1*, but also within the *IGH* enhancer region. It is important to note, however, that the aberrant rearrangements may arise from different errors in antibody production processes at different stages of B-cell differentiation. In MCL (Z-138) translocations occur due to errors in V(D)J rearrangement in immature B-cells prior to antigen stimulation, while in MM (U266) errors in both V(D)J rearrangement and class switch recombination can result in rearrangement events.

From the interaction profiles we found that in U266 the two most frequently interacting DHS are opposite ends of the Eα1 enhancer region. The DHS closest to the gene interacts most strongly. Interestingly, one DHS within the insert (arrow 'ii' in Fig. 4A) does not interact with the promoter and is not embedded within a region of H3K27ac marked chromatin. This suggests that the occurrence of DHS-promoter interactions depends strongly on the surrounding chromatin environment. In Z-138, the strongest interaction with the *CCND1* promoter is with a cluster of DHS within Eμ (*i.e.*, the closest enhancer genomically).

In Fig. 4B we show the chromatin states and DHS at three regions within the *CCND1* locus where there is significant remodelling compared to GM12878. In both U266 and Z-138 the *CCND1* gene body gains several DHS and the promoter-like chromatin state, *i.e.* a H3K4me3 broad domain appears. This is shifted slightly in an upstream direction in Z-138 compared to U266. From 10 distinct DHS within *CCND1*, 5 are common to both cancer cell lines, 3 are specific to Z-138 and 2 are specific to U266. In GM12878 cells there is only one DHS within the gene body, and none at the promoter (which has polycomb associated histone modifications). In other words, the different rearrangements lead to similar, but not identical, changes of the chromatin structure at *CCND1*.

Figure 4C shows the distribution of the 3D size of the *CCND1* gene body (measured by its radius of gyration, R_g, see Supplementary Methods). The difference between the cell lines is striking: in GM12878 the gene is on average more compact, and the variation is smaller than in the other cell lines. The snapshots in Fig. 4C show typical configurations: the differences in the size of the gene, the chromatin states, and the DHS pattern are clear. In GM12878, where the gene is mostly inactive, it has a crumpled structure. In U266 there are several DHS within the gene, which is more stretched out in the configuration shown. Protein mediated loops can form between the DHS, which would reduce the 3D size of the gene: variation in the number of such loops present at any one time leads to the large variation in R_g. The snap-

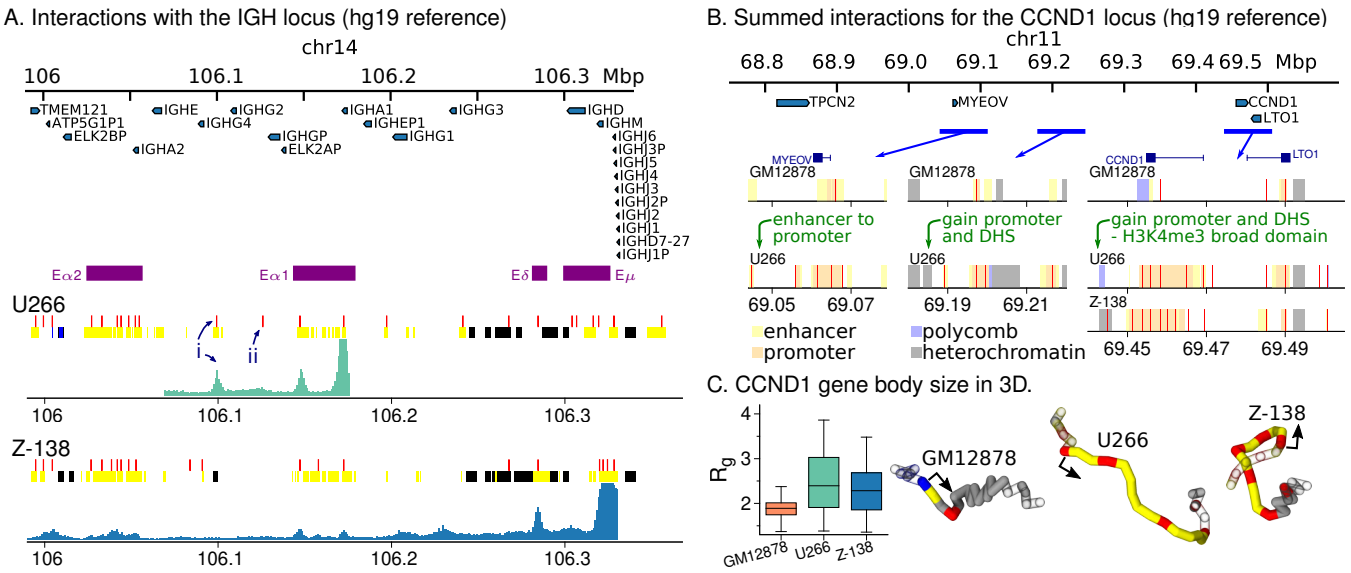

Figure 4. **A.** Plots showing 4C interactions between the *CCND1* promoter and the *IGH* locus from simulations of U266 and Z-138. These have been mapped back to the hg19 reference genome; only regions which are proximate to *CCND1* after the rearrangement show a signal. Enhancer positions are shown as purple blocks. Red, yellow, blue and black blocks show DHS, open chromatin, polycomb and heterochromatin regions from the simulation input data as in previous figures. **B.** Three chromosome 11 regions which show substantial chromatin remodelling in U266 and Z-138 cells are indicated with blue lines. Plots showing chromatin states as coloured blocks are shown for these regions as indicated by arrows. Yellow, orange, blue and back blocks indicate enhancer, promoter, polycomb and heterochromatin states respectively. Note that both enhancer and promoter states are associated with H3K27ac, and so are treated the same in simulations (open chromatin). Red lines indicate positions of DHS. Green text indicates how the chromatin is remodelled compared to GM12878. **C.** The radius of gyration gives a measure of the 3D size of the *CCND1* gene body; box plots show distributions across the population of structures for each cell type. Images to the right show typical snapshots of the gene and a 5 kbp flaking region on each side. The gene body is shown as a solid tube, with flanking regions shown as an outline. The arrow shows the transcription start site. The polymer is coloured according to the simulation input data: red indicates active protein binding sites (DHS), blue polycomb protein binding sites, and yellow are open chromatin (H3K27ac).

shot for the Z-138 case shows a configuration where a loop forms between a DHS within the gene body and a DHS in the upstream region. Since there are more DHS across the gene in Z-138 there are more possibilities for loops to form: smaller R_g values are therefore more likely and there is less variation.

The oncogene structure of *CCND1* is driven by chromatin states

We have observed that in these cell lines the genome rearrangement is accompanied by an extensive remodelling of chromatin (both chromatin states and DHS). In particular, in both U266 and Z-138 a H3K4me3 broad domain containing several DHS appears over *CCND1*. A likely scenario is that after the rearrangement, the *IGH* super-enhancers act as a strong recruiter of transcription factors, chromatin remodelers, etc., to the region. The resulting changes then in turn disrupt the wider 3D structure of the locus, leading to dysregulation of *CCND1*. The HiP-HoP model provides a unique opportunity to examine these two effects in isolation: the genomic rearrangement itself, and the remodelling of chromatin states.

Figure 5A shows Hi-C from a simulation where chromatin state and DHS data for GM12878 cells were used as input, but the hg19_u266 genomic insert rearrangement was included. In Fig. 5C we compare 4C results from GM12878 simulations with and without the insert; this shows that adding the insert has only a small effect on the interaction profile of the gene promoters. The profiles for *TPCN2* and *MYEOV* are unchanged. Also, despite the insert containing a DHS within a H3K27ac region, there is little interaction with the nearby *CCND1* promoter (which has a polycomb chromatin state). On the other hand, there is some interaction between the insert DHS and the *LTO1* promoter. Using active protein binding as a predictor of gene expression (Fig. 5E) suggests that without the chromatin remodelling at *CCND1* which is observed in real U266 cells, the presence of the insert alone would not lead to Cyclin D1 upregulation.

In Figs. 5B and D we show similar results from simulations where U266 input data is used with the intact hg19 genome (no insert). This allowed us to study the effect of the chromatin remodelling in the absence of the insert. We found that in this case the *CCND1* promoter still showed interaction peaks across the *CCND1* gene body, at regions downstream (in and around *LTO1*), and to a lesser extent at regions upstream (at *MYEOV* and a region halfway between the two). The presence of the insert has a bigger effect than the GM12878 case, but it is still modest. Beyond the loss of interactions between *CCND1* and the DHS within the insert itself, the main effect of removing the insert is an increase in interaction between *CCND1* and upstream DHS (including at *MYEOV*), consistent with their reduced genomic separation. The simulations predict that very little change in

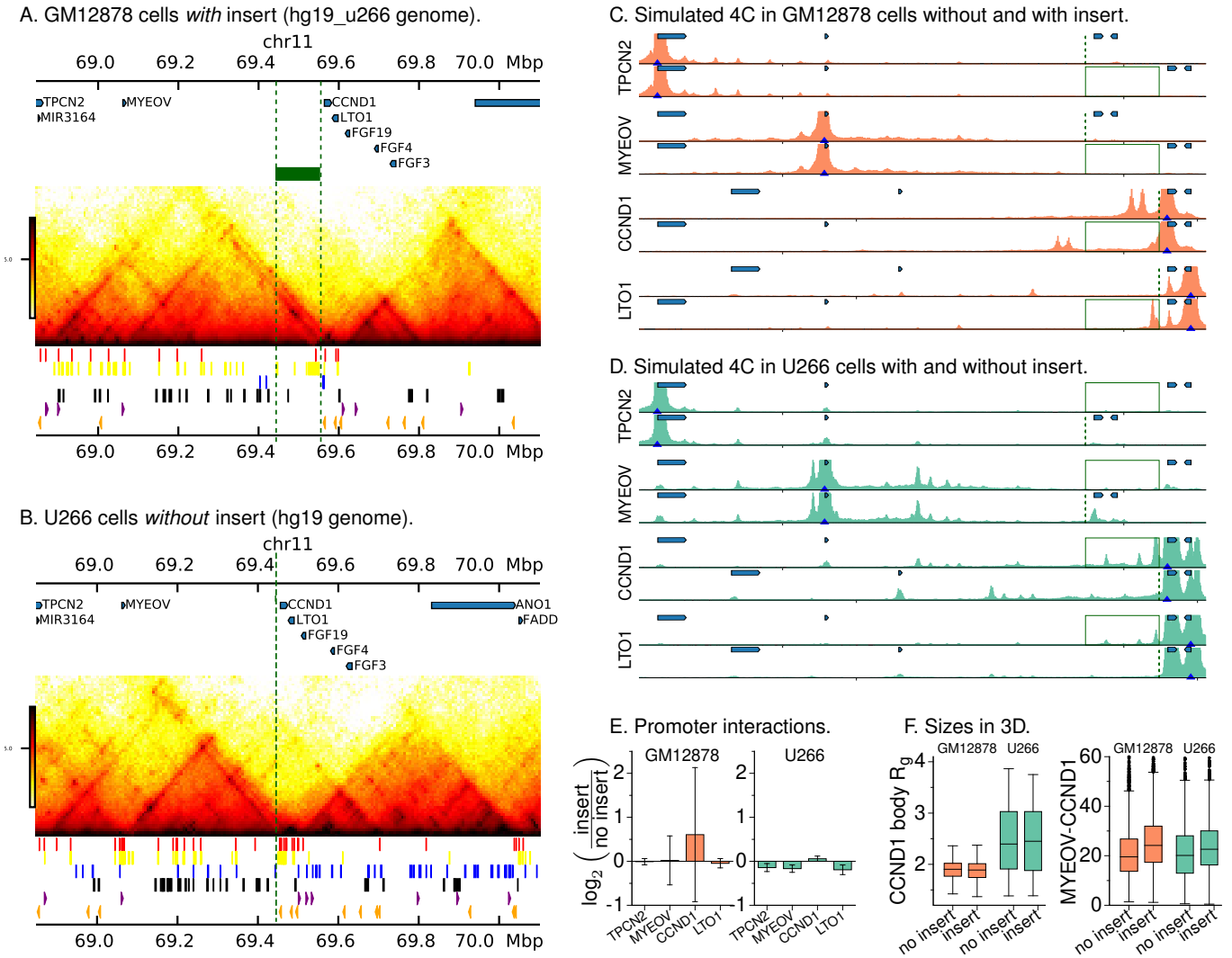

Figure 5. **A.** Simulated Hi-C data are shown for the rearranged (insert) *CCND1* locus (hg19_u266 genome), where the native chromatin state data (obtained from GM12878 cells) were used. These input data are shown below the colour map according to the scheme used in Fig. 2A. **B.** Simulated Hi-C data are shown for the intact (no insert) *CCND1* locus, where ectopic chromatin state data (obtained from U266 cells) were used. Again, the input data are shown below the map. **C.** Simulated 4C are shown for simulations using GM12878 chromatin states both with and without the insert. **D.** Simulated 4C are shown for simulations using U266 chromatin states both with and without the insert. **E.** As a rough predictor of expression levels, the proportion of cells where the gene promoter was bound by an active protein was measured. The plot shows the $\log_2$ ratio of values from simulations with and without the insert using input data from the indicated cell line. Error bars show the standard error. F. Box plots showing the distribution of: left, the radius of gyration of the *CCND1* gene body; and right, the 3D separation of *MYEOV* and *CCND1*. Data from simulations with and without the insert using input data from the indicated cell lines are shown.

gene expression level would result from removing the insert while keeping the U266 chromatin states (Fig. 5E). Figure 5F shows distributions of 3D single cell measurements. The 3D size of the *CCND1* gene changed very little due to the presence or absence of the insert, but there was a large difference between the cell lines. The 3D separation of *MYEOV* and *CCND1* showed relatively little change between the cell lines, but was slightly larger for cells where the insert was present (consistent with the increased genomic separation). Together these results suggest that it is chromatin remodelling which drives the changes in the 3D structure of the locus in terms of promoter-enhancer interactions, and this in turn drives up-regulation of Cyclin D1. The presence of the insert region otherwise has a more modest effect.

Conclusions

In this work we have adapted the HiP-HoP simulation model to study chromatin 3D structure and the effect of genome rearrangements in malignant and non-malignant B-cells. HiP-HoP simulations predict 3D structures from DNA accessibility and chromatin state data. We used the simulated structures to generate Hi-C and 4C-like population level data, and single cell-like measurements. Importantly Hi-C data are not used as an input, so in this sense the model is truly predictive for 3D structures. Here, by "rearranging" the input data we generate predictions for the effect of genome rearrangements which are found in MCL and MM cell lines.

We first confirmed that the HiP-HoP model gave good pre-

dictions for the *CCND1* gene locus by performing simulations of the healthy B-cell derived GM12878 lymphoblastoid cell line. The TAD pattern observed in Hi-C data [5] is clearly reproduced by the simulations, with boundaries at CTCF sites. In these cells *CCND1* is only expressed at very low levels; the promoter region has a polycomb associated chromatin state, and our simulations predicted interactions between the promoter and other polycomb regions.

Further simulations predicted that the TAD structure around *CCND1* is preserved in the U266 MM cell line which possesses a rearrangement where a super-enhancer from the *IGH* locus is inserted upstream of *CCND1* (Fig. 2). This is consistent with previously published low-resolution Hi-C data in that cell line [41]. In these cells the chromatin around *CCND1* is remodelled: several DHS are established within the gene body and promoter, and a broad region gains an active promoter chromatin state (H3K4me3 and H3K27ac). The simulations predicted that the *CCND1* promoter interacts with these new DHS in the gene body, as well as with a number of DHS within the inserted *IGH* region and the neighbouring gene *LTO1* (interactions which are not present in GM12878 cells).

We also performed simulations of the MCL cell line Z-138, where *CCND1* translocates with chromosome 14 becoming juxtaposed with the *IGH* Eμ enhancer (Fig. 3). In this cell line a TAD boundary formed by a cluster of CTCF binding sites downstream of *CCND1* was preserved after the translocation, as was a boundary to the left of the *IGH* super-enhancers (presumably in healthy cells this isolates the super-enhancers from other nearby genes). In other words, the simulations predict a fusion of the *CCND1* and *IGH* regions into a new TAD (an oncogenic TAD fusion). In this new arrangement the promoters of *CCND1* and *LTO1* readily interacted with each other and the proximate Eμ and Eδ super-enhancers, while weaker interactions were observed with the more distant Eα1 and Eα2.

In vivo, the genomic rearrangements are accompanied by chromatin remodelling; within our simulation scheme it was also possible to examine the effect of a rearrangement in the absence of remodelling. Simulations of GM12878 cells with the Eα1 enhancer DNA inserted *in silico* upstream of *CCND1* (with chromatin states otherwise unchanged) showed only very minor changes in terms of 3D structure compared with the unaltered genome (Fig. 5). We also performed a simulation using U266 chromatin states, but without the insert (*i.e.*, the epigenomic rearrangement is included, but not the genomic rearrangement). This showed little change compared to the U266 case with the insert. Together this suggests that it is the local chromatin remodelling, rather than the proximity of *CCND1* to the *IGH* enhancers *per se*, which drives gene deregulation. Or in other words, a genomic translocation leads to an epigenomic translocation, which drives Cyclin D1 over-expression. The remodelling includes the appearance of an H3K4me3 broad domain over *CCND1*; the results here support our previous work suggesting that such broad domains are associated with super-enhancer hijacking [9]. Importantly, the chromosomal locations which are predicted by the model to strongly interact with *CCND1* could be used for targeting in future experiments which aim to uncover the mechanisms through which broad domains are generated.

In summary, our work strongly suggests that genome rearrangements drive a subsequent epigenomic rearrangement, which in turn leads to deregulation and proto-oncogene activation. We have demonstrated that polymer physics-based modelling can be useful for understanding the structural consequences of genome rearrangements, and can help to focus future experimental work. It would be interesting to see if in the future, such simulations could also help us understand the mechanisms behind the epigenomic rearrangement. This would clearly be important for the development of any therapies which seek to interfere with that process.

Methods

We used the HiP-HoP model as detailed in Ref. [14]; full details are given in the Supplementary Methods. In brief, chromatin is represented as a chain of beads (each representing a 1 kbp region), and we evolve the configuration of this chain using a molecular dynamics scheme and the LAMMPS simulation software [42]. To improve simulation efficiency, in each simulation we include 40 Mbp of chromatin (40,000 beads) which includes 11 copies of the region of interest; for each region we perform two such simulations to generate 4400 individual configurations each of which can be said to represent a single cell. The chromatin concentration is roughly matched to that of a typical nucleus. From the configurations we generate simulated Hi-C and 4C data and single cell-like distance measurements. As detailed in the text, three different mechanisms are included to drive the chain into specific structures: diffusing bridge forming proteins, loop extrusion, and a heteromorphic polymer structure. DNase hypersensitive sites (DHS) are used to identify binding sites for an active protein (we use the simplifying assumption that all DHS are the same and bind this protein, which represents a general complex of polymerase and transcription factors). We use chromatin state data to identify binding sites for two species of repressive protein (*e.g.*, representing HP1 and polycomb repressive complexes), and to identify regions which have an open chromatin structure. ChIP-seq data for CTCF is used to identify direction dependent anchor sites for loop extrusion. Full details of the input data treatment and publicly available data sets used are given in the Supplementary Methods.

Author contributions

AM and BAW performed the Z-138 break-point mapping and sequencing; DK performed bioinformatics analysis and cell line qPCR experiments. NK performed cell line qPCR experiments and FISH probe generation. CAB designed and performed the simulations. DR, LJR and CAB designed the research and wrote the paper. All authors contributed to writing the paper.

Acknowledgements

Work in the DR lab is supported by a Wellcome Trust Seed Award in Science (206103/Z/17/Z). BMJ acknowledges funding from FEDER / Ministry of Science and Innovation - Spanish State Research Agency under the project RTI2018-094788-A-I00, and La Caixa Banking Foundation under the r project LCF/BQ/PI19/11690001. Work in the LJR lab was supported by CCLG Little Princess Trust (NK) and an MRC DiMeN DTP studentship (DK). CAB acknowledges support from European Research Council (ERC CoG 648050 THREEDCELLPHYSICS).

References

[1] H. K. Norton and J. E. Phillips-Cremins. *Crossed wires: 3D genome misfolding in human disease*. J Cell Biol **216**, 3441 (2017).

[2] D. Jung, et al. *Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus*. Annual Review of Immunology **24**, 541 (2006).

[3] E. N. KONTOMANOLIS, et al. *Role of Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review*. Anticancer Research **40**, 6009 (2020).

[4] J. R. Dixon, et al. *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature **485**, 376 (2012).

[5] S. S. Rao, et al. *A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping*. Cell **159**, 1665 (2014).

[6] D. Hnisz, et al. *Super-enhancers in the control of cell identity and disease*. Cell **155**, 934 (2013).

[7] Y. Gong, et al. *Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries*. Nature Communications **9**, 542 (2018).

[8] P. Thandapani. *Super-enhancers in cancer*. Pharmacology & Therapeutics **199**, 129 (2019).

[9] A. Mikulasova, et al. *Epigenomic translocation of H3K4me3 broad domains following super-enhancer hijacking*. bioRxiv 2020.02.12.938563 (2020).

[10] J. M. Engreitz, V. Agarwala, and L. A. Mirny. *Three-Dimensional Genome Architecture Influences Partner Selection for Chromosomal Translocations in Human Disease*. PLOS ONE **7**, 1 (2012).

[11] V. Roukos and T. Misteli. *The biogenesis of chromosome translocations*. Nature Cell Biology **16**, 293 (2014).

[12] S. Bianco, et al. *Polymer physics predicts the effects of structural variants on chromatin architecture*. Nature Genetics **50**, 662 (2018).

[13] K. C. Akdemir, et al. *Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer*. Nature Genetics **52**, 294 (2020).

[14] A. Buckle, et al. *Polymer Simulations of Heteromorphic Chromatin Predict the 3-D Folding of Complex Genomic Loci*. Molecular Cell **72**, 786 (2018).

[15] C. A. Brackey, D. Marenduzzo, and N. Gilbert. *Mechanistic modeling of chromatin folding to understand function*. Nature Methods **17**, 767 (2020).

[16] C. T. Watson and F. Breden. *The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease*. Genes & Immunity **13**, 363 (2012).

[17] A. Thibodeau, et al. *Chromatin interaction networks revealed unique connectivity patterns of broad H3K4me3 domains and super enhancers in 3D chromatin*. Scientific Reports **7**, 14466 (2017).

[18] S. Park, et al. *Broad domains of histone H3 lysine 4 trimethylation in transcriptional regulation and disease*. The FEBS Journal **287**, 2891 (2020).

[19] F. Serra, et al. *Restraint-based three-dimensional modeling of genomes and genomic domains*. FEBS Letters **589**, 2987 (2015).

[20] F. C. Mills, et al. *DNase I Hypersensitive Sites in the Chromatin of Human Mu Immunoglobulin Heavy-Chain Genes*. Nature **306**, 809 (1983).

[21] C. Chen and B. K. Birshtein. *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**, 1310 (1997).

[22] F. C. Mills, et al. *Enhancer Complexes Located Downstream of Both Human Immunoglobulin Calpha Genes*. The Journal of Experimental Medicine **186**, 845 (1997).

[23] The ENCODE Project Consortium. *An integrated encyclopedia of DNA elements in the human genome*. Nature **489**, 57 (2012).

[24] C. A. Brackley, et al. *Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization*. Proc. Natl. Acad. Sci. USA **110**, E3605 (2013).

[25] C. A. Brackley, et al. *Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains*. Nucleic Acids Res. **44**, 3503 (2016).

[26] C. A. Brackley. *Polymer compaction and bridging-induced clustering of protein-inspired patchy particles*. J. Phys. Condens. Matter **32**, 314002 (2020).

[27] G. Fudenberg, et al. *Formation of Chromosomal Domains by Loop Extrusion*. Cell Reports **15**, 2038 (2016).

[28] A. L. Sanborn, et al. *Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes*. Proceedings of the National Academy of Sciences USA **112**, E6456 (2015).

[29] I. F. Davidson, et al. *DNA loop extrusion by human cohesin*. Science (2019).

[30] J. Ernst, et al. *Mapping and analysis of chromatin state dynamics in nine human cell types*. Nature **473**, 43 (2011).

[31] E. Carrillo-de Santa-Pau, et al. *Automatic identification of informative regions with epigenomic changes associated to hematopoiesis*. Nucleic Acids Research **45**, 9244 (2017).

[32] V. I. Risca, et al. *Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping*. Nature **541**, 237 (2017).

[33] Z. Zhao, et al. *Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions*. Nature Genetics **38**, 1341 (2006).

[34] N. Krietenstein, et al. *Ultrastructural Details of Mammalian Chromosome Architecture*. Molecular Cell **78**, 554 (2020).

[35] L. Lu, et al. *Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases*. Molecular Cell **79**, 521 (2020).

[36] C. A. Davis, et al. *The Encyclopedia of DNA elements (ENCODE): data portal update*. Nucleic Acids Research **46**, D794 (2017).

[37] J. H. A. Martens and H. G. Stunnenberg. *BLUEPRINT: mapping human blood cell epigenomes*. Haematologica **98**, 1487 (2013).

[38] C. A. Brackley, et al. *Complex small-world regulatory networks emerge from the 3D organisation of the human genome*. bioRxiv 2020.05.12.091041 (2020).

[39] D. J. Downes, et al. *High-resolution targeted 3C interrogation of cis-regulatory element organization at genome-wide scale*. Nature Communications **12**, 531 (2021).

[40] P. Jares and E. Campo. *Advances in the understanding of mantle cell lymphoma*. British journal of haematology **142**, 149 (2008).

[41] P. Wu, et al. *3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations*. Nature Communications **8**, 1937 (2017).

[42] S. Plimpton. *Fast Parallel Algorithms for Short-Range Molecular Dynamics*. J. Comp. Phys. **117**, 1 (1995).