Unravelling higher order genome organisation [working title]

Results 3: Domain boundaries

Benjamin L. Moore

July 8, 2015

igmm
INSTITUTE OF GENETICS
& MOLECULAR MEDICINE

MRC Medical Research Council

THE UNIVERSITY OF EDINBURGH

CANCER RESEARCH UK

1 CHROMATIN DOMAIN BOUNDARIES

1.1 INTRODUCTION

Multiple studies have defined chromatin domains of different types, for example: chromosome compartments;[1] topological associating domains (TADs);[2] contact and loop domains;[3] physical domains;[4,5] and others.[6] The existence of these domains necessitates "boundary regions" either between consecutive domains or bookending more sparsely-positioned domains, however the functional relevance of said boundary regions is still open to debate.

In their study of topological domains, Dixon *et al.* identified average enrichments over TAD boundary regions in both human and mouse for various features including CTCF and Pol2.[2] Boundaries were also enriched for signs of active transcription, such as with the histone modification H3k36me3. These results, coupled with an observable enrichment for promoters at domain boundaries, have lead to the theory that boundaries may act as an additional layer of transcriptional control,[7] however an alternative theory could be that looping between enhancer elements and promoters results in an observable boundary through C-method experiments.[3] Another non-exclusive explanation is that if chromatin domains represent co-regulatory regions as is widely thought,[7–9] boundaries themselves could be mere side-effects and as such of limited biological interest.

An obvious experiment to resolve these opposing theories would be to delete a predicted boundary region and test for local changes in both contacts and expression. Such an experiment was performed on a region of the human X-chromosome containing the genes encoding the dosage-compensation long non-coding RNAs Xist and Tsix, which are separated by a TAD boundary.[10] This study found that while histone modifications within the body of a TAD could be removed without affecting the structure, deletion of a boundary did have an effect and lead to increased intradomain contacts.[10] Surpsingingly however, this effect was not total and some observable barrier remained, lending evidence that TADs may be centrally constrained, rather than by their borders.[10]

A second experiment used CRISPR genome editing to link TAD boundary changes with limb development disorders,[11] indicating that boundary changes could provide an underlying explanation for pathogenic non-coding structural variants.[12] Similarly, domain boundaries on X-chromosomes were found to be weakened following the disruption of condensation binding sites.[13] Together these studies suggest a complex scenario whereby TAD boundaries are an important structural feature, yet do not fully explain domain partitioning.

Computational analysis of boundaries has emerged during the time this work was completed. Border "strength", here defined by the ratio of total intra:inter-domain contacts, was found to correlate with increased occupancy of a combination of bound architectural proteins.[14]

Many questions remain about chromatin boundaries. For example, are the observed enrichments persistent across cell types and how do they compare across organisation strata, such as compartments and TADs? Through computational analysis of the set of boundaries re-called from published datasets, we can investigate these questions and probe boundary enrichments across a broad array of locus-level chromatin features.

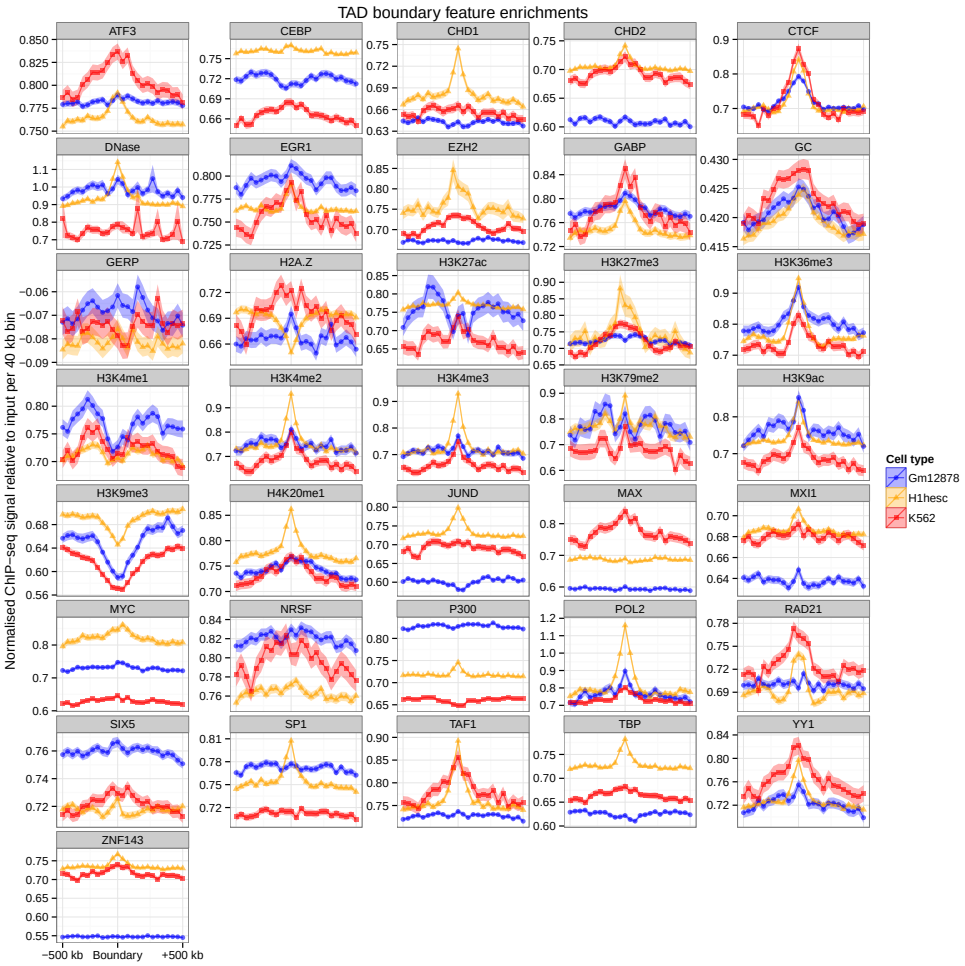

Figure 1: Spectra of TAD boundary enrichments and depletions. Placeholder

1.2 TAD AND COMPARTMENT BOUNDARIES

The mammalian genome is organized into TADs, predominantly self-interacting chromatin domains, with boundary regions reportedly associated with pronounced peaks and troughs of particular features within 500 kb of the predicted boundary.[2] Exploration of this phenomenon using a set of 24 mouse ESC chromatin features (and a smaller number of human ESC features) reportedly revealed enrichment peaks of CTCF, H3K4me3 and H3K36me3, as well as a pronounced dip in H3K9me3, suggesting that high levels of transcription may contribute to boundary formation.[2] However, it was unclear whether other features show unusual patterns in TAD boundary regions, and whether the constellation of features involved changes between cell types. The features associated with boundaries separating A and B compartments calculated from Hi-C eigenvectors have not been studied to our knowledge. The datasets assembled here, consisting of 35 matched chromatin features across three cell types, allow us to conduct the first comparative study of the constituents of human TAD and compartment boundary regions.

We derived TAD boundaries according to established methods (see Methods XX) for all three cell types under study. We then sought evidence for significantly enriched or depleted features at TAD boundary regions using a conservative approach (a nonparametric statistical test and Bonferroni multiple testing correction, see Methods XX).

Our findings confirmed the previously reported peaks (CTCF and POL2) and dip (H3K9me3) in ESC data, but also revealed substantial heterogeneity between cell types. CTCF binding was found enriched at TAD boundaries

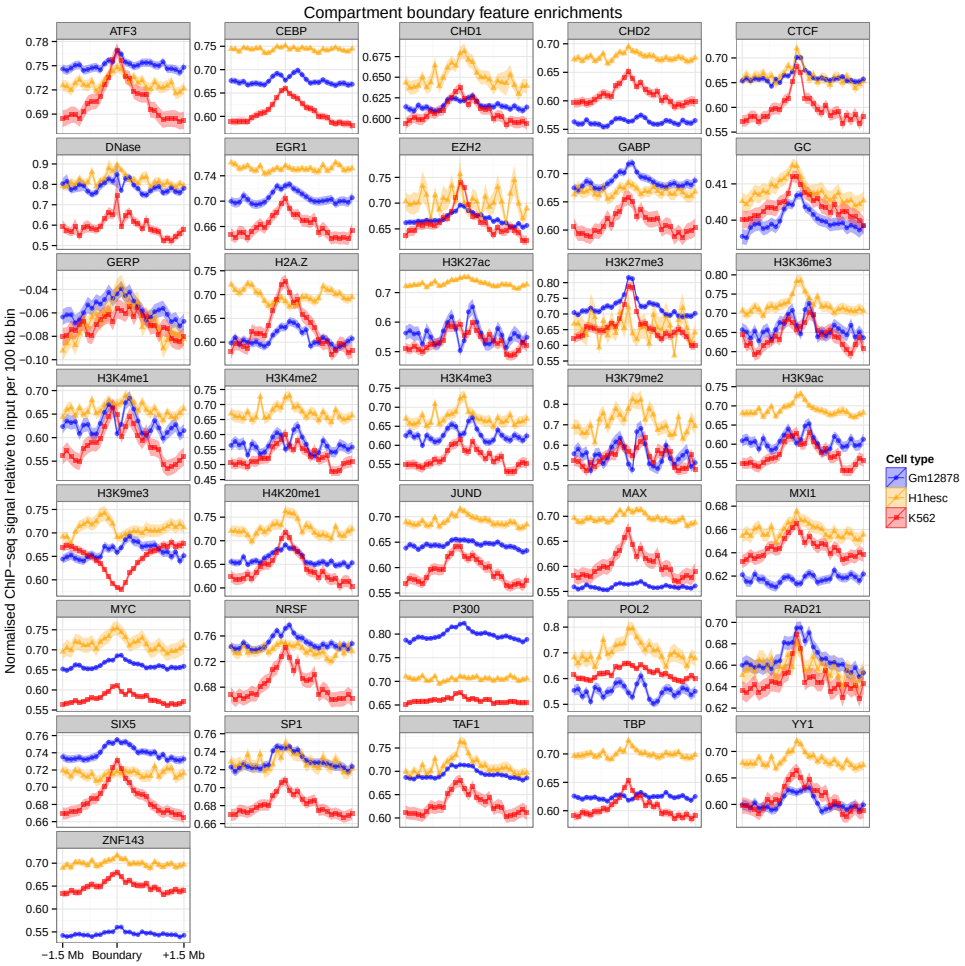

Figure 2: Spectra of compartment boundary enrichments and depletions. Placeholder

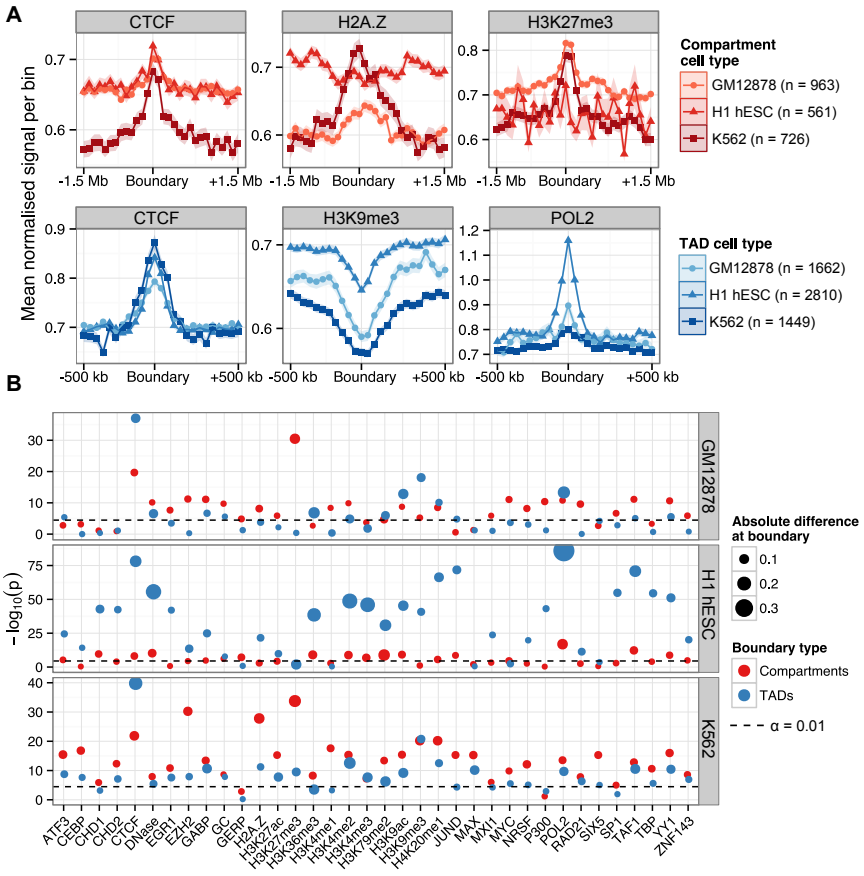

Figure 3: Compartment and TAD boundary enrichment summary in three human cell types. Placeholder

across all cell types, but other features, including H3K36me3 and H3K4me3, show dramatic peaks of enrichment in H1 hESC cells that are not seen consistently in other cell types (Figure 6, Additional file 1: Figure S12). Although the dip in H3K9me3 at TAD boundaries is seen in all cell types, the extent of the depletion varies and is weakest in H1 hESC cells. Many other features show significant, though often modest, enrichments in a particular cell type. However, overall the complexity of TAD boundaries (measured as the number of strongly enriched features) is notably higher in H1 hESC than in the other two, more differentiated, cell types (Figure 6), involving large increases in the binding of sequence specific factors such as SP1 and JUND.

Across all three cell types several features demonstrate consistent and statistically significant patterns at TAD boundaries (Figure 6, Figure S12), including peaks associated with active transcription of genes (POL2, H3K9ac) and dips in H3K9me3, as previously reported.[2] However other novel feature peaks of interest emerge across cell types, such as peaks of H4K20me1, a modification previously implicated in chromatin compaction.[15] We also observe consistent increases in GC content at TAD boundaries, at a scale that is difficult to reconcile with the presence of smaller-scale features such as repeat elements or CpG islands (Additional file 1: Figure S12).

Where neighbouring genomic regions occupy contrasting A and B nuclear compartments, the disparity implies the presence of a boundary region. Putative compartment boundaries were identified by using an HMM to infer the state sequence of A/B compartments across the genome based on observed principal component eigenvectors. Analogously to the TAD boundary analysis we then sought significant enrichments or depletions in 36 chromatin features over these compartment boundaries (Figure 6, Figure S13). Compartment boundaries display similar spectra of enrichments to previously studied TAD boundaries[2] but at lower resolution, reflecting

the different scales of these levels of organization (Figure 6B, Figure S13). Peaks associated with active promoters (POL2, TAF1, H3K9ac) are again evident. Parallel enrichments of CTCF, YY1 and H4K20me1 are also seen at compartment boundaries, as they were for TAD boundaries, in each cell type under study. In addition, compartment boundaries show enrichments of H3K79me2, which is known to play critical roles in cellular reprogramming.[16] Remarkably, H3K79me2 has also recently been shown to mark the borders of small (hundreds of bp) regions of open chromatin.[17] Thus there may be similarities in chromatin compaction boundaries at very different scales.

Certain features show intriguing contrasts between cell types the histone variant H2A.Z lacks any trace of enrichment at H1 hESC compartment boundaries, but is significantly enriched in the other two cell types (Figure 6A), consistent with reports describing H2A.Z relocation during cellular differentiation.[18] Compartment boundaries also show enrichment for the cohesin complex subunit RAD21 in the two hematopoietic cell types , and cohesin is another factor implicated in modulating nuclear architecture in partnership with CTCF.[19] Various other enrichments with very modest effect sizes are also evident at compartment boundaries (Figure 6B, Figure S13). In contrast to TAD boundaries, the composition of compartment boundaries appears least complex in H1 hESC, relative to the other two cell types. Overall compartment and TAD boundaries are associated with overlapping spectra of chromatin features across cell types. These involve DNA binding proteins implicated in chromosome architecture (CTCF, YY1, RAD21), but also implicate the initiation and repression of transcription as critical to boundary formation. However these two boundary classes occur at different scales, with patterns of informative features typically spanning regions up to 500 Kb for TAD boundaries, and patterns associated with compartment boundaries often spanning more than 1 Mb.

1.2.1 CTCF and YY1

Significant peaks in YY1 are evident in all cell types, which is intriguing given the evidence that YY1 and CTCF cooperate to affect long distance interactions.[20] Co-binding of CTCF with YY1 has also been shown to identify a subset of highly conserved CTCF sites.[21] Co-binding of CTCF and YY1 may also therefore be a contributing factor in the establishment of TAD boundaries, which appear to be broadly conserved across mammals.[2] To test this, we split our sets of TAD boundaries into those possessing ChiP-seq peaks (region peaks called by ENCODE[22]) for CTCF, YY1, both CTCF and YY1 (overlapping peaks) and neither. We then tested each boundary subset for genome-wide enrichments of the other features in our dataset (Figure S14). Unexpectedly, we found that boundaries marked by YY1 (without overlapping CTCF peaks) were generally most strongly-enriched for other features in our dataset. We also found that boundaries lacking both CTCF and YY1 peaks showed instead the strongest enrichments for RAD21 in each cell type (Figure S14), reinforcing previous findings that describe the distinct influences of CTCF and cohesin in organizing chromatin structure.[19,23,24]

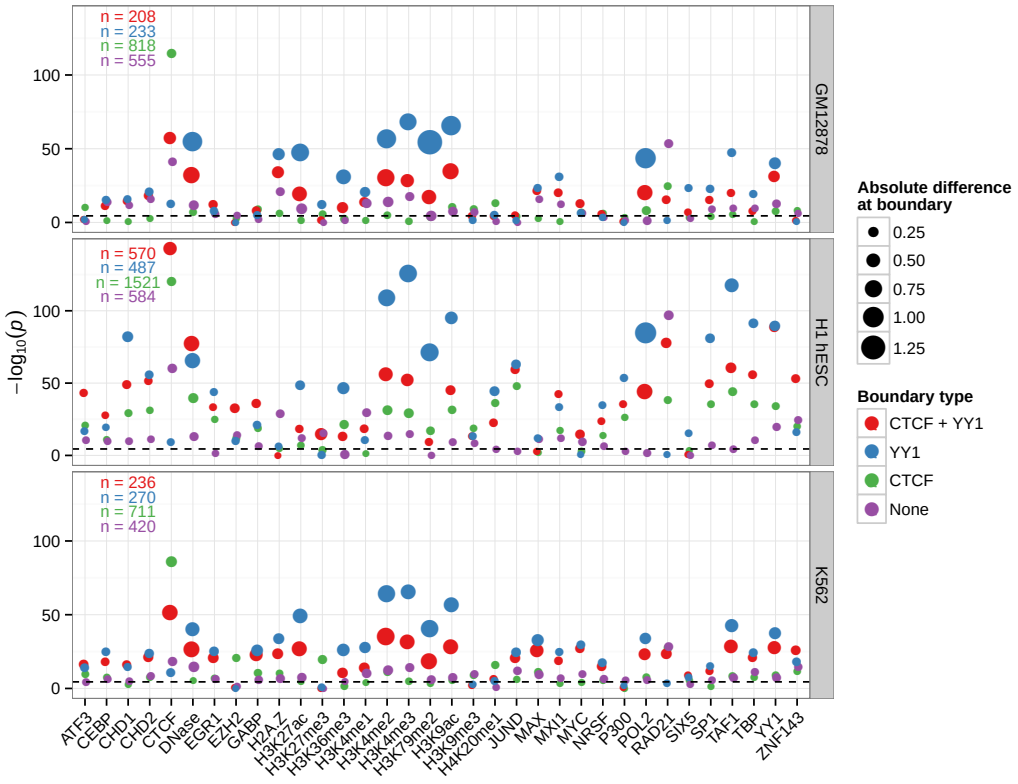

Figure 4: Distinct enrichments of CTCF and YY1 boundaries. Placeholder

1.2.2 Repeats

1.3 DE NOVO BOUNDARY PREDICTION

1.4 METATAD BOUNDARIES

1.5 OTHER BOUNDARIES

1.5.1 Giemsa bands

1.5.2 Superboundaries

Thus far compartment and TAD boundaries have been considered separately, however it is of interest to consider how these boundary regions interact across scales. Open questions remain about the co-occurence of these two boundary regions, and whether

REFERENCES

[1] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**(5950): 289–93.

[2] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**(7398): 376–80.

[3] Rao S, Huntley M, Durand N, Stamenova E, Bochkov I, Robinson J, Sanborn A, Machol I, Omer A, *et al.* (2014) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, **159**(7): 1665–1680.

[4] Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, **148**(3): 458–72.

[5] Hou C, Li L, Qin ZS, Corces VG (2012) Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Molecular Cell*, **48**(3): 471–484.

[6] Filippova D, Patro R, Duggal G, Kingsford C (2014) Identification of alternative topological domains in chromatin. *Algorithms for molecular biology : AMB*, **9**: 14.

[7] Sexton T, Cavalli G (2015) Review The Role of Chromosome Domains in Shaping the Functional Genome. *Cell*, **160**(6): 1049–1059.

[8] Le Dily F, Bau D, Pohl a, Vicent GP, Serra F, Soronellas D, Castellano G, Wright RHG, Ballare C, *et al.* (2014) Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes & Development*, **28**(19): 2151–2162.

[9] Nora EP, Dekker J, Heard E (2013) Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? *BioEssays*, **35**(9): 818–828.

[10] Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**(7398): 381–5.

[11] Lupiáñez D, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz J, *et al.* (2015) Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, pp. 1012–1025.

[12] Ren B, Dixon J (2015) A CRISPR Connection between Chromatin Topology and Genetic Disorders. *Cell*, **161**(5): 955–957.

[13] Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ (2015) Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*.

[14] Van Bortle K, Nichols MH, Li L, Ong CT, Takenaka N, Qin ZS, Corces VG (2014) Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome biology*, **15**(5): R82.

[15] Evertts AG, Manning AL, Wang X, Dyson NJ, Garcia Ba, Coller Ha (2013) H4K20 methylation regulates quiescence and chromatin compaction. *Molecular biology of the cell*, **24**(19): 3025–37.

[16] Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, *et al.* (2012) Chromatin-modifying enzymes as modulators of reprogramming. *Nature*, **483**(7391): 598–602.

[17] Chai X, Nagarajan S, Kim K, Lee K, Choi JK (2013) Regulation of the boundaries of accessible chromatin. *PLoS genetics*, **9**(9): e1003778.

[18] Ku M, Jaffe JD, Koche RP, Rheinbay E, Endoh M, Koseki H, Carr SA, Bernstein BE (2012) H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome biology*, **13**(10): R85.

[19] Zuin J, Dixon JR, van der Reijden MIJa, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPC, van de Werken HJG, Knoch Ta, *et al.* (2013) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, pp. 1–6.

[20] Atchison ML (2014) Function of YY1 in Long-Distance DNA Interactions. *Frontiers in immunology*, **5**(February): 45.

[21] Schwalie PC, Ward MC, Cain CE, Faure AJ, Gilad Y, Odom DT, Flicek P (2013) Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome biology*, **14**(12): R148.

[22] ENCODE (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**(7414): 57–74.

[23] Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, Lenhard B, Giorgetti L, Heard E, *et al.* (2013) Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome research*, **23**(12): 2066–77.

[24] Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong CT, Hookway Ta, Guo C, *et al.* (2013) Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell*, **153**(6): 1281–95.