# 1 REANALYSIS OF HI-C DATASETS

## 1.1 INTRODUCTION

Since the initial publication of the Hi-C technique in 2009,[1] there has been rapid advancement of both the technique itself and the resolution at which interaction frequencies have been analysed. From the proof-of-concept analysis at 1 megabase (Mb) and 100 kilobase (kb) resolution,[1] subsequent experiments achieved resolutions first of 40 kb[2], then 10 kb[3] and most recently 1 kb,[4] enabling bona fide genome-wide fragment-level analysis for the first time.

Such rapid progression in the field has resulted in a wide variety of public Hi-C datasets being available, albeit with differing qualities. With proper correction and at a suitable resolution, these interaction frequencies can be compared and contrasted both within and between species.

In this chapter we uniformly reprocessed publicly-available human Hi-C datasets in order to address fundamental questions about the stability of higher order genome organisation between cell populations from the same species. Previously Hi-C studies have compared across two samples, such as K562 against GM06990[1] or IMR90 against GM12878.[2] Here we make use of three Hi-C datasets corresponding to extensively-studied human cell lines: K562, GM12878 and H1 hESC. Together these make up the "Tier 1" cell lines studied by the ENCODE consortium,[5] hence have huge amounts of matched ChIP-seq and histone modification data available.

By combinatorial reanalysis of these cell-matched datasets, we can comprehensively investigate the quantitative relationships between higher order chromatin structure and locus-level chromatin features.

## 1.2 HI-C REPROCESSING

Each Hi-C dataset used in this work was reprocessed from raw sequencing reads using the same pipeline (Methods **??**). Briefly, raw sequencing reads were sourced from three different publications: Lieberman-Aiden *et al.*[1], Dixon *et al.*[2] and Kalhor *et al.*[6]. These reads were mapped to human genome build `hg19` using an iterative mapping procedure that maximised the number of uniquely mappable reads from each sample (Methods **??**).

Next a filtering step was applied, which removed those fragment pairs that were likely artifactual or erroneous (Methods **??**). A correction step was then applied, whereby biases such as mappability and GC content were removed to give each fragment equal visibility (Methods **??**). Overall these steps produced comparable maps of interaction frequency in different cell types, despite their differing origins (Fig. 1).
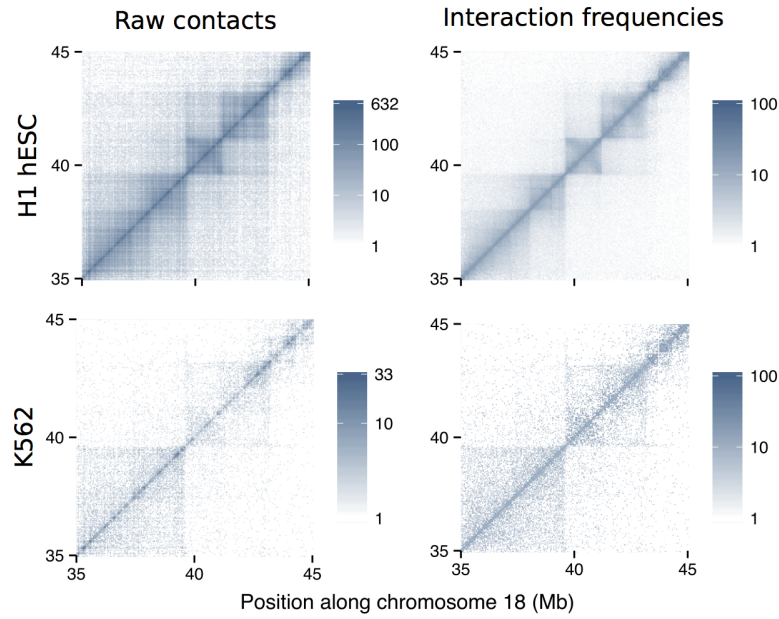
**Figure 1: Iterative correction converts raw counts to normalised interaction frequencies.** The sample with highest sequencing depth (H1 hESC) is shown alongside that with much lower sequencing (K562) both before and after iterative correction and normalisation procedures were applied (Methods **??**) at 40 kb resolution for a 10 Mb section of human chromosome 18. Fill gradients are on a $\log_{10}$ scale.

Figure 1 shows a 10 Mb region of chromosome 18 before and after filtering and normalisation in two different cell types. Self-interacting domains visible in the deeply-sequenced H1 hESC cell type also become more visible in the K562 cell type after normalisation. In addition many of the long-range and intra-domain contacts visible in each raw contact map are down-weighted during the normalisation procedure, indicating their prominence was enhanced by biases or other sources of noise in the experimental procedure (Fig. 1).

The relative sparsity of the K562 Hi-C contact map compared to that of the H1 hESC cell line should also be noted (Fig. 1). At the time much of this study was performed, deeply-sequenced Hi-C datasets for cell lines K562 and GM12878 were not available, thus the majority of analyses were performed at a lower resolution of 1 Mb to further reduce the impact of variable sequencing depth between cell lines (Methods **??**).

## 1.3 COMPARTMENT PROFILES

After uniformly reprocessing each Hi-C dataset and calling compartment eigenvector profiles (Methods **??**), we can compare these between three human cell lines. We find compartment profiles have a striking concordance (Fig. 2), despite the variable sources of both sample material and experimental data. This strong correlation of higher order chromatin structure (at the level of compartments) between three very different human cell types shows that genomic regions appear to be constitutively present in either the A or B compartments regardless of cell lineage.

This close correspondence also validates our approach of combining these different datasets, and suggests our uniform pipeline is successfully accounting for differences in sequencing
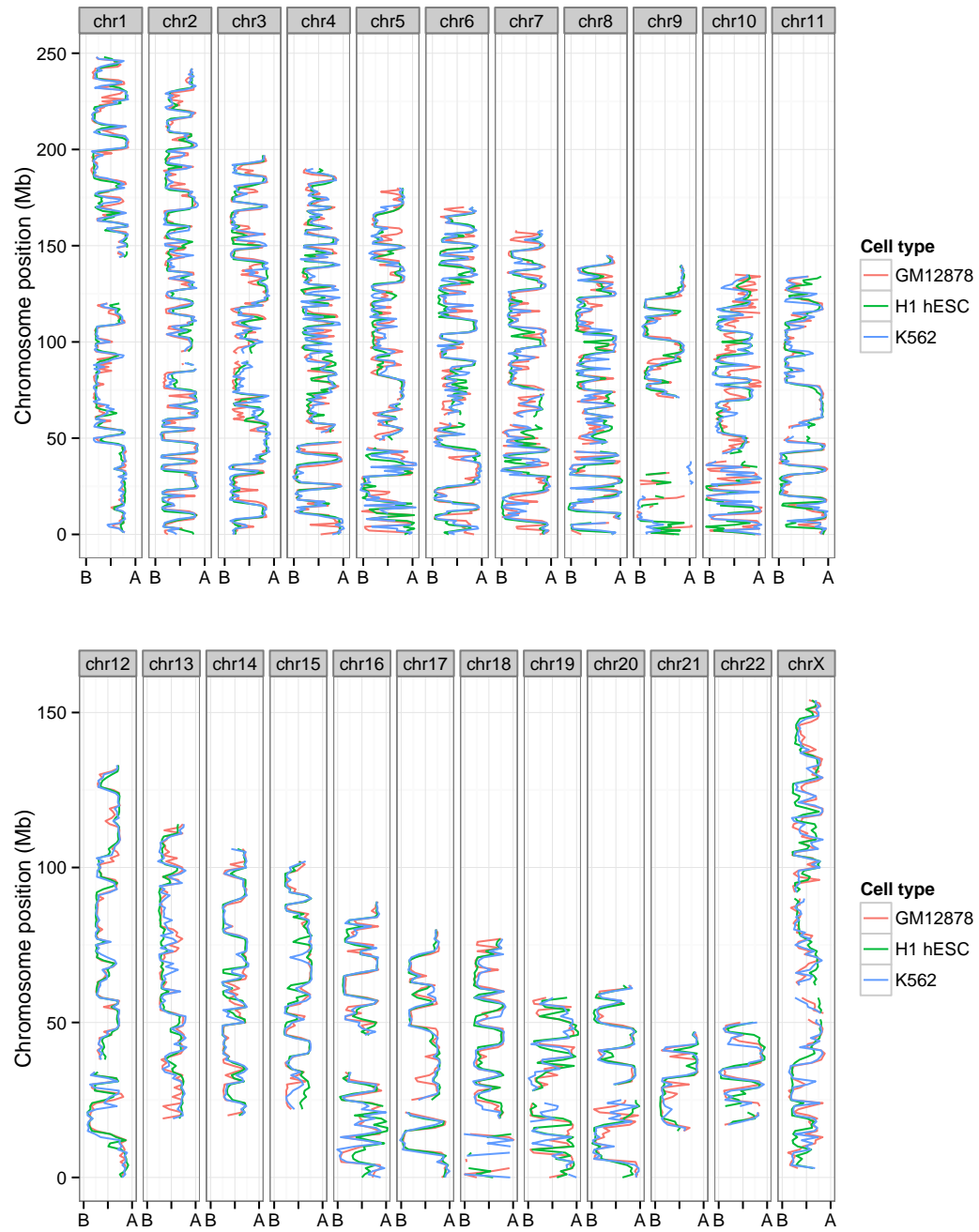
**Figure 2: Compartment profiles are observably well-correlated between human cell types and across all chromosomes.** Compartment eigenvectors are plotted along the lengths of each human chromosome (chrY and chrM are omitted) displaying strong concordance between three different human cell types.
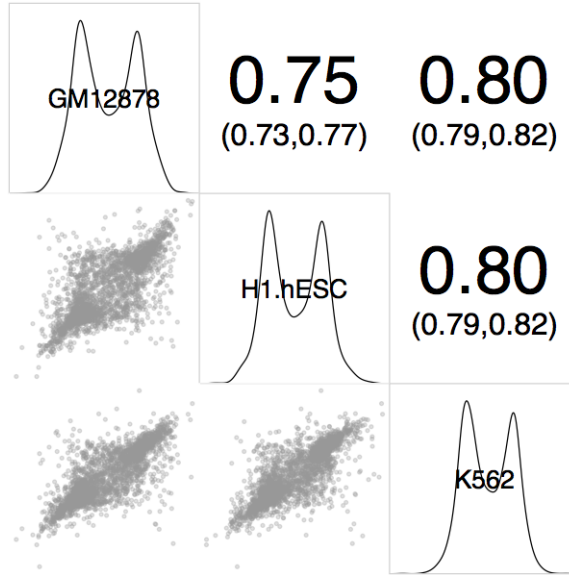
Figure 3: **Compartment eigenvectors are highly correlated between human cell types.** Megabase resolution compartment eigenvector values are shown in a plot matrix. *Upper triangle*: Pearson correlation coefficients between pairs, with 95% confidence intervals; *diagonal* Kernel density estimates of eigenvector values per cell type; *lower triangle*: *x-y* scatterplot of values.

depth and other batch effects. The pairwise Pearson correlation coefficients between these independent measures are all in the interval [.75, .8] (Fig. 3). Also of note is that each cell type independently shows a similar bimodal distribution of compartment eigenvector, indicative of the two distinct underlying A/B compartment states (Fig. 3).

## 1.4 DOMAIN CALLS

### 1.4.1 Compartments

The continuous compartment eigenvector is most commonly used as-is to classify A/B compartments by thresholding based on sign: typically the eigenvector is orientated such that positive values reflect A compartments and negative values B compartments.[1,6] However, given compartments are understood to be generally broad and alternating domains along a chromosome, often aligning with other large domains such as LADs, an improved classification method might penalise the calls of very short compartment calls, which may be the result of noise. For this reason, instead of using raw eigenvector values we consider observed values as emissions from unobserved underlying states (Fig. 4). We built a hidden Markov model (HMM; Methods **??**) to represent these states through a well-described probabilistic framework.[7]

Firstly we designed the HMM to have two unobserved states with univariate Gaussian distributed emissions (representing our A and B compartments; Fig. 5). To parameterise these states we used the iterative Baum-Welch algorithm, an Expectation-Maximisation procedure designed for HMMs. Having parameterised the HMM for each cell type, we then use the
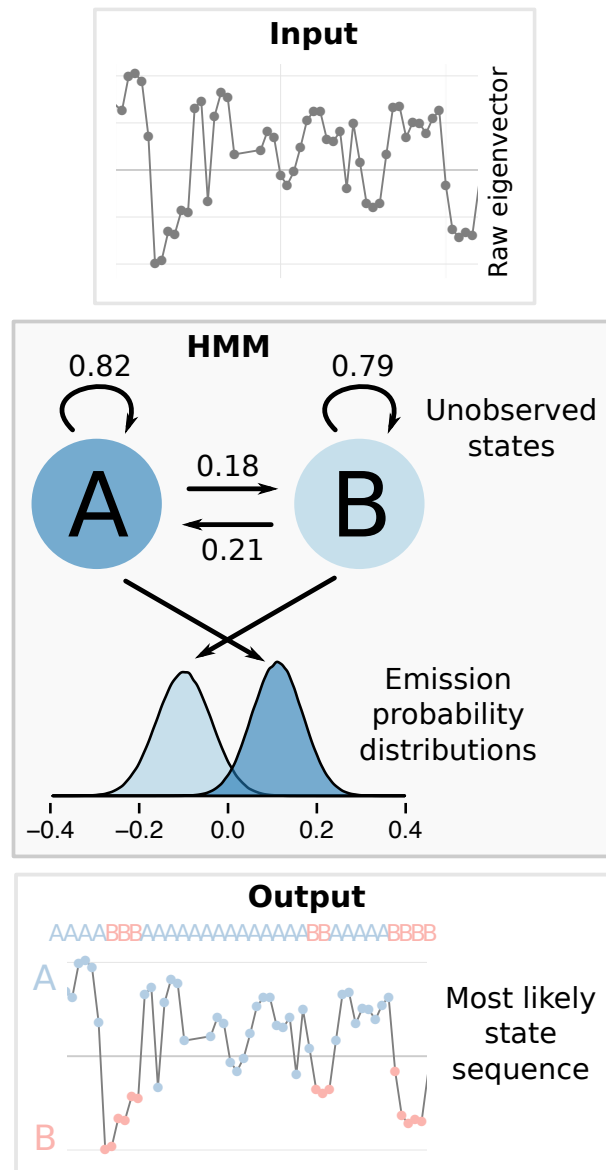
**Figure 4: Overview of HMM method for compartment calls.** Schematic of HMM method for compartment calls, showing transition and emissions distributions learned in the GM12878 cell type. A description of HMMs is given in Methods **??**. Emission distributions for each cell type are shown in Figure 5.
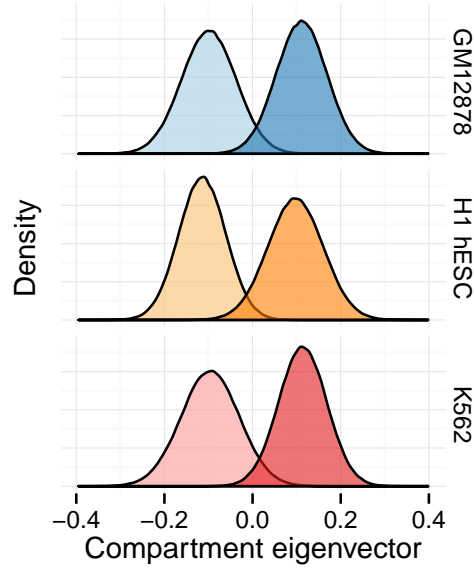
**Figure 5: Univariate Gaussian emission distributions for A and B unobserved HMM states.** Probability distributions for compartment eigenvector values per HMM state are shown for each cell type. Distributions centred below zero (lighter colours) show the distribution for the B compartment state, distributions centred above zero (darker) represent the A compartment state.

Viterbi algorithm to infer the most likely state sequence to have generated our observed data. This two-state sequence is then used to assign compartment identities to genomic bins. A schematic of this procedure is shown in Figure 4.

In practice, this acts to de-noise our compartment calls. Whereas single sign-changes along the series would (under a thresholding procedure) have resulted in a single-block compartment, these may now be modelled as noisy emissions from a single unobserved state. An exemplar region is shown in Fig. 6. This shows an approximately 50 Mb region from chromosome 8 with eigenvector data from the H1 hESC cell line. A simple thresholding method in this region calls a total of 12 regions, whereas our HMM method finds only 6 larger regions in the same window. The disparity is caused by very short and single-bin compartments being disfavoured by the HMM-based method (e.g. Fig. 6).

Having called compartments we can compare their properties across cell types. In each case, a majority of chromosome compartment sizes are in the range 0–10 Mb, with a handful compartments reaching up to 40 Mb in size (Fig. 7). Median sizes for compartments called in this work match those reported previously, with a median size of around 5 Mb.[8] Our slightly larger mean compartment sizes (up to 7.6 Mb in H1 hESC) may be due to our altered domain calling procedure (Fig. 6) and is clearly influenced by some large outliers (Fig. 7).

Next we compare compartments between the three cell types under study. To do this, we calculate the minimum absolute distance from each boundary in one cell type against those in a designated comparison cell type (we used H1 hESC). The cumulative distribution of this boundary differences is shown (Fig. 8) and compared to a null distribution of random compartment boundaries. We find boundaries are significantly more closely aligned across cell types than is expected by chance. For example, genome-wide approximately 37% of compartment boundaries in H1 hESC have a corresponding boundary within 100 kb in
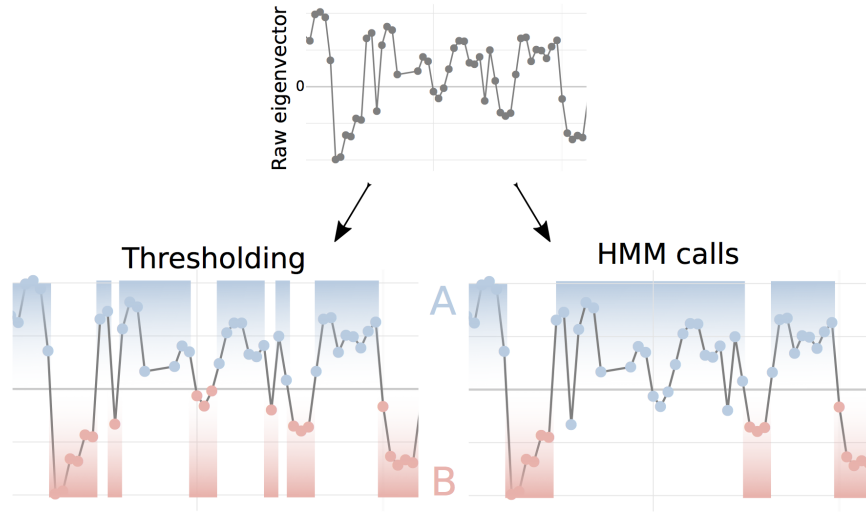
**Figure 6: Compartment calls by simple thresholding method or context-aware HMMs.** Chromosome compartments have previously been called through simple thresholding at 0,[1] in this work we also use an HMM-based method to call unobserved states that have emitted our noisy observed values (*right*).
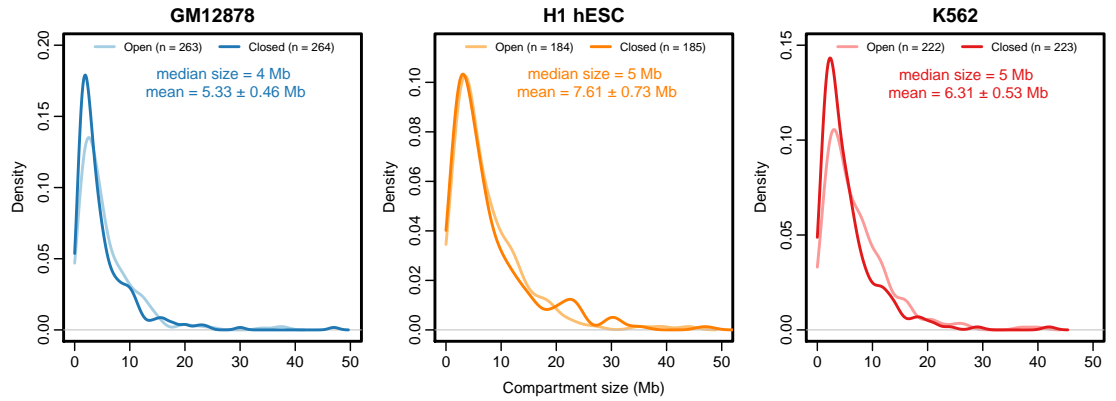


**Figure 7: Size distributions of compartment calls in three human cell types.** Kernel density estimates of compartment sizes (Open: A; Closed: B) are shown per cell type with summary statistics (*inset*), including mean compartment size with 95% confidence intervals.
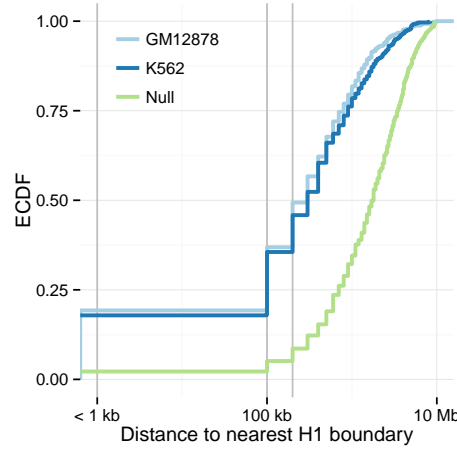
**Figure 8: Compartment boundaries are shared between cell types.** The empirical cumulative density functions (ECDF) of distances between H1 compartments and those called in GM12878 and K562 are shown. Vertical lines mark distances of 0, 1 and 2 bins. These distances were compared with a null distribution calculated from shuffled boundaries at a matched resolution.

GM12878 (and 35% in K562, but only 5% in random boundaries). Comparisons of the cumulative boundary distance distributions yield statistically significant differences relative to the null model (K-S test, K562: $D = 0.47$; $p \approx 0$; GM12878: $D = 0.49$; $p \approx 0$; Fig. 8).

### 1.4.2 TADs

Topological associating domains (TADs) are self-interacting blocks of the genome first described by Dixon *et al.*[2]. We applied the original TAD calling method without modification, which uses a measure of the directional contact bias of a fragment (Section **??** and Fig. **??**).

The Dixon *et al.*[2] method of calling TADs relies on the detection of boundaries,[4] thus is affected by sequencing depth: experiments with sparser contact matrices may not contain enough for a sufficiently high degree of bias to allow a boundary call. This is evident in our datasets even after normalisation, with the deeply-sequenced H1 hESC cell type having approximately 50% more TADs called than in the GM12878 cell type (Fig. 9). This effect could have been mitigated by down-sampling reads in the H1 cell type, but at a cost of reducing the quality of the best dataset under study. Instead this disparity should just be noted as a potential cofounder in downstream TAD analysis; at lower-resolution such as that used to calculate compartment eigenvectors (1 Mb) this sensitivity to sequencing depth is not evident (Figs. 2, 3).

Despite differing numbers, there is still detectable levels of conservation of TADs between cell types (Fig. 10). Genome-wide, 45% of all H1 TAD boundaries have a matching boundary in GM12878 in the same or an adjacent 40 kb bin (K562: 41%, null model: 22%; K–S test: $D = 0.2$, $p \approx 0$). To illustrate this conservation with a real example, a 20 Mb region of chromosome 2 is pictured (Fig. 11), highlighting the conservation between both TADs and compartment calls across the three cell types and at multiple scales: from chromosome-wide 1 Mb compartment eigenvectors, to TADs with individual boundaries resolved to 40 kb.
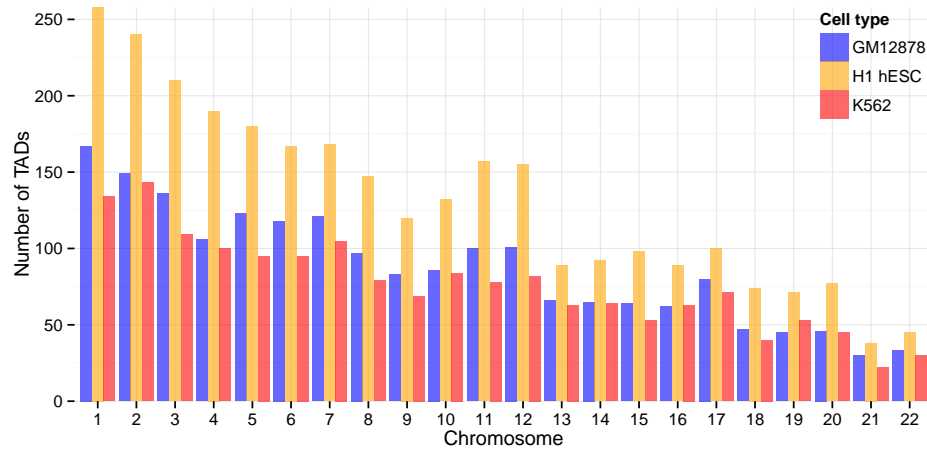
**Figure 9: The number of called TADs per cell type under study.** A greater number of TADs were called in H1 hESC (2,897 total) than in GM12878 (1,925) or K562 (1,677), due to the difference in sequencing depths in each experiment when matrices were binned at 40 kb resolution.
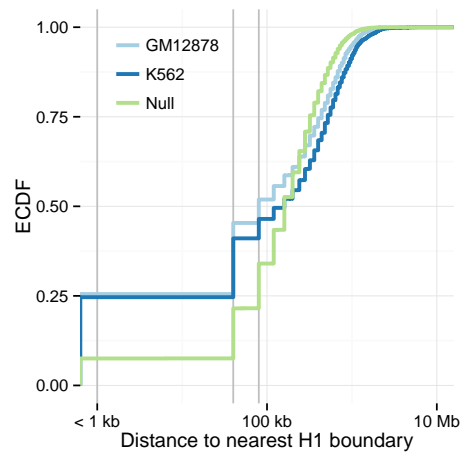


**Figure 10: TAD boundaries are shared between cell types.** The empirical cumulative density functions (ECDF) of distances between H1 TADs and those called in GM12878 and K562 are shown. Vertical lines mark distances of 0, 1 and 2 bins. These distances were compared with a null distribution calculated from randomly placed boundaries allocated at a matched resolution.
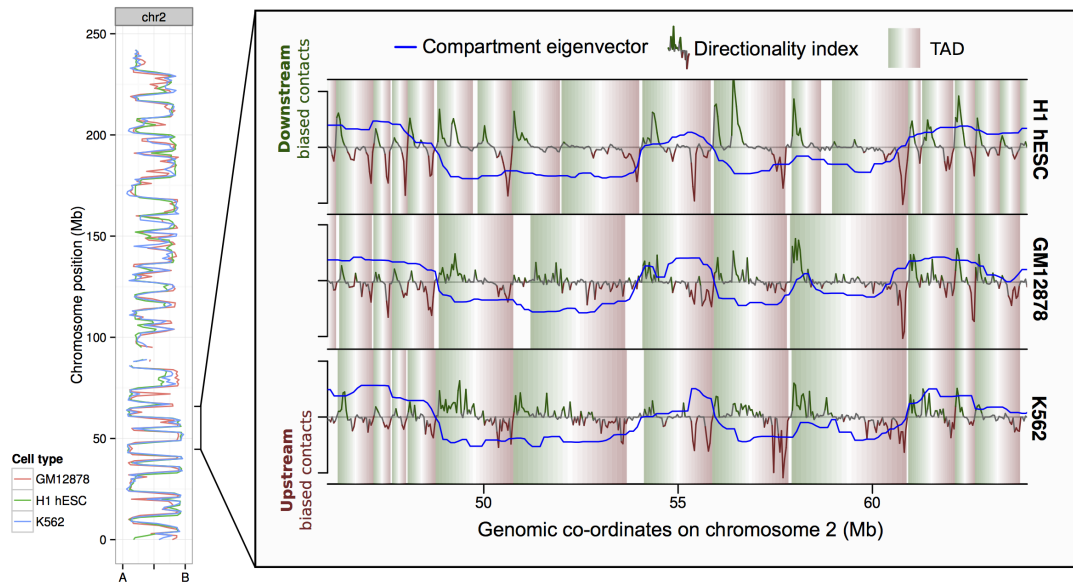
**Figure 11: Concordance of chromatin structure at multiple scales over three human cell types.** The eigenvector compartment profile is shown for chromosome 2 for three human cell types (*left*). At higher resolution, the zoomed region illustrates conservation of topological domains (TADs) over 20 Mb of the same chromosome.

## 1.5 DOMAIN EPIGENOMICS

The use of well-studied human cell types allows intersection with publicly-available epige-nomics datasets, such as those produce by the ENCODE consortium.[5] In total, 35 cell-matched ChIP-seq datasets were available for all three of the tier 1 ENCODE cell lines: GM12878, H1 hESC and K562 (see Methods **??**). In this section we integrate reprocessed Hi-C data and derived higher order domains with these various high-resolution datasets.

### 1.5.1 A/B compartments

The overwhelming majority of intersected chromatin features are significantly enriched in active A compartments relative to B compartments (Fig. 12). This is expected, A compartments represent the actively-transcribed and accessible portions of the genome, and have previously been shown to positively correlate with many of the features shown.[1,8]

Exceptions to this rule are few. However the repressive histone modification H3k9me3 is found more often in B compartments in two cell types, as is the P300 transcription factor (Fig. 12). Also of note is the histone variant H2A.Z which is significantly enriched in A compartments in GM12878 and K562, but this relationship is reversed in the embryonic stem cell line (Fig. 12). Recent evidence suggests specialised roles for H2A.Z in regulating both repression and activation during embryonic stem cell differentiation, acting as a "general facilitator".[9] Additionally H2A.Z has been reported to mark histone octamers for depletion, thereby permitting gene activation during differentiation.[10] Potentially, then, the H2A.Z enrichment in B compartments could be driven by regions soon to be de-repressed as the stem cell differentiates.
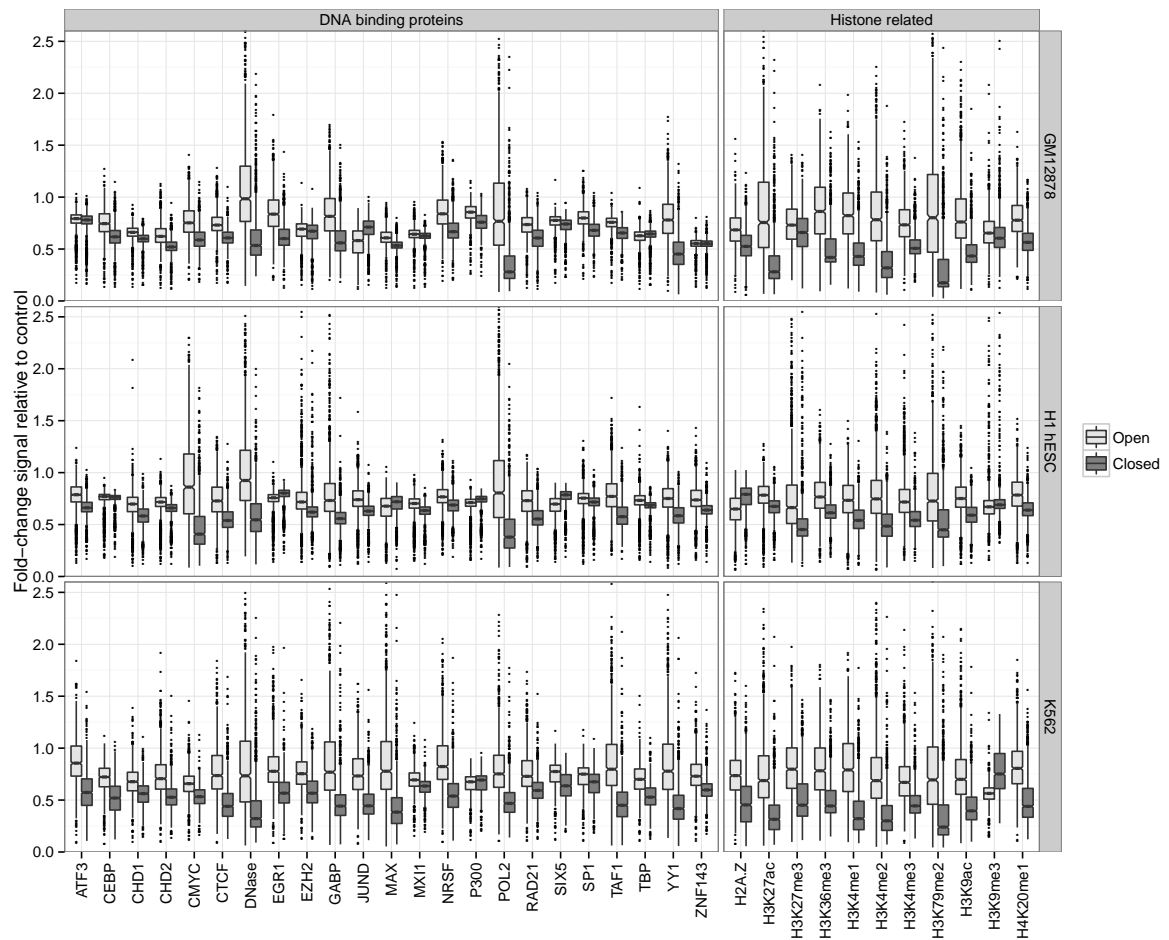
**Figure 12: The chromatin signatures of A/B compartments.**    Notched boxplots summarise the distribution of each feature over 1 Mb bins in open (A) and closed (B) compartments genome-wide.

### 1.5.2 TAD classes

Unlike compartments, initially TADs were not observably correlated with blocks of chromatin features.[2] Later studies have linked TADs with such enrichments, first in *Drosophila*[11] and later in human cells, where it was argued TADs are merely a low-resolution window to smaller "sub-compartments", bearing similar active and inactive marks to their much-larger namesakes.[4] Here we look for evidence that TADs called in our human cell types correspond to the "epigenomic domains" identified in *Drosophila* by Sexton *et al.*[11] Epigenomic domains were identified through supervised clustering of physical domains by their average enrichment for selected epigenomic features of known function, for example enrichment of H3K27me3 mark was used to call Polycomb (PcG) associated domains.

We found that TADs called across the three cell types used in this work could similarly be clustered into transcriptionally active (active), repressed heterochromatin (null) and polycomb-associated (PcG) domains, based on the patterns of DNase hypersensitivity, H3k9me3 and H3k27me3, respectively (Fig. 13). *Drosophila* physical domains were clustered into four categories, with three of those matching our annotations. The fourth *Drosophila* domain type was enriched for the HP1 protein (therefore likely centromeric) for which we did not have ENCODE ChIP-seq data in all human cell types under study.

This analysis reveals that active compartments typically cover both active and PcG-associated TADs, while B compartments appear more homogeneous and are composed mostly of H3k9me3-enriched heterochromatin even when considering fine-grained TAD structures rather than megabase-sized genomic blocks (Fig. 13).

These findings also link with recent work that suggested TADs are windows into "sub-compartments"[4] which more closely reflect the functional enrichments of compartments. However, in our data we did not find statistical support for the suggested 5 classes of sub-compartment; instead, an ensemble of algorithms for optimising the number of cluster centroids voted for two or three clusters of TADs (*data not shown*). This is not wholly surprising as Rao *et al.*[4] report sub-compartments only on extremely deep-sequenced samples, and at a scale of below that of TADs.

## 1.6 VARIABLE REGIONS

Despite the vast majority of the genome being in matched chromatin compartments across human cell types (Fig. 2), there are also regions of disagreement. Reasons for observable differences include technical errors and biases, but also more interesting functional explanations, where cell-type specific activation or repression is reflected in changes in higher order structure.

To conservatively call regions of variable structure (RVS), we used HMM-called compartment states and selected those which were either: i) open in one cell type and closed in both others or ii) closed in one cell type and open in both others. This left sets of RVS which could be considered as "flipped open" or "flipped closed" in a given cell type.
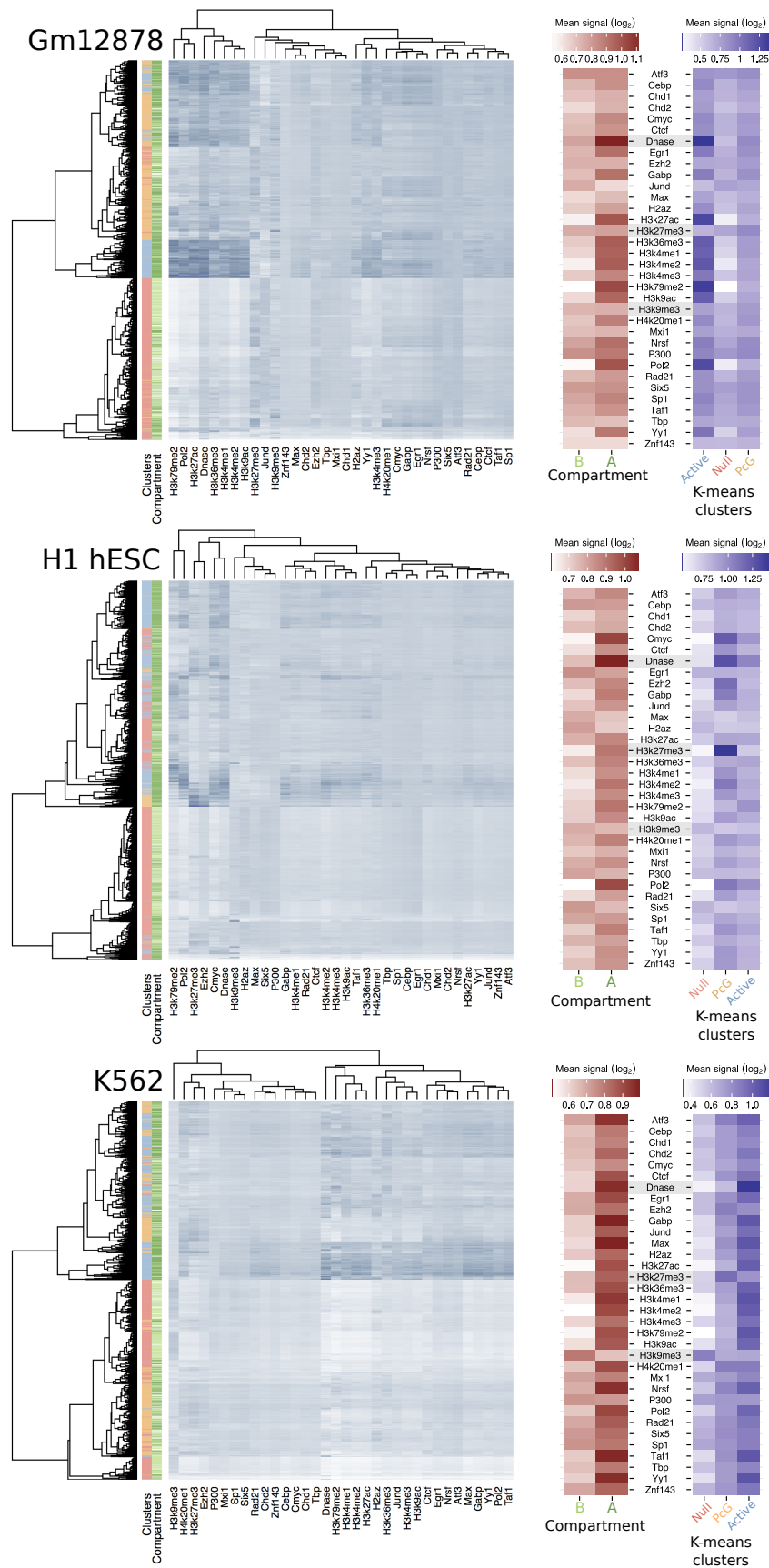
**Figure 13: TADs reflect epigenomic domains.** Following the Drosophila results of Sexton *et al.* [11], clustering of TAD domains by mean log2 signal of 34 ENCODE features distinguishes null, active and polycomb-associated (PcG) domains, as well as reflecting the encompassing A/B compartments.
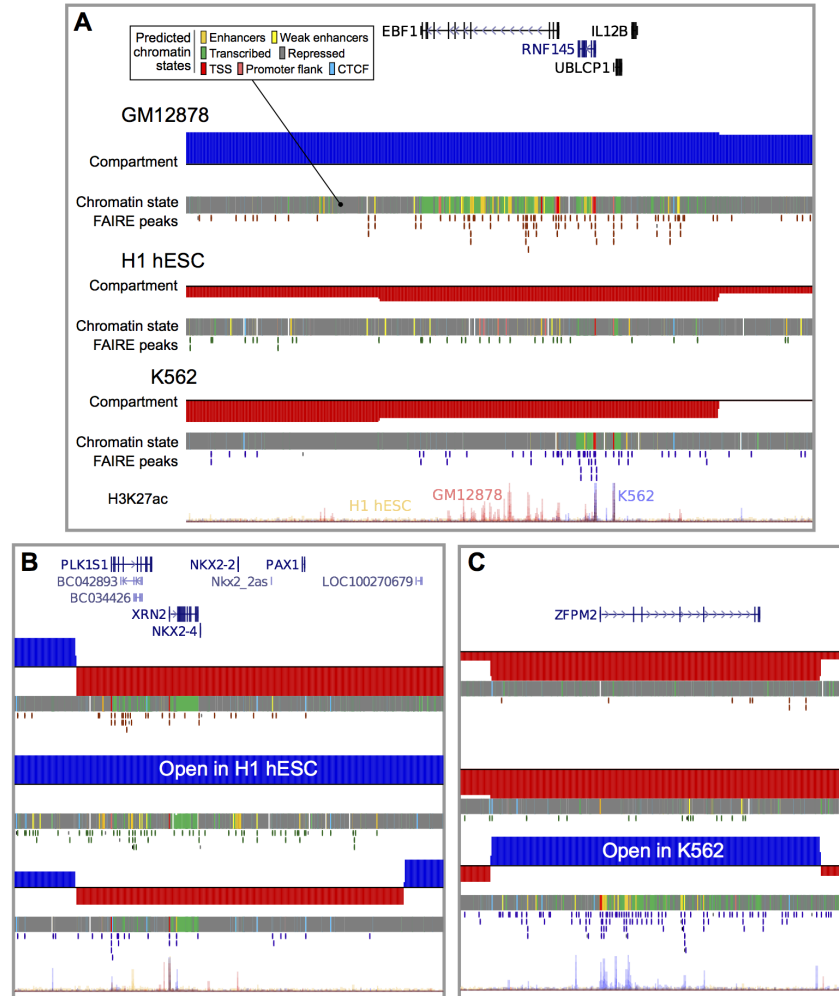
**Figure 14: Structurally variable regions indicate cell type specific biology.** Regions occupying the active A nuclear compartment in one cell type, but the repressive B compartment in the other two, were selected and ranked by the number of predicted active enhancers. Examples of particular interest from the top 5 regions per cell type are shown: (A) the chr5:158-159 Mb region which occupies the open A compartment in GM12878 cells, (B) the chr20:21-22 Mb region which is open in H1 hESC, (C) the chr8: 106-107 Mb region which is open in K562. Displayed tracks are: known (UCSC) genes, compartment eigenvectors, chromHMM/Segway combined chromatin state predictions, open chromatin FAIRE peaks, and H3K27ac signal.
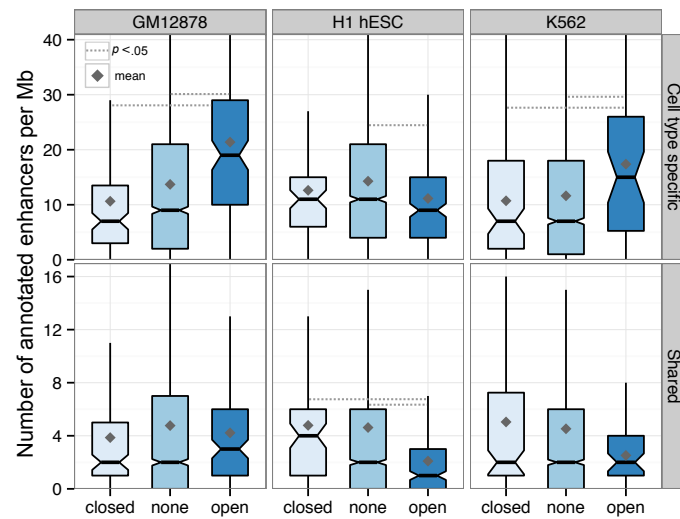
**Figure 15: Regions of variable structure are enriched for cell type specific enhancers.** Numbers of predicted enhancer states (cell type specific or shared between two or more cell types) are shown for regions with altered (open or closed) and non-altered (none) compartments in each cell type.

### 1.6.1 Chromatin state enrichment

Given our conservative definition of RVS (Section 1.6), such notable changes between transcriptionally permissive and repressive environments might be expected to be associated with cell-type-specific biology, such as functional chromatin states. To test this, we used consensus predicted chromatin state annotations, built from two machine learning algorithms, ChromHMM[12] and SegWay[13,14], and tested for enrichment or depletion in our set of RVS (Methods **??**).

We found that RVS show a striking enrichment for cell-type specific enhancers in both of our derived cell lines, but not in embryonic stem cells (Fig. 15). This observation can be explained as the undifferentiated cell type would not be expected to have lineage-specific enhancer contacts active in its pluripotent state. The same pattern was not seen for enhancers shared between two or more of the cell types under study. We observed a similar enrichment for cell-type-specific transcription but not for several other chromatin states including promoter activity (Fig. 16).

Together these state enrichments suggest the identified RVS reflect regions of cell-type specific biology, with heightened enhancer and transcriptional activity in their active cell type Fig. 16). Combined with the observed large-scale concordance of higher order chromatin organisation between cell types (Figs. 2, 11), these results reinforce a model of organisation whereby chromatin organisation is largely conserved and static across cell types, but also permits local flexibility in order to activate or repress regions of biological importance to a given cell type.
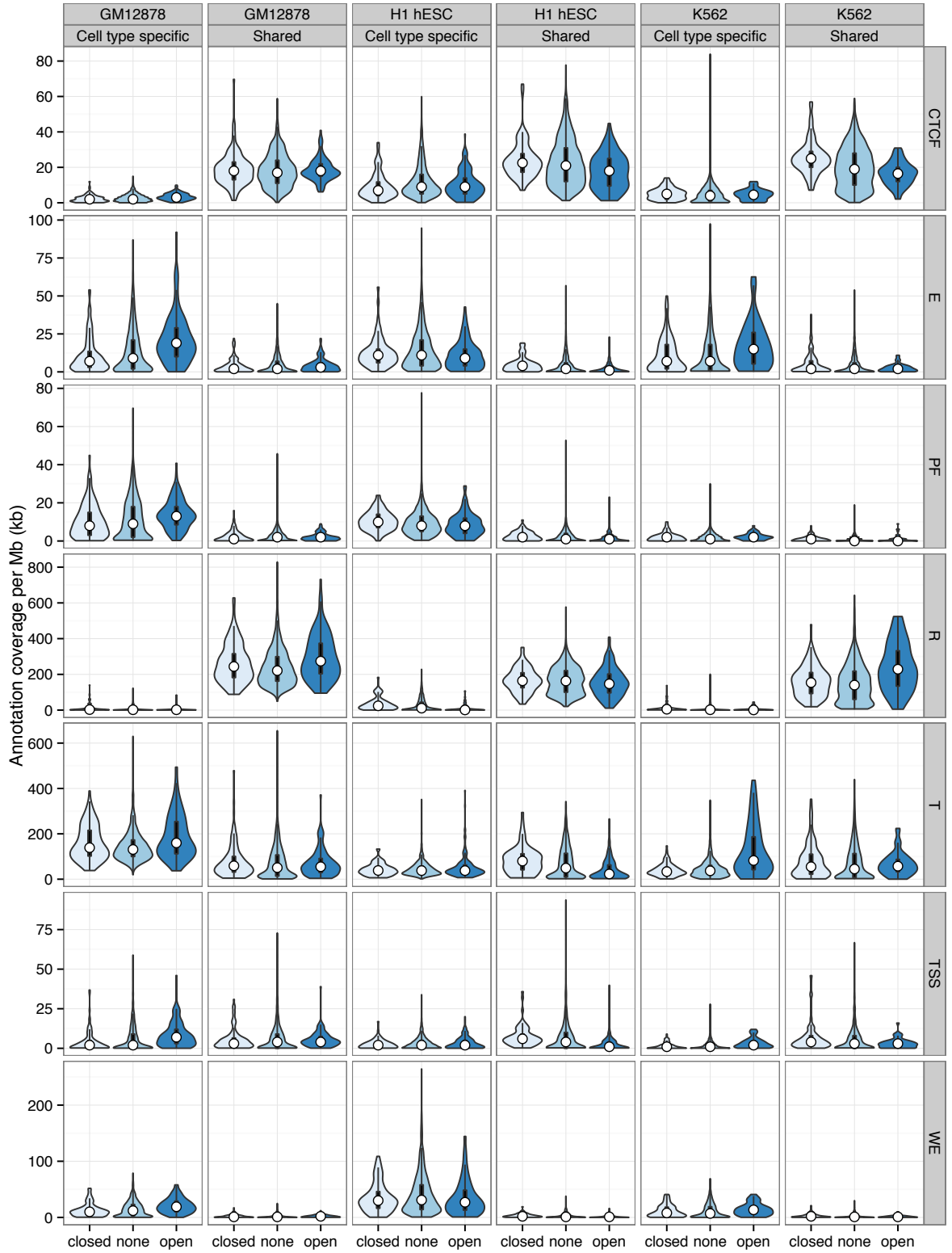
**Figure 16: Distributions of features across all predicted chromatin states in regions of variable higher order structure.** Distributions of the summed coverage of predicted chromatin states in each Mb per cell type are shown as bean plots. Predicted chromatin states are those from Hoffman *et al.*[14] and are labelled as: TSS: promoter and TSS; PF: promoter flanking region; E: enhancer; WE: weak enhancer or cis-regulatory element; CTCF: CTCF enriched element; T: transcribed region; R: repressed or low-activity.

### 1.6.2 Gene ontology enrichment

Specific examples of RVS highlight genes of interest (Fig. 14) but should be coupled with statistical evidence prior to suggestions of a general trend. For this reason we used Gene Ontology (GO) terms to test for functional enrichment within open RVS per cell type.

Functional enrichments of genes found in "flipped open" RVS in each cell type were calculated using DAVID[15,16] and filtered by false discovery rate (FDR < .05). This revealed slight enrichments for keywords "blood", "oxygen carrier" and "$\beta$ Haemoglobin" in the K562 cell type, a multipotent cell type which is known to show properties of an early erythrocyte, among others.[17] However, in the other two cell types we did not find significant enrichments across regions, except for artefacts caused by violations of the independence assumption used in GO term hypergeometric testing. Our RVS blocks were at least 1 Mb each so could contain more than one gene, thus enrichments were seen for those genes known to for paralogous clusters along a chromosome, such as olfactory receptors. The full results of these tests are given in the appendix (Tables **??**, **??**, **??**).

The size of RVS could also explain we do not capture the relationship hinted at by cherry-picked examples (Fig. 14). Given regions contain multiple (often unrelated) genes, we can imagine a case where a cell type specific locus is activated and moves into a more central position, disturbing adjacent genes which remain in a repressed state. Thus the cell type specific signals contained within the sum of all RVS in a given cell type could be obscured by the noise of adjacent genes captured in these broad compartment transitions.

### 1.6.3 Contact changes

A defining characteristic of active A compartment regions is a preferential bias in contacting other A compartment regions.[1] However, it is not clear whether cell-type-specific transitions in higher-order structure are solely compartment-level phenomena, or involve other structural strata. We therefore examined the genome-wide contact profiles of each region of variable cell-type-specific chromatin structure in detail.

If our cell-type-specific RVS are mediated by finer-scale structural levels (such as TADs) we might expect to see predominantly short-range contact changes in their underlying contact profile. Instead, we found that variable regions preferentially interact with other A compartment regions in the cell types in which they are active (Fig. 17), but not in the other cell types in which they are inactive. This supports the idea that these cell-type-specific regions are undergoing compartment-level transitions, disproportionately mediated by the formation of long-range contacts, while also not precluding additional changes at lower levels such as TADs.

## 1.7 NUCLEAR POSITIONING

Chromosome positioning within the nucleus is known to reflect gene density, with the most gene-dense chromosomes occupying the centre of the nucleus in human cells.[18] Kalhor *et al.*[6] used a Hi-C variant to recreate probability density maps of chromosome positions
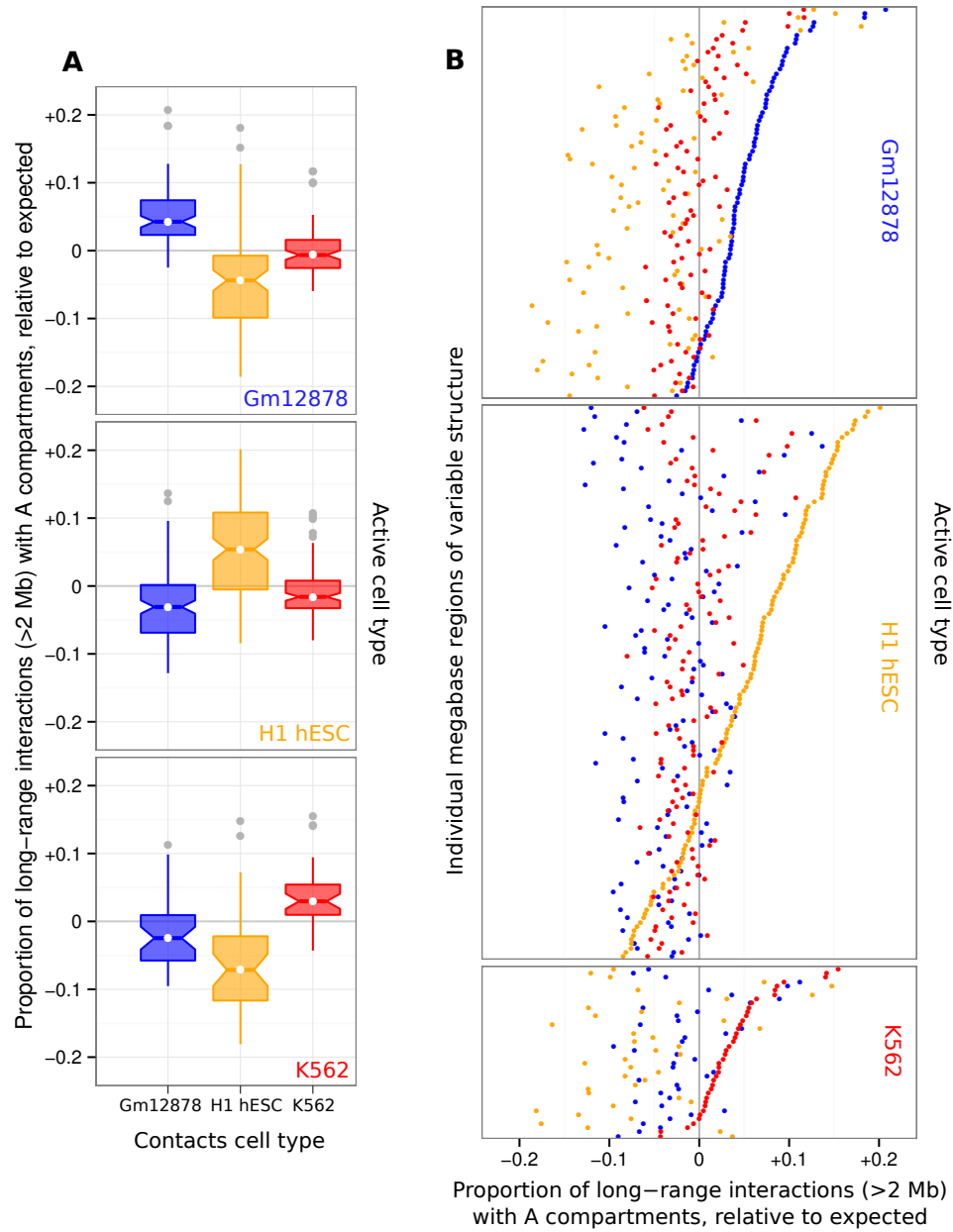
**Figure 17: Regions of variable higher order structure change their genome-wide contact profiles to favour active compartments.** Genome-wide normalised contacts were summed for each region of variable structure and the relative proportion of those that were with active / A compartments is shown across the three cell types used in this study. Proportions were subtracted from the genome-wide average per cell type, such that positive values indicate a greater than expected interaction bias with active compartments. These data are presented both as a summary boxplot (A) and with each individual region visualised, sorted by relative proportion of A contacts in active cell type (B).
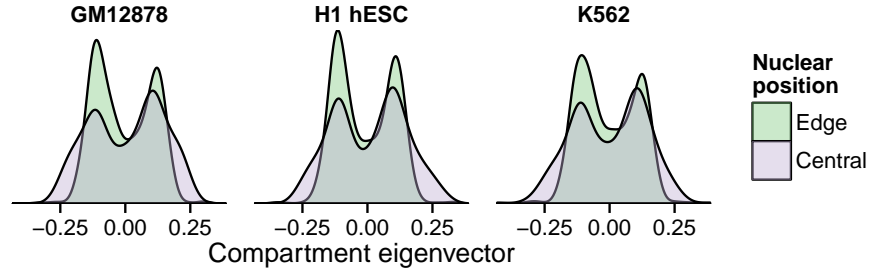
**Figure 18: Nuclear positioning of chromosomes relative to compartment eigenvectors.** Kernel density estimates showing peripheral chromosomes have a greater proportion of B compartments (negative eigenvectors) relative to centrally-position chromosomes. Positioning data from Boyle *et al.* [19] (Methods **??**).

which again reflected this feature of higher order chromatin organisation, and also reported active regions were more diffuse than inactive. A testable hypothesis with the eigenvector data used in this work is that active A compartments are enriched in the central nucleus of our human cell types, and B compartments are preferentially located in the nuclear periphery.

To test this, published data on chromosome positioning preference within the nucleus was used to label chromosomes as "central" or or "edge".[19] Chromosomes whose DAPI hybridisation signals were significantly enriched ($p \leq 2 \times 10^{-2}$) in the inner nuclear shell, as defined by Boyle *et al.* [19], made up the "central" group and included chromosomes 1, 16, 17, 19 and 22. Similarly the "edge" group had enriched signals ($p \leq 5 \times 10^{-3}$) in the outer shell relative to the inner nuclear shell and included chromosomes 2, 3, 4, 7, 8, 11, 13 and 18. The remaining chromosomes showed no significant preference to either inner or outer nuclear shells at $\alpha = 0.05$.[19]

We found that positive eigenvectors (reflecting A compartments) did show a modest relative enrichment in centrally-positioned chromosomes relative to those located at the nuclear periphery (Fig. 18). The significance of the difference in distribution of eigenvectors in the central and edge of the nucleus was determined by a two-sided Kolmogorov-Smirnov (K-S) test, with the null hypothesis that there is no difference between the empirical cumulative density functions of the central chromosome eigenvectors ($F_{central}$) and peripheral ($F_{edge}$). The difference was found to be statistically significant in each cell type ($H_0 : F_{edge} = F_{central}$; GM12878: $D = 0.11$, $p < 6 \times 10^{-4}$; H1 hESC: $D = 0.12$, $p < 8 \times 10^{-8}$; K562: $D = 0.10$, $p < 5 \times 10^{-3}$)

# 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] Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen Ca, Schmitt AD, Espinoza CA, Ren B (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, **503**(7475): 290–4.

[4] 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.

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

[6] Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L (2012) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature biotechnology*, **30**(1): 90–8.

[7] Eddy SR (2004) What is a hidden Markov model? *Nature biotechnology*, **22**(10): 1315–1316.

[8] Dekker J, Marti-Renom Ma, Mirny La (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics*, **14**(6): 390–403.

[9] Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q, Ge K, Levens D, Crane-Robinson C, Zhao K (2013) H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell*, **12**(2): 180–192.

[10] Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH (2012) Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell*, **151**(7): 1608–1616.

[11] 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.

[12] Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nature methods*, **9**(3): 215–6.

[13] Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes Ja, Noble WS (2012) Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods*, **9**(5): 473–476.

[14] Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes Ja, *et al.* (2013) Integrative annotation of chromatin elements from ENCODE data. *Nucleic acids research*, **41**(2): 827–41.

[15] Huang BDW, Lempicki R (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, (301): 1–43.

[16] Huang DW, Sherman BT, Lempicki Ra (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*, **37**(1): 1–13.

[17] Lozzio BB, Lozzio CB, Bamberger EG, Feliu AS (1981) A multipotential leukemia cell line (k-562) of human origin. *Experimental Biology and Medicine*, **166**(4): 546–550.

[18] Bickmore Wa (2013) The spatial organization of the human genome. *Annual review of genomics and human genetics*, **14**: 67–84.

[19] Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis Ja, Bickmore Wa (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human molecular genetics*, **10**(3): 211–9.