

論文の内容の要旨

論文題目 **A Computational Analysis of the Impact of 3D Chromosome Organization on its Gene Expression Regulation**

(染色体3次元構造がその遺伝子発現制御に及ぼす影響の情報解析)

氏 名 ナガイ ルイス アウグスト エイジ

To define three-dimensional (3D) chromatin structures in eukaryotic nuclei, Chromosome Conformation Capture (3C) sequencing technologies, such as the genome-wide 3C version (Hi-C), have emerged as a promising strategy, and revealed that the 3D structures non-randomly compacted have a functional impact on gene expression. For example, in B cells (B lymphocytes), the nuclear lamina interacting directly and indirectly with DNA and chromatin are disrupted during early lymphocyte development. Another study combining 3D fluorescence in situ and Hi-C analysis has shown that to establish a lineage specific signature, particular genome-wide structural transformations, i.e. chromatin compartment switches, are observed from loci of key developmental genes. In addition, the recent advances in 3C technologies have identified sub-compartment regions involved in B-cell fate determination.

B cells are central in the humoral immune system and the abnormal gene regulation in the cells is associated with cancer development. Diffuse large B-cell lymphoma, one of the most common type of cancer in B cells, represents 30-40% of all non-Hodgkin lymphoma. Genetic translocations on the chromosome structure deregulate B Cell CLL/Lymphoma 6 (Bcl6) gene in germinal-center response in mouse giving rise to different types of lymphoma. Moreover, a recent study using gene expression profiling revealed that PRDM1/BLIMP-1, a master regulator of plasma-cell differentiation, is inactivated in lymphoma which loss of its expression correlates with tumor cell proliferation.

Here, I sought to identify chromatin dynamics involved in the gene regulation of B-cell lymphoma. I combined different scales of genome structures from Hi-C of published data with mouse gene expression profiles (RNA-seq). I observed that the higher-order chromatin organizations characterized as compartments and topologically associating domains (TADs) are highly conserved among cells. However, the compartment switch from repressive in B cells to permissive in lymphoma involves a specific gene set (~8%) that exhibits increased gene expression levels in lymphoma. Interestingly, TAD boundaries are enriched with Prdm1 motif, suggesting coordination between the higher-order of chromatin structures and cancer development.

Results

Highly Conserved Folding Patterns on Chromatin Compartment Domains from Heterogeneous Resources

To examine the three-dimensional chromatin folding dynamics, I analyzed mouse Hi-C data from pro-B cell, B-cell lymphoma cell line and embryonic stem cell. Interaction maps from Hi-C data can provide information in multiple levels of genome organization hierarchy. The first level to examine chromatin interaction is the compartment domain. Eukaryotic genome is composed of two distinct

regions: the gene-rich and actively transcribed compartment A and the inactive compartment B. Overall, my analysis classified ~1.48 Gb of the mouse pro-B cell genome in B compartment, containing ~4,900 genes, whereas ~1.1 Gb was classified in A compartment, containing ~14,600 genes (Fig. 1a). To compare the folding patterns among the cells, I compared the A and B coordinates as previous described (Fig. 1). I found that 84.5% and 83.8% of the genomic coordinates, in pro-B cell and lymphoma respectively, remained stable when compared them in ES cell. Furthermore, I found a higher similarity (88.4%) between pro-B cell and lymphoma compartment coordinates. These results are consistent with the observation in a study that 90.7% of compartments are conserved in pre-pro-B and pro-B mice cells. In human, a previous study found a similarity degree of 64% among ES cell and four derived lineage compartments. While another study across 21 human cells and tissues, researchers observed 40.4% of conservation. These results suggest that my analysis achieves satisfactory chromatin organization structures by finding very similar chromatin compartment domains between pro-B cell and lymphoma using heterogeneous data resources.

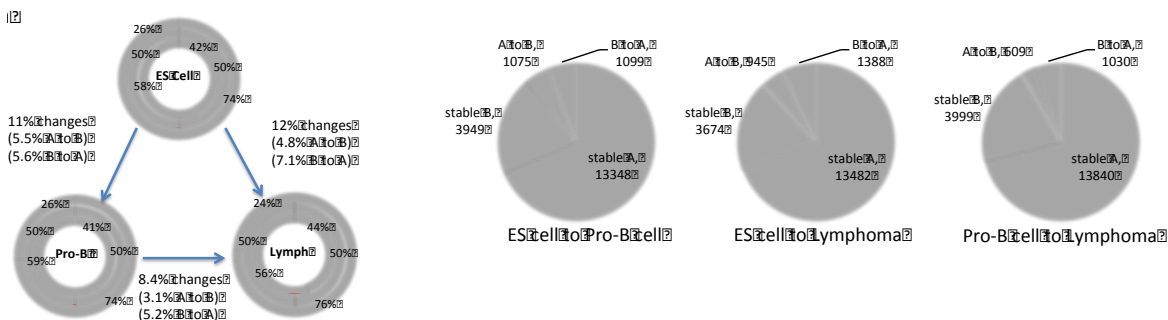


Figure 1. Chromatin organization at compartment level in ES cell, pro-B cell, and Lymphoma. (left) General information about genome size divided in A (blue) and B (red) per cell type. Outer: genome sequence in A or B; middle: amount of A and B compartment units identified in the whole genome; inner: genes contained in each compartment. (right) shows the number of genes in each compartment transition.

Reorganization of the mouse genomic compartments in mouse B-lymphoma

It has been known that compartment reorganizations were associated with the disruption of normal gene expression program leading to breast cancer. In order to investigate whether this phenomenon is also observed in mouse B cell-derived lymphoma, I identified chromatin compartments at 100 Kb resolution from normalized chromatin interaction matrices and obtained gene expression values from RNA-seq paired-end sequences. I examined regions switching compartments across cell types where 8% switched from inactive B-cell compartment to active lymphoma compartment, 11% between pro-B cell and ES cell, and 12% between lymphoma and ES cells (Fig. 1). This result is consistent with those of other studies; for example, the high gene expression level of *Myc* has changed from B compartment in ES cell to A compartment in B-cell lymphoma. Another gene described in the literature is *Ebf1*, as an important regulator that orchestrates B-cell fate, changed from B to A compartments.

To investigate the influence of chromatin compartmentalization frequency in chromosomes, I calculated a normalization score by dividing the sum of compartments in each chromosome by its chromosome size. The distribution of compartments throughout the genome was much more similar between pro-B cell and lymphoma than between those cells and ES cell. The genes located in the compartments switching from B to A tend to show increased expression levels, whereas the genes positioned in the A to B compartment change show the opposite tendency (Fig. 2). I further compared the expression levels of switching-gene with those of random genes located in stable regions. The overall tendency of gene expression in compartment changes is subtle due in part to that a subset of genes are affected by the compartment change. I selected all the genes from changes of repressive B compartment to active A compartment for a functional enrichment analysis. Pro-B cell and lymphoma were enriched for similar GO terms related to B cell function, such as natural killer cell activation involved in immune response, humoral immune response, B cell proliferation, and immune response process. To identify specific GO terms from each of pro-B cell and lymphoma, I next filtered out 846 common genes and found that pro-B cell were enriched with immune response terms, such as negative regulation of viral entry into host cell and proteolysis, whereas lymphoma was

enriched with sensory perception of chemical stimulus, V(D)J recombination, negative regulation of T cell apoptotic process.

Taken together, in a global level, my results indicate that the genes identified in A and B compartments have a high plasticity when compared GO term enrichment. In addition, changes in compartment corresponded to changes in gene expression, indicating that A and B compartments might have contributions to the cell differentiation although it might not be determinant.

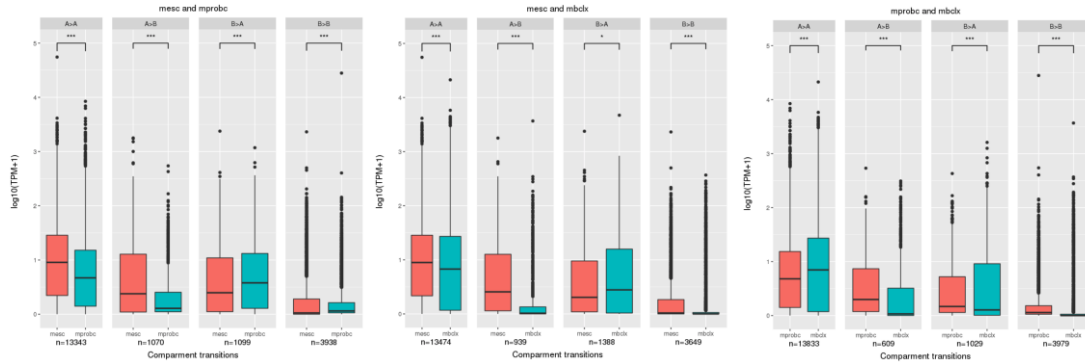


Figure 2. Chromatin reorganization and its gene expression levels in ES cell, B cell and B cell lymphoma.

Influence of topologically associating domains on compartment reorganization

I next examined the sub-compartment structure organized into dense and contiguous self-interacting regions know as TADs. Although TADs tend to be conserved across different cell types, chromatin interactions vary from cell to cell. Here I raised the question about the possibility that TADs contribute to the gene expression programs between pro-B cell and lymphoma.

At 40 kb resolution, my TAD calling classified the genome structures into similar numbers; 2,829 in lymphoma, 2,807 TADs in pro-B cell, and 2,808 in ES cell. Among them, the majority of TADs identified in a cell was conserved in another cell; by applying an approach previously described, I identified >70% overlapped TADs in a pair of samples, resulting in 2,348 (83.6% of total TADs) in lymphoma and ES cell, 2,319 (82.6%) in pro-B cell and ES cell, 2,235 (79%) in lymphoma and pro-B cell. In addition, TADs were largely identified from chromosome 7 in all samples, whereas lymphoma forms notably larger number of TADs in chromosome 14 and smaller number of TADs in chromosome X. These numbers support the study that the prostate cancer cell exhibits an increasing number of TADs compared to in its normal state. I also identified 301 unique TADs in pro-B cell, 191 in ES cell, and 198 in lymphoma, which is proportional to those found in a previous study; 65 unique out of 787 TADs identified in pro-B cell.

Collectively, my results show that the majority of TADs among pro-B cell, lymphoma and ES cell are highly conserved, whereas specific genomic regions are involved in the structural reorganization. In particular, B-cell lymphoma organizes the genome structure to ES-cell like TADs that are different from pro-B cell.

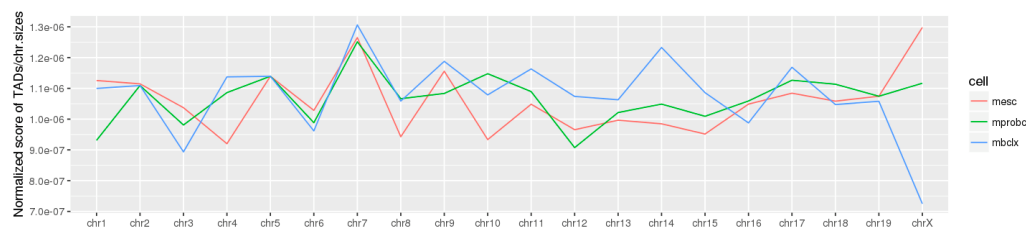


Figure 3. Normalized size of TADs per chromosome.

TAD boundaries suggest gene regulation function

Recent studies have revealed that TADs associate with CTCF and the protein complex cohesin by forming relative conserved structures across cell types to bring enhancers and specific genes close. Also, it has been observed that the disruption of TAD boundaries promotes gene expression leading to a physical malformation in mouse, suggesting the importance of CTCF in TAD boundaries. On the other hand, only 15% of TAD boundaries in mammals present CTCFs and 85% reside inside TADs. This scattered disposition points out that whereas CTCF can afford flexible adjustment to the chromatin conformation, the 3D chromatin organization is more likely to be influenced by a fine orchestration of cell-specific regulatory program. I then asked whether TAD boundaries of normal and cancer cells would exhibit CTCF enrichment, and whether genes located at boundaries would exhibit any variation of expression levels when compared to those located at intra-TAD regions.

Although I observed CTCF ($p < 0.01$) motif enriched at the boundaries in ES cell, I also found PRDM1 ($p = 0.01$), HRE ($p < 0.01$), and Meis Homeobox 1 ($p = 0.001$) enrichments that play important roles in normal development. The boundaries in pro-B cell were enriched with Nanog and PRDM1 ($p = 0.01$) motifs. In contrast, I could not profile the CTCF motif enrichment neither in lymphoma cell nor pro-B cell. Interestingly, the coding gene of PRDM1, that is associated with various cancer developments, exhibits the high expression level only in lymphoma even though PRDM1 was located in active compartment and enriched in TAD boundaries in all samples.

I next categorized genes based on their proximity to TAD boundaries. Remarkably, genes located around TAD boundaries ($< 40\text{Kb}$) show significantly higher expression levels ($p < 0.0001$) in all samples. This has been also observed in a high-resolution experiment in fruit flies. Overall, although TADs are highly conserved between cell types and often delimited by CTCF motifs, my results show the relationship of TAD boundaries with cancer-related transcription factors rather than with the CTCF motifs.

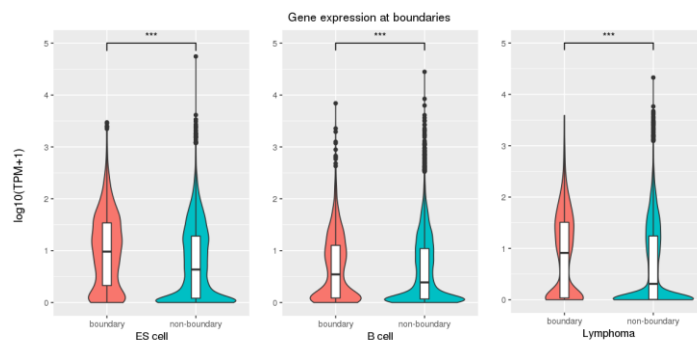


Figure 4. Expression levels of genes at TAD boundaries and intra-domain.

Conclusion

My results show that the majority of TADs among pro-B cell, lymphoma and ES cell are highly conserved, whereas specific genomic regions are involved in the compartment change. In particular, the switching compartment regions are associated with gene expression increase. I concluded that an unknown mechanism profoundly exists to restrict the structural and functional changes of genomic regions and cognate genes in a specific manner.