**1. Introduction :**

Contrary to conventional illustrations, genomic DNA is not linear, but is folded into compact chromosomal structures that likely impact gene expression of the embedded genes. On a global scale, regions with a high degree of compaction are classified as heterochromatin while uncondensed regions are called as euchromatin. These are respectively associated with lower and higher expression levels. Locally, areas where frequent DNA-DNA interactions occur as a results of their close proximity within the cellular nuclear space are called topologically associated domains (TADs). These domains are largely conserved across cell lines and they frequently contain smaller loop structures that promote contact between different genetic regulatory elements such as enhancers and promoters. Such chromatin loops are often found at TAD boundaries (Rao et al, 2014). They are enriched in binding sites of architectural proteins, including CTCF and cohesin (Pope et al, 2014) which are both thought to function in the delimitation of TADs since the boundaries of TADs represent genomic insulators by preventing DNA-DNA interactions across domains. In addition to modulating regulatory interactions between genomic elements, TAD boundaries are also found to be gene-dense and are enriched in highly transcribed genes (Ong et Corces, 2014).

A large proportion of the mammalian transcriptome does not code for proteins and to date, the number of known noncoding genes is more than 3 times that of protein-coding genes (Iyer et al, 2015). Among noncoding RNAs, long noncoding RNAs (>200bp) that do not overlap protein-coding genes are the most abundant (long intergenic noncoding RNAs, lincRNAs). Functional roles of the characterized lincRNAs are diverse. They can for example act as decoys, occupying the active sites of enzymes to reduce their activity, or conversely recruit proteins to specific sequences in the genome to increase their activity locally (Howard Chang and John Rinn 2013). Enhancer associated lincRNAs (elincRNAs) are another functional class of lincRNAs whose promoter region is associated with enhancers. Most active enhancers are transcribed (Arner et al, 2015) ant the resulting transcripts can recruit transcription factors to increase transcriptional activity. Enhancer-associated RNAs are often transcribed in bidirectionally and rapidly degraded, but some are transcribed preferentially or exclusively in one direction (Darrow et al, 2014). LincRNAs that are associated with enhancer activity (elincRNAs) and are transcribed preferentially in one direction are good candidates for studying the involvement of lincRNA in the regulation of DNA-DNA contacts. There are already some characterized elincRNAs, such as HOTTIP that have been showed to act as a link chromosomal interactions and transcription. HOTTIP nascent transcript promotes transcription of several HOXA genes by recruiting of histone modifying proteins to target genes via chromosomal looping (Wang et al, 2011). Although this elincRNA transforms information from high order chromatin organization into histone modifications, it it is not required for the formation of chromosomal looping, but this might not hold true for all elincRNAs.

LincRNAs promoters are enriched in enhancer marks (Popadin et al, 2013), which suggests their likely role in transcriptional regulation. Additionally, lincRNAs might be involved in the control of nuclear architecture as they have been shown to be enriched at TAD-boundaries to mediate promoter-enhancer interactions (Chen et al 2016). In addition, disrupted elincRNAs regulation has been linked to various disease diseases (Ounzain et al, 2014).

To investigate whether enhancer-associated lincRNAs are involved in the regulation of TAD nuclear organization, I used lincRNAs whose promoter region is associated with enhancer activity as elincRNAs in a human lymphoblastoid cell line (LCL) and studied their involvement in the regulation of DNA-DNA contacts specifically.

By using various bioinformatics tools to analyze publicly available multi-omics data from the ENCODE project, I investigated the molecular properties of these elincRNAs, their enrichment in different regulatory elements and their association with the amount of DNA-DNA interactions to examine their role in regulating TAD organization. In general, the study of elincRNAs is of particular interest as they have been linked to various disease phenotypes (Ounzain et al, 2014).

**2. Materials and methods :**

Unless stated otherwise, all statistical tests were performed using R 3.3.1(citation). Overlapping of genomic elements were done using either bedtools 2.26 or the “intervals” package (citation)in R. Manipulations on Hi-C contact matrixes were performed using the “Matrix” (citation)package.

*Genes:*

LincRNAs and protein-coding genes used come from the ENCODE website. The list of genes used in all analyses corresponds to genes expressed in the GM12878 lymphoblastoid cell line.

*TAD definition:*

The list of TADs used in the computations is based on that from Rao et al, 2014. They called the TADs based on Hi-C data across different human cell lines normalized and processed with their own protocol. Here, all the large TADs that completely encompassed smaller ones were removed to preserve the signal from the boundaries of the small TADs.

*TAD boundaries definition:*

Instead of defining TAD boundaries based on arbitrary thresholds, TADs were split into 10 bins of 10% their length. This threshold was chosen based on previous findings showing an increase in transcriptional activity at ~10% from the TAD border (Histogram from summary 3).

*Conservation and tissue specificity:*

The sequence conservation was previously calculated (Tan Yihong, J., unpublished) using **phastCons scores** (Siepel et al, 2005). Tissue specificity index (Tau) was computed following the method from Kryuchkova-Mostacci et al, 2016.

*Expression levels:*

Median expression levels were computed from the **GEUVADIS 1000 Genomes project expression matrix. The matrix contains expression data for 373 individuals.**

*DNA-DNA contacts:*

Contacts were calculated using Hi-C contact matrices from Rao et al (2014). All computations are performed on 5kb resolution matrices with a MAPQ score of at least 30. The matrices were normalized using the KR normalization vector provided by the authors, when analyzing only GM12878. When comparing between different cell lines, SQRTVC (square root vanilla coverage) was used for chromosome 9 of all cell lines, because the KR algorithm did not converge for chromosome 9 of K562.

Contacts in each gene were calculated in 2 ways: contacts between genes and the whole chromosome, and contacts between genes and their respective TADs.

In both cases, this was done using the sum of interactions multiplied by the length of the gene, divided by the resolution.

*Chip-seq:*

Chip-seq data for CTCF, RAD21 and SMC3 in GM12878 were retrieved from the ENCODE website. All three datasets were produced in the Snyder lab, at Stanford University.

*Enrichment of genetic elements:*

All enrichment tests were performed using the genome association tester (GAT) (Heger et al, 2013) version 1.2. All tests for enrichment of lincRNAs were performed using the intergenic space of the genome as a workspace. When testing for enrichment of protein-coding genes, the protein coding space of the genome was used as the workspace.

**3. Results**

* elincRNAs enriched at TAD boundaries compared to nelincRNAs ?

→if TRUE: Possible role in TAD organization. Need real experiments

to establish a causal link.

* Insulator proteins binding sites enriched in elincRNAs ?

→ if TRUE: Role as insulator ?

**4. Discussion:**

Found/didn’t found association between elincRNAs and….

 Although these results suggest a role for elincRNAs in TAD organization, they provide no information on their exact function of mechanistic role.