**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).

CONNECT THOSE § WITH A SMOOTH TRANSITION

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. HOTTIP is a good example of a characterized elincRNA affecting chromosomal structure. The nascent transcript recruits histone modification proteins to target genes via chromosomal looping. [CHECK HOTTIP IN MORE DETAILS]

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. Results**

* (?)Used bins instead of TAD boundaries to avoid arbitrary thresholds
* elincRNAs enriched at TAD boundaries compared to nelincRNAs ?

→if TRUE: Possible role in TAD organization.

* Insulator proteins binding sites enriched in elincRNAs ?

→ if TRUE: Role as insulator ?

**3. Methods :**

* TAD definition:TADs used in the computations come 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 algorithm and experimental protocol. The large encompassing TADs were removed to preserve the signal from the boundaries of the smaller ones inside.
* TAD boundaries definition: TAD were split into 10bins of 10% TAD length. The threshold was chosen based on previous findings showing an increase in transcriptional activity at ~10% from the TAD border (Histogram from summary 3)

\*conservation

Computed phastCons score (Siepel et al, 2005)

\*tissue specificity

Computed taus based on method from Kryuchkova-Mostacci et al 2016

\*expression levels

\*data

Pcgenes and lincRNAs expressed in LCL: ENCODE

Raw Hi-C data: aidenlab.org, data from Rao et al, 2014

ChIP-seq data: ENCODE

* Enrichment tests: Enrichment tests were performed using the Genome Association Tester (GAT) version 1. 2. [Workspaced, Annotations, Segments]
* Manipulation of .bed files: BEDTools 2.26 was used to find intersections between genomic segments.
* Statistical tests: Statistical tests were performed using the R software environment for statistical computing and graphics version 3.3.1.

**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.