Introduction :

Genomic DNA is folded onto itself, forming compact structures that likely impact gene expression. On a large 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. On a local level, areas where DNA-DNA interactions are especially frequent 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 elements such as enhancers and promoters. Such chromatin loops are found at domains boundaries (Rao et al, 2014) and are enriched binding sites of architectural proteins, including CTCF and cohesin. Both proteins are thought to play a role in the delimitation of TADs since the boundaries of TADs act as insulators, preventing DNA-DNA interactions across them. TAD boundaries are also gene-dense and enriched in highly transcribed genes (Ong et Corces, 2014).

CONNECT THOSE § WITH A SMOOTH TRANSITION

The largest part of transcripts in the mammalian genome do not code for proteins (Iyer et al, 2015). Among these non-coding RNAs, long non-coding RNAs (lncRNAs) are the longest (>200bp). Long intergenic non-coding RNAs (lincRNAs) are the most abundant of their class. They are characterized by being located outside protein-coding genes. LincRNAs promoters are enriched in enhancer marks (Popadin et al, 2013), which suggest a role in transcriptional regulation. LincRNAs might also play a role in the control of nuclear architecture as they have been shown to mediate promoter-enhancer interactions and are enriched in TAD-boundaries (Chen et al 2016). LincRNAs whose promoter region is associated with enhancers (elincRNAs) are therefore good candidates for studying their involvement in the regulation of DNA-DNA contacts.

In this work, we study the role of elincRNAs in the organization of TADs using bioinformatics tools and publicly available data from the ENCODE project. We investigate the properties of these elincRNAs and their frequence in different elements of the genome. In general, the study of elincRNAs is of particular interest as they have been linked to various disease phenotypes (Ounzain et al, 2014).

**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 ?

**Methods :**

* TAD definition:

🡪TADs used in the computations come from [ENCODE?] the large encompassing TADs were removed to avoid hiding the smaller ones inside.

* TAD boundaries definition:

🡪Instead of using an arbitrary threshold for defining TAD boundaries, TADs were split into 10 bins of 10% of their length.

enrichment test:

GAT 🡪 parameters

\*conservation

Computed phastCons score (Siepel et al, 2005)

\*tissue specificity

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

\*expression levels

\*data

TADs: ENCODE ?

Pcgenes and lincRNAs expressed in LCL: ENCODE

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

ChIP-seq data: ENCODE

**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.

**Tools :**

- GAT 1.2

- BEDtools 2.26

- R 3.3.1