Introduction :

Contrary to their conventional illustrations, genomic DNA is not linear structures 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 result 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 cohesion [ref?], 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 different domains. In addition to regulating 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 do not code for proteins and to date, the number of noncoding genes is more than 3 times that of protein-coding genes (Iyer et al, 2015). Among non-coding RNAs, long non-coding 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. For example …. give some examples of lincRNA roles (maybe read a review paper on lncRNAs, ex. Howard Chang and John Rinn 2013? I think, Nature?)

Here, you might want to discuss more on elincRNAs first. For example:

* most active enhancers are associated with transcriptional activity (read on FANTOM consortium papers)
* enhancer-associated transcripts are often transcribed bi-directionally, i.e. no selection on the direction of transcription, and these transcripts are often unstable and rapidly degraded
* some lincRNAs have been associated with enhancer activity and are likely transcribed in a preferred direction (uni-directional transcription), these are called elincRNAs

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

**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

TADs: called using the raw Hi-C data generated above across different human cell lines

ChIP-seq data: ENCODE

**Tools :**

- GAT 1.2

- BEDtools 2.26

- R 3.3.1

**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

General comments:

* try to use Endnote or other equivalent reference tools to organize your references, this will be very important and useful for you in the future