Do enhancer-associated long intergenic noncoding RNAs contribute to nuclear architecture ?

1. Introduction :

It was only discovered in the past decade that a surprisingly large proportion of the mammalian transcriptome does not code for proteins. To date, the number of annotated noncoding genes longer than 200 nucleotides (long noncoding RNA, lncRNA) exceeds that of protein-coding genes by at least 3 times (Iyer et al., 2015)⁠. Among lncRNAs, those that do not overlap protein-coding genes are the most abundant, referred to as long intergenic noncoding RNAs, lincRNAs. Functional and evolutionary analyses, together with extensive characterization of a handful of lincRNAs, demonstrate that some of these transcripts are involved in the regulation of gene expression programs, both transcriptionally and post-transcriptionally, and that they can contribute to organismal traits and diseases (Kornienko, Guenzl, Barlow, & Pauler, 2013)⁠. However, the mechanisms and functions, if any, for the majority of lincRNAs remain unknown (Rinn & Chang, 2012)⁠.

A high proportion of lincRNA arise from enhancers as transcription as mwidespreadfrequently Recently, lincRNAs associated with human trait variants have been shown to have enhancer-associated *cis*-regulatory roles on local gene expression regulation. Importantly, these loci are correlated with higher frequency of chromosomal interactions relative to other lincRNAs in a human lymphoblastoid cell line (LCL), suggesting that enhancer-associated lincRNAs (elincRNAs) may be involved in gene regulation through modulating chromatin architecture (Tan et al, 2016, under revision).

Indeed, spatial organization of the genome impacts gene regulation (Engreitz, Ollikainen, & Guttman, 2016)⁠. Specifically, the folding of genomic DNA into variably compacted chromosomal structures can strongly influence expression of the embedded genes (Gorkin, Leung, & Ren, 2014)⁠. Globally, regions with low degree of compaction are associated with high levels of active transcription, referred to as heterochromatin. On the other hand, relatively uncondensed and less transcriptionally active regions are called euchromatin (Passarge, 1979)⁠(Tamaru, 2010)⁠. Active chromatin are further compartmentalized into smaller domains where frequent DNA-DNA interactions occur as a result of their close spatial proximity, called topologically associating domains (TADs), which are key in modulating gene transcription programs. Recent findings show that boundary of TADs are essential for modulating and preventing gene regulatory interactions within and across TADs, respectively.  ~~They are often gene-dense and are enriched in highly transcribed genes~~ (Ong & Corces, 2014)⁠.

Detailed functional characterizations of elincRNAs have demonstrated the molecular mechanisms underlying their roles in the spatial organization of the genome. For example, Haunt is one such elincRNA (Yin et al., 2015)⁠⁠. This elincRNA is involved in the regulation of HoxA gene transcription by modulating intrachromosomal interactions, specifically, through mediating promoter-enhancer looping. These recent findings raise the question on what is the prevalence of elincRNAs that contribute to gene regulation through the modulation of chromosomal conformation.

~~Unlike most enhancer-associated noncoding RNAs, which are often transcribed bidirectionally and then rapidly degraded (Darrow & Chadwick, 2013)⁠, elincRNAs are more stably and preferentially transcribed in one direction (Marques et al., 2013)⁠. These distinct features of elincRNAs make them less likely to be spurious byproducts of pervasive transcription of the mammalian genomes and thereby, represent suitable candidates to study their potential regulatory functions. Those elincRNAs will therefore be the focus of my analysis.~~

~~Chromosomal contacts within TADs, often seen as looping structures, occur particularly at TAD boundaries and are crucial for establishing correct interactions between regulatory elements, such as enhancers and promoters (Gorkin et al., 2014)⁠. TAD boundaries are also enriched in architectural proteins, including CTCF (Pope et al., 2014)⁠, which functions to delimit TAD borders by acting as genomic insulators that prevent DNA-DNA interactions across multiple TADs. Cohesin, another architectural protein, is also enriched at TAD boundaries. It is a multi-protein complex that is thought to be involved in establishing enhancer-promoter interactions (Ji et al., 2016)⁠. While most CTCF sites are shared between different cell types and species (Ji et al., 2016)⁠, cohesin binding at gene regulatory elements is often cell-type specific (Hadjur et al., 2009)⁠.~~

Using various bioinformatics tools to analyze publicly available multi-omics data from human lymphoblastoid cell lines (LCLs)~~from the ENCODE project (ENCODE Project et al., 2012)⁠ and data from whole-genome chromosome conformation capture (Hi-C) experiments (Rao et al., 2014)⁠~~, I investigated the molecular properties of elincRNAs. Specifically, I examined their enrichment in regulatory elements that are key in TAD regulation and their association with the amount of chromsomal interactions to gain initial insight into their roles in gene regulation within topological domains. My analyses show that elincRNAs are associated with high density of DNA :DNA contacts within TADs and are significantly enriched in protein binding sites important for TAD regulation. Importantly, elincRNAs are strongly enriched at chromosomal loop anchors, where promoter-enhancer interactions occur, supporting the hypothesis that they may contribute to gene regulation by establishing contacts between gene regulatory elements and modulating chromosomal organization.

2. Results

Enhancer-associated lincRNAs (elincRNAs) in human lymphoblastoid cell lines (LCLs) were identified based on overlap with LCL enhancers (GM12878, (ENCODE Project et al., 2012)) at their putative promoter regions (estimated as the 1kb upstream from their transcriptional start site). LincRNAs whose promoter region also overlapped other predicted regulatory regions, specifically, promoter elements (ENCODE Project et al., 2012), in LCLs were excluded from the analysis (elincRNAs=236 and other LCL-expressed lincRNAs=1756).

elincRNAs show similar expression levels as other lincRNAs

Unlike most enhancer-associated noncoding RNAs (eRNAs), which are often transcribed bidirectionally and then rapidly degraded by the nuclear exosome (Darrow & Chadwick, 2013) (Lam, Li, Rosenfeld, & Glass, 2014)⁠, elincRNAs are more stably and preferentially transcribed in one direction and are often spliced and polyadenylated (Marques et al., 2013)⁠. First, to investigate if that if LCL-expressed elincRNAs share similar expression profiles as the lowly-expressed eRNAs relative to other expressed genes, I compared expression levels of elincRNAs to that of other lincRNAs and protein-coding genes (Figure 1). I found that elincRNAs are no more lowly expressed compared to other lincRNAs in human LCLs (GM12878, two-tailed Mann-Whitney U test, p=0.258), suggesting elincRNAs may have distinct properties compared to other enhancer-associated transcripts.

elincRNA transcripts are less conserved than other lincRNAs

To gain insights into elincRNAs evolution, I investigated their nucleotide conservation in primates and placental mammals using phastCons scores (REF), a measure of nucleotide conservation (Methods). I found that exons of elincRNAs are less conserved than other LCL-expressed lincRNAs as well as protein coding genes (p= ??, two-tailed Mann-Whitney U test, Figure 2). Interestingly, a set of trait-relevant and enhancer-associated lincRNAs showed that although exons of these lincRNAs did not seem to have evolved under purifying selection relaive to other LCL-expressed lincRNAs across mammalian and primate evolution, their sequences are constraint specfically during recent human evolution (Tan et al 2016, under revision). Therefore, although my result may suggest that elincRNA transcripts were not evolving under constaint across broad mammalian evolution, their conservation across modern human evolution remain to be investigated.

elincRNAs promoter regions are enriched at loop anchors and cohesin binding

Next, to examine whether elincRNA are associated with the regulation of chromosmal architecture, I first investigated the co-localization of elincRNAs with regulatory regions essential in establishing chromosomal interactions within topologically associating domains (TADs). Specifically, significant enrichment of elincRNAs was found at loop anchors (fold enrichment=XX, p=XX), where enhancer-promoter gene regulatory interactions occur (Ji et al., 2016), than what would be expected if they were randomly distributed across the intergenic regions of the human genome (Methods).

Previous studies have demonstrated that TAD regulatory elements are also enriched at the loop anchors within TADs [REF]. Specifically, the cohesin protein complex has been previously shown to be important for intra-TAD gene regulation in a cell type-specific manner [REF]. Also enriched at loop anchors, the CTCF transcription factor is another central player in the regulation of chromatin architecture and gene expression. According to a recent model (Ji et al., 2016)⁠, loops mediated by CTCF or CTCF and cohesin collectively have insulating properties that segregate interactions across TADs and are important for the structural maintenance of TAD. In contrast, loops containing only cohesin binding sites are crucial in mediating regulatory chromosomal interactions where cohesin depletion is associated with disrupted promoter-enhancer interactions within TADs [REF].

Significant e were found forCTCF and cohesin binding peaks (cohesin protein subunits, SMC3 and RAD21 [REF for SMC3/RAD21 as subunit of cohesin] with respectively, relative to respectively, As msiteshuman or independently, collectively in chromosomal regulationIfurther d theindependent of elincRNAs in and cohesin binding by obtaining mutually exclusive bindings sites of the proteins.This revealed a much stronger enrichment of cohesin binding sites in elincRNA loci (fold enrichment=XX, p=XX) compared to CTCF binding sites (fold enrichment=XX, p=XX, Figure 7).much more frequently , supporting their roles in modulating promoter-enhancer looping

elincRNAs promoter regions are not enriched at TAD boundaries

As loop anchors are frequently found at TAD boundaries, I next investigated whether similar enrichment would be observed for elincRNAs at TAD boundaries. I estimated TAD boundary regions by calculating……. (Methods, Figure ?). Although elincRNAs promoter regions are enriched at loop anchors relative to other LCL-expressed lincRNAs, no significant enrichment were found for these loci at TAD boundaries (fold enrichment=XX, p=XX, Methods, Figure 3). Despite the absence of significant elincRNA enrichment at TAD boundaries, they elincRNAs tend to be more frequently found near to the end of the TADs and are depleted at the center (bin 5, 0.37 fold, q=0.06) relative to other LCL-expressed lincRNAs (Figure 3). The trend is consistent with their enrichment at TAD boundaries-enriched loop anchors (fold enrichment=1.74, q<0.001, supplementary files).

This observed lack of significant enrichment of elincRNAs at TAD boundaries may be a consequence of the poor resolution of the current Hi-C technology, which is restrained to a maximum of 5 Kb resolution [REF], as well as limitations in the method used to define boundary regions. Particularly, I defined TAD boundaries by extendeing inward from borders of TADs (see methods for details), therefore all genes located in outside of defined TADs but are close to a TAD border are excluded from the analysis.

elincRNAs are associated with high DNA-DNA contacts within TADs

To further support their role in regulating promoter-enhancer contacts, I investigated whether elincRNAs are associated with regions with higher DNA-DNA interactions. To this end, I measured the average amount of contact in their respective TADs (see material and methods for details). I found that elincRNAs are frequently embedded within TADs with higher density of contacts compared to other lincRNAs in GM12878 (Fold difference=1.24, two-tailed Mann-Whitney U test, p<0.001, Figure 8). In addition, fold difference in the amount of DNA contacts within TADs harbouring these elincRNA loci were less pronounced in 3 other cell lines relative to other LCL-expressed lincRNAs (fold difference=1.05, 1.07, and 1.04, p=XX, XX, 0.472 in HUVEC, K562 and NHEK, respectively) compared to GM12878. This provides evidence that the association between elincRNA expression and chromosomal contacts is likely cell line-dependent.

Although my findings do not provide insights into the mechanisms through which elincRNAs may promote chromosomal architecture, their associated high DNA-DNA contacts, together with their enrichment in loop anchors and cohesin binding suggest a role for elincRNAs in the modulaion of promoter-enhancer looping within topological domains.

3. Figures and tables

Figure 1: Median expression levels of elincRNAs compared with other lincRNA and protein-coding genes (PCG) in GM12878. Median expression values are displayed in the boxes. Two-tailed Mann-Whitney test, \*\*\*P<0.001; – not significant

Figure 2: Comparison of sequence conservation through mammalian and primate evolution between elincRNAs, other lincRNAs and PCG. Averaged phastCons score is used as a measure. The green horizontal line represents the median conservation of ancestral repeats, which are assuming to be evolving neutrally. Two tailed Mann-Whitney test, \*\*\*P<0.001.

Figure 3: Enrichment of elincRNA promoter regions at TAD boundaries and loop anchors compared to other lincRNAs. Fold enrichments and associated q-values are displayed on the bars.

Figure 4: Enrichment of elincRNAs across TADs compared to other lincRNAs. Each bar represent a bin of 10% TAD length. The log10 of q-values are put in color codes to give an estimation of the confidence in each value.

Figure 5: Enrichment in architectural proteins in elincRNA promoter regions, compared to other lincRNA. Fold enrichment and associated q-values are displayed on the bars. SMC3 and RAD21 are subunits of cohesin.

Figure 6: Proportions of overlap between RAD21, SMC3 and CTCF peaks in the human genome.

Figure 7: Enrichment of CTCF and cohesin exclusive binding sites in promoter regions of elincRNAs compared to other lincRNAs. Fold enrichment and associated q-values are displayed on the bars.

Figure 8: Mean amount of DNA-DNA contact within TADs for elincRNAs compared to other lincRNAs across different cell lines. Set of genes as defined in GM12878 are used for all comparisons. Two tailed Mann-Whitney test, \*\*\*P<0.001; \*P<0.05; – non-significant

Figure 9: Schematic representation of TADs and loops and typical patterns observed in Hi-C matrices in corresponding to these structures. Center: Example of a Hi-C matrix visualized in Juicebox (Durand et al., 2016)⁠. These matrices are symmetric and only upper/lower triangles are therefore used to simplify the visualization. The darker pixels on the matrix contain more interactions. Left: Two separate TADs are observed as high interactions triangle on the matrix. These are examplified by regions of compacted DNA where frequent interactions occur, while interactions across TADs (i.e. between the blue and green triangles) are less frequent. A and B are the borders of the first TAD while C and D are the borders of the second TAD. Boundaries are the rectangles expanding inwards from the borders. Right: Representation of a loop with A and B being the anchors of the loop where strong contact is observed. Unlike TADs, the contact is not particularly high in the region between the two anchors, therefore loops are seen as a sharp increase in contacts deviating from the matrix diagonal.

Figure 10: Visual representations of the algorithms used to compute contacts in Hi-C matrices. As they are symmetric, matrices are represented as upper triangles for simplicity reasons. Left: Method used to compute the mean of all interactions in TADs. Each TAD is taken as a submatrix (upper triangles of the submatrices are depicted in blue) and the mean value in the submatrix is computed. Right: Schematic representation of the algorithm used to measure insulation. A diamond (blue) of width w set to 100kb is slid on all position along the diagonal. For each position, the sum in the diamond is computed and later used to define boundaries. The sum in the diamond at position d (dotted line) represents a measure of all interactions across position d (i.e. between elements before and after position d).

Figure 11: Example of the calculated sums of interactions through a TAD, how boundaries were extended until they reach the threshold and the corresponding TAD in the Hi-C matrix, visualized in Juicebox. The solid vertical lines represent the TAD borders, the horizontal dashed lines represent the threshold required to stop extending boundaries and the transparent areas represent the final boundaries. All blue elements relate to the left side, while all green elements relate to the right side.

Figure S1: Expression of elincRNAs compared with other lincRNAs and PCG in different cell lines. Sets of genes as defined in GM12878 are used in all comparisons. Two-tailed Mann-Whitney test, \*\*\*P<0.001; – not significant

Figure S2: Tissue specificity of elincRNA compared to other lincRNA and protein-coding genes. Tissue specificity index (Tau) is used as a measure of tissue specificity, with 1 being the highest possible specificity and 0 being the lowest. Median values are displayed on the boxes. Two tailed Mann-Whitney test, \*\*\*P<0.001; – non-significant

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.

Most active enhancers are transcribed (Arner et al., 2015)⁠ and the resulting transcripts are hypothesized to recruit transcription factors to increase transcriptional activity.

t is still uncertain whether this is achieved by the transcript itself, or the act of transcription, but experiments showed evidences for a trancript-dependent role of eRNAs in the formation of promoter-enhancer loops, as reducing their abundance reduced promoter-enhancer contacts (Li et al., 2013)⁠.

5. Materials and methods:

Unless stated otherwise, all statistical tests were performed using R 3.3.1 (R Core Team, 2016)⁠. Overlapping of genomic elements were done using either bedtools 2.26 (Quinlan & Hall, 2010)⁠or the “intervals” package (Bourgon, 2015)⁠ in R. Manipulations on Hi-C contact matrices were performed using the “Matrix” package (Bates & Maechler, 2016)⁠.

Genes

LincRNAs and protein-coding genes used were retrieved from the ENCODE website. The list of genes used in all analyses corresponds to genes expressed in the GM12878 lymphoblastoid cell line. Subcategories of genes were defined based on overlap between their promoter region, defined as the 1kb region upstream of the transcription start site and regulatory elements available on ENCODE (ENCODE Project et al., 2012)⁠. These regulatory elements are predicted computationally from Chip-seq data by a hidden Markov-model. Only predicted active promoters where considered when using promoters, and all enhancers when considering enhancers. The 2 categories of lincRNAs that are used throughout this report are elincRNAs, defined as overlapping enhancers but no promoters in their promoter region, and other lincRNAs defined as overlapping neither enhancer nor promoters in their promoter region.

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 encompass smaller ones were removed to preserve the signal from the boundaries of the small TADs. Boundaries from very large TADs would otherwise contain the signal from smaller TADs inside, generating noise.

Hi-C data and normalization

Contacts were calculated using Hi-C contact matrices from Rao et al (2014). All computations are performed on 5kb resolution matrices constructed from all read pairs mapping to the genome with a MAPQ score of at least 30. The matrices were normalized using the KR normalization vector provided by the authors whenever possible. 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 probably as a result of the high sparsity of the matrix. I chose SQRTVC as a substition for KR as the authors reported this method to yield very close results to KR.

The normalization procedure consists in dividing each entry in the contact matrix M by a corresponding value in the normalization vector V:

$M^\*\_{i,j}=\frac{M\_{i,j}}{V\_{KR}[\frac{i}{res}]\*V\_{KR}[\frac{j}{res}]}$

Where $M\_{i,j}$ is an entry from the raw matrix and $M^\*\_{i,j}$ corresponding normalized entry.

TAD boundaries definition

Boundaries are extended from TAD borders towards the interior of TADs using a custom algorithm (Figure 10). The method used to define boundaries relies on the assumption that boundaries are insulated regions. In other words, there are few interactions between elements before and after the boundaries. The insulation is measured by sliding a diamond on every position along the matrix diagonal and computing the sum in the diamond at each position. Lower values represent more insulated regions. The size of the diamond has been set to an arbitrary threshold of 100kb, considered reasonable as the median length of filtered TADs is 140kb.

More formally, the algorithm can be described as sliding a diamond of width w along the diagonal of a square matrix M of n dimensions on all positions d between w and n-(w-1). Those latter limits are set to prevent the diamond from getting out of the matrix. At each position, the sum of all values in the diamond is stored in a vector V. This can be rewritten as:

$\left\{\begin{matrix}1\leq w\leq \frac{n}{2}+1 \\ \forall d\in\left \{ w , ... , n-\left ( w-1 \right ) \right \}\end{matrix}\right. V\_{d}=\sum\_{i=d-(w-1)}^{d}\sum\_{j=d}^{d+(w-1)}M\_{i,j}$

The sums from the diamond are then used to compute boundaries. For all TADs, boundaries are extended inwards from the borders as long as the value of V does not exceed a threshold defined as the starting value (at the border) plus 10% of the maximum value in the TAD (Figure 11).

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 et al, under revision) through mammalian and primate evolution using phastCons scores (Siepel et al., 2005)⁠ and averaged phastCons score were used as a measure of sequence conservation. Tissue specificity index (Tau) was computed following the described in Kryuchkova & Robinson-Rechavi (2015⁠), considering only genes with expression above a cutoff of 0.1 RPKM.

Expression levels

Processed median expression data for elincRNAs and protein-coding genes in 4 different cell lines were calculated by Tan et al, under revision. The original data comes from ENCODE (ENCODE Project et al., 2012)⁠.

DNA-DNA contacts

For each gene overlapping a TAD, the mean contact inside the respective TAD was used as a measure. For single genes that overlap several TADs, the contacts are computed for each TAD independently. The mean contact in a TAD is computed by taking the arithmetic mean in a square submatrix spanning from the beginning to the end of the TAD in the intrachromosomal matrix.

Chip-seq

Chip-seq peaks for CTCF, RAD21 and SMC3 in GM12878 were retrieved from the ENCODE website (ENCODE Project et al., 2012)⁠. The CTCF and cohesin exclusive peaks were obtained by using the intersect and subtract tools from the bedtools suite.

Enrichment of genetic elements

All enrichment tests were performed using the genome association tester (GAT) (Heger et al., 2013)⁠ version 1.2. This program allows to test if genomic segments of interest are found in a desired set of annotations more often than expected if they were distributed randomly in a workspace. All tests using lincRNAs as annotations or segments were performed using the intergenic space of the genome as a workspace. When testing for enrichment of anchors at boundaries, the whole genome was used as the workspace. For all tests, the number of samples was set to 10,000, the number of buckets was consequently adjusted to 270,000 and segments overlap was used as the measure.

6. References:

Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., … Hayashizaki, Y. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science (New York, N.Y.), 347(6225), 1010–4. http://doi.org/10.1126/science.1259418

Bates, D., & Maechler, M. (2016). Matrix: Sparse and Dense Matrix Classes and Methods. manual. Retrieved from https://cran.r-project.org/package=Matrix

Bourgon, R. (2015). intervals: Tools for Working with Points and Intervals. manual. Retrieved from https://cran.r-project.org/package=intervals

Darrow, E. M., & Chadwick, B. P. (2013). Boosting transcription by transcription: enhancer-associated transcripts. Chromosome Research, 21(6), 713–724. article. http://doi.org/10.1007/s10577-013-9384-6

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Systems, 3(1), 99–101. http://doi.org/10.1016/j.cels.2015.07.012

ENCODE Project, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57–74. http://doi.org/nature11247 [pii]\n10.1038/nature11247

Engreitz, J. M., Ollikainen, N., & Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nature Reviews Molecular Cell Biology, 17(12), 756–770. http://doi.org/10.1038/nrm.2016.126

Gorkin, D. U., Leung, D., & Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell, 14(6), 771–775. http://doi.org/10.1016/j.stem.2014.05.017

Guil, S., & Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. Nature Structural & Molecular Biology, 19(11), 1068–1075. http://doi.org/10.1038/nsmb.2428

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., … Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature, 460(7253), 410–413. http://doi.org/10.1038/nature08079

Heger, A., Webber, C., Goodson, M., Ponting, C. P., & Lunter, G. (2013). GAT: A simulation framework for testing the association of genomic intervals. Bioinformatics, 29(16), 2046–2048. http://doi.org/10.1093/bioinformatics/btt343

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., … Chinnnaiyan, A. M. (2015). The Landscape of Long Noncoding RNAs in the Human Transcriptome. Nat Genet., 47(3), 199–208. http://doi.org/10.1038/ng.3192.The

Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., … Young, R. A. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell, 18(2), 262–275. http://doi.org/10.1016/j.stem.2015.11.007

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., & Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. BMC Biology, 11(1), 59. http://doi.org/10.1186/1741-7007-11-59

Kryuchkova, N., & Robinson-Rechavi, M. (2015). A benchmark of gene expression tissue-specificity metrics. bioRxiv, (January), 27755. http://doi.org/10.1093/bib/bbw008

Lam, M. T. Y., Li, W., Rosenfeld, M. G., & Glass, C. K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends in Biochemical Sciences, 39(4), 170–182. http://doi.org/10.1016/j.tibs.2014.02.007

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., … Rosenfeld, M. G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature, 498(7455), 516–20. http://doi.org/10.1038/nature12210

Lupiáñez, D. G., Spielmann, M., & Mundlos, S. (2016). Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Trends in Genetics, 32(4), 225–237. http://doi.org/10.1016/j.tig.2016.01.003

Marques, A. C., Hughes, J., Graham, B., Kowalczyk, M. S., Higgs, D. R., & Ponting, C. P. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. Genome Biology, 14(11), R131. http://doi.org/10.1186/gb-2013-14-11-r131

Ong, C., & Corces, V. G. (2014). CTCF : an architectural protein bridging genome topology and function. Nature Publishing Group, 15(4), 234–246. http://doi.org/10.1038/nrg3663

Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. American Journal of Human Genetics, 31(2), 106–15. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1685768&tool=pmcentrez&rendertype=abstract

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., … Gilbert, D. M. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature, 515(7527), 402–405. http://doi.org/10.1038/nature13986

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842. http://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2016). R: A Language and Environment for Statistical Computing. manual, Vienna, Austria. Retrieved from https://www.r-project.org/

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., … Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell, 159(7), 1665–1680. http://doi.org/10.1016/j.cell.2014.11.021

Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. Annual Review of Biochemistry, 81, 145–166. http://doi.org/10.1146/annurev-biochem-051410-092902

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., … Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Research, 15(8), 1034–1050. http://doi.org/10.1101/gr.3715005

Tamaru, H. (2010). Confining euchromatin/heterochromatin territory: Jumonji crosses the line. Genes and Development, 24(14), 1465–1478. http://doi.org/10.1101/gad.1941010

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., … Chang, H. Y. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature, 472(7341), 120–4. http://doi.org/10.1038/nature09819

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., … Shen, X. (2015). Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. Cell Stem Cell, 16(5), 504–516. http://doi.org/10.1016/j.stem.2015.03.007