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

1. Introduction :

It was only recently discovered that a surprisingly large proportion of the mammalian transcriptome does not code for proteins. To date, the number of annotated noncoding genes is more than 3 times higher than protein-coding genes (Iyer et al., 2015)⁠. Amongst all noncoding RNAs, long noncoding RNAs (>200bp) that do not overlap protein-coding genes are the most abundant (long intergenic noncoding RNAs, lincRNAs). Functional and evolutionary analyses, together with extensive characterization of a handful of lincRNAs, demonstrate that these transcripts are involved in gene regulation processes transcriptionally and post-transcriptionally, and that they can contribute to organismal traits and diseases (Kornienko, Guenzl, Barlow, & Pauler, 2013)⁠. However, the mechanisms of function, if any, for the majority of lincRNAs remain unknown (Rinn & Chang, 2012)⁠.

It is thought that spatial organization of the genome is an important factor in gene regulation (Engreitz, Ollikainen, & Guttman, 2016)⁠. Indeed, contrary to the traditional view, genomic DNA is not linear, but is folded into variably compact chromosomal structures that likely impact expression of the embedded genes (Gorkin, Leung, & Ren, 2014)⁠. On a global scale, regions with a high degree of compaction are classified as heterochromatin while relatively uncondensed regions are called euchromatin (Passarge, 1979)⁠. These are respectively associated with lower and higher levels of active transcription (Tamaru, 2010)⁠. Chromosomes are further compartmentalized into smaller domains, called topologically associated domains (TADs). The amount of DNA-DNA interactions is high within TADs as a result of their close spatial proximity, and low across TADs. TAD boundaries are the regions lying at the borders of TADs (Figure 9) and have been shown to be essential for gene regulation. They are often gene-dense and are enriched in highly transcribed genes (Ong & Corces, 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. TAD boundaries are also enriched in cohesin binding, 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)⁠.

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)⁠. Deletion of TAD boundaries often disrupts those interactions, resulting in gene misexpression and disease phenotypes (Lupiáñez, Spielmann, & Mundlos, 2016)⁠. Recently, some lincRNAs such as Haunt (Yin et al., 2015)⁠ or Firre (Engreitz et al., 2016)⁠, were reported to promote intra- or interchromosomal interactions, either by forming promoter-enhancer looping, or by regulating chromatin structural organization⁠. Furthermore, lincRNAs associated with human traits have also been shown to have enhancer-associated cis-regulatory roles and their loci are correlated with a higher density of DNA contact within TADs in a human lymphoblastoid cell line (LCL) (Tan et al, 2016, under revision). It has been shown that most active enhancers are transcribed, generating noncoding products, including lincRNAs (Guil & Esteller, 2012)⁠. This raises the question whether lincRNAs with enhancer-like activities (elincRNAs) contribute to gene regulation and the organization of these dynamic chromosomal structures. Those elincRNAs will therefore be the focus of my analysis.

Unlike most enhancer-associated noncoding RNAs, which are often transcribed bidirectionally and then rapidly degraded (Darrow & Chadwick, 2013)⁠, elincRNAs are transcribed preferentially in one direction (Marques et al., 2013)⁠ and are thereby likely good candidates to study the involvement of lincRNAs in the regulation of gene-enhancer interactions within chromatin domains. There are already a few functionally characterized elincRNAs, such as HOTTIP, that have been shown to act as a link between chromatin interactions and transcription (Wang et al., 2011)⁠. In this example, pre-established chromosomal looping is required for the nascent HOTTIP transcript to recruit the histone modifying complex WDR5-MLL1 to the the HOXA gene cluster, thus maintaining its active transcription.

Using various bioinformatics tools to analyze publicly available multi-omics data from the ENCODE project, I investigated the molecular properties of elincRNAs, their enrichment in different regulatory elements and their association with the amount of DNA-DNA interactions to gain insight into their roles in gene regulation within topological domains in human LCLs. My analysis shows elincRNAs are associated with high density of chromosomal contacts within TADs and are significantly enriched in loop anchors where promoter-enhancer interactions occur. I also find that they are strongly enriched in cohesin binding, supporting the idea that they may contribute to gene regulation by establishing contacts between regulatory elements and modulating chromosomal organization.

2. Results

LincRNAs were divided into 3 categories based on overlap with enhancers, as predicted by ENCODE in GM12878 LCLs (ENCODE Project et al., 2012)⁠ in their putative promoter regions (estimated as the region1kb upstream from their transcriptional start site) (Table 1, Methods). I excluded those that also overlapped predicted promoter elements as they are likely to have different molecular attributes and functions (Marques et al., 2013)⁠.

elincRNAs show similar expression levels to other lincRNAs

Enhancer-associated RNAs in general are known to have relatively low transcript abundances as they tend to be rapidly degraded by the exosome (Lam, Li, Rosenfeld, & Glass, 2014)⁠. I first tested if that is the case for my set of elincRNA compared to other lincRNAs and protein-coding genes (Figure 1). Although the elincRNAs median expression levels were lower relative to other lincRNAs , there was no significant difference (two-tailed Mann-Whitney U test, p=0.258) in GM12878. This similarity of expression between elincRNAs and other lincRNAs may be linked to the distinct features of lincRNAs. Notably, they are unidirectionally transcribed and often polyadenylated, whereas eRNAs comprise many short, non-polyadenylated unstable transcripts that are transcribed bidirectionally (Darrow & Chadwick, 2013)⁠.

elincRNAs show lower conservation than other lincRNAs

ElincRNAs are less conserved than other LCL-expressed lincRNAs as well as protein coding genes (Figure 2). These differences are observed both when looking at conservation in mammals and in primates. These observations are consistent with previous findings in mice (Marques et al., 2013) showing that elincRNAs are almost under neutral evolution. Interestingly, other studies recently showed (Tan et al, under revision) that exons of trait relevant lincRNAs are less conserved than other LCL-expressed lincRNAs through mammalian and primate evolution as well, but have been under strong purifying selection during recent human evolution. It would be worth investigating if this is true for elincRNAs as well.

ElincRNAs promoter regions are enriched at loop anchors but not TAD boundaries

Although elincRNAs promoter regions are not significantlyenriched at TAD boundaries relative to other LCL-expressed lincRNAs , they are enriched at anchors (Figure 3), where interactions between enhancer and promoter elements are known to occur (Ji et al., 2016)⁠. Although loop anchors are themselves enriched at TAD boundaries (1.74 fold, q<0.001, supplementary files), the lack of significant enrichment of elincRNAs at TAD boundaries may be a consequence of the method used to defined boundaries and the poor resolution of the current Hi-C technology. Notably, boundaries are extended from the TAD borders to the inside of TADs (see methods for details), therefore genes that are close to a TAD, but outside the border are not detected.

Despite the absence of elincRNA enrichment at TAD boundaries, dividing TADs into 10 equally sized bins (Figure 4) reveals that elincRNAs tend to be more frequently found near the end of the TAD and are depleted at the center of the TADs (bin 5, 0.37 fold, q=0.06) relative to other LCL-expressed lincRNAs. The trend is consistent with their enrichment at loop anchors, which are enriched at TAD boundaries.

ElincRNAs are enriched in cohesin

CTCF and cohesin are often called architectural or insulator proteins, as they are thought to prevent TADs from interacting with each other while increasing interactions within TADs. Enrichment tests for binding of these proteins in elincRNAs revealed that CTCF, SMC3 and RAD21 binding peaks were all highly enriched in elincRNA promoter regions (5.2-7.1 fold, q<0.001) but only slighly enriched in other lincRNAs (1.1-1.3 fold, q= 0.04-0.30) (Figure 5).

Most binding peaks for CTCF and cohesin overlap in the genome (Figure 6) and to determine if the enrichment of CTCF peaks is a consequence of the overlap with cohesin peaks, I performed enrichment tests for CTCF- and cohesin-exclusive binding (Figure 7). The difference in fold enrichment for cohesin binding sites in elincRNAs compared to other lincRNAs much stronger when looking at exclusive binding sites, while it decreased for CTCF. This suggests elincRNAs specifically may be involved in the formation of cohesin-only loops. According to a recent model (Ji et al., 2016)⁠, loops mediated by CTCF or CTCF and cohesin have insulator properties and are important for the structural maintenance of boundaries and the formation of insulated neighbourhood within TADs, while cohesin only loops mediate promoter-enhancers interactions. My results may point towards a role of elincRNAs in the formation of promoter-enhancer loops.

ElincRNAs are associated with high DNA-DNA contacts in TADs

Enhancer associated RNAs (eRNAs) are thought to stabilize long range promoter-enhancer interactions, as enhancers forming loops with TSS are more likely to express these eRNAs (Sanyal, Lajoie, Jain, & Dekker, 2012)⁠. To investigate whether elincRNAs are associated with regions of higher DNA-DNA contact, I used the average amount of contact in their respective TAD as a proxy (see material and methods for details). I find that elincRNAs are associated with TADs presenting higher amounts of contacts than other lincRNAs (Figure 8) in GM12878 (two-tailed Mann-Whitney U test, p<0.001), with a 1.24 fold increase in median contacts. When comparing the contacts for the same sets of genes in 3 other cell lines, these results are consistent, but less significant in HUVEC and K562 (two-tailed Mann-Whitney U test, p <0.05) with respectives fold increases in median contacts of 1.05 and 1.07 for elincRNAs compared to other lincRNAs. The amount of contacts was not significantly higher for elincRNAs in NHEK (two-tailed Mann-Whitney U test, p = 0.472) although the trend was still going in the same direction with a fold increase in median contacts of 1.04.

The strength of this association seems to be very cell line-dependent, but always pointing towards a higher amount of contacts for elincRNAs. Although these results do not give any insights into the mechanisms through which elincRNAs may promote contact, high DNA-DNA contacts, together with the enrichment of cohesin suggest a role for elincRNAs in the establishment of promoter-enhancer looping inside TADs.

3. Figures and tables

Table 1 : Categories of lincRNAs defined by overlaps of regulatory elements in their promoter regions.

|  |  |  |
| --- | --- | --- |
| Name | Overlap status | N. lincRNAs |
| elincRNA | Overlaps enhancers only | 236 |
| plincRNA | Overlaps promoter only. | 480 |
| Other lincRNA | Overlaps neither promoter, nor enhancers | 1756 |

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. 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.

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 Yihong, J., unpublished) using phastCons scores (Siepel et al., 2005)⁠. Tissue specificity index (Tau) was computed following the described in Kryuchkova & Robinson-Rechavi (2015⁠), considering only genes with expression above a 0.1 [RPKM?] cutoff.

Expression levels

Median expression levels were computed from [the GEUVADIS 1000 Genomes project expression matrix. The matrix contains expression data for 373 individuals.] → encode, source ?

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.

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 data for CTCF, RAD21 and SMC3 in GM12878 were retrieved from the ENCODE website.

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. 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.

Bourgon, R. (2015). intervals: Tools for Working with Points and Intervals. manual.

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.

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.

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

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature, 489(7414), 109–113. http://doi.org/10.1038/nature11279

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