10.10.2016 :

Hi-C experiments used to define TADs (Rao et al, 2014) are subject to biases caused by chromatine accessibility, nucleosome occupancy, restriction site-density and alignability. These bias have been corrected by coverage normalization. (and other algorithms, see supplemental data)

11.10.2016

Because of the large number of overlaps between TADs, I merged the overlapping TADs together to consider only the large ones. (used merge from BEDtools)

27.10.2016

#############################  
  
27/10/16  
  
Change of project direction due to differences in pcgene and lincRNA lengths, thus difficulty in determining threshold to call TADbound genes.  
  
New proposal:  
Do enhancer-associated lincRNAs (elincRNAs) contribute to chromosomal organization of TADs?  
  
  
  
Method:  
1. Determine lincRNAs/pcgenes (whose (a) promoter region or (b) promoter+gene body) that overlap predicted enhancer elements in LCL (elincRNA or ePCgenes).  
        - gene promoter region: +/- 1kb from gene TSS  
        - LCL enhancer elements: predicted by Encode project in GM12878  
        - also get non-elincRNAs/non-ePCgenes  
          
2. Test enrichment of elincRNA/ePCgenes at TAD boundaries.  
        - divide TAD into 10/20 bins (based on DNA interactions from Hi-C matrix)  
        - use GAT (genome association tester):  
                (a) segment: elincRNA/ePCgene/nonelincRNA/nonePCgene loci; annotation: TAD bins; workspace: whole human genome (hg19)  
                (b) segment: TAD bins; annotation: elincRNAs/ePCgenes/nonelincRNA/nonePCgene; workspace: whole human genome (hg19)  
                (c) segment: elincRNA/nonelincRNA; annotation: TAD bins; workspace: intergenic space of the genome (no pcgene)  
                (d) segment: TAD bins; annotation: elincRNA/nonelincRNA; workspace: intergenic space of the genome (no pcgene)  
                (e) segment: ePCgene/nonePCgene; annotation: TAD bins; workspace: all pcgene space   
                (f) segment: TAD bins; annotation: ePCgene/nonePCgene; workspace: intergenic all pcgene space   
                  
3. If you see an enrichment of elincRNAs/ePCgenes at TAD boundaries, is this specific to enhancer elements with bi-directional transcription or uni-directional transcription?  
        - Background: enhancers are all associated with transcriptional activity, and this activity is not specific to either direction, typically bi-directional  
        - elincRNAs/ePCgenes are transcribed only in one direction  
        - repeat tests in #2 for all predicted enhancer elements, is the enrichment at TAD boundaries still there?        
  
4. Functional characterization of these genes: expression level, subcellular fractionation, tissue specificity, conservation  
        - elincRNA vs. non-elincRNA  
        - ePCgene vs. non-ePCgene  
        - elincRNA vs. ePCgene  
          
5. Are elincRNAs enriched to overlap CTCF and cohesin binding sites?  
        - Use GAT to test for enrichment  
                (a) segment: elincRNA/nonelincRNA; annotation: CTCF/cohesin binding sites; workspace: intergenic space  
                (b) segment: CTCF/cohesin binding sites; annotation: elincRNAs/nonelincRNA; workspace: intergenic space  
                (c) segment: ePCgene/nonePCgene; annotation: CTCF/cohesin binding sites; workspace: all pcgene space  
                (d) segment: CTCF/cohesin binding sites; annotation: ePCgenes/nonePCgene; workspace: all pcgene space  
  
6. Is the expression levels of elincRNAs correlated with the amount of chromosomal interaction (Hi-C data) happening at the gene locus. Do the same for non-elincRNAs and pcgenes and compare.  
  
7. If #6 is true.  
On a more global scale, is the expression level of the lincRNA correlated with how much DNA:DNA interaction is happening within that locus across different tissues and cell types?  I can try to call expression of all lincRNAs in different human cell lines (ones with available matching Hi-C data). Then, you can download and measure chromosomal interactions at lincRNA locus in each cell line. Basically, you would create two large matrices, one for lincRNA expression levels and one for Hi-C contact at lincRNA locus. Then, you can test whether there is any correlations between the two. Finally, you can divide the lincRNAs into LCL TAD-bound and nonTAD-bound, and see if there is any differences there.   
        - can start with just one other cell line