Constructing a Transcription Factor Regulatory Network with Functional Impact on Cardiac Conduction and Rhythm Disorders

Ben F. Habermeyer

University of Pennsylvania

Constructing a Transcription Factor Regulatory Network with Functional Impact on Cardiac Conduction and Rhythm Disorders

Transcription constitutes the first step of gene expression and is highly regulated by a family of proteins known as transcription factors (TFs). TFs are broadly defined by their ability to (1) bind to DNA in a sequence-specific manner and (2) regulate transcription. TFs can have a greater than 1000-fold preference for sequence-specific binding to short DNA “motif” sequences, often represented in the form of a probability weight matrix (PWM) showing the information content of each base in the motifs (Lambert et al., 2018). These TF binding sequences are located within two classes of functional DNA elements: promoters and enhancers. Promoters are located upstream of the transcription start site (TSS) and include the core-promoter which directs the recruitment of RNA polymerase II to initiate transcription, whereas enhancer elements may be located up to 1Mb away from the core promoter either upstream or downstream of the TSS (Zabidi & Startk, 2016). TF binding to these elements may impact transcription by directly recruiting RNA polymerase, as is the case for TATA-binding protein which binds to the TATA-box within the core promoter, or by recruiting coregulators – often multi-domain proteins or protein complexes with domains involved chromatin binding, nucleosome remodeling, and covalent modifications of histones and other proteins involved in transcription (Farnham, 2011). While promoter-proximal TFs act in cis, TFs associated with distant enhancer elements physically interact with core promoters in trans through chromatin looping and are influenced by 3-D chromatin architecture (Pennacchio et al., 2013).

TFs are represented by 8% of all human protein-coding genes whose mutations are often deleterious, indicating their profound functional importance. Indeed, many exert control over developmental patterning, drive cell differentiation, or control pathways such as immune responses (Lambert et al., 2018). Furthermore, TFs which bind to enhancer elements exhibit cell type and temporal specificity and are largely responsible for shaping tissue specific gene-expression (Hu & Tee, 2017). Over 95% of genome-wide association study (GWAS) associated SNPs are located outside of coding sequences with more than 75% associated with DNAse hypersensitivity sites indicative of regulatory regions, indicating that mutations in enhancer elements may contribute significantly to human disease (Maurano et al., 2012). Therefore, a fundamental question to understanding gene expression at the transcriptional level is how genes are regulated by the coordinated efforts of many TFs. That is, how can the network of genes, regulatory elements, and transcription factors forming functional interactions be uncovered?

Advances in molecular biology and sequencing technology have led to the recent advent of bioassays which may elucidate such TF networks. A technology important to understanding genome-wide TF binding to enhancers and promoters is chromatin immunoprecipitation sequencing (ChIP-Seq). In this method, cell samples are treated with formaldehyde to crosslink DNA with any proteins, including TFs which may be bound to it. Next, the cells are lysed and the crosslinked chromatin is purified and sheared by sonication or micrococcal nuclease so that small (typically 200-600 bp) fragments of DNA remain, including the proteins bound to that DNA. A specific antibody is then applied to immunoprecipitate (IP) any DNA-protein fragments containing the TF or histone modification of interest. The crosslinking is then reversed by heating to release the DNA, which is then purified and prepared for short-read sequencing (Park, 2009). Alignment of sequencing reads to a reference genome is performed using an algorithm such as bowtie, and binding ‘peaks’ containing many reads which map across the same genomic loci indicative of TF-DNA interactions are identified using an algorithm such as MACS , which compares enrichment of sequencing reads at a given genomic position between IP and control samples (Langmead et al., 2009, Zhang et al., 2008). ChIP-seq is fundamental to constructing TF networks as it details the entire set of regulatory elements to which a TF binds, and it also may be used to assess the function of regulatory elements through IP on histone modifications. Thousands of ChIP-seq experiments have been performed on TF’s and histone modifications in many tissues and cell types as part of the ENCODE project, indicating the assay’s relevance to functional annotation of the genome and to studying disease (Dunham, et al., 2012). However, ChIP-seq is limited by its protein-centric approach as it may only assay one TF at a time, and its resulting binding profiles are highly-tissue and cell-type specific. Thus, many experiments must be performed to exhaustively cover a set of transcription factors binding to a gene of interest’s regulatory elements. ChIP-seq is further limited by its inability to connect trans-acting regulatory elements such as enhancers to target genes. More than 40% of known enhancer regions do not interact with their nearest promoters and instead ‘jump’ to their target promoters through chromatin looping, bypassing several intervening genes (Li, et al., 2012). While ChIP-seq is powerful for identifying predicted enhancer sequences, it is limited in its ability to elucidate the gene target(s) of the predicted enhancer. To fully construct a gene’s regulatory network, a method is therefore needed to connect its promoter with distal enhancer elements.

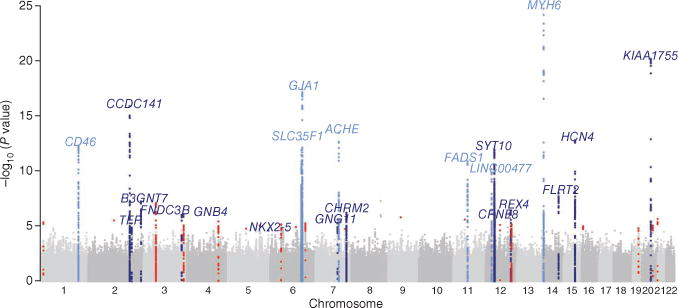
Questions regarding genomic organization and how the 3-D arrangement of DNA elements relates to functional interaction are answered using chromatin conformation capture (3C) based approaches (Davies, Higgs, & Hughes, 2017). More specifically, genome-wide linkage of genes’ cis-acting promoters to distal enhancer elements can be answered using promoter-capture Hi-C (CHi-C). In this method, formaldehyde is again used to crosslink DNA to protein factors, resulting in covalent links between spatially adjacent DNA elements. The DNA is cut using a restriction enzyme such as HindIII, and the DNA overhangs are filled and biotinylated. Biotinylated DNA segments in close proximity are then blunt-end ligated to each other. Crosslinks are reversed with proteinase or heating, and the DNA is sheared. Streptavidin is used to pull down biotinylated fragments which contain the DNA of two distal sequences ligated together (van Berkum et al., 2010). Unique to CHi-C, these fragments are then subject to solution hybridization selection, originally developed for exon sequencing, in which oligonucleotides complementary to known gene promoters are used to select for only fragments containing promoters (Misfud et al., 2015). This enrichment allows for greater sequencing depth of functionally important promoter-enhancer interactions. Fragments are then amplified and sequenced using paired-end sequencing, aligned to a reference genome, and significant interactions are identified using pipelines such as CHiCAGO, which corrects for background interactions and sequencing artifacts (Cairns, et al., 2016). The result of a CHi-C experiment is a set of promoters and their physical interacting partners in the genome, likely distal regulatory elements. Promoters regulating genes of interest may then be examined for regulatory partners to construct an interactome. This “all vs. all” method is DNA-centric, detailing how DNA segments interact but not the protein factors driving such an interaction, and is hence limited in its ability to declare the function of regulatory elements.

In this project, I combine these two genome-wide, protein-centric and DNA-centric approaches to map gene regulation on the molecular level and construct a TF regulatory network, focusing on the regulatory network in developing heart cells implicated in cardiac conduction and rhythm disorders (CRD). There is little information supporting the notion that genomic regions associated with adult phenotypes such as heart rate play important roles in cardiomyocytes differentiated from human embryonic stem cells (hESC-CMs), which have become a popular model to elucidate functional genomics of cardiac disease but have an immature phenotype. More generally, it is essential, in the context of potential cardiac regenerative medicine approaches, to understand how TFs drive specific interactions between promoters and enhancers in hESC-CMs to regulate gene expression patterns underlying cardiomyocyte development (Choy, et al, 2018). To construct a transcriptional regulatory network of TFs and genomic elements, I identified functional elements within CHi-C data which intersect significant CRD-associated SNPs from a GWAS. I then integrated these elements with ChIP-seq data and chromatin accessibility assays from the ENCODE project to construct a network of TFs which are shown to bind to these genomic elements, thus providing strong evidence for regulatory potential. My results show that integration of GWAS SNPs with genome-wide TF binding and regulatory element interactions in hESC-CMs may be used to (1) identify enhancers and transcription factors important in the regulation of genes within which significant SNPs lie, (2) identify novel enhancer elements at the genomic location of significant SNPs which do not reside in genes, and the downstream genes they affect, and (3) construct a ‘big picture’ network of genes, enhancer elements, and TFs in hESC-CMs development whose collective action impacts heart rate in adults.

# Methods

## Overview of Open-Source Data

Genomic loci important to the development of CRD in adult humans were identified from a two-stage GWAS was performed with 181,171 individuals to identify SNPs associated with perturbed heart rate features indicative of CRD (den Hoed et al., 2013). 21 SNPs were determined significant and are presented in Figure 1.



*Figure 1.* Manhattan plot of GWAS SNPs and association with heart rate phenotype. Significant SNPs (P < 5E-8) are annotated with their nearest human gene. Color represents stage of the 2-stage study in which SNP was identified (light blue stage 1, dark blue stage 2).

CHi-C data was obtained in browser extensible data (BED) format from the Gene Expression Omnibus, accession number GSE100720, with coordinates mapped to the hg19 genome assembly (Choy et al., 2018). Three biological replicates of 20 million hESC-CMs were prepared and CHi-C was performed using the HiCUP pipeline (Wingett et al., 2015). CHiCAGO was used to identify promoter-genome interactions, and 107,145 unique promoter-interacting regions were identified, encompassing 18,169 unique promoters.

Genomic data including TF and histone ChIP-seq and DNAse peaks were obtained from the Cistrome Data Browser (CistromeDB) (Mei et al., 2017). This browser is a wrapper for ENCODE data, taking its raw sequencing read data sets and performing the ChiLin chromatin profiling pipeline on them (Qin et al., 2016). This pipeline maps the reads from all datasets to the newest hg38 genome assembly, performs quality control analysis, and provides a useful interface to download BED file results or query for interaction with genome coordinates or gene names.

**Intersection of GWAS SNPs and CHi-C Interactions.** The UCSC genome browser LiftOver tool was used to convert CHi-C BED hg19 genomic coordinates to hg38 assembly coordinates for subsequent integration with CistromeDB (Kent, et al., 2002). Locations of the 21 SNPs associated with CRD were compiled using the UCSC genome browser and written to a BED file. GoShifter was used to assess significant overlap between trait-associated SNPs and genomic annotations from the CHi-C data (Tryanka, et al., 2015). BEDTools was used to find intersecting SNPs with CHi-C reads and the interacting DNA elements of each SNP-containing region (Quinlan & Hall, 2010).

**Intersection of SNP-Containing Interactions with ENCODE Data.** Genome coordinates from CHi-C dataset intersecting or interacting with a SNP region were entered in the CistromeDB toolkit to assess intersection with called peaks from TF binding, histone modification, and chromatin accessibility assays. Analysis was narrowed to only include interactions with data collected from cardiomyocytes or hESC cell types. Any intersection with the CHi-C regions was recorded.

**Motif Enrichment of SNP-Containing Interactions.** CHi-C chromosome coordinates spanning a CRD-associated SNP or interacting with such a region were converted into DNA sequences in FASTA format using UCSC. These sequences were examined for novel, ungapped motif enrichment using MEME suite (Bailey, et al., 2009). As a background for nucleotide frequency and motif distribution, 50 CHi-C reads which did not interact with the SNPs were randomly selected and converted to FASTA format. A discriminative search was conducted using MEME within the SNP CHi-C sequences against the custom background looking for motifs of length 7-24bp to encompass the size of most human TF binding motifs. Significant motifs were entered into the TOMTOM tool which matches a motif’s PWM to a database of transcription factors. The HOCOMOCO v11 database of human TFs was used to find the most probable binding factor of motifs enriched in the target sequences (Kulakovskiy, et al., 2015).

**Results**

Using GoShifter, a significant overlap of CRD GWAS SNPs with the CHi-C reads was observed (Bonferroni-adjusted p = 0.045), indicating enrichment for regulatory regions controlling genes involved in cardiac rhythm. Three of the 21 SNPs associated with CRD intersected with aligned CHi-C reads. These three SNPs formed a total of 21 interactions with a median read length of 2400bp containing potential regulatory elements. A summary of the intersecting SNPs is provided in Table 1 below. Two SNPs are within functionally important elements of genes and hence directly impact the gene’s product or transcription, whereas the other is in an intronic region suggesting its interacting partner could produce impact the CRD phenotype.

Table 1

*SNPs Associated with CRD Intersecting with CHi-C Interactions*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SNP | hg38 Locus | Effect | Nearest Gene | Affected Gene | # Interacting Reads |
| rs4140885 | chr2:187468087 | Within intron | TFP1 | ZSWIM2 | 1 |
| rs13245899 | chr7:100899261 | Near promoter | ACHE | ACHE | 5 |
| rs365990 | chr14:23392352 | Missense mutation | MYH6 | MYH6 | 15 |

Regulatory regions within SNP-intersecting CHi-C reads were identified on the basis of (1) the location of the read in relation to known genes, (2) the presence of open chromatin from DNAse assays, (3) enhancer or promoter-associated histone modifications, and (4) TF binding. Following these classification guidelines, 11 regulatory regions were identified encompassing 3 gene promoters and 8 associated enhancer elements and are summarized in Table 2, with check marks indicating the histone mark was present in at least one experiment of the relevant cell types. All 11 genomic regions showed open chromatin from DNAse digestion results and bound to at least 1 transcription factor from ENCODE ChIP-seq data. The 3 reads encompassing gene promoters each spanned the respective gene’s upstream 5’ end. One read containing an enhancer element was located on chromosome 2 and spanned an intronic sequence, whereas the other 7 were not located within genes. The reads also intersected with many modified histone peaks from Chip-Seq experiments. All 3 promoters showed the presence H3K27me3, indicating that these genes can be repressed for developmental control. H3K4me1 and H3K4me3 marks associated with poised/active enhancer and promoter elements, respectively, were seen in these reads, as were H3K27ac marks associated with active enhancers (Bannister & Kouzarides, 2011). Transcriptionally active regions marked by H4K5ac and regions marked by H3K9me3 transcriptional elongation as well as gene poor heterochromatin were also seen in predicted enhancer elements (Choy et al., 2018)

Table 2

*Regulatory Regions and their Histone Modifications*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Regulatory Element | Chromosomal position | H3K27me3 | H3K4me1 | H3K27ac | H3K4me3 | H4K5ac | H3K9me3 |
| Promoter | chr2:186846152-186854172 | ✓ | ✓ | ✓ | ✓ |  |  |
| Promoter | chr7:100888404-100902940 | ✓ |  | ✓ |  |  |  |
| Promoter | chr14:23388605-23392917 | ✓ | ✓ |  |  |  |  |
| Enhancer | chr2:187467461-187468552 |  |  |  |  | ✓ |  |
| Enhancer | chr7:101059257-101080334 | ✓ | ✓ | ✓ |  |  |  |
| Enhancer | chr7:101218044-101218857 |  |  | ✓ | ✓ |  | ✓ |
| Enhancer | chr14:23326999-23328981 |  |  |  |  |  | ✓ |
| Enhancer | chr14:23377635-23382458 | ✓ | ✓ |  | ✓ |  | ✓ |
| Enhancer | chr14:23435539-23436924 |  | ✓ | ✓ |  |  | ✓ |
| Enhancer | chr14:23440020-23444181 | ✓ | ✓ | ✓ | ✓ | ✓ |  |
| Enhancer | chr14:23450335-23454718 | ✓ | ✓ | ✓ | ✓ |  | ✓ |

For each of the 3 regions which intersected SNPs, the corresponding interacting regions from CHi-C were located on the same chromosome. The median distance between interacting reads was 59kb, indicating the sequences contained distal regulatory elements rather than the promoter of an adjacent gene. A schematic of the long-range regulatory elements is provided in Figure 2. For simplicity, enhancer regions are referred to as “E” followed by their chromosome number and order on the linear chromosome.

A close up of a map

Description automatically generated

*Figure 2.* Schematic of Long-Range Chromatin Looping Impacting Gene Regulation Important to CRD.

Intersection of ENCODE TF ChIP-seq peaks with these regions containing 8 enhancers and 3 promoters resulted in 53 interactions between 32 unique TFs. Enhancers were found to bind to slightly fewer (4.3) TFs as compared to promoters (5). Based on the number of enhancer and promoter regions a TF bound to, the most important TFs to the regulation of these 3 genes were CTCF, which interacted with 4 regulatory regions, as well as GATA4, YY1, SRF, and SMAD3, each of which was found to bind 3 regulatory regions. A transcriptional network of TFs and their interacting promoter and enhancer elements is presented in Figure 3.

A close up of a map

Description automatically generated

*Figure 3.* Network of TFs and Regulatory Elements. Size of Node Represents Relative Number of Interacting Partners

Motif enrichment was performed within the SNP-interacting CHi-C reads to confirm the presence of motifs for TFs from ChIP-seq data. Significant enrichment was found for many of the TFs (E-value < 0.05) including SMAD3, SP1-2, and SRF. This analysis also found many significantly enriched motifs and corresponding TFs which were not found in the ChIP-seq intersecting analysis, suggesting that even more factors may be implicated in the regulatory network. An example is the enriched motif in Figure 4, (E-value = 3.5E-8) and its best-match TF ZN770 from HOCOMOCO (E-value=3.62E-11). CistromeDB does not have any ChIP-seq experiments for ZN770 conducted in cardiomyocytes or hESCs, so this TF was not included in the network analysis.

A picture containing indoor

Description automatically generated

*Figure 4.* ZN770 Motif (upper) vs Enriched Motif in SNP-Interacting CHi-C Reads (lower)

**Discussion**

Uncovering full complement of TFs and their interacting regulatory elements is a fundamental, unanswered question with vast implications to understanding tissue-specific gene regulation, development, and disease. Twin studies with electrocardiogram (ECG) data have shown that genetic factors contribute to variation in heart rate, with heritability estimates ranging from 55 to 77% (den Hoed, et al., 2013). Hence, the development of CRD likely has a genetic background rooted in regulation of gene expression. In this project, I attempted to answer the question of constructing a comprehensive network of TFs and regulatory elements, focusing on a subset of interactions which are important to proper heart development. Intersecting CHi-C data from hESC-CMs with significant SNPs associated with CRD from a GWAS, I identified 3 loci which are parts of long-distance chromatin interactions within developing cardiomyocytes. Two of these regions are extremely deleterious, and hence the SNP itself is likely causal of CRD symptoms. These include a nonsynonymous mutation in an exon of MYH6 whose protein product is a myosin heavy chain, and a SNP just upstream of the 5’ end in ACHE which encodes acetylcholinesterase. Both genes have OMIM pages with affected phenotypes exhibiting CRD symptoms. The third SNP-containing region is within a large intronic sequence of a pseudogene, suggesting the SNP is not associated with the gene itself, but rather the region acts as a regulatory element for a distal gene. This region, termed E2 within chromosome 2, is located within the TFP1 pseudogene, and interacts with the ZSWIM2 gene which produces a ubiquitin ligase. Open chromatin within the TFP1 exon shown by DNAse, and intersection of this region with CEBPB enhancer binding protein ChIP-seq peaks suggest that this region is acting as an enhancer for distal genes. While no other ChIP-seq regions were found to interact with this enhancer, there is evidence this region is a unique regulator of ZSWIM2, whose faulty regulation could produce CRD phenotypes.

Of the 8 enhancer regions identified in this project, 5 are present on the UCSC genome browser’s regulation tracks as known regulatory elements, confirming CHi-C’s ability to link functional enhancers with promoters. By intersecting these enhancer regions and their interacting promoters with ChIP-seq peaks from experiments conducted in cardiomyocytes and hESC-CMs, I was able to uncover a network of binding to these regions and identify TFs which are important to the regulation of genes associated with CRD during heart development. Of the most prominent TFs by number of interactions with regulatory elements, CTCF bound the most, agreeing with what is known about CTCF and its role in mediating chromatin looping and subdividing topologically-associating domains within the nucleus (Ghirlando & Felsenfeld, 2016). The next most enriched TFs include GATA4 and SRF. The GATA factors GATA4 and GATA6 redundantly regulate the onset of cardiac differentiation. Similarly, SRF modulates expression of multiple cardiac genes (Bruneau, 2013). The regulatory network constructed therefore connects functionally-important TFs to target genes which they modulate in the development of heart cells.

Constructing transcriptional regulatory networks from CHi-C and ChIP-seq data presents a unique opportunity to get a holistic view of gene regulation. Data from multiple GWAS may be overlapped with the results from these molecular tools to get a subnetwork of interactions implicated in a specific disease. However, the methods presented have some limitations. CHi-C uses restriction enzymes to digest DNA, resulting in long DNA sequences (average 4kb) which may span multiple regulatory elements, and therefore may mask specific interactions. Furthermore, both CHi-C and ChIP-seq are highly tissue-specific – this prevents them from capturing functional interactions across multiple time points or tissues, and the resulting TF network construction is limited by the availability of ChIP-seq data in the tissue of interest. As I have shown, this analysis may be complemented by motif enrichment for TF binding sites in CHi-C reads, but this is not a substitute for experimental data.

**Conclusion & Future Directions**

The results of this project show that CHi-C and ChIP-seq data may be used to answer 3 fundamental questions related to transcriptional regulation, each of which have been exemplified in this project. First, these data sets in combination with GWAS SNPs may uncover regulatory elements of genes whose perturbation leads to detrimental phenotypes, as exemplified by the 7 enhancer regions identified for ACHE and MYH6. Second, this data may be used to identify novel functions of SNPs not associated with functional genes, as shown by the interaction between a probable enhancer in TFP1 and its interacting promoter of the gene ZSWIM2. Lastly, these results may be combined to produce a network of transcription factors, genes, and their regulatory elements important to development and disease. Building off this work one could examine the ubiquitin ligase product of ZSWIM2 to identify its functional importance in the development of CRD, since its dysregulation appears to cause significant association with disease. More ChIP-seq and GWAS experiments can be conducted and intersected with this CHi-C dataset to uncover transcriptional networks impacting other cardiac phenotypes.

References

Lambert, S. A., Jolma, A., Campitelli, L. F., Taipale, J., Hughes, T. R., … Weirauch, M. T. (2018). The Human Transcription Factors. *Cell, 172*(4), 650-665.

Zabidi, M. A., & Stark, A. (2016). Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Cell 32*(12), 801-814.

Frietze S., Farnham P.J. (2011). Transcription Factor Effector Domains. In: Hughes T. (eds) A Handbook of Transcription Factors. Subcellular Biochemistry, vol 52. Springer, Dordrecht

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: Five Essential Questions. *Nature Reviews Genetics* *14*(4), 288-295.

Hu, Z., & Tee, W. (2017). Enhancers and Chromatin Structures: Regulatory Hubs in Gene Expression and Diseases. *Bioscience Reports 37*(2),

Maurano, M, T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., … Reynolds, A. P. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science 337*(6099), 1190-1195.

Park, P. (2009). ChIP-seq: Advantages and Challenges of a Maturing Technology. *Nature Reviews Genetics* *10*(1), 669-680.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and Memory Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biology 10*(25),

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology, 9(9), 137

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C., A., … Doyle, F. (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature 489*(57), 57-74.

Li, G. Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., … Wang, P. (2012) Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell 148*(1-2), 84-98.

Davies, K. Oudelaar, A., Higgs, D, & Hughes, J. (2017). How to Best Identify Chromosomal Interactions: A Comparison of Approaches. *Nature Methods 14*(2), 125-134.

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., … Lander, E. S. (2010). Hi-C: A Method to Study the Shree-Dimensional Architecture of Genomes. *Journal of Visualized Experiments: JoVE*, (39), 1869.

Misfud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., … Ferreira, L. (2015). Mapping Long-Range Promoter Contacts in Human Cells with High-Resolution Capture Hi-C. *Nature 47*(1), 598-606.

Cairns, J., Freire-Pritchett, P., Wingett, S. W., Várnai, C., Dimond, A., Plagnol, V., … Spivakov, M. (2016). CHiCAGO: Robust Detection of DNA Looping Interactions in Capture Hi-C Data. *Genome biology, 17*(1), 127.

Choy, M., Javierre, B. M., Williams, S. G., Baross, S. L., Liu, Y., … Wingett, S. W. (2018) Promoter Interactome of Human Embryonic Stem Cell-Derived Cardiomyocytes Connects GWAS Regions to Cardiac Gene Networks. *Nature Communications* *9*(2526),

den Hoed, M., Eijgelsheim, M., Esko, T., Brundel, B. J., Peal, D. S., Evans, D. M., … Loos, R. J. (2013). Identification of Heart Rate-Associated Loci and their Effects on Cardiac Conduction and Rhythm Disorders. *Nature genetics, 45*(6), 621–631.

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., & Andrews, S. (2015). HiCUP: Pipeline for Mapping and Processing Hi-C Data. *F1000Research, 4*(1), 1310.

Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., … Zang, C. (2017). Cistrome Data Browser: A Data Portal for ChIP-Seq and Chromatin Accessibility Data in Human and Mouse, *Nucleic Acids Research, 45*(D1), 658–662.

Qin, Q., Mei, S., Wu, Q., Sun, H., Li, L., Taing, L., … Liu, X. S. (2016). ChiLin: A Comprehensive ChIP-seq and DNase-seq Quality Control and Analysis Pipeline. *BMC Bioinformatics, 17*(1), 404.

Kent W., Sugnet C., Furey T., Roskin K., Pringle T., Zahler A., Haussler D. (2002). The Human Genome Browser at UCSC. *Genome Res.12*(6), 996-1006.

Trynka, G., Westra, W., Slowikowski, K., Hu, X., Xu, H., Stranger, B., … Klein, R. (2015). Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-Coding Variants within Complex-Trait Loci. *The American Journal of Human Genetics, 97*(1), 139-152.

Quinlan, A., & Hall, I. (2010). BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics, 26*(6), 841–842.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic acids* *research, 37*(Web Server issue), W202–W208.

Kulakovskiy, I. V., Vorontsov, I. E., Yevshin, I. S., Soboleva, A. V., Kasianov, A. S., Ashoor, H., … Makeev, V. J. (2015). HOCOMOCO: Expansion and Enhancement of the Collection of Transcription Factor Binding Sites Models. *Nucleic acids research, 44*(D1), D116–D125.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of Chromatin by Histone Modifications. *Cell research, 21(3),* 381–395.

Ghirlando, R., & Felsenfeld, G. (2016). CTCF: Making the Right Connections. *Genes & Dev, 30*(1), 881-891.

Bruneau, B. (2013). Signaling and Transcriptional Networks in Heart Development and Regeneration. *Cold Spring Harbor Laboratory Press*