**Experience of Using ENCODE and other public data sets to identify regulatory elements with STARR-Seq data**

Intelligent identification of candidate enhancers in specific biological contexts requires the appropriate integration of existing ENCODE data with the appropriate cell-type and disease specific data sets. Gerstein Lab has a considerable track record of participating in, and collaborating with, the existing ENCODE DAC and AWG groups. Logically one must first identify and choose the candidate enhancers to be test, before experimental assays are performed to test them. One can incorporate new types of datasets to ENCODE. For example, whole genome STARR-seq data in K562 cells produced by Kevin White’s laboratory are already contributing to the algorithms being developed by the ENCODE DAC and functional characterization AWG.

Figure 1 outlines the high-level work flow of our approach for nominating candidate enhancers for functional characterization testing in the Center. These data include histone, chromatin accessibility, expression profiling and transcription factor ChIP-Seq experiments. In general, we use machine learning algorithms to identify candidate enhancers genome-wide. The resulting candidate enhancers are further processed by computational pipelines that identify genetic variants and predict their effects on candidate enhancer function, and by algorithms that match candidate enhancers to candidate target genes.

In general, we begin by integrating existing ENCODE data and results with existing data that include (but are not limited to) chromatin modification profiling, chromatin accessibility profiling, transcription factor mapping, whole genome sequencing, GWAS, and eQTL studies. From these analyses we identify an initial set of candidate enhancers that we further characterize. Datasets that we apply our methods are listed in tabular form below (Table 1).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cell/Tissue Type** | **K4Me/K27Ac/EP300** | **Dnase/Faire/5C** | **RNAseq** | **Roadmap**  **K4me/K27ac/Dnase** |
| GM12878 (ENCODE) | 6b | 5b | >10b | 2a |
| K562 (ENCODE) | 12b | 11b | >10b | 2a |
| HeLa S3 (ENCODE) | 5b | 5b | 9b | 0 |
| HEPG2 (ENCODE) | 4b | 5b | >10b | 6a |
| HUVEC (ENCODE) | 14b | 5b | 2b | 0 |
| A549 (ENCODE) | 4b | 1b | >10b | 4a |
| MCF-7 (ENCODE) | 8b | 11b | >10b | 1a |
| SK-N-SH (ENCODE) | 7b | 3b | >10b | 0 |

Table 1. Identified ENCODE and GWAS datasets for analysis. Existing data generated by ENCODE and GWAS studies

have been identified. The number of each type of study, the marks identified and the cells/tissue associated with the data. Datasets are procured from the a) NIH Roadmap Epigenomics Consortia ([www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics](http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics), Bernstein Nat Biotech 2010), b) ENCODE consortium ([www.encodeproject.org,](http://www.encodeproject.org/) ENCODE Consortium Nature 2012), c) GWAS studies ([www.gwascentral.org](http://www.gwascentral.org/))

**Experience on Genome-wide identification of candidate enhancers:**

Using the ENCODE and auxiliary datasets outlined in Table 1, we predict high-confidence candidate enhancers for downstream functional analysis. Prior to inputting the data into our machine learning algorithms for candidate enhancer identification (described below), we processed the datasets in Table 1 using tools developed in our lab, namely PeakSeq (69) and MUSIC (70), which have been applied by the ENCODE consortium and to ENCODE and Roadmap Epigenomics Consortium (RMEC) data. The main bulk of the datasets are listed in Table 1, featuring existing functional genomics datasets from ENCODE and RMEC projects. Specifically, we utilize peaks from histone marks and transcription factors to locate the regulatory regions. We use the activating marks and transcription factors that associate with enhancers (H3K4me1, H3K27ac, H3K9ac, P300, DNase/FAIRE).

These data are then input into enhancer prediction algorithms. A variety of enhancer prediction methods have been employed by the ENCODE consortium to examine existing data sets. For example, Ren and

colleagues developed a Random-Forest based algorithm, RFECS (Random Forest based Enhancer

identification from Chromatin States)(71). Park and colleagues developed a supervised machine learning method to identify and classify enhancers using chromatin marks across multiple metazoan species studied in ENCODE and modENCODE(72). As part of the ENCODE and modENCODE projects, Gerstein, Snyder and colleagues have developed methods that integrate ChIP-seq, chromatin, conservation, sequence and gene annotation data to identify gene-distal enhancers(73), which they have partially validated(74). In addition, the Gerstein lab has also developed a tool that utilizes the pattern within the histone marks to predict active regulatory regions in each tissue or cell line. The performance of some of these enhancer prediction algorithms was compared by the DAC as part of the ENCODE Enhancer challenge and the pattern recognition- based algorithm developed by the Gerstein lab was one of the top performing algorithms for enhancer prediction in mouse forebrain. Some of the top performing methods are used to predict enhancers initially. We assess the datasets for each ENCODE cell line individually, and all in combination. From these analyses we identify both cell/tissue-specific candidate enhancers and candidate enhancers that are shared across cell types, including among ENCODE cell lines. For each cell line or tissue type we can expect tens of thousands of candidate enhancers, and in some cases more than 100,000, based on previous results(71, 72, 74). We also expect 50-70% of these enhancers to be cell type specific based on previous results (22, 74)**.**

Although we initially use the approach outlined above for genome-wide candidate enhancer identification, it is worth noting that presently the ENCODE Functional Characterization AWG is evaluating this and other approaches for enhancer identification, with candidate enhancer validation experiments being performed by the Kevin White lab using whole genome STARR-seq and by the Len Pennacchio lab using in vivo mouse reporter assays. We also have experience in ensemble approaches to combine predictions from the best performing methods based on the results from the current ENCODE Functional Characterization AWG, and from other future Functional Characterization Centers in the ENCODE consortium with whom we are working to test a common set of candidate elements.

**Experience in Characterizing Variation in Enhancer Predictions:**

Previously we have extensively analyzed patterns of variation in noncoding regions, along with their coding targets, creating the tool ncVAR for assessing genetic variation in TFBSs (75). In recent studies (62), we

have integrated and extended these methods to develop a prioritization pipeline called FunSeq (and

subsequently FunSeq2). FunSeq prioritizes variants with respect to their deleterious impact on many different types of noncoding functional elements, including TF binding sites, regulatory elements, and regions of open chromatin. It identifies the regions under strong selective pressure as estimated using the variant frequencies computed from the whole genome sequencing data in 1000 Genomes Project and uses these regions as sensitive and ultra-sensitive non-coding regions of the genome. For each noncoding mutation in a regulatory element, FunSeq analyzes the target of the affected regulatory element. Then it scores the impact of the variants and prioritizes them based on a number of factors like network connectivity and motif disruption. It identifies deleterious variants in many noncoding functional elements, including TF binding sites, enhancer elements, and regions of open chromatin corresponding to DNase I hypersensitive sites. We determine linkage of variants identified in candidate enhancers with variants identified in the GWAS datasets in Table 1. We thus use FunSeq to annotate candidate enhancers.

We have also characterized the regulatory elements in terms of their association with human diseases.

On this front, we have developed LARVA that can identify recurrently damaging non-coding mutations and prioritize them with respect to their significance. To estimate significance, LARVA utilizes models that estimate background mutation frequencies in non-coding elements using as features the functional genomics datasets from ENCODE and RMEC projects (Table 1). We use the whole genome sequencing datasets from 1000 Genomes Project, and polymorphism datasets from dbSNP and Exome Aggregation Consortium (ExAC) projects as reference backgrounds to filter out the non-causative mutations. We integrate the tissue specific expression quantitative trait datasets (eQTL) from GTex Project to generate evidence for the causal variants generated by the mutation STARR-Seq experiments. Extensive tissue-specific and allele-specific expression, vascular cell epigenome mapping, and in vitro functional studies identifying causal variants and enhancers are needed to validate the bioinformatic findings.

**Experience on Linking Candidate Enhancers to Targets:**

We have previously developed computational pipelines for identification of targets of candidate regulatory elements, including methods that can successfully identify targets for gene-distal and gene-proximal regulatory

elements. Our methods utilize the correlation between the gene expression levels and the activity of the

regulatory element to identify significantly correlating activity. We also utilize sets of chromatin conformation datasets, generated from experiments including 4C, 5C, Hi-C, and ChIA-Pet. Furthermore, the Gerstein lab recently developed a method, named ENGINE, for utilizing these datasets in a machine learning framework for assigning targets to regulatory elements. ENGINE is a new version of the component of FunSeq that performs enhancer-target matching (<http://papers.gersteinlab.org/papers/funseq2>). Table 1 shows that there are many conformation datasets from ENCODE and RMEC datasets that we can utilize by combining the correlation based target estimation with conformation datasets. In particular, the conformation datasets will narrow down the possible targets of candidate enhancers to a subset of regions in which each candidate enhancer interacts. We expect this to dramatically decrease the false positive rate of the correlation. Specifically, ENGINE computes the correlation of the activity at the candidate enhancer region (using ENCODE and RMEC datasets) with the expression levels of the genes that each candidate enhancer region has contact with. Then, ENGINE uses the expression levels and several statistics about the shapes of the histone modification and transcription factor binding signals as additional features and builds a random forest based prediction model to score the candidate target genes. We utilize the validated sets of enhancer-target gene linkages for training ENGINE. In our previous efforts, we have utilized ENCODE data to build a set of such linkages between candidate regulatory elements and their target genes.

**D. Bibliography**

1. Benoist C, Chambon P. In vivo sequence requirements of the SV40 early promotor region. Nature.

1981;290(5804):304-10. PubMed PMID: 6259538.

2. Khoury G, Gruss P. Enhancer elements. Cell. 1983;33(2):313-4. PubMed PMID: 6305503.

3. Banerji J, Olson L, Schaffner W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell. 1983;33(3):729-40. PubMed PMID: 6409418.

4. Gillies SD, Morrison SL, Oi VT, Tonegawa S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell. 1983;33(3):717-28.

PubMed PMID: 6409417.

5. Queen C, Baltimore D. Immunoglobulin gene transcription is activated by downstream sequence elements. Cell. 1983;33(3):741-8. PubMed PMID: 6409419.

6. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. Nat Rev Genet. 2013;14(4):288-95. Epub 2013/03/19. doi: nrg3458 [pii]

10.1038/nrg3458. PubMed PMID: 23503198.

7. Levine M. Transcriptional enhancers in animal development and evolution. Current biology : CB.

2010;20(17):R754-63. doi: 10.1016/j.cub.2010.06.070. PubMed PMID: 20833320.

8. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15(4):272-86. doi: 10.1038/nrg3682. PubMed PMID: 24614317.

9. Jeziorska DM, Jordan KW, Vance KW. A systems biology approach to understanding cis-regulatory module function. Seminars in cell & developmental biology. 2009;20(7):856-62. doi:

10.1016/j.semcdb.2009.07.007. PubMed PMID: 19660565.

10. Yuh CH, Bolouri H, Davidson EH. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science. 1998;279(5358):1896-902. PubMed PMID: 9506933.

11. mod EC, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, Washietl S, Arshinoff BI, Ay F, Meyer PE, Robine N, Washington NL, Di Stefano L, Berezikov E,

Brown CD, Candeias R, Carlson JW, Carr A, Jungreis I, Marbach D, Sealfon R, Tolstorukov MY, Will S, Alekseyenko AA, Artieri C, Booth BW, Brooks AN, Dai Q, Davis CA, Duff MO, Feng X, Gorchakov AA, Gu T, Henikoff JG, Kapranov P, Li R, MacAlpine HK, Malone J, Minoda A, Nordman J, Okamura K, Perry M, Powell

SK, Riddle NC, Sakai A, Samsonova A, Sandler JE, Schwartz YB, Sher N, Spokony R, Sturgill D, van Baren

M, Wan KH, Yang L, Yu C, Feingold E, Good P, Guyer M, Lowdon R, Ahmad K, Andrews J, Berger B, Brenner

SE, Brent MR, Cherbas L, Elgin SC, Gingeras TR, Grossman R, Hoskins RA, Kaufman TC, Kent W, Kuroda MI, Orr-Weaver T, Perrimon N, Pirrotta V, Posakony JW, Ren B, Russell S, Cherbas P, Graveley BR, Lewis S, Micklem G, Oliver B, Park PJ, Celniker SE, Henikoff S, Karpen GH, Lai EC, MacAlpine DM, Stein LD, White KP, Kellis M. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science.

2010;330(6012):1787-97. doi: 10.1126/science.1198374. PubMed PMID: 21177974; PMCID: PMC3192495.

12. Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, Li Z, Ishii H, Spokony RF, Chen J, Hwang L, Cheng C, Auburn RP, Davis MB, Domanus M, Shah PK, Morrison CA, Zieba J, Suchy S, Senderowicz L, Victorsen A, Bild NA, Grundstad AJ, Hanley D, MacAlpine DM, Mannervik M, Venken K, Bellen H, White R, Gerstein M, Russell S, Grossman RL, Ren B, Posakony JW, Kellis M, White KP. A cis-regulatory map of the Drosophila genome. Nature. 2011;471(7339):527-31. doi:

10.1038/nature09990. PubMed PMID: 21430782; PMCID: PMC3179250.

13. Negre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG, Feng X, Ahmad K, Russell S, White RA, Stein L, Henikoff S, Kellis M, White KP. A comprehensive map of insulator elements for the

Drosophila genome. PLoS Genet. 2010;6(1):e1000814. doi: 10.1371/journal.pgen.1000814. PubMed PMID:

20084099; PMCID: PMC2797089.

14. Slattery M, Ma L, Spokony RF, Arthur RK, Kheradpour P, Kundaje A, Negre N, Crofts A, Ptashkin R, Zieba J, Ostapenko A, Suchy S, Victorsen A, Jameel N, Grundstad AJ, Gao W, Moran JR, Rehm EJ, Grossman RL, Kellis M, White KP. Diverse patterns of genomic targeting by transcriptional regulators in Drosophila melanogaster. Genome research. 2014;24(7):1224-35. doi: 10.1101/gr.168807.113. PubMed PMID: 24985916; PMCID: PMC4079976.

15. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung MS,

Clawson H, Contrino S, Dannenberg LO, Dernburg AF, Desai A, Dick L, Dose AC, Du J, Egelhofer T, Ercan S, Euskirchen G, Ewing B, Feingold EA, Gassmann R, Good PJ, Green P, Gullier F, Gutwein M, Guyer MS, Habegger L, Han T, Henikoff JG, Henz SR, Hinrichs A, Holster H, Hyman T, Iniguez AL, Janette J, Jensen M, Kato M, Kent WJ, Kephart E, Khivansara V, Khurana E, Kim JK, Kolasinska-Zwierz P, Lai EC, Latorre I, Leahey A, Lewis S, Lloyd P, Lochovsky L, Lowdon RF, Lubling Y, Lyne R, MacCoss M, Mackowiak SD, Mangone M, McKay S, Mecenas D, Merrihew G, Miller DM, 3rd, Muroyama A, Murray JI, Ooi SL, Pham H, Phippen T, Preston EA, Rajewsky N, Ratsch G, Rosenbaum H, Rozowsky J, Rutherford K, Ruzanov P, Sarov M, Sasidharan R, Sboner A, Scheid P, Segal E, Shin H, Shou C, Slack FJ, Slightam C, Smith R, Spencer WC, Stinson EO, Taing S, Takasaki T, Vafeados D, Voronina K, Wang G, Washington NL, Whittle CM, Wu B, Yan KK, Zeller G, Zha Z, Zhong M, Zhou X, mod EC, Ahringer J, Strome S, Gunsalus KC, Micklem G, Liu XS, Reinke V, Kim SK, Hillier LW, Henikoff S, Piano F, Snyder M, Stein L, Lieb JD, Waterston RH. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. 2010;330(6012):1775-

87. doi: 10.1126/science.1196914. PubMed PMID: 21177976; PMCID: PMC3142569.

16. Niu W, Lu ZJ, Zhong M, Sarov M, Murray JI, Brdlik CM, Janette J, Chen C, Alves P, Preston E, Slightham C, Jiang L, Hyman AA, Kim SK, Waterston RH, Gerstein M, Snyder M, Reinke V. Diverse

transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans. Genome research.

2011;21(2):245-54. doi: 10.1101/gr.114587.110. PubMed PMID: 21177963; PMCID: PMC3032928.

17. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature.

2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PMCID: PMC3439153.

18. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science.

1998;281(5373):60-3. PubMed PMID: 9679020.

19. Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell. 2015;162(4):900-10. doi: 10.1016/j.cell.2015.07.038. PubMed PMID: 26276636; PMCID: PMC4642453.

20. Swamynathan SK, Piatigorsky J. Orientation-dependent influence of an intergenic enhancer on the promoter activity of the divergently transcribed mouse Shsp/alpha B-crystallin and Mkbp/HspB2 genes. J Biol

Chem. 2002;277(51):49700-6. doi: 10.1074/jbc.M209700200. PubMed PMID: 12403771.

21. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren

B. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012;488(7409):116-20. doi:

10.1038/nature11243. PubMed PMID: 22763441.

22. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA

elements in the human genome. Nature. 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID:

22955616; PMCID: 3439153.

23. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM. In vivo enhancer analysis of human conserved non-coding sequences. Nature.

2006;444(7118):499-502. doi: 10.1038/nature05295. PubMed PMID: 17086198.

24. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. ChIP-seq accurately predicts tissue-specific activity of enhancers.

Nature. 2009;457(7231):854-8. doi: 10.1038/nature07730. PubMed PMID: 19212405; PMCID: 2745234.

25. Dickel DE, Visel A, Pennacchio LA. Functional anatomy of distant-acting mammalian enhancers. Philosophical transactions of the Royal Society of London Series B, Biological sciences.

2013;368(1620):20120359. doi: 10.1098/rstb.2012.0359. PubMed PMID: 23650633; PMCID: 3682724.

26. Morcillo P, Rosen C, Dorsett D. Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics. 1996;144(3):1143-54. PubMed PMID: 8913756; PMCID: 1207607.

27. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, Myeroff

L, Lutterbaugh J, Jarrar A, Kalady MF, Willis J, Moore JH, Tesar PJ, Laframboise T, Markowitz S, Lupien M, Scacheri PC. Epigenomic enhancer profiling defines a signature of colon cancer. Science.

2012;336(6082):736-9. doi: 10.1126/science.1217277. PubMed PMID: 22499810; PMCID: PMC3711120.

28. Fraser P. Transcriptional control thrown for a loop. Current opinion in genetics & development.

2006;16(5):490-5. doi: 10.1016/j.gde.2006.08.002. PubMed PMID: 16904310.

29. Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Current opinion in genetics & development. 2005;15(2):136-44. doi:

10.1016/j.gde.2005.02.005. PubMed PMID: 15797196.

30. Song SH, Hou C, Dean A. A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Molecular cell. 2007;28(5):810-22. doi: 10.1016/j.molcel.2007.09.025. PubMed PMID: 18082606;

PMCID: 2195932.

31. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell. 2012;149(6):1233-44. doi:

10.1016/j.cell.2012.03.051. PubMed PMID: 22682246; PMCID: 3372860.

32. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Consortium F, Forrest AR, Carninci P, Rehli M, Sandelin A, Kawaji H, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jorgensen M, Dimont E, Arner E, Schmid C, Schaefer U, Medvedeva YA, Plessy C, Vitezic M, Severin J, Semple CA, Ishizu Y, Young RS, Francescatto M, Alam I, Albanese D, Altschuler GM, Arakawa T, Archer JA, Arner P, Babina M, Rennie S, Balwierz PJ, Beckhouse AG, Pradhan-Bhatt S, Blake JA, Blumenthal A, Bodega B, Bonetti A, Briggs J, Brombacher F, Burroughs AM, Califano A, Cannistraci CV, Carbajo D, Chen Y, Chierici M, Ciani Y, Clevers HC, Dalla E, Davis CA, Detmar M, Diehl AD, Dohi T, Drablos F, Edge AS, Edinger M, Ekwall K, Endoh M, Enomoto H, Fagiolini M, Fairbairn L, Fang H, Farach-Carson MC, Faulkner GJ, Favorov AV, Fisher ME, Frith MC, Fujita R, Fukuda S, Furlanello C, Furuno M, Furusawa J, Geijtenbeek TB, Gibson AP, Gingeras T, Goldowitz D, Gough J, Guhl S, Guler R, Gustincich S, Ha TJ, Hamaguchi M, Hara M, Harbers M, Harshbarger J, Hasegawa A, Hasegawa Y, Hashimoto T, Herlyn M, Hitchens KJ, Ho Sui SJ, Hofmann OM, Hoof I, Hori F, Huminiecki L, Iida K, Ikawa T, Jankovic BR, Jia H, Joshi A, Jurman G, Kaczkowski B, Kai C, Kaida K, Kaiho A, Kajiyama K, Kanamori-Katayama M, Kasianov AS, Kasukawa T, Katayama S, Kato S, Kawaguchi S, Kawamoto H, Kawamura YI, Kawashima T, Kempfle JS, Kenna TJ, Kere J, Khachigian LM, Kitamura T, Klinken SP, Knox AJ, Kojima M, Kojima S, Kondo N, Koseki H, Koyasu S, Krampitz S, Kubosaki A, Kwon AT, Laros JF, Lee W, Lennartsson A, Li K, Lilje B, Lipovich L, Mackay-sim A, Manabe R, Mar JC, Marchand B, Mathelier A, Mejhert N, Meynert A, Mizuno Y, de Lima Morais DA, Morikawa H, Morimoto M, Moro K, Motakis E, Motohashi H, Mummery CL, Murata M, Nagao-Sato S, Nakachi Y, Nakahara F, Nakamura T, Nakamura Y, Nakazato K, van Nimwegen E, Ninomiya N, Nishiyori H, Noma S, Nozaki T, Ogishima S, Ohkura N, Ohmiya H, Ohno H, Ohshima M, Okada-Hatakeyama M, Okazaki Y, Orlando V, Ovchinnikov DA, Pain A, Passier R, Patrikakis M, Persson H, Piazza S, Prendergast JG, Rackham OJ, Ramilowski JA, Rashid M, Ravasi T, Rizzu P, Roncador M, Roy S, Rye MB, Saijyo E, Sajantila A, Saka A, Sakaguchi S, Sakai M, Sato H, Satoh H, Savvi S, Saxena A, Schneider C, Schultes EA, Schulze-Tanzil GG, Schwegmann A, Sengstag T, Sheng G, Shimoji H, Shimoni Y, Shin JW, Simon C, Sugiyama D, Sugiyama T, Suzuki M, Suzuki N, Swoboda RK, t Hoen PA, Tagami M, Takahashi N, Takai J, Tanaka H, Tatsukawa H, Tatum Z, Thompson M, Toyoda H, Toyoda T, Valen E, van de Wetering M, van den Berg LM, Verardo R, Vijayan D, Vorontsov IE, Wasserman WW, Watanabe S, Wells CA, Winteringham LN, Wolvetang E, Wood EJ, Yamaguchi Y, Yamamoto M, Yoneda M, Yonekura Y, Yoshida S, Zabierowski SE, Zhang PG, Zhao X, Zucchelli S, Summers KM, Suzuki H, Daub CO, Kawai J, Heutink P, Hide W, Freeman TC, Lenhard B, Bajic VB, Taylor MS, Makeev VJ, Hume DA, Hayashizaki Y. An atlas of active enhancers across human cell types and tissues. Nature.

2014;507(7493):455-61. doi: 10.1038/nature12787. PubMed PMID: 24670763.

33. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G,

Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh

N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ,

Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature.

2015;518(7539):317-30. doi: 10.1038/nature14248. PubMed PMID: 25693563; PMCID: PMC4530010.

34. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D,

Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D,

Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer

A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75-82. doi: 10.1038/nature11232. PubMed PMID: 22955617; PMCID:

3721348.

35. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of

Regulatory Elements) isolates active regulatory elements from human chromatin. Genome research.

2007;17(6):877-85. doi: 10.1101/gr.5533506. PubMed PMID: 17179217; PMCID: 1891346.

36. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. Mapping and analysis of chromatin state dynamics in nine

human cell types. Nature. 2011;473(7345):43-9. doi: 10.1038/nature09906. PubMed PMID: 21441907; PMCID:

3088773.

37. May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Afzal V, Simpson PC, Rubin EM, Black BL, Bristow J, Pennacchio LA, Visel A. Large-scale discovery of enhancers from human heart tissue. Nature genetics. 2012;44(1):89-93. doi: 10.1038/ng.1006. PubMed PMID: 22138689; PMCID: 3246570.

38. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu XD. Reprogramming transcription by distinct classes of enhancers functionally

defined by eRNA. Nature. 2011;474(7351):390-4. doi: 10.1038/nature10006. PubMed PMID: 21572438;

PMCID: 3117022.

39. Moorman C, Sun LV, Wang J, de Wit E, Talhout W, Ward LD, Greil F, Lu XJ, White KP, Bussemaker

HJ, van Steensel B. Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(32):12027-32.

doi: 10.1073/pnas.0605003103. PubMed PMID: 16880385; PMCID: PMC1567692.

40. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type- specific gene expression. Nature. 2009;459(7243):108-12. doi: 10.1038/nature07829. PubMed PMID:

19295514; PMCID: 2910248.

41. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, Rubenstein JL, Rubin EM, Pennacchio LA, Visel A. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell. 2013;155(7):1521-31. doi:

10.1016/j.cell.2013.11.033. PubMed PMID: 24360275; PMCID: 3989111.

42. Sakabe NJ, Savic D, Nobrega MA. Transcriptional enhancers in development and disease. Genome biology. 2012;13(1):238. doi: 10.1186/gb-2012-13-1-238. PubMed PMID: 22269347; PMCID: 3334578.

43. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet.

2015;24(R1):R102-10. doi: 10.1093/hmg/ddv259. PubMed PMID: 26152199; PMCID: PMC4572001.

44. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu

H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N,

Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated

variation in regulatory DNA. Science. 2012;337(6099):1190-5. Epub 2012/09/08. doi:

10.1126/science.1222794. PubMed PMID: 22955828; PMCID: 3771521.

45. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nature genetics. 2013;45(2):124-30. doi:

10.1038/ng.2504. PubMed PMID: 23263488; PMCID: PMC3826950.

46. Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE, Stephens M, Gilad Y, Pritchard JK. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012;482(7385):390-4. doi: 10.1038/nature10808. PubMed PMID: 22307276; PMCID: PMC3501342.

47. Miller CL, Anderson DR, Kundu RK, Raiesdana A, Nurnberg ST, Diaz R, Cheng K, Leeper NJ, Chen CH, Chang IS, Schadt EE, Hsiung CA, Assimes TL, Quertermous T. Disease-Related Growth Factor and Embryonic Signaling Pathways Modulate an Enhancer of TCF21 Expression at the 6q23.2 Coronary Heart Disease Locus. PLoS Genet. 2013;9(7):e1003652. Epub 2013/07/23. doi: 10.1371/journal.pgen.1003652. PubMed PMID: 23874238; PMCID: 3715442.

48. Miller CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patlolla B, Assimes TL, Kaiser FJ, Perisic L, Hedin U, Maegdefessel L, Schunkert H, Erdmann J, Quertermous T, Sczakiel G. Coronary Heart Disease-Associated Variation in TCF21 Disrupts a miR-224 Binding Site and miRNA-Mediated Regulation. PLoS Genet.

2014;10(3):e1004263. Epub 2014/03/29. doi: 10.1371/journal.pgen.1004263. PubMed PMID: 24676100.

49. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Perisic L, Hedin U, Betsholtz C, Ruusalepp A, Franzen O, Assimes TL, Montgomery SB, Schadt EE, Bjorkegren JLM, Quertermous T. Integrative fine-

mapping of regulatory variants and mechanisms at coronary heart disease loci. in review. 2015.

50. Turner AW, Martinuk A, Silva A, Lau P, Nikpay M, Eriksson P, Folkersen L, Perisic L, Hedin U, Soubeyrand S, McPherson R. Functional Analysis of a Novel Genome-Wide Association Study Signal in SMAD3 That Confers Protection From Coronary Artery Disease. Arteriosclerosis, thrombosis, and vascular biology. 2016. doi: 10.1161/ATVBAHA.116.307294. PubMed PMID: 26966274.

51. Turner AW, Nikpay M, Silva A, Lau P, Martinuk A, Linseman TA, Soubeyrand S, McPherson R. Functional interaction between COL4A1/COL4A2 and SMAD3 risk loci for coronary artery disease.

Atherosclerosis. 2015;242(2):543-52. doi: 10.1016/j.atherosclerosis.2015.08.008. PubMed PMID: 26310581.

52. Beaudoin M, Gupta RM, Won HH, Lo KS, Do R, Henderson CA, Lavoie-St-Amour C, Langlois S, Rivas

D, Lehoux S, Kathiresan S, Tardif JC, Musunuru K, Lettre G. Myocardial Infarction-Associated SNP at 6p24

Interferes With MEF2 Binding and Associates With PHACTR1 Expression Levels in Human Coronary Arteries. Arteriosclerosis, thrombosis, and vascular biology. 2015;35(6):1472-9. doi: 10.1161/ATVBAHA.115.305534.

PubMed PMID: 25838425; PMCID: PMC4441556.

53. Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA. Regulatory variation in a

TBX5 enhancer leads to isolated congenital heart disease. Hum Mol Genet. 2012;21(14):3255-63. doi:

10.1093/hmg/dds165. PubMed PMID: 22543974; PMCID: PMC3384386.

54. Savic D, Park SY, Bailey KA, Bell GI, Nobrega MA. In vitro scan for enhancers at the TCF7L2 locus. Diabetologia. 2013;56(1):121-5. doi: 10.1007/s00125-012-2730-y. PubMed PMID: 23011354; PMCID: PMC3525810.

55. Savic D, Ye H, Aneas I, Park SY, Bell GI, Nobrega MA. Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism. Genome research. 2011;21(9):1417-25. Epub 2011/06/16. doi:

10.1101/gr.123745.111. PubMed PMID: 21673050; PMCID: 3166827.

56. Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome research. 2010;20(9):1191-7. Epub

2010/07/16. doi: 10.1101/gr.105361.110. PubMed PMID: 20627891; PMCID: 2928497.

57. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, Yao K, Kehoe SM, Lenz HJ, Haiman CA, Yan C, Henderson BE, Frenkel B, Barretina J, Bass

A, Tabernero J, Baselga J, Regan MM, Manak JR, Shivdasani R, Coetzee GA, Freedman ML. The 8q24

cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nature genetics.

2009;41(8):882-4. Epub 2009/06/30. doi: ng.403 [pii]

10.1038/ng.403. PubMed PMID: 19561607.

58. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei G, Yan J, Niittymaki I, Mecklin JP, Jarvinen H, Ristimaki A, Di-Bernardo M, East P, Carvajal-Carmona L, Houlston RS, Tomlinson I, Palin K, Ukkonen E, Karhu A, Taipale J, Aaltonen LA. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nature genetics. 2009;41(8):885-90. Epub 2009/06/30. doi: ng.406 [pii]

10.1038/ng.406. PubMed PMID: 19561604.

59. Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, Stephens RM, Harris TJ, Munroe

DJ, Wu X. Long-range enhancers on 8q24 regulate c-Myc. Proceedings of the National Academy of Sciences

of the United States of America. 2010;107(7):3001-5. Epub 2010/02/06. doi: 10.1073/pnas.0906067107. PubMed PMID: 20133699; PMCID: 2840341.

60. Sazonova O, Zhao Y, Nurnberg S, Miller C, Pjanic M, Castano VG, Kim JB, Salfati EL, Kundaje AB, Bejerano G, Assimes T, Yang X, Quertermous T. Characterization of TCF21 Downstream Target Regions

Identifies a Transcriptional Network Linking Multiple Independent Coronary Artery Disease Loci. PLoS Genet.

2015;11(5):e1005202. doi: 10.1371/journal.pgen.1005202. PubMed PMID: 26020271; PMCID: PMC4447360.

61. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013;339(6122):957-9. doi: 10.1126/science.1229259. PubMed

PMID: 23348506; PMCID: PMC4423787.

62. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, Das J, Abyzov A, Balasubramanian S, Beal K, Chakravarty D, Challis D, Chen Y, Clarke D, Clarke L, Cunningham F, Evani US, Flicek P, Fragoza R, Garrison E, Gibbs R, Gumus ZH, Herrero J, Kitabayashi N, Kong Y, Lage K, Liluashvili V, Lipkin SM, MacArthur DG, Marth G, Muzny D, Pers TH, Ritchie GR, Rosenfeld JA, Sisu C, Wei X, Wilson M, Xue Y, Yu F, Genomes Project C, Dermitzakis ET, Yu H, Rubin MA, Tyler-Smith C, Gerstein M. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science. 2013;342(6154):1235587. doi: 10.1126/science.1235587. PubMed PMID: 24092746; PMCID: PMC3947637.

63. Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, Etchin J, Lawton L, Sallan SE, Silverman LB, Loh ML, Hunger SP, Sanda T, Young RA, Look AT. Oncogene regulation. An

oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science.

2014;346(6215):1373-7. doi: 10.1126/science.1259037. PubMed PMID: 25394790; PMCID: PMC4720521.

64. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W. Genome-wide analysis of noncoding regulatory mutations in cancer. Nature genetics. 2014;46(11):1160-5. doi: 10.1038/ng.3101. PubMed PMID: 25261935; PMCID: PMC4217527.

65. Fredriksson NJ, Ny L, Nilsson JA, Larsson E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. Nature genetics. 2014;46(12):1258-63. doi:

10.1038/ng.3141. PubMed PMID: 25383969.

66. Melton C, Reuter JA, Spacek DV, Snyder M. Recurrent somatic mutations in regulatory regions of human cancer genomes. Nature genetics. 2015;47(7):710-6. doi: 10.1038/ng.3332. PubMed PMID: 26053494;

PMCID: PMC4485503.

67. Puente XS, Bea S, Valdes-Mas R, Villamor N, Gutierrez-Abril J, Martin-Subero JI, Munar M, Rubio- Perez C, Jares P, Aymerich M, Baumann T, Beekman R, Belver L, Carrio A, Castellano G, Clot G, Colado E,

Colomer D, Costa D, Delgado J, Enjuanes A, Estivill X, Ferrando AA, Gelpi JL, Gonzalez B, Gonzalez S,

Gonzalez M, Gut M, Hernandez-Rivas JM, Lopez-Guerra M, Martin-Garcia D, Navarro A, Nicolas P, Orozco M, Payer AR, Pinyol M, Pisano DG, Puente DA, Queiros AC, Quesada V, Romeo-Casabona CM, Royo C, Royo

R, Rozman M, Russinol N, Salaverria I, Stamatopoulos K, Stunnenberg HG, Tamborero D, Terol MJ, Valencia

A, Lopez-Bigas N, Torrents D, Gut I, Lopez-Guillermo A, Lopez-Otin C, Campo E. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015;526(7574):519-24. doi: 10.1038/nature14666. PubMed PMID: 26200345.

68. Kron KJ, Bailey SD, Lupien M. Enhancer alterations in cancer: a source for a cell identity crisis. Genome Med. 2014;6(9):77. doi: 10.1186/s13073-014-0077-3. PubMed PMID: 25473436; PMCID:

PMC4254433.

69. Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat

Biotechnol. 2009;27(1):66-75. doi: 10.1038/nbt.1518. PubMed PMID: 19122651; PMCID: PMC2924752.

70. Harmanci A, Rozowsky J, Gerstein M. MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. Genome biology.

2014;15(10):474. doi: 10.1186/s13059-014-0474-3. PubMed PMID: 25292436; PMCID: PMC4234855.

71. Rajagopal N, Xie W, Li Y, Wagner U, Wang W, Stamatoyannopoulos J, Ernst J, Kellis M, Ren B. RFECS: a random-forest based algorithm for enhancer identification from chromatin state. PLoS Comput Biol.

2013;9(3):e1002968. doi: 10.1371/journal.pcbi.1002968. PubMed PMID: 23526891; PMCID: PMC3597546.

72. Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, Parker SC, Gu T, Kundaje A, Riddle NC, Bishop E, Egelhofer TA, Hu SS, Alekseyenko AA, Rechtsteiner A, Asker D, Belsky JA, Bowman SK, Chen QB, Chen RA, Day DS, Dong Y, Dose AC, Duan X, Epstein CB, Ercan

S, Feingold EA, Ferrari F, Garrigues JM, Gehlenborg N, Good PJ, Haseley P, He D, Herrmann M, Hoffman MM, Jeffers TE, Kharchenko PV, Kolasinska-Zwierz P, Kotwaliwale CV, Kumar N, Langley SA, Larschan EN, Latorre I, Libbrecht MW, Lin X, Park R, Pazin MJ, Pham HN, Plachetka A, Qin B, Schwartz YB, Shoresh N, Stempor P, Vielle A, Wang C, Whittle CM, Xue H, Kingston RE, Kim JH, Bernstein BE, Dernburg AF, Pirrotta V, Kuroda MI, Noble WS, Tullius TD, Kellis M, MacAlpine DM, Strome S, Elgin SC, Liu XS, Lieb JD, Ahringer J, Karpen GH, Park PJ. Comparative analysis of metazoan chromatin organization. Nature. 2014;512(7515):449-

52. doi: 10.1038/nature13415. PubMed PMID: 25164756; PMCID: PMC4227084.

73. Yip KY, Alexander RP, Yan KK, Gerstein M. Improved reconstruction of in silico gene regulatory networks by integrating knockout and perturbation data. PLoS One. 2010;5(1):e8121. doi:

10.1371/journal.pone.0008121. PubMed PMID: 20126643; PMCID: PMC2811182.

74. Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, Rozowsky J, Birney E, Bickel P, Snyder

M, Gerstein M. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. Genome biology. 2012;13(9):R48. doi: 10.1186/gb-2012-13-9-r48.

PubMed PMID: 22950945; PMCID: PMC3491392.

75. Mu XJ, Lu ZJ, Kong Y, Lam HY, Gerstein MB. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. Nucleic Acids Res.

2011;39(16):7058-76. doi: 10.1093/nar/gkr342. PubMed PMID: 21596777; PMCID: PMC3167619.

76. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013;339(6123):1074-7. Epub 2013/01/19. doi:

science.1232542 [pii]