

# HACER: an atlas of human active enhancers to interpret regulatory variants

Jing Wang<sup>1,2</sup>, Xizhen Dai<sup>1</sup>, Lynne D. Berry<sup>1</sup>, Joy D. Cogan<sup>3</sup>, Qi Liu<sup>1,2,\*</sup> and Yu Shyr<sup>1,2,\*</sup>

<sup>1</sup>Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN, USA, <sup>2</sup>Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA and <sup>3</sup>Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN, USA

Received August 15, 2018; Revised September 07, 2018; Editorial Decision September 11, 2018; Accepted September 12, 2018

## ABSTRACT

Recent studies have shown that disease-susceptibility variants frequently lie in cell-type-specific enhancer elements. To identify, interpret, and prioritize such risk variants, we must identify the enhancers active in disease-relevant cell types, their upstream transcription factor (TF) binding, and their downstream target genes. To address this need, we built HACER (<http://bioinfo.vanderbilt.edu/AE/HACER/>), an atlas of Human Active Enhancers to interpret Regulatory variants. The HACER atlas catalogues and annotates in-vivo transcribed cell-type-specific enhancers, as well as placing enhancers within transcriptional regulatory networks by integrating ENCODE TF ChIP-Seq and predicted/validated chromatin interaction data. We demonstrate the utility of HACER in (i) offering a mechanistic hypothesis to explain the association of SNP rs614367 with ER-positive breast cancer risk, (ii) exploring tumor-specific enhancers in selective *MYC* dysregulation and (iii) prioritizing/annotating non-coding regulatory regions targeting *CCND1*. HACER provides a valuable resource for studies of GWAS, non-coding variants, and enhancer-mediated regulation.

## INTRODUCTION

Enhancers are distal regulatory DNA regions essential for the precise spatiotemporal control of gene expression (1–6). They are bound by transcription factors (TFs) and activate gene expression through interaction with the gene promoter in a cell-type-specific manner (7). Sequence variants within enhancers can alter transcription factor binding and/or disrupt enhancer-promoter interactions, resulting in gene expression dysregulation and disease (8–11). For instance, three enhancer variants have been shown to reduce *RET* expression by disrupting SOX10, GATA2 and

RARB binding and thus increase Hirschsprung disease risk (12). As another example, cancer-risk SNP rs6983267 has been found to increase TCF7L2 binding and enhancer activity to elevate *c-MYC* expression in colorectal cancer cells (13,14). Recent genome-wide association studies (GWAS) have found >88% of disease-risk variants lie in non-coding regions (15), especially enriched in enhancers (16). To identify, interpret, and prioritize enhancer risk variants, we first must identify active enhancers in disease-relevant cell types, their upstream transcription factor binding and their downstream target genes.

Genome-wide cell-type-specific enhancers can be identified based on clusters of TF binding and certain histone modification patterns observed in ChIP-Seq and/or based on accessible ‘open’ chromatin identified through DNase-Seq and FAIRE-Seq (17–23). These data types and approaches form the basis of several enhancer databases (24–27) (Figure 1). For example, ENCODE combines DNase and H3K27ac signals to predict enhancer-like regions across 47 human cell types (<http://zlab-annotations.umassmed.edu/enhancers/>). The Ensembl Regulatory Build applies a genome segmentation algorithm to DNase-Seq and ChIP-Seq datasets for 18 human cell types to assign the regulatory state of each base pair, including enhancers (24). The Segway encyclopedia provides functional elements annotation (such as promoters and enhancers) of 164 human cell types using ChIP-Seq, DNase-Seq, FAIRE-Seq and Repli-Seq (BioRxiv: <https://doi.org/10.1101/086025>). DENdb applies five methods to ChIP-Seq histone modification data to predict enhancers in 15 human cell-lines (25). dbSUPER (26) and SEA (27) are two super-enhancer databases that combine ChIP-Seq signals for TF binding and H3K27ac data for 102 and 99 human cell types, respectively.

Recent studies have shown that bi-directional enhancer RNA (eRNA) production, strongly correlated with enhancer activity (28,29), is a more direct and reliable indicator than TF binding or histone markers (30–32). The FANTOM5 project used Cap Analysis of Gene Expression (CAGE) tags to detect 43,011 putative enhancers based

\*To whom correspondence should be addressed. Tel: +1 615 322 6618; Fax: +1 615 936 2602; Email: qi.liu@vanderbilt.edu  
Correspondence may also be addressed to Yu Shyr. Tel: +1 615 936 0682; Fax: +1 615 936 2602; Email: yu.shyr@vanderbilt.edu

Database	Type	Data source	# of cell type	Expression	TF binding	Predicted					Validated		GWAS
	Identification	TF	Nearest	50 kb	FANTOM5	eQTL	4DGenome	Other studies	Enhancer	Gene	Enhancer-disease		
<b>HACER</b>	enhancer	GRO-Seq, PRO-Seq, CAGE	265	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<b>ENCODE</b>	enhancer	ChIP-Seq, DNase-Seq	47										
<b>Ensembl</b>	enhancer	ChIP-Seq	18										
<b>DENdb</b>	enhancer	ChIP-Seq, DNase-Seq	15	✓	✓	✓					✓		
<b>Segway</b>	enhancer	ChIP-Seq, DNase-Seq, FAIRE-Seq, Repli-Seq	164										
<b>dbSUPER</b>	super enhancer	ChIP-Seq	102				✓						
<b>SEA</b>	super enhancer	ChIP-Seq	99		✓	✓	✓						
<b>EnhancerAtlas 2.0</b>	enhancer	ChIP-Seq, DNase-Seq, CAGE, FAIRE-Seq	179		✓								
<b>HEDD</b>	enhancer-disease	ChIP-Seq, CAGE	111		✓			✓	✓				✓

Figure 1. Summary of features distinguishing HACER from exiting enhancer databases.

on bidirectional eRNA pairs (29). EnhancerAtlas 2.0 (33) and HEDD (34) are two comprehensive enhancer resources, which combine a large number of datasets including ChIP-Seq histone marker data and FANTOM55 CAGE profiles for 179 and 111 cell types, respectively. GeneHancer is a database of enhancer and enhancer–gene associations derived from multiple sources, embedded in the framework of GeneCards (35). Enhancer and enhancer–gene association across cell types are aggregated to generate a confidence score, which makes it difficult to explore cell-type-specific enhancer and interactions.

In comparison to CAGE signals, which are often dominated by highly abundant and stable RNA, nascent RNA sequencing approaches such as GRO-Seq and PRO-Seq are more sensitive to unstable eRNAs, thus offering increased coverage of enhancer regions (36); however, GRO/PRO-Seq data are either not used or not processed in a standard way to identify active enhancers. Even if enhancers are detected, they are scattered in the literature, and have not been collected in any database.

To study enhancer function, one fundamental step is to link enhancers with their upstream regulators and downstream target genes. TF ChIP-Seq provides a map of binding sites in enhancer regions, but connecting enhancers with their target genes remains challenging. The earliest and most common strategy has been to assign enhancers to the nearest gene (28,37–39) or to genes within a certain distance (40,41). Studies have shown, however, that enhancers can skip the nearest gene to regulate a more distal one, and the distance can be quite large (42). Recently, the FANTOM5 project used expression correlation between eRNA and promoters to predict regulatory links (29). The GTEx project employed expression quantitative trait locus (eQTL)

analysis to identify the impact on target gene expression of single-nucleotide polymorphisms (SNPs) within an enhancer (43). Compared with predictive approaches, chromosome conformation capture-based technologies (such as 4C, 5C, Hi-C, ChIA-PET, HiChIP and Capture Hi-C) provide more reliable data to find target genes (42). Although rapid progress in these technologies has led to a dramatic increase in chromatin interaction data, existing databases still provide limited information on enhancer-mediated regulation, especially for chromatin contacts detected by high-throughput experiments (Figure 1).

To fill these gaps and facilitate study of regulatory variants, we developed HACER, an atlas of human active and in-vivo-transcribed enhancers. HACER catalogues and annotates 1 676 284 enhancers in 265 human cell lines by integrating FANTOM5 CAGE profiles and reprocessing publicly available GRO/PRO-Seq data. To place enhancers within regulatory networks, HACER identifies 772 902 TF–enhancer bindings based on reanalysis of ENCODE ChIP-Seq data, as well as integrating data for a large number of predicted chromatin interactions, and most importantly, validated interactions from high-throughput experiments. Notably, HACER annotates ~6.5 million validated enhancer-promoter interactions, ~1.4 million from 4DGenome and ~5.1 million from chromosome conformation capture-based technologies (Table 1). HACER provides query tools to interpret disease risk SNPs, to explore disease-specific enhancers and enhancer–gene interactions, and to annotate/prioritize non-coding regulatory variants. HACER further allows for visualization of transcriptional regulatory networks made up of TFs, enhancers, and target genes, along with risk SNPs and eQTL variants. Cell-type-specific knowledge of TF–enhancer–gene interactions

**Table 1.** Database statistics

HACER	#
Cell line	265
Tissue	42
Enhancer	1 676 284
TF–enhancer binding	772 902
Validated enhancer–promoter interaction	6 460 619
eQTL	1 581 613
GWAS	3 435

bridges the gap between gene regulation and human disease and will greatly facilitate deciphering of the functional role of enhancer variants in disease risk. HACER is available at <http://bioinfo.vanderbilt.edu/AE/HACER/>.

## MATERIALS AND METHODS

### Active enhancer identification and quantification

HACER identifies enhancers based on bidirectional eRNA pairs detected in GRO-Seq, PRO-Seq, and CAGE profiles. We conducted an extensive search of GRO/PRO-Seq datasets in the NCBI GEO database and downloaded all available raw sequence data for human cell lines under normal culture conditions. In total, we obtained 211 GRO-Seq and 41 PRO-Seq datasets. After adapter trimming and low quality sequence removal by cutadapt (44), GRO-Seq reads or reverse-complemented PRO-Seq reads from FastX tools (45), longer than 15 bp, were aligned to human genome hg19 using Bowtie2 (46). Reads mapped to rRNA loci and reads with mapping quality less than 10 were removed. NRSA (<http://bioinfo.vanderbilt.edu/NRSA/>) was applied to identify active enhancers and to quantify their transcription (47). Enhancers detected via CAGE and their associated enhancer expression matrix in each cell type were downloaded from FANTOM5 (29,48). A CAGE enhancer was defined to be active in a cell type if transcribed in that cell type. Enhancer expression was quantified by log<sub>2</sub>-transformed number of reads per bp (log<sub>2</sub> density). Enhancers annotated in HACER were further cross-referenced to existing enhancer databases, including VISTA (<https://enhancer.lbl.gov/>) (49), ENCODE Enhancer-like Regions (<http://zlab-annotations.umassmed.edu/enhancers>), The Ensembl Regulatory Build ([https://useast.ensembl.org/info/genome/funcgen/regulatory\\_build.html](https://useast.ensembl.org/info/genome/funcgen/regulatory_build.html)) (24), and chromatin state segmentation by ChromHMM from ENCODE/Broad (<https://genome.ucsc.edu/cgi-bin/hgTrackUi?g=wgEncodeBroadHmm&db=hg19>) (50,51), as well as the super-enhancer database dbSUPER (<http://asntech.org/dbsuper/index.php>) (26). All the data in HACER is based on genome build hg19.

### TF–enhancer binding

To map TF binding sites within enhancers, we reanalysed ENCODE TF ChIP-Seq data. Seventeen cell lines with active enhancers annotated in HACER have corresponding ChIP-Seq data for 156 TFs in ENCODE. We downloaded peak files from ENCODE and assigned TF peaks in a cell

type to active enhancers in the same cell type under the following criteria: the enhancer, instead of a promoter, is closest to the peak; and the enhancer is within  $\pm 1$  kb from the peak. The TF was then considered to bind the enhancer. Both the closest site to enhancer, and the site with the highest score if multiple peaks were assigned to the enhancer were recorded for each TF–enhancer pair. In total, 772 902 TF–enhancer bindings, involving 156 TFs and 81 478 enhancers, were detected and annotated in HACER.

### Enhancer–gene interactions

HACER provides the most complete resource to-date of enhancer–gene interactions, including 8 525 861 predicted and 6 460 619 experimentally validated links. The predicted interactions are derived from four strategies: (i) assigning enhancers to the nearest gene; (ii) linking enhancers to genes within a distance of 50 kb; (iii) FANTOM5 (29) and (iv) GTEx (V7) (43). FANTOM5 predicts links based on the assumption that transcriptional activity of the enhancer and of the putative target gene transcriptional start site (TSS) are correlated across human cells (29). GTEx uses expression quantitative trait locus (eQTL) analysis to estimate the effect of enhancer variants on gene expression. Experimentally validated interactions were collected from 4DGenome (52) and chromatin interaction studies including Hi-C, ChIA-PET, HiChIP and Capture Hi-C (53–58). 4DGenome is a comprehensive database of chromatin interactions compiled through literature curation, covering both low and high-throughput experimental assays. HACER annotates an additional  $\sim 5.1$  million chromatin interactions based on studies not collected in 4DGenome, such as interactions from high resolution Capture Hi-C in two human blood cell types (53).

### Assignment of GWAS SNPs to enhancers

NHGRI-EBI GWAS Catalog data (hg19) was downloaded from <https://www.ebi.ac.uk/gwas/downloads> (59), and includes SNP ID, disease/trait, *P*-value for the association, odds ratio, and PMID. GWAS SNPs are assigned to HACER enhancers if the SNP falls within an enhancer region.

### Database implementation

The HACER website runs on a Linux-based Apache web server. PHP is used for server-side scripting. The database is organized and managed by MySQL. The web pages are constructed using HTML5 and rendered using the CSS library Bootstrap. A cross-platform JavaScript library jQuery is used to provide a responsive user-friendly front-end interface.

## RESULTS

### Statistics and usage

HACER currently catalogues and annotates 1 676 284 active, *in-vivo*-transcribed enhancers across 265 human cell types. It includes 772 902 TF–enhancer bindings, 6 460 619 experimentally validated enhancer–promoter interactions,

1 581 613 eQTL variants, 66 942 FANTOM5 eRNA-promoter co-expressions, and 3 435 GWAS SNPs (Table 1). A user-friendly web interface was developed to browse, query, and download the enhancer data; to report new GRO/PRO/CAGE datasets; and to send feedback (Supplementary Figure S1).

### Browsing HACER

Users can browse all enhancers by cell line (Supplementary Figure S2A). When the cell line of interest is selected, all enhancers in this cell line are listed in tabular format with the option to filter and customize by annotation (enhancer ID, genomic coordinates, nearest active gene, distance to nearest gene, detection technique, cell type, and/or cross-reference to existing enhancer databases) (Supplementary Figure S2B). Clicking the enhancer ID reveals further detailed information for each enhancer, organized on five tabs for different types of information: basic, TF binding, target genes, GWAS, and eQTLs (Supplementary Figures S2C–G). The ‘Basic’ tab not only provides basic information for the enhancer (name, cell line, location, closest active gene, etc.), but also displays the enhancer sequence, a box-plot illustrating all eRNAs expressed in the cell line with a highlighted dot representing the enhancer RNA level, and cross-reference to existing enhancer databases (Supplementary Figure S2C). The ‘TF binding’ tab lists TF binding sites within the enhancer, including both the closest site and the highest score of the binding site (Supplementary Figure S2D). The ‘Target genes’ tab lists all target genes interacting with the enhancer based on FANTOM5, 4DGenome, or chromatin interaction studies (Supplementary Figure S2E). The ‘GWAS’ (Supplementary Figure S2F) and ‘eQTL’ (Supplementary Figure S2G) tabs present, respectively, GWAS SNPs and eQTL variants falling within the enhancer region.

### Querying HACER

HACER provides four strategies to query the database: SNP-centric, gene-centric, genomic coordinate-centric, and batch query. When queried on a GWAS risk SNP, HACER returns the cell-type-specific enhancers in which the SNP is located, along with their transcription factor binding and target genes. The transcriptional regulatory network of TF–enhancer–target genes is visualized along with the SNP, providing insight into potential mechanism(s) through which the SNP increases disease risk (*Case study 1*). When queried on a gene, HACER returns all enhancers targeting this gene, providing insight into disease-specific enhancers and enhancer–gene interactions (*Case study 2*). When queried on a genomic region, HACER finds active enhancers overlapping the region. Batch query allows the user to input a set of non-coding variants or enhancers, which HACER then prioritizes by their functional importance (*Case study 3*).

*Case study 1: interpret breast-cancer risk SNP rs614367.* Breast cancer is the most common cancer among women worldwide. Genome-wide association studies have successfully identified multiple breast cancer susceptibility loci (60), with rs614367 among those with the strongest association specific to estrogen receptor (ER)-positive disease

(61,62). The SNP rs614367 is located in an intergenic region with multiple flanking genes, including *CCND1*, *MYEOV*, *ORAOV1*, *FGF19*, *FGF4* and *FGF3*, all of which are potential breast cancer susceptibility genes. To find the gene that mediates the association between the SNP and breast cancer risk, we queried HACER on rs614367 and limited the search of active enhancers and chromatin interactions only in MCF7, an ER-positive breast cancer cell line. Within genomic coordinates chr11:69316698–69335855, we found four active enhancers harbouring the SNP. These enhancers are bound by E2F1/GATA3/MYC/TCF7L2/ZNF217 and interact with *CCND1* and *MYEOV* in the MCF7 cell line (Supplementary Figure S3). One eQTL was predicted to affect *GAL* expression; however, this link was detected in colon but not breast tissue (blue line in Supplementary Figure S3). Thus HACER narrowed the target genes to two candidates: *MYEOV* and *CCND1*. *CCND1* functions as a mediator of estrogen-induced cell proliferation, and *CCND1* expression together with inactivation of pRb are features of tumors with poor response to endocrine therapies (63). Therefore, *CCND1* is highly likely to be the gene that mediates rs614367 breast cancer susceptibility: variants in the enhancer region (either rs614367 or other functional variants) may be hypothesized to drive risk by disrupting E2F1/GATA3/MYC/TCF7L2/ZNF217 binding and affecting *CCND1* expression. This hypothesis has been validated by a recent extensive fine-scale mapping of this region in 89 050 European subjects, which found *EKL4* and *GATA3* as mediators of the regulatory effect of this region on *CCND1* expression (64).

*Case study 2: explore tumor-specific enhancers targeting MYC.* MYC is one of the most commonly activated oncogenes in a broad spectrum of human cancers (65). Dysregulation of MYC is often caused by tumor-specific super-enhancers in the region surrounding the MYC gene (41,66–73). HACER allows for ready exploration of diverse tumor-specific enhancers that regulate the MYC gene. When we query HACER on the MYC gene in any cell type, we find 86 MYC enhancers that are formed within the 3 Mb MYC locus (chr8:127574516–130664889) (Supplementary Figure S4A). If we limit enhancers and enhancer-promoter interactions to HCT-116, a colon cancer cell line, we find seven enhancers upstream of the MYC gene (Supplementary Figure S4B); the first four and the last of these previously have been labelled as super-enhancers in HCT-116 (54,69). If we limit the query to K562, a leukemia cell line, we find two enhancers downstream of the MYC gene (Supplementary Figure S4C). One is ~440 kb downstream of MYC TSS, and the other is located at the distal end of the 3 Mb region (~1.9 Mb from MYC TSS) and has previously been labelled as a super-enhancer (54). Both regions have been found to regulate MYC expression in mouse leukaemia cell lines (74,75). In short, HACER is able to identify diverse tumor-specific MYC enhancers/super-enhancers, differing in size and location, and consistent with previous studies (54,76).

*Case study 3: prioritizing non-coding regulatory variants.* Identification of non-coding drivers from thousands of variants is difficult. HACER prioritizes non-coding regulatory

variants based on two assumptions: (i) variants within enhancer elements targeting disease genes are highly likely to be drivers, and (ii) regions with more annotation available are more likely to be functional. When queried on a list of non-coding regions (batch query) and a set of disease genes, HACER highlights the regions targeting the disease genes and ranks all regions by the availability of information on cell-type-specific enhancers, TF binding, target genes, and GWAS risk SNPs. As an example shown in Supplementary Figure S5, HACER is queried on a list of non-coding regions, including a breast cancer risk region (chr11:69328760–69328765), and ‘CCND1’ as the target gene of interest. HACER highlights the breast cancer risk region targeting *CCND1* and ranks all query regions by the availability of functional information (Supplementary Figure S5B). A network view is also provided to show upstream TFs and downstream targets of each query (Supplementary Figure S5B).

### Downloading enhancers, reporting new datasets and sending feedback

HACER provides a Download page for users to download all enhancers by cell type. To download enhancers, click on the cell type of interest; the resulting file lists every enhancer in this cell type with comprehensive genomic annotation (e.g., chromatin location, target genes, detection methods). HACER also provides a Contact page for users to report new PRO-Seq, GRO-Seq and CAGE datasets. A hyperlink to the raw data should be provided, from which we will download, and process the data to identify, annotate and deposit enhancers into HACER. Users also are encouraged to send feedback via the Contact page.

### DISCUSSION AND FUTURE DEVELOPMENT

HACER not only catalogues and annotates active, *in-vivo*-transcribed enhancers from a large number of human cell types, but also provides insight into enhancer-mediated regulation. Integrating active enhancer, TF–enhancer binding, and enhancer–promoter interaction data within a single data repository enables the study of complex transcriptional regulation that drives cell-specific gene activities. HACER provides tools to find the potential mechanism underlying the association of enhancer variants with disease risk, to study enhancer activity and enhancer–gene regulation across different cell lines, and to link non-coding variants with disease genes for narrowing down potential causal variants. HACER represents a valuable and unique resource for studies on regulatory variants and enhancer-mediated regulation.

Few cell lines have matched GRO/PRO/CAGE, TF ChIP-Seq and chromatin interaction data, which greatly limits the ability to study enhancer function in a cell-type-specific manner and to explore enhancer roles in human disease. We will update HACER immediately if new datasets are reported through the contact page. Meanwhile, we plan to update HACER semi-annually to match the latest GRO/PRO/CAGE-Seq and chromatin interaction data from public databases and literature. Because enhancers and TSSs both produce bidirectional transcripts, enhancers

(especially long enhancers) can be difficult to distinguish from alternative TSSs or long non-coding RNAs (lncRNAs) (36), resulting in possible false positive enhancer calls. In the future, we plan to add histone marker H3K4me3 and H3K4me1 along with RNA-Seq data to help distinguish enhancers from novel TSSs and to reduce false positives.

Disruption of enhancer function leads to disease via alteration in target gene regulation. Currently, HACER links enhancers with disease based on disease-associated genetic variants from the GWAS Catalog; gene-disease association data have not been integrated. To study the role of enhancers in human disease, we plan to add gene-disease association data from MalaCards (77), DISEASES (78) and DisGeNET (79) and provide tools to explore enhancer-disease association directly.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### FUNDING

National Cancer Institute [5U01 CA163056-05 to Y.S.]; Cancer Center Support Grant [2P30 CA068485-19 to Y.S.]. Funding for open access charge: National Cancer Institute [5U01 CA163056-05 to Y.S.].

*Conflict of interest statement.* None declared.

### REFERENCES

- Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W. *et al.* (2015) Chromatin architecture reorganization during stem cell differentiation. *Nature*, **518**, 331–336.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y. *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and Enhancer/Promoter function. *Cell*, **162**, 900–910.
- Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E. *et al.* (2013) Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature*, **504**, 306–310.
- Blackwood, E.M. and Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action. *Science*, **281**, 60–63.
- Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A. and Bejerano, G. (2013) Enhancers: five essential questions. *Nat. Rev. Genet.*, **14**, 288–295.
- Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T. and Kraus, W.L. (2011) A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, **145**, 622–634.
- Spitz, F. and Furlong, E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J. *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190–1195.
- Chen, C.Y., Chang, I.S., Hsiung, C.A. and Wasserman, W.W. (2014) On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC Med. Genomics*, **7**, 34.
- Khurana, E., Fu, Y., Chakravarty, D., Demicheli, F., Rubin, M.A. and Gerstein, M. (2016) Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.*, **17**, 93–108.
- Zhang, F. and Lupski, J.R. (2015) Non-coding genetic variants in human disease. *Hum. Mol. Genet.*, **24**, R102–R110.
- Chatterjee, S., Kapoor, A., Akiyama, J.A., Auer, D.R., Lee, D., Gabriel, S., Berrios, C., Pennacchio, L.A. and Chakravarti, A. (2016)

- Enhancer variants synergistically drive dysfunction of a gene regulatory network in hirschsprung disease. *Cell*, **167**, 355–368.
13. Wright, J.B., Brown, S.J. and Cole, M.D. (2010) Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol. Cell. Biol.*, **30**, 1411–1420.
  14. Pomerantz, M.M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M.P., Doddapaneni, H., Beckwith, C.A., Chan, J.A., Hills, A., Davis, M. *et al.* (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.*, **41**, 882–884.
  15. Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9362–9367.
  16. Corradin, O. and Scacheri, P.C. (2014) Enhancer variants: evaluating functions in common disease. *Genome Med.*, **6**, 85.
  17. Berman, B.P., Nibu, Y., Pfeiffer, B.D., Tomancak, P., Celniker, S.E., Levine, M., Rubin, G.M. and Eisen, M.B. (2002) Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 757–762.
  18. Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E. and Taipale, J. (2006) Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*, **124**, 47–59.
  19. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **459**, 108–112.
  20. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854–858.
  21. Song, L. and Crawford, G.E. (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.*, **2010**, pdb prot5384.
  22. Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G. and Aerts, S. (2015) Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. *PLoS Genet.*, **11**, e1004994.
  23. Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., Graf, S., Huss, M., Keefe, D. *et al.* (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.*, **21**, 1757–1767.
  24. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T. and Flicek, P.R. (2015) The ensembl regulatory build. *Genome Biol.*, **16**, 56.
  25. Ashoor, H., Kleftogiannis, D., Radovanovic, A. and Bajic, V.B. (2015) DENdb: database of integrated human enhancers. *Database*, **2015**, bav085.
  26. Khan, A. and Zhang, X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, **44**, D164–D171.
  27. Wei, Y., Zhang, S., Shang, S., Zhang, B., Li, S., Wang, X., Wang, F., Su, J., Wu, Q., Liu, H. *et al.* (2016) SEA: a super-enhancer archive. *Nucleic Acids Res.*, **44**, D172–D179.
  28. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182–187.
  29. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmid, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.
  30. Hah, N., Murakami, S., Nagari, A., Danko, C.G. and Kraus, W.L. (2013) Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.*, **23**, 1210–1223.
  31. Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A. *et al.* (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, **474**, 390–394.
  32. Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., Chapelle, A.L., Heidemann, M., Hintermair, C. *et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.*, **18**, 956–963.
  33. Gao, T., He, B., Liu, S., Zhu, H., Tan, K. and Qian, J. (2016) EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics*, **32**, 3543–3551.
  34. Wang, Z., Zhang, Q., Zhang, W., Lin, J.R., Cai, Y., Mitra, J. and Zhang, Z.D. (2018) HEDD: Human Enhancer Disease Database. *Nucleic Acids Res.*, **46**, D113–D120.
  35. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M. *et al.* (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*, **2017**, 1–17.
  36. Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A. and Lis, J.T. (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.*, **46**, 1311–1320.
  37. Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C. *et al.* (2013) Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell*, **51**, 310–325.
  38. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X. *et al.* (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, **498**, 516–520.
  39. Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M. *et al.* (2013) Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, **498**, 511–515.
  40. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. and Zhao, K. (2012) Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.*, **22**, 490–503.
  41. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I. and Young, R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
  42. Yao, L., Berman, B.P. and Farnham, P.J. (2015) Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. *Crit. Rev. Biochem. Mol. Biol.*, **50**, 550–573.
  43. Consortium, G.T. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
  44. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, **17**, 10–12.
  45. Pearson, W.R., Wood, T., Zhang, Z. and Miller, W. (1997) Comparison of DNA sequences with protein sequences. *Genomics*, **46**, 24–36.
  46. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
  47. Wang, J., Zhao, Y., Zhou, X., Hiebert, S.W., Liu, Q. and Shyr, Y. (2018) Nascent RNA sequencing analysis provides insights into enhancer-mediated gene regulation. *BMC Genomics*, **19**, 633.
  48. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S. *et al.* (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.*, **16**, 22.
  49. Visel, A., Minovitsky, S., Dubchak, I. and Pennacchio, L.A. (2007) VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic acids Res.*, **35**, D88–D92.
  50. Ernst, J. and Kellis, M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.*, **28**, 817–825.
  51. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.
  52. Teng, L., He, B., Wang, J. and Tan, K. (2015) 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics*, **31**, 2560–2564.
  53. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W.,

- Ewels, P.A. *et al.* (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.*, **47**, 598–606.
54. Schuijers, J., Manteiga, J.C., Weintraub, A.S., Day, D.S., Zamudio, A.V., Hnisz, D., Lee, T.I. and Young, R.A. (2018) Transcriptional dysregulation of MYC reveals common Enhancer-Docking mechanism. *Cell Rep.*, **23**, 349–360.
55. Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J. *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, **148**, 84–98.
56. Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A. *et al.* (2016) Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, **351**, 1454–1458.
57. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Rusczycki, B. *et al.* (2015) CTCF-Mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, **163**, 1611–1627.
58. Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H. *et al.* (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, **462**, 58–64.
59. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.
60. Ghousaini, M., Pharoah, P.D.P. and Easton, D.F. (2013) Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? *Am. J. Pathol.*, **183**, 1038–1051.
61. Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghousaini, M., Hines, S., Healey, C.S. *et al.* (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.*, **42**, 504–507.
62. Lambrechts, D., Truong, T., Justenhoven, C., Humphreys, M.K., Wang, J., Hopper, J.L., Dite, G.S., Apicella, C., Southey, M.C., Schmidt, M.K. *et al.* (2012) 11q13 is a susceptibility locus for hormone receptor positive breast cancer. *Hum. Mutat.*, **33**, 1123–1132.
63. Lange, C.A. and Yee, D. (2011) Killing the second messenger: targeting loss of cell cycle control in endocrine-resistant breast cancer. *Endocr. Relat. Cancer*, **18**, C19–C24.
64. French, J.D., Ghousaini, M., Edwards, S.L., Meyer, K.B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M.J., O'Reilly, M., Hillman, K.M. *et al.* (2013) Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am. J. Hum. Genet.*, **92**, 489–503.
65. Gabay, M., Li, Y. and Felsher, D.W. (2014) MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb. Perspect. Med.*, **4**, a014241.
66. Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G. *et al.* (2013) Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell*, **24**, 777–790.
67. Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S. and Engreitz, J.M. (2016) Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*, **354**, 769–773.
68. Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S.A., Belver, L., Wendorff, A.A., Xu, L., Castillo-Martin, M., Llobet-Navas, D., Cordon-Cardo, C. *et al.* (2014) A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.*, **20**, 1130–1137.
69. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
70. Lin, C.Y., Erkek, S., Tong, Y., Yin, L., Federation, A.J., Zaparka, M., Haldipur, P., Kawachi, D., Risch, T., Warnatz, H.J. *et al.* (2016) Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature*, **530**, 57–62.
71. Liu, M., Maurano, M.T., Wang, H., Qi, H., Song, C.Z., Navas, P.A., Emery, D.W., Stamatoyannopoulos, J.A. and Stamatoyannopoulos, G. (2015) Genomic discovery of potent chromatin insulators for human gene therapy. *Nat. Biotechnol.*, **33**, 198–203.
72. Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I. and Young, R.A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320–334.
73. Zhang, X., Choi, P.S., Francis, J.M., Imielinski, M., Watanabe, H., Cherniack, A.D. and Meyerson, M. (2016) Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.*, **48**, 176–182.
74. Shi, J., Whyte, W.A., Zepeda-Mendoza, C.J., Milazzo, J.P., Shen, C., Roe, J.S., Minder, J.L., Mercan, F., Wang, E., Eckersley-Maslin, M.A. *et al.* (2013) Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.*, **27**, 2648–2662.
75. Rathert, P., Roth, M., Neumann, T., Muerdter, F., Roe, J.S., Muhar, M., Deswal, S., Cerny-Reiterer, S., Peter, B., Jude, J. *et al.* (2015) Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature*, **525**, 543–547.
76. Bahr, C., von Paleske, L., Uslu, V.V., Remeseiro, S., Takayama, N., Ng, S.W., Murison, A., Langenfeld, K., Petretich, M., Scognamiglio, R. *et al.* (2018) A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature*, **553**, 515–520.
77. Rappaport, N., Nativ, N., Stelzer, G., Twik, M., Guan-Golan, Y., Stein, T.I., Bahir, I., Belinky, F., Morrey, C.P., Safran, M. *et al.* (2013) MalaCards: an integrated compendium for diseases and their annotation. *Database*, **2013**, bat018.
78. Pletscher-Frankild, S., Palleja, A., Tsafou, K., Binder, J.X. and Jensen, L.J. (2015) DISEASES: text mining and data integration of disease-gene associations. *Methods*, **74**, 83–89.
79. Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F. and Furlong, L.I. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, **2015**, bav028.