# The adapted Activity-By-Contact model for enhancer-gene assignment and its application to single-cell data

Dennis Hecker[1,2,3], Fatemeh Behjati Ardakani[1,2,3], Marcel H. Schulz[1,2,3*]

**1 Institute of Cardiovascular Regeneration, Goethe University and University Hospital, 60590 Frankfurt am Main, Germany,**
**2 Cardio-Pulmonary Institute, Goethe University, Frankfurt am Main, Germany**
**3 German Centre for Cardiovascular Research, Partner site Rhein-Main, 60590 Frankfurt am Main, Germany**

*** marcel.schulz@em.uni-frankfurt.de**

## Abstract

Identifying regulatory regions in the genome is of great interest for understanding the epigenomic landscape in cells. One fundamental challenge in this context is to find the target genes whose expression the regulatory regions affect. A recent successful method is the Activity-By-Contact (ABC) model [10] which scores enhancer-gene interactions based on enhancer activity and the contact frequency of an enhancer to its target gene. It requires two types of assays to measure enhancer activity, which limits the applicability. The ABC score describes regulatory interactions entirely from a gene's perspective, and does not account for all the candidate target genes of an enhancer. Moreover, there is no implementation available which would allow for an integration with transcription factor (TF) binding information or an efficient analysis of single-cell data. We show that the ABC-model is comparably accurate with only one assay to measure enhancer activity. Further, we demonstrate that it can yield a higher accuracy by adapting the enhancer activity according to the number of contacts the enhancer has to its candidate target genes. We combined our adapted ABC-model with TF binding information and illustrate an analysis of a single-cell ATAC-seq data set of the human heart. We were able to characterise cell type-specific regulatory interactions as well as to prioritise candidate TFs that drive cell type-specific expression. All executed processing steps are incorporated into our new computational pipeline STARE. The software is available at `https://github.com/schulzlab/STARE`.

## 1 Introduction

Unravelling the mechanisms behind gene expression regulation is a central task in epigenomics. Enhancers are key players in this process. They are accessible regions in the genome, which can be bound by transcription factors (TFs) in a sequence-specific fashion. Those TFs have a variety of functions: recruit other cofactors, remodel chromatin conformation, cause changes in epigenetic modifications or directly interact with the transcription machinery, affecting gene expression (31; 14; 22). Many methods exist for annotating enhancers, *e.g.*, using open-chromatin assays like DNase-, ATAC- or NOMe-seq (45; 5; 18). Histone modifications associated with enhancer activity like H3K27ac or H3K4me1 also aid enhancer annotation (15; 7). Besides a plethora of methods to define enhancers, another ongoing challenge is to identify target genes of enhancers, which is essential for understanding their function. These enhancer-gene interactions can span large distances and are insufficiently explained by linear proximity (51; 43). Many approaches exist to predict target genes of enhancers, *e.g.*, using correlation of enhancer activity and gene expression (47; 11; 41), or correlation of accessibility across samples [33]. Another possibility is to include chromatin contact data, *e.g.* Hi-C [25], to call chromatin loops for annotating enhancer-gene links (34; 40; 52). Although loops correlate with gene expression, their anchors only cover a fraction of active promoters and enhancers and their removal impacts

expression of only few genes (30; 35; 43). Fulco et al. (2019) [10] combined measurements of enhancer activity with chromatin contact data and proposed the Activity-By-Contact (ABC) model. The assumption is that active enhancers that frequently contact a gene's promoter are more likely to affect a gene's regulation. The ABC-model requires DNase-seq, H3K27ac ChIP-seq data and a Hi-C matrix. The Hi-C matrix can be substituted with a matrix averaged over multiple cell types, or with a quantitative function describing the distance-contact relationship [10]. The original ABC-model formulation is entirely gene-centric, which means it does not take the candidate target genes of an enhancer into account.
We could show that, instead of using both DNase- and H3K27ac ChIP-seq data for the ABC-model, one assay for measuring enhancer activity yields a similar accuracy. Further, we propose an adaptation to the ABC-score, which describes enhancer activity in a gene-specific manner. We validated the adaptation on data sets of experimentally tested enhancer-gene interactions in K562 cells. We developed STARE, a fast implementation of the ABC-score with an approach to quantify TF binding affinity for genes. STARE can compute enhancer-gene interactions from single cell chromatin accessibility data, illustrated on data of the human heart. With only one data modality at single-cell resolution, we identified cell type-specific regulatory interactions and prioritised TFs for expression regulation.

# 2 Methods

## 2.1 Activity-By-Contact score

We use the terms enhancer and regulatory region interchangeably. The principle of the ABC-model is that an enhancer, which is highly active and has a high contact frequency with a gene, is likely to regulate it [10]. The ABC-score represents the relative contribution of an enhancer $r$ to the regulation of gene $g$, measured by the enhancer's activity $A_r$ and its contact frequency $C_{r,g}$ with the promoter of $g$. For each candidate enhancer the activity is multiplied with the contact frequency and this product is taken relative to the sum over all candidate enhancers $R_g$ in a predefined window around the transcription start site (TSS) of a gene:

$$ABC(r,g) = \frac{A_r \cdot C_{r,g}}{\sum\limits_{i \in R_g} A_i \cdot C_{i,g}} \,. \tag{1}$$

By definition the scores per gene sum up to 1. In practice, a cut-off is used to select valid interactions. The ABC-model allows a many-to-many relationship: a gene may link to multiple enhancers and an enhancer may link to multiple genes. We tested three types of epigenomic assays to approximate $A_r$ by counting sequencing reads in the enhancer: DNase-seq, ATAC-seq, or H3K27ac ChIP-seq. The contact $C_{r,g}$ is taken from a normalised chromatin contact frequency matrix. A pseudocount is added to each $C_{r,g}$, so that all candidates $R_g$ are taken into account (see Supplementary Material).

## 2.2 Adapted Activity-By-Contact-score

We present an adaptation of the ABC-score, to account for all candidate target genes of an enhancer. The activity of an enhancer conceptually represents all regulatory interactions an enhancer has. However, an enhancer can interact with different genes and not all genes are equally likely to be brought into vicinity of the enhancer. We assume that an enhancer's regulatory input is a function of the number of contacts with all potential target genes. Target genes that are often in contact with the enhancer would receive more regulatory input than genes with fewer contacts. Thus, the activity $A_r$ could be denoted in a relative manner:

$$A_r = \sum_{g \in G_r} A_{r,g} \,, \tag{2}$$

where $A_{r,g}$ denotes the relative regulatory activity of enhancer $r$ towards gene $g$ and $G_r$ denotes the set of all genes that are located within a predefined window around $r$. Since the values of $A_{r,g}$ are not known, we propose an approximation by using chromatin contacts:

$$A_{r,g} \approx A_r \cdot \frac{C_{r,g}}{\sum_{j \in G_r} C_{r,j}} \,. \tag{3}$$

We approximate the regulatory activity $A_{r,g}$ by the relative fraction of its contact $C_{r,g}$ to the sum over the contacts to all genes $G_r$ in a window around the enhancer. Thus, the adapted ABC-score $\overline{ABC}_{r,g}$ becomes:

$$\overline{ABC}_{r,g} = \frac{A_{r,g} \cdot C_{r,g}}{\sum\limits_{i \in R_g} A_{i,g} \cdot C_{i,g}} \,. \tag{4}$$

In comparison to Eq. (1), the activity $A_r$ of an enhancer is replaced with its gene-specific relative activity $A_{r,g}$, see Eq. (3). Thus, the adapted $\overline{\text{ABC}}$-score does not only function in a gene-centric way, by considering the relative number of contacts for each gene to its candidate enhancers, but also in an enhancer-centric way, by adapting the activity to the relative number of contacts of an enhancer's candidate target genes. This adapted $\overline{\text{ABC}}$-score therefore uses a reduced enhancer activity estimate, in particular for enhancers in contact with many genes, and prevents them from being accounted with their full activity for all genes within the window.

## 2.3 Validation on CRISPR-screens

To validate different assays for measuring enhancer activity and to test our adapted $\overline{\text{ABC}}$-score, we examined the performance on experimentally validated enhancer-gene interactions. We made use of two CRISPR-screens for K562 cells. Fulco et al. (2019) [10] used a CRISPRi-FlowFISH approach and collected data from other CRISPR-based studies. Gasperini et al. (2019) [12] used a single-cell CRISPR screen, introducing guide RNAs at a high multiplicity of infection, followed by single-cell RNA-seq. Unlike Fulco et al. (2019) [10], we do not divide interactions into enhancer-gene and promoter-gene pairs and do not exclude interactions, where the expression decreased after enhancer perturbation. We tested four different setups for measuring enhancer activity: 1) DNase-seq, 2) H3K27ac ChIP-seq, 3) the geometric mean of DNase-seq and H3K27ac ChIP-seq, 4) and ATAC-seq. In addition, we assessed a K562 Hi-C matrix, and a Hi-C matrix averaged across 10 cell types, constructed by Fulco et al. (2019) [10]. For each combination we evaluated the ABC- and the $\overline{\text{ABC}}$-scoring approach and calculated precision-recall curves (more details in Supplementary Material).

### 2.3.1 Intersection with eQTLs

Further, we compared the agreement of the regular ABC- and the $\overline{\text{ABC}}$-score with expression quantitative trait loci (eQTL) data. We intersected ABC-scored interactions in K562 cells with significant eQTL-gene pairs identified in whole blood from the GTEx portal [48]. Since eQTLs are limited to variants present in the analysed population, we compared the number of eQTL-gene pairs found by the ABC- and $\overline{\text{ABC}}$-score as fraction of all eQTL-gene pairs, where the variant was located within a candidate enhancer.

## 2.4 TF affinities and summarisation on gene-level

In addition to enhancer-gene interactions, STARE aims to describe a TF's regulatory impact on a gene (Fig. S1). There are two steps required. First, genomic regions that influence the regulation of a gene have to be identified. This can be done via ABC-scoring as described before, or in a more simplistic approach by taking all open regions within a defined window around the gene's TSS into account. Utilising the ABC-model allows STARE to function with any desired window size around the TSS. Second, the affinities of TFs to the identified regions have to be quantified. Instead of relying on calling TF binding sites, we use the tool TRAP, which calculates relative binding affinities for TFs in a genomic region. TRAP describes TF binding with a biophysical model to predict the number of TF molecules that bind in a sequence [36]. The higher the affinity, the more likely a TF is to bind. Retaining low binding affinities can hold valuable information (38, 20) and omits selection of an arbitrary threshold. For all analyses presented in this manuscript we used a non-redundant collection of 818 human TF motifs in the form of position frequency matrices (PFMs) from JASPAR 2022, HOCOMOCO v11, and the work of Kheradpour and Kellis (6; 21; 19). When converting

PFMs to position-specific energy matrices required by TRAP, we take the average nucleotide content of the candidate regulatory regions as background.
To summarise the TF affinities in enhancers per gene, we combine them with the predicted enhancer-gene interactions. The summarisation depends on how the region-gene mapping was done. For the window-based approach, TF affinities in all regions around the gene's TSS are summed:

$$af_{g,tf} = \sum_{r \in R_g} \frac{af_{r,tf}}{ml_{tf}} \cdot A_r \cdot e^{-\frac{d_{r,g}}{d_0}} , \tag{5}$$

where $af_{g,tf}$ is the affinity of TF $tf$ summarised for gene $g$. $R_g$ is the set of all regions $r$ that were located within the window around $g$. $af_{r,tf}$ is the affinity of $tf$ in $r$, $ml_{tf}$ is the motif length of $tf$ and $A_r$ is the activity for $r$. The affinity is corrected for the distance $d_{r,g}$ of $r$ to the TSS of $g$ by an exponential decay function, as proposed by Ouyang et al. (2009), where $d_0$ is set to a constant of 5000 bp.
When the $\overline{\text{ABC}}$-score was used to assign regions to genes, the summarisation changes, as there is more epigenomic information available. Regions close to the TSS ($\leq 2500$ bp) are always included, independent of their $\overline{\text{ABC}}$-score, as they are very informative for the expression regulation of a gene [39]. They are also scaled differently:

$$af_{g,tf} = \sum_{r \in R_g} \frac{af_{r,tf}}{ml_{tf}} \cdot \begin{cases} A_r \cdot e^{-\frac{d_{r,g}}{d_0}}, & \text{if } d_{r,g} \leq 2500\ bp \\ A_{r,g}, & \text{otherwise} \end{cases} , \tag{6}$$

where $R_g$ is a set of regions that was linked to the gene with the $\overline{\text{ABC}}$-score and $A_{r,g}$ is the adapted activity (Eq. 3). For regions close to the TSS the base activity $A_r$ is corrected with the exponential decay function, as the contact frequency would likely be the contact of the region with itself and thus $A_{r,g}$ could be erroneous. When using the regular ABC-score, the affinity scaling changes as follows:

$$af_{g,tf} = \sum_{r \in R_g} \frac{af_{r,tf}}{ml_{tf}} \cdot \begin{cases} A_r \cdot e^{-\frac{d_{r,g}}{d_0}}, & \text{if } d_{r,g} \leq 2500\ bp \\ A_r \cdot \frac{C_{r,g}}{C_{max}}, & \text{otherwise} \end{cases} . \tag{7}$$

The regions close to the TSS are scaled in the same way as before. For all other regions we divide the contact frequency $C_{r,g}$ by the maximum contact that was measured for all region-gene pairs $C_{max}$. The reasoning is to incorporate the contact frequency and to have both multipliers for $A_r$ in the range of $[0, 1]$. Essentially, the activity multipliers for $\overline{\text{ABC}}$ and ABC differ only in how the contact is scaled: for $\overline{\text{ABC}}$ relative to all gene contacts of the respective enhancer and for ABC to the contacts of all enhancer-gene pairs.
In addition to the TF affinities, we report three additional gene features: the number of regions considered per gene, the regions' average distance to the TSS, and the regions' average base pair length, as all three can be predictive for gene expression [42].

## 2.5 Application to single-cell data

To test the capability of $\overline{\text{ABC}}$-scored enhancer-gene interactions combined with TF affinities to capture regulatory cell type-specific information, we analysed a single-cell data set of the human heart from Hocker et al. (2021) [16], providing single-nuclei (sn) ATAC-seq, as well as snRNA-seq data. The candidate enhancers were pooled and ATAC-seq RPKM was measured for each cell type. For chromatin contacts we tested H3K27ac HiChIP data of the left ventricle [3], as well as an average Hi-C matrix [10]. Regions known to accumulate an anomalous amount of sequencing reads were excluded (2, 46). As we had enhancer activity and enhancer contact data at hand, we determined enhancer-gene interactions for each cell type with the $\overline{\text{ABC}}$-score (cut-off 0.02, 5 MB window around the 5' TSS), and summarised TF affinities for each gene in each cell type based on those interactions. To assess the predictability of gene expression by gene-TF affinities, we used INVOKE [38], which implements a linear regression model based on gene-TF scores, and selects TFs that are most predictive for gene expression. We trained prediction models for multiple setups for each cell type to compare their prediction accuracy. We repeated this process for cell type-specific (CS) genes, defined

as genes where the z-score on expression between cell types was $\geq 2$ and TPM $\geq 0.5$. The gene-TF matrices were limited to TFs expressed in a cell type (TPM $\geq 0.5$). Schmidt et al. (2020) [40] applied INVOKE on similar data types for bulk data, but restricted information of distant enhancers to those connected to promoters via loops. With the ABC-approach we have a finer resolution of regulatory interactions and can integrate the contact frequency into the summarisation of TF affinities.

#### 2.5.1 Comparison to co-accessibility analysis

A common approach to identify regulatory interactions in single-cell ATAC-seq data is to call co-accessible regions. Hocker et al. (2021) [16] ran Cicero [33] on their snATAC-seq data to derive pairs of regions with correlated accessibility, limited to a distance of 250 kb. Whenever either side of a co-accessible region pair overlapped a 400 bp window around the 5' TSS of a gene, we considered it as an enhancer-gene interaction for that gene. We tested how informative the resulting 97,020 co-accessible interactions are for our gene expression prediction model. Summarisation of TF affinities was done according to Eq. 5, and the affinities of regions close to the TSS ($\leq 2500$ bp) were included, as described in Sec. 2.4.

# 3 Results

## 3.1 One activity assay yields similar accuracy

The original formulation of the ABC-score requires two types of assays to measure enhancer activity, namely DNase-seq and H3K27ac ChIP-seq [10]. We tested the performance of the ABC-scoring principle on two validation data sets (10; 12), using different assays for measuring enhancer activity. On both data sets the combination of DNase-seq and H3K27ac ChIP-seq performed better than either of them alone (Fig. 1a+b left panels; Tab. 1). However, the combination achieved only a slight improvement over the usage of DNase-seq data alone. Quantifying enhancer activity with ATAC-seq resulted in a worse performance on the Fulco validation data (n=3,999), but a clear improvement on the larger Gasperini data set (n=37,738), surpassing the combination of DNase-seq and H3K27ac ChIP-seq. Using contact information averaged across multiple cell types, instead of K562-specific data, changed the performance marginally (Tab. S1).

## 3.2 Adapted $\overline{\text{ABC}}$-score improves interaction prediction

We propose an adapted $\overline{\text{ABC}}$-score (Eq. 4), where the activity of an enhancer is described in a gene-specific manner (Fig. 1c, Eq. 3). On both validation data sets and for all combinations of activity measurements the $\overline{\text{ABC}}$-score outperformed the regular ABC-score (p-value$\approx$0.0078 Wilcoxon signed-rank test) (Fig. 1a+b right panels; Tab. 1). The difference was more pronounced in the Gasperini validation data. Using an average Hi-C matrix changed the accuracy marginally for both ABC-scores (Fig. S2; Tab. S1). We could reproduce the higher accuracy of the $\overline{\text{ABC}}$-score in a direct comparison to the implementation of Fulco et al. (2019) [10] (Fig. S2; Tab. S2). The gene *CD58* is an example, where the $\overline{\text{ABC}}$-score was better for identification of enhancer-gene interactions. Out of nine candidate enhancers that were perturbed by Gasperini et al (2019) [12], one showed a significant impact on gene expression. While the $\overline{\text{ABC}}$-score correctly classified that particular interaction, the ABC-score called four false positive interactions (Fig. 1d).
Further, we examined the correlation between each of the ABC-scoring approaches and the change in gene expression as measured in the CRISPR-screens (Fig. S3a+b:i-ii). While the Spearman correlation coefficient was similar between ABC and $\overline{\text{ABC}}$ for the Fulco data (both $\rho \approx$-0.59), the $\overline{\text{ABC}}$-score yielded a higher correlation for the Gasperini data (ABC $\rho \approx$-0.47; $\overline{\text{ABC}}$ $\rho \approx$-0.52). We observed that the adapted score has a different distribution than the regular one. The mean was unchanged, as the scores sum up to 1 for all genes, but the $\overline{\text{ABC}}$-score had a higher standard deviation (ABC$\approx$0.00815; $\overline{\text{ABC}}$ $\approx$0.01709). Also, the $\overline{\text{ABC}}$-score had a broader distribution of the highest scores (Fig. S3c), enabling a finer resolved cut-off selection (Fig. S3a+b:iii). Although it uses the same information, the $\overline{\text{ABC}}$-score performed better than the regular ABC-score.

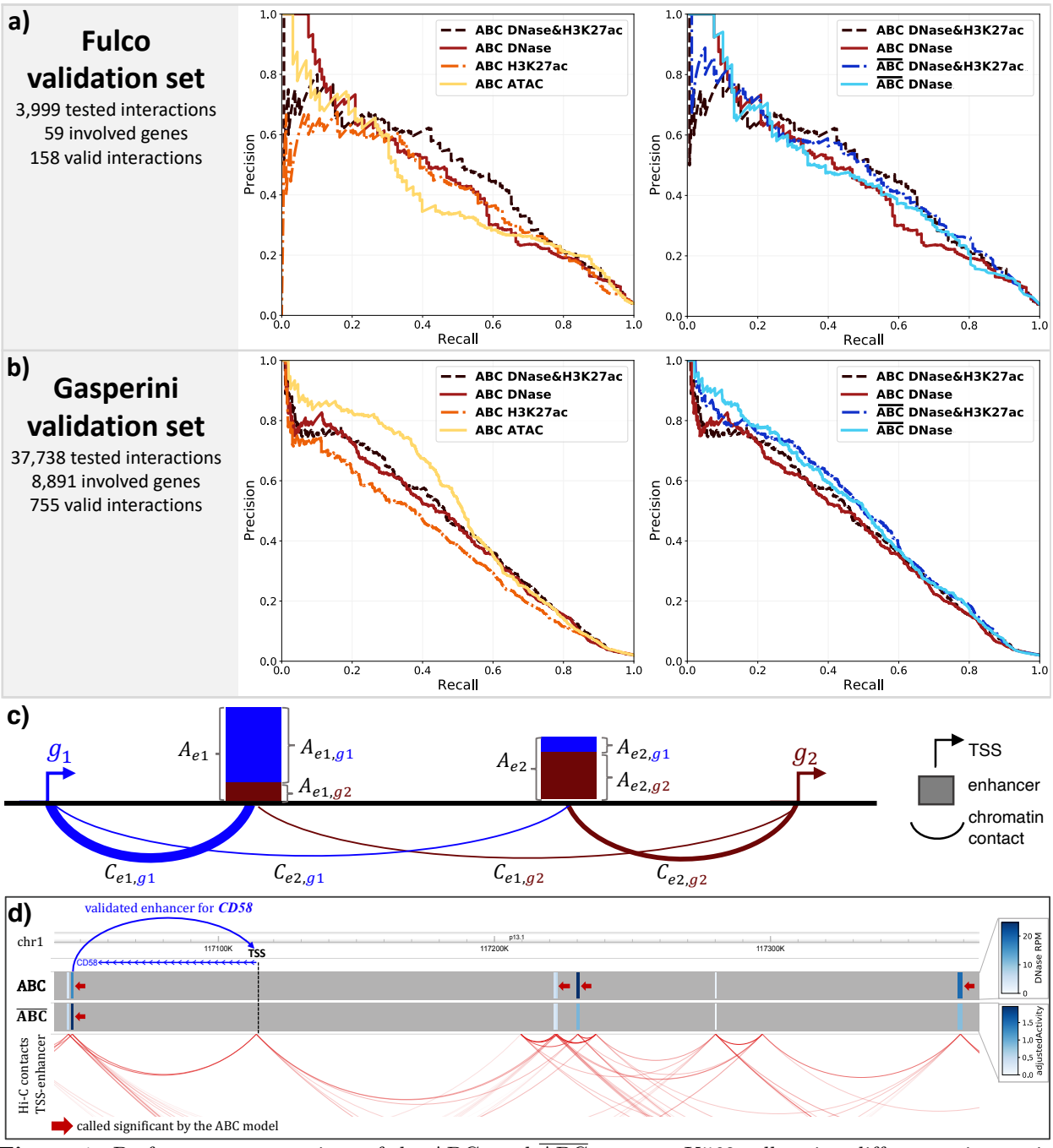


**Figure 1.** Performance comparison of the ABC- and $\overline{\text{ABC}}$-score on K562 cells using different epigenomic assays. For chromatin contacts a K562 Hi-C matrix (5kb resolution) was used [34]. **(a, b)** Precision-recall (PR) curves of both ABC-scores on experimentally validated enhancer-gene links **(a)** Fulco et al. (2019) [10]; **(b)** Gasperini et al. (2019) [12]. The respective areas under the PR curves can be found in Tab 1. **(c)** Schema for the adapted $\overline{\text{ABC}}$-score, which distributes the activity of an enhancer among its scored genes, dependent on contact frequencies. **(d)** Example gene *CD58* where the adapted $\overline{\text{ABC}}$- performs better than the ABC-score, based on interactions from Gasperini et al. (2019) [12]. The activity, or respectively the adapted activity, of the enhancers is represented by their colour. The score cut-off was set to 0.02 for both approaches. Visualisation with the WashU Epigenome Browser [24].

**Table 1.** AUPRC (area under the precision recall curve) using ABC and $\overline{\text{ABC}}$ for identifying regulatory interactions on two validation data sets, with different assays of enhancer activity. A K562 Hi-C matrix was used for contacts (more details Sec. 2.3).

| Validation data | Fulco et al. (2019) 158 valid out of 3,999 interactions | | | | Gasperini et al. (2019) 755 valid out of 37,738 interactions | | | |
|---|---|---|---|---|---|---|---|---|
| Assay(s) for enhancer activity | DNase & H3K27ac | DNase | H3K27ac | ATAC | DNase & H3K27ac | DNase | H3K27ac | ATAC |
| ABC | 0.460 | 0.445 | 0.399 | 0.405 | 0.442 | 0.433 | 0.382 | 0.492 |
| $\overline{\text{ABC}}$ | 0.466 | 0.453 | 0.440 | 0.422 | 0.479 | 0.480 | 0.444 | 0.522 |

#### 3.2.1 $\overline{\text{ABC}}$-score recovers more eQTL-gene pairs

We examined how many eQTLs found in whole blood are covered by ABC interactions in K562 cells. 150,771 eQTLs had their variant located in a candidate enhancer and thus could potentially be detected by an ABC interaction. We took a varying number of the highest scored interactions for both the ABC- and the $\overline{\text{ABC}}$-approach, and intersected them with the eQTL-gene pairs. For each selected number of top ranked interactions the $\overline{\text{ABC}}$-version captured a higher fraction of eQTL-gene pairs (Fig. S3d).

### 3.3 Regulatory interactions in single-cell data

Using just one epigenome assay for the ABC-score enables direct application on high-resolution snATAC-seq data, which has the potential to improve the prediction of cell-type specific interactions. As a proof of concept, we analysed a human heart snATAC-seq data set, comprising eight cell type clusters [16]. The candidate enhancers of the cell types were pooled to a set of 286,777 regions with a summarised ATAC-seq measurement for each defined cell type (Fig. 2a). On average we predicted 416,813 $\overline{\text{ABC}}$-scored interactions (SD $\approx$ 12,575) over all cell types. Approximately 23.9% of a cell type's interactions were shared with all other cell types. Each cell type also featured unique interactions, although this was highly variable. Atrial cardiomyocytes (aCM) and ventricular cardiomyocytes (vCM) formed the largest intersection of interactions found in only two cell types, consisting of 41,472 interactions (Fig. 2b). The average median of enhancers per expressed gene (TPM $\geq$ 0.5) across cell types was 4.75 (SD $\approx$ 0.43) (Fig. S4a). Despite all cell types having the same candidate enhancers and a shared contact measurement, their predicted enhancer-gene interactions appeared to be considerably distinct.

#### 3.3.1 Linking epigenetic features to cell type-specific expression

We trained cell type-specific gene expression prediction models based on gene-TF affinity matrices, constructed with different approaches (Fig. 3). Using the $\overline{\text{ABC}}$-score resulted in the highest Pearson correlation coefficient across all cell types ($\mu \approx$0.57), with similar results using the average Hi-C matrix ($\mu \approx$0.56). The regular ABC-score had a slightly lower performance ($\mu \approx$0.54), whereas interactions identified via co-accessibility resulted in the lowest performance ($\mu \approx$0.50). Since the co-accessible interactions were limited to a distance of 250 kb, we also tested the $\overline{\text{ABC}}$-score in a 250 kb window, which did not change the performance in the prediction model ($\mu \approx$0.57, Fig. S5a).
We defined a set of CS genes for each cell type ($\mu$=1,404 genes, $\sigma \approx$582 genes) and analysed those in more detail. The CS genes were mostly unique to a cell type (Fig. S4b), and GO term enrichment returned cell-type appropriate terms (Fig. S4c). To further characterise the sets of CS genes, we examined additional attributes and compared CS genes to non-CS genes (Fig. 3b; Fig. S4d). CS genes had more assigned enhancers than non-CS genes in all cell types except for LC. This matches the finding of Fulco et al. (2019) [10], who described a higher number of enhancers for tissue-specific than for ubiquitously expressed genes. Furthermore, CS genes tended to have a higher percentage of unique interactions, meaning the fraction of interactions of a gene, that were exclusively found in that cell type, was higher. Most cell types had a slightly higher average activity in enhancers linked to CS genes than in enhancers linked to non-CS genes, except for aCM and FB. There was no clear trend visible for the average contact frequency or TSS distance of the assigned enhancers.

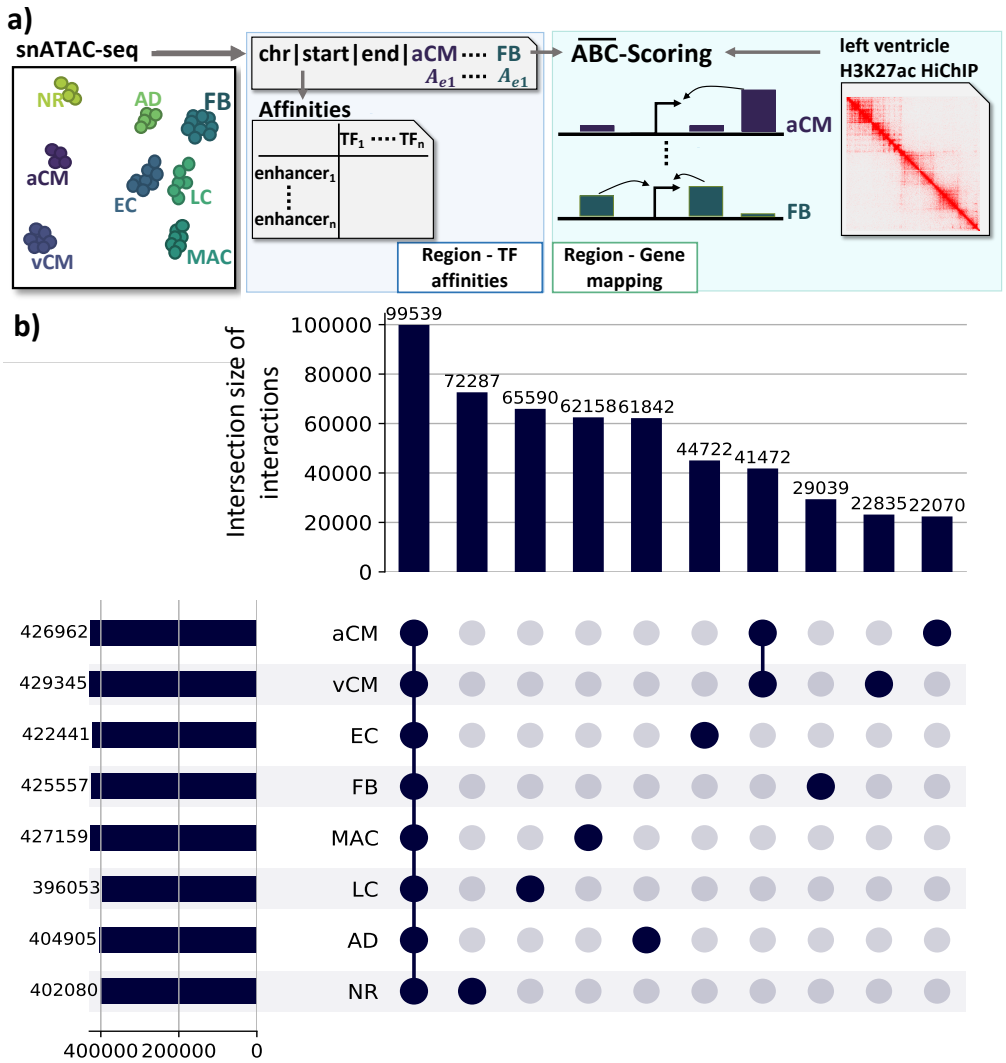


**Figure 2.** Analysis of single-cell heart data in combination with TF affinity annotation. **(a)** Schema of data processing. snATAC-seq from Hocker et al. (2021) [16] was used to identify candidate enhancers and their activity in the annotated cell types. For each cell type $\overline{\text{ABC}}$-interactions were called. Enhancer-gene contacts were retrieved from left ventricle H3K27ac HiChIP data [3]. **(b)** Upset plot of the enhancer-gene interactions called in each cell type. Only the 10 largest intersections are shown. **aCM**: atrial cardiomyocytes; **vCM**: ventricular cardiomyocytes; **EC**: endothelial cells; **FB**: fibroblasts; **MAC**: macrophages; **LC**: lymphocytes; **AD**: adipocytes; **NR**: nervous cells.

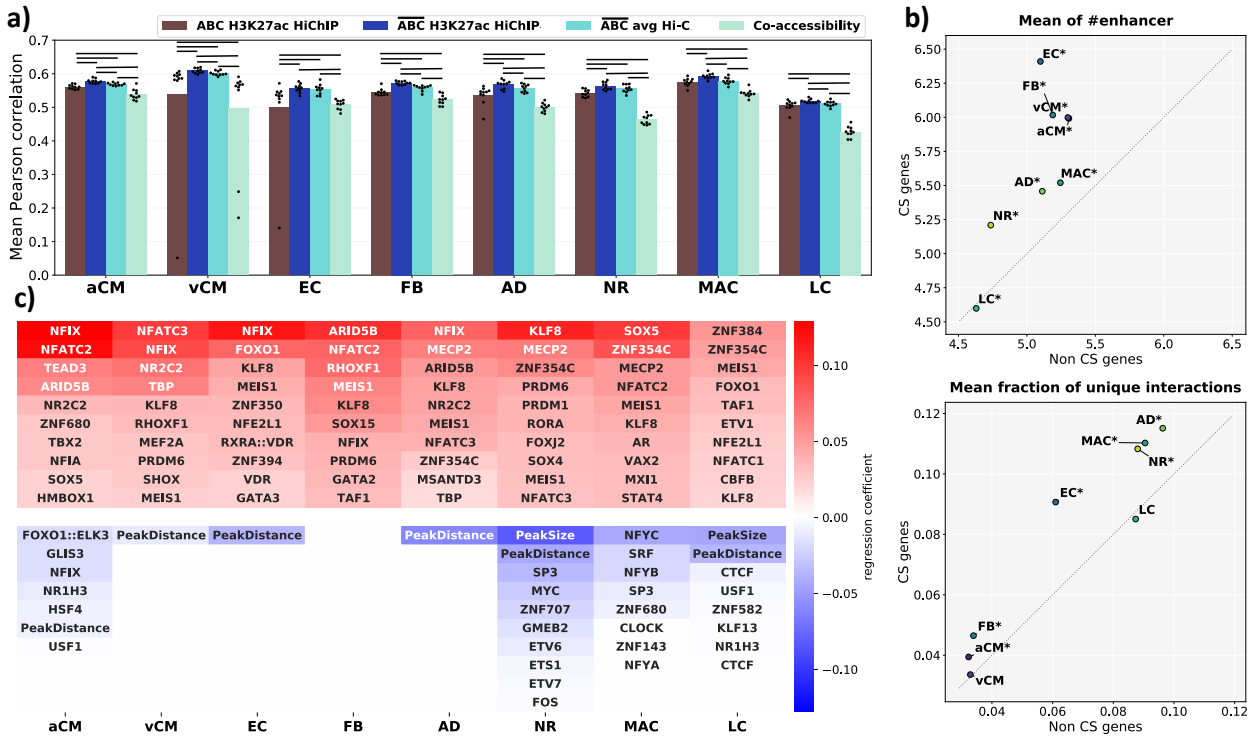


**Figure 3.** Gene expression prediction on single-cell data and characterisation of CS genes. **(a)** Accuracy of a gene expression prediction model based on different gene-TF affinity matrices. The model was trained on all genes with available expression values. The Pearson correlation coefficient is shown as average over a 10-fold outer cross-validation. A horizontal line above the bars indicate significance (p-value $\leq 0.05$, Mann-Whitney U test). snATAC-seq data was used for all approaches. ABC/$\overline{\text{ABC}}$ H3K27ac HiChIP: regular/adapted ABC scoring with H3K27ac HiChIP as contact data; $\overline{\text{ABC}}$ avg Hi-C: adapted $\overline{\text{ABC}}$ with an average Hi-C matrix as contact data; Co-accessibility: enhancer-gene links defined by Cicero [33], see Sec. 2.5.1. The respective Spearman correlation coefficients, as well as additional approaches and their performance in a training on CS genes only, are presented in Fig. S5. **(b)** Comparison of attributes between CS and not cell type-specific (non-CS) genes (*: p-value $\leq 0.05$, Mann-Whitney U test). **(c)** TFs with the highest absolute regression coefficients in an expression prediction model on CS genes only. Motif variant suffixes are not shown, different TF motifs can have positive and negative correlations in the same cell. **PeakSize**: average base pair length of the regions assigned to a gene; **PeakDistance**: average distance of regions to the TSS. See Fig. 2 for cell type abbreviations.

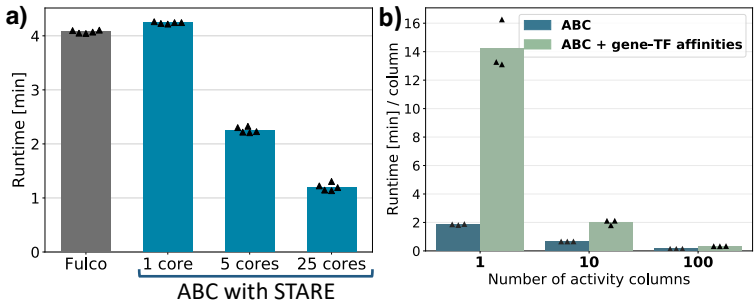


**Figure 4.** Runtime of the STARE pipeline. **(a)** Runtime of the original ABC pipeline by (Fulco et al. (2019)) [10] and our ABC-scoring implementation with a different number of cores. Any writing of output files was omitted. Calculations were done for K562 cells, scoring 155,679 candidate enhancers for 24,734 genes in a 10 MB window. The bars show the mean of five runs. **(b)** Runtime divided by the number of activity columns for the ABC-scoring alone and when additionally calculating the gene-TF affinity matrix. Writing and compression of output files is included. Calculations were done on single-cell heart data [16] with 286,777 candidate enhancers, for 57,278 genes in a 5 MB window. 818 TFs were assessed. The bars' height represent the mean of three runs.

We repeated the training of gene expression models but limited it to CS genes and observed an overall decrease in prediction accuracy (Fig. S5b). Using the $\overline{\text{ABC}}$-score with H3K27ac HiChIP data performed best ($\mu \approx 0.48$). Replacing HiChIP with average Hi-C data led to a small decrease in performance ($\mu \approx 0.46$). The prediction model returns a regression coefficient for each TF, indicative of TF relevance (Fig. 3c). We found many TFs with known roles, such as NFIX which regulates murine skeletal muscle regeneration through Myostatin expression (27; 37). NFATC2 and NFATC3 were strongly associated TFs in cardiomyocytes and NFAT-calcineurin signalling plays a role in regulating hypertrophic growth response in cardiomyocytes (29; 50). The average size of regions associated to a gene and their average distance to the TSS had negative associations with gene expression in some cell types. Overall, we were able to characterise cell type-specific expression regulation and could suggest plausible TFs.

### 3.4 Runtime of STARE

We optimised STARE's runtime by allowing multiple steps to be run in parallel (Fig. 4a) and to omit redundant calculations, if multiple cell entities with differing activity representations are processed. The runtime per activity column decreases when multiple columns are handled in the same run, allowing large data sets to be processed in a few minutes (Fig. 4b).

## 4 Discussion

We present a variation of identifying regulatory enhancer-gene interactions with the ABC-model. In its original study the ABC-model already showed a better accuracy for detecting validated enhancer-gene links than other approaches [10]. We were able to demonstrate that, unlike suggested in the original work, one assay for measuring enhancer activity works similarly well, specifically using DNase- or ATAC-seq without H3K27ac ChIP-seq data. Further, using averaged contact data yielded a high performance as well, which broadens the applicability of the ABC-score to all data sets, where a measurement of enhancer activity is available. Especially for single-cell epigenomics it is challenging to measure multiple modalities in the same cell.

We propose an adapted $\overline{\text{ABC}}$-score where the activity of an enhancer is described in a gene-specific manner by weighting it relative to the number of enhancer-gene contacts. This adaptation resulted in an improvement in identifying experimentally validated enhancer-gene interactions and more accurate expression prediction. The accuracy of different scoring approaches was partially inconsistent between the validation data sets, especially when using ATAC-seq for enhancer activity. Although both data sets are based on a dCas9-KRAB

system, there are differences in the experimental setups and scope. Fulco et al. (2019) [10] relied on RNA-FISH for measuring RNA expression, testing interactions for 30 genes, and combined that with previous results totalling interactions for 65 genes. Gasperini et al. (2019) [12] introduced multiple guide RNAs and measured expression via single-cell RNA-seq, allowing quantification of interactions for 9,810 genes. Details in candidate enhancer selection, filtering steps, and processing might lead to biases or varying sensitivity in the identification of enhancer-gene pairs. Further, it is debatable whether CRISPR-screens are able to detect all regulatory interactions, as true enhancers with small effect sizes may be overlooked. Moreover, perturbations of individual enhancers are presumably not capable of accounting for shadow enhancers with redundant functionality (43; 44). More large-scale validation data could consolidate and contextualise our findings, but are currently unavailable. Importantly, any model to annotate enhancer-gene interactions is only a prediction and likely not capturing the whole regulatory complexity of genes. The ABC-model requires two data types, which makes it applicable in a range of scenarios, but it might also miss out relevant epigenetic information. Further, the model assumes all genes are regulated in the same way.

We present the STARE framework to derive gene-TF affinities. After mapping candidate enhancers to genes, using either the ABC-score or a window-based approach, STARE summarises TF affinities on a gene level. Unlike other methods aiming to determine regulatory relations of TFs to genes (23; 49), STARE does not require scarce TF ChIP-seq data. It uses a motif-based biophysical model [36] to determine TF affinities in accessible regions. As consequence, unlike other methods [26], it is able to retain low affinity binding information. Patel et al. (2021) [32] use similar data as STARE and rely on a graph-based approach, but they do not incorporate active regulatory regions and analysis is limited to a window marked by the most distant CTCF peaks within 50 kb of the gene body. With the ABC-model, STARE specifically determines candidate enhancers within any selected window size. Notably, our model assumes an additive influence of TFs, which is not likely to accurately capture biological reality [53]. Furthermore, there is potential redundancy on two levels: the aforementioned functional redundancy of enhancers and the redundancy of TF binding motifs (13; 8).

We applied our framework to single-cell data with clustered cell types of the human heart [16], where the activity of a uniform set of candidate enhancers was measured for each cell type. The assumption was that cell type-specificity is mainly driven by activity of regulatory regions and less by spatial chromatin conformation. Although chromatin contacts were also found to change upon cell differentiation (9; 54), the 3D conformation of the genome is described as less dynamic and more as a scaffold to enable and stabilise regulatory interactions (43; 17). We were able to unravel differences in regulatory interactions across cell types and to characterise regulation of cell type-specific genes, despite using bulk chromatin contact data.

We demonstrated a downstream application example of STARE with a linear expression model, which allows to identify candidate regulatory TFs for further evaluation. Considering the small training sets and that the model assumes the same regulation for all genes, the prediction yielded a reasonable performance with meaningful prioritisation of TFs. Schmidt et al. (2020) [40] used a similar expression prediction approach and incorporated distant enhancers. However, their work relies on annotated loops, which are not likely to cover all relevant regulatory interactions. In addition, their strategy is not able to integrate contact frequencies into TF affinities, nor to derive cell type-specific interactions from bulk contact data. There are other tools for predicting gene expression explicitly in individual cells that incorporate TF information, such as SCENIC [1], ACTION [28] or TRIANGULATE [4], but none of them considers long-range enhancer-gene interactions. A compelling approach would be to combine these expression prediction tools with information on distant enhancers.

STARE represents a currently unique form of deriving TF affinities on a gene level: it combines enhancer-gene links called by the ABC-score with a non hit-based TF annotation. Prospectively, we would like to apply STARE on individual cells instead of clustered cell types, which would require additional steps to account for the drastic sparsity of most single-cell measurements. It would also be highly interesting to further investigate the importance of chromatin contacts in cell type-specificity, once the resolution and availability for such data are advanced.

## Availability and genome version

The presented results are provided via Zenodo (`https://doi.org/10.5281/zenodo.5841991`). All data is in hg19. For details on the used data see Supplementary Material.


## Acknowledgements

We thank Nina Baumgarten for advice and discussions on the implementation. We acknowledge the ENCODE Consortium and the lab of Michael Snyder in Stanford who generated the ATAC-seq data sets.

## Funding

This work has been supported by the DZHK (German Centre for Cardiovascular Research, 81Z0200101) and the Cardio-Pulmonary Institute (CPI) [EXC 2026] ID: 390649896, and the DFG SFB (TRR 267) Noncoding RNAs in the cardiovascular system, Project-ID 403584255.


## References


1. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, and S. Aerts. SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11):1083–1086, Nov. 2017.

2. H. M. Amemiya, A. Kundaje, and A. P. Boyle. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports*, 9(1):9354, Dec. 2019.

3. C. G. Anene-Nzelu, W. L. W. Tan, C. J. M. Lee, Z. Wenhao, A. Perrin, A. Dashi, Z. Tiang, M. I. Autio, B. Lim, E. Wong, H. S. Tan, B. Pan, M. P. Morley, K. B. Margulies, T. P. Cappola, and R. S.-Y. Foo. Assigning Distal Genomic Enhancers to Cardiac Disease–Causing Genes. *Circulation*, 142(9):910–912, Sept. 2020.

4. F. Behjati Ardakani, K. Kattler, T. Heinen, F. Schmidt, D. Feuerborn, G. Gasparoni, K. Lepikhov, P. Nell, J. Hengstler, J. Walter, and M. H. Schulz. Prediction of single-cell gene expression for transcription factor analysis. *GigaScience*, 9(11):1–14, Oct. 2020.

5. J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, 109(1), Jan. 2015.

6. J. A. Castro-Mondragon, R. Riudavets-Puig, I. Rauluseviciute, R. Berhanu Lemma, L. Turchi, R. Blanc-Mathieu, J. Lucas, P. Boddie, A. Khan, N. Manosalva Pérez, O. Fornes, T. Leung, A. Aguirre, F. Hammal, D. Schmelter, D. Baranasic, B. Ballester, A. Sandelin, B. Lenhard, K. Vandepoele, W. W. Wasserman, F. Parcy, and A. Mathelier. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 50:D165–D173, 2022.

7. M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *SYSTEMS BIOLOGY*, 107(50):21931–21936, 2010.

8. D. A. Cusanovich, B. Pavlovic, J. K. Pritchard, and Y. Gilad. The Functional Consequences of Variation in Transcription Factor Binding. *PLOS Genetics*, 10(3):13, 2014.

9. J. Fraser, M. Rousseau, S. Shenker, M. A. Ferraiuolo, Y. Hayashizaki, M. Blanchette, and J. Dostie. Chromatin conformation signatures of cellular differentiation. *Genome Biology*, 10(4):R37, 2009.

10. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R. Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M. Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander, and J. M. Engreitz. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics*, 51(12):1664–1669, Dec. 2019.

11. T. Gao and J. Qian. EAGLE: An algorithm that utilizes a small number of genomic features to predict tissue/cell type-specific enhancer-gene interactions. *PLOS Computational Biology*, 15(10):e1007436, Oct. 2019.

12. M. Gasperini, A. J. Hill, J. L. McFaline-Figueroa, B. Martin, S. Kim, M. D. Zhang, D. Jackson, A. Leith, J. Schreiber, W. S. Noble, C. Trapnell, N. Ahituv, and J. Shendure. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell*, 176(1-2):377–390, Jan. 2019.

13. A. Gitter, Z. Siegfried, M. Klutstein, O. Fornes, B. Oliva, I. Simon, and Z. Bar-Joseph. Backup in gene regulatory networks explains differences between binding and knockout results. *Molecular Systems Biology*, 5(1):276, Jan. 2009.

14. D. H. Gonzalez. Introduction to Transcription Factor Structure and Function. In *Plant Transcription Factors*, pages 3–11. Elsevier, 2016.

15. N. D. Heintzman, G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243):108–112, May 2009.

16. J. D. Hocker, O. B. Poirion, F. Zhu, J. Buchanan, K. Zhang, J. Chiou, T.-M. Wang, Q. Zhang, X. Hou, Y. E. Li, Y. Zhang, E. N. Farah, A. Wang, A. D. McCulloch, K. J. Gaulton, B. Ren, N. C. Chi, and S. Preissl. Cardiac cell type–specific gene regulatory programs and disease risk association. *Science Advances*, 7(20):eabf1444, May 2021.

17. E. Ing-Simmons, R. Vaid, X. Y. Bing, M. Levine, M. Mannervik, and J. M. Vaquerizas. Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *Nature Genetics*, 53(4):487–499, Apr. 2021.

18. T. K. Kelly, Y. Liu, F. D. Lay, G. Liang, B. P. Berman, and P. A. Jones. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Research*, 22(12):2497–2506, Dec. 2012.

19. P. Kheradpour and M. Kellis. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Research*, 42(5):2976–2987, Mar. 2014.

20. J. F. Kribelbauer, C. Rastogi, H. J. Bussemaker, and R. S. Mann. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annual Review of Cell and Developmental Biology*, 35(1):357–379, 2019. PMID: 31283382.

21. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, and V. J. Makeev. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-seq analysis. *Nucleic Acids Research*, 46:D252–D259, 2018.

22. S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch. The Human Transcription Factors. *Cell*, 172(4):650–665, Feb. 2018.

23. X. Lan, H. Witt, K. Katsumura, Z. Ye, Q. Wang, E. H. Bresnick, P. J. Farnham, and V. X. Jin. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Research*, 40(16):7690–7704, Sept. 2012.

24. D. Li, S. Hsu, D. Purushotham, R. L. Sears, and T. Wang. WashU Epigenome Browser update 2019. *Nucleic Acids Research*, 47(W1):W158–W165, July 2019.

25. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, E. S. Lander, and J. Dekker. Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, Oct. 2009.

26. R. C. McLeay, T. Lesluyes, G. Cuellar Partida, and T. L. Bailey. Genome-wide in silico prediction of gene expression. *Bioinformatics*, 28(21):2789–2796, Nov. 2012.

27. G. Messina, S. Biressi, S. Monteverde, A. Magli, M. Cassano, L. Perani, E. Roncaglia, E. Tagliafico, L. Starnes, C. E. Campbell, M. Grossi, D. J. Goldhamer, R. M. Gronostajski, and G. Cossu. Nfix Regulates Fetal-Specific Transcription in Developing Skeletal Muscle. *Cell*, 140(4):554–566, 2010.

28. S. Mohammadi, V. Ravindra, D. F. Gleich, and A. Grama. A geometric approach to characterize the functional identity of single cells. *Nature Communications*, 9(1):1516, Dec. 2018.

29. J. D. Molkentin, J.-R. Lu, C. L. Antos, B. Markham, J. Richardson, J. Robbins, S. R. Grant, and E. N. Olson. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell*, 93(2):215–228, 1998.

30. E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, and B. G. Bruneau. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, 169(5):930–944, May 2017.

31. C. O. Pabo and R. T. Sauer. Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu. Rev. Biochem*, 61:1053–95, 1992.

32. N. Patel and W. S. Bush. Modeling transcriptional regulation using gene regulatory networks based on multi-omics data sources. *BMC Bioinformatics*, 22(1):200, Dec. 2021.

33. H. A. Pliner, J. S. Packer, J. L. McFaline-Figueroa, D. A. Cusanovich, R. M. Daza, D. Aghamirzaie, S. Srivatsan, X. Qiu, D. Jackson, A. Minkina, A. C. Adey, F. J. Steemers, J. Shendure, and C. Trapnell. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular Cell*, 71(5):858–871.e8, 2018.

34. S. Rao, M. Huntley, N. Durand, E. Stamenova, I. Bochkov, J. Robinson, A. Sanborn, I. Machol, A. Omer, E. Lander, and E. Aiden. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680, Dec. 2014.

35. S. S. Rao, S.-C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K.-R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, and E. L. Aiden. Cohesin Loss Eliminates All Loop Domains. *Cell*, 171(2):305–320, Oct. 2017.

36. H. G. Roider, A. Kanhere, T. Manke, and M. Vingron. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics*, 23(2):134–141, Jan. 2007.

37. G. Rossi, C. Bonfanti, S. Antonini, M. Bastoni, S. Monteverde, A. Innocenzi, M. Saclier, V. Taglietti, and G. Messina. Silencing Nfix rescues muscular dystrophy by delaying muscle regeneration. *Nature Communications*, 8(1):1055, 2017.

38. F. Schmidt, N. Gasparoni, G. Gasparoni, K. Gianmoena, C. Cadenas, J. K. Polansky, P. Ebert, K. Nordström, M. Barann, A. Sinha, S. Fröhler, J. Xiong, A. Dehghani Amirabad, F. Behjati Ardakani, B. Hutter, G. Zipprich, B. Felder, J. Eils, B. Brors, W. Chen, J. G. Hengstler, A. Hamann, T. Lengauer, P. Rosenstiel, J. Walter, and M. H. Schulz. Combining transcription factor binding affinities with open-chromatin data for accurate gene expression prediction. *Nucleic Acids Research*, 45(1):54–66, Nov. 2016.

39. F. Schmidt, F. Kern, P. Ebert, N. Baumgarten, and M. H. Schulz. TEPIC 2—an extended framework for transcription factor binding prediction and integrative epigenomic analysis. *Bioinformatics*, 35(9):1608–1609, May 2019.

40. F. Schmidt, F. Kern, and M. H. Schulz. Integrative prediction of gene expression with chromatin accessibility and conformation data. *Epigenetics & Chromatin*, 13(1):4, Dec. 2020.

41. F. Schmidt, A. Marx, N. Baumgarten, M. Hebel, M. Wegner, M. Kaulich, M. Leisegang, R. Brandes, J. Göke, J. Vreeken, and M. Schulz. Integrative analysis of epigenetics data identifies gene-specific regulatory elements. *Nucleic Acids Research*, 49(18):10397–10418, Oct. 2021.

42. F. Schmidt and M. H. Schulz. On the problem of confounders in modeling gene expression. *Bioinformatics*, 35(4):711–719, 2019.

43. S. Schoenfelder and P. Fraser. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics*, 20(8):437–455, Aug. 2019.

44. D. Singh and S. V. Yi. Enhancer Pleiotropy, Gene Expression, and the Architecture of Human Enhancer–Gene Interactions. *Molecular Biology and Evolution*, 38(9):3898–3909, Aug. 2021.

45. L. Song and G. E. Crawford. DNase-seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells. *Cold Spring Harbor Protocols*, 2010(2):pdb.prot5384, Feb. 2010.

46. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, Sept. 2012.

47. The FANTOM Consortium, R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A. R. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–461, Mar. 2014.

48. The GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, 369:1318–1330, 2020.

49. S. Wang, H. Sun, J. Ma, C. Zang, C. Wang, J. Wang, Q. Tang, C. A. Meyer, Y. Zhang, and X. S. Liu. Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nature Protocols*, 8(12):2502–2515, Dec. 2013.

50. B. J. Wilkins and J. D. Molkentin. Calcineurin and cardiac hypertrophy: Where have we been? where are we going? *The Journal of Physiology*, 541(1):1–8, 2002.

51. L. Yao, B. P. Berman, and P. J. Farnham. Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. *Critical Reviews in Biochemistry and Molecular Biology*, 50(6):550–573, Nov. 2015.

52. X. Yi, Z. Zheng, H. Xu, Y. Zhou, D. Huang, J. Wang, X. Feng, K. Zhao, X. Fan, S. Zhang, X. Dong, Z. Wang, Y. Shen, H. Cheng, L. Shi, and M. J. Li. Interrogating cell type-specific cooperation of transcriptional regulators in 3D chromatin. *iScience*, 24(12):103468, Dec. 2021.

53. J. Zeitlinger. Seven myths of how transcription factors read the cis-regulatory code. *Current Opinion in Systems Biology*, 23:22–31, Oct. 2020.

54. C. Zhang, Z. Xu, S. Yang, G. Sun, L. Jia, Z. Zheng, Q. Gu, W. Tao, T. Cheng, C. Li, and H. Cheng. tagHi-C Reveals 3D Chromatin Architecture Dynamics during Mouse Hematopoiesis. *Cell Reports*, 32(13):108206, Sept. 2020.