



Research article

GNet: An integrated context-aware neural framework for transcription factor binding signal at single nucleotide resolution prediction

Jujuan Zhuang, Kexin Feng*, Xinyang Teng and Cangzhi Jia*

School of Science, Dalian Maritime University, Dalian, Liaoning 116026, China

* **Correspondence:** Email: kexin1004_0919@163.com, cangzhijia@dlmu.edu.cn.

Abstract: Transcription factors (TFs) are important factors that regulate gene expression. Revealing the mechanism affecting the binding specificity of TFs is the key to understanding gene regulation. Most of the previous studies focus on TF-DNA binding sites at the sequence level, and they seldom utilize the contextual features of DNA sequences. In this paper, we develop an integrated spatiotemporal context-aware neural network framework, named GNet, for predicting TF-DNA binding signal at single nucleotide resolution by achieving three tasks: single nucleotide resolution signal prediction, identification of binding regions at the sequence level, and TF-DNA binding motif prediction. GNet extracts implicit spatial contextual information with a gated highway neural mechanism, which captures large context multi-level patterns using linear shortcut connections, and the idea of it permeates the encoder and decoder parts of GNet. The improved dual external attention mechanism, which learns implicit relationships both within and among samples, and improves the performance of the model. Experimental results on 53 human TF ChIP-seq datasets and 6 chromatin accessibility ATAC-seq datasets shows that GNet outperforms the state-of-the-art methods in the three tasks, and the results of cross-species studies on 15 human and 18 mouse TF datasets of the corresponding TF families indicate that GNet also shows the best performance in cross-species prediction over the competitive methods.

Keywords: transcription factor binding site; gated highway neural network; encoder-decoder architecture; external attention mechanism

1. Introduction

Transcription factors (TFs) are important molecules that control gene expression. They either activate or inhibit gene transcription by binding to specific DNA sequences [1,2]. Revealing how TFs recognize and bind to specific DNA sequences is crucial for understanding and further studying the function of cis-regulatory elements in regulatory genomics research. The DNA fragments that bind to specific TFs and are evolutionarily conserved are called transcription factor binding sites (TFBSs) [3]. Identifying TFBSs and their corresponding motifs are fundamental in the field of regulatory genomics [4]. With the rapid development of high-throughput sequencing technology, various approaches for quantifying TFBSs have emerged. For example, a protein-bound microarray (PBM) [5] allows for the study of DNA-protein interaction on the genome scale in a high-throughput manner. Systematic evolution of ligands by exponential enrichment coupled with massively parallel sequencing (SELEX-seq) [6] has become a preferred method to study the in vitro binding between TFs and DNA through large-scale mining of TF information. Chromatin immunoprecipitation sequencing (ChIP-seq) [7] provides the opportunity to study DNA-TF interactions on a genome-wide scale. An assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) [8] describes putative accessible regions of the genome that normally work with TFs and RNA polymerase. These methods provide us with a large amount of available data for studying TFBSs-associated tasks.

In recent years, multiple computational methods have been introduced for the identification of TFBSs and their motifs [9–12]. For example, Kmer SVM [9] and gkm-SVM [10] are support vector machine (SVM) models with K-mer and gapped K-mer characteristics of the DNA sequence to predict TFBSs. Multiple expectation maximizations for motif elicitation (MEME) [11] and simple, thorough, rapid, enriched motif elicitation (STREME) [12] apply the expectation maximization (EM) algorithm and suffix trees to store data, and discover TF-DNA binding motifs by searching for repeated and ungapped fragments in sequences, respectively. Deep neural network (DNN) models [13] have rapidly become the mainstream algorithms in many fields due to its high efficiency and applicability, which has also been introduced to impressively predict TFBSs [14]. DeepBind uses deep learning to analyze how proteins bind to DNA and RNA [15]. DeepSEA is based on deep learning and was developed to predict sequence noncoding compilation effects and can also be used to predict TFBSs [16]. DanQ combines convolutional and bidirectional long short-term memory networks to build a de novo construction for predicting non-coding functions from sequences [17]. DeepGRN combines single attention modules and paired attention modules to predict TFBSs [18]. TFregulomeR is a computational tool that can reveal the context-specific characteristics and functions of transcription factors [19]. With one-hot encoding, BPNet [20] predicts base resolution ChIP-nexus binding profiles of pluripotency TFs by introducing expanded convolutional neural networks (CNN), and FCNsignal [21] constructs a fully convolutional neural network (FCN) model that can simultaneously predict base resolution signals, distinguish binding or non-binding regions, and predict corresponding TF-binding motifs. Although these DNN-based approaches have shown remarkable performances in the tasks of TFBSs prediction, location, and motifs recognition from genomic sequences, they seldom utilize the contextual features of DNA sequences and are limited to one-hot encoding feature representations and a onefold network.

The depth of the network is a major obstacle to training DNN models [22,23]. Deep highway neural network models [24] can effectively reduce the training difficulties caused by network depth with gated linear units [25], which greatly alleviates the vanishing gradient problems by applying a

linear path for the gradients. The gated mechanism has been widely used in the study of biological sequences [26,27]. ACNet [28] utilizes a gated residual network to extract spatial information for a polyadenylation (Poly(A)) signal prediction based on co-occurrence embedding. PASNet [29] applies the gated convolution network to automatically extract the underlying patterns and identify polyadenylation signals (PAS) from the genome sequences. Deep motif (DeMo) [30] is a deep convolution/highway multilayer perceptron (MLP) framework that can classify TFBSs tasks in genomic sequences. Meanwhile, in recent years, attention mechanisms, which can extract more global spatial long-distance dependence and solve the problem of information overload by calculating the importance weight vector and allocating computing resources to more important tasks, have been introduced into many deep learning tasks such as computer vision and natural language processing [26,27]. Various types of attention mechanisms have emerged in different fields [31–32]; for example, self-attention [31] maps the input to three features-Q, K, and V-to consider the correlation between the different parts of one sample and obtain the weighted features. However, self-attention ignores potential correlations among different samples and has quadratic complexity. External attention [33] considers the correlations between all samples and generates a weighted output by referring to two cascaded linear layers as memory units, which has linear complexity. In order to derive the joint implicit information at different positions of nucleotide sequences, we improve the external attention mechanism and named this as a dual external attention (DEA) to obtain the relationships between segments within a sample and between samples.

In our work, we develop an integrated context-aware neural framework, GNet, based on a gated highway network to consider TF binding signals prediction at single nucleotide resolution, to determine the binding or non-binding regions at a sequence level, and understand motif recognition problems. GNet extracts spatial and temporal patterns based on attention and gated highway networks to obtain the co-occurrence embedding of the signals. Specifically, with one-hot coding, nucleotide chemical properties and nucleotide density as input features, while infiltrating the gated idea into the whole network and coupling with the improved DEA mechanism, GNet positively performs in three tasks, namely signal regression, sequence classification and motif recognition on ChIP-seq data, and chromatin accessibility ATAC-seq data, which outperforms several state-of-the-art methods. In addition, we conduct a cross-species studies on 15 human and 18 mouse TF data, and GNet also shows the best performance over the competition method.

2. Materials and methods

2.1. Data collection and preprocessing

ChIP-seq is a powerful tool for studying protein-DNA interactions in the body owing to its rapid and efficient genome-wide detection of DNA regions that interact with specific TFs. In this work, we collect 53 human TF ChIP-seq datasets from the ENCODE project [34] (<https://www.encodeproject.org/>), including 21 from the GM12878 cell line, 20 from the K562 cell line, and 12 from the HeLa-S3 cell line, as shown in Supplementary Table S1. To ensure data quality, we preferentially select data with biological replication. With the same operation as in [21], the peak of each sequence, as the center, is extended to a length of 1000 bp, and the obtained sequence and the corresponding signal (p-value) are regarded as “positive” samples. Sequences of the same length at 3000 bp upstream of the peak and the corresponding signals are selected as “negative” samples. We ensure

that the signal value of the “negative” samples is smaller than that of the corresponding “positive” samples. Hg38 is used as the reference genome for selection. The signal values are normalized by $\log_{10}(1 + \text{signal})$ to reduce the influence of large difference. We randomly select 20% of the positive and negative samples for each TF as test data, 10% of the remaining samples as validation data, and the rest as training data.

We also collect 6 chromatin accessibility ATAC-seq datasets from ENCODE, which are from A549, GM12878, HepG2, IMR90, K562, and MCF7 cell lines. For each dataset, we preprocess the above cell lines as ChIP-seq data. We download 18 mouse TF ChIP-seq datasets from ENCODE, which are from the same TF family as human datasets with mm10 as the reference genome. Supplementary Table S1 describes the details of these datasets.

In addition, we randomly select in vitro PBM data of 8 mouse TFs from the Dialogue for Reverse Engineering Assessments and Methods 5 (DREAM5) [35] project to further verify the effect of our improved DEA mechanism. The 8 datasets are from two distinct microarray designs, named HK and ME (the abbreviation of the designers’ name). Each of these datasets contain more than 40,000 35 bp probe sequences with corresponding PBM probe intensities. We normalize the data according to the total signal strength.

2.2. Feature representation

In our framework, each DNA sequence is represented as a matrix of dimension $L \times 8$, where L is the length of the DNA sequence and 8 ($4 + 3 + 1$) corresponds to the sum of the dimensions of one-hot encoding, the 3D coordinates coding of nucleotide chemical features, and the nucleotide density coding of each nucleotide. For DNA sequence $N = \{N_1, N_2, \dots, N_L\}$, the three encodings are as follows.

2.2.1. One-hot encoding

DNA sequence N is encoded as an $L \times 4$ -matrix in one-hot form, where 4 is the dimension of the binary one-hot vector ($A = [1, 0, 0, 0]$, $C = [0, 1, 0, 0]$, $G = [0, 0, 1, 0]$, and $T = [0, 0, 0, 1]$) corresponding to nucleotide $\{A, C, G, T\}$ at that position.

2.2.2. Nucleotide chemical property (RFHC)

Nucleotides have different chemical properties according to different ring structures, hydrogen bonds, and chemical functions, etc. [36]. For example, A and G contain two rings, while C and T contain one ring. G and C can form strong hydrogen bonds, while A and T form weak hydrogen bonds. A and C are classified as amino groups, while G and T are classified as ketones. We use a 3-dimension binary vector (x_i, y_i, z_i) to represent the chemistry of each nucleotide N_i , where $x_i = 1$ means that N_i contains two rings and $x_i = 0$ shows that N_i contains one rings; similarly, y_i, z_i represents the characteristic of hydrogen bonds and chemical function of N_i , respectively. The specific formulas are as follows:

$$x_i = \begin{cases} 1, & \text{if } N_i \in \{A, G\} \\ 0, & \text{other nucleotides} \end{cases} \quad y_i = \begin{cases} 1, & \text{if } N_i \in \{A, T\} \\ 0, & \text{other nucleotides} \end{cases} \quad z_i = \begin{cases} 1, & \text{if } N_i \in \{A, C\} \\ 0, & \text{other nucleotides} \end{cases} \quad (1)$$

Therefore, each sequence is transformed into an $L \times 3$ -matrix according to the 3D binary encoding ($A = [1, 1, 1]$, $C = [0, 0, 1]$, $G = [1, 0, 0]$, and $T = [0, 1, 0]$).

2.2.3. Nucleotide density (ND)

ND takes into account the position and frequency information of nucleotides on the sequence, which is also an effective coding method [37,38]. The density D_i of nucleotide N_i at position i in sequence N is defined as follows:

$$D_i = \frac{1}{\|L_i\|} \sum_{j=1}^i f(N_j), \quad (2)$$

$$f(N_j) = \begin{cases} 1, & \text{if } N_j = N_i, \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where $\|L_i\|$ is the length of the sequence of the final nucleotide N_i and i stands for the position of the nucleotide N_i . Therefore, each sequence can be encoded as an $L \times 1$ matrix.

2.3. Network architecture

GNet is a DNN model with an encode-decode architecture, with details as follows.

2.3.1. The gated highway network unit

As the depth of the neural network increases, the gradient back propagation of network training becomes more and more difficult [29], and the gated mechanism is an effective method to improve the efficiency of network training [24,39]. Gated highway neural models can control the information flow by capturing large context multi-level patterns using the linear shortcut connections with neither extra parameters nor computational cost, and thus can greatly alleviate the vanishing gradient problems since a linear path is provided for the gradients.

In this work, we infiltrate the idea of a gated mechanism into each component of GNet. Specifically, we introduce two gates, T and C , to control the information flow of the first input and the second input, respectively. For simplicity, we set $C = 1 - T$. The specific formula of the gated network unit is as follows:

$$\begin{aligned} H^\zeta &= C(x^\zeta, W_C^\zeta) \otimes X_1^\zeta \oplus T(x^\zeta, W_T^\zeta) \otimes X_2^\zeta \\ &= (1 - T(x^\zeta, W_T^\zeta)) \otimes X_{first}^\zeta \oplus T(x^\zeta, W_T^\zeta) \otimes X_{second}^\zeta, \\ &= (1 - f_S(x^\zeta)) \otimes X_{first}^\zeta \oplus f_S(x^\zeta) \otimes X_{second}^\zeta \end{aligned} \quad (4)$$

where X_{first}^ζ and X_{second}^ζ are two different inputs of the ζ -th gated network unit, $\zeta \in \{1, 2, 3, 4\}$,

\otimes is the matrix multiplication, \oplus is the element-wise addition operation, x^ζ is the input that controls the ζ -th gate, and $f_S(\cdot)$ is the sigmoid function for dynamically calculating the gate channel value. In particular,

$$H^\zeta = \begin{cases} X_{first}^\zeta, & \text{if } f(x^\zeta) = 0 \\ X_{second}^\zeta, & \text{if } f(x^\zeta) = 1 \end{cases} \quad (5)$$

In the subsequent encoding architecture, we design several gated units with different additional effects in convolution layers and gated recurrent unit (GRU) layers, see the encoder architecture section for details.

2.3.2. Dual-external-attention mechanism

An attention mechanism is one of the common modules used to derive the global dependency pattern features in natural language processing, which can allocate resources according to the importance weight. An external-attention mechanism [33] was developed to solve the two major pain points of self-attention, that is, high computational complexity and the absence of relationships between samples [31]. It has linear complexity and an implicit consideration for the relationship between different samples. However, it ignores the relationship among the component elements in one sample. DEA, proposed in this paper, is an improved external-attention in which the relationship among component elements is considered by multiplying the input and its transpose, as shown in Figure 1. The specific formula of DEA is as follows:

$$f_{attention}(X) = Line(Normal(Softmax(X \otimes X^T)) \otimes X), \quad (6)$$

where X is the input, $Normal()$ represents the normalization process, and \otimes is the matrix multiplication. DEA can combine the advantages of self-attention and external attention to learn implicit relationships within a sample and among samples, and the computational cost is lower than that of self-attention.

2.3.3. The encoder architecture

DNN can extract high-quality contextual information from DNA sequences without feature engineering, a variety of TFBSs-associated prediction models based on DNN have been proposed for extracting discerning patterns of DNA sequences automatically [17–19]. The feature extraction architecture of GNet consists of three blocks, each containing a convolution layer with a gated highway network unit, a maximum pooling layer, and a dropout layer [40]. CNN can capture the local spatial pattern features of the sequence, and the gated highway network increases the flexibility of contextual information extraction and alleviates the problem of gradient backflow obstruction. The gated convolution highway network unit in the first block is denoted as the formula:

$$\begin{aligned} H^1 &= (1 - f_S(x^1)) \otimes X_{first}^1 \oplus f_S(x^1) \otimes X_{second}^1 \\ &= (1 - f_S(\Delta(ReLU(W^1 * X^1 + b^1)))) \otimes \Delta(ReLU(W^1 * X^1 + b^1)) , \\ &\quad \oplus f_S(\Delta(ReLU(W^1 * X^1 + b^1))) \otimes \nabla(ReLU(W^1 * X^1 + b^1)) \end{aligned} \quad (7)$$

where W^1 , X^1 , and b^1 are the weight matrix, input, and bias vector of the 1-st convolution, respectively, and $\Delta()$ and $\nabla()$ represent the first 1/2 and last 1/2 components of the convolutional layer output matrix, respectively. Here we take the first 1/2 components of the output matrix of the convolution layer as the input x^1 to generate the first gate.

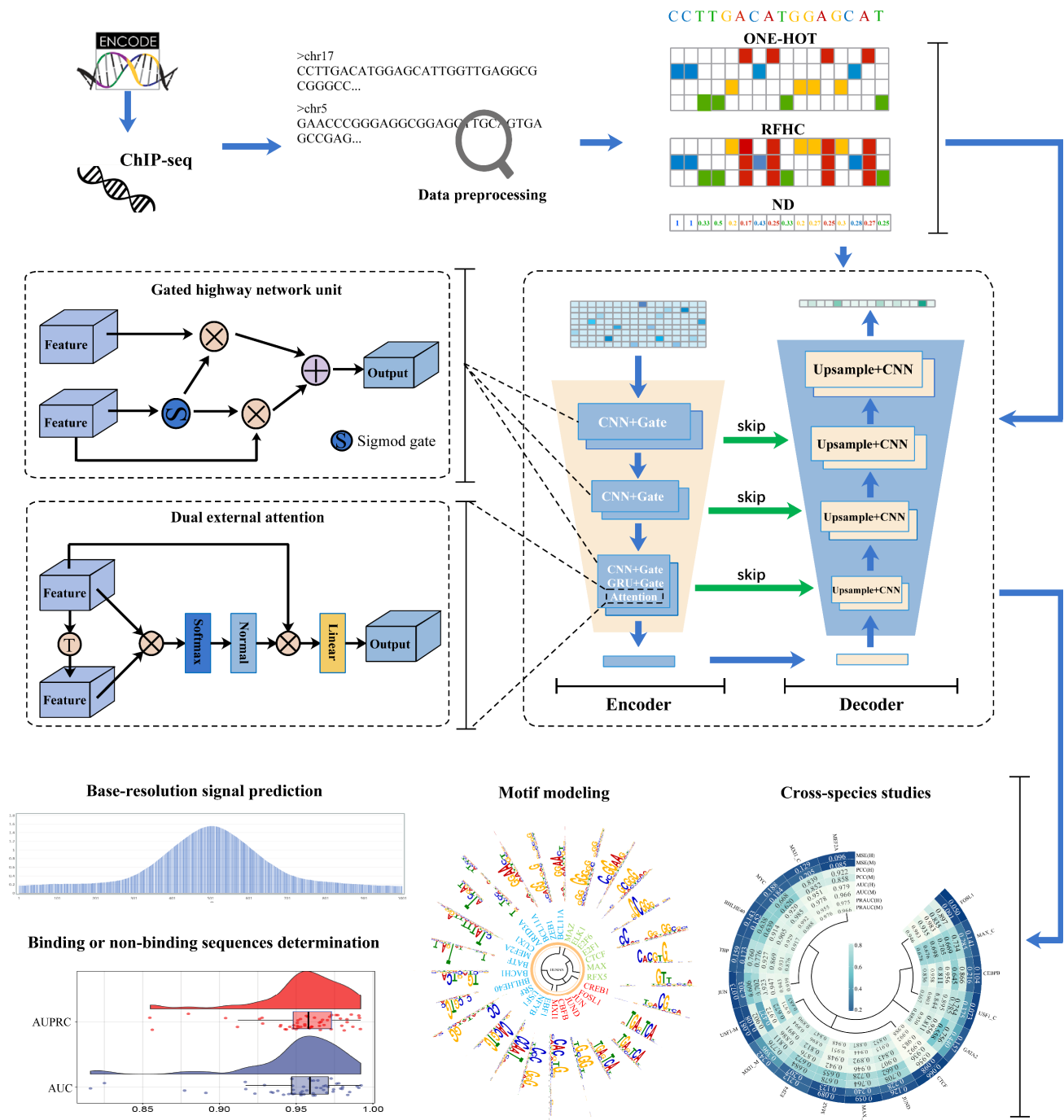


Figure 1. The framework of GNet. The three coding matrices of DNA sequence are spliced and fed into GNet with encoder-decoder-skip architecture, in which, the encoder part integrates the gated highway network to capture large context multi-level patterns and DEA mechanism to learn implicit relationships inter-sample and among samples. The decoder layer is connected with the corresponding encoder layer with a skip identity connection to make full use of the context information. GNet is applied for TF binding signal prediction under single nucleotide resolution, determination of binding or non-binding sequences, motif recognition and cross-species research between human and mouse.

The convolution gated network units in the second and the third blocks are as follows:

$$\begin{aligned} H^\tau &= (1 - f_S(x^\tau)) \otimes X_{first}^\tau \oplus f_S(x^\tau) \otimes X_{second}^\tau \\ &= (1 - f_S(\text{Relu}(W^\tau * X^\tau + b^\tau))) \otimes X^\tau \\ &\quad \oplus f_S(\text{Relu}(W^\tau * X^\tau + b^\tau)) \otimes \text{Relu}(W^\tau * X^\tau + b^\tau) \end{aligned} \quad (8)$$

where W^τ and b^τ are the weight matrix and bias vector of the τ -th block of the encoder part, respectively, $\tau = 2$ and 3 , and X^τ is the output of the $\tau - 1$ block. Here we take the output matrix of the τ -th convolution layer as the input x^τ to generate the τ -th gate. In this way, the information of the previous block can be fully considered to increase the multilinearity of the calculation and capture more context information.

In the third block, after the dropout layer, we add a bidirectional GRU [41] with a gated highway network unit:

$$\begin{aligned} H^4 &= (1 - f_S(x^4)) \otimes X_{first}^4 \oplus f_S(x^4) \otimes X_{second}^4 \\ &= (1 - f_S(\overrightarrow{\text{GRU}}(W^4 * X^4 + b^4))) \otimes \overrightarrow{\text{GRU}}(W^4 * X^4 + b^4) \\ &\quad \oplus f_S(\overleftarrow{\text{GRU}}(W^4 * X^4 + b^4)) \otimes \overleftarrow{\text{GRU}}(W^4 * X^4 + b^4) \end{aligned} \quad (9)$$

where $\overrightarrow{\text{GRU}}()$ and $\overleftarrow{\text{GRU}}()$ are the GRU forward and backward operations, respectively. The output matrix of the forward operation of the GRU is used as the input to generate the fourth gate. A GRU with a gated mechanism can better extract the long-term dependence of nucleotides. The output of the GRU is fed into the DEA layer, which allows the model to focus on elements that contribute more to the prediction by assigning different weights to each position. Finally, the global context information of the sequence is captured by the global average pooling layer.

2.3.4. The decoder architecture with gated mechanism

The decoder architecture that contains four blocks is the process of up-sampling, where each block contains an upper sampling layer with a linear interpolation algorithm, a batch normalization (BN) layer, and a CNN layer. In each of the three blocks, there is a skip identity connection with the corresponding encoder layer to make full use of the context information. We apply up-sampling to recover the down-sampled eigenmatrix, and then integrate the positional information in the encoded process with the gated identity connection. The last layer of convolution is used to convert the features into predictable signals of dimension $1 \times l$. The process for this part is as follows:

$$X_{out}^\zeta = \text{linear}(X_{in}^\zeta), \quad (10)$$

$$X_{out}^\zeta = X_{out}^\zeta + X^\zeta, \quad (11)$$

$$X_{out}^\zeta = \text{Relu}(\text{BN}(X_{out}^\zeta)) * W^\zeta + b^\zeta, \quad (12)$$

where $linear()$ represents a linear interpolation operation, X_{in}^ζ is the input of the network, and X^ζ is the position information derived from the corresponding encoded part.

2.4. The implementation of the model

GNet has been executed on Tesla P40 using python 3.6, cuda10.0 with pytorch1.1 backend. GNet is an integrated framework, in which the regression task at a single nucleotide resolution is fundamental, so the mean square error (MSE) at a single nucleotide resolution is used as the loss function:

$$loss = \frac{1}{N \times L} \sum_{i=1}^N \sum_{j=1}^L (y_{ij} - \hat{y}_{ij})^2 + \lambda \|\theta\|_2, \quad (13)$$

where N and L are the number of samples and the length of the sequence, respectively, and \hat{y}_{ij} and y_{ij} are the predicted signal value and the real value, respectively. In order to avoid overfitting, L_2 regularization ($\|\theta\|_2$) is adopted, where λ represents the regularization parameter. In our work, we apply an Adam optimizer [42] to train our network, with the minimum batch size 500 and a learning rate decay with a decay rate 0.9 for every 10 epochs. The learning rate, β value of Adam, and regularization parameters are randomly selected from $\{0.001, 0.0001\}$, $\{0.9, 0.99, 0.999\}$, and $\{0, 0.001\}$, respectively, and model with the best performance on the validation data set is saved for testing. The source code for GNet is available at <https://github.com/keke0419/GNet-main>.

2.5. Evaluation indicators

As an integrated framework, GNet is applied to handle three different tasks, so we use different indicators for evaluation. For signal modeling tasks at a single nucleotide resolution, the mean square error (MSE) and Pearson correlation coefficient (PCC) are chosen as evaluation indexes. For the determination task of binding or non-binding regions, we use the area under the receiver operating characteristic curve (AUC) and the area under the precision recall rate curve (AUPRC) as evaluation indicators. Additionally, for the motif recognition and prediction task, $-\log_2(p\text{-value})$, $-\log_2(E\text{-value})$, and $-\log_2(q\text{-value})$ of the testing are used as evaluation indicators.

3. Experimental results

We tested GNet on three tasks with 53 TF ChIP-seq datasets and 6 chromatin accessibility ATAC-seq datasets, as well as cross-species studies between human and mouse TFS on 15 human and 18 mouse TF ChIP-seq datasets.

3.1. Predicting base resolution signals on ChIP-seq data

We conducted a regression analysis on 53 human TF ChIP-seq datasets from ENCODE to predict the base resolution signal and compared GNet with two of the most advanced competitive approaches, BpNet [20] and FCNsignal [21], on the test sets of 53 TF ChIP-seq data with MSE and PCC as evaluation metrics. BpNet [20] takes a one-hot encoding of the sequence as an input, and constructs a dilated CNN model to predict the base resolution ChIP-nexus binding profiles of pluripotency TFs.

FCNsignal [21] constructs a CNN-dominated encoding and decoding architecture with a one-hot encoding of sequences as an input to predict TF-DNA binding signals at base resolution level. Figure 2 and Supplementary Table S2 shows the performance comparison of the three algorithms.

GNet performed remarkable on these 53 TF ChIP-seq datasets, with the lowest average MSE of 0.121, compared with a value of 0.133 for BPNet and a value of 0.127 for FCNsignal, and the largest average PCC of 0.798, in contrast to a value of 0.768 for BPNet and a value of 0.780 for FCNsignal. This indicates that, compared with the two models that use only one-hot encoding as features and are dominated by CNN, infiltrating the idea of a gated mechanism into the whole network, and considering the physicochemical and the density properties of nucleotide is effective for identifying TFBSs.

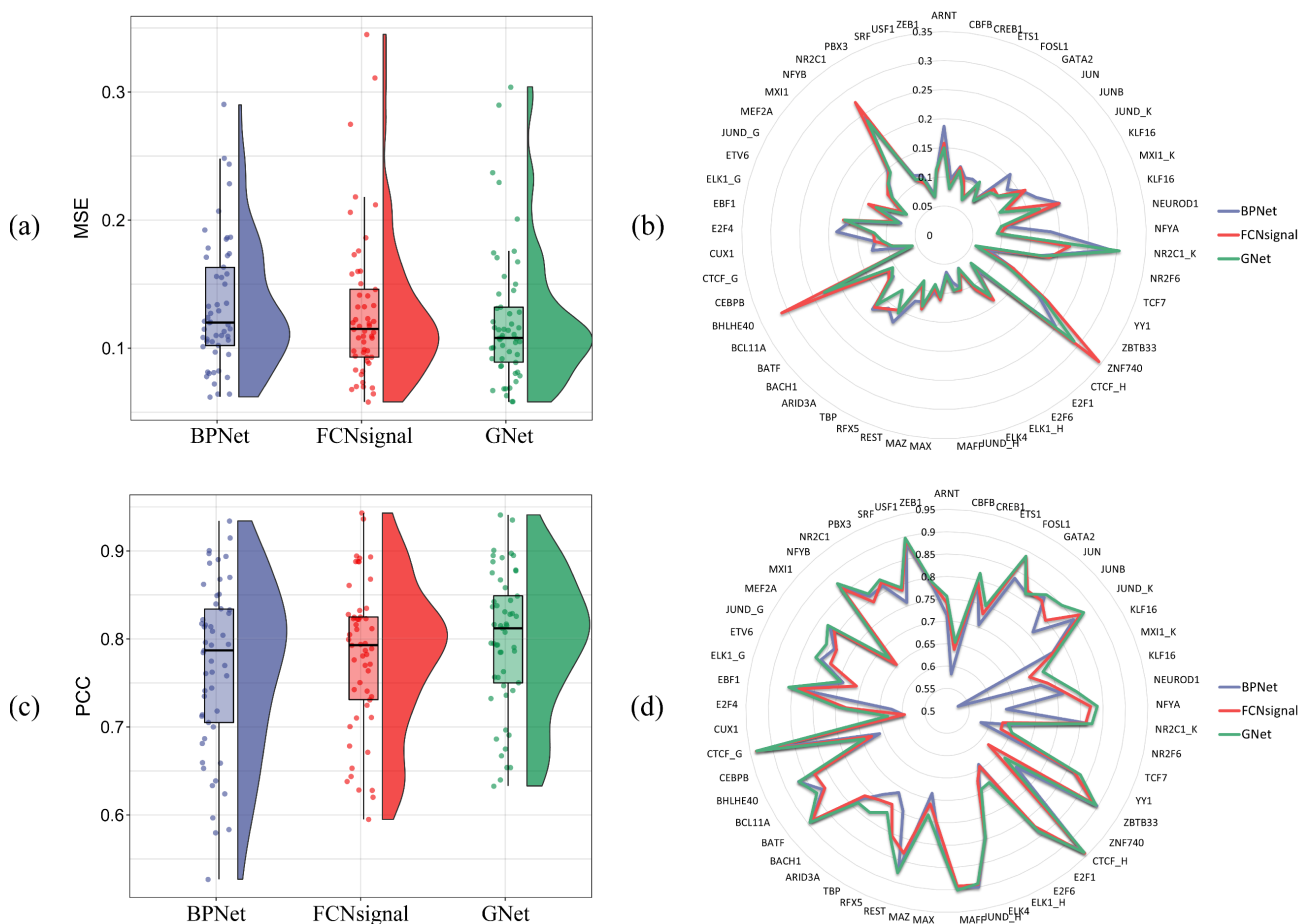


Figure 2. The performance of GNet in TF binding signal prediction at single nucleotide level on 53 ChIP-seq datasets compared with BPNet and FCNsignal. (a),(b) show the MSE values of the three methods, the smaller the value, the better the model. (c),(d) display the PCC values of the three methods, the higher value is better.

In addition, to demonstrate the performance of our model in predicting TF signals at a single nucleotide resolution, we visualized TF signal values predicted by GNet and the two competitive methods and compared them with real signal values. Figure 3(a) shows a comparison of two randomly chosen TFs. GNet thoroughly fits the TF signal values, and the fitting degree is better than the other competitive methods.

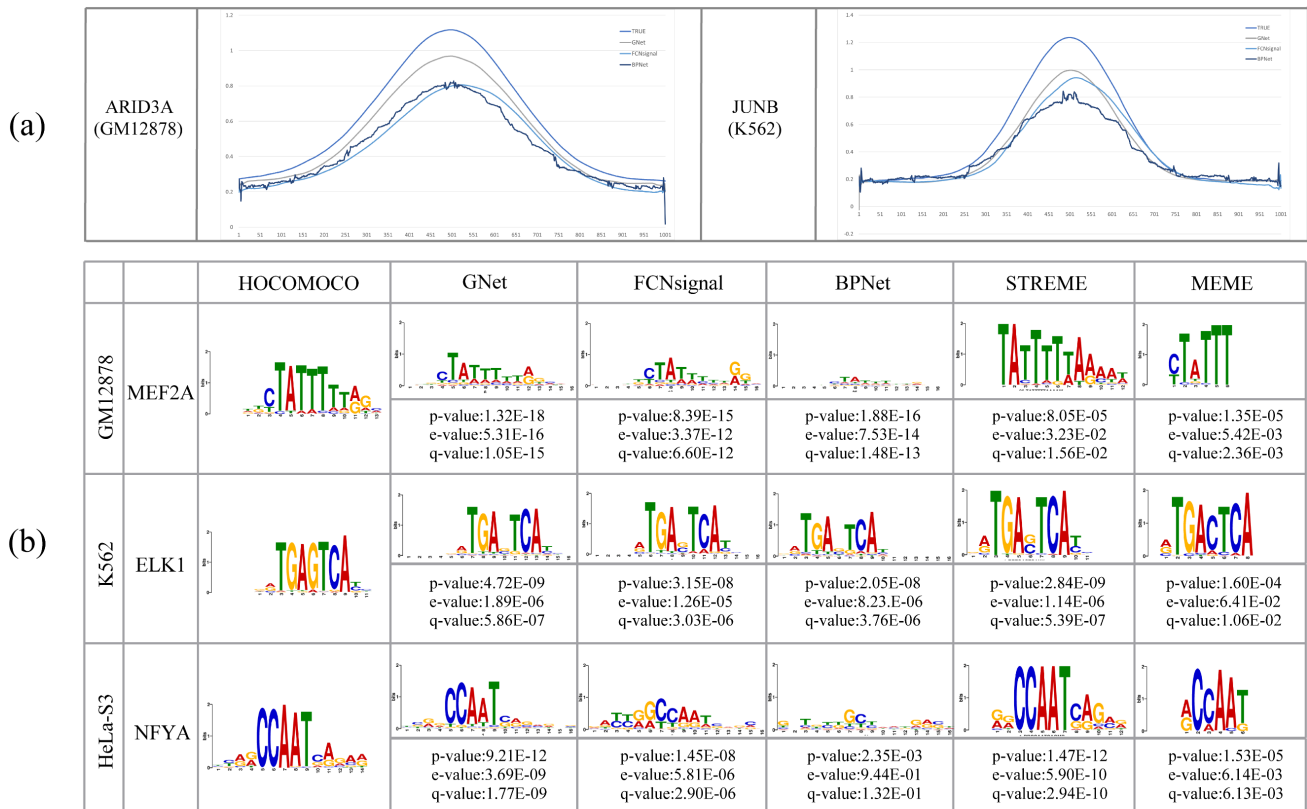


Figure 3. Visualization of model effects and predicted motifs. (a) Comparison of TF signals predicted by models at single nucleotide resolution. We compare GNet with two competitive methods and the real experimental signals on randomly selected 2 ChIP-seq datasets. (b) Comparison of motifs predicted by different models, and compared with the experimental verification on randomly selected 3 ChIP-seq datasets.

3.2. Determining the binding or non-binding sequence on ChIP-seq data

The problem here is how to use the predicted signal values at a base resolution to determine either binding or non-binding sequences. In [21], the authors confirmed that the maximum value of signals could reflect the openness of TF-DNA binding; the higher the signal value, the higher the degree of openness. Based on the same strategy, we distinguished binding or non-binding sequences with GNet.

We compared our model with the two competitive methods mentioned in Section 3.1 with AUC and AUPRC for evaluation. As shown in Figure 4 and Supplementary Table S2, the average AUC and AUPRC of GNet on 53 datasets are greater than the two competitive methods with an AUC of 0.951 and an AUPRC of 0.954. In particular, the AUC values of GNet are higher than FCNsignal and BPNet on 46 and 48 datasets, respectively, and the AUPRC values are higher than FCNsignal and BPNet on 44 and 50 datasets, respectively.

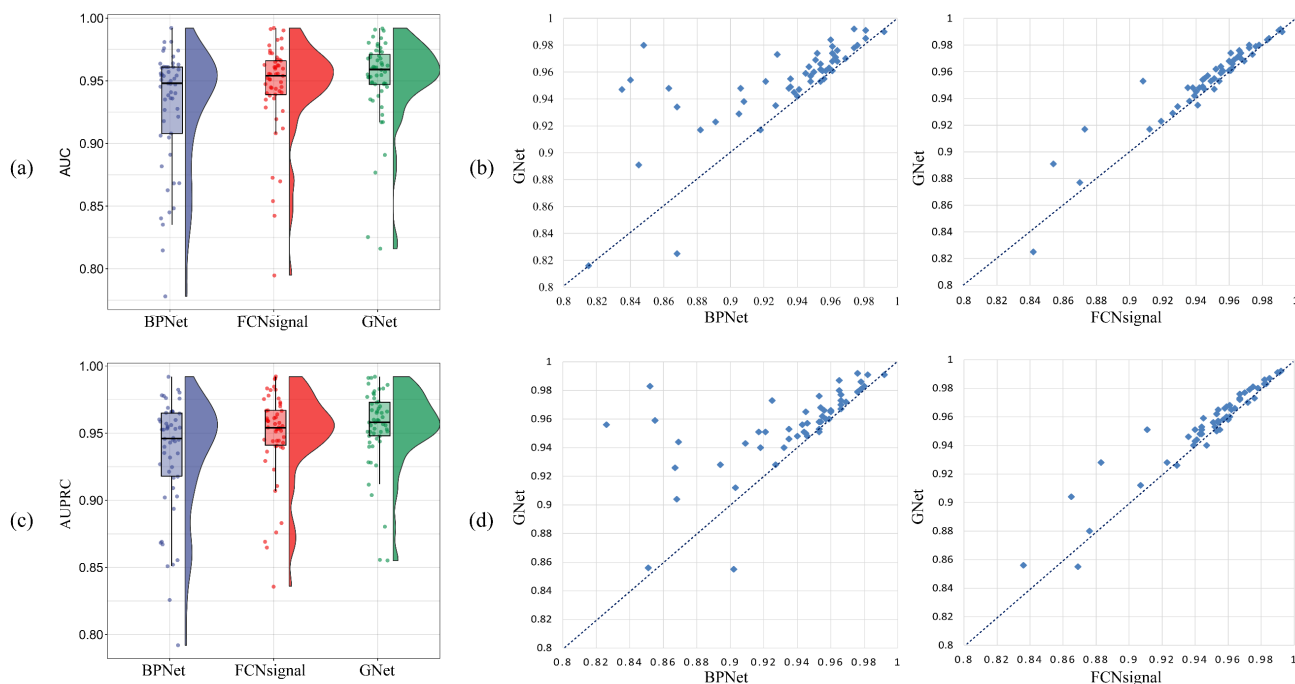


Figure 4. The performance of GNet in binding or non-binding sequence determination on 53 ChIP-seq datasets compared with BPNet and FCNsignal. (a),(b) present the AUC values of the three algorithms. (c),(d) show the AUPRC values of the three methods.

3.3. Predicting motifs on ChIP-seq data

We adopt the same motif identification strategy as [21,43]; the position with the maximum predicted signal value in the sequence is locked and extended 49 bp forward and 50 bp backward to get a 100 bp segment as a possible binding region. We use the training weight of the first convolution layer in the coding architecture to the possible binding region with a sliding window to get value and select the sequence segment of the region with the maximum value; these sequence segments are aligned to calculate the corresponding position frequency matrix (PFM) to match with experimentally verified motifs in the standard database HOCOMOCO [44].

In addition to the two competitive methods mentioned above, we also select two motif recognition tools, MEME [11] and STREME [12], as the benchmark methods, where MEME uses the EM algorithm to discover a new, unencapsulated motif by searching for recurring, ungapped sequence fragments in the sequence, and STREME stores data with suffix trees, and uses it to efficiently count matches to a PWM of a candidate motif, to identify rich or relatively rich ungapped motifs of fixed length in the sequence. We use the negative logarithm with base 2 of p - value, E - value, and q - value generated in TOMTOM [45] as our evaluation indicators. The results are shown in Figure 5(a) and Table S3.

In 53 ChIP-seq datasets, GNet outperforms the four competitive methods with the lowest average p - value, E - value, and q - value, so that the average values of the three indicators are all higher than the collective. Moreover, GNet found the largest number of motifs, specifically 9, 5, 7, and 4 more than the competitive methods MEME, STREME, BPNet, and FCNsignal, respectively. Figure 3(b) shows partial TF motifs predicted with the five methods. The motifs predicted by GNet are closer to

the experimentally verified motifs than that predicted by other methods. Figures 1 and S1 shows the visualization of part of predicted motifs.

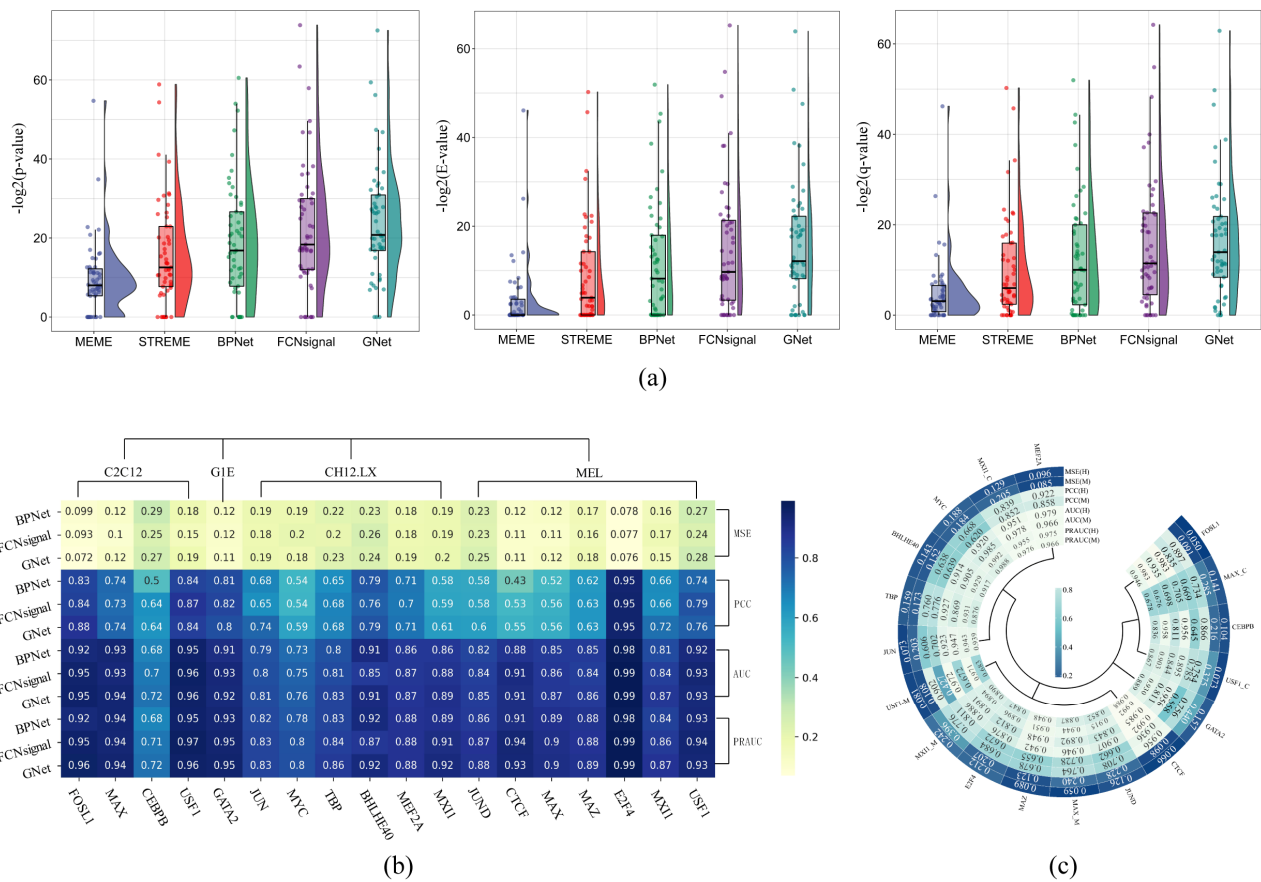


Figure 5. The performance of GNet in motif identification on 53 ChIP-seq datasets compared with BpNet, FCNsignal, MEME and STREME, and cross-species studying between human and mouse on corresponding TFs compared with BpNet and FCNsignal. (a) shows the results of hypothesis testings in motif identification, the larger the values of the three indexes are, the closer the predicted motif is to the standard motif. (b) Heat map displays the results of GNet trained on human ChIP-seq data and tested on corresponding TF mouse data compared with the two competitive methods. (c) Results of GNet trained on mouse ChIP-seq data and tested on corresponding TF mouse and human test datasets respectively.

3.4. The performance of GNet on ATAC-seq data

ATAC-seq data are commonly used to test chromatin accessibility and to determine gene expression regulation mechanisms. In this study, to verify the overall performance of GNet, we downloaded 6 ATAC-seq datasets from ENCONA as independent evaluation data, on which we tested our model in signal prediction at a single nucleotide resolution and the binding or non-binding sequences prediction were compared with the two competitive methods FCNsignal and BpNet. The evaluation indexes are MSE, PCC, AUC, and AUPRC. The comparison results are shown in Table 1.

Table 1. Comparison of GNet, BpNet and FCNsignal in regression and classification tasks on 6 ATAC-seq datasets.

Model	indicator	K562	MCF7	IMR90	HepG2	GM12878	A549	Mean
GNet	MSE	0.221	0.281	0.254	0.217	0.318	0.495	0.298
	PCC	0.804	0.818	0.816	0.805	0.814	0.757	0.802
	AUC	0.944	0.951	0.957	0.951	0.956	0.937	0.949
	AUPRC	0.949	0.954	0.958	0.952	0.958	0.944	0.953
FCNsignal	MSE	0.216	0.278	0.217	0.206	0.318	0.479	0.286
	PCC	0.787	0.818	0.804	0.784	0.797	0.736	0.788
	AUC	0.943	0.946	0.943	0.945	0.954	0.933	0.944
	AUPRC	0.948	0.949	0.952	0.947	0.958	0.941	0.949
BpNet	MSE	0.215	0.240	0.244	0.246	0.296	0.412	0.276
	PCC	0.767	0.818	0.771	0.763	0.796	0.730	0.774
	AUC	0.925	0.950	0.929	0.916	0.942	0.941	0.934
	AUPRC	0.931	0.954	0.931	0.919	0.946	0.949	0.938

Note: Bold numbers are the best results.

The average PCC, AUC, and AUPRC of GNet are all higher than those of the competitive methods; in particular, GNet achieves the best PCC, AUC, and AUPRC values on the 5 datasets, except for A456t. This shows that GNet has a good robustness, even on ATAC-seq data.

3.5. The performance on cross-species studies

It was demonstrated that the genes of humans and mice are more than 85 percent similar, and about 80 percent of the proteins encoded by genes are homologous. In recent years, many human and mouse genome databases have been developed, among which HOCOMOCO [44] is specially designed for the study of human and mouse TFs. We downloaded 15 human TFBSs ChIP-seq datasets and 18 mouse datasets of the same corresponding TF families from ENCODE. For the same TF family, we trained GNet on human TFBSs datasets and tested on mouse datasets; the results of base-resolution signal prediction and the binding or non-binding sequence distinction tasks are shown in Figure 5(b).

Compared with the results tested on human TFBSs test data, the results of GNet tested on mouse datasets are generally good, and even for some TFs, the results are better than those tested on human datasets, such as for mouse TF CTCF, GNet obtains a PCC of 0.95 compared to the value of 0.941 for the human test set. The average PCC, AUC, and AUPRC values of GNet are higher than those of the competitive methods on these two cross-species tasks.

We also trained GNet on mouse data and tested on human data. As shown in Figure 5(c), the models trained on mouse TFBSs data performed relatively good on human data, and PCC values are even 30% greater than those tested on mouse data. This indicates that GNet performs well in a cross-species study of human and mouse TF.

4. Discussion

In this section, we conducted ablation analyses for the experiment by comparing the effects of an improved DEA and the other attention mechanisms, the effect of a network with and without the

gated mechanism, and the influence of nucleotide density and nucleotide chemical features added in this architecture.

4.1. The effects of dual external attention mechanism

We compared the performance of our model in signal prediction at a single nucleotide resolution and a binding or non-binding sequences prediction with attention-free, self-attention [31], external attention, and DEA mechanisms, respectively, on several randomly selected datasets (7 ChIP-seq datasets and 2 ATAC-seq datasets) without changing the other settings in the architecture. The results are shown in Table 2, from which we can see that GNet with a built-in DEA obtains the best performance. In addition, the models with an external attention and attention-free behave similarly, that is, an external attention mechanism which focuses on the relationship between different samples does not only contribute much to feature extraction of TFBS-associated tasks. The DEA mechanism that can obtain the implicit relationships within and among samples has a lower computation complexity and higher accuracy than self-attention. This indicates that DEA is more effective in feature extraction of TFBS-associated tasks.

Table 2. Comparison of the models with attention-free, self-attention, external attention and DEA respectively.

Model	indicator	HeLa-S3		K562		GM12878		ATAC-seq		Mean
		E2F1	ELK1	NFYA	MXI1	CEBPB	ZEB1	K562	GM12878	
DEA	MSE	0.132	0.063	0.092	0.104	0.058	0.112	0.221	0.318	0.1375
	PCC	0.838	0.691	0.841	0.732	0.696	0.791	0.804	0.814	0.7759
	AUC	0.993	0.974	0.983	0.954	0.917	0.961	0.944	0.956	0.9603
	AUPRC	0.993	0.986	0.985	0.959	0.928	0.960	0.949	0.958	0.9648
Self-attention	MSE	0.131	0.068	0.094	0.121	0.085	0.121	0.196	0.291	0.1384
	PCC	0.837	0.680	0.833	0.725	0.684	0.789	0.806	0.816	0.7713
	AUC	0.993	0.971	0.969	0.956	0.911	0.961	0.943	0.956	0.9575
	AUPRC	0.992	0.984	0.978	0.960	0.919	0.961	0.947	0.958	0.9624
External-attention	MSE	0.127	0.061	0.105	0.131	0.059	0.102	0.228	0.298	0.1389
	PCC	0.842	0.684	0.82	0.702	0.672	0.795	0.786	0.816	0.7646
	AUC	0.993	0.965	0.974	0.939	0.875	0.961	0.94	0.957	0.9505
	AUPRC	0.993	0.981	0.978	0.947	0.892	0.959	0.945	0.959	0.9568
Attention-free	MSE	0.136	0.070	0.095	0.110	0.057	0.112	0.191	0.341	0.1390
	PCC	0.824	0.676	0.836	0.709	0.698	0.780	0.805	0.81	0.7673
	AUC	0.993	0.971	0.982	0.950	0.917	0.956	0.945	0.958	0.9590
	AUPRC	0.993	0.984	0.984	0.956	0.926	0.954	0.949	0.961	0.9634

Note: Bold numbers are the best results.

To further verify the effect of our improved DEA mechanism, we randomly downloaded PBM data of 8 mice TFs from DREAM5 project and applied DeepBind [18] as the evaluation model to evaluate the performance of the DEA mechanism by predicting the probe strength; R^2 (reflecting the proportion that all variations of dependent variables can be explained by independent variables through

a regression relationship, the closer it is to 1, the better the interpretation of independent variables to dependent variables) and PCC are selected as the evaluation indexes. We trained DeepBind with self-attention, external attention, and DEA mechanisms via 5-fold cross validation, and the results are shown in Supplementary Table S4. The average effect of DeepBind combined with DEA is better than that of the model with an external attention mechanism and is as good as that of the combined mechanism with self-attention. This confirms that our improved DEA mechanism takes less time and works not worse than self-attention, at least in TFBS-associated tasks.

4.2. The effect of introducing the gated highway unit and different types of features

We compared the effects of our model with and without the gated highway unit in signal prediction at a single nucleotide resolution and the binding or non-binding sequences prediction on the same datasets as the last subsection, without changing the other settings in the architecture. As shown in Table 3, GNet with a built-in gated highway unit performs better than the model without a gated mechanism. This indicates that, without an extra computational cost, our gated highway neural unit can efficiently capture large contextual multilevel patterns.

At the same time, we also compared GNet with the model using only a one-hot encoding as the input. Table 3 shows that our model integrating a one-hot encoding, nucleotide density, and nucleotide chemical features as the input performs better. A one-hot coding is not sufficient to cover all intrinsic features of DNA sequences; density, as well as chemical properties of sequences, are also important in TFBS-associated tasks.

Table 3. The comparison results of the models with or without gated highway units, and only one-hot encoding.

Model	indicator	HeLa-S3		K562		GM12878		ATAC-seq		Mean
		E2F1	ELK1	NFYA	MXI1	CEBPB	ZEB1	K562	GM12878	
GNet	MSE	0.132	0.063	0.092	0.104	0.058	0.112	0.221	0.318	0.1375
	PCC	0.838	0.691	0.841	0.732	0.696	0.791	0.804	0.814	0.7759
	AUC	0.993	0.974	0.983	0.954	0.917	0.961	0.944	0.956	0.9603
	AUPRC	0.993	0.986	0.985	0.959	0.928	0.960	0.949	0.958	0.9648
No gated mechanism	MSE	0.136	0.066	0.094	0.108	0.075	0.115	0.217	0.292	0.1379
	PCC	0.831	0.679	0.828	0.723	0.68	0.792	0.801	0.795	0.7661
	AUC	0.991	0.964	0.971	0.955	0.897	0.96	0.943	0.95	0.9539
	AUPRC	0.991	0.981	0.978	0.959	0.909	0.959	0.947	0.953	0.9596
Only one-hot	MSE	0.131	0.063	0.088	0.108	0.063	0.110	0.238	0.405	0.1508
	PCC	0.840	0.691	0.844	0.733	0.676	0.787	0.801	0.816	0.7735
	AUC	0.992	0.960	0.976	0.955	0.914	0.956	0.945	0.958	0.9570
	AUPRC	0.991	0.980	0.980	0.958	0.922	0.955	0.950	0.961	0.9621
FCNsignal	MSE	0.141	0.068	0.098	0.117	0.058	0.108	0.216	0.318	0.1443
	PCC	0.825	0.671	0.822	0.700	0.678	0.789	0.787	0.797	0.7586
	AUC	0.992	0.961	0.978	0.944	0.873	0.960	0.943	0.954	0.9506
	AUPRC	0.991	0.982	0.983	0.945	0.883	0.959	0.948	0.958	0.9561

Note: Bold numbers are the best results.

5. Conclusions

In this work, we propose an integrate context-aware neural framework, named GNet, based on a gated mechanism and improved external attention to consider TF binding signals prediction at a single nucleotide resolution, determination of binding or non-binding regions at the sequence level and motif recognition problems respectively. Most previous studies have only performed regression or classification tasks at the sequence level, rarely at the single nucleotide level, and few model studies have integrated multiple tasks such as a regression and classification. On 53 human TF ChIP-seq datasets and 6 chromatin accessibility ATAC-seq datasets, the experimental results show that GNet has an excellent performance in the three tasks, is superior to several competitive methods, and the results of cross-species studies on 15 human and 18 mouse TF data show that GNet also shows the best performance in cross-species prediction over the competitive methods.

GNet performs well in these three tasks and cross-species studies, which can extract spatial and temporal patterns benefiting from the following aspects: the addition of two new features, gated mechanism including gated highway network unit and the skip identity connection in decoder architecture, and improved attention mechanisms. RFHC coding considers the three chemical properties of nucleotides, and ND coding considers the position and frequency information of nucleotides, adding these two types of features effectively avoids the loss of information. At the same time, the gated mechanism can increase the flexibility of the model, extract more effective contextual information and reduce the gradient backflow blocking problem. Combined with the improved attention mechanism DEA, the model can be trained to learn the implicit relationships inter-sample and among samples to improve the model performance.

Although GNet model has achieved excellent performance, there are still some limitations, such as not deeply constructing the decoder framework. In future work, we will consider introducing K-mer or Word2vec [46] embedding to consider the dependence between nucleotides. At the same time, complementary sequences, reverse sequences and complementary reverse sequences of DNA have been proved to play a certain role in TFBSs prediction [47]. Deconvolution has also been widely used in various decoding networks [48,49]. Meanwhile, the algorithm of learning how to embed DNA sequence and TF tag into the same space is also generated [50], which provides new visions and opportunities for our subsequent studying.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors thank those who contributed to this paper, as well as the reviewers for their careful reading and valuable suggestions.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. G. Badis, M. F. Berger, A. A. Philippakis, S. Talukder, A. R. Gehrke, S. A. Jaeger, et al., Diversity and complexity in DNA recognition by transcription factors, *Science*, **324** (2009), 1720–1723. <https://doi.org/10.1126/science.1162327>
2. A. Jolma, J. Yan, T. Whittington, J. Toivonen, K. R. Nitta, P. Rastas, et al., DNA-binding specificities of human transcription factors, *Cell*, **152** (2013), 327–339. <https://doi.org/10.1016/j.cell.2012.12.009>
3. P. J. Mitchell, R. Tjian, Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins, *Science*, **245** (1989), 371–378. <https://doi.org/10.1126/science.2667136>
4. L. Elnitski, V. X. Jin, P. J. Farnham, S. J. Jones, Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques, *Genome Res.*, **16** (2006), 1455–1464. <https://doi.org/10.1101/gr.4140006>
5. M. F. Berger, A. A. Philippakis, A. M. Qureshi, F. S. He, P. W. Estep, M. L. Bulyk, Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities, *Nat. Biotechnol.*, **24** (2006), 1429–1435. <https://doi.org/10.1038/nbt1246>
6. A. Jolma, T. Kivioja, J. Toivonen, L. Cheng, G. Wei, M. Enge, et al., Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities, *Genome Res.*, **20** (2010), 861–873. <https://doi.org/10.1101/gr.100552.109>
7. T. S. Furey, ChIP–seq and beyond: New and improved methodologies to detect and characterize protein–DNA interactions, *Nat. Rev. Genet.*, **13** (2012), 840–852. <https://doi.org/10.1038/nrg3306>
8. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position, *Nat. Methods*, **10** (2013), 1213–1218. <https://doi.org/10.1038/nmeth.2688>
9. C. Fletez-Brant, D. Lee, A. S. McCallion, M. A. Beer, kmer-SVM: a web server for identifying predictive regulatory sequence features in genomic data sets, *Nucleic Acids Res.*, **41** (2013), W544–W556. <https://doi.org/10.1093/nar/gkt519>
10. M. Ghandi, D. Lee, M. Mohammad-Noori, M. A. Beer, Enhanced regulatory sequence prediction using gapped k-mer features, *PLoS Comput. Biol.*, **10** (2014), e1003711. <https://doi.org/10.1371/journal.pcbi.1003711>
11. T. L. Bailey, N. Williams, C. Mischel, W. W. Li, MEME: Discovering and analyzing DNA and protein sequence motifs, *Nucleic Acids Res.*, **34** (2006), W369–W373. <https://doi.org/10.1093/nar/gkl198>
12. T. L. Bailey, STREME: Accurate and versatile sequence motif discovery, *Bioinformatics*, **37** (2021), 2834–2840. <https://doi.org/10.1093/bioinformatics/btab203>
13. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, **521** (2015), 436–444. <https://doi.org/10.1038/nature14539>
14. D. Berrar, W. Dubitzky, Deep learning in bioinformatics and biomedicine, *Briefings Bioinf.*, **22** (2021), 1513–1514. <https://doi.org/10.1093/bib/bbab087>
15. B. Alipanahi, A. Delong, M. T. Weirauch, B. J. Frey, Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning, *Nat. Biotechnol.*, **33** (2015), 831–838. <https://doi.org/10.1038/nbt.3300>

16. J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning–based sequence model, *Nat. Methods*, **12** (2015), 931–934. <https://doi.org/10.1038/nmeth.3547>
17. D. Quang, X. Xie, DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences, *Nucleic Acids Res.*, **44** (2016), e107–e107. <https://doi.org/10.1093/nar/gkw226>
18. C. Chen, J. Hou, X. Shi, H. Yang, J. A. Birchler, J. Cheng, DeepGRN: Prediction of transcription factor binding site across cell-types using attention-based deep neural networks, *BMC Bioinf.*, **22** (2021), 38. <https://doi.org/10.1186/s12859-020-03952-1>
19. Q. X. X. Lin, D. Thieffry, S. Jha, T. Benoukraf, TFregulomeR reveals transcription factors' context-specific features and functions, *Nucleic Acids Res.*, **48** (2020), e10–e10. <https://doi.org/10.1093/nar/gkz1088>
20. Ž. Avsec, M. Weilert, A. Shrikumar, S. Krueger, A. Alexandari, K. Dalal, et al., Base-resolution models of transcription-factor binding reveal soft motif syntax, *Nat. Genet.*, **53** (2021), 354–366. <https://doi.org/10.1038/s41588-021-00782-6>
21. Q. Zhang, Y. He, S. Wang, Z. Chen, Z. Guo, Z. Cui, et al., Base-resolution prediction of transcription factor binding signals by a deep learning framework, *PLoS Comput. Biol.*, **18** (2022), e1009941. <https://doi.org/10.1371/journal.pcbi.1009941>
22. P. J. Werbos, Generalization of backpropagation with application to a recurrent gas market model, *Neural Networks*, **1** (1988), 339–356. [https://doi.org/10.1016/0893-6080\(88\)90007-X](https://doi.org/10.1016/0893-6080(88)90007-X)
23. R. K. Srivastava, K. Greff, J. J. C. S. Schmidhuber, Training very deep networks, *arXiv preprint*, (2015), arXiv:1507.06228. <https://doi.org/10.48550/arXiv.1507.06228>
24. J. G. Zilly, R. K. Srivastava, J. Koutník, J. Schmidhuber, Recurrent highway networks, *arXiv preprint*, (2016), arXiv:1607.03474. <https://doi.org/10.48550/arXiv.1607.03474>
25. Y. N. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks, *arXiv preprint*, (2016), arXiv:1612.08083. <https://doi.org/10.48550/arXiv.1612.08083>
26. D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, *arXiv preprint*, (2014), arXiv:1409.0473. <https://doi.org/10.48550/arXiv.1409.0473>
27. K. Xu, J. L. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, et al., Show, attend and tell: Neural image caption generation with visual attention, *arXiv preprint*, (2015), arXiv:1502.03044. <https://doi.org/10.48550/arXiv.1502.03044>
28. Y. Guo, C. Li, D. Zhou, J. Cao, H. Liang, Context-aware dynamic neural computational models for accurate Poly(A) signal prediction, *Neural networks*, **152** (2022), 287–299. <https://doi.org/10.1016/j.neunet.2022.04.025>
29. Y. Guo, D. Zhou, W. Li, J. Cao, R. Nie, L. Xiong, et al., Identifying polyadenylation signals with biological embedding via self-attentive gated convolutional highway networks, *Appl. Soft Comput.*, **103** (2021), 107133. <https://doi.org/10.1016/j.asoc.2021.107133>
30. J. Lanchantin, R. Singh, Z. Lin, Y. Qi, Deep motif: Visualizing genomic sequence classifications, *arXiv preprint*, (2016), arXiv:1605.01133. <https://doi.org/10.48550/arXiv.1605.01133>
31. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, *arXiv preprint*, (2017), arXiv:1706.03762. <https://doi.org/10.48550/arXiv.1706.03762>
32. R. Li, Z. Wu, J. Jia, Y. Bu, H. Meng, Towards discriminative representation learning for speech emotion recognition, in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, (2019), 5060–5066. <https://doi.org/10.24963/ijcai.2019/703>

33. M. H. Guo, Z. N. Liu, T. J. Mu, S. M. Hu, Beyond self-attention: External attention using two linear layers for visual tasks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **45** (2023), 5436–5447. <https://doi.org/10.1109/TPAMI.2022.3211006>
34. E. A. Feingold, P. J. Good, M. S. Guyer, S. Kamholz, L. Liefer, K. Wetterstrand, The ENCODE (ENCyclopedia Of DNA elements) project, *Science*, **306** (2004), 636–640. <https://doi.org/10.1126/science.1105136>
35. M. T. Weirauch, A. Cote, R. Norel, M. Annala, Y. Zhao, T. R. Riley, et al., Evaluation of methods for modeling transcription factor sequence specificity, *Nat. Biotechnol.*, **31** (2013), 126–134. <https://doi.org/10.1038/nbt.2486>
36. B. Manavalan, S. Basith, T. H. Shin, L. Wei, G. Lee, Meta-4mCpred: A Sequence-based meta-predictor for accurate DNA 4mC site prediction using effective feature representation, *Mol. Ther. Nucleic Acids*, **16** (2019), 733–744. <https://doi.org/10.1016/j.omtn.2019.04.019>
37. Y. Yang, Z. Hou, Y. Wang, H. Ma, P. Sun, Z. Ma, et al., HCRNet: High-throughput circRNA-binding event identification from CLIP-seq data using deep temporal convolutional network, *Briefings Bioinf.*, **23** (2022), bbac027. <https://doi.org/10.1093/bib/bbac027>
38. K. Liu, W. Chen, iMRM: A platform for simultaneously identifying multiple kinds of RNA modifications, *Bioinformatics*, **36** (2020), 3336–3342. <https://doi.org/10.1093/bioinformatics/btaa155>
39. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, USA, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
40. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.*, **15** (2014), 1929–1958.
41. K. Cho, B. V. Merriënboer, D. Bahdanau, Y. J. C. S. Bengio, On the properties of neural machine translation: Encoder-decoder approaches, in *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, ACL, Doha, Qatar, (2014), 103–111. <https://doi.org/10.3115/v1/W14-4012>
42. D. Kingma, J. J. C. S. Ba, Adam: A method for stochastic optimization, *arXiv preprint*, (2014), arXiv:1412.6980. <https://doi.org/10.48550/arXiv.1412.6980>
43. Q. Zhang, S. Wang, Z. Chen, Y. He, Q. Liu, D. S. Huang, Locating transcription factor binding sites by fully convolutional neural network, *Briefings Bioinf.*, **22** (2021), bbaa435. <https://doi.org/10.1093/bib/bbaa435>
44. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, et al., HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis, *Nucleic Acids Res.*, **46** (2018), D252–D259. <https://doi.org/10.1093/nar/gkx1106>
45. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. S. Noble, Quantifying similarity between motifs, *Genome Biol.*, **8** (2007), R24. <https://doi.org/10.1186/gb-2007-8-2-r24>
46. T. Mikolov, K. Chen, G. Corrado, J. J. C. S. Dean, Efficient estimation of word representations in vector space, *arXiv preprint*, (2013), arXiv:1301.3781. <https://doi.org/10.48550/arXiv.1301.3781>
47. L. Deng, H. Wu, X. Liu, H. Liu, DeepD2V: A novel deep learning-based framework for predicting transcription factor binding sites from combined DNA sequence, *Int. J. Mol. Sci.*, **22** (2021), 5521. <https://doi.org/10.3390/ijms22115521>

48. M. D. Zeiler, G. W. Taylor, R. Fergus, Adaptive deconvolutional networks for mid and high level feature learning, in *2011 International Conference on Computer Vision*, IEEE, Barcelona, Spain, (2011), 2018–2025. <https://doi.org/10.1109/ICCV.2011.6126474>
49. M. D. Zeiler, D. Krishnan, G. W. Taylor, R. Fergus, Deconvolutional networks, in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, San Francisco, USA, (2010), 2528–2535. <https://doi.org/10.1109/CVPR.2010.5539957>
50. H. Yuan, M. Kshirsagar, L. Zamparo, Y. Lu, C. S. Leslie, BindSpace decodes transcription factor binding signals by large-scale sequence embedding, *Nat. Methods*, **16** (2019), 858–861. <https://doi.org/10.1038/s41592-019-0511-y>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)