Predicting PRDM9 binding sites by a convolutional neural network and verification using genetic recombination map

TAKAHIRO NAKAMURA[1] TOSHINORI ENDO[1,2]
NAOKI OSADA[1,2,a)]

Abstract: PR domain-containing 9 (PRDM9) is a zinc-finger protein that binds to specific DNA motifs and induces the crossing-over between chromosomes, resulting in a high recombination rate around binding sites. In this study, we developed a strategy to evaluate the prediction accuracy of PRDM9 binding site by examining the correlation with local recombination rate to avoid the effect of overfitting to one type of data. We compared the methods using position-specific weight matrix (PWM), which has been commonly used in previous studies, and convolutional network (CNN), which has recently performed well. Approximately 170,000 genomic DNA fragments of humans (301 bp each) containing the Chromatin Immuno-Precipitation with high-throughput sequencing (ChIP-seq) peak of PRDM9 of B-allele in the HEK293T cell line were used for constructing PWM and positive data to train CNN. We found that CNN outperformed PWM in terms of area under the curve, and the prediction scores of CNN correlated more strongly with the local recombination rate than PWM. We also investigated the potential PRDM9 binding sites missed by the ChIP-seq experiments but labeled as positive in CNN and discuss the reason for the difference in performances.

Keywords: binding site prediction, CNN, PRDM9, genetic recombination

1. Introduction

Understanding sequence specificities and genomic binding sites recognized by DNA-binding proteins is crucial for elucidating the regulatory mechanisms of genes, such as transcription and selective splicing. Computational models created from experimentally determined sequence data have been used to predict binding sites on a genome-wide basis [1]. In general, when evaluating a model, the prediction accuracy is verified using test data obtained in advance, apart from the training data used to tune the model parameters. However, overfitting to noise was reported in a modeling using protein-binding microarray data [2]. In addition, it is difficult to use only one experimental data to evaluate the generalized performance of a model. To amend the effect of overfitting on one type of training data, a validation method that does not require test data is desirable.

PR domain-containing 9 (PRDM9) is a zinc finger protein that induces chromosomal recombination in mammalian germ cells by localizing in a narrow (1–2 kb) hotspot, where recombination frequently occurs in early meiosis [3]–[5]. The binding of PRDM9 induces topoisomerase sporulation-specific 11, which forms a double-strand break, and crossing-over occur at the site [6], [7]. The binding sites can be identified by Chromatin Immuno-Precipitation with high-throughput sequencing (ChIP-seq) experiment. In this study, we focus on the correlation between PRDM9 binding and genetic recombination to validate the prediction accuracy.

For binding prediction, position-specific weight matrix (PWM) is one of the most widely used method to identify protein binding sites [8]. PWM is a matrix of data representing the DNA motifs to which a protein binds and is created on the basis of the frequency of nucleotides at each position. Meanwhile, CNN can learn complex features from large amount of DNA sequences and have been reported to predict DNA-protein binding better than other methods, including PWM [2].

In this study, we examined the correlation between the predicted binding score for a genomic fragment and the recombination rate of the region to test whether CNN outperforms PWM in terms of the correlation strength between prediction scores and local recombination rates. We used the DNA sequence of the binding site of human PRDM9 and recombination map. First, we compared the performance of the PWM and CNN in terms of area under the curve (AUC) using test data. We observed a stronger correlation between the prediction scores and local recombination rates in CNN than in PWM. In addition, we revealed that the DNA fragments with high prediction scores in CNN had higher local recombination rates, even without ChIP-seq peaks. Further, we examined the feature that CNN recognizes and discussed possible binding mechanisms of PRDM9.

2. Materials and Methods

2.1 Materials

We retrieved data of ChIP-seq experiments performed by Altemose et al. (2017) to obtain DNA-PRDM9 binding sequence data [7] (https://www.ncbi.nlm.nih.gov/geo/, GSE99407). The experiments obtained 170,198 nonoverlapping peak fragments by transfecting HEK293T cell line with N-terminal YFP-tagged human PRDM9 of B-allele (hereafter, PRDM9) and conducted ChIP-seq against YFP. Each peak was included at the center of a 301-bp region. We assigned 10% of the peak regions (17,009) to positive test data and the remaining 90% (153,189) to positive

1 Graduate School of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
2 Faculty of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
a) nosada@ist.hokudai.ac.jp

training data for CNN. For negative data, 100-bp-length sequences were extracted without overlaps from human genomes (hg19) that are at least 1,000 bp away from the center of ChIP-seq peaks.

We used a recombination map estimated by pyrho [9] considering the population demography using genomic polymorphism data. In this study, we used the estimations based on the demography of the CEU (Utah residents with northern and western European ancestry) population.

2.2 Preparation for test data

To verify the prediction accuracy of CNN and PWM, we created PRDM9-binding test data. Positive test data were generated by cutting out a 100-bp DNA fragment containing 31 bp around the peak (ChIP-seq core region) at a random position within the fragment. Negative test data were randomly sampled so that the positive data constituted 2% of total test data based on the ratio of the ChIP-seq-positive to -negative regions, yielding 833,441 negative test regions.

2.3 PWM used in the study and its scoring

We used PWMs obtained from the same ChIP-seq data and constructed using the Bayesian de novo motif-finding algorithm [10] (Additional file of Altemose et al. (2017 [7]). In total 17 PWMs were available.

For scoring, the log-likelihood ratio score S, which is obtained based on the occurrence probability of a motif considering the nucleotide frequency in the background sequence, was calculated. The nucleotide frequencies of the background sequences were calculated for each chromosome and strand. $S_{i,j,k}$ is calculated by summing up the log-likelihood values for all sites of the j-th PWM, starting from the k-th site of the i-th fragment for both strands [11].

Next, $S_{i,j,k}$ was summarized to the score for the i-th fragment and the j-th PWM, $S_{i,j}$, expressed as follows:

$$S_{i,j} = \max_{k} S_{i,j,k} \quad (1)$$

There are two ways to summarize the score for the i-th fragment. One is to accept the maximum score among N PWMs, $S_{\max,i}$, expressed as follows:

$$S_{\max,i} = \max_{1 \le j \le N} S_{i,j} \quad (2)$$

The other method is to sum the scores of all N PWMs, $S_{\mathrm{sum},i}$, expressed as follows:

$$S_{\mathrm{sum},i} = \sum_{j=1}^{N} 2^{S_{i,j}} \quad (3)$$

Next, we determined how many PWMs should be used for the prediction. All PWMs were ranked by the probability of hitting within 100 bp of ChIP-seq peaks. We added PWMs one by one from the one with the maximum score and evaluated the performance in terms of AUC using $S_{\max,i}$ and $S_{\mathrm{sum},i}$. As a result, AUC was maximized using up to the 15th PWM with $S_{\mathrm{sum},i}$, and the condition was adopted as the final PWM scoring (Table 1).

Table 1 AUC transition by changing cutoff of the PWM number

PWM	Max Method	Sum Method
Until 1st	0.8330	0.8330
Until 2nd	0.8363	0.8383
Until 3rd	0.8329	0.8358
Until 4th	0.8250	0.8279
Until 5th	0.8301	0.8339
Until 6th	0.8364	0.8403
Until 7th	0.8414	0.8451
Until 8th	0.8449	0.8488
Until 9th	0.8469	0.8510
Until 10th	0.8503	0.8557
Until 11th	0.8528	0.8584
Until 12th	0.8512	0.8565
Until 13th	0.8568	0.8615
Until 14th	0.8577	0.8623
Until 15th	0.8582	0.8629
Until 16th	0.8203	0.8294
Until 17th	0.7941	0.8040

2.4 CNN and its scoring

We modified the equivariant Bayesian convolutional network (EBCN) [12] written in Python using deep learning libraries: TensorFlow-GPU (Version 1.13.1) [13] and Keras-GPU (Version 2.3.1). The model employs Monte Carlo (MC) dropout [14] in which the dropout is performed during not only training but also prediction, and the average of the repeated output predictions is produced. In addition, the model is internally adjusted so that the output for an input forward sequence is equal to the output of its reverse complement sequence. These features result in a clean separation of the internal representations [15] and a small inconsistency in the model accuracy from training to training. The input is a DNA fragment of constant length, which is converted into a matrix of 4 × input length by one-hot vectorizing the nucleotides (Table 2). Then, the matrix undergoes multiple convolutional and pooling layers to extract features and outputs the scores (0–1) for PRDM9 binding through dense layer with softmax function.

Table 2 One-hot encoding for nucleotides

Nucleotide	One-hot vector
A	$[1,0,0,0]^{\mathrm{T}}$
C	$[0,1,0,0]^{\mathrm{T}}$
G	$[0,0,1,0]^{\mathrm{T}}$
T	$[0,0,0,1]^{\mathrm{T}}$

2.5 Creating training data for CNN and the network optimization

Positive training data were created using 90% of the ChIP-seq positive region. To construct robust CNN, we tried three training methods with some changes to the positive fragments. First, DNA fragments of 100-bp length with the ChIP-seq peak in its center were used for training (training method 1). Next, DNA fragments of 100-bp length with ChIP-seq core at a random

position were used as positive training data (training method 2). Finally, we trained the model using three-fold augmented positive fragments (training method 3). In training method 3, for each fragment, the peak was randomly replaced within the fragment for three times. Negative data were increased to match the number of positive augmented data.

We employed the EBCN structure (recombination topology) developed by Brown and Lunter (2019) [12] and further optimized the network structure and hyperparameters. We tuned number of filters in each convolutional layer, kernel size of the first convolutional layer, number of internal convolutional layers, kernel size of inner convolutional layers, pool size, learning rate, coefficient of the L2 normalization term, dropout rate, batch size, optimizer and activation function of the convolutional layer (Table 3). In particular, batch size was optimized from among 32, 64, 128, 256, 512, 1024, and 2048, optimizer was from among SGD [16], Momentum [17], and Adam [18], and activation function was from among ELU [19], ReLU [20], SELU [21], and LReLU [22]. The hyperparameter auto-optimization library, Optuna, was employed [23].

Table 3 List of hyperparameters and searching range and step

Hyperparameters	Minimum	Maximum	Step
The number of filters	2	100	2
The kernel size of first filter	2	40	2
The number of internal conv. layers	1	3	1
The kernel size of internal filter	2	4	1
Pool size	2	6	2
L2 coefficient	0	0.01	0.0001
Learning rate	0.0001	0.1	None
Dropout rate	0	1	None

Owing to time constraints, approximately 10% of the training data were used for each training method. The epoch was set to 20, and the trial was conducted 100 times. After the hyperparameter search, if the network structure conflicts with equivariance, we shaved off at most 1 bp of the input sequence and completed structures of CNN for each training method (Tables 4–9).

Table 4 Optimized network structure of final CNN for training method 1

Layers	Parameters
Equivariant Conv1D Layer 1	Filters = 34; kernel size = 22; activation function = LReLU
Equivariant MC Dropout Layer 1	Dropout rate = 0.1731
Spatial MaxPooling1D Layer	Pool size = 6
Equivariant Conv1D Layer 2	Filters = 34; kernel size = 4; activation function = LReLU
Equivariant MC Dropout Layer 2	Dropout rate = 0.1731
Equivariant Conv1D Layer 3	Filters = 34; kernel size = 4; activation function = LReLU
Equivariant MC Dropout Layer 3	Dropout rate = 0.1731
Equivariant Conv1D Layer 4	Filters = 34; kernel size = 4; activation function = LReLU
Equivariant MC Dropout Layer4	Dropout rate = 0.1731
Reverse Complement Sum Pooling Layer	None
Global Spatial MaxPooling1D Layer	None
Dense Layer	Activation function = Softmax

Table 5 Hyperparameters of backpropagation for training method 1

Hyperparameter	Optimal Value
Learning rate	1.659×10^{-3}
L2 coefficient	0.0011
Batch size	64
Optimizer	Adam

Table 6 Optimized network structure of final CNN for training method 2

Layers	Parameters
Equivariant Conv1D Layer 1	Filters = 54; kernel size = 40; activation function = LReLU
Equivariant MC Dropout Layer 1	Dropout rate = 0.2360
Spatial MaxPooling1D Layer	Pool size = 6
Equivariant Conv1D Layer 2	Filters = 54; kernel size = 2; activation function = LReLU
Equivariant MC Dropout Layer 2	Dropout rate = 0.2360
Equivariant Conv1D Layer 3	Filters = 54; kernel size = 2; activation function = LReLU
Equivariant MC Dropout Layer 3	Dropout rate = 0.2360
Reverse Complement Sum Pooling Layer	None
Global Spatial MaxPooling1D Layer	None
Dense Layer	Activation function = Softmax

Table 7 Hyperparameters of backpropagation for training method 2

Hyperparameter	Optimal Value
Learning rate	9.652×10^{-2}
L2 coefficient	0.0033
Batch size	32
Optimizer	Momentum

Table 8 Optimized network structure of final CNN for training method 3

Layers	Parameters
Equivariant Conv1D Layer 1	Filters = 88; kernel size = 14; activation function = LReLU
Equivariant MC Dropout Layer 1	Dropout rate = 0.3980
Spatial MaxPooling1D Layer	Pool size = 2
Equivariant Conv1D Layer 2	Filters = 88; kernel size = 3; activation function = LReLU
Equivariant MC Dropout Layer 2	Dropout rate = 0.3980
Reverse Complement Sum Pooling Layer	None
Global Spatial MaxPooling1D Layer	None
Dense Layer	Activation function = Softmax

Table 9 Hyperparameters of backpropagation for training method 3

Hyperparameter	Optimal Value
Learning rate	1.737×10^{-3}
L2 coefficient	0.0
Batch size	128
Optimizer	Adam

3. Results

3.1 CNN outperformed PWM in prediction accuracy using test data

We trained CNN in all three training methods. The training was repeated five times, 50 epochs each, and the model that had the maximum accuracy was adopted as the final model for each of three training method. We compared the AUC of the three CNN training methods and the PWM, using the test data. The AUC of PWM was 0.8629. The AUCs of training methods 1–3 were 0.9125, 0.8950, and 0.9167, respectively. The all CNN training methods outperformed PWM. Training method 3 showed the highest AUC, and the PWM showed the lowest AUC. In the following analyses, we use the results with training method 3 as a representative.

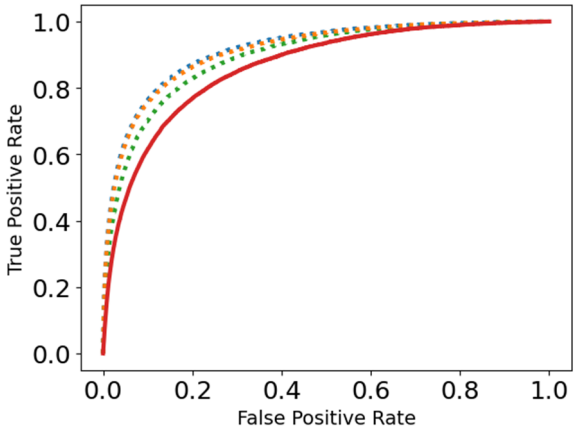

Figure 1 Receiver operating characteristic (ROC) curve with test data

ROC curves are shown for PWM and three training methods. The dashed orange, dashed green, and dashed blue lines represent the ROC curves of training method 1–3, respectively. The solid red line indicates the ROC curve of PWM.

3.2 CNN again outperformed PWM in prediction accuracy validated using recombination map

To evaluate model prediction accuracy using recombination rate, the entire autosomal genome was divided into fragments of 100-bp length from the start position, and fragments without ambiguous nucleotide (N) were used. For each fragment, the average recombination rate (cM/bp) was calculated from the recombination map. We scored the probability of PRDM9 binding for each fragment using CNN and PWM. The correlation coefficient (ρ) between the prediction score and recombination rate was statistically significant for CNN (ρ = 0.180, p-value < 1.0×10^{-16}) and for PWM (ρ = 0.138, p-value < 1.0×10^{-16}). The correlation coefficient was stronger for CNN than for PWM.

The scored fragments were sorted in descending order by prediction score, and fragments were grouped into 10 bins (10% of data for each bin). The mean and standard error of recombination rate in each bin was calculated and shown in Figure 2. For CNN, the average recombination rate was high in bins with high prediction scores and low in bins with low prediction scores (Fig. 2B). To observe the correlation within the group of high scores, we grouped the top 10% of fragments into 1% bins and examined the average recombination rate. Again, CNN showed a stronger correlation than PWM (Fig. 2A).

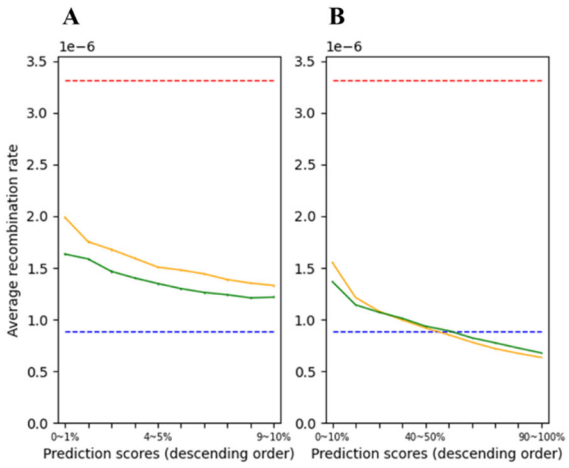

Figure 2 Correlation between recombination rate and prediction score for all fragments

The yellow and green lines represent the mean recombination rate in bins for CNN score and PWM score, respectively. The error bars are drawn in the figure but are nearly invisible because of the large sample sizes. The average recombination rate of all ChIP-seq-negative and -positive fragments are shown in the blue and red dashed lines, respectively.

3.3 Can PWM and CNN detect potential recombination hotspots from ChIP-seq-negative fragments?

To examine whether PWM and CNN detect potential recombination hotspots missed by the ChIP-seq experiments, we plot the prediction score and average recombination rate in each bin for ChIP-seq-negative fragments (Figure 3). Although the correlation became somewhat weaker, we observed a statistically significant correlation between prediction scores and recombination rates. The correlation coefficient between the prediction score and the recombination rate (ρ) was 0.129 (p-value $< 1.0\times10^{-16}$) in PWM and 0.169 (p-value $< 1.0\times10^{-16}$) in CNN. The results indicate that PWM and CNN properly capture sequence features of recombination hotspots.

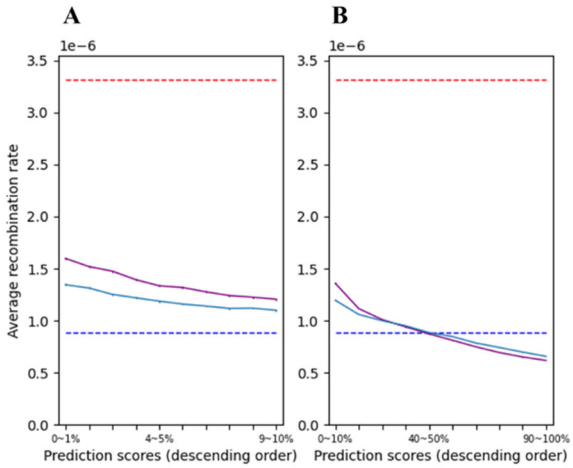

Figure 3 Correlation between recombination rate and prediction score for top 10%

The purple and light blue lines represent the mean recombination rate in bins for CNN score and PWM score to ChIP-seq-negative fragments, respectively. The error bars were drawn in the figure but invisible because of the large sample size. The average recombination rate of all ChIP-seq-negative and -positive fragments are shown in blue and red dashed lines, respectively.

3.4 CNN captures sequence features surrounding PRDM9-binding site

The results shown in Subsections 3.1 and 3.2 indicated that CNN outperformed PWM in predicting PRDM9 binding, and the result presented in Subsection 3.3 suggested the possibility of detecting PRDM9 binding sites undetected by ChIP-seq experiments. However, the high recombination rate of fragments with high prediction scores did not necessarily mean that the fragments contain actual PRDM9 binding sites. There are two possible reasons. One possibility is that CNN finds the potential binding sites missed in the ChIP-seq experiments due to differences in the conditions of PRDM9 binding between actual germ cells and cultured cells, or it detects binding sites of other alleles rather than B-allele used for the experiments. Another possibility is that CNN detects the recombination hotspots using surrounding features of PRDM9 rather than the binding motif.

To test the latter possibility, we examined the distance from ChIP-seq-negative and CNN-positive fragments to the nearest ChIP-seq peak. To define CNN-positive fragments, the optimal threshold was decided using the ROC curve (Figure 1). The point whose coordinates were closest to (0,1) was used for the threshold [24]. We obtained 4,118,963 CNN-positive and ChIP-seq-negative fragments and 22,062,475 CNN-negative and ChIP-seq-negative fragments with the threshold. We randomly extracted equal numbers of CNN-positive and -negative fragments without overlap (411,886 fragments each) from the ChIP-seq-negative fragments and examined the distance to the nearest ChIP-seq peak (Figure 4). The results showed that the CNN-positive DNA fragments were significantly closer to the ChIP-seq peak than the CNN-negative fragments (p-value $< 1.0\times10^{-16}$, Mann–Whitney U Test).

For comparison, we conducted the same procedure using PWM-positive fragments. We obtained 5,579,323 PWM-positive and ChIP-seq-negative fragments and 20,602,115 PWM-negative and ChIP-seq-negative fragments. We randomly extracted equal numbers of PWM-positive and PWM-negative fragments without overlap (557,922 fragments each). The CNN-positive fragment is still significantly closer to ChIP-seq peaks than the PWM-positive fragments (p-value $< 1.0\times10^{-16}$, Mann–Whitney U Test) (Figure 5).

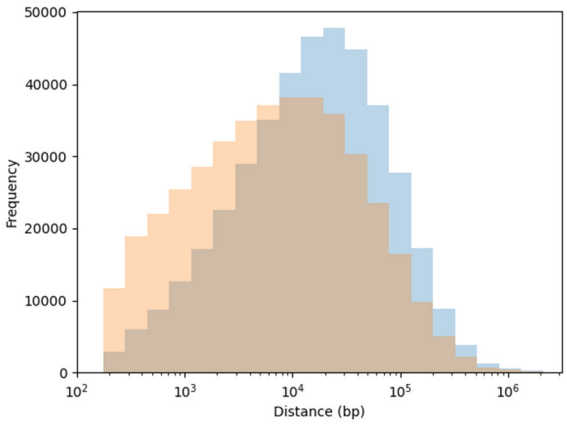

Figure 4 Distance to the nearest ChIP-seq peak center from ChIP-seq-negative and CNN-positive fragments
Orange histogram represents the frequency of distance between the center of CNN-positive fragment and the center of the ChIP-seq peak. Blue histogram represents the frequency of distance between the center of CNN-negative fragment and the center of the ChIP-seq peak.

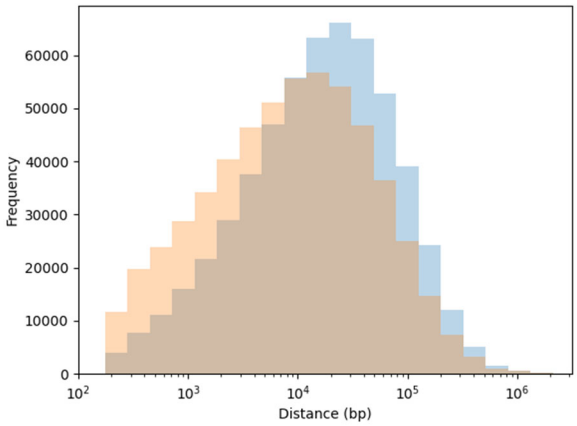

Figure 5 Distance to the nearest ChIP-seq peak center from ChIP-seq-negative and PWM-positive fragments
Orange histogram represents the frequency of distance between the center of PWM-positive fragment and the center of the ChIP-seq peak. Blue histogram represents the frequency of distance between the center of PWM-negative fragment and the center of the ChIP-seq peak.

4. Discussion

4.1 Advantages of validation using recombination map

In this study, we compared the performance of CNN and PWM in terms of the correlation between prediction scores and recombination rates other than validating prediction models using test data. In these two methods, CNN consistently outperformed PWM and the method using recombination map was confirmed its usefulness for validation at least. Also, the method allows to validate models without the influence of overfitting to the ChIP-seq data and verifying the prediction accuracy even for ChIP-seq-negative data. This validation method is not limited to PWM or CNN but can be applied to a variety of prediction models.

4.2 Comparison between CNN and PWM

CNN outperformed PWM in both validation methods, which used test data and recombination rate. The reason was that, as suggested by the distances from the CNN-positive and ChIP-seq-negative fragments to the nearest ChIP-seq peak (Fig. 4), CNN could predict on the basis of a comprehensive judgment of the surrounding features in addition to directly recognizing the binding site by the PRDM9 zinc finger protein. Further, the fact that the discriminatory performance to the ChIP-seq-positive fragments was improved by adding weaker motifs of PWMs suggested the usefulness of information other than direct protein-binding sites. Nevertheless, CNN that used only information about 100 bp around the peak (training method 1) resulted in higher accuracy than the PWM. These results suggested that CNN is a better modeling method for predicting the binding site of PRDM9.

4.3 Features recognized by CNN and the formation of PRDM9 binding sites

Our analysis suggested that CNN captures some unknown features around the ChIP-seq peak other than the zinc finger binding motif of PRDM9. For example, biased gene conversion would form GC-rich regions around recombination hotspots [25], and PRDM9 binding motifs are highly enriched with the cytosine nucleotide. Therefore, PRDM9 binding sites tend to locate in regions with high GC content, and CNN might have recognized such features. In addition, CNN potentially detected a different PRDM9 binding site rather than the ChIP-seq peak or similar binding sites. Further verification will be needed to make an assertion.

5. Concluding Remarks

We developed a strategy to evaluate the accuracy of PRDM9 binding site prediction by examining the correlation with local recombination rate to avoid the effect of overfitting to one type of data. We evaluated PWM and CNN methods for detecting PRDM9 binding sites using not only test data but also recombination map and found that CNN outperformed in both cases. This validation method is applicable for not only PWM and CNN but also variety of models. Further, the genomic distance between the ChIP-seq peak and CNN-positive fragments suggested that CNN recognized not only the binding motif of the zinc finger but also the features of surrounding sequences when predicting the PRDM9 binding to DNA fragments.

Reference

[1] Y. Zeng, M. Gong, M. Lin, D. Gao, and Y. Zhang, “A review about transcription factor binding sites prediction based on deep learning,” *IEEE Access*, vol. 8, 2020.

[2] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, “Predicting the sequence specificities of DNA- and RNA-

binding proteins by deep learning," *Nat. Biotechnol.*, vol. 33, no. 8, pp. 831–838, 2015.

[3] F. Baudat *et al.*, "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice," *Science*, vol. 327, no. 5967, pp. 836–840, Feb. 2010.

[4] S. Myers *et al.*, "Drive against hotspot motifs in Primates implicates the PRDM9 gene in meiotic recombination," *Science*, vol. 327, no. 5967, pp. 876–879, Feb. 2010.

[5] E. D. Parvanov, P. M. Petkov, and K. Paigen, "Prdm9 controls activation of mammalian recombination hotspots," *Science*, vol. 327, no. 5967, p. 835, 2010.

[6] M. J. Neale and S. Keeney, "Clarifying the mechanics of DNA strand exchange in meiotic recombination," *Nature*, vol. 442, no. 7099, pp. 153–158, Jul. 2006.

[7] N. Altemose *et al.*, "A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis," *eLife*, vol. 6, pp. 1–46, 2017.

[8] X. Xia, "Position weight matrix, Gibbs sampler, and the associated significance tests in motif characterization and prediction," *Scientifica (Cairo)*, vol. 2012, 917540, 2012.

[9] J. P. Spence and Y. S. Song, "Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations," *Sci. Adv.*, vol. 5, no. 10, eaaw9206, 2019.

[10] B. Davies *et al.*, "Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice," *Nature*, vol. 530, no. 7589, pp. 171–176, 2016.

[11] M. Megraw, V. Baev, V. Rusinov, S. T. Jensen, K. Kalantidis, and A. G. Hatzigeorgiou, "MicroRNA promoter element discovery in Arabidopsis," *Rna*, vol. 12, no. 9, pp. 1612–1619, 2006.

[12] R. C. Brown and G. Lunter, "An equivariant Bayesian convolutional network predicts recombination hotspots and accurately resolves binding motifs," *Bioinformatics*, vol. 35, no. 13, pp. 2177–2184, 2019.

[13] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016.

[14] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," 33rd Int. Conf. Mach. Learn. ICML 2016, vol. 3, pp. 1651–1660, 2016.

[15] A. Shrikumar, P. Greenside, and A. Kundaje, "Reverse-complement parameter sharing improves deep learning models for genomics," *bioRxiv*, 2017.

[16] S. Ruder, "An overview of gradient descent optimization algorithms," pp. 1–14, 2016.

[17] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Iclr*, pp. 1–15, Dec. 2014.

[19] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 4th Int. Conf. Learn. Represent. ICLR Conf. Track *Proc.*, vol. 2016, pp. 1–14, 2016–.

[20] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," Proc. Fourteenth Int. Conf. Artif. Intell. Stat., vol. 15, pp. 315–323, 2011.

[21] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *Adv. Neural Inf. Process. Syst.*, vol, vol. 2017, pp. 972–981, 2017–Decem, pp.

[22] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *ICML Work. Deep Learn. Audio, Speech Lang. Process.*, vol. 28, 2013.

[23] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min, pp. 2623–2631, 2019.

[24] A. K. Akobeng, "Understanding diagnostic tests 3: Receiver operating characteristic curves," *Acta Paediatr.*, vol. 96, no. 5, pp. 644–647, 2007.

[25] L. Duret, A. Eyre-Walker, and N. Galtier, "A new perspective on isochore evolution," *Gene*, vol. 385, pp. 71–74, 2006.