

# A systematic evaluation of Hi-C data enhancement methods for enhancing PLAC-seq and HiChIP data

Le Huang<sup>†</sup>, Yuchen Yang<sup>†</sup>, Gang Li<sup>id</sup>, Minzhi Jiang, Jia Wen, Armen Abnoui, Jonathan D. Rosen, Ming Hu<sup>id</sup> and Yun Li<sup>id</sup>

Corresponding authors: Yun Li, Departments of Genetics, Biostatistics and Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. E-mail: [yunli@med.unc.edu](mailto:yunli@med.unc.edu); Ming Hu, Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195, USA. E-mail: [hum@ccf.org](mailto:hum@ccf.org)

<sup>†</sup>Le Huang and Yuchen Yang contributed equally to this work.

## Abstract

The three-dimensional organization of chromatin plays a critical role in gene regulation. Recently developed technologies, such as HiChIP and proximity ligation-assisted ChIP-Seq (PLAC-seq) (hereafter referred to as HP for brevity), can measure chromosome spatial organization by interrogating chromatin interactions mediated by a protein of interest. While offering cost-efficiency over genome-wide unbiased high-throughput chromosome conformation capture (Hi-C) data, HP data remain sparse at kilobase (Kb) resolution with the current sequencing depth in the order of  $10^8$  reads per sample. Deep learning models, including HiCPlus, HiCNN, HiCNN2, DeepHiC and Variationally Encoded Hi-C Loss Enhancer (VEHiCLE), have been developed to enhance the sequencing depth of Hi-C data, but their performance on HP data has not been benchmarked. Here, we performed a comprehensive evaluation of HP data sequencing depth enhancement using models developed for Hi-C data. Specifically, we analyzed various HP data, including Smc1a HiChIP data of the human lymphoblastoid cell line GM12878, H3K4me3 PLAC-seq data of four human neural cell types as well as of mouse embryonic stem cells (mESC), and mESC CCCTC-binding factor (CTCF) PLAC-seq data. Our evaluations lead to the following three findings: (i) most models developed for Hi-C data achieve reasonable performance when applied to HP data (e.g. with Pearson correlation ranging 0.76–0.95 for pairs of loci within 300 Kb), and the enhanced datasets lead to improved statistical power for detecting long-range chromatin interactions, (ii) models trained on HP data outperform those trained on Hi-C data and (iii) most models are transferable across cell types. Our results provide a general guideline for HP data enhancement using existing methods designed for Hi-C data.

**Keywords:** deep learning, evaluation, HiChIP, PLAC-seq, Hi-C, enhancement

## Introduction

Mammalian genome folds into a complex three-dimensional (3D) structure in the nucleus, facilitating cis-regulatory elements to regulate genes up to megabase away [1]. The unbiased genome-wide high-throughput chromosome conformation capture (Hi-C) technology has been widely adopted for studying chromatin spatial organization [2]. However, Hi-C usually requires billions of reads to achieve kilobase (Kb) resolution, which is cost-prohibitive [3, 4]. Most existing Hi-C data are of ~500 million or fewer raw reads, preventing subsequent Kb resolution analysis. To enhance Hi-C data, several computational methods, including HiCPlus [5], HiCNN [6], HiCNN2 [7], DeepHiC [8] and variationally encoded Hi-C loss enhancer (VEHiCLE) [9] have been recently proposed. All five methods are based on deep neural network with different architectures. Specifically, HiC-

Plus uses three layers of convolution neural networks (CNN) [10] to construct the mapping from low-depth Hi-C data to high-depth Hi-C data; HiCNN adopts a 54-layer CNN with skip connections [11]; HiCNN2 extends HiCNN and ensembles three deep learning models [7]; DeepHiC utilizes generative adversarial networks (GAN) framework [12]; and VEHiCLE pretrains a variational autoencoder [13] model and fine-tunes a GAN model. This is an active research area with multiple more recent methods developed for enhancing Hi-C data [14, 15].

In 2016, HiChIP and proximity ligation-assisted ChIP-Seq (PLAC-seq) technologies [16, 17] were proposed to measure protein-mediated chromatin interactions. While offering higher signal-to-noise ratio (SNR) and better cost-efficiency over genome-wide unbiased Hi-C data, HP data are still sparse at Kb resolution with the current sequencing depth of typically several hundred million

**Le Huang** is a PhD student in the Curriculum in Bioinformatics and Computational Biology at the University of North Carolina at Chapel Hill.

**Yuchen Yang** is an associate professor in the School of Ecology at Sun Yat-sen University.

**Gang Li** is a PhD student in the Department of Statistics and Operations Research at the University of North Carolina at Chapel Hill.

**Minzhi Jiang** is a PhD student in the Department of Applied Physical Sciences at the University of North Carolina at Chapel Hill.

**Jia Wen** is a postdoctoral researcher in the Department of Genetics at the University of North Carolina at Chapel Hill.

**Armen Abnoui** was a postdoctoral researcher in the Department of Quantitative Health Sciences at Lerner Research Institute, Cleveland Clinic Foundation, and now is a senior innovation software engineer at NovaSignal.

**Jonathan D. Rosen** is a PhD student in the Department of Biostatistics at the University of North Carolina at Chapel Hill.

**Ming Hu** is an assistant staff in the Department of Quantitative Health Sciences at Lerner Research Institute, Cleveland Clinic Foundation.

**Yun Li** is a professor in the Departments of Genetics, Biostatistics and Computer Science at the University of North Carolina at Chapel Hill.

**Received:** November 4, 2021. **Revised:** March 30, 2022. **Accepted:** March 31, 2022

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

raw reads per sample. Computationally enhancing the depth of HP data can facilitate downstream analysis, such as identification of long-range enhancer–promoter interactions, and prioritization of putative causal genes of genetic variants associated with human complex diseases and traits. No method has been developed for HP data enhancement and the aforementioned methods developed for Hi-C data [i.e. HiCPlus, HiCNN, HiCNN2, DeepHiC and VEHICLE [5–9] have not been evaluated for their performance on HP data yet.

To benchmark the performance of these methods when applied to HP data, we conducted a systematic evaluation using seven publicly available HP datasets, namely, Smc1a HiChIP data from the human lymphoblastoid cell line GM12878 [17], and H3K4me3 PLAC-seq data from five cell types including the mouse embryonic stem cells (mESC) [18] and four human fetal brain cell types [19], and mESC CCCTC-binding factor (CTCF) PLAC-seq data [18]. We focused on three aspects in our evaluation: (i) the relative performance among the assessed methods; (ii) whether training with HP data leads to improved performance than training with Hi-C data and (iii) transferability of the trained models across datasets.

## Results

### Overview of the evaluation framework

In this study, we mainly applied three existing methods (HiCNN2, HiCPlus and DeepHiC) designed for Hi-C data to enhance the sequencing depth of HP data (Figure 1). We also explored HiCNN and VEHICLE but chose not to include them for most assessments because HiCNN has a highly similar performance as HiCNN2 (Supplementary Figures S1 and S2, see Supplementary Data available online at <https://academic.oup.com/bib>); and VEHICLE's specific features tailored for Hi-C data render it suboptimal for HP data (Supplementary Figure S3, see Supplementary Data available online at <https://academic.oup.com/bib>). First, we generated low-depth HP ('baseline') datasets from mESC H3K4me3 PLAC-seq data and GM12878 Smc1a HiChIP data with different downsampling ratios (**Data preprocessing in Methods section**). We similarly generated low-depth mESC CTCF PLAC-seq data and H3K4me3 PLAC-seq data for four human brain cell types with downsampling ratio of 0.125. Next, we split each dataset into training and testing datasets with the training dataset consisting of chromosomes 1, 2, 3, 5, 7 and 9 (chromosome 2 was used as the validation data, as part of the training procedure, to select the best model); and the testing dataset containing all the other chromosomes (chromosomes 4, 6, 8, 10–19 for mESC or chromosomes 4, 6, 8, 10–22 for human cell types). Then on the training datasets, we applied each method to train models using the low-depth (i.e. baseline) input data and the high-depth (i.e. full data without downsampling) target data and subsequently applied the trained models to the low-depth testing data

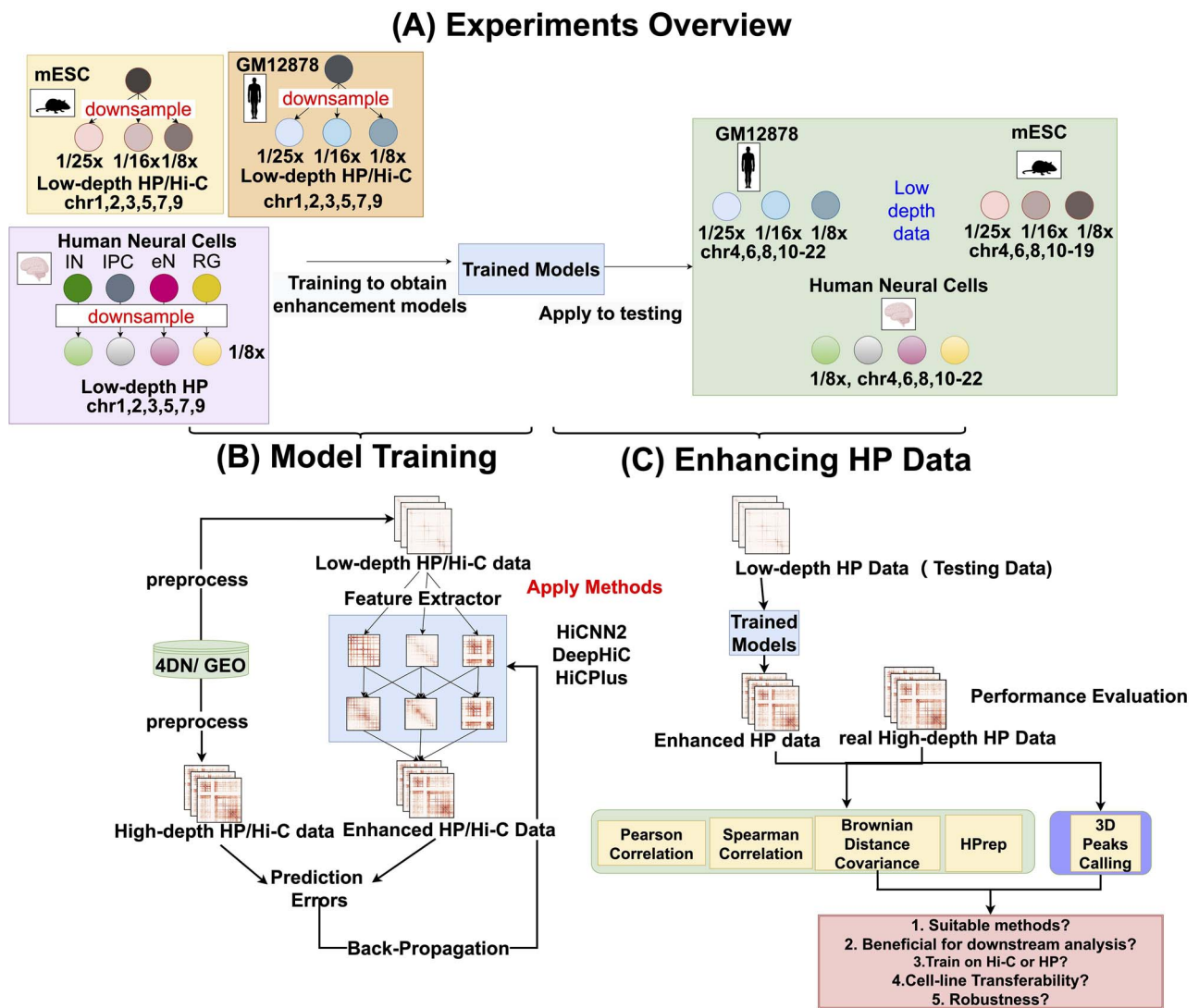
to obtain an enhanced high-depth data (Figure 1). Finally, we calculated the similarity between the enhanced HP data and the high-depth HP data (i.e. full data without downsampling for the testing chromosomes, which serves as the working truth). Specifically, we assessed similarity using four metrics: Pearson's correlation coefficient, Spearman's rank correlation coefficient, Brownian distance covariance (or distance correlation) [20] and HPRep, a new method to assess the reproducibility of HP data [21]. For presentation brevity, we have decided to only show the Pearson correlation coefficient results in the main text as the other statistics reach qualitatively the same conclusions. In addition, we performed 3D peak calling before and after enhancement to assess the impact of enhancement on the detection of chromatin interactions.

Next, we compared the performance of each method using the model trained on HP data to that trained on Hi-C data (detailed in later section **Hi-C or HP data for training**). Lastly, we evaluated the transferability of each method by enhancing HP data across different cell types detailed in the later section **Model transferability**).

Since HP data measure protein-mediated chromatin interactions, we evaluated enhancement results only for bin pairs where at least one bin contains the protein of interest. Specifically, we defined bin pairs where both bins contain the protein of interest as the 'AND' set, and bin pairs where only one bin contains the protein of interest as the 'XOR' set, following our previous work [18, 21]. We removed bin pairs where neither of the two bins contains the protein of interest from our downstream analysis, and only applied abovementioned similarity metrics (Pearson's correlation, Spearman's correlation, Brownian distance and HPRep) to bin pairs in the 'AND' and 'XOR' sets. Noticeably, in HP data, bin pairs in the 'AND' set usually show higher contact frequency than bin pairs in the 'XOR' set due to double ChIP enrichment. We thus evaluated similarity for the 'AND' and 'XOR' sets separately.

### Performance comparison of different methods

We benchmarked the performance of the three methods (HiCPlus, HiCNN2 and DeepHiC) in terms of enhancing low-depth HP data at 10 Kb resolution. Note that we decided not to include HiCNN for most evaluations because HiCNN and HiCNN2 perform highly similarly (Supplementary Figures S1 and S2, see Supplementary Data available online at <https://academic.oup.com/bib>). For HiCNN2, which is an ensemble method with three models (HiCNN2-1, HiCNN2-2 and HiCNN2-3), we present only HiCNN2-1 for the rest of the manuscript because the three HiCNN2 models perform almost indistinguishably (Supplementary Figures S1 and S2, see Supplementary Data available online at <https://academic.oup.com/bib>). Evaluations of the three methods (namely, HiCPlus, HiCNN2-1 and DeepHiC) suggest that they perform reasonably, all significantly outperforming the low-depth HP data (Figure 2, Supplementary Figures

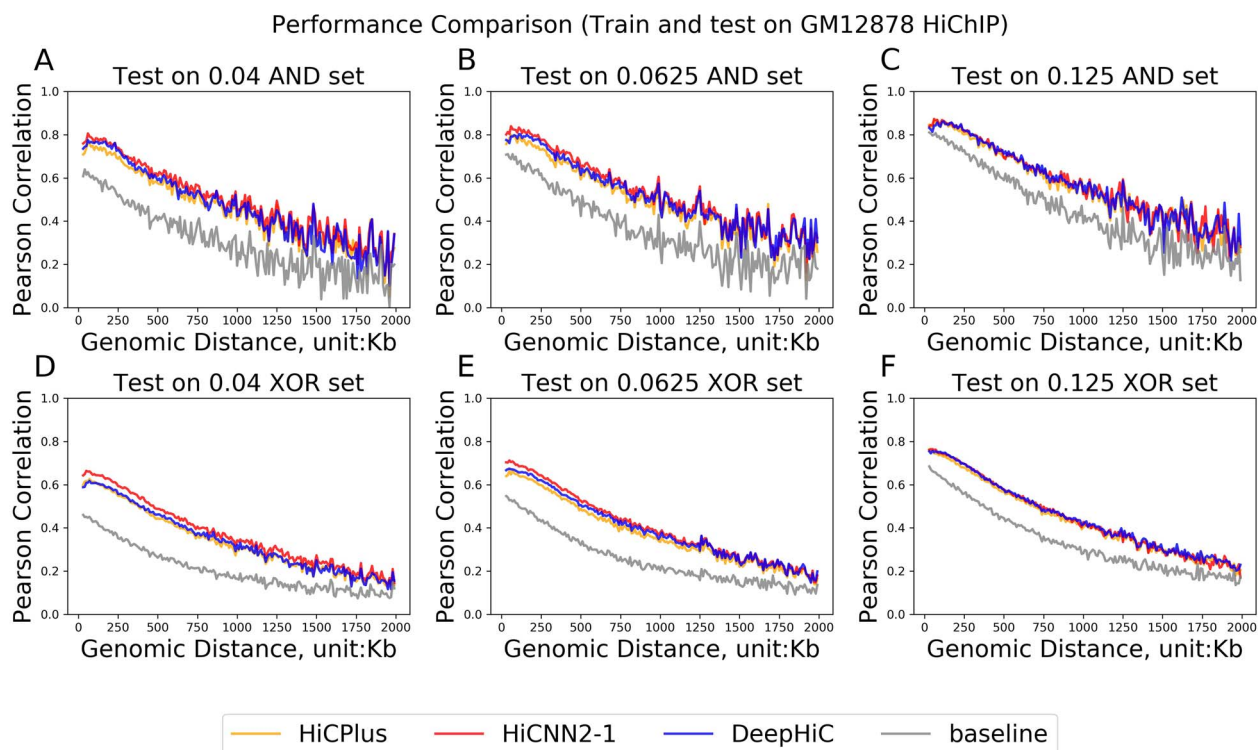


**Figure 1.** Overview of experimental design. **(A)** Experiments overview: we applied each deep learning method to Hi-C or HP dataset from GM12878 cell line, mESC and four different human neural cell types to train enhancement models. We then applied those models to enhance the testing datasets. The yellow, orange and purple blocks on the left side represent training datasets (trained on chromosome 1, 3, 5, 7, 9 and validated on chromosome 2) from mESCs, GM12878 and human brain cells, respectively. The tree structure in each block contains high-depth datasets (which are the target datasets, shown as parental nodes in darker colors) and low-depth datasets (which are the input datasets, shown as offspring nodes in lighter colors). Note that the low-depth datasets were created by downsampling from high-depth datasets (see details in Methods). On the right side, the green block represents the testing datasets. **(B)** Model training: this panel shows the overall training procedure. The deep learning models learn the features (blue block) which can enhance the sequencing depth of input dataset. The loss (prediction error) measures the difference between an estimated value and its true value, and the gradient of loss can optimize the parameters of the neural networks (see ‘The Principle of Deep Learning’ section under Methods). **(C)** Enhancing HP data. This panel shows that we first applied pre-trained models on testing datasets and then evaluated the performance of each model by comparing the enhanced datasets (prediction) with their corresponding high-depth datasets (ground truth) with four metrics (Pearson correlation coefficient, Spearman’s rank correlation coefficient, Brownian distance covariance [20] and HPrep [21]). We additionally evaluated the impact of enhancement on 3D peak calling.

S4 and S5, see Supplementary Data available online at <https://academic.oup.com/bib>). For example, when using HiCNN2-1 to enhance the GM12878 HiChIP data by 25x (i.e. from 0.04 depth to full), Pearson correlation coefficients for the ‘AND’ set are 0.70–0.81, which are 0.09–0.24 higher than the low-depth data, when the genomic distance is 20–250 Kb (Figure 2A). Similarly, when using HiCPlus to enhance the mESC PLAC-seq data by 25x, Pearson correlation coefficients for the ‘XOR’ set are 0.44–63, 0.16–0.21 higher than the low-depth data, when the genomic distance is 50–500 Kb

(Supplementary Figure S5, see Supplementary Data available online at <https://academic.oup.com/bib>). VEHICLE shows inferior performance when enhancing GM12878 HiChIP data (Supplementary Figure S3, see Supplementary Data available online at <https://academic.oup.com/bib>) possibly due to certain features tailored for Hi-C data that are no longer suitable for HP data. For instance, VEHICLE performs Knight-Ruiz (KR) normalization for Hi-C data. However, because HP technologies enrich chromatin contacts at the region mediated by the protein of interest, the equal visibility assumption made





**Figure 2.** Methods comparison when enhancing GM12878 HiChIP data. Three enhancement methods are compared: HiCPlus, HiCNN2-1 and DeepHiC. Left panel (A and D) shows performance in 0.04 downsampled data, middle panel (B and E) in 0.0625 downsampled data and right panel (C and F) in 0.125 downsampled data. Performance is quantified with the Pearson correlation coefficient (y-axis). X-axis is the genomic distance in Kb unit. Top row (A–C) shows performance among bin pairs in the AND set and the bottom row (D–F) shows performance among bin pairs in the XOR set. The gray line represents the baseline (i.e. low-depth data without any enhancement).

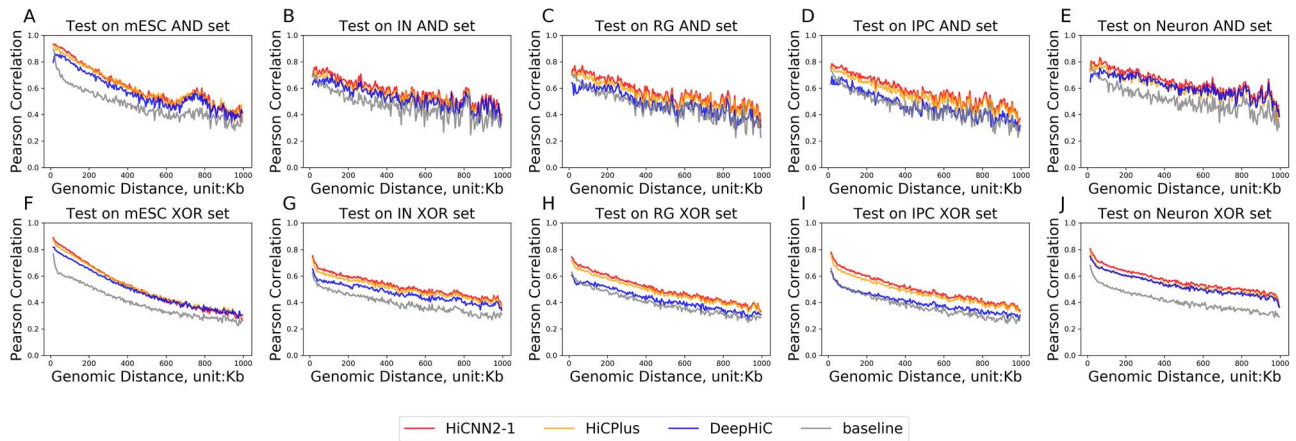
in KR normalization is invalid for HP data. In addition, VEHICLE requires all diagonal bin pairs to have nonzero counts. Therefore, we decided not to pursue further with VEHICLE enhancement.

In addition, the three methods perform similarly for most of the seven HP datasets evaluated, with Pearson correlation differences largely within a difference of 0.1 (Figures 2 and 3, Supplementary Figures S4 and S5, see Supplementary Data available online at <https://academic.oup.com/bib>). For example, when down-sampling ratio is 1/25 and the distance is 20 Kb–1.25 Mb for the GM12878 HiChIP data, HiCNN2-1 improves Pearson correlation by 0.024–0.048 and 0.012–0.059 on the ‘XOR’ set, compared with HiCPlus and DeepHiC, respectively (Figure 2D). For another example, when down-sampling ratio is 1/16 and the distance is 250 Kb–1.5 Mb for the mESC PLAC-seq data, HiCNN2-1 and HiCPlus show highly similar performance and improve Pearson correlation by 0.01–0.09 on the ‘AND’ set, compared to DeepHiC (Supplementary Figure S5B, see Supplementary Data available online at <https://academic.oup.com/bib>). When enhancing some cell types, for instance, radial glia (RG) and intermediate progenitor cells (IPCs), DeepHiC is substantially worse than HiCNN2-1 and HiCPlus with a difference of 0.2 in Pearson correlation (Figure 3C, D and H, I). The inferior performance of

DeepHiC may be due to mode collapse issues [22, 23] for GAN models (more details in the **Discussion** section).

### 3D peak calling

3D peak calling, or the detection of statistically significant long-range chromatin interactions, is one of the important downstream analyses for various types of chromatin conformation data, including HP data. To evaluate the impact of HP data enhancement, we further applied our model-based analysis of PLAC-seq and HiChIP (MAPS) pipeline [18] to identify significant chromatin interactions before and after enhancement, and compared them with chromatin interactions detected from the full data. We treated 3D peak calling results derived from the full data (without any down-sampling) as the truth. Specifically, we defined true peaks as bin pairs with MAPS false discovery rate (FDR) <1%, contacts  $\geq 12$ , and signal to noise ratio (SNR, i.e. the ratio of observed count over expected count)  $\geq 2$ ; and we defined true background bin pairs as those with MAPS FDR >10% and contacts  $\geq 12$ . We found that even high-depth input HP data (e.g. 0.5 down-sampled GM12878 HiChIP data in Figure 4 or 0.5 downsampled mESC PLAC-seq data in Supplementary Figure S6, see Supplementary Data available online at <https://academic.oup.com/bib>, where the raw sequencing depth is ~322 million and ~568



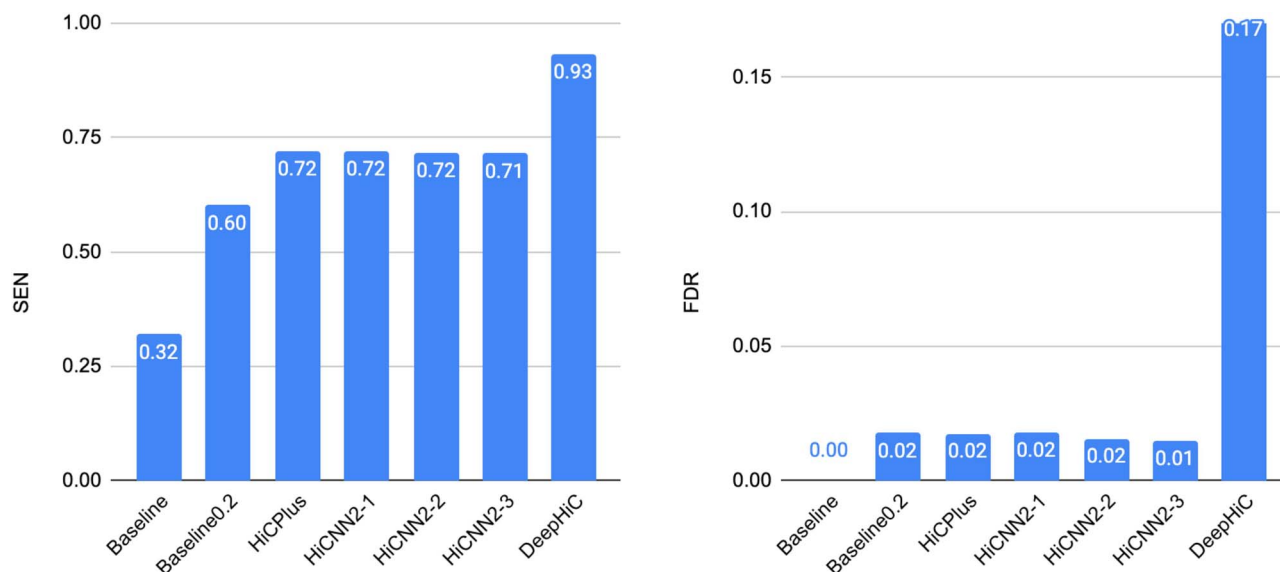
**Figure 3.** Methods comparison when enhancing HP data of various cell types. Three enhancement methods are compared: HiCPlus, HiCNN2-1 and DeepHiC. Downsampling ratio is 0.125 for all five cell types evaluated: mESC (1st column), interneurons (IN, 2nd column), RG (3rd column), IPC (4th column) and eN (5th and rightmost column). Performance is quantified with the Pearson correlation coefficient (y-axis). X-axis is genomic distance in Kb unit. Top row (A–E) shows performance among bin pairs in the AND set and the bottom row (F–J) shows performance among bin pairs in the XOR set. The gray line represents the baseline (i.e. low-depth data without any enhancement).

million, respectively) benefit from enhancement in that enhanced datasets can improve the power of 3D peak calling. For example, for 0.5 downsampled GM12878 HiChIP data (Figure 4), the baseline (i.e. downsampled data before enhancement) has a low sensitivity of 0.32, while the enhanced data (using HiCPlus or HiCNN2) improve the sensitivity to 0.71–0.72, while maintaining the desired FDR 1%. DeepHiC increases sensitivity even more drastically but fails to maintain the desired 1% FDR. Similarly, we observed clear improvement with enhanced datasets for 0.5 downsampled mESC PLAC-seq data (Supplementary Figure S6, see Supplementary Data available online at <https://academic.oup.com/bib>). Observing that the FDRs from baseline is essentially 0, we relaxed the MAPS-FDR threshold to 0.2 for the baseline, which led to an actual FDR of 0.02, comparable to that after enhancement. With the relaxed FDR threshold, the power of the baseline increased substantially, from 0.32 to 0.6 but still clearly lower than 0.71 after enhancement (Figure 4). Similar patterns were observed for 0.5 downsampled mESC data (Supplementary Figure S6, see Supplementary Data available online at <https://academic.oup.com/bib>).

We additionally examined SNR (again defined as the ratio of observed count over expected count) before and after enhancement, both compared to SNRs from the full data without enhancement. We found that SNR estimates from baseline data without enhancement are significantly lower than those from full data (Supplementary Figure S7B, see Supplementary Data available online at <https://academic.oup.com/bib>). Treating the estimates from full data as the working truth, these results indicate that baseline data tend to underestimate the magnitude of 3D peaks. Data enhancement mitigates the underestimation issue, with enhanced data producing SNR estimates more closely approaching the working truth (Supplementary Figure S7C, see Supplementary Data available online at <https://academic.oup.com/bib>).

Although we observed a significant difference in SNR estimates both at 3D peaks (Supplementary Figure S7E, see Supplementary Data available online at <https://academic.oup.com/bib>) and at background bin pairs (Supplementary Figure S7F, see Supplementary Data available online at <https://academic.oup.com/bib>), we noticed that the absolute difference is more pronounced among 3D peaks. Specifically, mean and median SNR at 3D peaks are 4.43 and 4.07 at baseline, 4.90 and 4.37 after HiCNN2-1 enhancement and 4.99 and 4.38 when using the full data (Supplementary Figure S7E, see Supplementary Data available online at <https://academic.oup.com/bib>). In contrast, the mean and median SNR at background bin pairs are 0.97 and 0.94, 0.98 and 0.97, and 1.01 and 0.99, respectively, with only  $\leq 0.05$  absolute difference. The statistical significance at background bin pairs is driven primarily by the huge number of background bin pairs (Supplementary Figure S7F, see Supplementary Data available online at <https://academic.oup.com/bib>).

Encouraged by the power improvement in 3D peak calling genome-wide, we proceeded to examine two specific loci in mESCs where previous studies [24, 25] have established enhancer–promoter interactions. These two loci are *Med13l* and *Mtnr1a* loci (Figure 5). From Figure 5, we observe that baseline without enhancement fails to identify many 3D peaks, including the most important EPIs. After HP data enhancement, we were able to rescue some of the EPIs. For example, for the two bin pairs corresponding to EPIs at the *Med13l* locus (illustrated with black arrows), full data identified both; baseline identified only one, while every enhanced data were able to rescue the missed one (Figure 5A–E). Similarly, for the bin pair corresponding to EPI at the *Mtnr1a* locus (illustrated with black arrows), full data identified it; baseline failed to detect it, while again every enhanced data managed to rescue the signal (Figure 5G–K). Interestingly, simply multiplying the baseline matrix



**Figure 4.** 3D peak calling in 0.5 downsampled GM12878 HiChIP data. Left panel (A) shows sensitivity. Right panel (B) shows FDR. The truth (3D peaks or not) is established by peak calling via MAPS from full data without any downsampling. Specifically, true peaks are bin pairs with MAPS FDR <1%, contacts  $\geq 12$  and signal to noise ratio (SNR, i.e. the ratio of observed count over expected count)  $\geq 2$ . True background bin pairs are those with MAPS FDR >20% and contacts  $\geq 12$ . Baseline0.2 bars show the 3D peak calling performance when relaxing the MAPS-FDR threshold from 1 to 20%.

with a constant of 8 can rescue the EPIs at the *Med13l* locus (Figure 5F), suggesting that simple amplification of the contact frequency matrix may help 3D peak calling when the input data are of low depth. However, this simple strategy still fails to detect the EPI at the *Mtnr1a* locus (Figure 5L), showcasing the advantage of data enhancement.

Finally, we assessed whether the identified chromatin interactions relate to gene expression. As shown in [Supplementary Figure S8](https://academic.oup.com/bib) (see Supplementary Data available online at <https://academic.oup.com/bib>), we found that genes with promoters involving 3D peaks show significantly higher expression levels than genes whose promoters do not involve any 3D peaks. The fact that genes with 3D peaks identified from baseline data without any enhancement are expected as they tend to be the lower hanging fruits with stronger magnitude of chromatin interactions that can be detected by low-depth data.

### Hi-C or HP data for training?

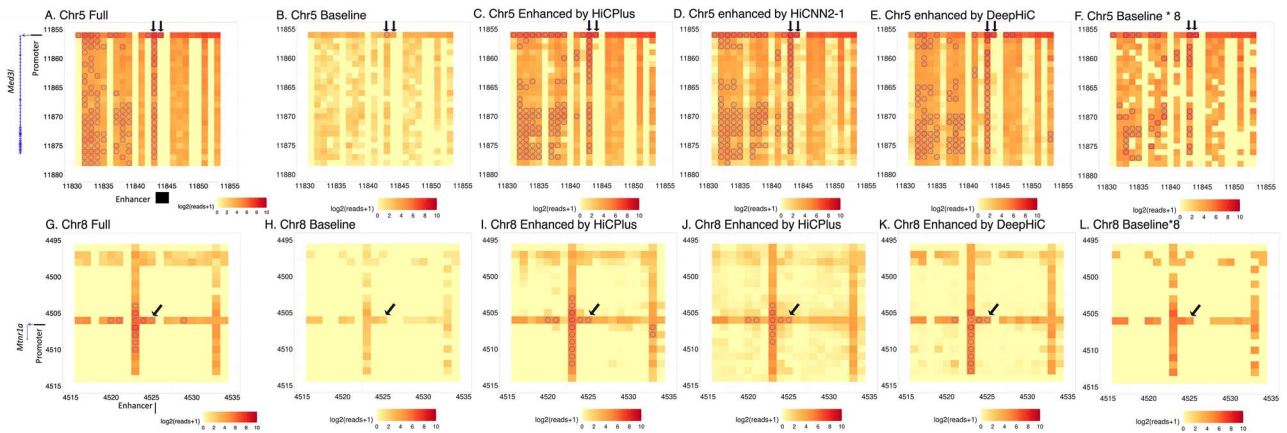
We then evaluated the robustness and relative performance of HP depth enhancement for each method when training models with different assays, specifically Hi-C or HP. For enhancing the GM12878 HiChIP data, HiCPlus, HiCNN2 and DeepHiC, all showed comparable or improved enhancement by using models trained on HP data than trained on Hi-C data (Figure 6; [Supplementary Figures S9–S11](https://academic.oup.com/bib), see Supplementary Data available online at <https://academic.oup.com/bib>). For example, when enhancing the GM12878 HiChIP data by 8X (i.e. enhancing 1/8 downsampled data to full data), Pearson correlation coefficients using the HiCPlus model trained on HP data improved by up to 0.11 for the ‘AND’ set and 0.04 for the ‘XOR’ set, compared with models trained on Hi-C data (distance: 250 Kb–2 Mb, Figure 6A and

D). Similarly, HiCNN2 and DeepHiC models trained on HP data demonstrated overall improved or comparable performance than those trained on Hi-C data, with more obvious improvement than HiCPlus. For example, within 250 Kb–2 Mb distance, HiCNN2 (Figure 6B and E) and DeepHiC (Figure 6C and F) improve Pearson correlation coefficient by 0.15 and 0.18 for the ‘AND’ set and 0.08 and 0.11 for the ‘XOR’ set, compared with the aforementioned 0.11 and 0.04 for HiCPlus (Figure 6A and D).

However, when enhancing the mESC PLAC-seq data, we observed mixed results using models trained on HP data versus those trained on Hi-C data, ([Supplementary Figures S12–S14](https://academic.oup.com/bib), see Supplementary Data available online at <https://academic.oup.com/bib>). Specifically, HiCPlus showed similar performance using HP data (light yellow) or Hi-C data (dark yellow) for training ([Supplementary Figures S12–S14 left panels](https://academic.oup.com/bib), see Supplementary Data available online at <https://academic.oup.com/bib>); HiCNN2-1’s HP trained models (light red) outperformed its Hi-C-trained models (dark red) ([Supplementary Figures S12–S14 middle panels](https://academic.oup.com/bib), see Supplementary Data available online at <https://academic.oup.com/bib>), while DeepHiC’s HP-trained models (light blue) were inferior to its Hi-C-trained models (dark blue) ([Supplementary Figures S12–S14 right panels](https://academic.oup.com/bib), see Supplementary Data available online at <https://academic.oup.com/bib>).

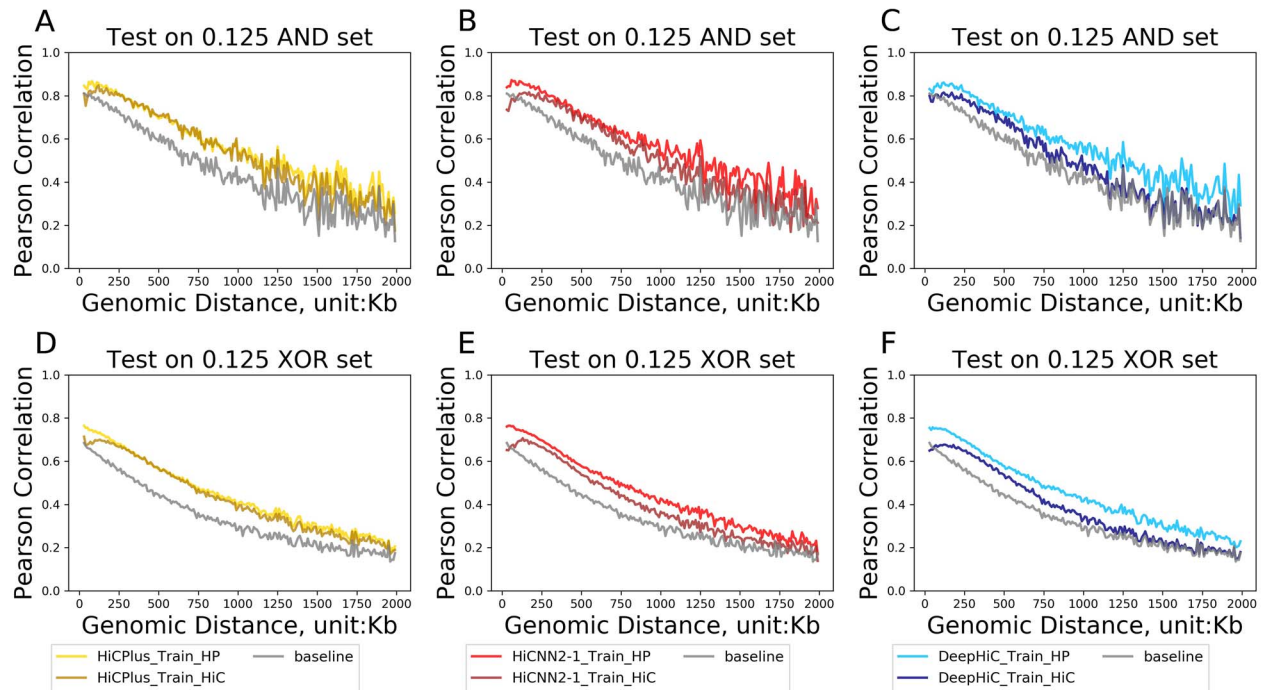
One possible explanation for DeepHiC’s better performance of mESC Hi-C-trained models is the much higher sequencing depth of mESC Hi-C data relative to mESC PLAC-seq data. Specifically, mESC Hi-C data is 4.59x that of mESC PLAC-seq data, in terms of informative reads (Table 1). Such drastic depth difference could render the models trained on Hi-C data more advantageous than those trained on HP data for all three methods ([Supplementary Figures S12–S14](https://academic.oup.com/bib), see Supplementary





**Figure 5.** 3D peak calling at *Med13l* and *Mtnr1a* loci. 3D peaking calling results from MAPS are shown. Top panel (A–F) is for the *Med13l* locus and bottom panel (G–L) is for the *Mtnr1a* locus. From left to right, we show MAPS peak calling results from the full data (without any downsampling), baseline (0.125 downsampled mESC data without enhancement), HiCPlus enhanced data, HiCNN2-1-enhanced data, DeepHiC enhanced data and baseline $\times$ 8 (by simply multiplying the baseline matrix with a constant 8). 3D peaks, bin pairs with MAPS FDR < 1%, are indicated by blue circles. For the full data (leftmost column), we further require contacts  $\geq 12$ , while for baseline and enhanced data, we relax the criterion to contacts  $\geq 2$ . The gene track is shown on the very left margin and the enhancer regions are shown at the bottom of the left panel as black rectangles. Gene and enhancer information are visualized with the help of WashU Epigenome Browser [32]. Bin pairs corresponding to the annotated enhancer–promoter regions are marked by black arrows.

#### HP vs Hi-C (Train on GM12878 HiChIP/HiC test on GM12878 HiChIP)



**Figure 6.** HiChIP-trained versus Hi-C-trained models when enhancing GM12878 HiChIP data by 8x. Performance is assessed by Pearson correlation coefficient. Each subfigure represents the performance of one of three read depth enhancement methods (HiCNN2-1, HiCPlus and DeepHiC) for a certain set (AND set or XOR set). In each subfigure, we show how Pearson correlation coefficient (y-axis) changes with genomic distance (x-axis), where the distance ranges from 20 Kb to 2 Mb with an increment of 10 Kb. The gray line represents the baseline (i.e. low-depth data without any enhancement).

Data available online at <https://academic.oup.com/bib>). Particularly in DeepHiC, we observed an obvious advantage of Hi-C-trained models over HP-trained models. To reduce the impact of the different sequencing depths, we downsampled mESC Hi-C data so that its informative reads are comparable to those in mESC PLAC-seq data (**Generating HiC\_Downsampled data in**

**Methods section**). After downsampling, we obtained 59.2 million informative reads in downsampled Hi-C data (HiC\_downsampled), matching that (also 59.2 million) in PLAC-seq data (Table 1). We then retrained DeepHiC models using the downsampled Hi-C data. With a comparable number of informative reads, models trained on the downsampled Hi-C data showed worse

**Table 1.** Read counts for HP and Hi-C datasets

	Read counts for HP and Hi-C dataset						
	<sup>a</sup> Raw reads in full data	<sup>a</sup> Informative reads <sup>b</sup> in full data	<sup>a</sup> Informative reads <sup>b</sup> in downsampled data				
Downsampling ratio	None	None	0.04	0.0625	0.125	0.25	0.5
GM12878 HiChIP <sup>a</sup>	643 644 994	28 762 260	1 149 807	1 797 116	3 593 523	7 191 222	14 187 319
mESC H3K4me3 PLAC-seq <sup>a</sup>	1 135 198 787	59 229 165	2 370 623	3 700 378	7 404 377	14 807 009	29 613 854
mESC CTCF PLAC-seq <sup>a</sup>	345 816 091	17 372 561	694 895	1 085 777	2 171 563	4 343 135	8 686 277
IN H3K4me3 PLAC-seq <sup>c</sup>	2 747 206 906	13 387 780	535 499	836 728	1 673 465	3 346 938	6 693 884
IPC H3K4me3 PLAC-seq <sup>c</sup>	1 837 960 692	15 171 775	606 862	948 226	1 896 464	3 792 937	7 585 882
eN H3K4me3 PLAC-seq <sup>c</sup>	1 740 000 000	20 547 587	821 895	1 284 213	2 568 439	5 136 889	10 273 789
RG H3K4me3 PLAC-seq <sup>c</sup>	1 487 624 144	14 180 735	567 217	886 285	1 772 582	3 545 177	7 090 363
GM12878 Hi-C	6 524 520 477	256 378 089	10 257 207	16 023 348	32 057 797	64 087 648	128 177 761
mESC Hi-C	7 260 480 082	272 146 960	10 885 391	17 002 222	34 005 844	68 045 585	136 067 650
GM12878 ratio (Hi-C/HP)	NA	8.914	8.921	8.916	8.921	8.912	9.035
mESC ratio (Hi-C/HP)	NA	4.595	4.592	4.595	4.593	4.595	4.595
mESC Hi-C_downsampled <sup>a</sup>	NA	59 228 684	2 369 794	3 700 223	7 405 280	13 369 886	26 757 578

This table contains the read count information for all datasets used in this study. Leftmost column shows the dataset names. IN, interneurons; IPC, intermediate progenitor cell; eN, excitatory neuron; RG, radial glia. <sup>a</sup>Retaining only bin pairs with genomic distance between 10 Kb and 2 Mb. <sup>b</sup>#Informative reads is the number of bin pairs after removing invalid self-ligation read pairs, short-range reads, blacklist regions, bins with mappability <0.9 and all 'NOT' pairs. <sup>c</sup>Retaining only bin pairs with genomic distance between 5 Kb and 1 Mb.

or comparable performance than those trained on HP data (Supplementary Figure S15, see Supplementary Data available online at <https://academic.oup.com/bib>). Although worse performance is expected, the magnitude of performance impairment is drastic. For example, when enhancing by 8x (middle panel) within a distance 500 Kb–1 Mb, the Pearson correlation is 0.22–0.37 with DeepHiC models trained on the downsampled Hi-C data, compared with 0.50–0.67 with models trained on HP data and 0.55–0.76 with models trained on full Hi-C data. These results suggest that the DeepHiC method is more sensitive to the sequencing depth of Hi-C data than HiCNN2 and HiCPlus.

### Model transferability

Although many HP datasets have been generated recently [16–19], deeply sequenced HP datasets are only available to limited cell types, making it infeasible to train models separately for each cell type. One potential solution is to use models pretrained on available datasets from other cell type(s).

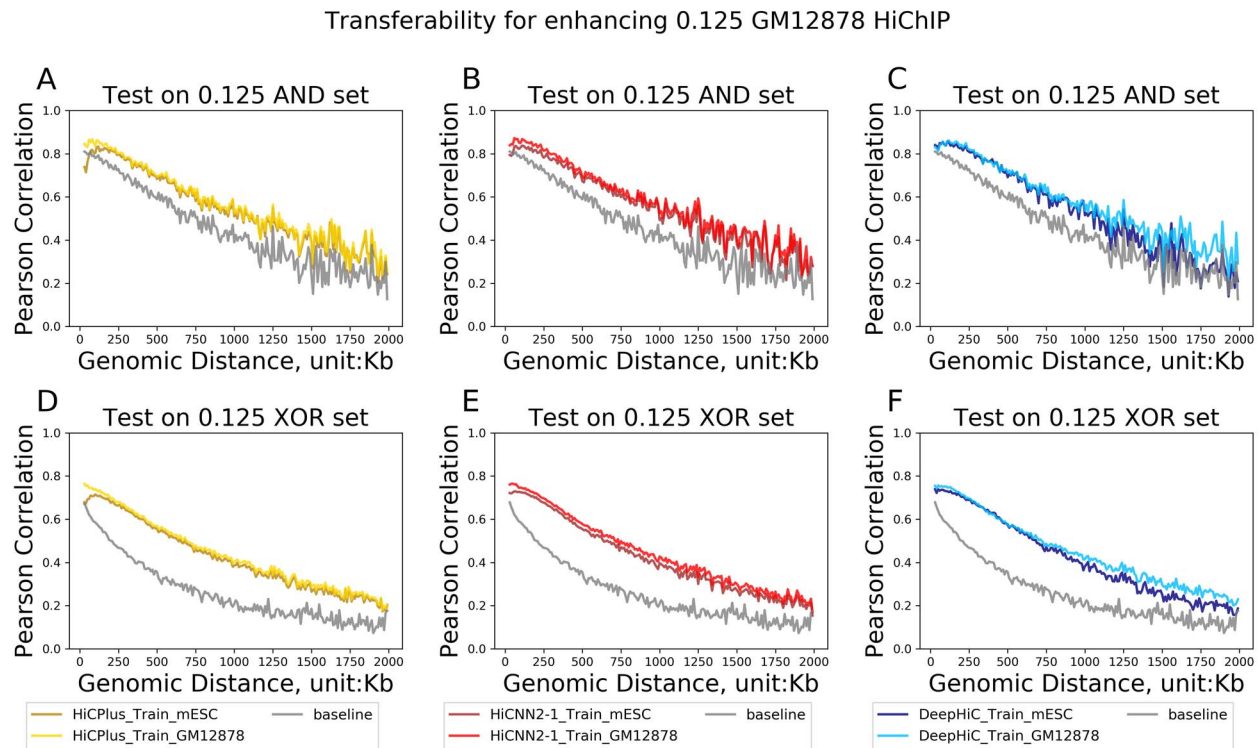
For enhancing the GM12878 HiChIP data, HiCPlus performed similarly with either GM12878-trained or mESC-trained models (the left panel of Figure 7; the left panels of Supplementary Figures S16–S18, see Supplementary Data available online at <https://academic.oup.com/bib>). Comparatively, HiCNN2-1 and DeepHiC showed slightly higher or higher accuracy using GM12878-trained models than mESC-trained models (the middle and right panels of Figure 7), with the difference more obvious for DeepHiC (the right panel of Figure 7).

For enhancing the mESC PLAC-seq data, the performance of HiCPlus, HiCNN2 and DeepHiC using models trained on the GM12878 HiChIP data was comparable to or even slightly better than using models trained on the mESC PLAC-seq data (Supplementary Figures S19–S21, see Supplementary Data available online at

<https://academic.oup.com/bib>). Specifically, for HiCPlus and HiCNN2-1, the two sets of models were nearly indistinguishable in terms of enhancing the mESC PLAC-seq data (left and middle panels of Supplementary Figures S19–S21, see Supplementary Data available online at <https://academic.oup.com/bib>). Interestingly, DeepHiC achieved even slightly better performance when using models trained on the GM12878 HiChIP data (right panel of Supplementary Figures S19–S21, see Supplementary Data available online at <https://academic.oup.com/bib>). One plausible reason is that the models for all three methods are originally developed and fine-tuned for the GM12878 Hi-C data.

Encouraged by the promising transferability results between mESC H3K4me3 PLAC-seq data and GM12878 Smc1a HiChIP data, we proceeded with transferability assessment across more cell types. Since HiCNN2-1 and HiCPlus achieved similarly best transferability performance, we presented only HiCNN2-1 results for brevity. Specifically, we enhanced 0.125 downsampled HP data for each of the six cell types [namely, GM12878 and mESC, and four human brain cell types (19)] including RG, IPCs, excitatory neurons (eN) and interneurons (IN)] using models trained from the corresponding cell type as well as using models trained from each of the other five cell types. For training and testing, we used the same chromosome splitting as illustrated in Figure 1. Results shown in Figure 8 and Supplementary Figures S22 and S23 (see Supplementary Data available online at <https://academic.oup.com/bib>) further support that the models learned are transferable across cell types. Specifically, Pearson correlation coefficients are almost indistinguishable when enhanced with models trained from the matching cell type or from other cell types (Figure 8 and Supplementary Figure S22, see Supplementary Data available online at <https://academic.oup.com/bib>), and all the models lead to similar performance in 3D peak



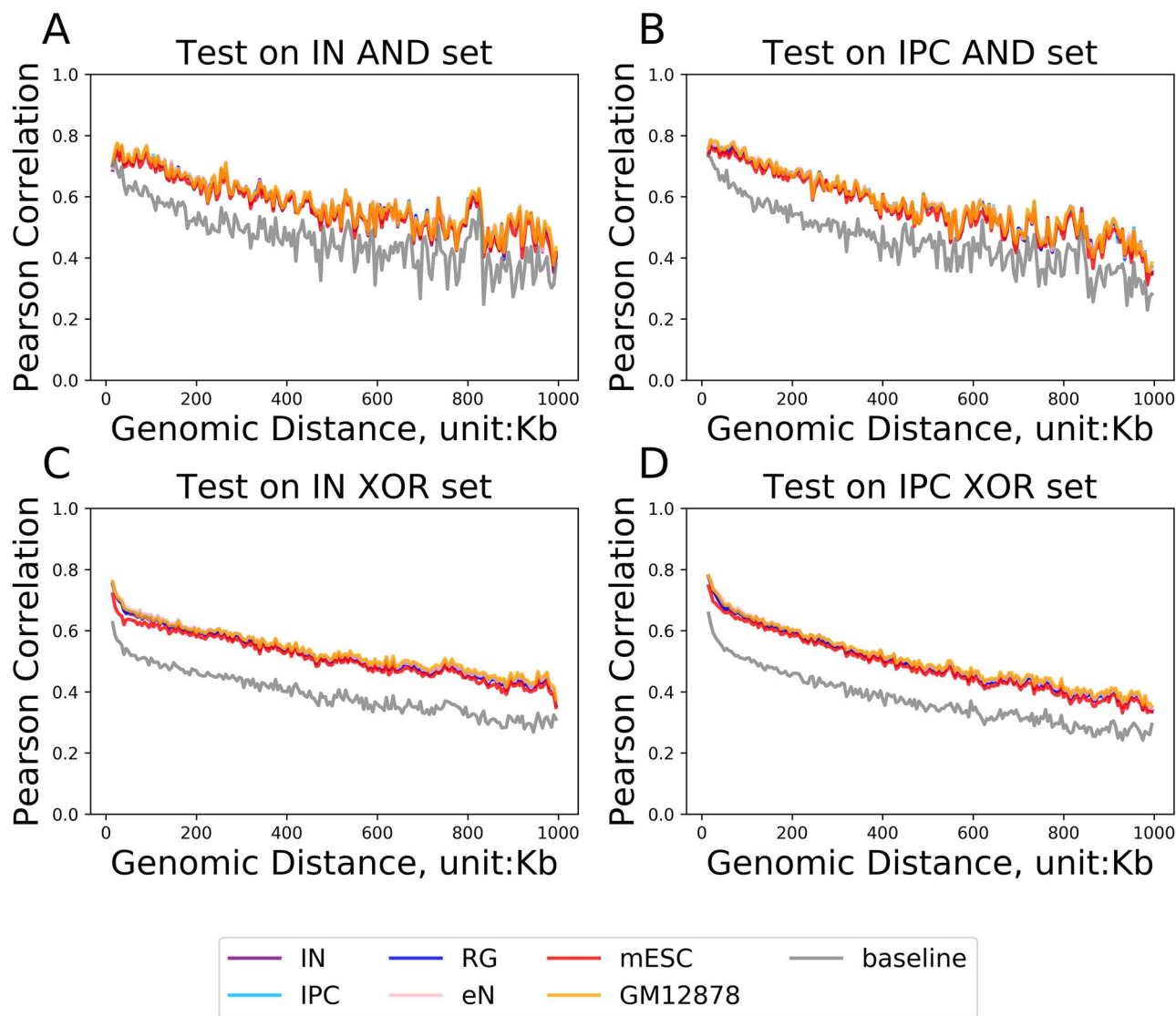


**Figure 7.** Model transferability when enhancing GM12878 HiChIP data by 8 $\times$ . Each subfigure compares two enhanced GM12878 HiChIP data: one using models trained with GM12878 HiChIP data (Train\_GM12878) and the other using models trained with mESC PLAC-seq data (Train\_mESC). The evaluation metric is Pearson correlation coefficient. Different colors in the subfigures represent different methods (yellow: HiCPlus, red: HiCPlus and blue: DeepHiC) while darker color represents models trained with mESC PLAC-seq and lighter color represents models trained with GM12878 HiChIP data. In each subfigure, we show how the evaluation metric (y-axis) changes with genomic distance (x-axis), where the distance ranges from 20 Kb to 2Mb with an increment of 10 Kb. The gray line represents the baseline (i.e. low-depth data without any enhancement).

calling (Supplementary Figure S23, see Supplementary Data available online at <https://academic.oup.com/bib>). For example, to detect chromatin interactions in IN, 0.125 downsampled PLAC-seq data (before any enhancement) had essentially no power at all (sensitivity to detect IN of IN-specific 3D peaks is 0.00, left most bars labeled 'Baseline' in Supplementary Figure S23C and G, see Supplementary Data available online at <https://academic.oup.com/bib>); in contrast, enhanced RG data using models trained with IN data resulted in a sensitivity of 0.49 (or 0.48) for IN (or IN-specific) 3D peaks (magenta bars in Supplementary Figure S23C and G, see Supplementary Data available online at <https://academic.oup.com/bib>); similarly and importantly, enhanced IN data using models trained with data in any of the other five cell types resulted in a comparable sensitivity of 0.47–0.59 (or 0.43–0.63) for IN (or IN-specific) 3D peaks (blue bars in Supplementary Figure S23C and G, see Supplementary Data available online at <https://academic.oup.com/bib>).

We further evaluated transferability in terms of capturing cell type-specific features, examining gene expression and open chromatin status in the corresponding cell types. Specifically, we compared the distribution of gene expressions for three groups of genes: (i) genes with 3D peak(s) looping to their promoters identified at 'baseline' (low-depth data without enhancement); (ii) genes without any 3D peaks at 'baseline' but with 3D

peak(s) after enhancement, separately for enhanced data using models trained with each of the six cell types and (iii) genes without any 3D peaks even with the full data ('background'). Not surprisingly, as shown in Supplementary Figure S24A–D (see Supplementary Data available online at <https://academic.oup.com/bib>), 'baseline' identified only few lower hanging fruits, and thus expression levels are the highest; genes with 3D peaks identified only after enhancement [whether using models trained with the matching cell type (yellow boxplots) or different cell types (nonbaseline and nonbackground cyan boxplots)], reassuringly, had only slightly lower expression levels, drastically higher than those 'background' genes. Similar patterns are observed when restricting only to cell type-specifically expressed genes (Supplementary Figure S24E–H, see Supplementary Data available online at <https://academic.oup.com/bib>). Following similar logic, we assessed 3D peaks in terms of their overlap with cell type-specific assay for transposase-accessible chromatin using sequencing (ATAC-seq) peaks, observing similar patterns (Supplementary Figure S24I–L, see Supplementary Data available online at <https://academic.oup.com/bib>). In particular, models trained with matching cell types (yellow bars) resulted in a similar proportion of overlap with cell type-specific ATAC-seq peaks as those trained with different cell types (nonbaseline and non-background cyan bars), suggesting that models trained



**Figure 8.** Model transferability when enhancing two neural cell types. All results are from HiCNN2-1 models. We test (i.e. perform enhancement) on two cell types: IN (left subfigures **A** and **C**) and IPC (right subfigures **B** and **D**), with a downsampling ratio of 0.125. The enhancement models are trained using HP data from each of the following six cell types: IN, IPC, RG, eN, mESC or GM12878. The gray line represents the baseline (i.e. low-depth data without any enhancement).

from unmatching cell types can similarly retain cell type-specific features.

Finally, we explored model transferability across different proteins of interest by cross-applying models learned from H3K4me3 and CTCF PLAC-seq data in mESC. We used the same mESC CTCF PLAC-seq data as in Juric *et al.* [18]. We observed almost indistinguishable performance when using models trained with the same protein of interest or the other protein (Supplementary Figure S25, see Supplementary Data available online at <https://academic.oup.com/bib>), both visibly better than without enhancement (i.e. baseline). These results suggest that the enhancement models learned are likely transferable also across different proteins of interest, with the caveat that our assessments only involved three different proteins: Smc1a above for GM12878, CTCF here for mESC and H3K4me3. In the future, more high-depth HP data with different proteins of interest will allow us to

perform a more comprehensive assessment across various proteins.

### Model robustness

Throughout the manuscript so far, we have used chromosomes chr1, 3, 5, 7 and 9 as training; chromosome 2 as the validation (part of the training procedure to select the best model); and the remaining chromosomes as testing. In addition, when creating low-depth input data, we performed downsampling only once. We evaluated model robustness by swapping training and testing, by using leaving-one-chromosome-out and by performing downsampling five times. Results presented in Supplementary Figures S26 and S27 (see Supplementary Data available online at <https://academic.oup.com/bib>) show that the models trained are robust, resulting in highly similar Pearson correlation coefficient decay profiles.

## Discussion

While several computational methods have been developed for enhancing the depth of Hi-C data, tools tailored for HP data depth enhancement are still lacking. In this study, we evaluated three methods (HiCPlus, HiCNN2 and DeepHiC) developed for Hi-C data, when applying them to enhance HP data. Our results showed that all three methods performed similarly on enhancing HP datasets when training on the HP data from the same cell type, with HiCNN2 and HiCPlus outperforming DeepHiC in most scenarios. We further assessed the robustness of enhancement when models were trained with Hi-C or HP data from the same cell type. We found that enhancement using models trained on ultra-high-depth Hi-C data achieved similar or even better performance than using models trained on HP data. However, when the sequencing depth of Hi-C data and HP data used for training were comparable, models trained on HP data exhibited better performance than those trained on Hi-C data. These results suggest that users can train models with high-depth Hi-C data for HP data enhancement if similar high-depth HP data are not available for training. We note that the terminology ‘Hi-C data *resolution* enhancement’ prevails in the literature. We have, however, decided to use ‘data *depth* enhancement’ to avoid ambiguity since *resolution* is also commonly used to indicate the bin size of the analysis unit.

Transferability across datasets (e.g. cell types, proteins of interest) is important because in practice there are limited cell types sequenced with HiChIP or PLAC-seq techniques. Our analysis across six cell types, three proteins of interest and two organisms, showed promising transferability results for enhancing HP data, consistent with the existing literatures [5–7] for enhancing Hi-C data. For example, models trained using high-depth GM12878 data can lead to better enhancement results in mESC than models trained with mESC data. More evaluations are needed in the future to draw stronger conclusions. Such evaluations will become possible when more high-depth HP data are generated both for training better models and for evaluations. Note that the actual meaningful information, specifically where the nonzero or zero contacts reside or where the chromatin interactions locate, differs across cell types, organisms and proteins of interest in HP dataset. The observed promising transferability results suggest that the rules learned to enhance lower depth data to higher depth are shared across cell types (even across organisms). Together, results presented under sections **Hi-C or HP data for training?** and **Model transferability** suggest that HiCNN2 or HiCPlus models *pretrained* from high-depth Hi-C or HP data can be directly applied to enhance HP from various cell types.

For performance evaluation, we used three standard metrics, Pearson correlation coefficients, Spearman correlation coefficients and Brownian distance covariance [20]. Brownian distance covariance is a multivariate

dependence coefficient which measures dependency of two random vectors of arbitrary and not necessarily equal dimensions, providing more general quantification of independence than linear correlation by Pearson correlation [20]. In addition, similarity metrics tailored for Hi-C data, such as HiCRep [26] and HiC-spector [27] have been widely used. We have recently extended HiCRep to HPRep, tailored for HP data after adjusting for ChIP enrichment biases [21]. Applying HPRep to evaluate the similarity between enhanced data and full data led to findings consistent with what was revealed by Pearson correlation: for example, the three deep learning methods behave better than baseline and they all have similar performance.

Overall, all three methods evaluated are able to generate enhanced data exceeding the baseline (i.e. low-depth data without enhancement), both in terms of enhancing the contact frequency matrix (as quantified by the correlation metrics) and probably more importantly in terms of improving power to detect chromatin interactions. Among the three, we recommend HiCNN2 and HiCPlus, both consistently exhibiting similar performance, superior to DeepHiC and VEHICLE, when applied to enhance HP data. Note that DeepHiC and VEHICLE, both employ the GAN model, which has been known to suffer from mode collapse problem [22, 23]. Due to the nature of HP data, multimodal distribution is expected because of the systematic difference between AND and XOR bin pairs, which might explain why DeepHiC and VEHICLE perform suboptimally in HP data enhancement. Not surprisingly, with increased downsampling ratio, enhanced data from very shallow depth data showed more pronounced improvement over the baseline. When the sequencing depth is high, there is less room for improvement, particularly when using methods developed for Hi-C data that do not consider ChIP enrichment bias of HP data. For example, when we enhanced HP data on higher depth data (e.g. 1/4 and 1/2 downsampled data), we found that ‘enhanced’ data from all three methods are comparable or even slightly worse than the baseline when measured by correlation metrics, while theoretically, enhancement methods can still improve 1/4 and 1/2 data. In addition, we observed HP data enhanced using these methods show lower correlation than Hi-C data enhanced by these methods. For example, the Pearson correlation coefficients are in the range of 0.95–0.96 within 500 Kb for HiCNN and HiCPlus on GM12878 1/8 ratio on chromosomes 6 and 12 (Figure S3 in HiCNN paper [6]) but enhanced HP data in our evaluations showed Pearson correlation <0.81. Furthermore, the improvement in HP data (as reflected by the correlation decay with distance figures) is not as smooth as in Hi-C data, which might be caused by unbalanced read distribution due to protein immunoprecipitation in HP data. Therefore, methods developed for Hi-C data are not optimal for HP data. Development of methods tailored to HP data is warranted.



## Methods

### Data preprocessing

All our assessed deep learning methods require training data, testing data and validation data as input. In our study, for the mESC PLAC-seq data, we assigned chromosomes 1, 3, 5, 7 and 9 as the training data, chromosome 2 as the validation data and chromosomes 4, 6, 8, 10–19 as the testing data. For the GM12878 HiChIP data, we assigned chromosomes 1, 3, 5, 7 and 9 as the training data, chromosome 2 as the validation data and chromosomes 4, 6, 8, 10–22 as the testing data. Here, validation data was used to select the best model (details in The principle of deep learning in Methods section). Both mESC PLAC-seq data and GM12878 HiChIP data consist of two parts: high-depth data and low-depth data, referring to the original/full HP data without downsampling and downsampled data, respectively. We used the low-depth data as the input for each deep learning method, and the high-depth data to calculate the loss function (details in The principle of deep learning in Methods section).

Specifically, we applied the following steps to generate low-depth data:

- (i) Converting read pairs into bin pairs. We followed our previous study [18, 21] to preprocess the HP data to retain only long-range read pairs (intrachromosomal contacts >1 Kb). We then randomly selected a subset of the read pairs with downsampling ratios 1/25, 1/16 or 1/8. Downsampling was implemented using command 'SAMtools views ratio' [28]. Next, we binned the original read pairs and downsampled read pairs into 10 Kb bin pairs, resulting in high-depth data and low-depth data, respectively.
- (ii) Converting high-depth and low-depth bin pairs into contact matrices. For each chromosome, based on whether the bins containing the protein of interest [H3K4me3 ChIP-seq peaks for mESC PLAC-seq data (18) and Smc1a ChIP-seq peaks for GM12878 HiChIP data (17)], we further grouped bin pairs into three categories: the 'AND' set (bin pairs where both bins contain the protein of interest), the 'XOR' set (bin pairs where only one bin contains the protein of interest) and the 'NOT' set (bin pairs where neither bins contains the protein of interest). Since HP technologies measure protein-mediated long-range chromatin interactions, we only focused on the 'AND' and 'XOR' sets for downstream analysis. In addition, we filtered out bin pairs with either end overlapping with the encyclopedia of DNA elements (ENCODE) blacklist regions [29] or with low mappability (mappability score <0.9) [30]. After filtering, we created a 10 Kb bin resolution contact matrix for each chromosome. In this work, we only used 10 Kb bin pairs with 1D genomic distance less than 2Mb in our analysis.
- (iii) Converting contact matrices into training data, testing data and validation data. According to the required format of each deep learning method,

we split the contact matrix for each chromosome into multiple submatrices. Different deep learning methods adopt different splitting strategies as their default configuration. DeepHiC splits the high-depth and low-depth contact matrices into nonoverlapping  $40 \times 40$  submatrices with stride size  $40 \times 40$ . In contrast, HiCNN, HiCNN2 and HiCPlus partition the low-depth data with overlapping  $40 \times 40$  submatrices with a stride size  $34 \times 34$  (the overlapping region between two consecutive submatrices is  $6 \times 40$ ). Next, HiCNN, HiCNN2 and HiCPlus partition the high-depth data into nonoverlapping  $28 \times 28$  submatrices with stride size  $28 \times 28$ . The overlapping submatrices split by HiCNN, HiCNN2 and HiCPlus imply that all inferred regions (i.e. the  $34 \times 34$  core regions) have flanking information. We applied each method with its default matrix splitting strategy. With those submatrices, we constructed three types of tensors: for training data, testing data and validation data, respectively. Here, training data is the tensor concatenating data from five chromosomes (1, 3, 5, 7 and 9), validation data is a tensor of chromosome 2, and testing data contains data from chromosomes 4, 6, 8, 10–19 or chromosomes 4, 6, 8, 10–22, for the mESC PLAC-seq data or the GM12878 HiChIP data, respectively.

### The principle of deep learning

All three deep learning methods (HiCNN2, HiCPlus and DeepHiC) evaluated in this study are supervised learning algorithms, which can be formulated as the following:

$$y = f_{\theta}(x), \quad (1)$$

where  $x$  represents the low-depth data in the training dataset (see Data preprocessing),  $y$  represents the enhanced data and  $\theta$  represents the parameters of neural network  $f(\bullet)$ , which approximates the mapping  $f : x \rightarrow y$  by learning from the training dataset. Each parameter  $\theta_i$  ( $i = 1, 2, \dots, n$ ) (with  $n$  being the total number of parameters of the neural network) can be optimized by the gradient descent algorithm (e.g. stochastic gradient descent or Adam [31]) in Equation (2)

$$\theta_i = \theta_i - r \frac{\partial}{\partial \theta_i} J(\theta), \quad (2)$$

where  $r$  is the learning rate, which controls the step size of gradient descending,  $J(\theta)$  is the predefined loss function and  $\frac{\partial}{\partial \theta_i} J(\theta)$  is the gradient of  $\theta_i$  which is calculated by backpropagation algorithm [10]. In HiCNN2 and HiCPlus,  $J(\theta)$  is the mean squared error as specified in Equation (3)

$$J(\theta) = \frac{1}{m} \sum_{j=1}^m (f_{\theta}(x_j) - y_j)^2, \quad (3)$$

where  $m$  is the sample size and it is the product of batch size (hyperparameter), the width of the submatrix

$N$  and the height of the submatrix  $N$ ;  $y = \{y_1, y_2, \dots, y_m\}$  is a batch of target data [high-depth submatrices, Data preprocessing (3) in Methods section],  $y_j \in \mathbb{R}^{N \times N}$  is the high-depth matrix;  $x_j \in \mathbb{R}^{N \times N}$  represents the low-depth matrix;  $f_\theta(x_j) \in \mathbb{R}^{N \times N}$  is the enhanced matrix;  $f_\theta(\bullet)$  is the neural network which represents the mapping  $f$  of  $x$  to  $y$  ( $f: x \rightarrow y$ ).

In HiCNN2 or HiCPlus, the training loss (represented by Equation 3) is optimized by the gradient descent algorithm (equation 2) iteratively. Each method trains the network using multiple epochs, with the default being 500 and 40 000 for HiCNN2 and HiCPlus, respectively. One epoch involves passing all batches completely through the neural network. In each epoch, HiCNN2 or HiCPlus uses validation loss to evaluate whether to retain the current trained model or not. Specifically, the algorithm calculates the validation loss between full data (viewed as the target data) and the enhanced data using Equation 3 and updates to the current model only when the validation loss decreases.

### Generating HiC\_downsampled data

In addition, we conducted additional experiments for evaluation of transferability, where models were trained on the mESC Hi-C data with comparable sequencing depth as the mESC PLAC-seq data, which we referred to as HiC\_downsampled data. We generated the HiC\_downsampled data by downsampling read counts within 2Mb genomic distance of mESC Hi-C data to 59.2 million, matching the total number of reads (59.2 million) in the 'AND' and 'XOR' sets of the corresponding mESC PLAC-seq data (Table 1).

### Data Availability

We downloaded GM12878 Smc1a HiChIP dataset [15], H3K4me3 PLAC-seq dataset in mESCs [16] (GSE119663), H3K4me3 PLAC-seq datasets for four human brain cell types [19], mESC CTCF PLAC-seq data [18], GM12878 Hi-C dataset [4] and mESC Hi-C dataset [3]. We also obtained ChIP-seq peaks for different cell lines (GM12878 Smc1a ChIP-seq peaks: <https://www.encodeproject.org/files/ENCFF686FLD/>, mESC H3K4me3 ChIP-seq peaks: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3380558>) as well as RNA-seq and ATAC-seq data for the human brain cell types from [19].

#### Key Points

- We focused on computationally enhancing the sequencing depth of data derived from HiChIP and PLAC-seq experiments.
- We evaluated three deep learning-based sequencing depth enhancement methods developed for Hi-C data.
- We provided practical guidelines of method of choice, type of training data to use and transferability across cell lines.

### Supplementary Data

Supplementary data are available online at <https://academic.oup.com/bib>.

### Acknowledgements

We thank Dr F.Y. and J.W. for providing advice and assistance on HiCPlus. We also thank the developers of HiCNN (Drs Z.W. and T.L.) for their help on HiCNN. We are also grateful to DeepHiC authors for answering our questions on GitHub, to VEHICLe developers (Drs J.C., M.H.) for advice to run the program on HP data and to Haidong Yi for debugging the tools. Finally, we thank Li lab members for providing feedback on the earlier version of the manuscript.

### Funding

National Institutes of Health (R01GM105785 and U01DA052713 to Y.L., UM1HG011585 and R35HG011922 to M.H.).

### References

1. Li Y, Hu M, Shen Y. Gene regulation in the 3D genome. *Hum Mol Genet* 2018;**27**:R228–33.
2. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, NY)* 2009;**326**:289–93.
3. Bonev B, Mendelson Cohen N, Szabo Q, et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* 2017;**171**:557–572.e24.
4. Rao SSP, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;**159**:1665–80.
5. Zhang Y, An L, Xu J, et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat Commun* 2018;**9**:750.
6. Liu T, Wang Z. HiCNN: a very deep convolutional neural network to better enhance the resolution of Hi-C data. *Bioinformatics* 2019;**35**:4222–8.
7. Liu T, Wang Z. HiCNN2: enhancing the resolution of Hi-C data using an ensemble of convolutional neural networks. *Genes* 2019;**10**:862.
8. Hong H, Jiang S, Li H, et al. DeepHiC: a generative adversarial network for enhancing Hi-C data resolution. *PLoS Comput Biol* 2020;**16**:e1007287–7.
9. Highsmith M, Cheng J. Vehicle: a variationally encoded hi-c loss enhancement algorithm for improving and generating hi-c data. *Sci Rep* 2021;**11**:1–13.
10. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press, Cambridge, MA, 2016.
11. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, New York, US, 2016. pp. 770–8.
12. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst* 2014;**27**:2672–80.
13. Kingma DP, Welling M. Auto-encoding variational bayes. *2nd International Conference on Learning Representations (ICLR)*. 2014.

14. Hu Y, Ma W. EnHiC: learning fine-resolution Hi-C contact maps using a generative adversarial framework. *Bioinformatics* 2021;**37**:i272–9.
15. Liu Q, Lv H, Jiang R. hicGAN infers super resolution Hi-C data with generative adversarial networks. *Bioinformatics* 2019;**35**: i99–i107.
16. Fang R, Yu M, Li G, et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* 2016;**26**:1345–8.
17. Mumbach MR, Rubin AJ, Flynn RA, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 2016;**13**:919–22.
18. Juric I, Yu M, Abnoui A, et al. MAPS: model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLoS Comput Biol* 2019;**15**:e1006982.
19. Song M, Pebworth M-P, Yang X, et al. Cell-type-specific 3D epigenomes in the developing human cortex. *Nature* 2020;**587**:644–9.
20. Székely GJ, Rizzo ML. Brownian distance covariance. *Ann Appl Stat* 2009;**3**:1236–65.
21. Rosen JD, Yang Y, Abnoui A, et al. HPrep: quantifying reproducibility in HiChIP and PLAC-seq datasets. *Curr Issues Mol Biol* 2021;**43**:1156–70.
22. Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans. *Adv Neural Inf Process Syst* 2016;**29**:2226–34.
23. Srivastava A, Valkov L, Russell C, et al. Veegan: reducing mode collapse in gans using implicit variational learning. *Adv Neural Inf Process Syst* 2017;**30**:3308–18.
24. Schoenfelder S, Furlan-Magaril M, Mifsud B, et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* 2015;**25**: 582–97.
25. Moorthy SD, Davidson S, Shchuka VM, et al. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res* 2017;**27**:246–58.
26. Yang T, Zhang F, Yardımcı GG, et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* 2017;**27**:1939–49.
27. Yan K-K, Yardımcı GG, Yan C, et al. HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps. *Bioinformatics* 2017;**33**:2199–201.
28. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;**25**: 2078–9.
29. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep* 2019;**9**:9354.
30. Hu M, Deng K, Selvaraj S, et al. HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics* 2012;**28**:3131–3.
31. Kingma DP, Ba J. Adam: a method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*. 2015.
32. Zhou X, Lowdon RF, Li D, et al. Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat Methods* 2013;**10**:375–6.