Supplementary Material

To provide a more comprehensive understanding of our PC-Loop framework and the experiments conducted, we present this supplementary material that outlines additional experimental details, extended analyses, and provides access to the datasets and code.

Supplementary Methods

1. Feature Extraction Regions

In the main manuscript, we detailed the seven regions for PC-Loop-epigenomics and the four regions for PC-Loop-sequence. Here, we provide a more in-depth description of each region, including their precise boundaries and the rationale for their selection.For instance, loop start, loop end, and flanking regions were defined based on the location of the CTCF binding motifs, as they play a crucial role in chromatin loop formation. We used ENCODE ChIP-seq datasets to identify CTCF occupancy sites in different cell lines, which supported the use of these features in model training.

2. LightGBM Hyperparameter Tuning

To determine the optimal parameters for LightGBM, we performed an extensive hyperparameter tuning process using grid search for learning rate, tree depth, and number of leaf nodes. This supplementary section includes the evaluation ranges for parameters L (learning rate), D (maximum depth), and N (number of leaf nodes), and their impact on model performance.

3. Feature Importance Analysis

In the main manuscript, we ranked the feature importance to highlight the contributions of various features. Based on the Gini impurity scores, we demonstrated that features related to histone modifications (e.g., H3K36me3 and H2AFZ) were significant across multiple datasets, supporting our hypothesis on the role of the epigenome in predicting chromatin loop interactions.

4. Details on Negative Loop Pair Generation

Negative loop samples were generated by randomly pairing CTCF binding sites that were not involved in ChIA-PET interactions on the same chromosome. We ensured that these random pairings had a similar distance distribution to the positive loop samples, preventing potential

biases during model training.

5. **Performance Metrics for Different Cell Lines**

Extended performance evaluation of PC-Loop-epigenomics, PC-Loop-sequence, and PC-Loop-integrate on additional cell lines not covered in the main manuscript is provided. Supplementary Table S1 details the F1 score, AUC, precision, and recall for each dataset, including independent test data from IMR90 and HUVEC cell lines. These results further validate the robustness of PC-Loop in accurately predicting CTCF-mediated chromatin loops.

6. **Supplementary Code and Dataset Access**

Scripts used for data preprocessing, feature extraction, and training the LightGBM model are available at [repository link]. The repository also includes pre-trained models for GM12878, K562, HeLa-S3, and MCF7 cell lines, along with a README file containing detailed usage instructions.Additionally, the processed datasets that form the basis of our model, including positive and negative loop pairs for each cell line, are provided to facilitate reproducibility and further analysis by the community.

Supplementary Tables and Figures

Table S1: Performance Evaluation of Lollipop, PC-Loop-sequence, PC-Loop-epigenomics, and PC-Loop-integrate on Three Cell Lines

Cell Line	Method	F1 score	AUC	Precision	Recall
K562	PCLoop-sequence	0.58	0.89	0.72	0.49
K562	PCLoop-epigenomics	0.86	0.99	0.86	0.86

K562	PCLoop-integrate	0.91	0.99	0.90	0.92
K562	Lollipop	0.80	0.98	0.85	0.76
GM12878	PCLoop-sequence	0.66	0.92	0.73	0.60
GM12878	PCLoop-epigenomics	0.80	0.98	0.83	0.78
GM12878	PCLoop-integrate	0.87	0.99	0.89	0.86
GM12878	Lollipop	0.76	0.97	0.84	0.70
HeLa-S3	PCLoop-sequence	0.63	0.92	0.72	0.55
HeLa-S3	PCLoop-epigenomics	0.83	0.98	0.86	0.81
HeLa-S3	PCLoop-integrate	0.89	0.99	0.89	0.89
HeLa-S3	Lollipop	0.79	0.98	0.87	0.73

Acknowledgments

We sincerely thank the Advanced Computing Center of Yunnan University for providing computational resources. All data used in this study were obtained from publicly available sources, such as the ENCODE project, to ensure transparency and reproducibility.