LoopHunter: Enhancing Chromatin Loop Annotation by Focusing on Larger Regions in Hi-C Data

*Yusen Hou, Yanlin Zhang**

Hong Kong University of Science and Technology (Guangzhou)

INTRODUCTIONS

Chromatin loops, which bring distant loci into close contact, play a crucial role in gene expression and regulation (Fig. 1). Although several methods have been developed for annotating loops from Hi-C contact maps, these methods remain unsatisfactory, particularly in accurately identifying loops from low coverage Hi-C contact maps at high resolutions.

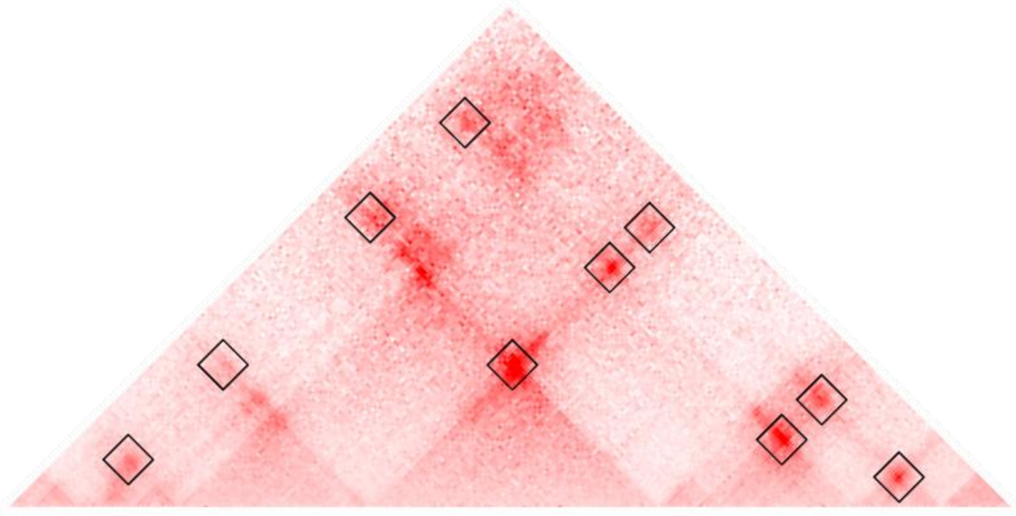

Fig 1. Chromatin loop examples (GM12878 Hi-C).

Motivation:

As chromatin loops manifest as small blob-shaped patterns on Hi-C contact maps, existing tools focus on analyzing contact pairs within a small area, such as a 21×21 window. However, they fail at capturing:

- Indistinct patterns in sparse regions:
  The blob-shaped patterns are often unclear in low coverage regions, leading to inaccurate loop detection.
- Broader patterns:
  Many chromatin loops exhibit broader patterns, including stripes, particularly in loops associated with the formation of Topologically Associating Domains, which current tools largely ignore.

Contribution:

- Integration long-range information:

Unlike traditional methods which only focus on local areas, LoopHunter considers much large regions (i.e., 224x224) to capture the broader footprint of chromatin loops.

- Leveraging knowledge distillation:

LoopHunter employs knowledge distillation from a teacher model (RefHiC [5]) and can make accurate chromatin loop annotations across a wide spectrum of cell types and coverage levels.

METHOD OVERVIEW

- Input: LoopHunter takes 224×224 sub-matrices as input;
- Model: LoopHunter employs a combination of axial-attention transformers and convolutional blocks to capture multi-scale data characteristics (Fig. 2);
- Training and prediction: LoopHunter is trained using knowledge distillation to learn from a teacher model (RefHiC) and makes dense predictions across the entire input region.
- Clustering: After obtaining the loop scores for each pixel, LoopHunter utilizes density-based clustering algorithm [5] to identify cluster centroid as the final predictions.

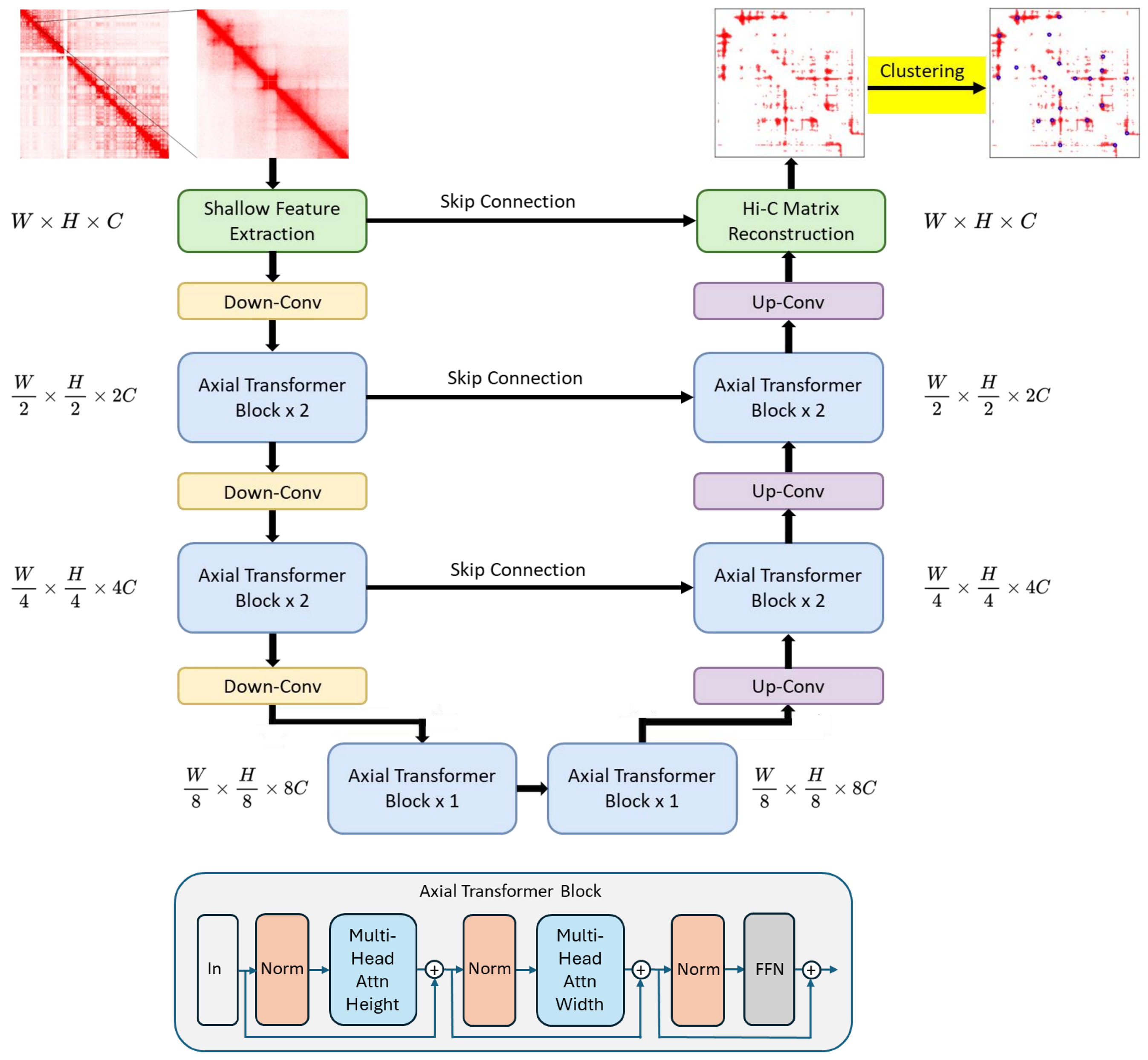

Fig 2. Overview of the LoopHunter network for loop scoring, followed by clustering algorithm.

RESULTS

LoopHunter accurately detects chromatin loops.

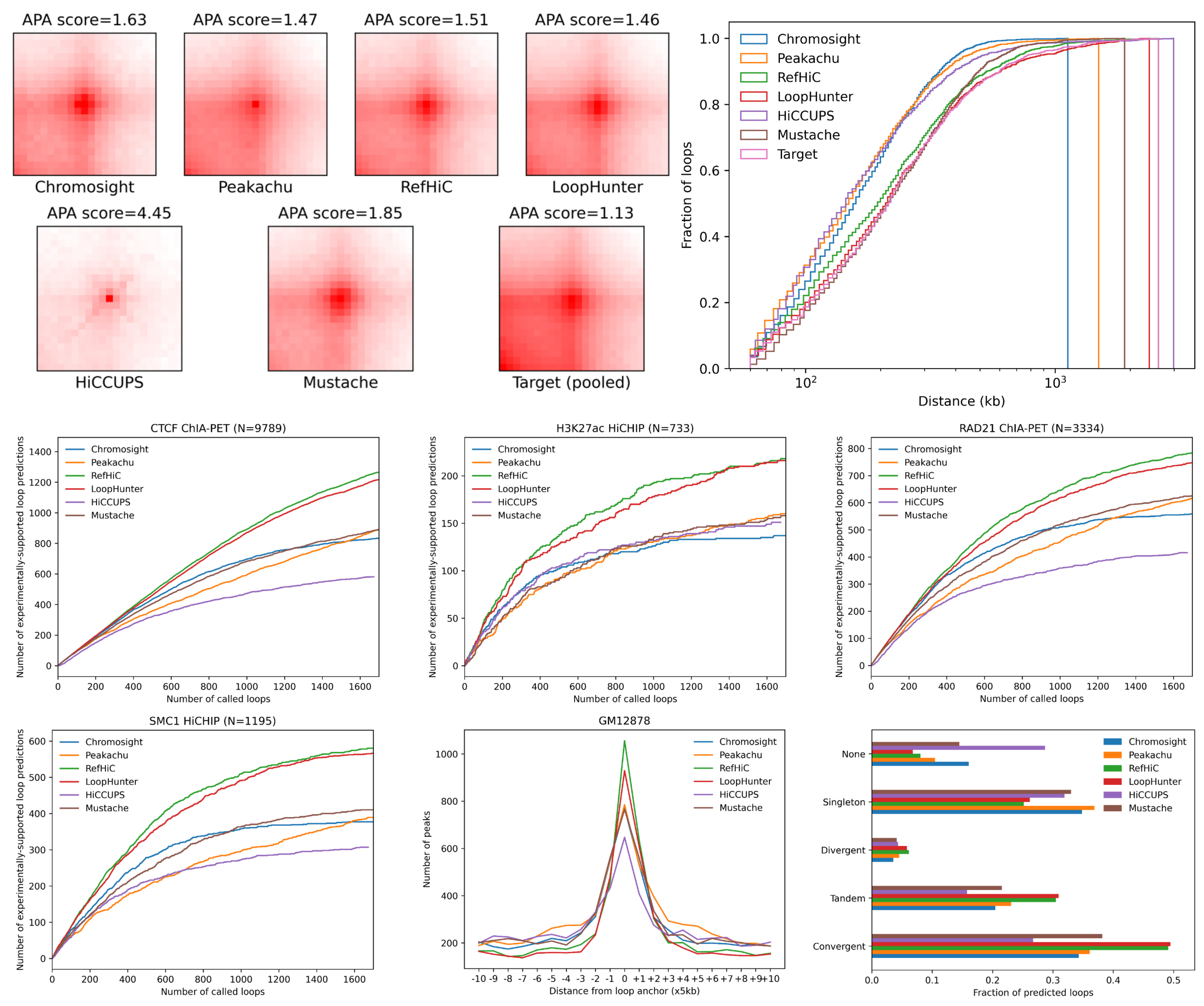

Fig 3. Comparison of LoopHunter and existing tools on GM12878 Hi-C data (500M read pairs).

LoopHunter identify chromatin loops across sequencing depths.

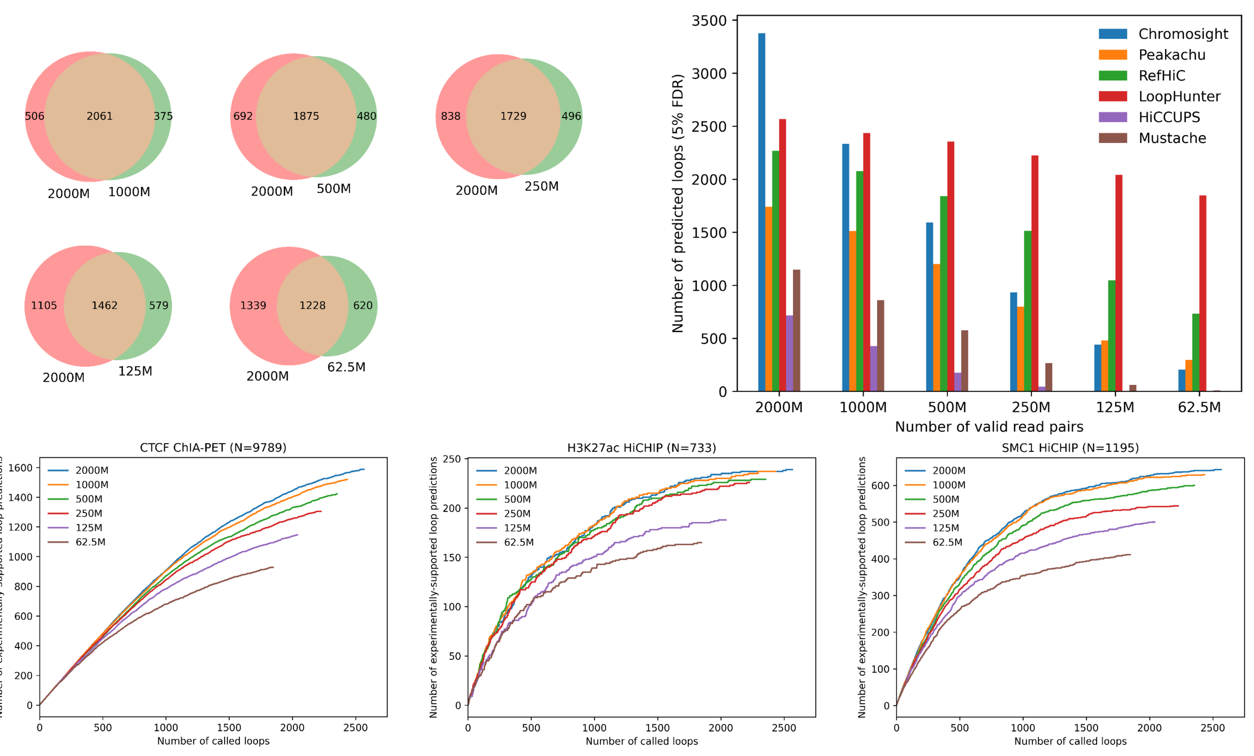

Fig 4. Comparison of LoopHunter and existing tools [1-5] on GM12878 Hi-C data with different valid read pairs.

LoopHunter excels across diverse cell types and species.

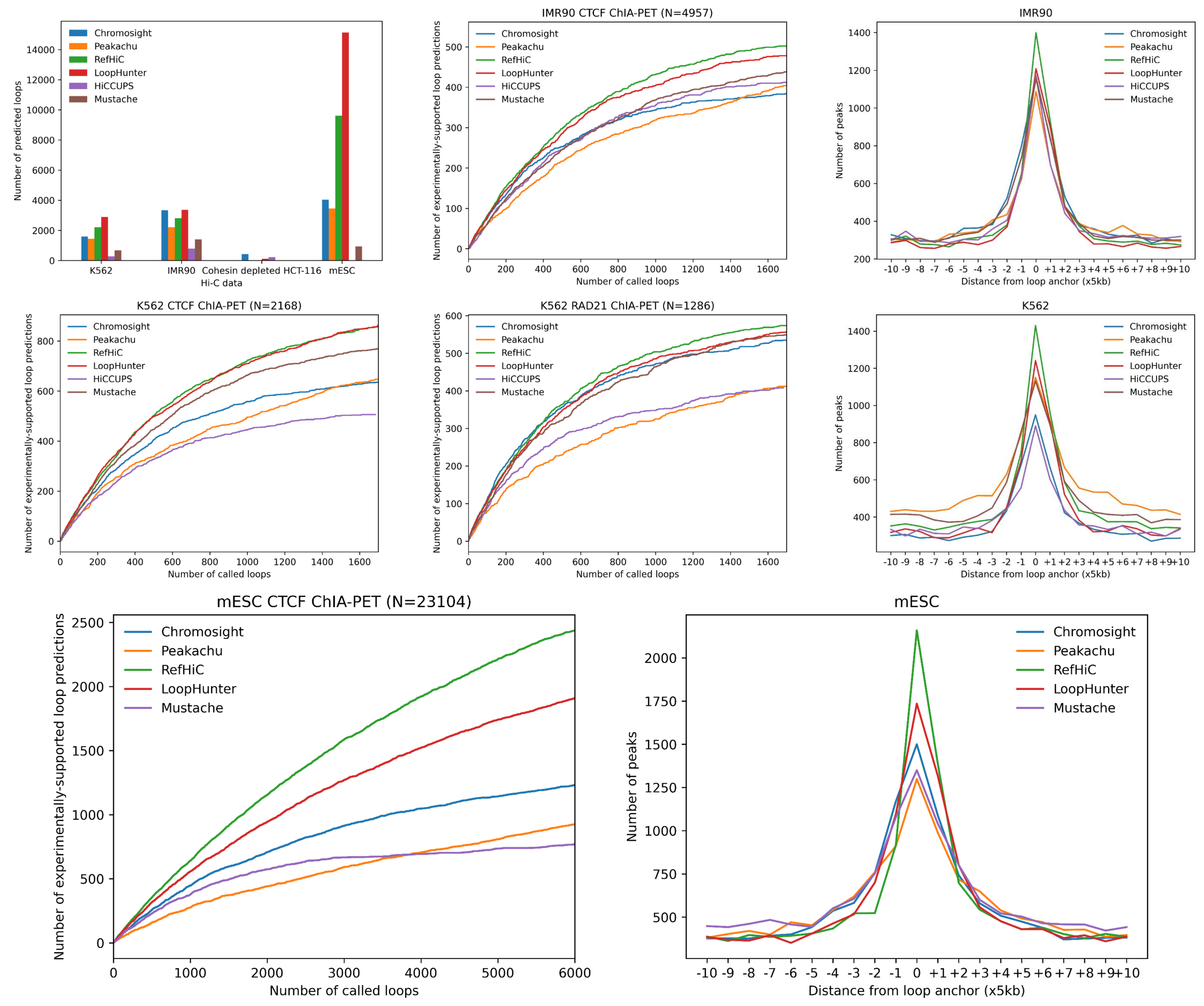

Fig 5. Comparison of LoopHunter and existing tools [1-5] on different Hi-C data from human K562, IMR90, and cohesin-depleted HCT-116 cells, as well as mouse ESC.

CONCLUSIONS

In this study, we introduce LoopHunter, an axial attention-based deep learning model to annotate chromatin loops from Hi-C contact.

LoopHunter distills knowledge from a reference-enabled approach, RefHiC [5], and excels by analyzing larger areas to effectively capture long-range information around loops.

This approach enhances the accuracy of loop detection across different coverages and cell types without requiring retraining, demonstrating significant improvements over traditional tools that often focus on smaller areas and miss broader patterns.

REFRERENCES

[1] Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).

[2] Salameh, T. J. et al. A supervised learning framework for chromatin loop detection in genome-wide contact maps. Nat. Commun. 11, 1–12 (2020).

[3] Roayaei Ardakany, A. et al. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. Genome Biol. 21, 1–17 (2020).

[4] Matthey-Doret, C. et al. Computer vision for pattern detection in chromosome contact maps. Nat. Commun. 11, 1–11 (2020).

[5] Zhang, Y. et al. Reference panel guided topological structure annotation of Hi-C data. Nat Commun 13, 7426 (2022).

ACKNOWLEDGEMENTS

We thanks Dr. Lujia Wang, Jiayong Zhou, Duohan Wu and Zirui Wang for useful discussions.