Paper Summaries

HiCPlus[1]: HiCPlus uses CNNs to find the high-resolution correspondent of low-resolution Hi-C data. In order to achieve this, it facilitates chromatin interaction analysis in different cell types. The model is trained on high-resolution datasets, enabling it to learn and enhance repeating patterns in low-level representations. HiCPlus produces enhanced Hi-C matrices highly similar to the true high-resolution matrices. The findings are important as they make the identification and analysis of significant chromatin interactions easier, which are crucial to our understanding of the 3D genome organization.

hicGAN[2]: hicGAN is a GAN network with a CNN generator and discriminator. Inherently, this paper is motivated by the idea that a properly trained discriminator can learn the latent properties of a Hi-C matrix, and that it can better train the generator than pixel-wise metrics. Similar to HiCPlus, the authors continue to approach the problem with a CV-inspired perspective. They also introduce various tools for interpreting Hi-C images to compare real world applicability of different scaled matrices. They present good performance in terms of metrics. However, we may be able to achieve better performance by integrating cell type-specific information.

Epiphany[3]: Though Hi C measurements remain costly and difficult to retrieve, other 1D epigenomic assays have become standard and extensively studied. Epiphany is an attempt to predict Hi-C matrices from these epigenomic tracks, implicitly incorporating cell type specific information into their predictions. Like hicGAN, the authors also focus not just on per-pixel metrics but on the interpretability of the actual predicted outputs.

DeepPHiC[4]: PHi-C data is expensive to get and therefore may not be available for tissues or cell types of interest. DeepPHiC aims to further promoter-capture Hi-C (pHi-C) data through analyzing promoter-promoter (PP) and promoter-enhancer (PE) interactions and predicts chromatin interaction. This architecture is a multimodal deep learning framework with deeply connected convolution layers that utilizes transfer learning. DeepPHi-C also utilizes a style similar to ResNet, with connections. DeepPHi-C also consists of two modules, a feature extractor module and a classifier module.

DeepC[5]: Utilizes transfer learning in order to predict genome folding from mega-base DNA sequence. Two step training process; the first step uses CNN (similar architecture to DeepSea) on 936 dataset to predict chromatin features. Second step network is us trained to predict Hi-C

interaction and preseeded with hidden weights from the previous step. Researchers found that the transfer learning allowed the model to be resilient to low quality Hi-C data.

References

[1] Zhang, Y., An, L., Xu, J. et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. Nat Commun 9, 750 (2018). https://doi.org/10.1038/s41467-018-03113-2

[2] Qiao Liu, Hairong Lv, Rui Jiang, hicGAN infers super resolution Hi-C data with generative adversarial networks, Bioinformatics, Volume 35, Issue 14, July 2019, Pages i99–i107, https://doi.org/10.1093/bioinformatics/btz317

[3] Yang, R., Das, A., Gao, V.R. et al. Epiphany: predicting Hi-C contact maps from 1D epigenomic signals. Genome Biol 24, 134 (2023). https://doi.org/10.1186/s13059-023-02934-9

[4] Hong H, Jiang S, Li H, Du G, Sun Y, Tao H, Quan C, Zhao C, Li R, Li W, Yin X, Huang Y, Li C, Chen H, Bo X. DeepHiC: A generative adversarial network for enhancing Hi-C data resolution. PLoS Comput Biol. 2020 Feb 21;16(2):e1007287. doi: 10.1371/journal.pcbi.1007287. PMID: 32084131; PMCID: PMC7055922.

[5] Schwessinger R, Gosden M, Downes D, Brown RC, Oudelaar AM, Telenius J, Teh YW, Lunter G, Hughes JR. DeepC: predicting 3D genome folding using megabase-scale transfer learning. Nat Methods. 2020 Nov;17(11):1118-1124. doi: 10.1038/s41592-020-0960-3. Epub 2020 Oct 12. PMID: 33046896; PMCID: PMC7610627.