Artificial Intelligence Prediction Across 12,000 Samples Shows Widespread Increased Gene-Gene Chromatin Interactions in Cancers that Constitute Therapeutic Vulnerabilities

Fuying Dao[a], **Benjamin Lebeau**[a], **Melissa Jane Fullwood**[*a]

[a] *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*

F. D: fuying.dao@ntu.edu.sg; *M. J. F: mfullwood@ntu.edu.sg

1. Introduction

Aberrant gene expression patterns are a hallmark of cancer, driving key features such as proliferation, invasion, and metastasis [1]. While cancer transcriptomes are shaped by epigenomic and genomic alterations [2], growing evidence highlights the role of altered 3D genome organization in cancer [3, 4]. Chromatin interactions, including enhancer-promoter, promoter-promoter (gene-gene interactions, GGIs), and silencer-promoter loops, bring distal regulatory elements into proximity, influencing transcription [5-8]. Despite efforts to map chromatin interaction landscapes across tissues and cell lines, a comprehensive understanding remains limited due to high costs and logistical challenges [9]. Artificial intelligence (AI), particularly deep learning, offers a powerful approach to model 3D genomic features [10, 11] However, no prior work has leveraged only RNA-Seq data to investigate chromatin interactions or explored AI as a tool for identifying drugs targeting 3D genome organization [12-15]. We reasoned that an AI-based method for predicting GGIs could address key questions in 3D genome organization: What are the global chromatin interaction patterns across the wide range of normal and cancer cells? How can drugs that change and reverse cancer-specific GGIs be effectively screened?

Here, we introduce AI4Loop, a bidirectional long short-term memory (Bi-LSTM) model that integrates multi-scale RNA-Seq data to predict cell type-specific GGIs. AI4Loop demonstrated robust generalization across different cell types and accurately distinguished Acute Myeloid Leukemia (AML) samples from normal controls. Using AI4Loop's efficient computation, we created a compendium of GGIs across nearly 12,000 samples, spanning diverse cell types and cancer subtypes. We showed that cancer cells tend to strengthen their GGIs and that GGIs are more predictive of cancer subtypes than RNA expression, indicating that GGIs are highly cell type specific. Furthermore, AI4Loop identified drugs that modulate GGIs by constructing a drug-perturbation GGI atlas from 50,000 drug-treated samples. Experimental confirmation by Hi-C further showed that the antibiotics eperezolid and radezolid induced the loss of oncogenic GGIs and cell viability assays showed they led to cancer cell death. These drugs would have been very difficult to discover by traditional assays. These findings establish AI4Loop as a rapid and effective platform for elucidating GGI dynamics, with potential applications in cancer identification, drug discovery and personalized treatment strategies.

2. Rationale

We reasoned that an artificial intelligence (AI)-powered framework for predicting GGIs from RNA-Seq data would enable rapid scanning of GGIs across large cohorts of cancer and normal samples as well as drug screening surveys. This would first enable us to understand whether GGIs are gained, lost or unchanged in cancers, and second, elucidate which drugs or small chemicals may be able to reverse GGI changes seen in cancer. This AI framework would facilitate the discovery of drugs for altering GGIs in cancer that would have been impossible to discover previously.

3. Results

First, we developed AI4loop, a novel deep learning method for predicting GGIs from RNA-Seq data only.

Second, using AI4Loop, we systematically mapped GGIs across 12,000 patient samples spanning 32 cancer types from The Cancer Genome Atlas (TCGA). GGI-based cancer classification was more predictive than traditional RNA expression, indicating GGIs have a high degree of cell type specificity.

Third, we found unprecedented, widespread oncogenic gains of GGIs across nearly all cancers, revealing that almost all tumors strengthen GGIs to sustain aberrant transcriptional programs.

Fourth, we further extended this approach to identify drugs that modulate GGIs by constructing a drug-perturbation GGI atlas from 50,000 drug-treated samples from the Connectivity Map (CMap) and LINCS Unified Environment (CLUE) database. We identified eperezolid and radezolid, two antibiotics that significantly disrupted cancer-acquired GGIs, which was confirmed by Hi-C experiments.

Taken together, our findings establish AI4Loop as a scalable platform for uncovering chromatin interaction-based vulnerabilities in cancer.

4. Conclusion

Our study represents the first large-scale, systematic mapping of GGIs in cancer and demonstrates the power of AI-driven approaches. AI4Loop provides an innovative strategy for studying chromatin interactions and identifying therapeutic targets at unprecedented scale and resolution. Our discovery of GGI-disrupting drugs opens new avenues for targeted cancer therapies that would not have been possible without the use of AI in Science, and provides a framework for further AI in Science drug discovery approaches.

Artificial Intelligence Prediction Across 12,000 Samples Shows Widespread Increased Gene-Gene Chromatin Interactions in Cancers that Constitute Therapeutic Vulnerabilities

Fuying Dao[a], **Benjamin Lebeau**[a], **Melissa Jane Fullwood**[*a]

[a] *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*

F. D: fuying.dao@ntu.edu.sg; *M. J. F: mfullwood@ntu.edu.sg

Fig. 1: Graphical Abstract – AI4Loop predicts gene-gene interactions (GGIs) from RNA-Seq data using a sliding-window-based Bi-LSTM model, enabling large-scale chromatin loop analysis, pan-cancer mapping, and drug screening to identify compounds that disrupt oncogenic GGIs.

Data and materials availability

The GGI profiles of cancer and healthy samples and codes for AI4Loop training and prediction are available at https://github.com/DaoFuying/AI4Loop. All other data are available in the main text or the supplementary materials and deposited to the Nanyang Technological University's repository (https://doi.org/10.21979/N9/ORBU74). The drug-perturbed Hi-C datasets are generated during the current study are available in GEO under accession number GSE287383.

Acknowledgments

This research is supported by the National Research Foundation Singapore under the AI Singapore Programme (AISG Award No: AISG3-GV-2023-014) and by the Ministry of Education, Singapore under its Academic Research Fund Tier 1 Thematic (RT5/22), both awarded to M.J.F (PI). This research is also supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research Investigator Award STaR (MOH-000709).

References

[1] L. Zhang *et al.*, "Gene expression profiles in normal and cancer cells," *Science,* vol. 276, no. 5316, pp. 1268-72, May 23 1997, doi: 10.1126/science.276.5316.1268.

[2] A. S. Adler, M. Lin, H. Horlings, D. S. Nuyten, M. J. van de Vijver, and H. Y. Chang, "Genetic regulators of large-scale transcriptional signatures in cancer," *Nat Genet,* vol. 38, no. 4, pp. 421-30, Apr 2006, doi: 10.1038/ng1752.

[3] J. Schuijers *et al.*, "Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism," *Cell Rep,* vol. 23, no. 2, pp. 349-360, Apr 10 2018, doi: 10.1016/j.celrep.2018.03.056.

[4] S. S. Rao *et al.*, "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell,* vol. 159, no. 7, pp. 1665-80, Dec 18 2014, doi: 10.1016/j.cell.2014.11.021.

[5] J. M. Dowen *et al.*, "Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes," *Cell,* vol. 159, no. 2, pp. 374-387, Oct 9 2014, doi: 10.1016/j.cell.2014.09.030.

[6] M. J. Fullwood *et al.*, "An oestrogen-receptor-alpha-bound human chromatin interactome," *Nature,* vol. 462, no. 7269, pp. 58-64, Nov 5 2009, doi: 10.1038/nature08497.

[7] G. Li *et al.*, "Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation," *Cell,* vol. 148, no. 1-2, pp. 84-98, Jan 20 2012, doi: 10.1016/j.cell.2011.12.014.

[8] Y. Cai *et al.*, "H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions," *Nat Commun,* vol. 12, no. 1, p. 719, Jan 29 2021, doi: 10.1038/s41467-021-20940-y.

[9] X. Zhang *et al.*, "Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers," *Nat Genet,* vol. 48, no. 2, pp. 176-82, Feb 2016, doi: 10.1038/ng.3470.

[10] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature,* vol. 596, no. 7873, pp. 583-589, Aug 2021, doi: 10.1038/s41586-021-03819-2.

[11] Y. Zhang, L. Boninsegna, M. Yang, T. Misteli, F. Alber, and J. Ma, "Computational methods for analysing multiscale 3D genome organization," *Nat Rev Genet,* vol. 25, no. 2, pp. 123-141, Feb 2024, doi: 10.1038/s41576-023-00638-1.

[12] S. Singh, Y. Yang, B. Poczos, and J. Ma, "Predicting enhancer-promoter interaction from genomic sequence with deep neural networks," *Quant Biol,* vol. 7, no. 2, pp. 122-137, Jun 2019, doi: 10.1007/s40484-019-0154-0.

[13] J. Zhou, "Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale," *Nat Genet,* vol. 54, no. 5, pp. 725-734, May 2022, doi: 10.1038/s41588-022-01065-4.

[14] J. Tan *et al.*, "Cell-type-specific prediction of 3D chromatin organization enables high-throughput in silico genetic screening," *Nat Biotechnol,* vol. 41, no. 8, pp. 1140-1150, Aug 2023, doi: 10.1038/s41587-022-01612-8.

[15] R. Yang *et al.*, "Epiphany: predicting Hi-C contact maps from 1D epigenomic signals," *Genome Biol,* vol. 24, no. 1, p. 134, Jun 6 2023, doi: 10.1186/s13059-023-02934-9.