

# RepliChrom: Interpretable machine learning predicts cancer-associated enhancer-promoter interactions using DNA replication timing



Fuying Dao<sup>1,2</sup>, Benjamin Lebeau<sup>2</sup>, Crystal Ling<sup>2</sup>, Mi Yang<sup>3</sup>, Xueqin Xie<sup>1</sup>,  
Melissa Jane Fullwood<sup>2,4,\*</sup>, Hao Lin<sup>1,\*</sup>, Hao Lyu<sup>1,\*</sup>

<sup>1</sup> Department of Clinical Laboratory,  
Sichuan Clinical Research Center for Cancer,  
Sichuan Cancer Hospital & Institute, Sichuan Cancer Center,  
School of Life Science and Technology,  
University of Electronic Science and Technology of China

<sup>2</sup> School of Biological Sciences, Nanyang Technological University

<sup>3</sup> The Clinical Hospital of Chengdu Brain Science Institute,  
School of Life Science and Technology,  
University of Electronic Science and Technology of China

<sup>4</sup> Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A\*STAR), Singapore



Fuying Dao, Benjamin Lebeau, Crystal Chia Yin Ling, Mi Yang, Xueqin Xie, Melissa Jane Fullwood, Hao Lin, Hao Lyu. 2025. RepliChrom: Interpretable machine learning predicts cancer-associated enhancer-promoter interactions using DNA replication timing. *iMeta* 4: e70052. <https://doi.org/10.1002/imt2.70052>



# Introduction

- **Multiscale 3D genome organization**

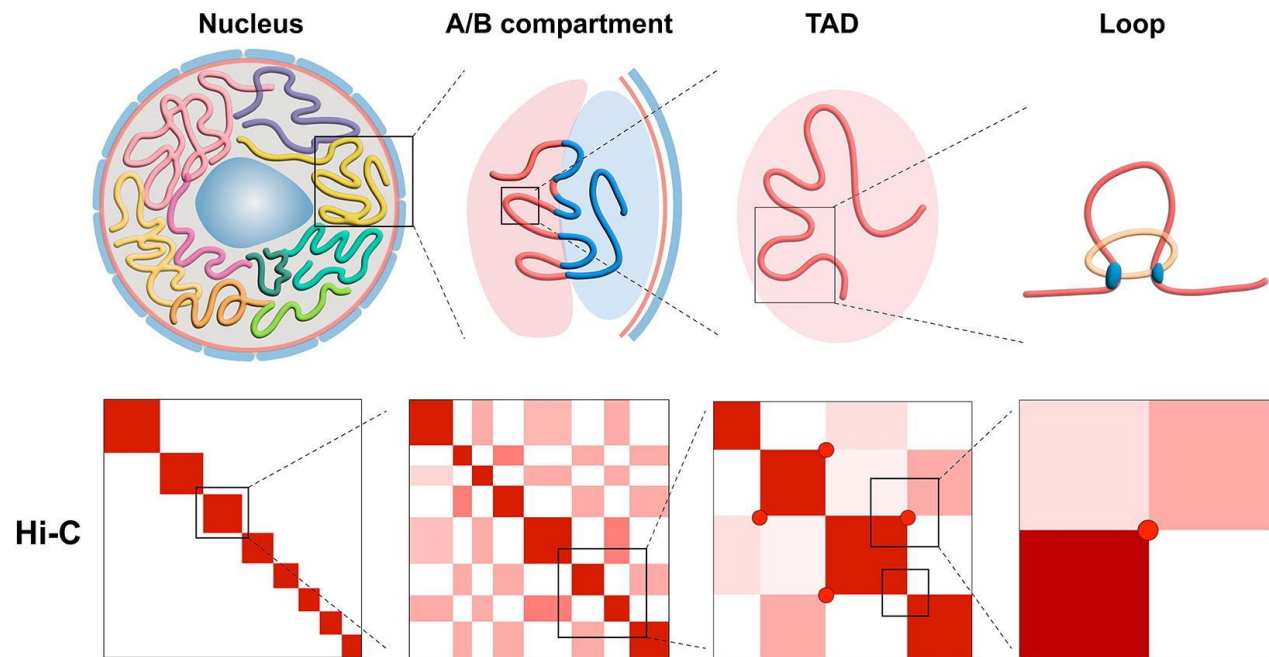


Figure 1. Different scales of 3D genome architecture.  
Bonev et al., *Nat Rev Genet*, 2016

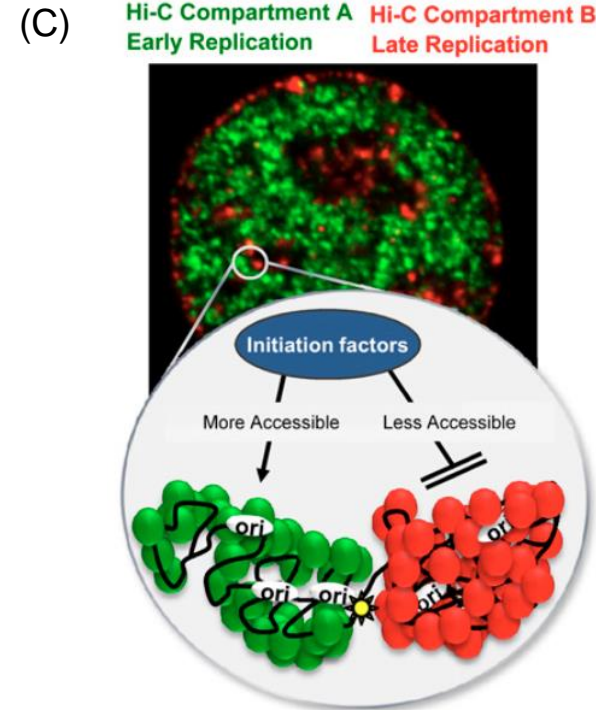
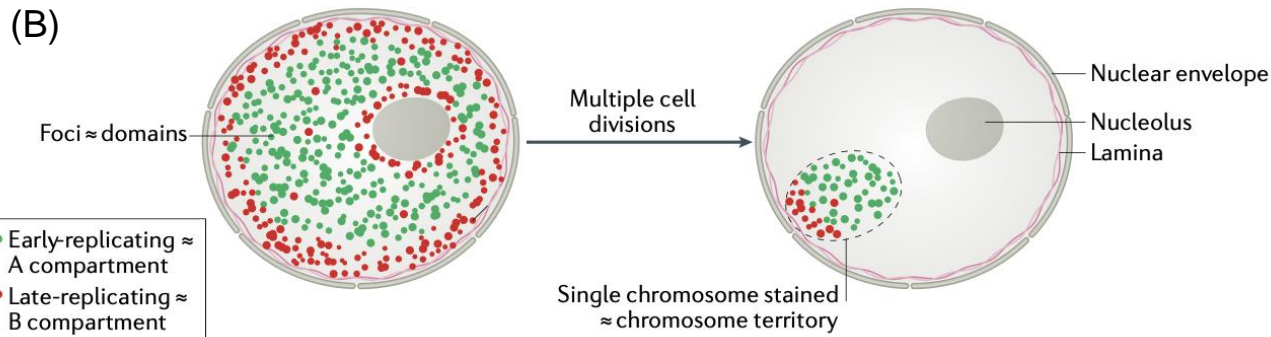
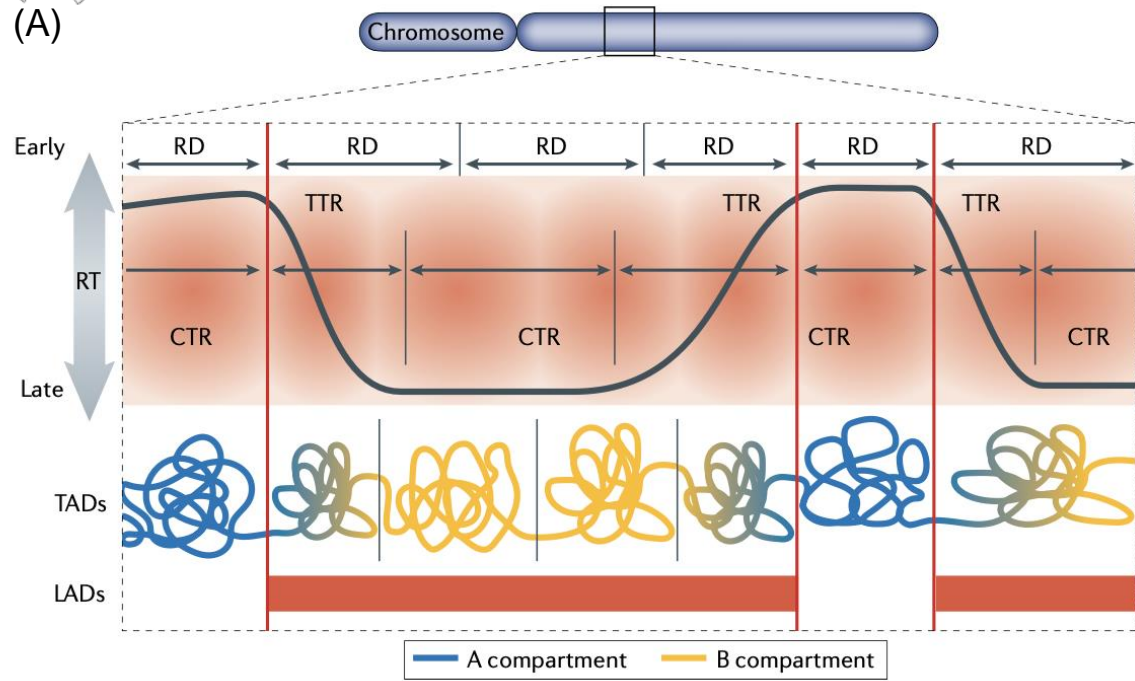
- **Machine learning offers a viable alternative**

Computational methods have greatly advanced our understanding of 3D genome organization. However, these methods all require DNA and/ or other epigenetic datasets, which are also laborious to produce.

- Human genome is organized into different levels of structure including TAD (Topologically Associated Domains) and chromatin interactions.
- Chromatin interactions, including enhancer-promoter (EPI), promoter-promoter, and silencer-promoter loops, bring distal regulatory elements into proximity, influencing transcription and cell fate determination.
- Chromatin interactions can be identified by the “C” methods which rely on proximity ligation of formaldehyde fixed and digested chromatin, e.g. Hi-C, ChIA-PET. Aiden et al., *Science*, 2009; Fullwood et al., *Nature*, 2009

- “C” methods are costly and technically demanding, limiting their application in large cohorts or clinical samples.

# DNA replication timing (RT) reflects 3D genome organization



David M. Gilbert

San Diego Biomedical Research Institute

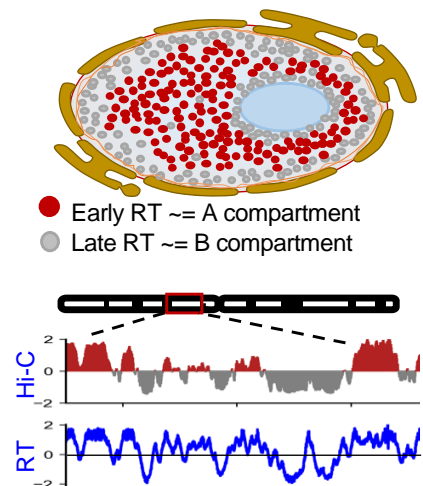
We hypothesize that RT may serve as an additional predictive signal for chromatin interactions, beyond conventional DNA sequence and epigenetic features.

Figure 2. DNA replication timing reflects 3D genome organization.  
Marchal C, Sima J, Gilbert DM. *Nat Rev Mol Cell Biol.* 2019

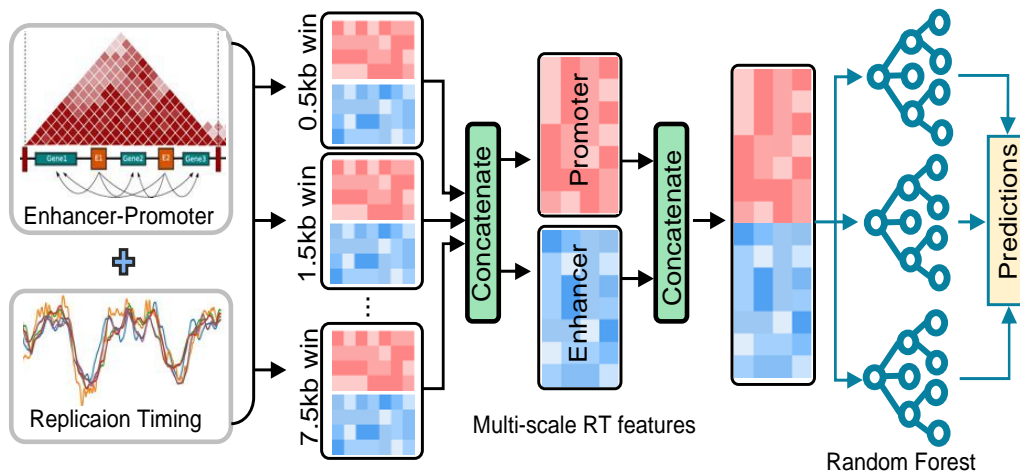


# Overview of RepliChrom

(A) Replication timing (RT) reflects genome organization

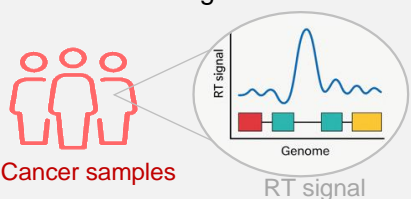


(B) RepliChrom: RT signals to predict enhancer-promoter interactions (EPIs)

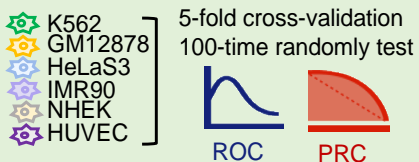


(C)

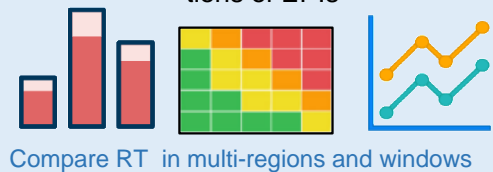
Abnormal RT signal in cancer



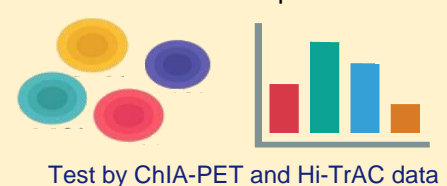
Training of RepliChrom is strongly predictive across cell lines



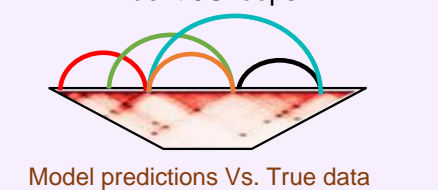
RT feature analysis optimized predictions of EPIs



RepliChrom accurately identifies EPIs in other C-techniques data



RepliChrom can identify independent 5C loops



RepliChrom application in acute lymphoblastic leukemia (ALL) samples

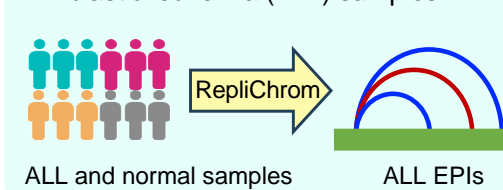


Figure 3. Overview of the RepliChrom model. (A) DNA replication timing (RT) reflects 3D genome organization. (B) The model takes Hi-C loop and RT data as input, generating a dataset of enhancer-promoter interactions (EPIs) and distance-matched non-EPIs. Next, multi-scale RT information was then extracted based on multi-bins, followed by classification model training using Random Forest. (C) Analysis in this study include RT Characteristic Analysis, RepliChrom model training and evaluation, RT features analysis, RepliChrom application in ChIA-PET and Hi-TrAC data, RepliChrom application in 5C data, and RepliChrom application in cancer.

# Abnormal RT Signal Across Cell Lines and Cancer

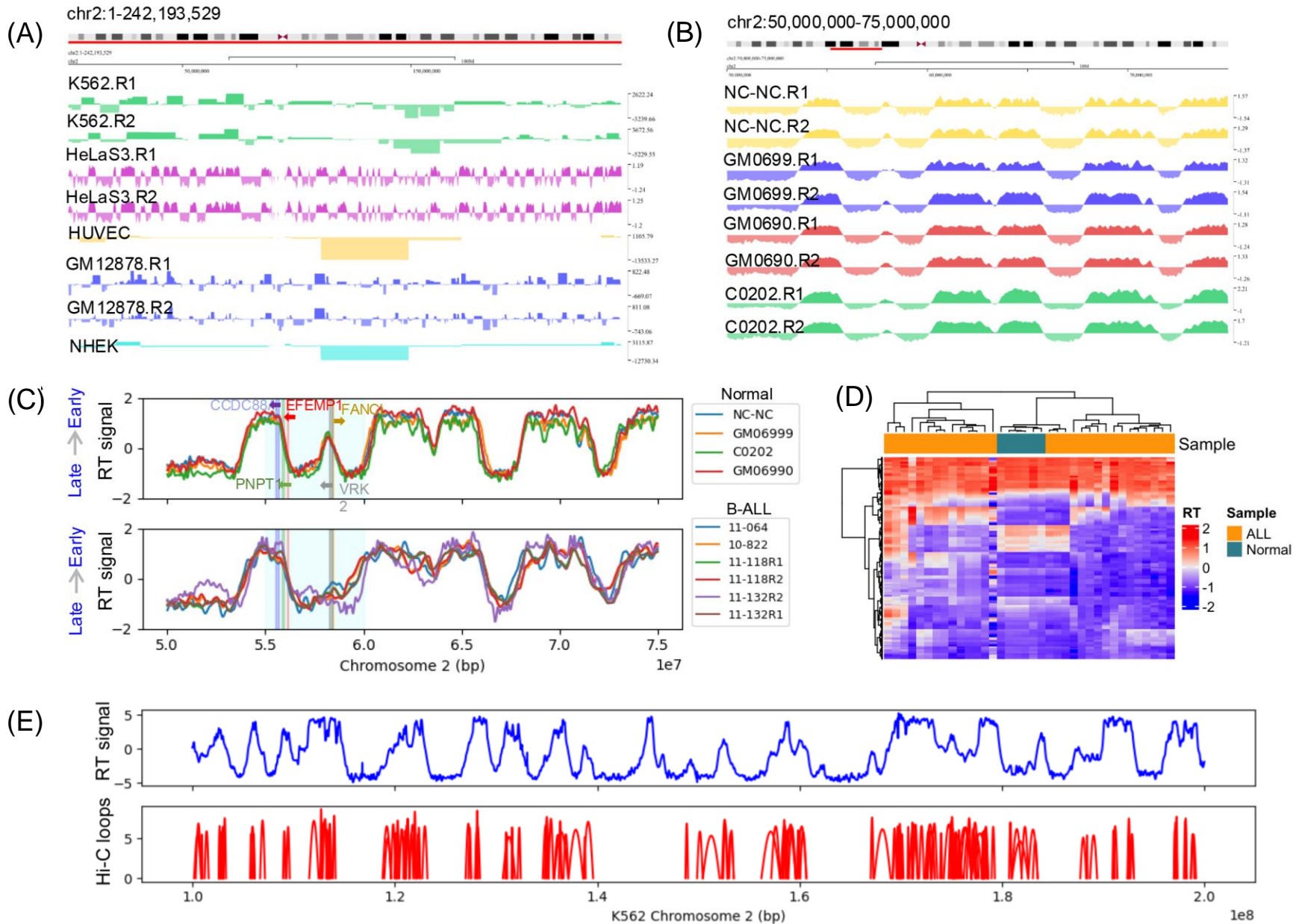


Figure 4. Analysis of replication timing (RT) profiles across different cell types and conditions. (A) Visualization of RT profiles on chromosome 2 in five cell lines using the Nucleosome Browser tool. (B) RT profiles of four non-leukemic B lymphoblastoid cell lines (C0202, NC-NC, GM06990, GM06999) over a 25-Mb segment of chromosome 2. (C) Comparison of RT profiles between normal mature B lymphocytes and B-ALL patient samples. (D) Unsupervised hierarchical clustering of ALL and normal samples based on RT information from the 55-60 Mb region on chromosome 2. (E) Correlation of RT and Hi-C signals in the K562 cell line on chromosome 2, demonstrating the alignment of RT with peaks of chromatin interaction intensity.

# RepliChrom Robustly Predicts EPIs Using Only RT Profiles

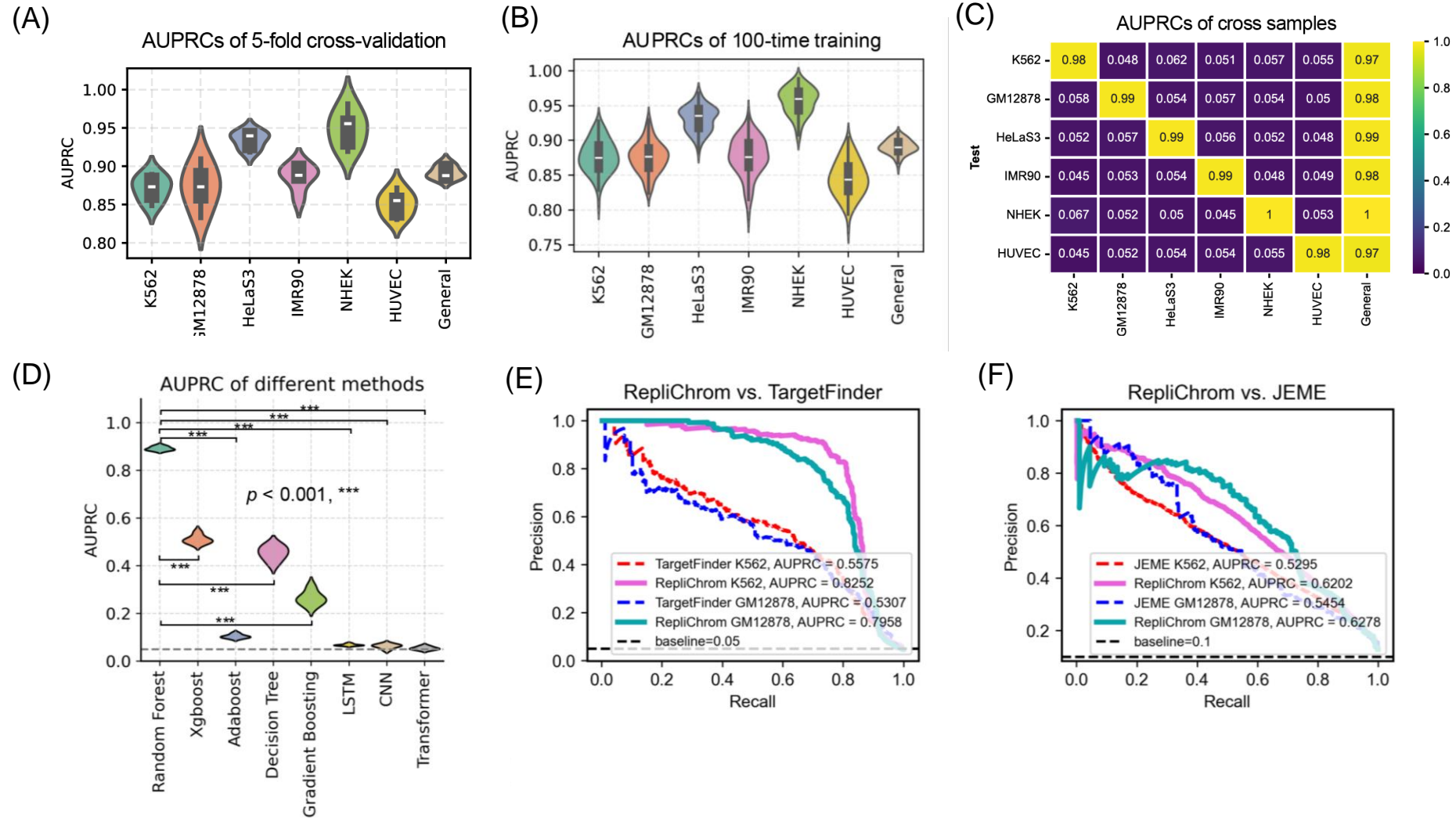


Figure 5. RepliChrom can predict enhancer-promoter interactions (EPIs) in both in-sample and cross-sample test, which also outperform other computational methods.



# RepliChrom Generalizes Across Platforms data

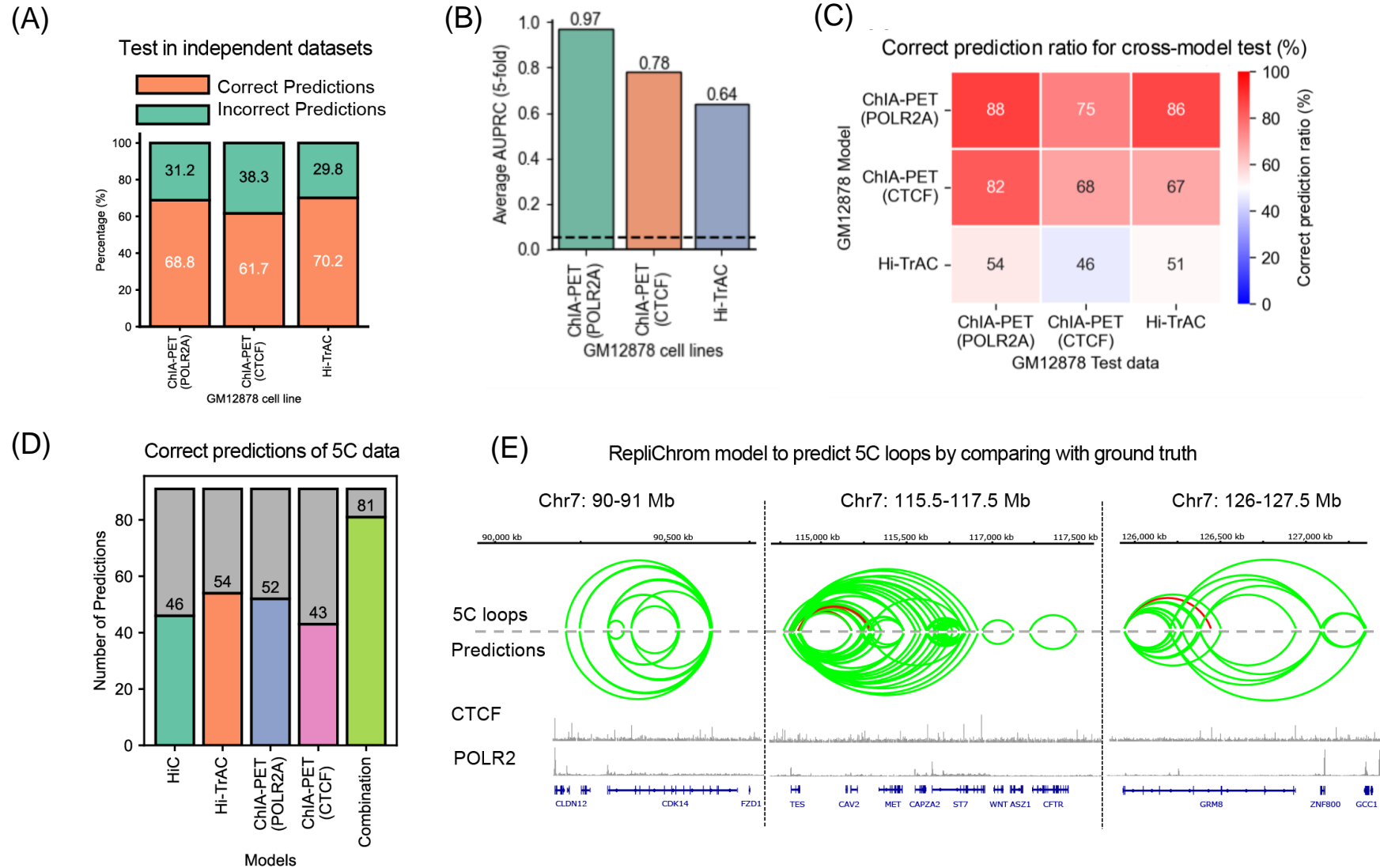


Figure 6. RepliChrom demonstrates strong cross-platform generalization in predicting chromatin interaction loops.



# RepliChrom Reveals oncogene-EPIs at Leukemia

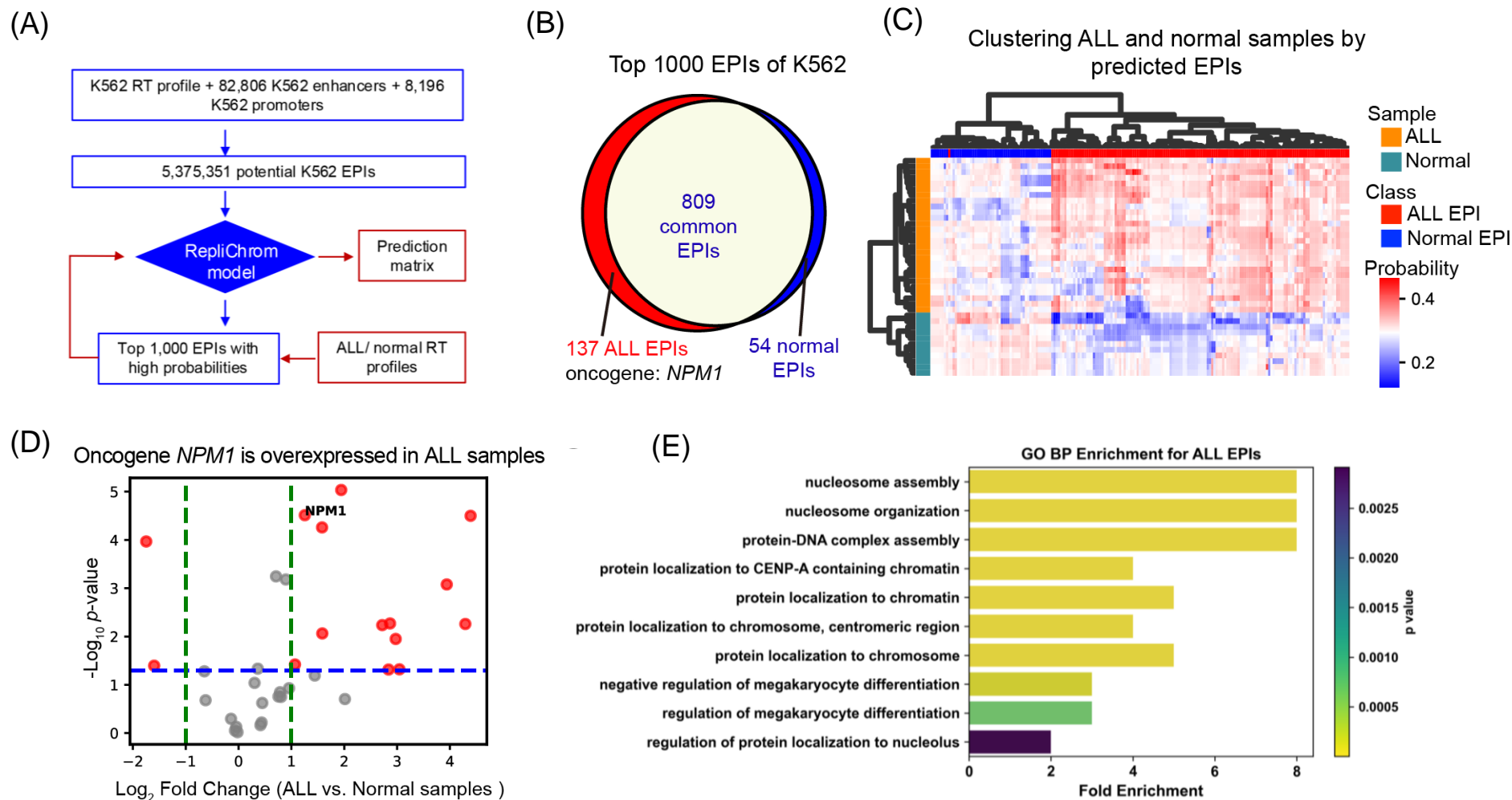


Figure 7. Replichrom model application in acute lymphoblastic leukemia (ALL) samples. known oncogenes such as *NPM1* were found among the ALL-specific EPIs.



# Summary

- ❑ We developed RepliChrom, an interpretable machine learning framework that leverages DNA RT to predict EPIs across diverse human cell lines
- ❑ Promoter-associated replication timing signals are key predictors of enhancer-promoter interactions
- ❑ RepliChrom generalizes well across platforms (Hi-C, Hi-TrAC, ChIA-PET, 5C)
- ❑ RepliChrom predicted cancer-associated EPIs could serve as a novel layer of epigenetic regulation in cancer and offer insights for biomarker discovery.



“**iMeta**” launched by iMeta Science Society in 2022, **impact factor (IF) 23.8**, ranking top 107/21973 in world and 2/161 in the **microbiology**. It aims to publish innovative and high-quality papers with broad and diverse audiences. **Its scope is similar to Cell, Nature, Science, Nature Biotechnology/Methods/Microbiology/Medicine/Food**. Its unique features include video abstract, bilingual publication, and social media with 600,000 followers. Indexed by **SCIE/ESI, PubMed, Google Scholar** etc.

“**iMetaOmics**” launched in 2024, with a **target IF>10**, and its scope is similar to **Nature Communications, Cell Reports, Microbiome, ISME J, Nucleic Acids Research, Briefings in Bioinformatics**, etc.

“**iMetaMed**” launched in 2025, with a **target IF>15**, similar to **Med, Cell Reports Medicine, eBioMedicine, eClinicalMedicine** etc.



Society: <http://www.imeta.science>

Publisher: <https://wileyonlinelibrary.com/journal/imeta>

iMeta: <https://wiley.atyponrex.com/journal/IMT2>

Submission: iMetaOmics: <https://wiley.atyponrex.com/journal/IMO2>

iMetaMed: <https://wiley.atyponrex.com/journal/IMM3>



[iMetaScience](#)



[iMetaScience](#)



[office@imeta.science](mailto:office@imeta.science)  
[imetaomics@imeta.science](mailto:imetaomics@imeta.science)



[Promotion Video](#)

Update  
2025/5/21