



RepliChrom: 基于DNA复制时序的癌症增强子-启动子互作可解释性预测模型

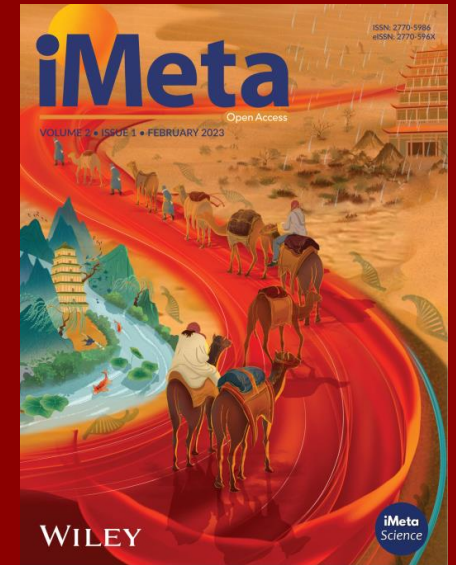
刀福英^{1,2}, Benjamin Lebeau², Crystal Ling², 杨宓³, 谢雪琴¹,
Melissa Jane Fullwood^{2,4,*}, 林昊^{1,*}, 吕昊^{1,*}

¹ 电子科技大学生命科学与技术学院, 四川省肿瘤医院

² 南洋理工大学生命科学院

³ 电子科技大学生命科学与技术学院, 成都脑科学研究院临床医院

⁴ 新加坡科技研究局 (A*STAR) 分子与细胞生物学研究所



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>

• 多尺度三维基因组结构

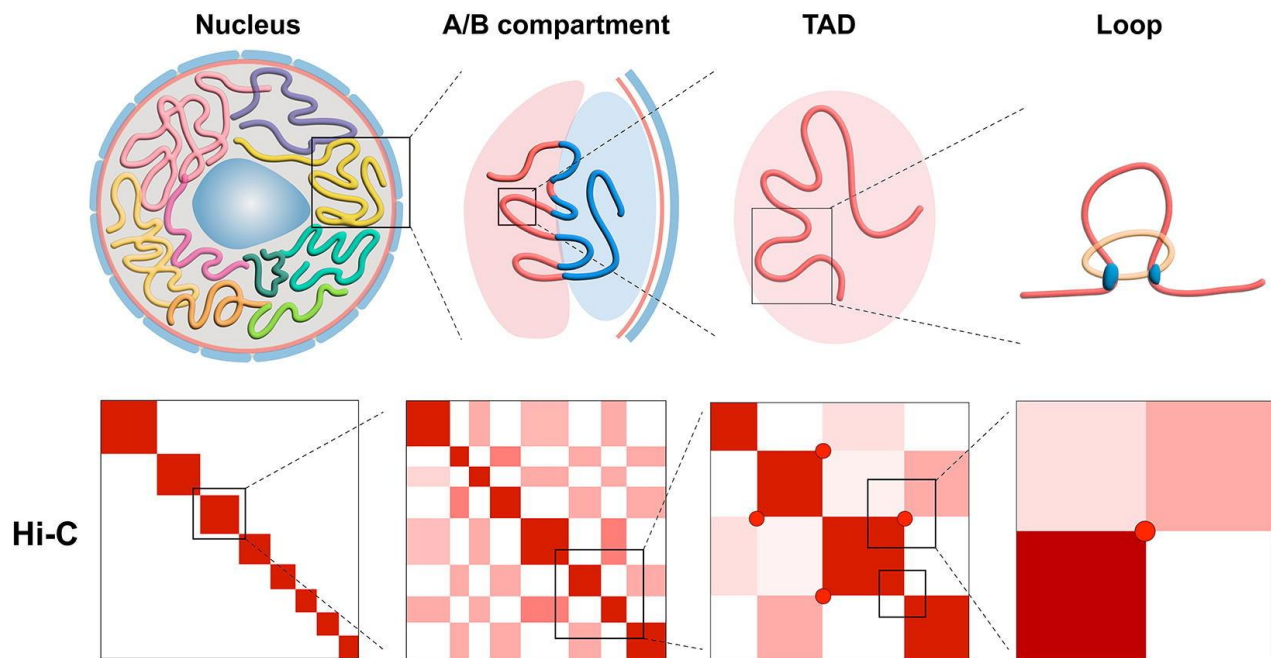


图1. 多尺度三维基因组结构。
Bonev et al., *Nat Rev Genet*, 2016

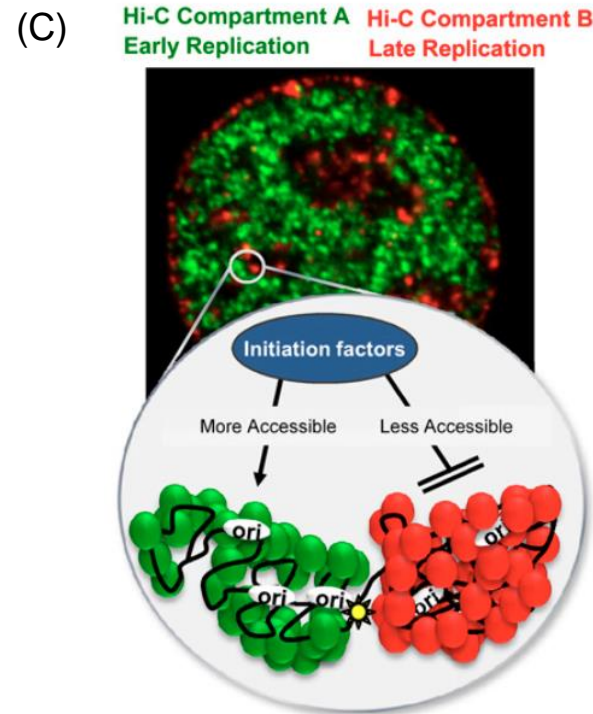
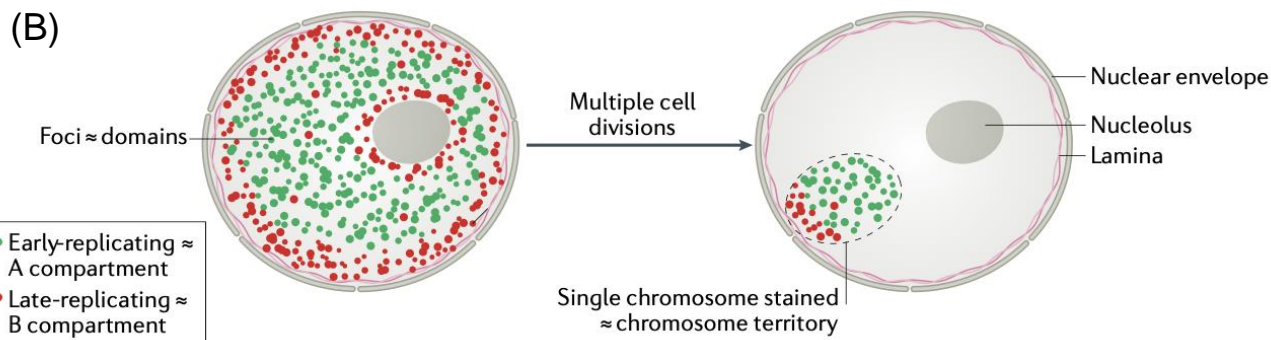
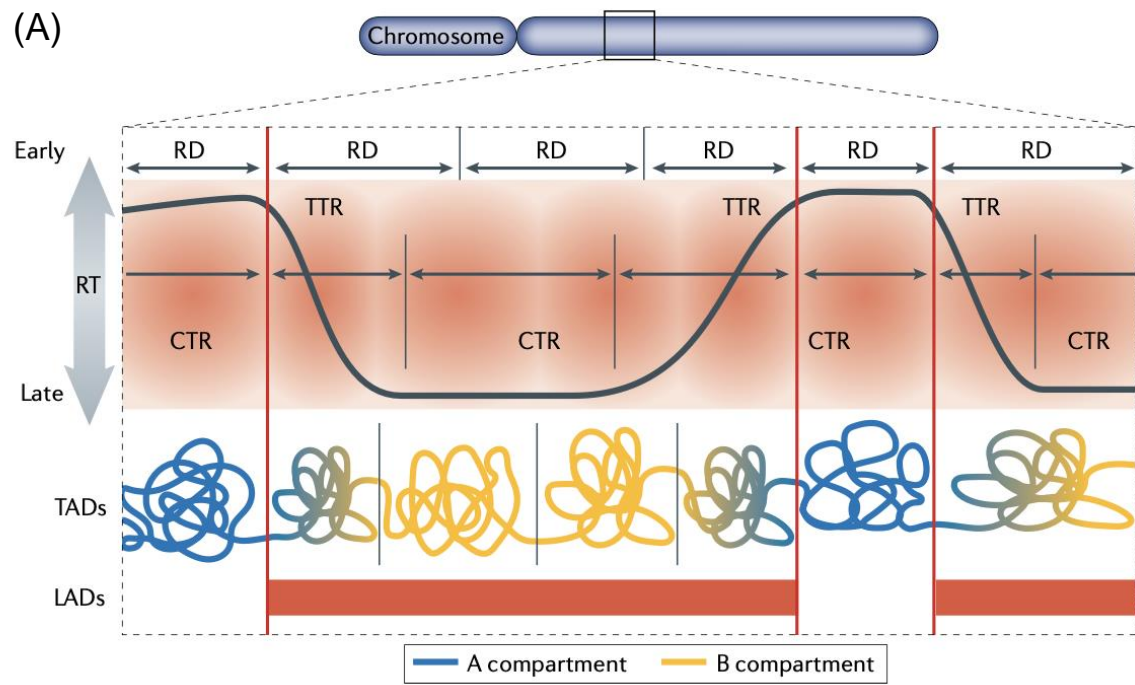
- 人类基因组由不同层次的结构组成，包括拓扑相关域和染色质相互作用。
 - 三维染色质互作包括增强子-启动子、启动子-启动子和沉默子-启动子环，使远端调控元件彼此接近，从而影响转录和细胞命运决定。
 - 染色质互作可以通过高通量测序技术“C”方法识别，其方法依赖于甲醛固定和消化染色质的邻近连接，例如Hi-C和ChIA-PET。
- 然而，这些“C”方法成本高昂且技术要求高，限制了其在大样本或临床样本中的应用。

• 机器学习提供了可行的替代方案

计算方法极大地推进了人们对三维基因组结构的理解，但是这些方法都需要DNA序列或其它表观遗传数据集，而这些数据集的生成也十分困难，为此当前的预测模型存在一定缺陷。



DNA复制时序对应三维基因组结构



David M. Gilbert

San Diego Biomedical Research Institute

本研究假设复制时序信号可以作为三维染色质互作的额外预测信号，超越传统的 DNA 序列和表观遗传特征。

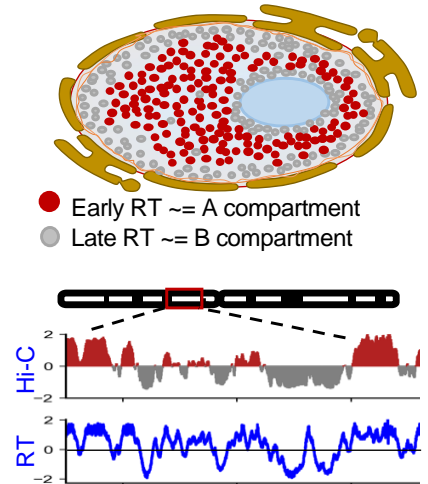
图2. DNA复制时序反映基因组的三维结构。

Marchal C, Sima J, Gilbert DM. *Nat Rev Mol Cell Biol.* 2019

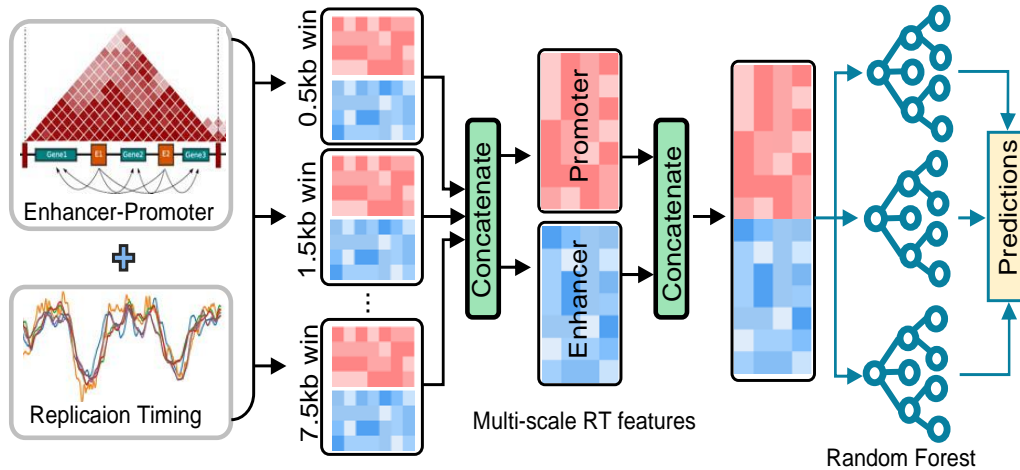


RepliChrom预测框架概括

(A) Replication timing (RT) reflects genome organization

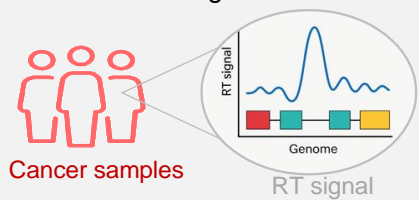


(B) RepliChrom: RT signals to predict enhancer-promoter interactions (EPIs)

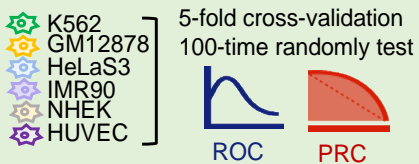


(C)

Abnormal RT signal in cancer



Training of RepliChrom is strongly predictive across cell lines



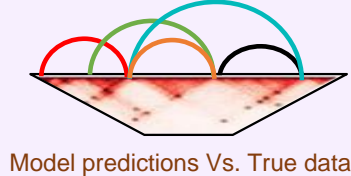
RT feature analysis optimized predictions of EPIs



RepliChrom accurately identifies EPIs in other C-techniques data



RepliChrom can identify independent 5C loops



RepliChrom application in acute lymphoblastic leukemia (ALL) samples

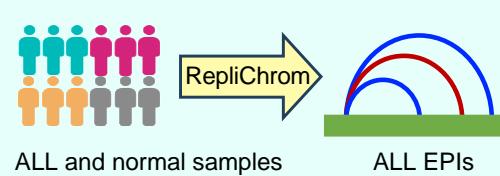


图3. RepliChrom预测模型概述。

(A) DNA复制时序能够反映三维基因组结构。

(B) 模型以Hi-C染色质环和复制时序信号为输入，生成增强子-启动子互作正样本和距离匹配的负样本数据集。随后提取多尺度复制时序特征，并使用随机森林进行分类模型训练。

(C) 本研究的分析内容包括：复制时序特性分析、RepliChrom模型训练与评估、复制时序特征重要性评估、RepliChrom在ChIA-PET和Hi-TrAC数据中的应用、在5C数据中的应用，以及在癌症样本中的应用。



不同细胞系和癌症中的异常复制时序信号

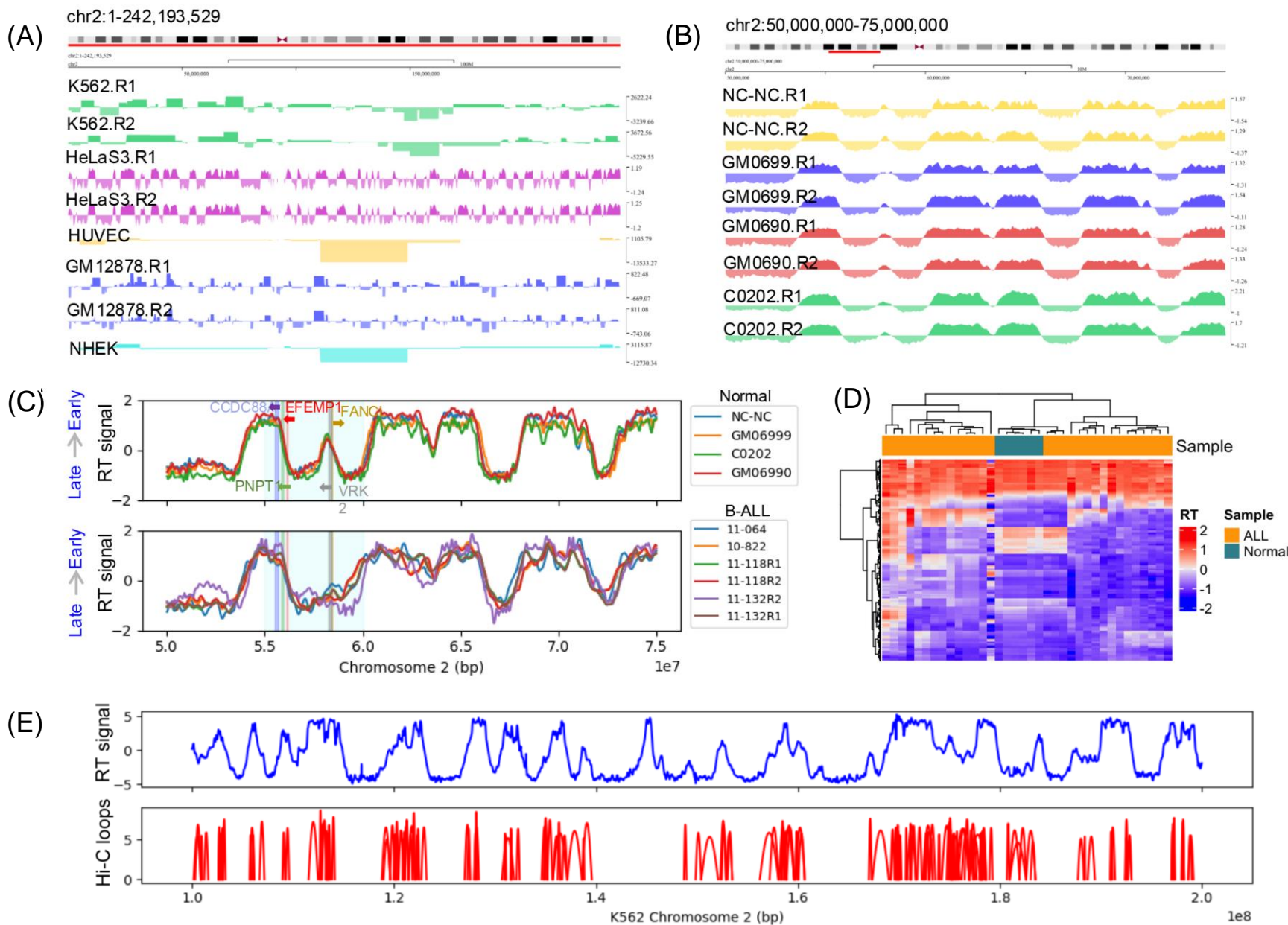


图4. 不同细胞类型和条件下复制时序谱的特性分析。

(A)可视化不同细胞系中复制时序谱。

(B) 四种非白血病B淋巴母细胞系

(C0202、NC-NC、GM06990、

GM06999) 在2号染色体长25 Mb区域上的复制时序信号。

(C) 比较正常成熟B淋巴细胞与B型急性性淋巴白血病患者样本之间复制时序差异。

(D) 基于2号染色体上55-60 Mb区域的复制时序信息, 对白血病和正常样本进行无监督层次聚类分析。

(E) 在K562细胞系中, 2号染色体上复制时序信号与Hi-C信号具有高度一致性。



RepliChrom仅使用复制时序信号准确预测染色质互作

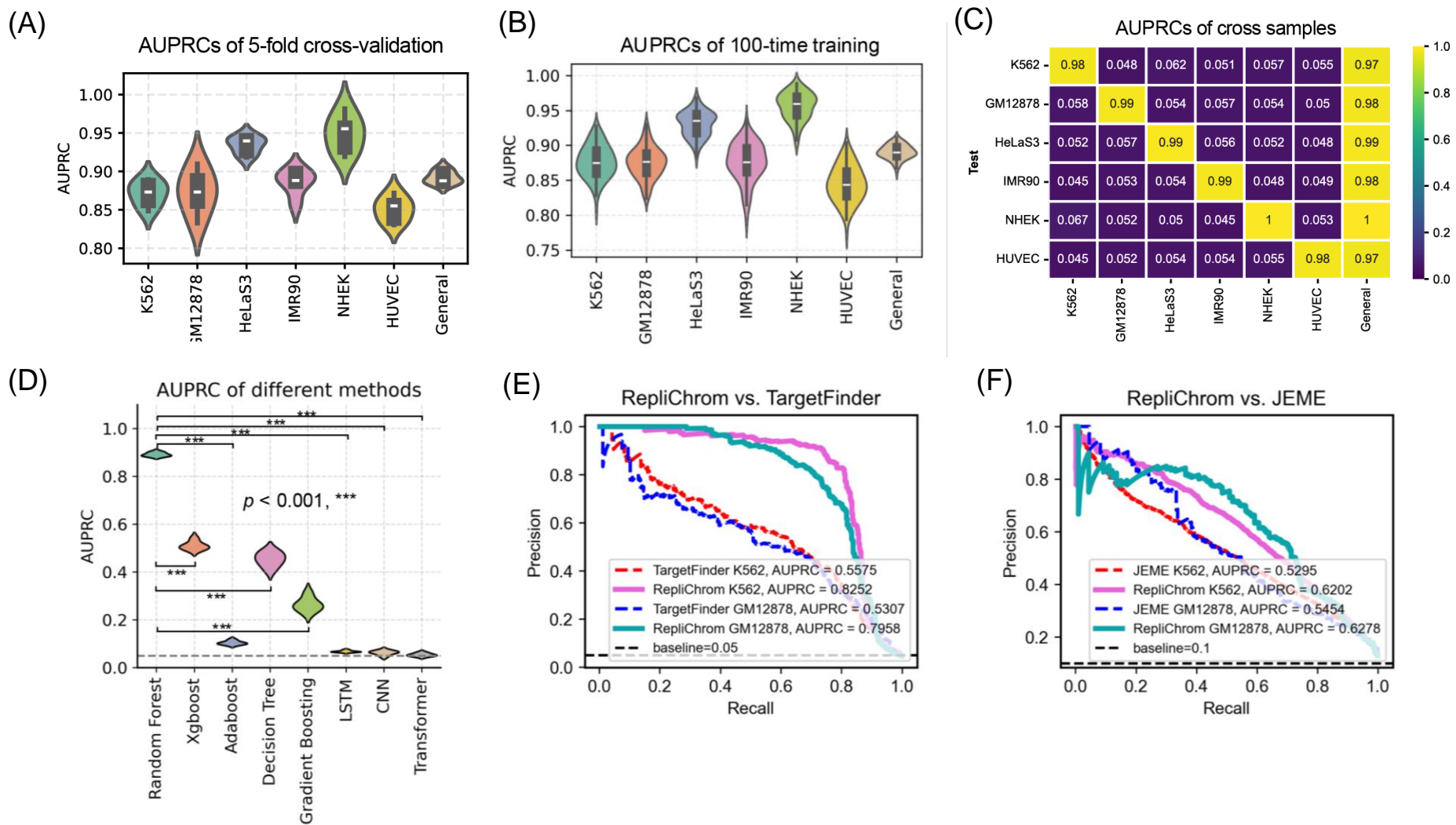


图5. RepliChrom模型能够在样本内和跨样本测试中准确预测增强子-启动子互作，且性能优于其他计算方法。



RepliChrom适用于不同测序平台获取的数据集

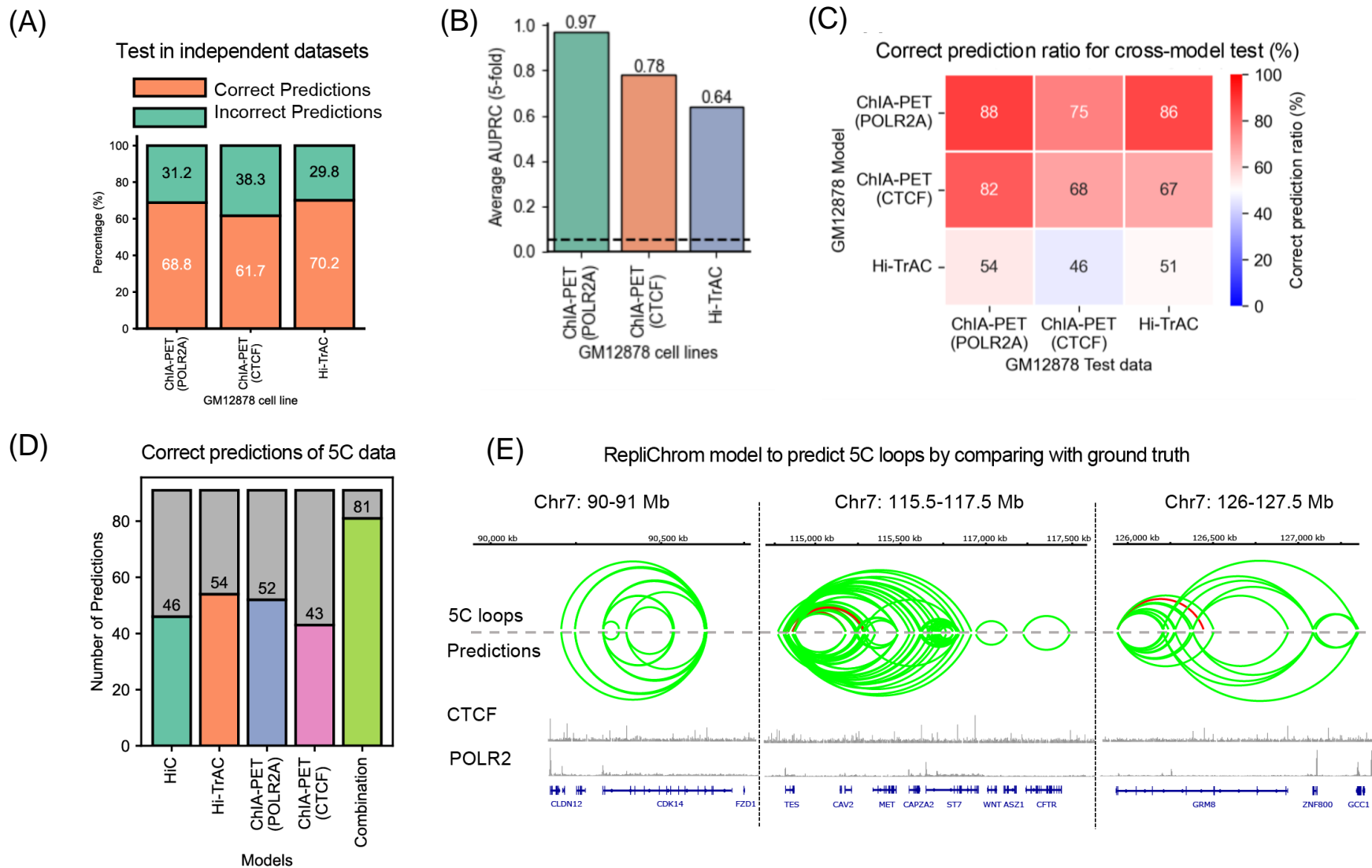


图6. RepliChrom在预测染色质互作环方面展现出良好的跨平台泛化能力。

RepliChrom 揭示白血病中的癌基因增强子-启动子互作

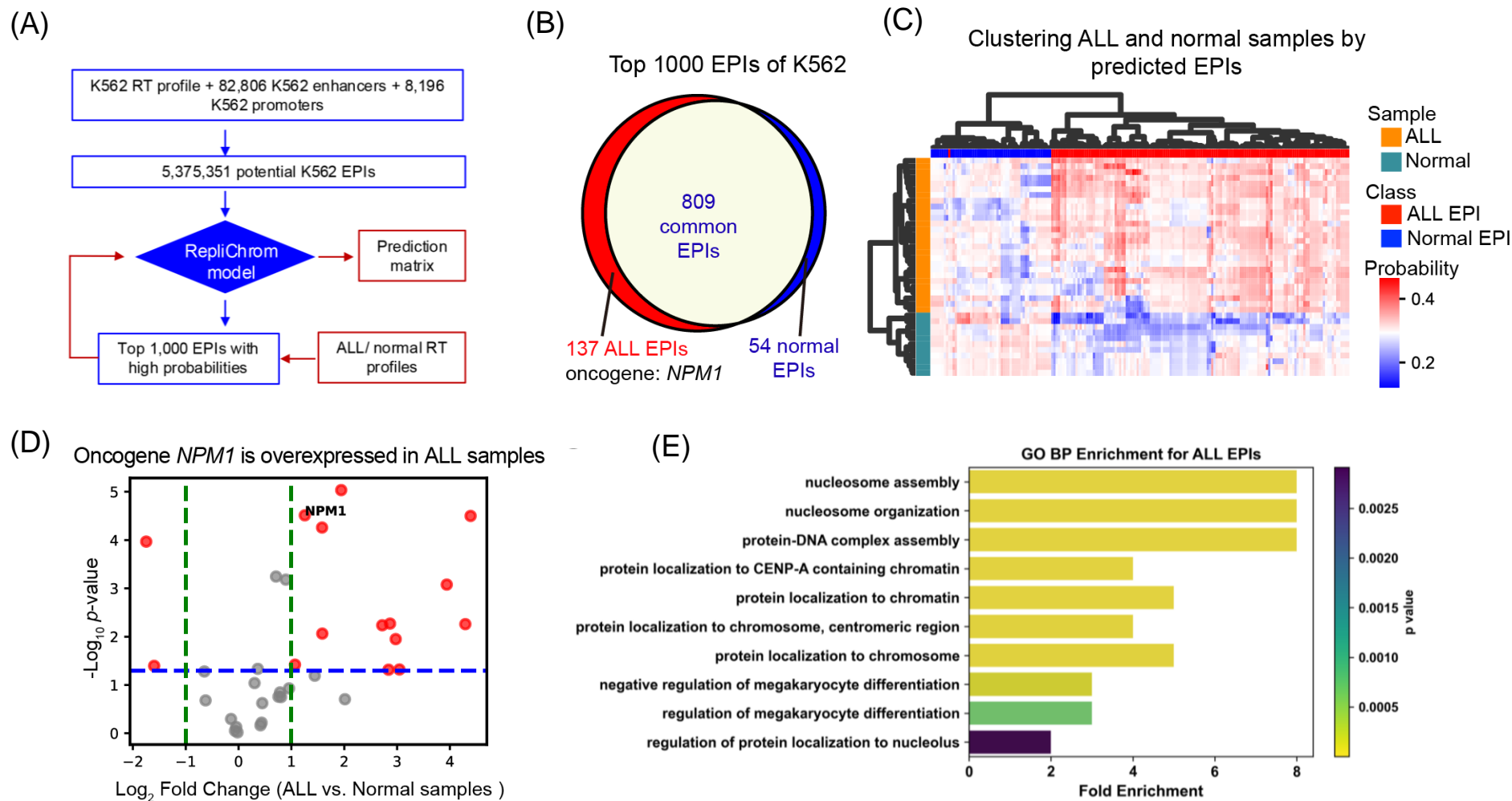
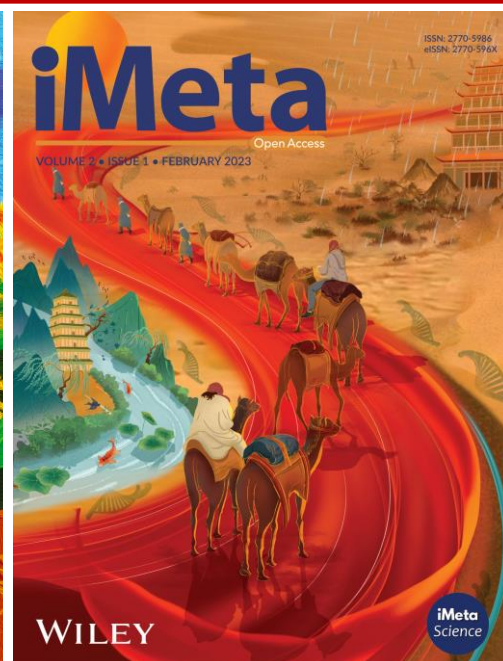
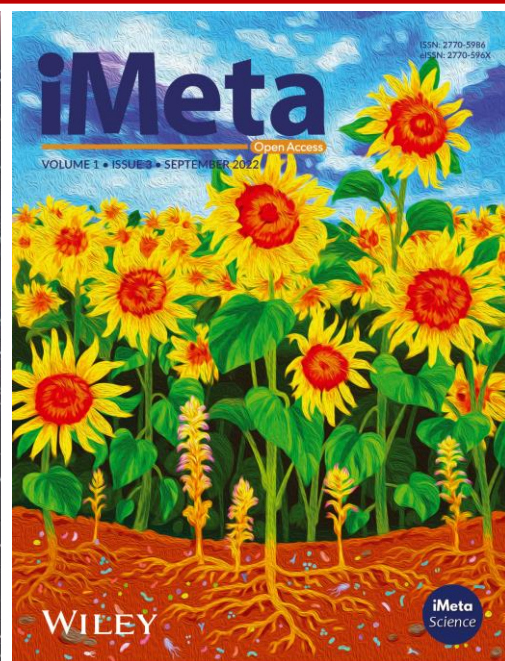


图7. RepliChrom模型应用于急性淋巴白血病样本中，发现癌基因*NPM1*存在于白血病特有的增强子-启动子互作中。



总结

- ❑ 本研究开发了一个可解释的机器学习框架RepliChrom，仅利用 DNA 复制时序信号预测三维基因组增强子-启动子互作；
- ❑ 启动子复制时序信号是增强子-启动子互作的关键预测因子；
- ❑ RepliChrom在多个平台数据集（如 Hi-C、Hi-TrAC、ChIA-PET、5C）上具有良好的泛化能力。
- ❑ RepliChrom能够预测癌症相关染色质互作结构，这可能代表一种新的表观遗传调控层面的标志物，为肿瘤生物标志物的发现提供新思路。



iMeta期刊(影响因子**23.8**)由宏科学、千名华人科学家和威立出版, 主编刘双江和傅静远教授。目标为生物/医学/环境综合期刊群(对标Cell/Nature/Science), 任何高影响力的研究、方法和综述均欢迎投稿, 重点关注技术、信息和组学等前沿交叉学科, 已被SCIE、PubMed等收录, IF 23.8位列全球SCI期刊前千分之五, 微生物学研究类全球第一, 中科院生物学双1区Top; 外审平均21天, 投稿至发表中位数57天。
子刊***iMetaOmics***(主编赵方庆和于君教授)、***iMetaMed*** 定位IF>10和15的综合、医学期刊, 欢迎投稿!



主页: <http://www.imeta.science>

出版社: <https://wileyonlinelibrary.com/journal/imeta>

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

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

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



office@imeta.science

imetaomics@imeta.science



宣传片



[iMeta](#)



更新日期
2025/5/21