ECCFP: a consecutive full pass based bioinformatic analysis for eccDNA identification from long-read sequencing data

Wang Li[1,2#], Biyuan Miao[1,3#], Jun Zhang[1#], Qingsong Zeng[1#], Tangxuan Zhang[1#], Zetong Wu[1], Yusheng Song[4], Mengting Li[5], Lilu Guo[6], Juan Luo[3], Jingxiang Xu[3], Tianliang Liu[7], Shifu Chen[7], Jia Gu[8], Shaogui Wan[1*]

1 Institute of Genomics and Precision Medicine, School of Medical Technology, Gannan Medical University, Ganzhou 341000, China
2 School of Medical and Information Engineering, Gannan Medical University, Ganzhou 341000, China
3 School of Basic Medicine, Gannan Medical University, Ganzhou 341000, China
4 Department of Interventional Radiology, Ganzhou Hospital-
5 Institute of Neurology, Jiangxi Academy of Clinical Medical Sciences, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, China
6 Department of Computer Science, Xidian University, Xi'an, China
7 LifeX Institute, School of Medical Technology, Gannan Medical University, Ganzhou 341000, China

Wang Li, Biyuan Miao, Jun Zhang, Qingsong Zeng, Tangxuan Zhang, Zetong Wu, Yusheng Song, et al. 2026. ECCFP: a consecutive full pass based bioinformatic analysis for eccDNA identification from long-read sequencing data. ***iMetaOmics*** 2: e70080. https://doi.org/10.1002/imo2.70080

Introduction

Extrachromosomal Circular DNA (eccDNA)

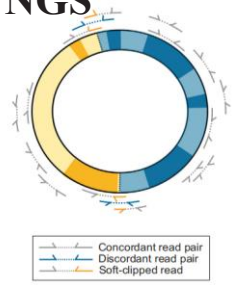
eccDNA is a circular DNA molecule that exists independently of chromosomes

Functions of eccDNA

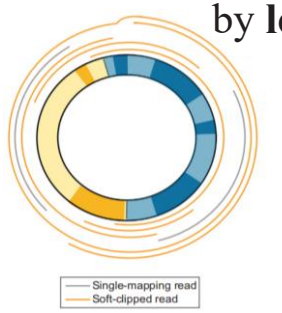
- Drives copy-number variation and gene rearrangements
- Functioning as mobile regulatory elements that involves gene expression regulation
- Plays roles in precision medicine, crop improvement, aging assessment, immunomodulation, and prenatal screening

eccDNA has the potential to be a biomarker for liquid biopsy and monitoring

eccDNA detection by **NGS**



eccDNA detection by **long-read sequencing**



Noer et al. trends in cancer 2022

Current eccDNA identification primarily relies on **high-throughput sequencing technologies**

NGS-Based Detection Technology

Main approach: Relies on indirect inference using split reads and discordant reads spanning junction sites

Key bottleneck: Difficulty in accurately identifying eccDNA derived from repetitive sequences or with chimeric structures due to short read lengths.

TGS-Based Detection Technology

Technical advantage: Long reads enable direct resolution of the complete molecular structure of eccDNA

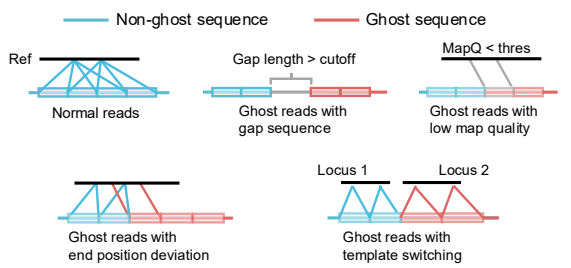
Current challenge: Existing TGS-based bioinformatics tools still have significant shortcomings in detection sensitivity, false-positive control, and result accuracy

Highlights

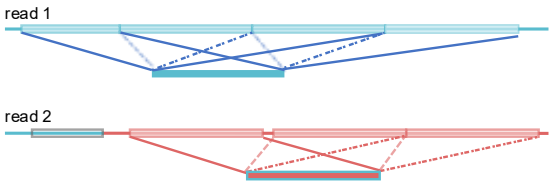
ECCFP: a consecutive full pass based bioinformatic analysis for eccDNA identification from long-read sequencing data

Development of ECCFP

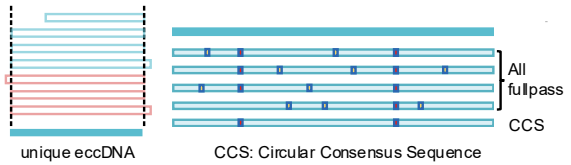
Category of ghost sequence



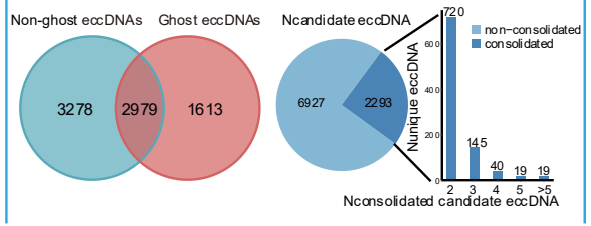
Identification of candidate eccDNA



Generation of unique eccDNA and CCS

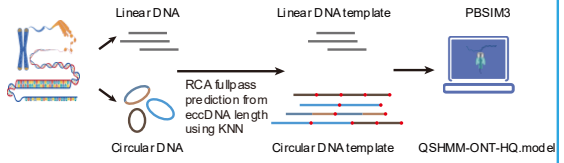


Characterization of eccDNA

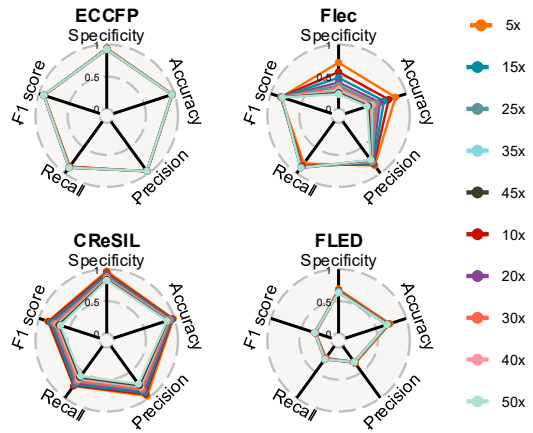


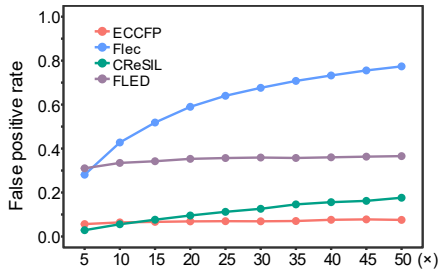
Performance Evaluation of ECCFP using Simulated Data

Construction of simulated data



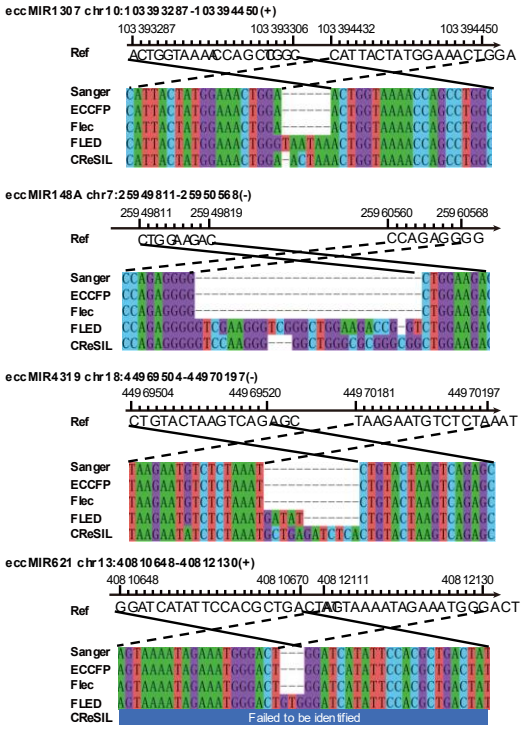
Evaluation of multi-dimensional performance using simulated data



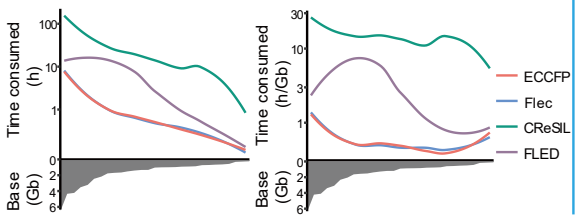


Performance Evaluation of ECCFP using Real Sequencing Data

Validation of the junction sequence using spike-in eccMIRs



Evaluation of the running efficiency using publicly available datasets



- ECCFP is an innovative tool that utilizes all consecutive full passes from individual read and the Boyer-Moore voting algorithm to improve the eccDNA detection
- ECCFP exhibits advantages in overall performance, sensitivity, and accuracy with high runtime efficiency, compared to other pipelines.
- ECCFP is an open-source project available at https://github.com/WSG-Lab/ECCFP, featuring easy installation and comprehensive usage tutorials.

Workflow of ECCFP

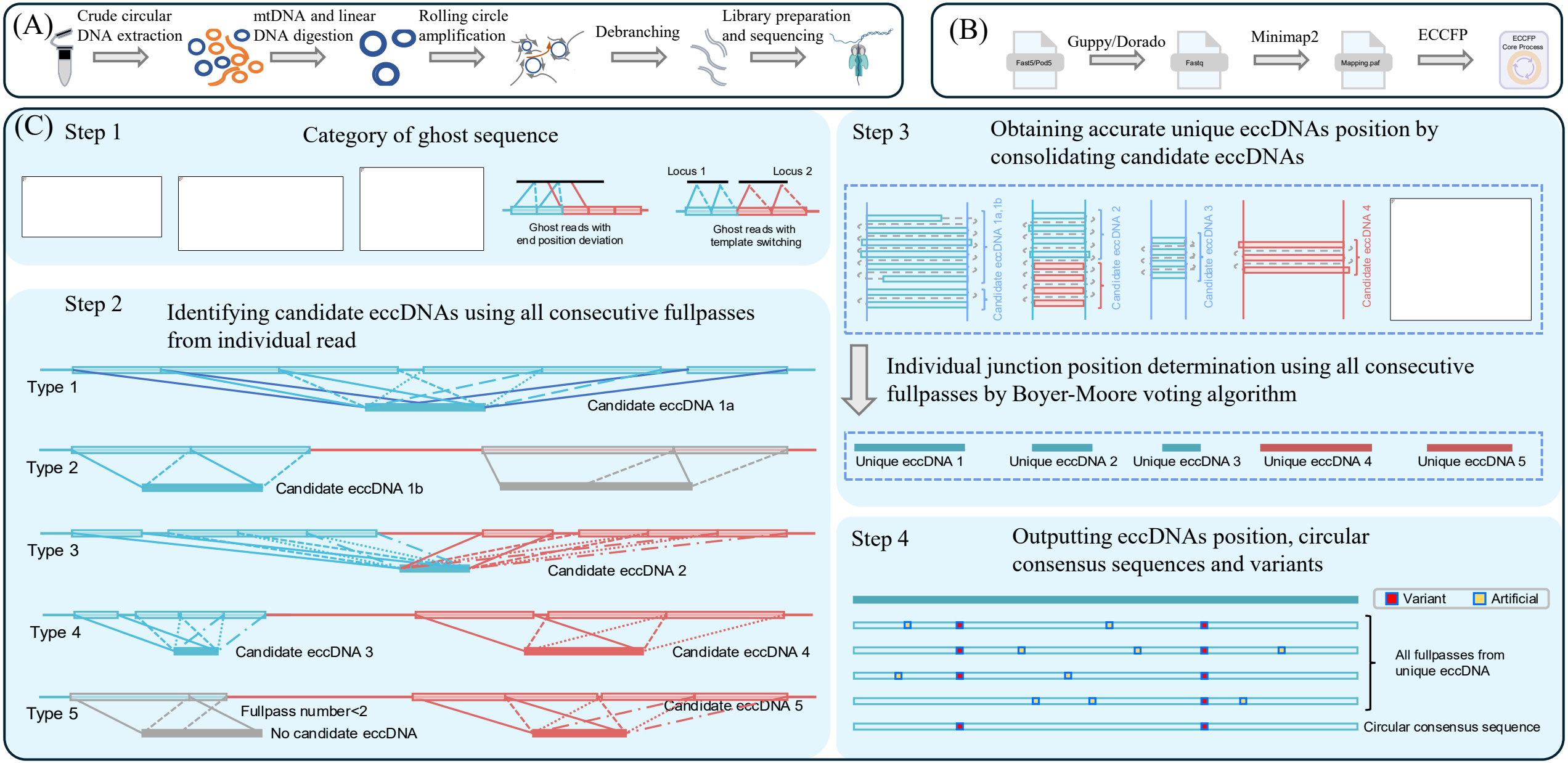


Figure 1. The workflow of eccDNA identification by ECCFP

ECCFP increases the number of eccDNA detection by including ghost sequences

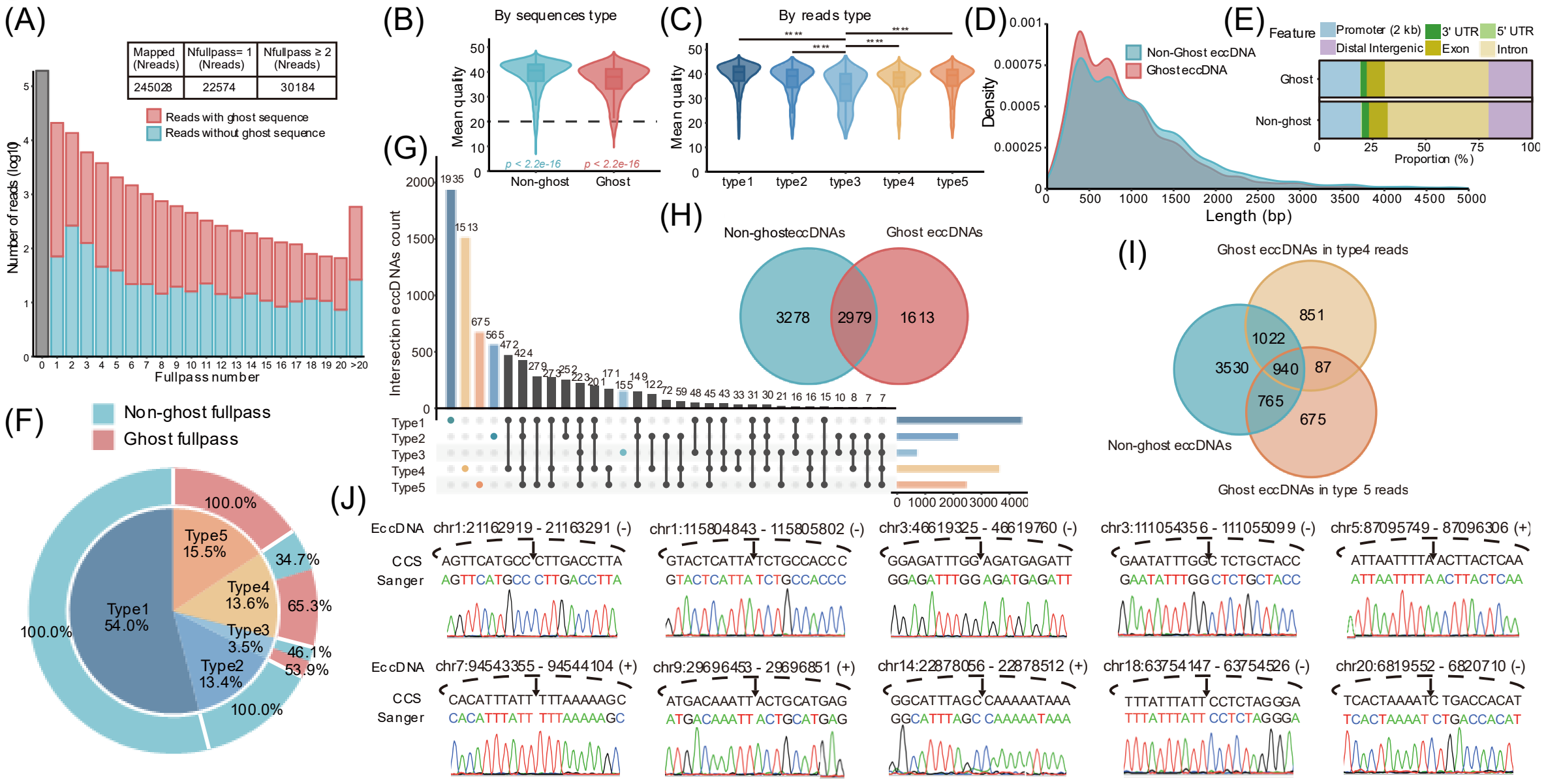


Figure 2. The number of eccDNA detection by ECCFP

ECCFP improves the position of eccDNA detection by Boyer-Moore voting algorithm

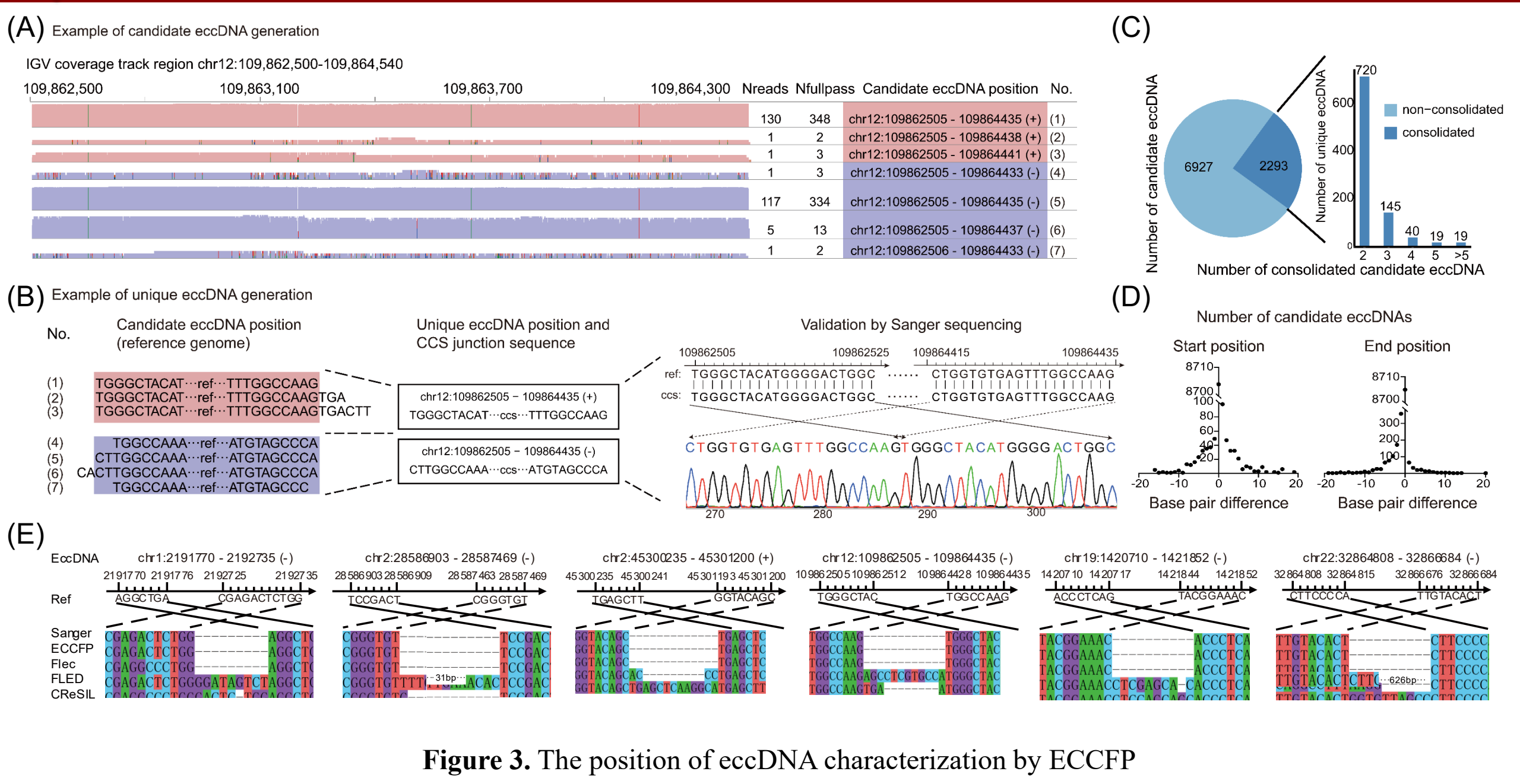


Figure 3. The position of eccDNA characterization by ECCFP

Performance evaluation using spike-in synthetic circular DNA

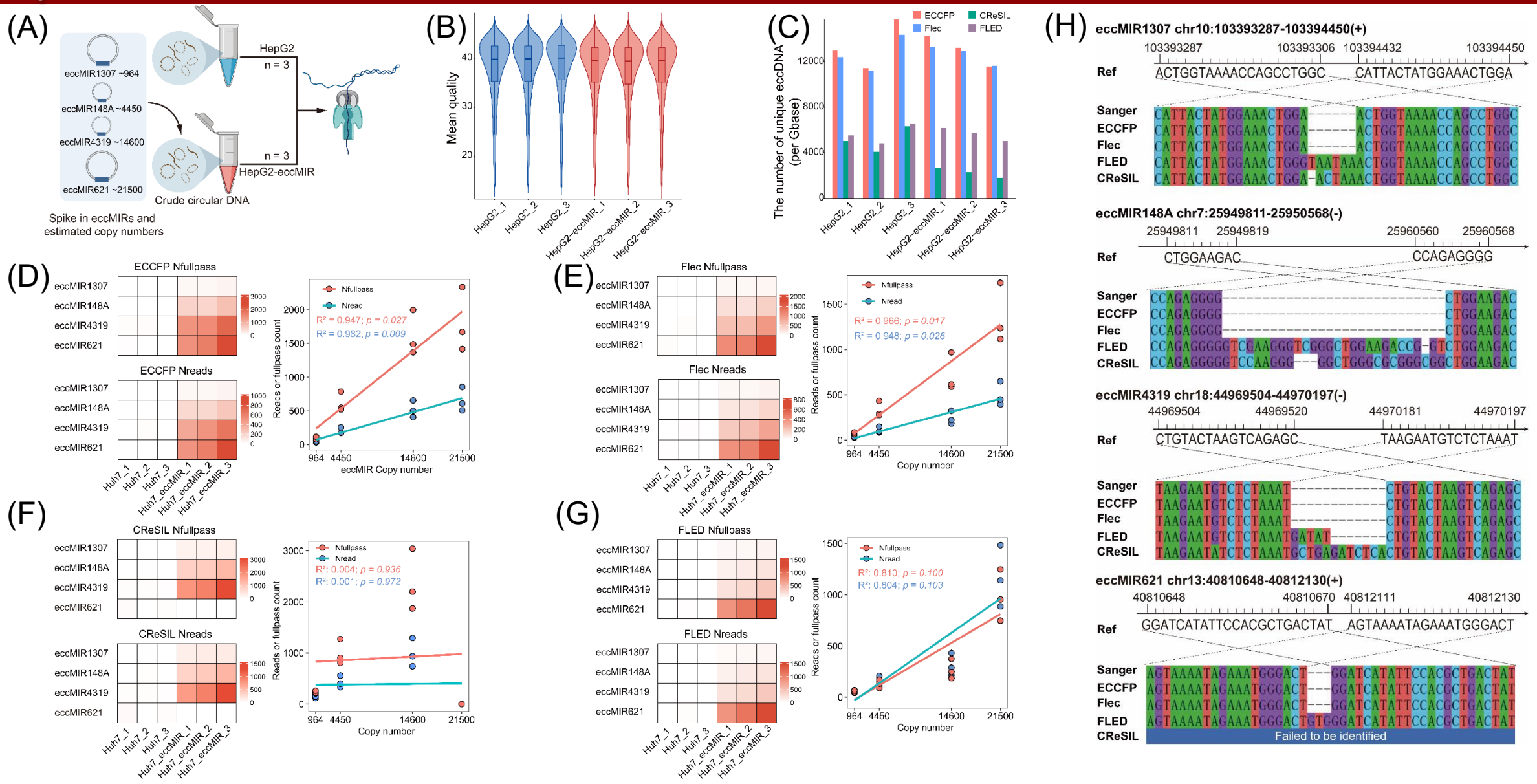


Figure 4. Evaluation of eccDNA detection by ECCFP using spiking in artificially synthesized circular DNA

Performance evaluation using simulated datasets

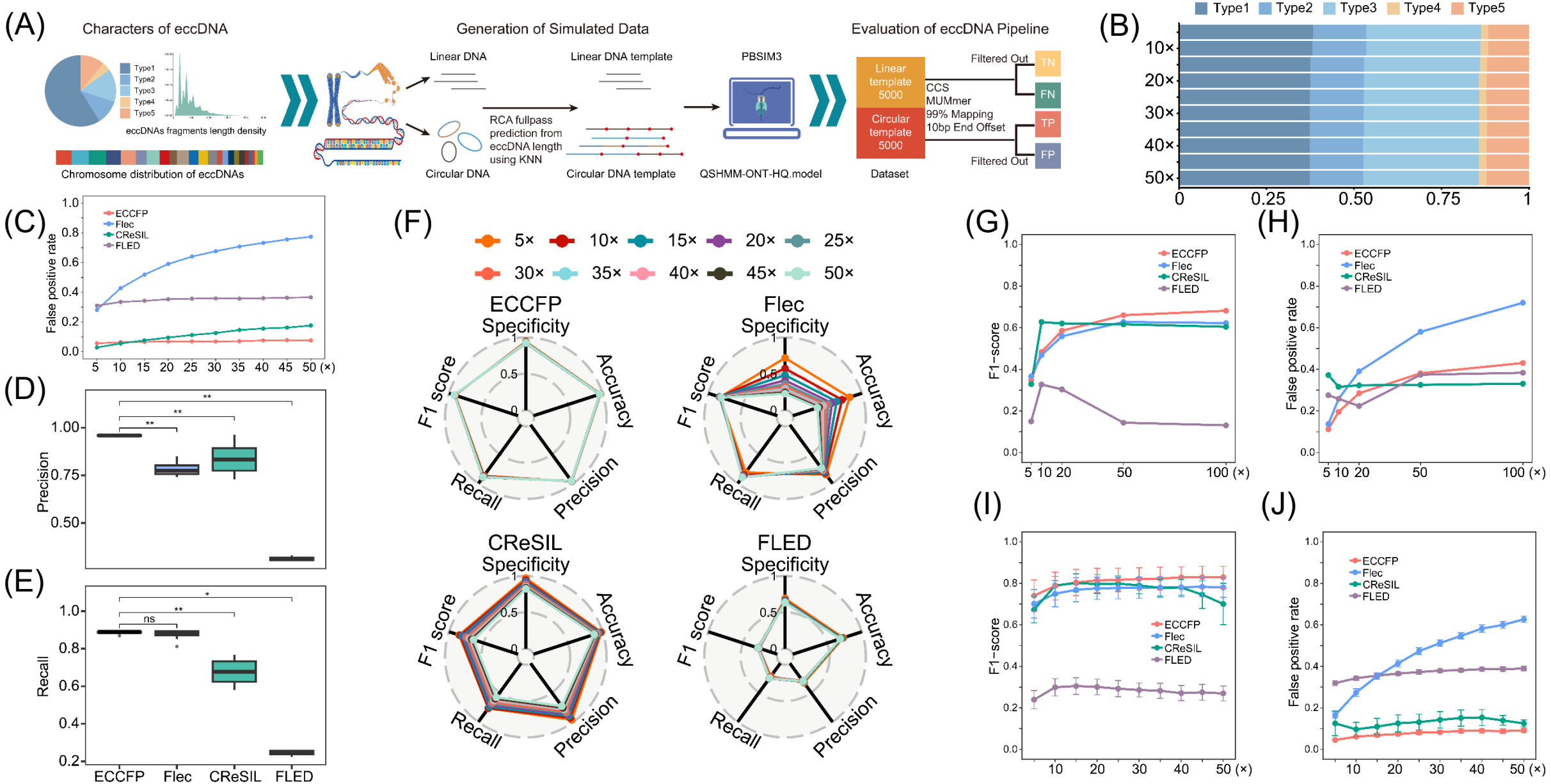


Figure 5. The performance evaluation of ECCFP using simulated data

Performance evaluation using public sequencing datasets

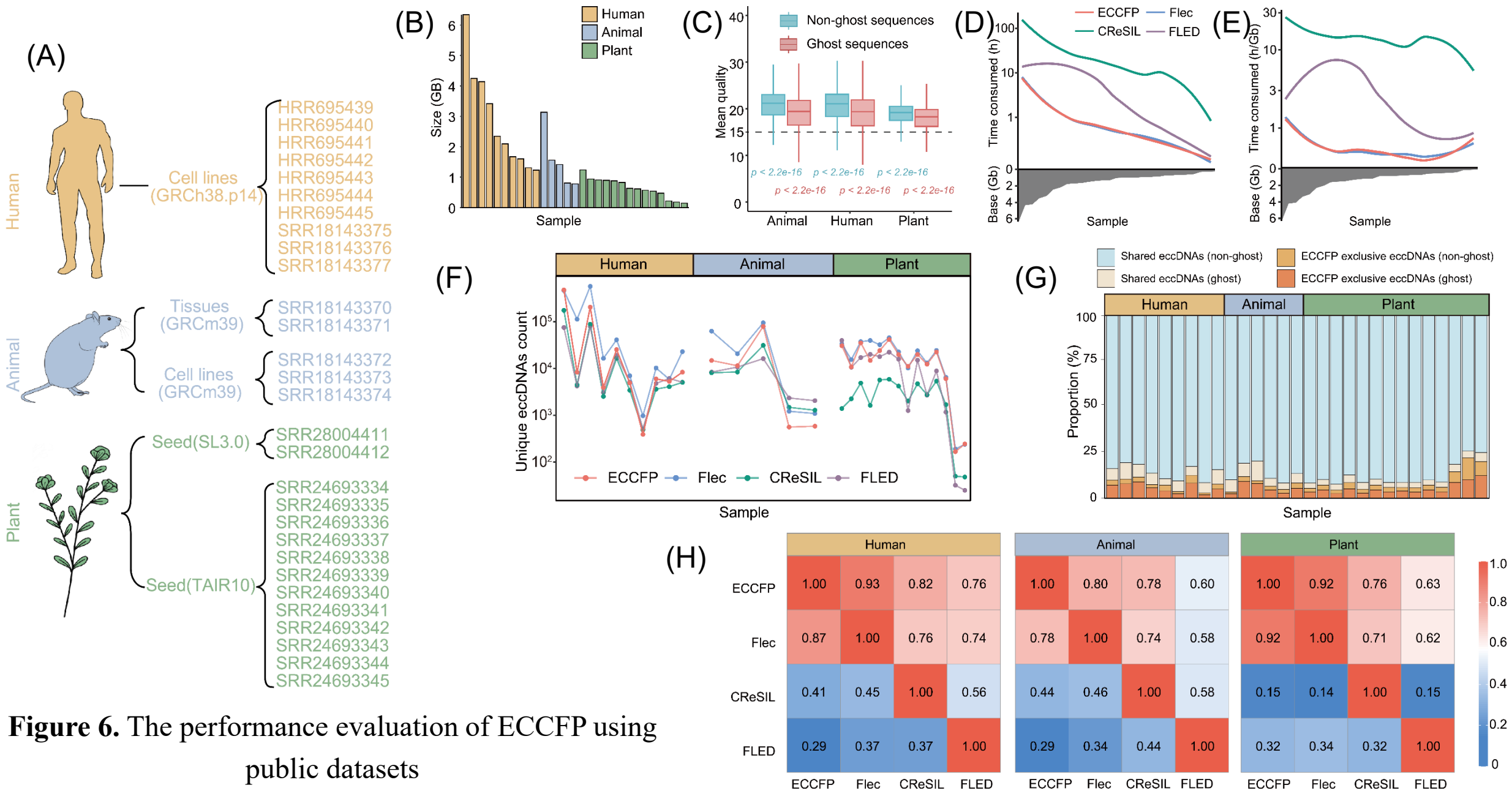


Figure 6. The performance evaluation of ECCFP using public datasets

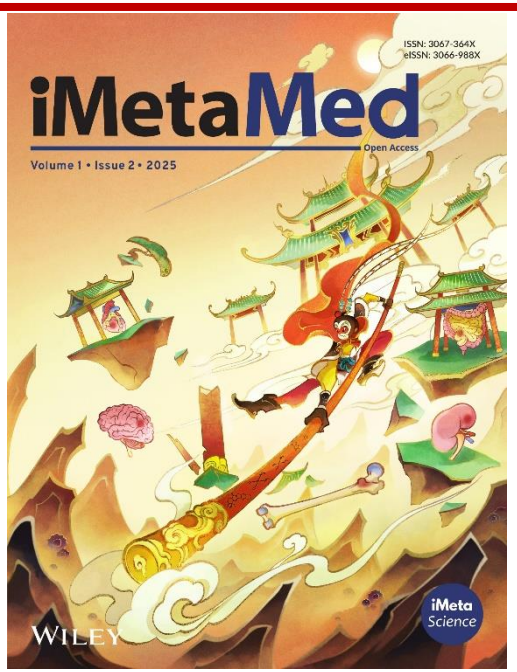
Summary

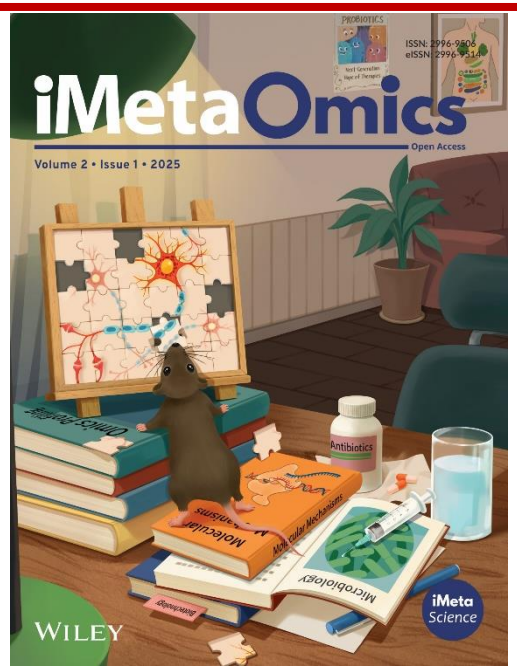
- This study presents **ECCFP**, a novel TGS-based method for eccDNA detection.
- ECCFP employs a **voting algorithm** to integrate support from multiple reads, achieving **precise localization of eccDNA junction sites** and effectively reducing false positives **without significantly increasing algorithmic complexity.**
- Evaluations on **simulated datasets, spike-in experimental datasets, and multi-species public datasets** demonstrate that ECCFP holds **clear advantages in sensitivity, accuracy, and runtime efficiency.**
- ECCFP is an **open-source and user-friendly** bioinformatics pipeline, poised to advance research on eccDNA as a clinical biomarker.

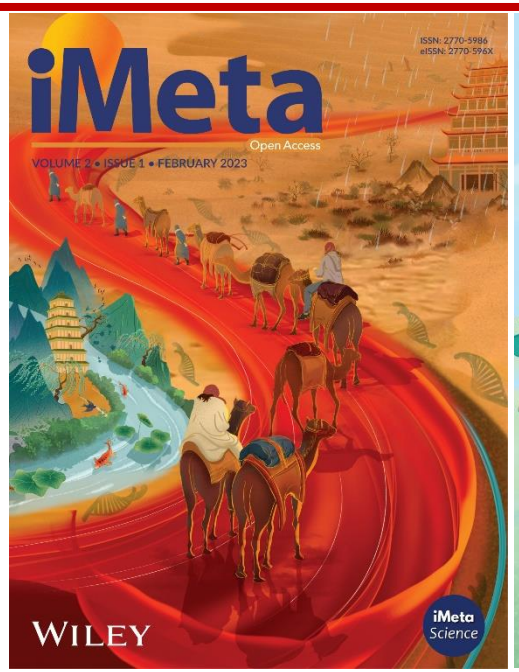
Wang Li, Biyuan Miao, Jun Zhang, Qingsong Zeng, Tangxuan Zhang, Zetong Wu, Yusheng Song, et al. 2026. ECCFP: a consecutive full pass based bioinformatic analysis for eccDNA identification from long-read sequencing data. ***iMetaOmics*** 2: e70080. https://doi.org/10.1002/imo2.70080

iMeta: To be top journals in biology and medicine

WILEY







"***iMeta***" launched in 2022 by iMeta Science Society, impact factor (IF) **33.2**, ranking top 65/22249 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 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
imetaomics@imeta.science

Promotion Video

Update 2025/7/6