



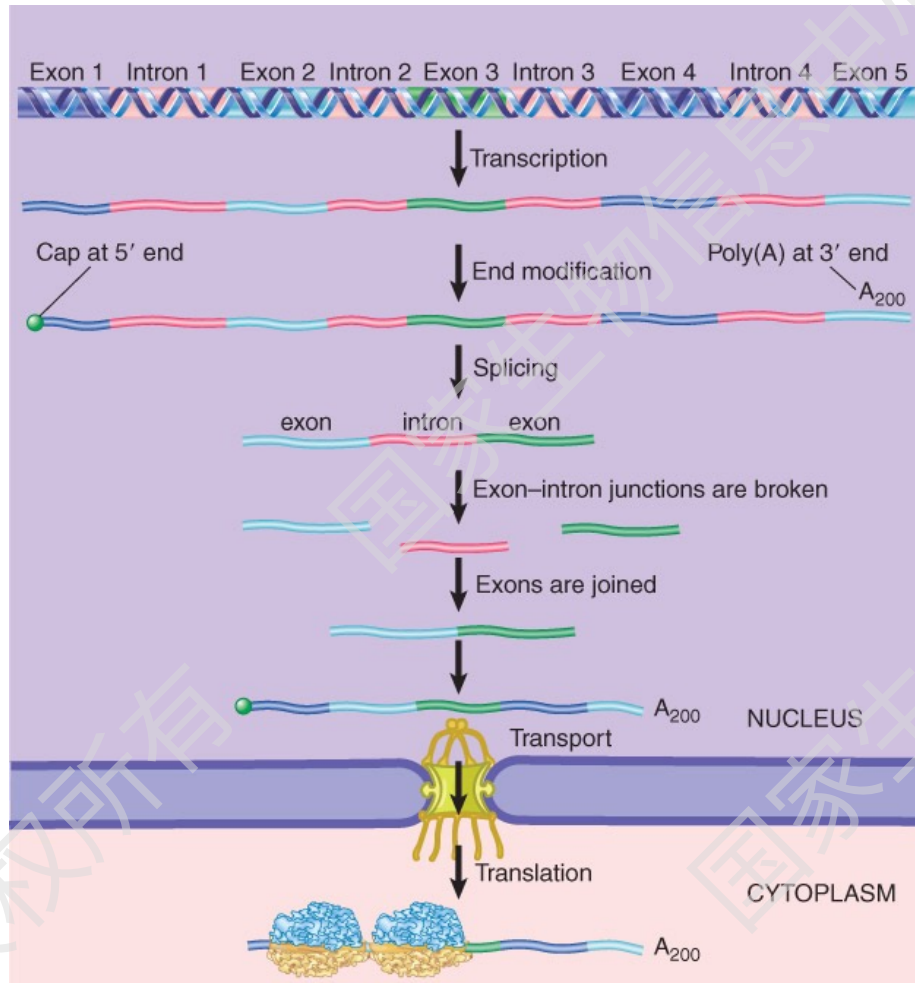
中国科学院北京基因组研究所 (国家生物信息中心)
BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES / CHINA NATIONAL CENTER FOR BIOINFORMATION

Detection and prediction of splicing regulation and full-length transcript isoforms

高远

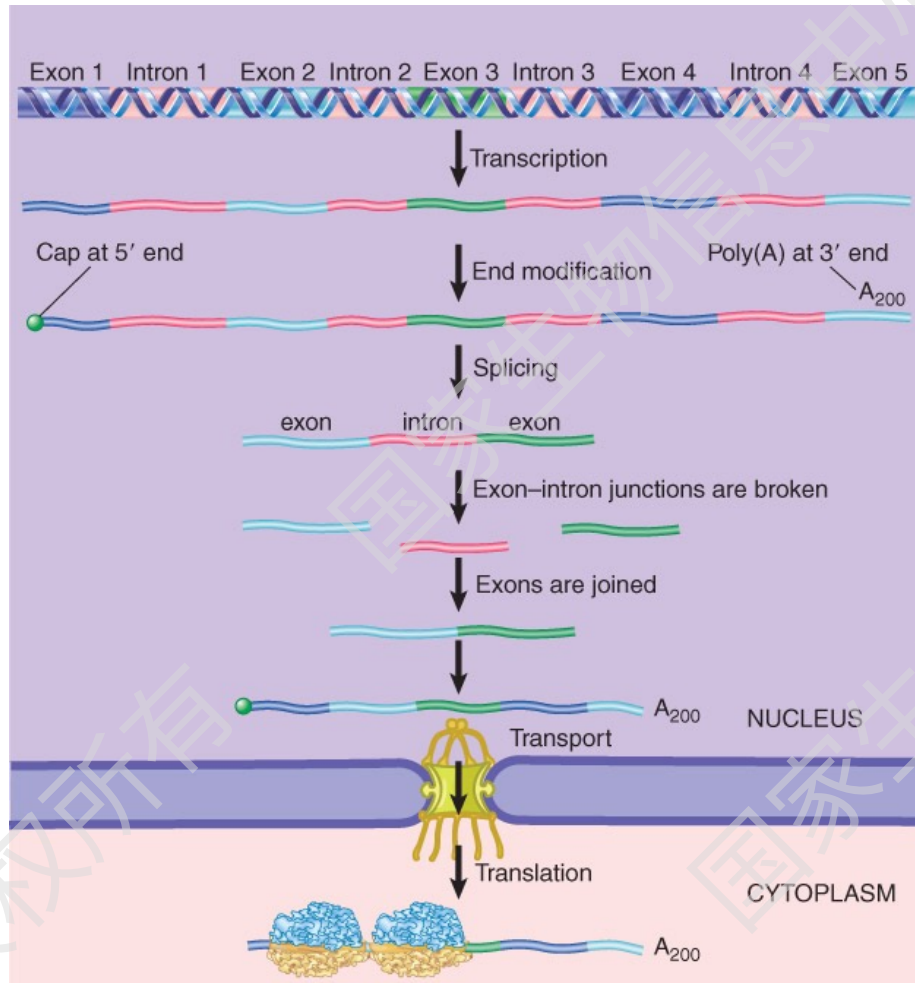
2025.11.21

Background: RNA splicing in eukaryotes

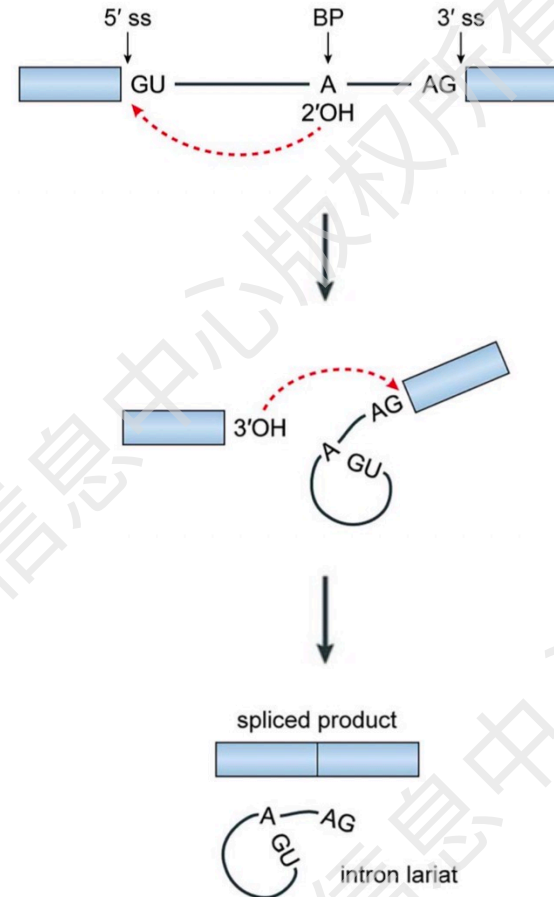


Lewin's GENES X

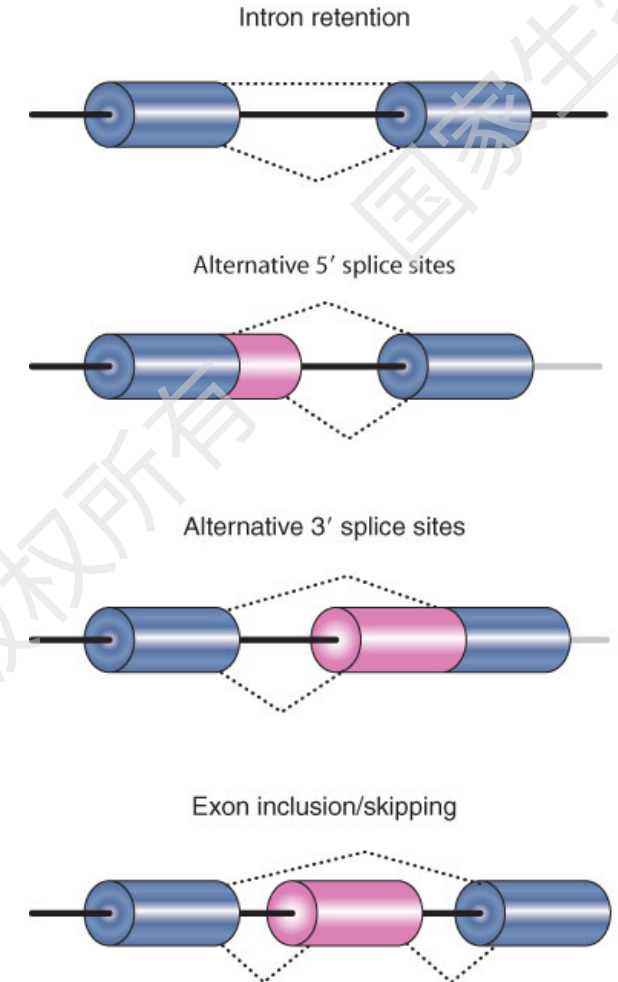
RNA splicing in eukaryotes: alternative splicing



Lewin's GENES X

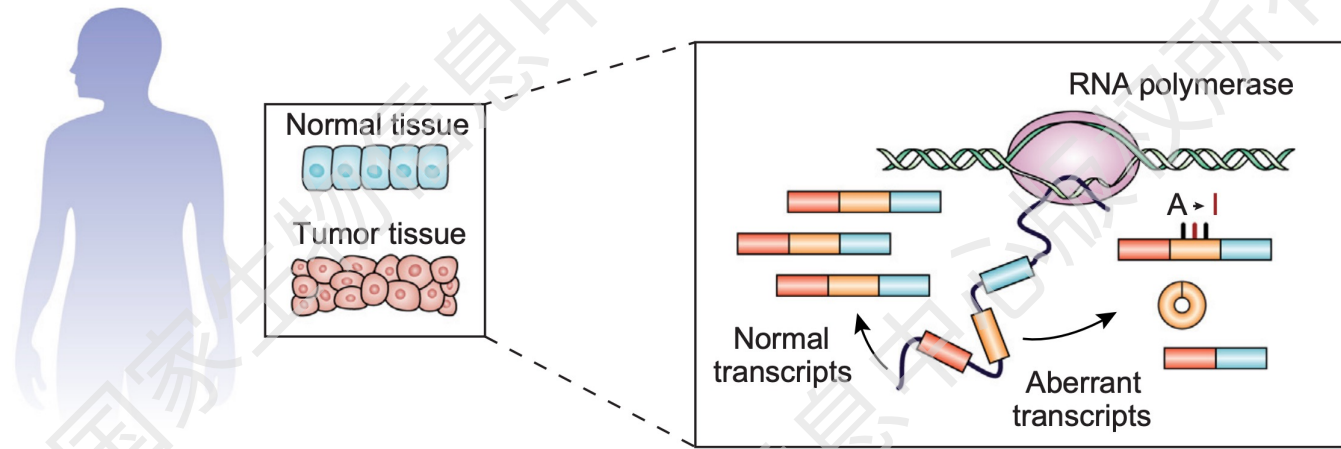


Wang et al. *Brief. Bioinform.*, 2023



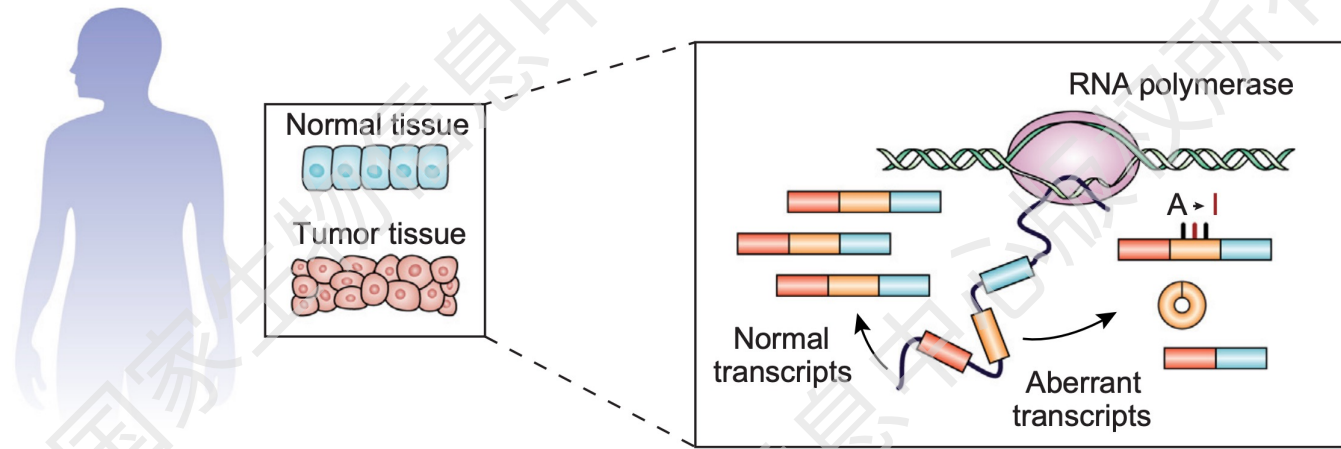
Lewin's GENES X

Alternative splicing events can be identified and quantified by mapping short-read data to the reference genome

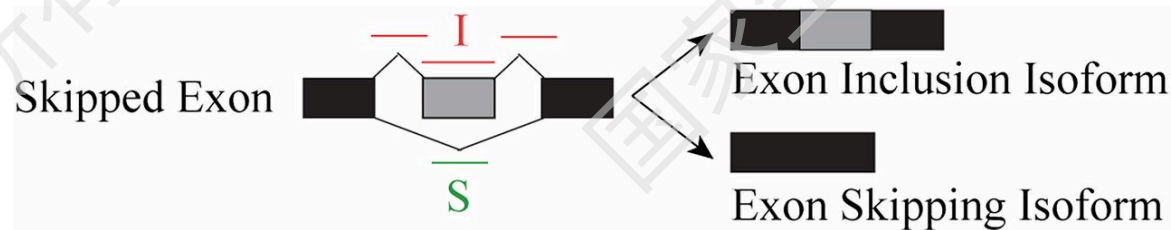


Pan et al. *Trends Pharmacol. Sci.*, 2021

Alternative splicing events can be identified and quantified by mapping short-read data to the reference genome



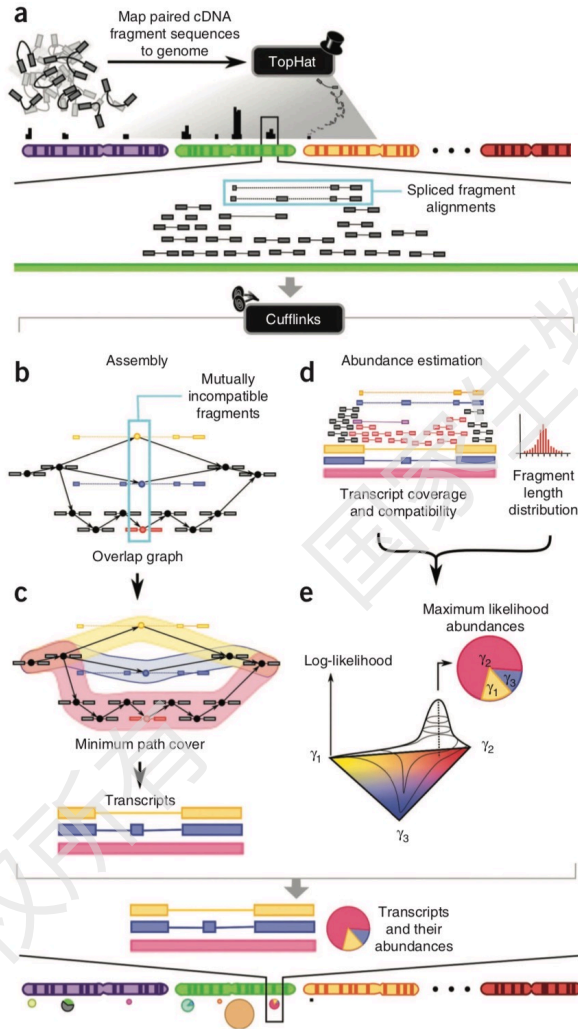
Pan et al. *Trends Pharmacol. Sci.*, 2021



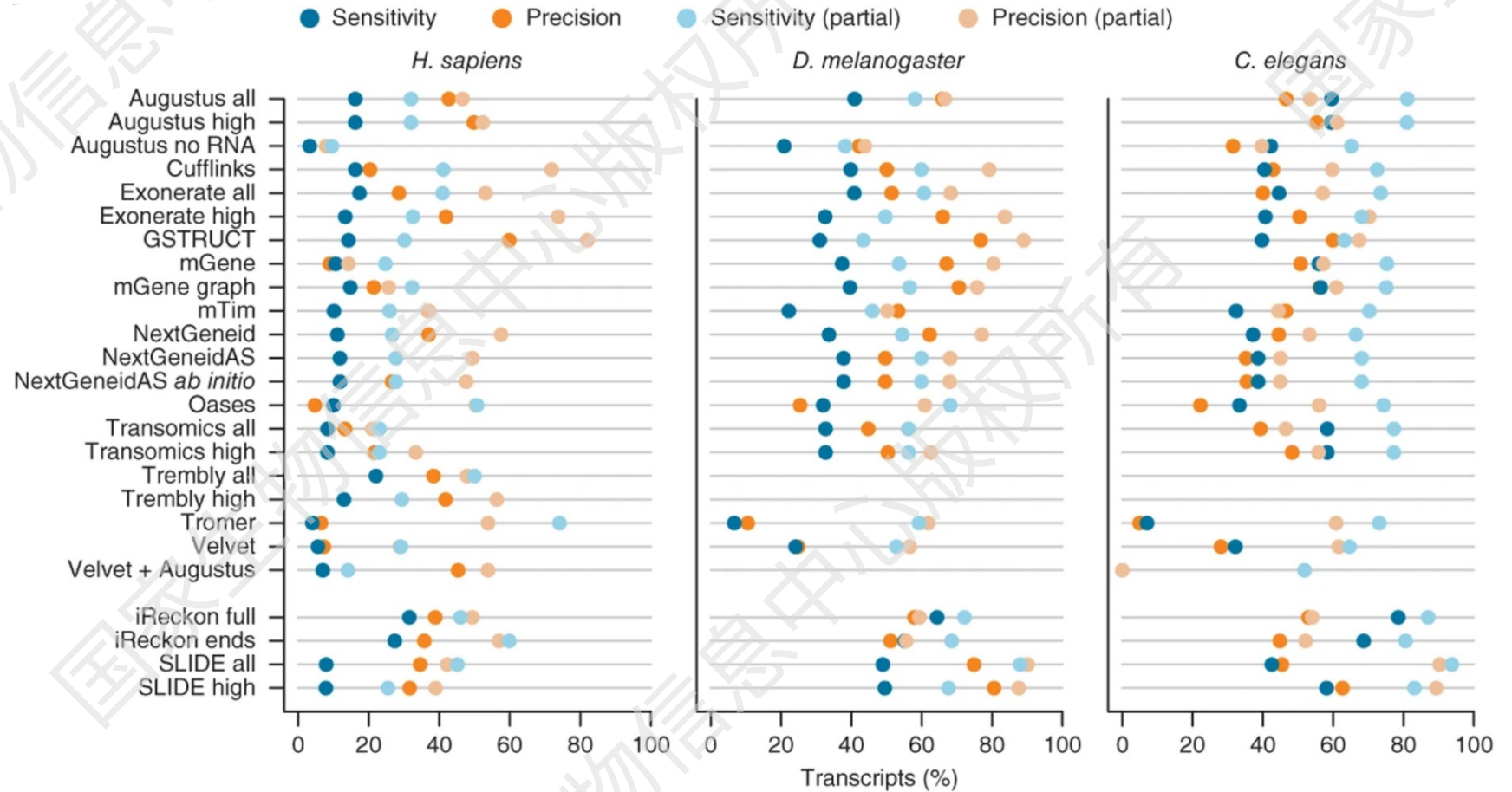
Percent-splice-in (PSI) in rMATS:

$$\hat{\psi} = (I/l_I) / (I/l_I + S/l_S)$$

Short-read RNA-seq is inherently limited in its ability to discover and quantify transcript isoforms

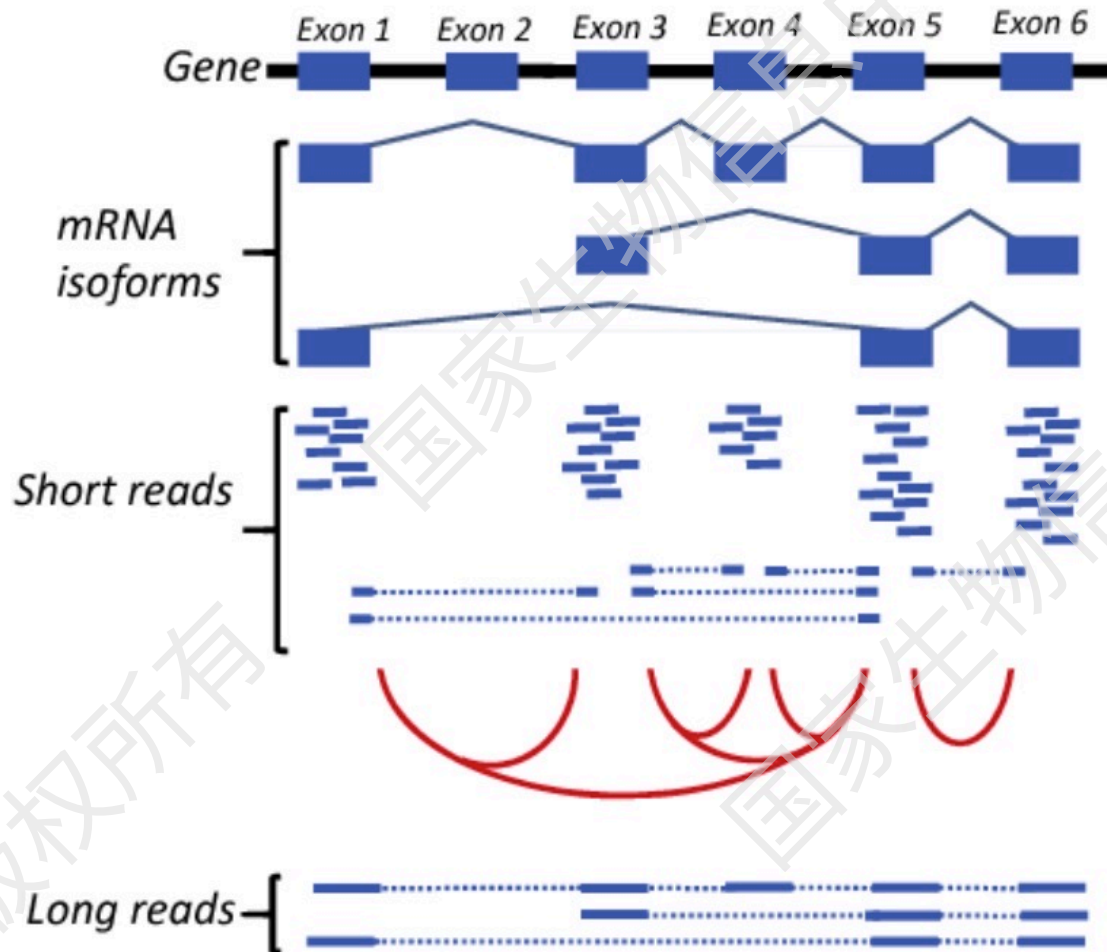


Trapnell *et al. Nat. Biotechnol.*, 2010



Steijger *et al. Nat. Methods*, 2013

Advantages and challenges of long-read data in full-length isoform analysis



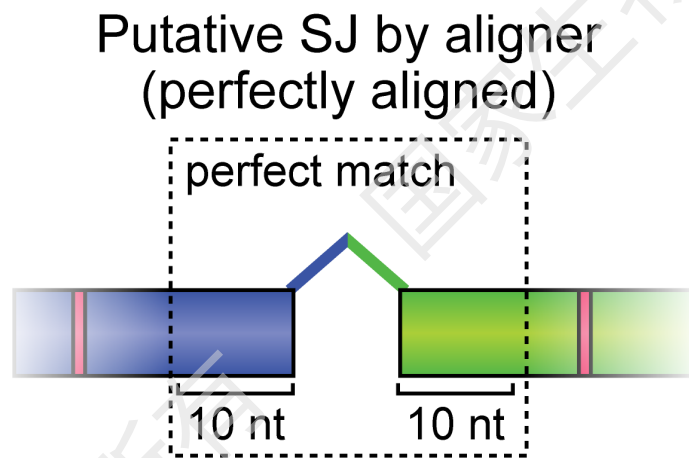
van Dijk et al. *Trends Genet.*, 2018

Table 1. Evaluation of junction accuracy on 2D ONT reads

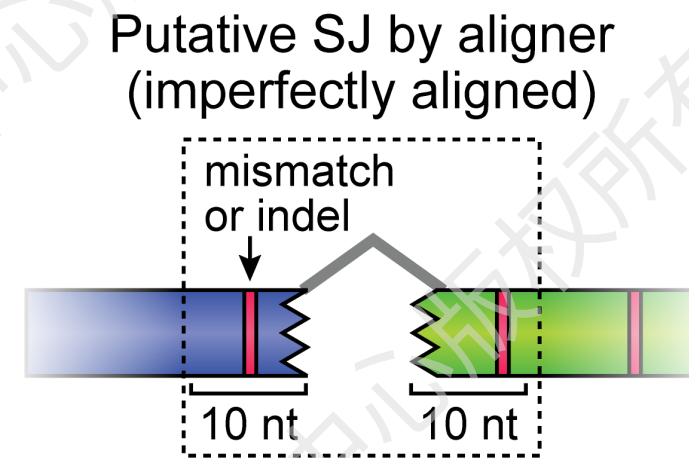
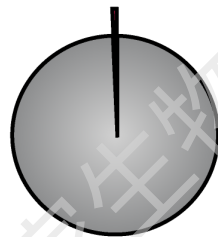
	GMAP	minimap2	SpAln	STAR
Run time (CPU min)	631	15.9	2076	33.9
Peak RAM (GByte)	8.9	14.5	3.2	29.2
# aligned reads	103 669	104 199	103 711	26 479
# chimeric alignments	1904	1488	0	0
# non-spliced alignments	15 854	14 798	17 033	10 545
# aligned introns	692 275	693 553	692 945	78 603
# novel introns	11 239	3113	8550	1214
% exact introns	83.8	94.0	87.9	55.2
% approx. introns	91.8	96.9	92.5	82.4

Li et al. *Bioinformatics*, 2018

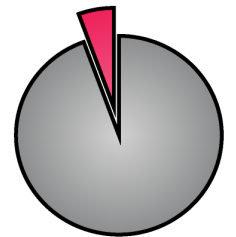
How to know whether a putative SJ is reliable or not?



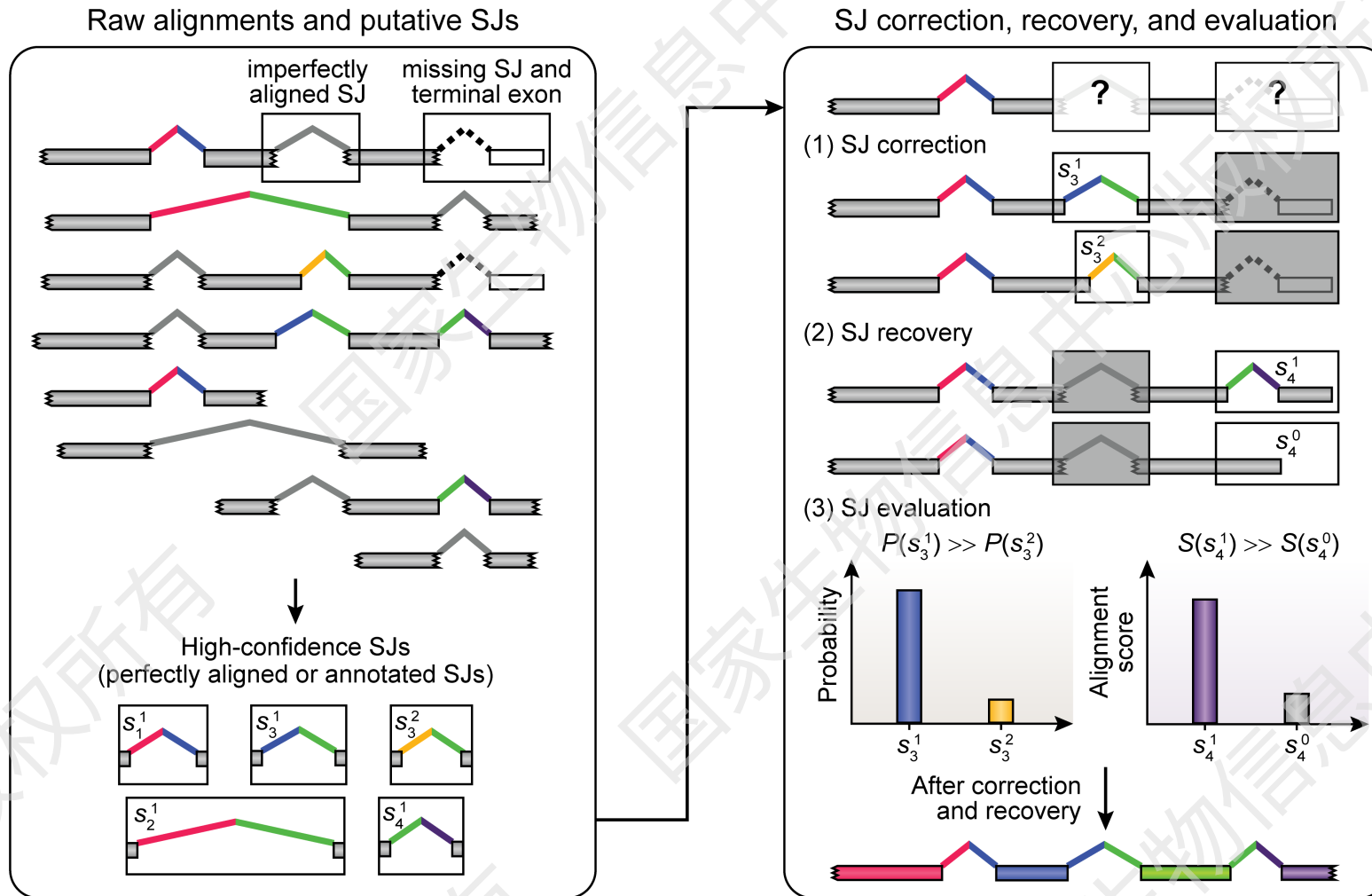
Incorrect SJ:
0.18-0.24%



Incorrect SJ:
8.8-13.0%



Correction of splice junctions for full-length isoform analysis



multivariate hypergeometric distribution

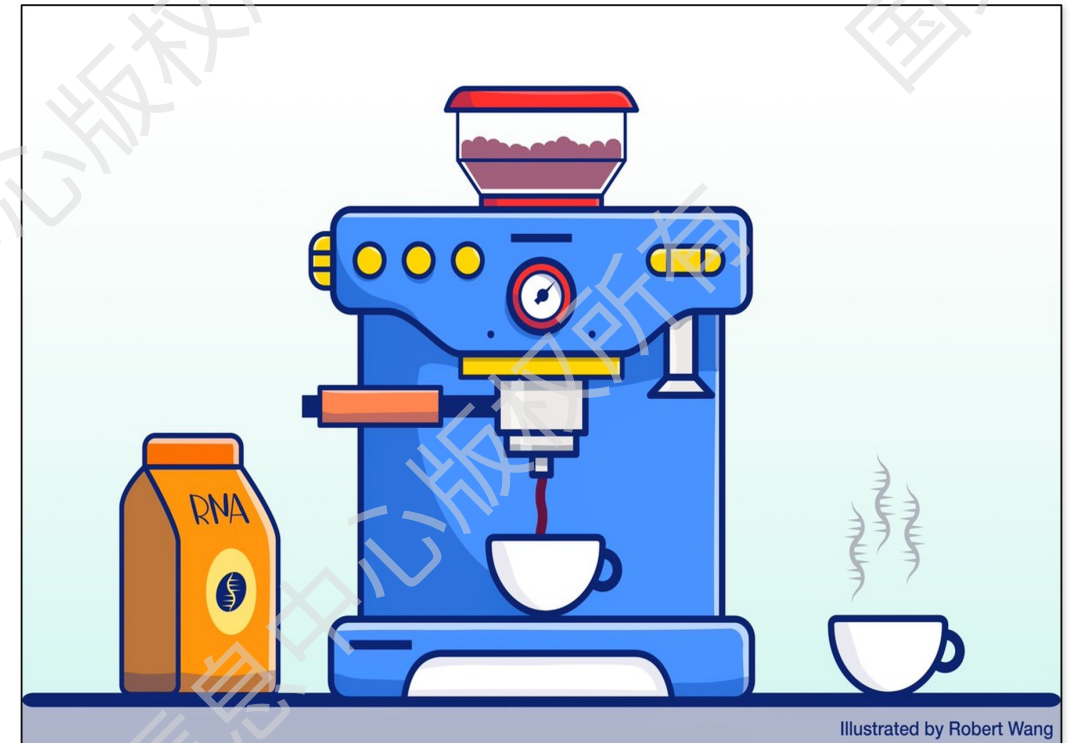
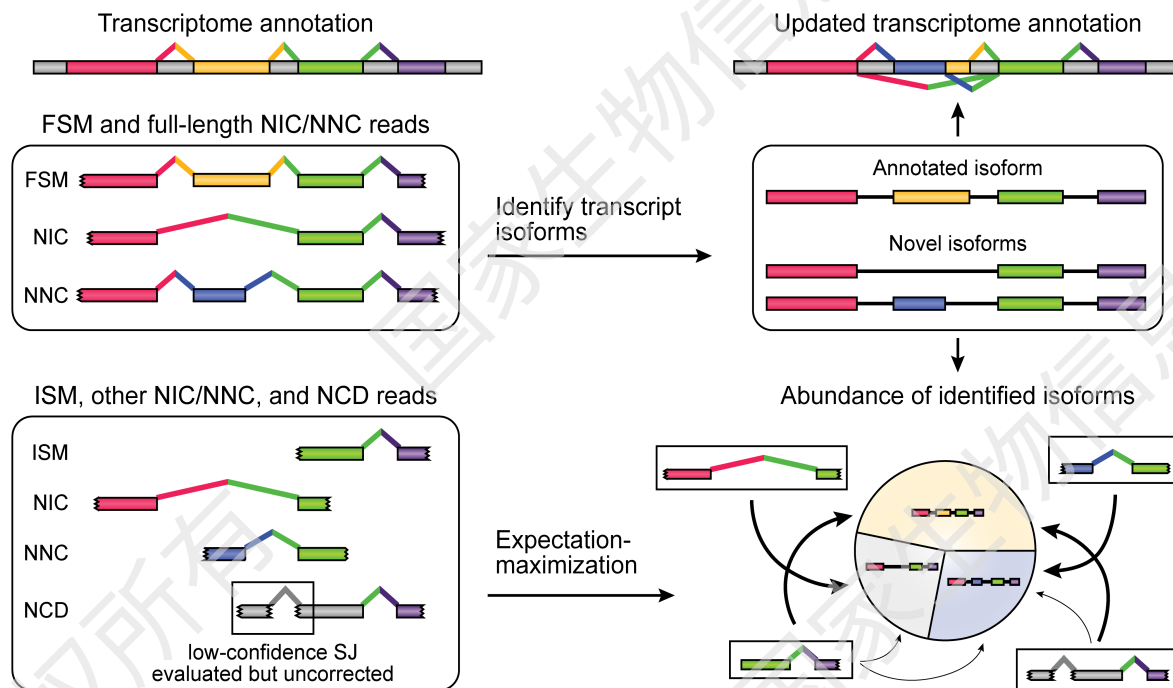
$$P(s_i^l) = \frac{\prod_j \binom{K_j + k_{ij}^l}{k_{ij}^l}}{\binom{\sum_j (K_j + k_{ij}^l)}{\sum_j k_{ij}^l}}$$

j : mismatch, insertion, deletion, or match

k : No. of the alignment type within a window of 20 nt surrounding the candidate SJ

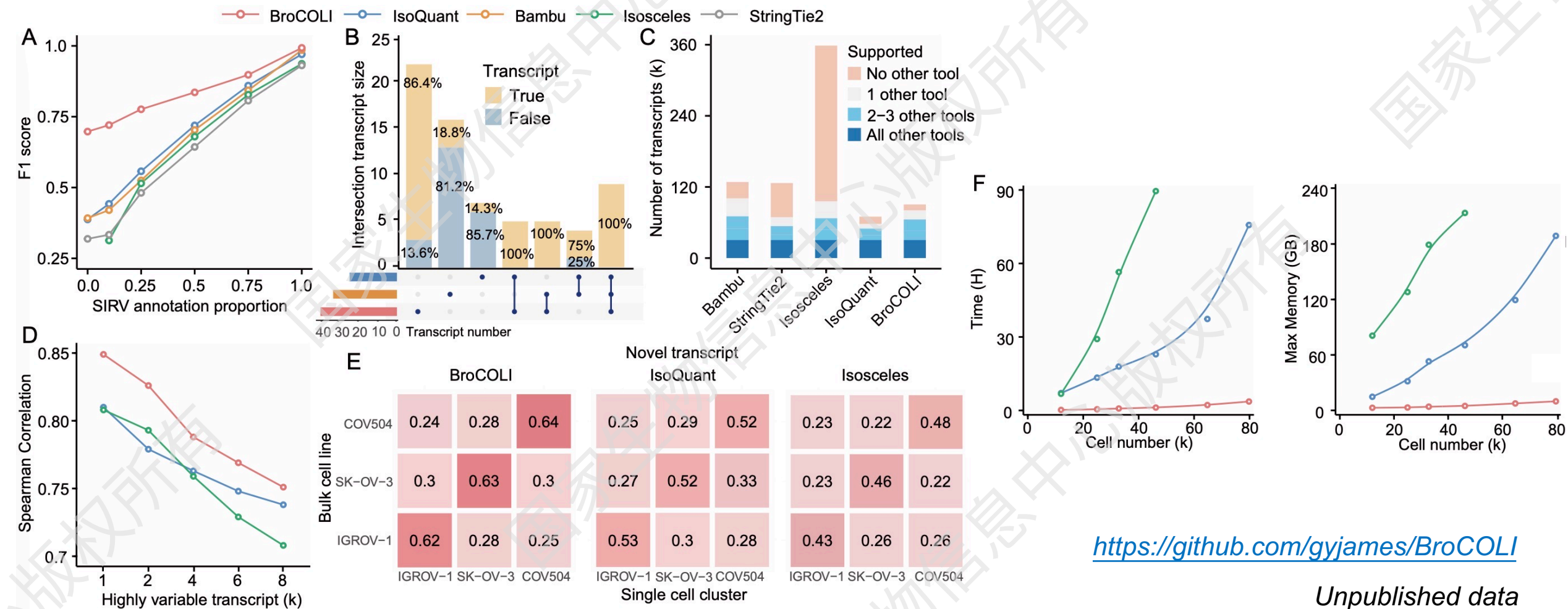
K : No. of the alignment type more than 10 nt away from any of the putative SJ positions

ESPRESSO (Error Statistics Promoted Evaluator of Splice Site Options): full-length transcript isoform analysis of long reads

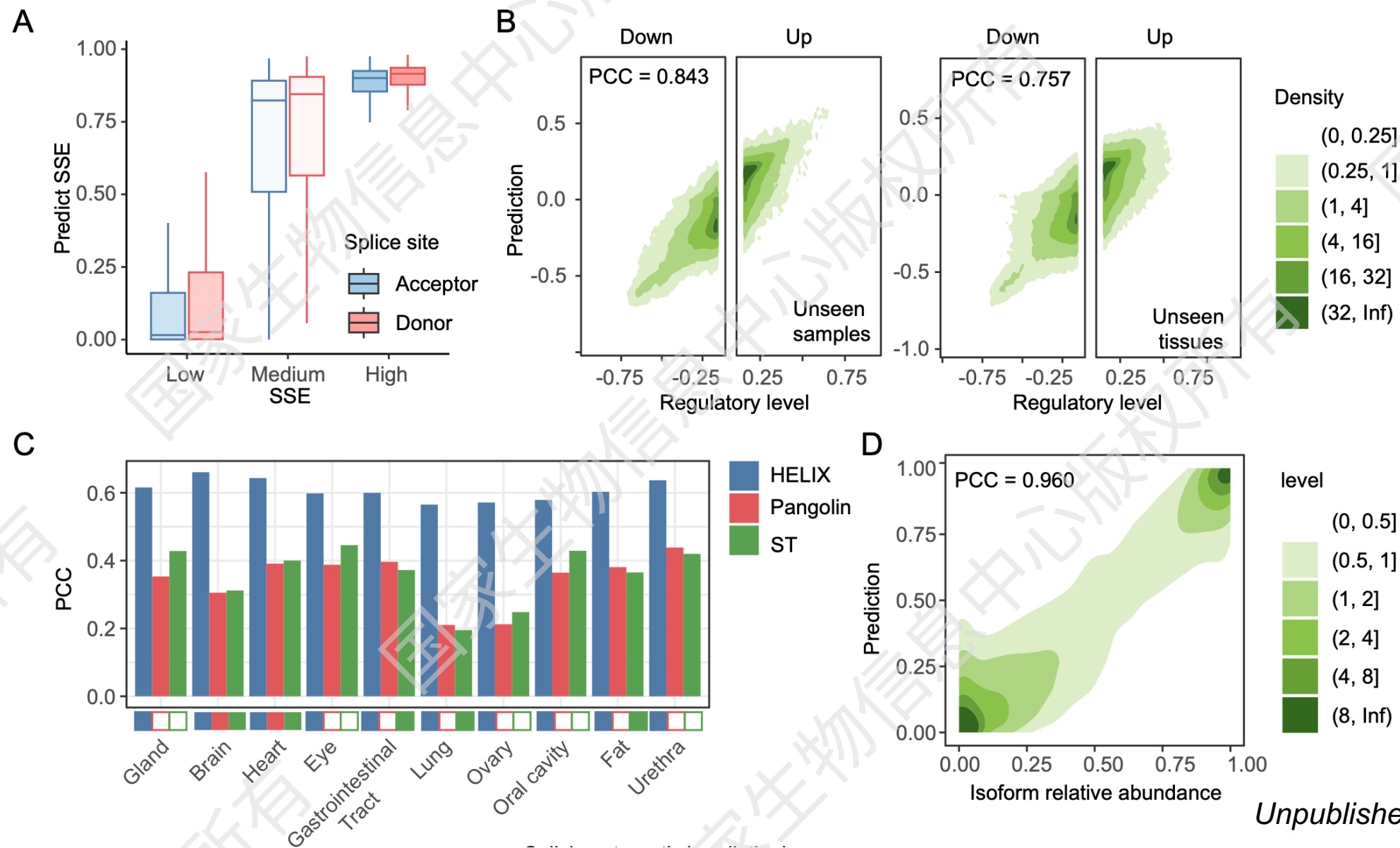


Gao et al. Science Advances, 2023

BroCOLI: a fast and lightweight solution for full-length analysis of bulk, single-cell and spatial transcriptomic data

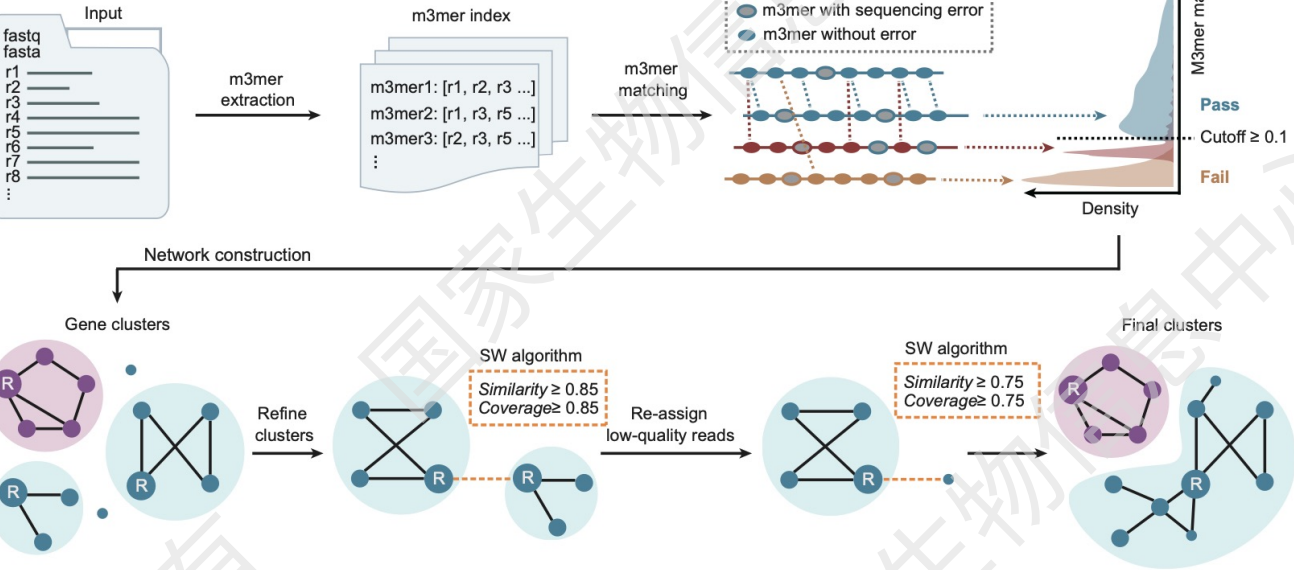


HELIX: AI model for regulation of splicing and full-length isoforms

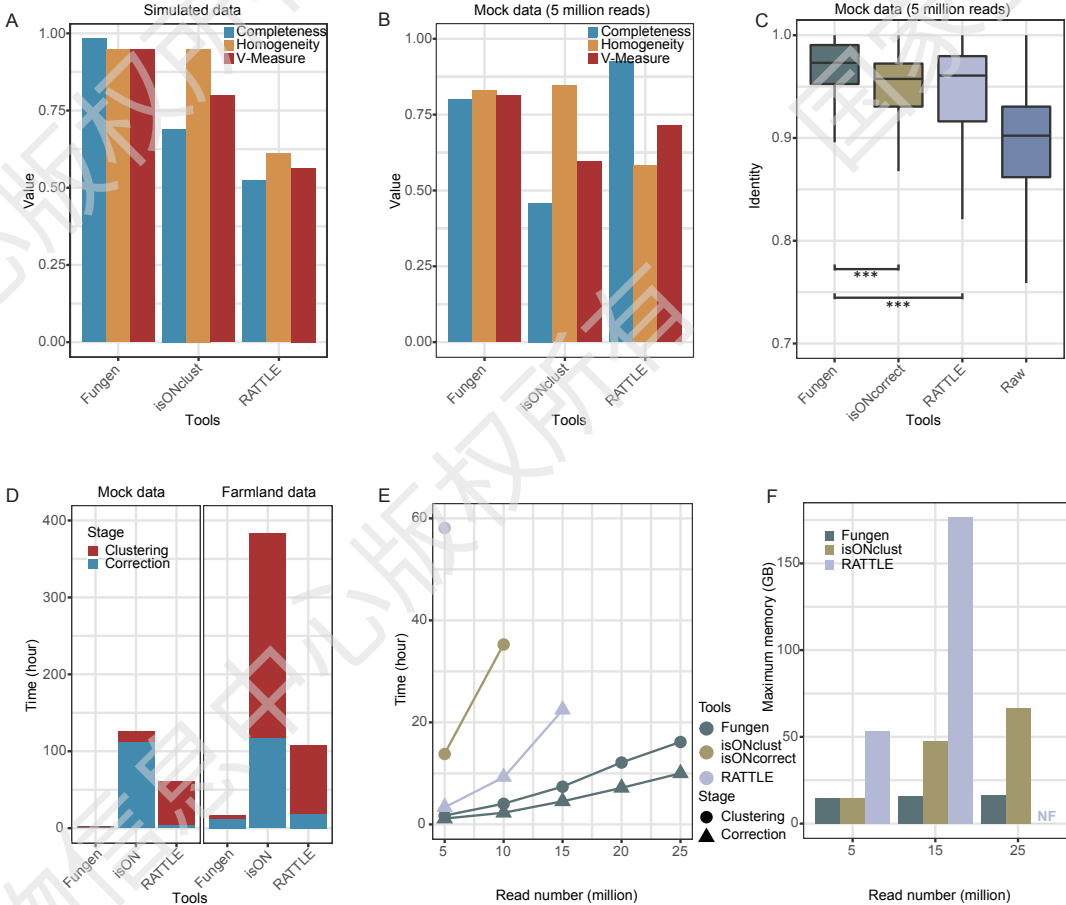
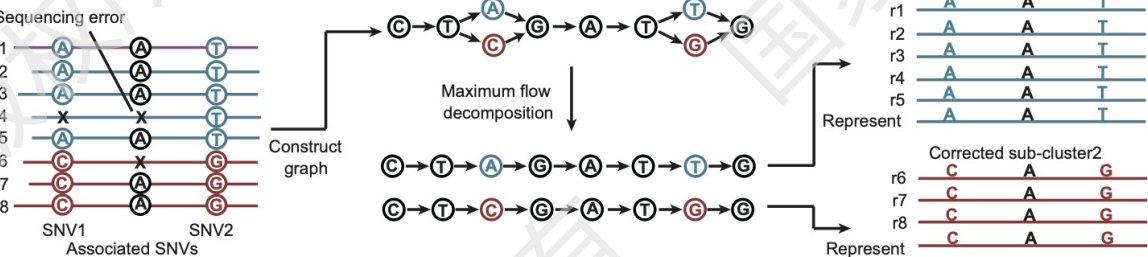


Fungen: clustering and correcting long-read metatranscriptomic data for exploring eukaryotic microorganisms

Stage 1: Clustering

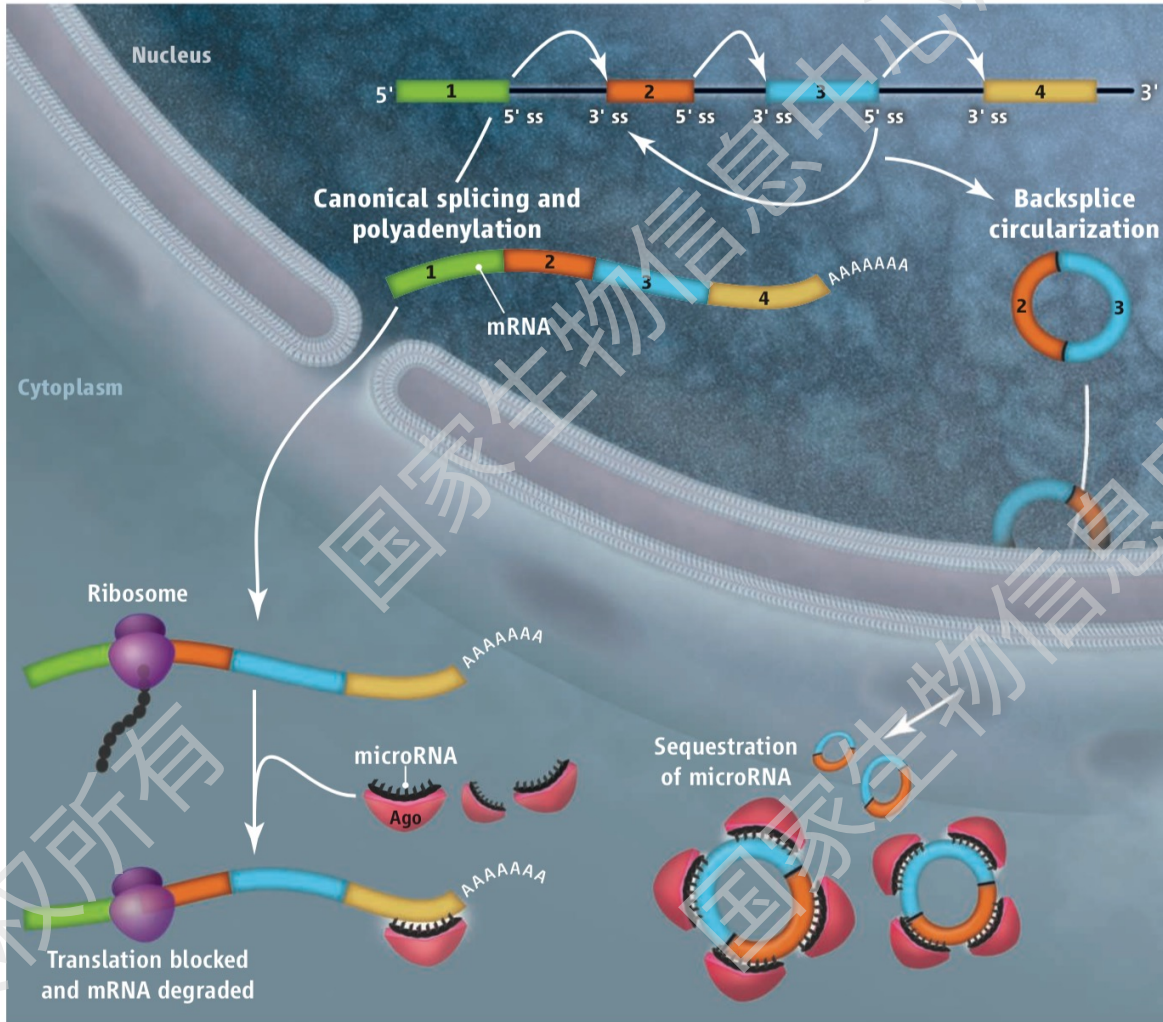


Stage 2: Correction

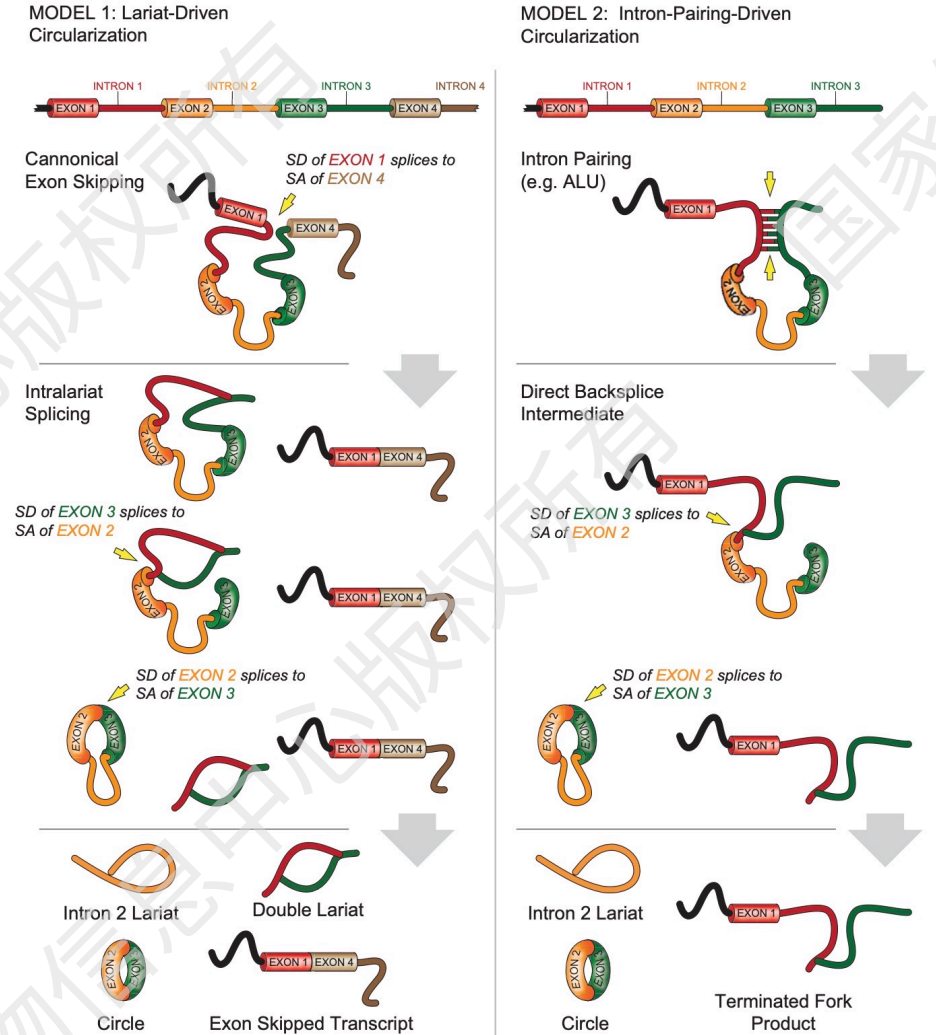


<https://ngdc.cncb.ac.cn/bit/fungen>

A special form of alternative splicing: circRNA



Wilusz et al. Science, 2013



Jeck et al. RNA, 2013

[illegible]

E Pathogenic dsRNAs → RNase L activation → Global circRNA degradation → Active PKR and NF90/NF110

I Low level translation: *circFAM120A* mRNA (m⁷G, (A)_n) → Inactive PKR and NF90/NF110 → *circFAM120A* → IGF2BP2 → High level translation

F mPTP "closed" → SCAR → ATP5B → mROS → CypD

G miR-7 → miR-7-targeted mRNAs → Cdr1as

H *circZNF609* (m⁷G, (A)_n) → HuR → mRNA stability increased → CKAP5 protein → Sustained cell-cycle progression

Nucleus:

- A** *ci-ankrd52* DNA hybrid (R-loop) → RNase H1 → Resolve R-loop for transcription
- B** *circSEP3* DNA hybrid (R-loop) → *circSEP3* → Exon6
- C** *circMBL* → MBL → Feedback-loop regulation
- D** Foreign DNA → cGAS → *cia-cGAS* → Immune responses

Liu et al. Cell, 2022

CIRI: circular RNA identifier

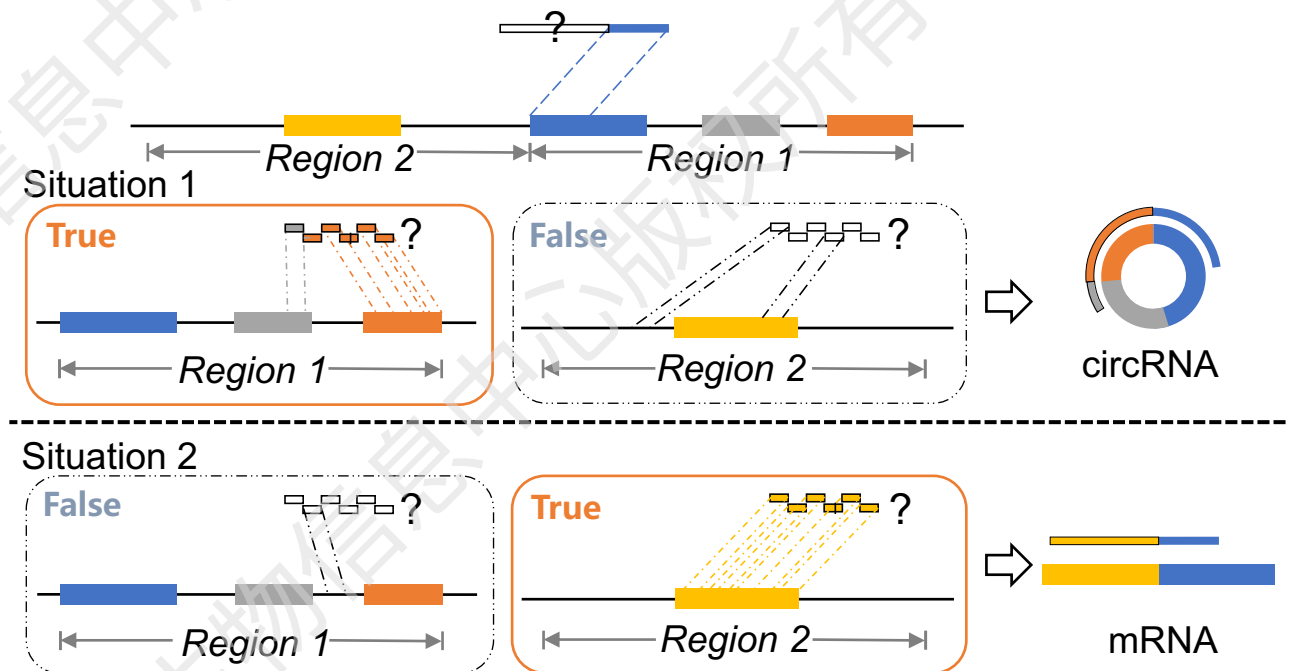
- CIRI

- paired chiasmic clipping signal
- *de novo* detection
- applicable to all 2nd Gen. sequencing read lengths



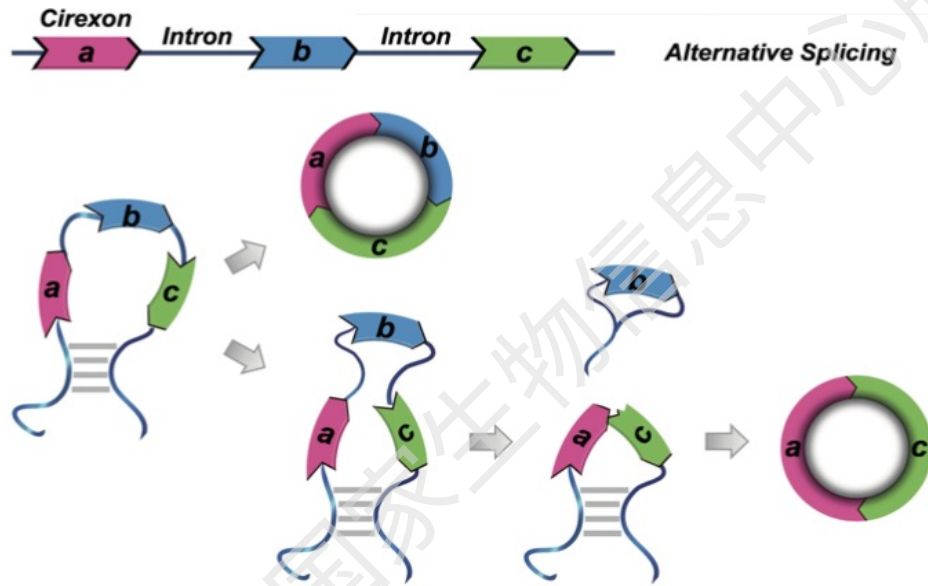
- CIRI2

- maximum likelihood estimation based on multi-seed matching
- higher reliability
- faster

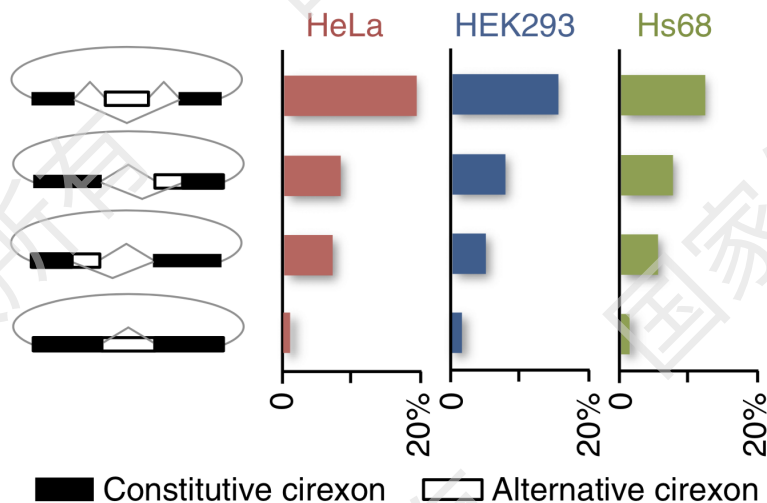


Gao et al. *Brief. Bioinform.*, 2018

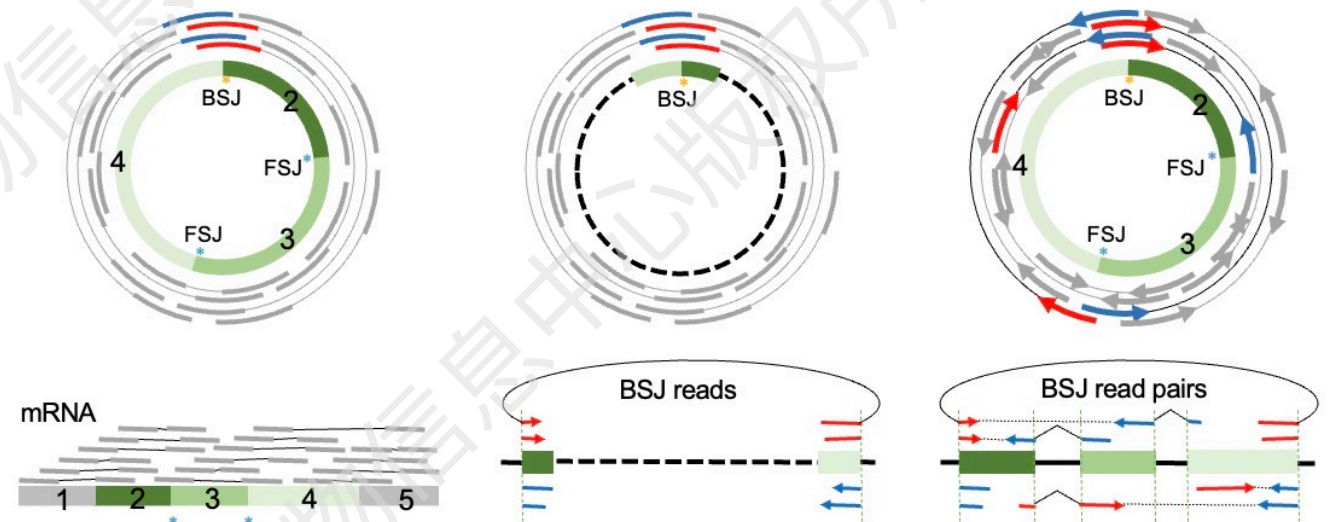
CIRI-AS: alternative splicing within circRNA



- Four types of AS events can be detected within circRNA
- Diversity of circRNAs
 - Alternative back-splicing
 - Alternative splicing

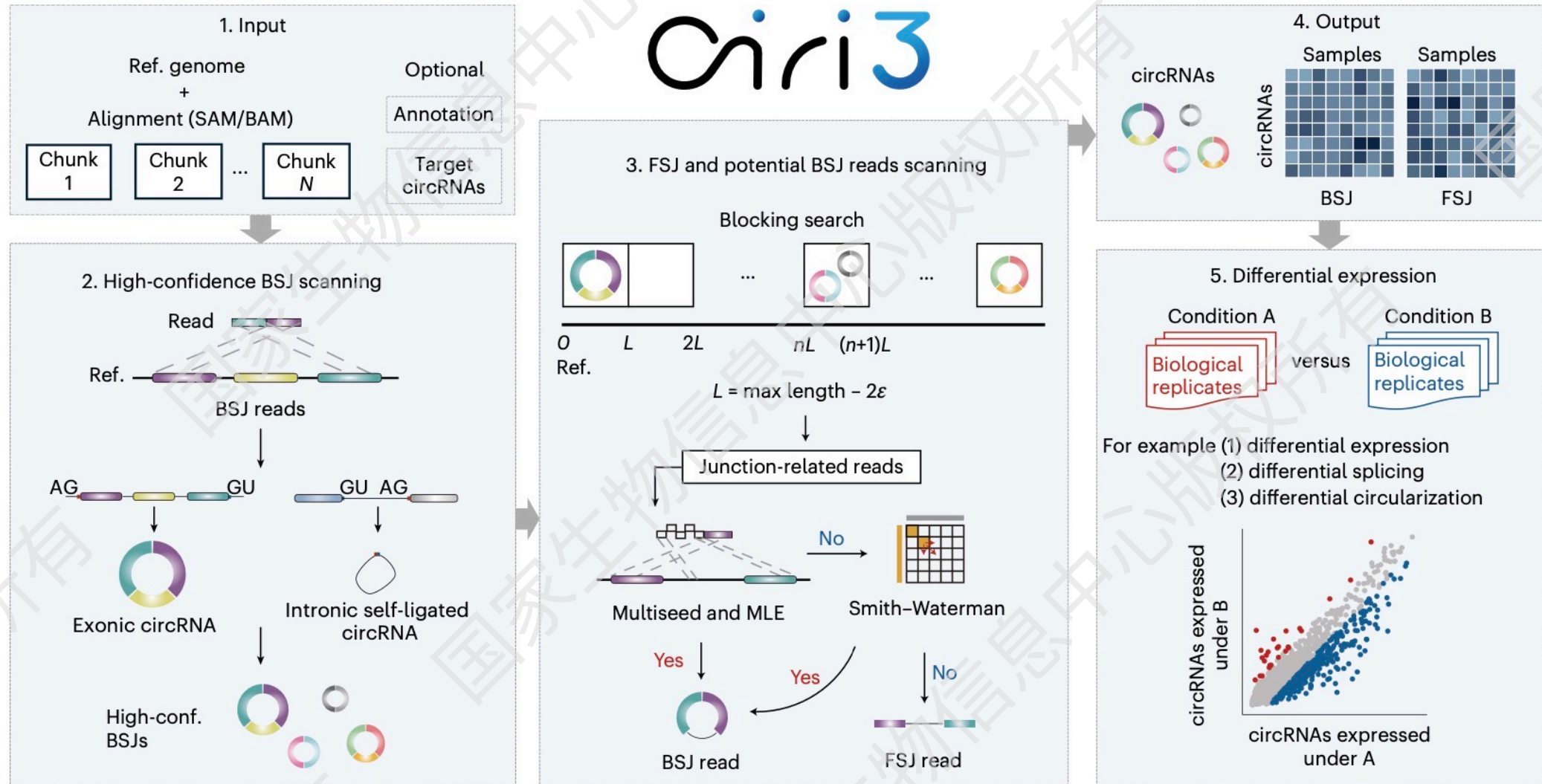


Gao et al. Nat. Commun., 2016

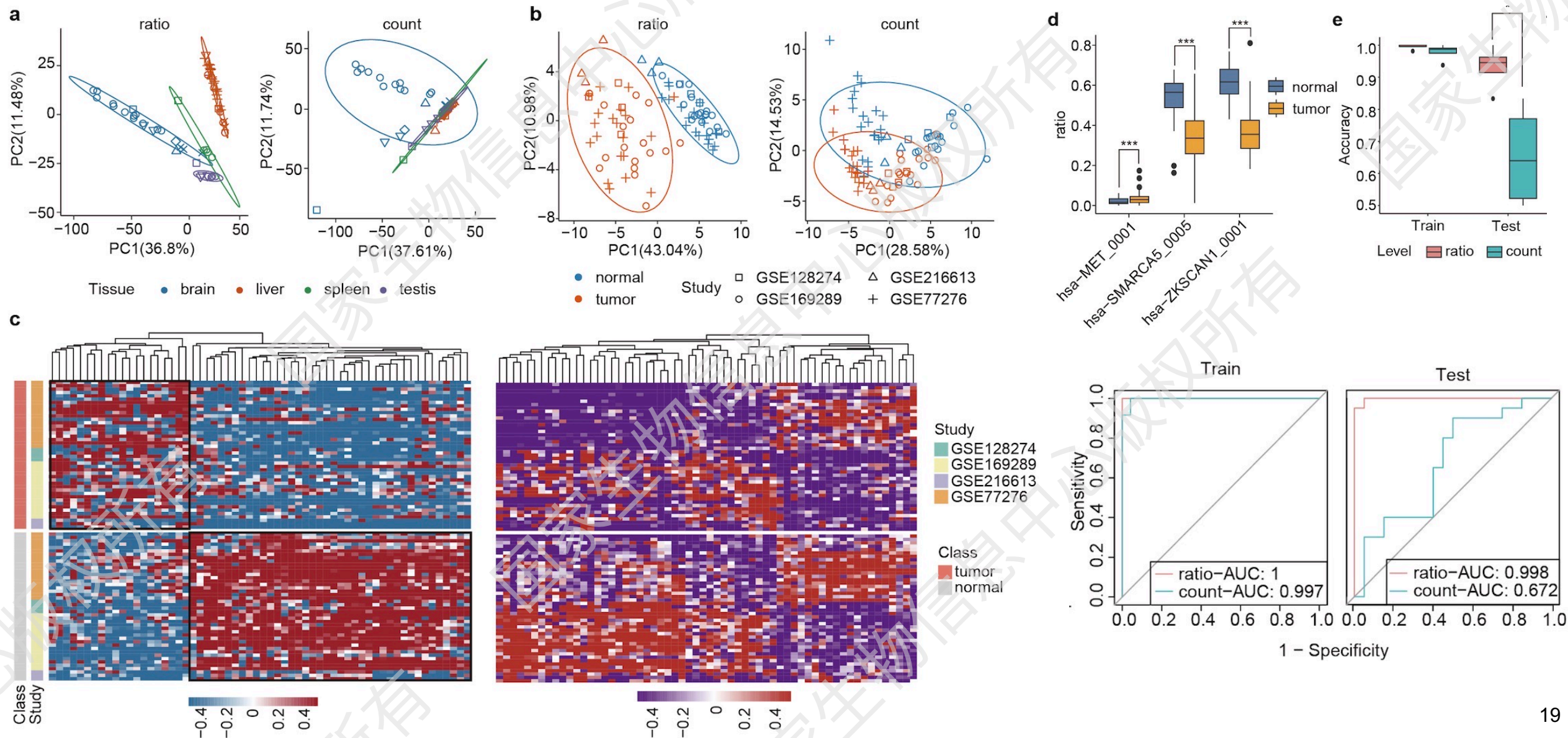


Gao et al. Trends Genet., 2018

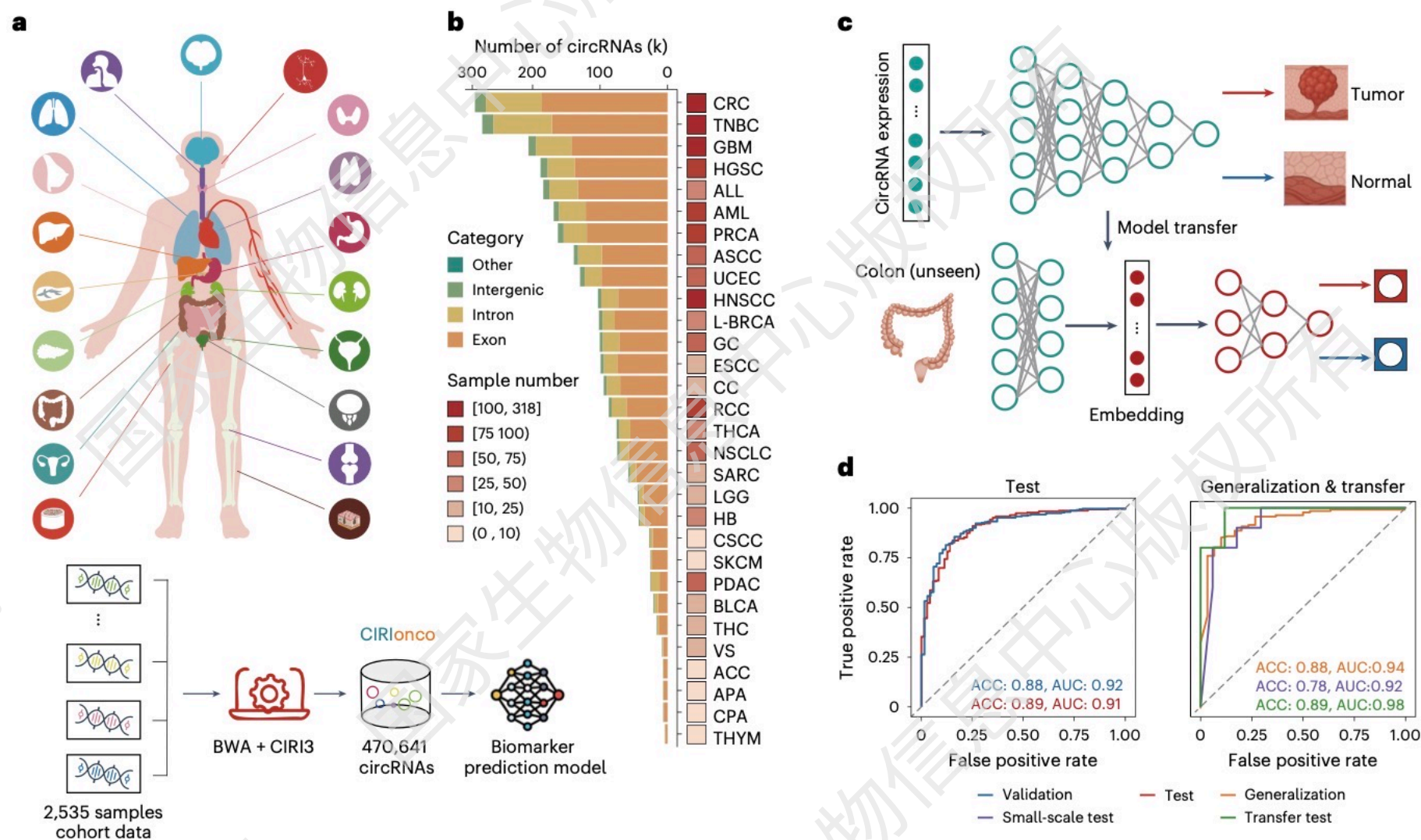
CIRI3: scalable and accurate detection and quantification of circular RNAs in extremely large RNA-seq datasets



BSJ ratio is a useful metric for circRNA regulation



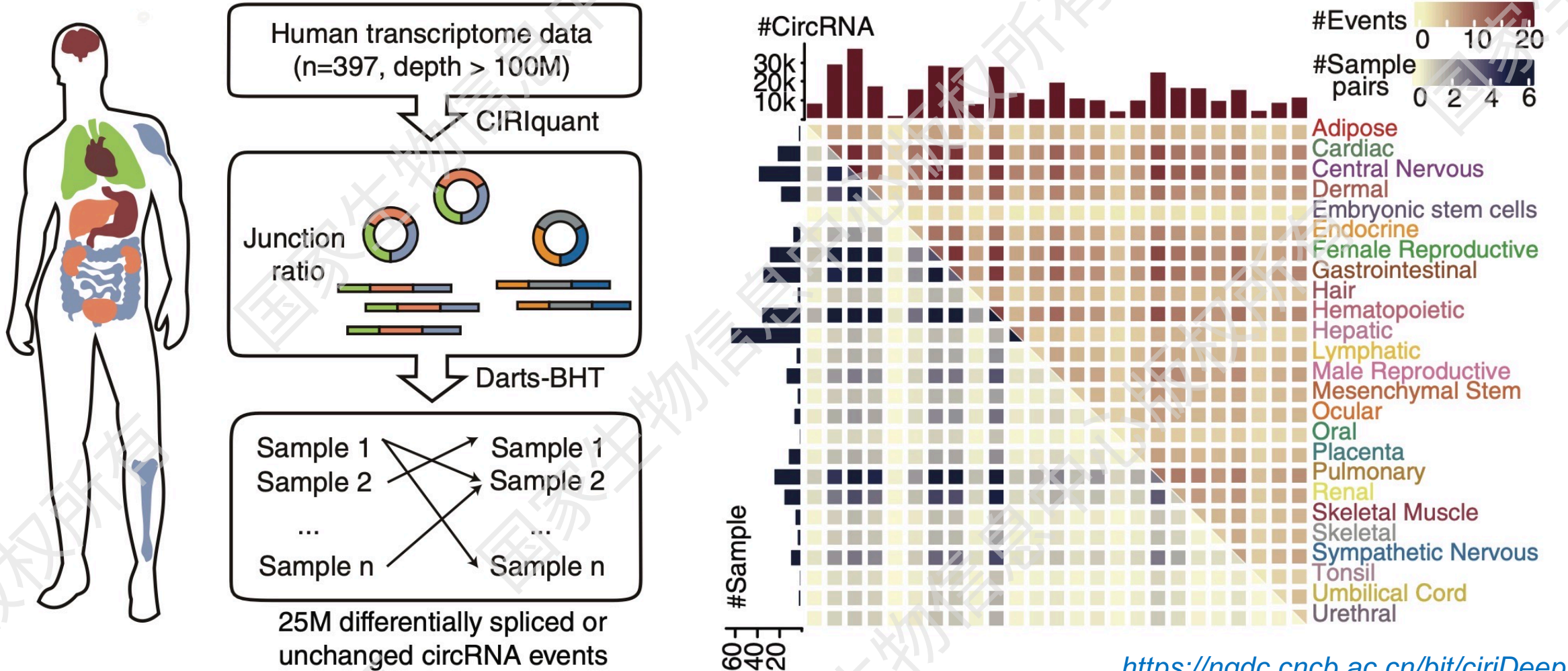
CIRIonco: diagnostic potential of circRNAs in cancer



<https://ngdc.cncb.ac.cn/cirionco/browse/>

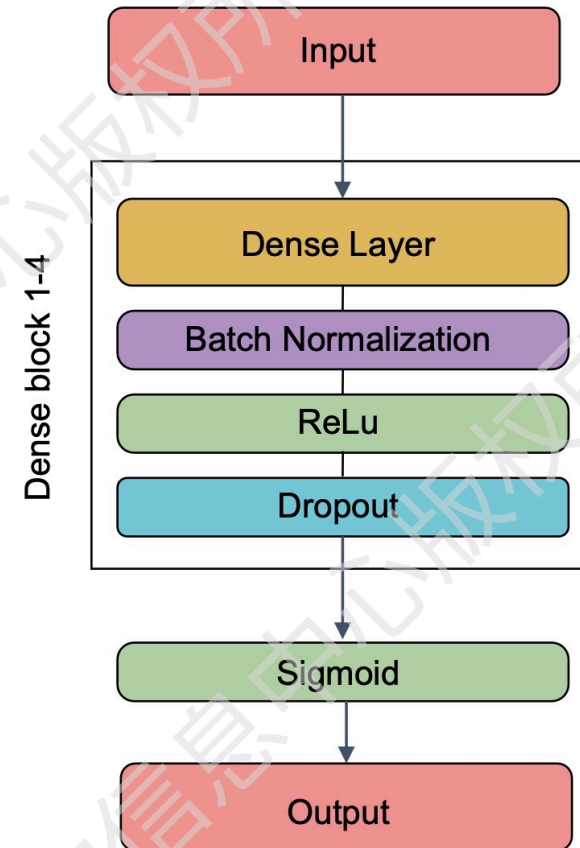
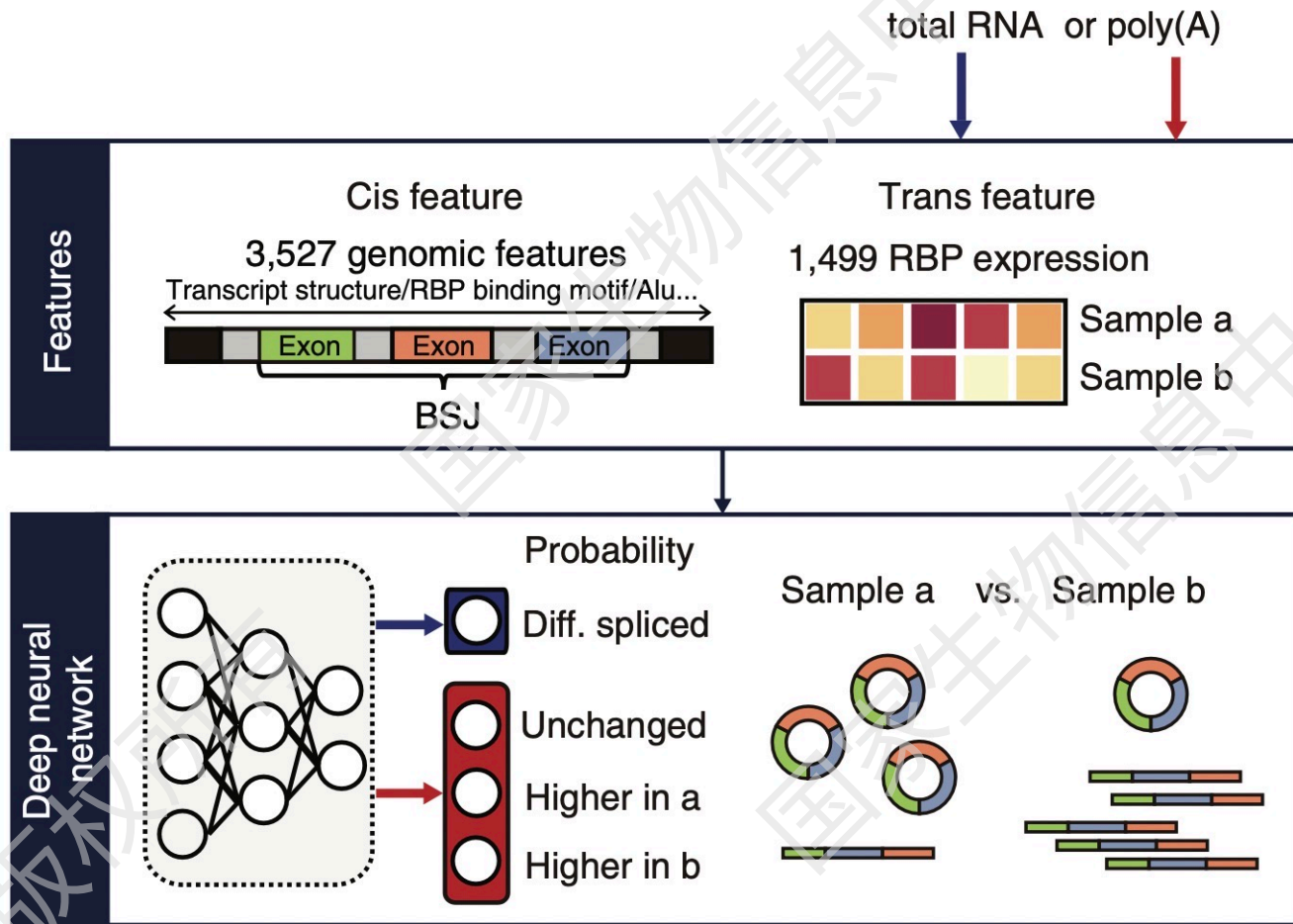
Zheng et al. Nature Biotechnology, 2025

CIRI-Deep: a deep learning model for analysis of circRNAs



<https://ngdc.cncb.ac.cn/bit/ciriDeep>

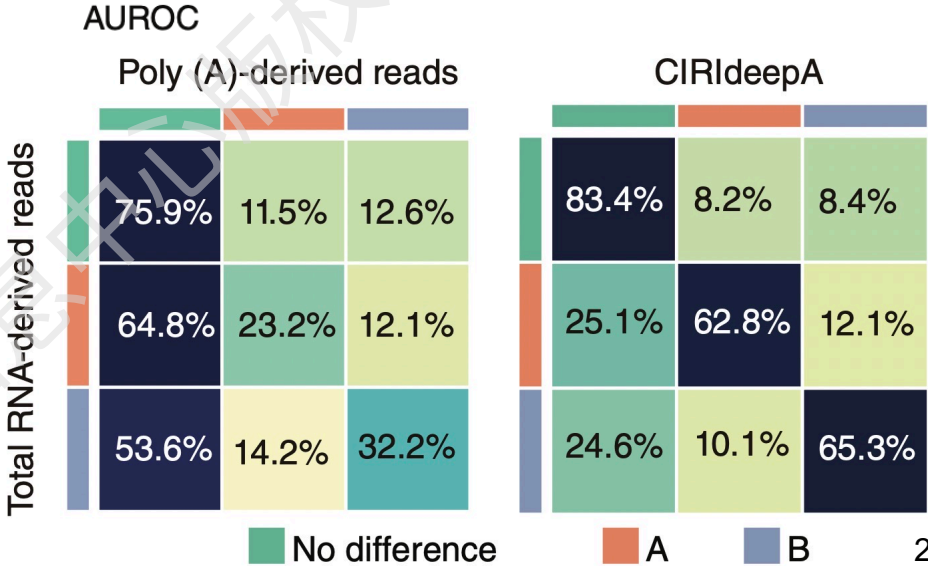
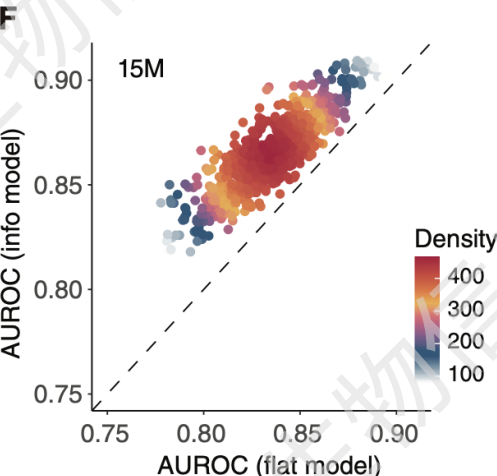
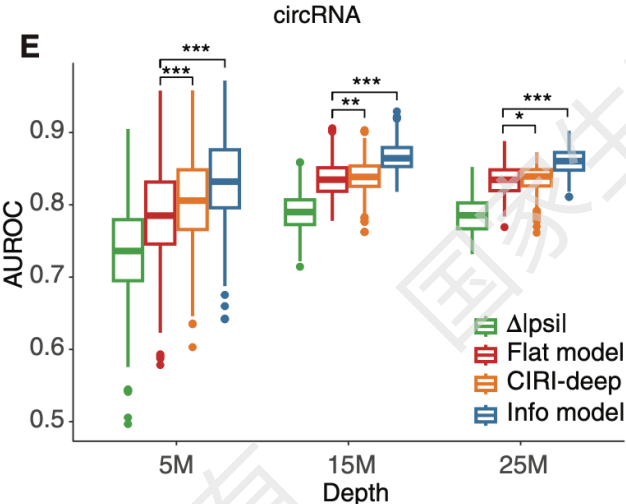
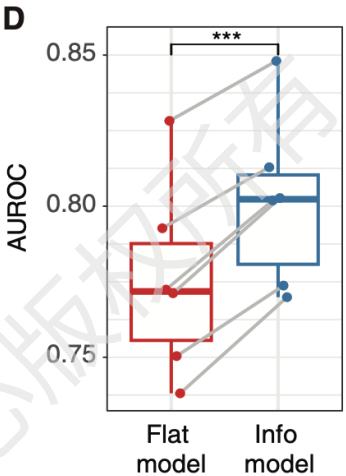
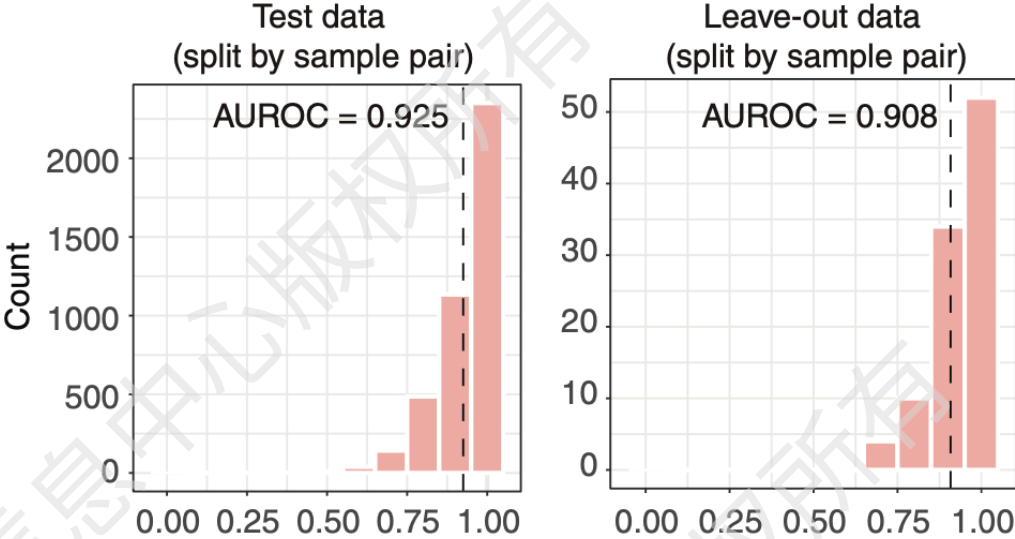
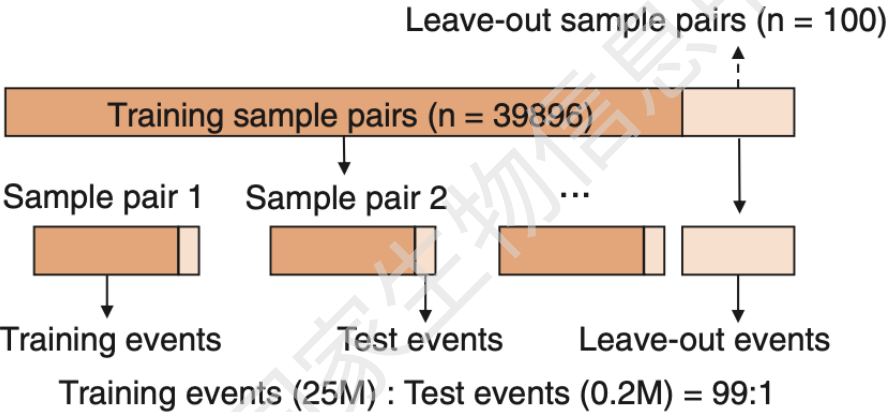
Schematic framework of CIRI-deep



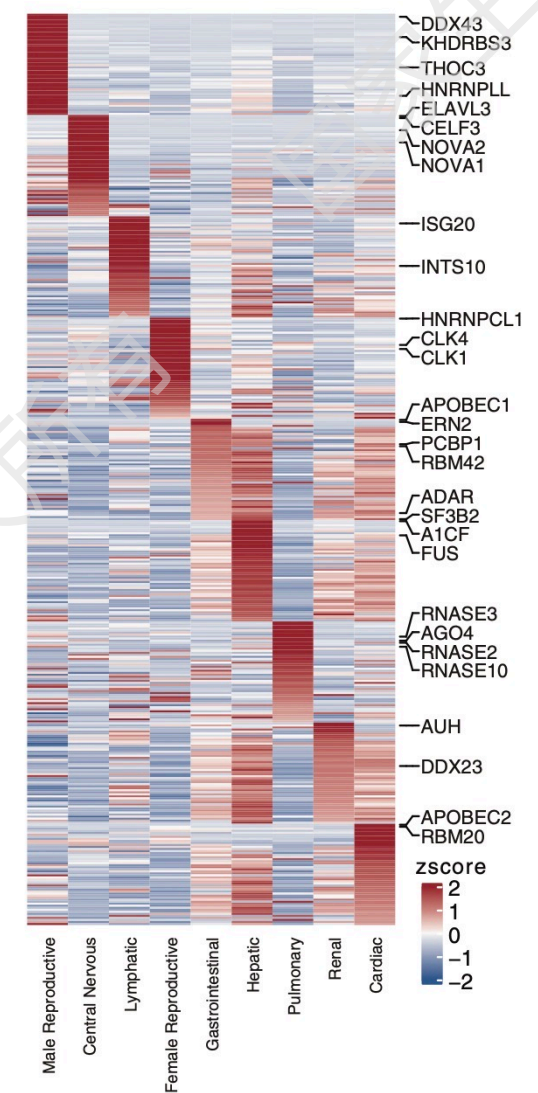
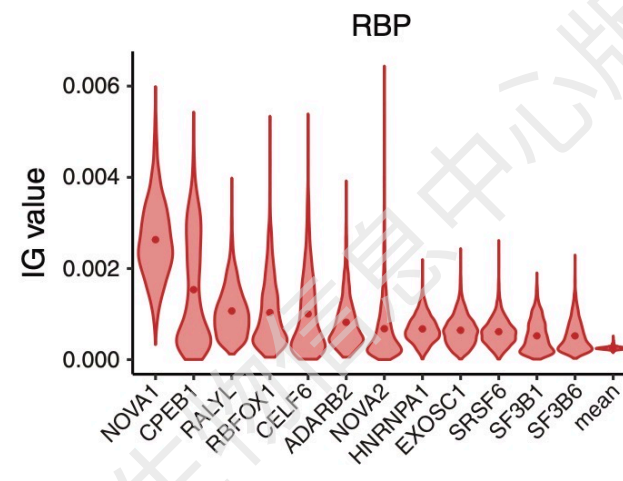
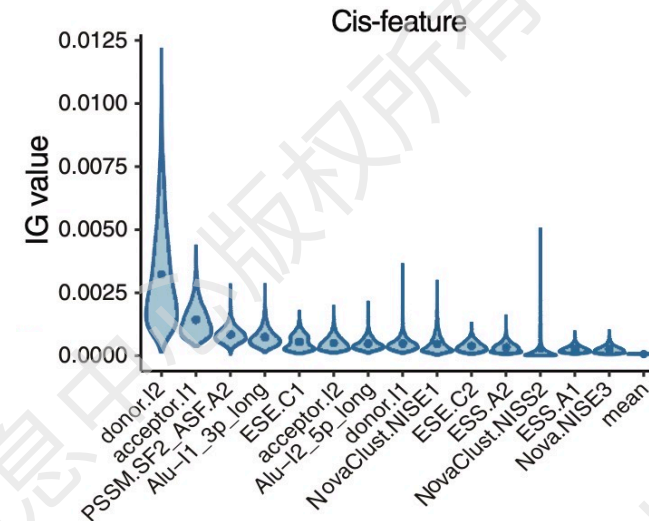
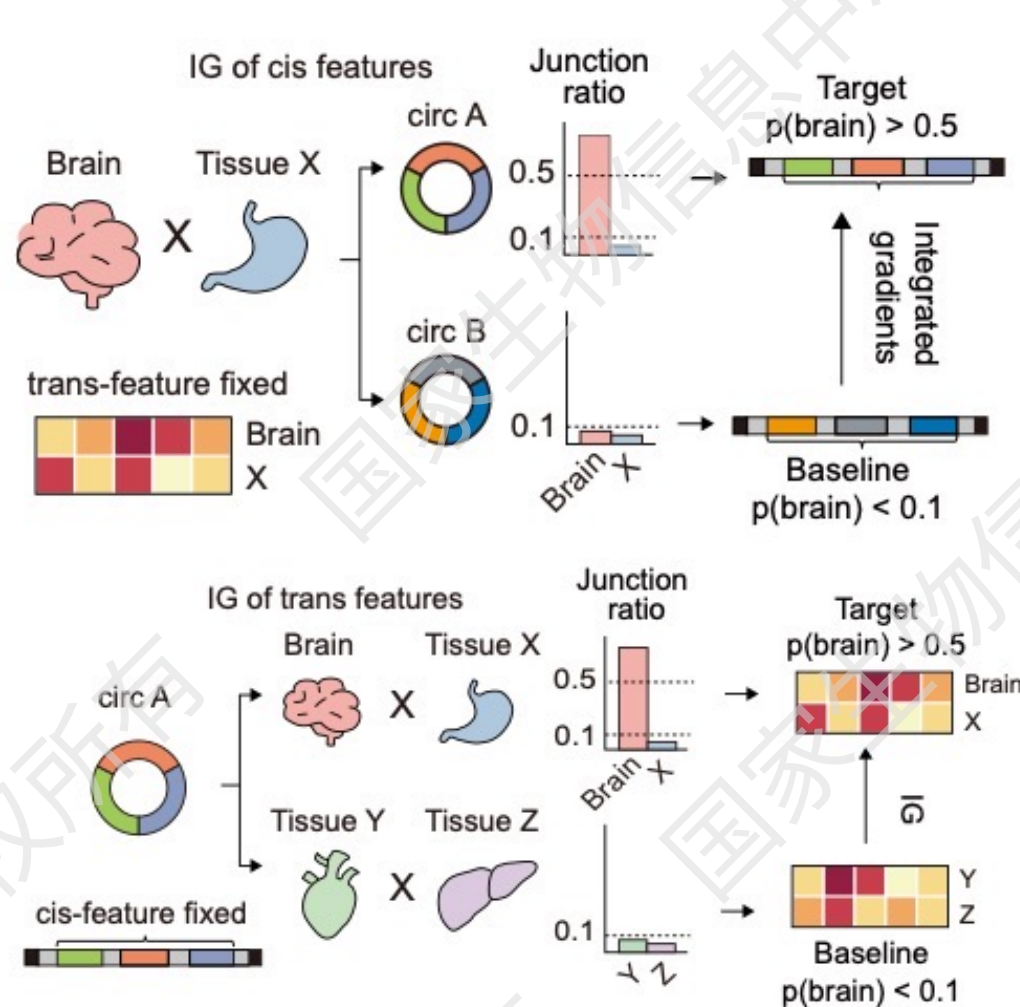
<https://ngdc.cncb.ac.cn/bit/ciriDeep>

Zhou et al. Advanced Science, 2024

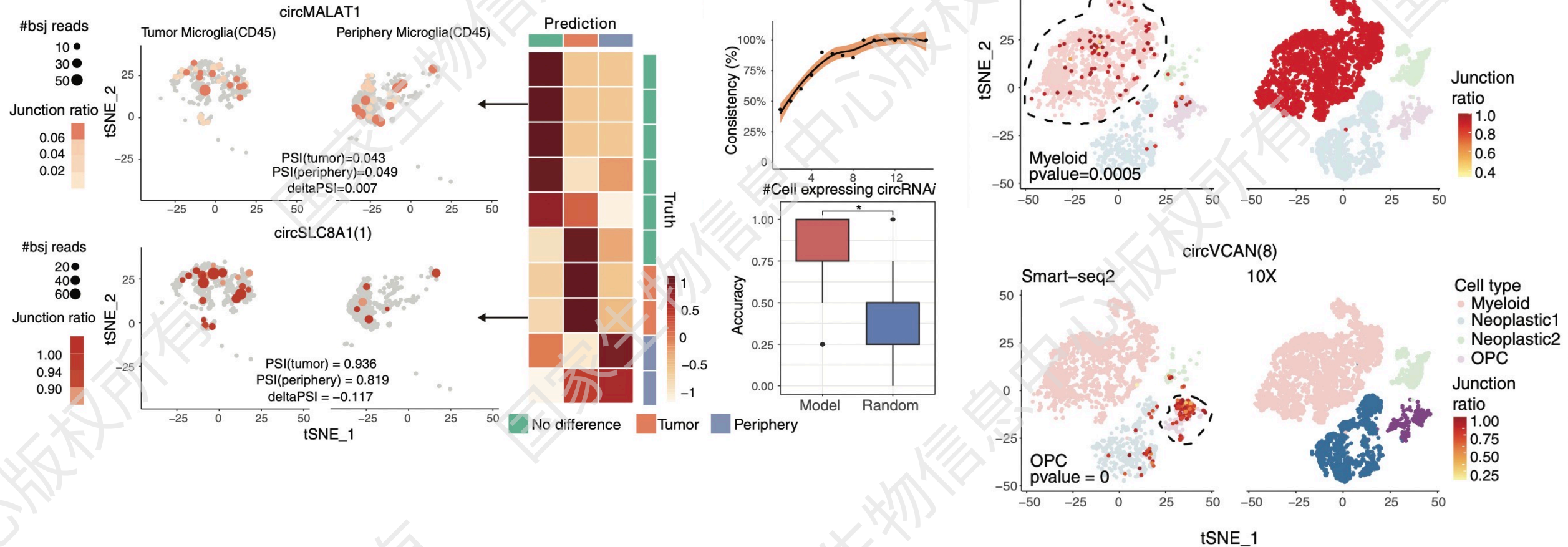
CIRI-deep accurately predicts differentially-spliced circRNA



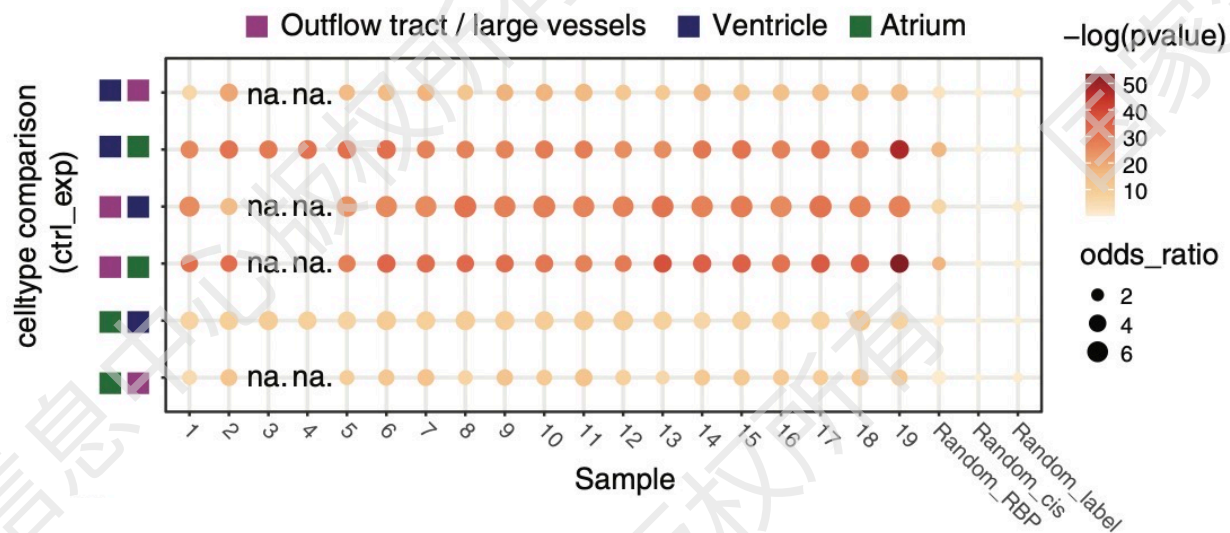
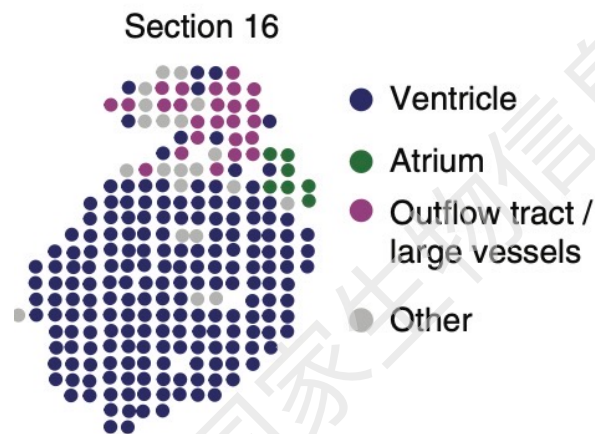
Identify significant tissue-specific cis and trans features with integrated gradients



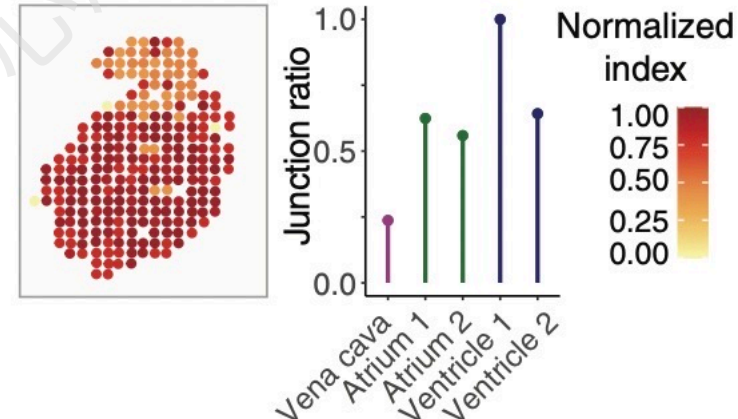
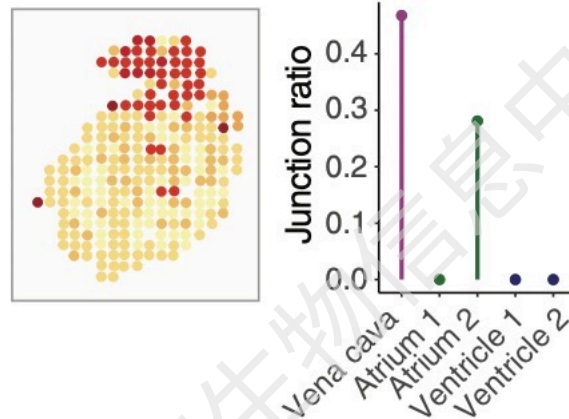
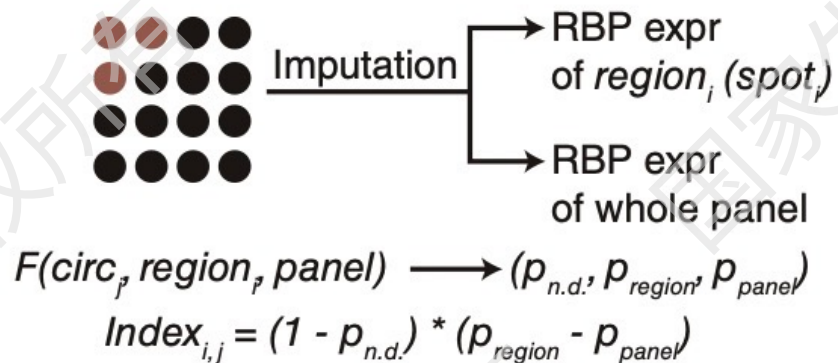
Prediction of differentially spliced circRNA between scRNA-seq clusters



Calculating circRNA index for junction ratio map



For each $region_i$ ($spot_i$), $circRNA_i$:



Acknowledgement

Xing Lab

Prof. Yi Xing
Zhicheng Pan
Feng Wang
Robert Wang
Prof. Lan Lin
Eric Kutschera
Yang Xu
Stephan Xie
Yuanyuan Wang
Kathryn E. Kadash-Edmondson

Zhao Lab

Prof. Fangqing Zhao
Jinyang Zhang
Yi Zheng
Peifeng Ji
Wanying Wu
Lingling Hou
Wenyu Shi

Zihan Zhou
Xin Zheng
Wei Pan
Weizhi Yuan
Bingqi Wu
Weiwei Zhang
Mengxue Peng
Shixin Chen
Wenao Hu
Yukun Zhang
Jie Zhang



北京市科学技术委员会



中华人民共和国科学技术部
Ministry of Science and Technology of the People's Republic of China

