Short human eccDNAs are predictable from sequences

中央研究院
ACADEMIA SINICA

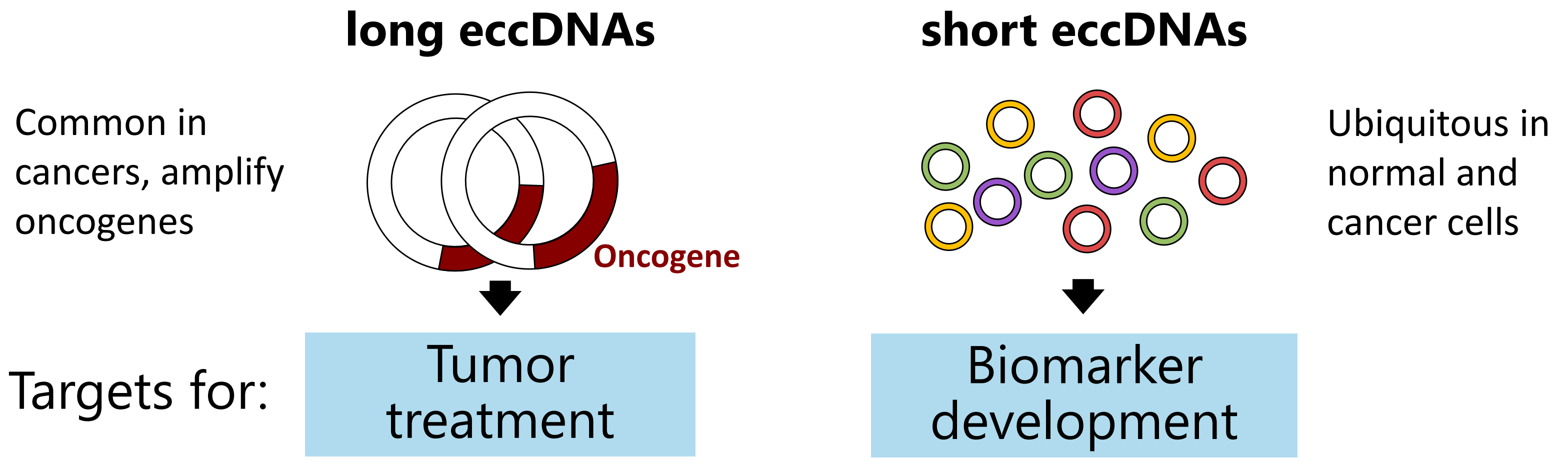
K.-L. Chang[1], J.-H. Chen[1], T.-C. Lin[1], J.-Y. Leu[2], C.-F. Kao[3], J. Y. Wong[1,✉] & H.-K. Tsai[1,✉]

[1] Institute of Information Science, [2] Institute of Molecular Biology, [3] Institute of Cellular and Organismic Biology, Academia Sinica, Taiwan.
✉ wongjinyung@gmail.com; hktsai@iis.sinica.edu.tw

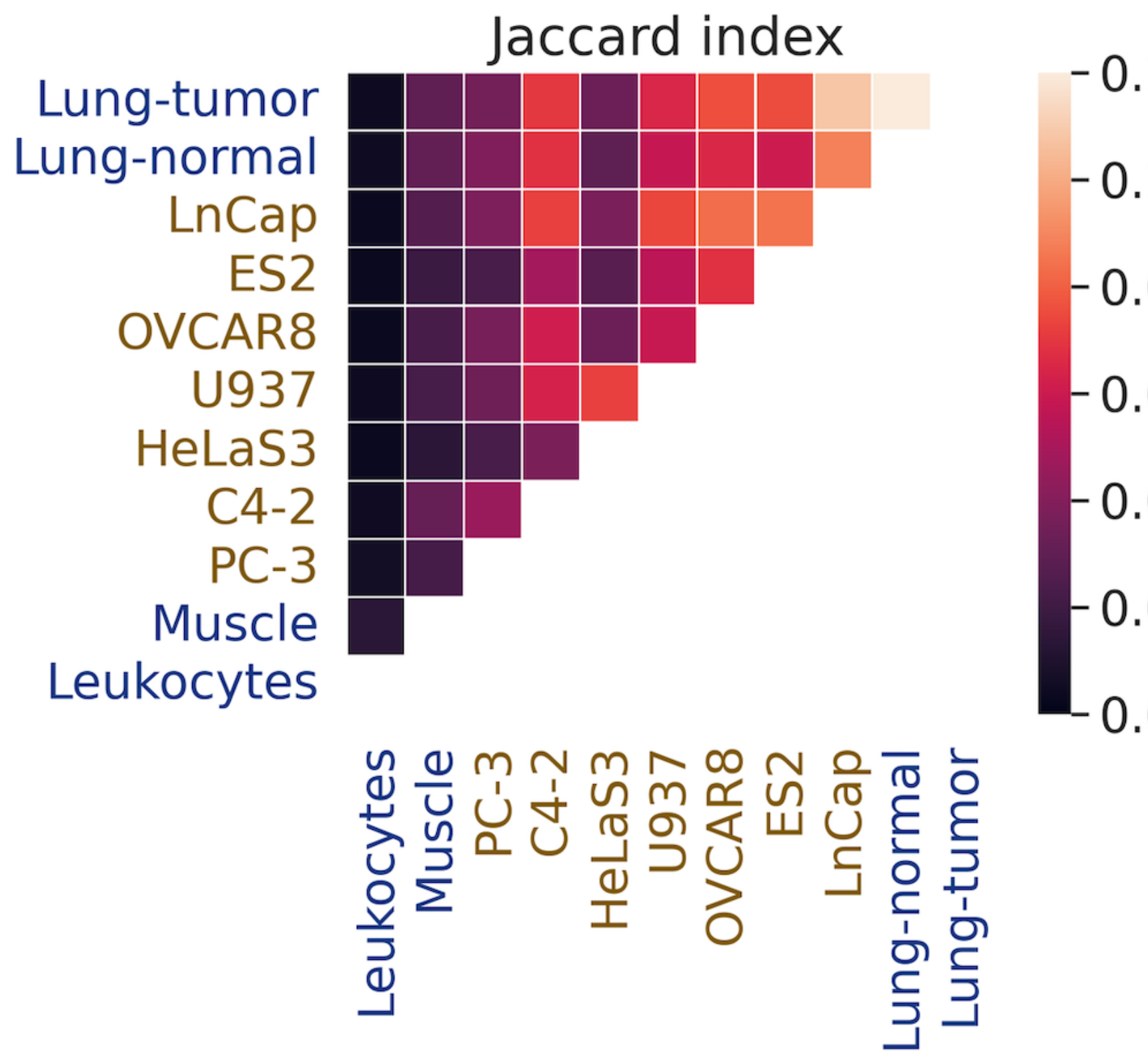
Introduction:
eccDNAs are emerging clinical targets

What are eccDNAs? Why they are important?

- **e**xtra**c**hromosomal **c**ircular **DNA**s
- Chromosomal origins
- Two research focuses:



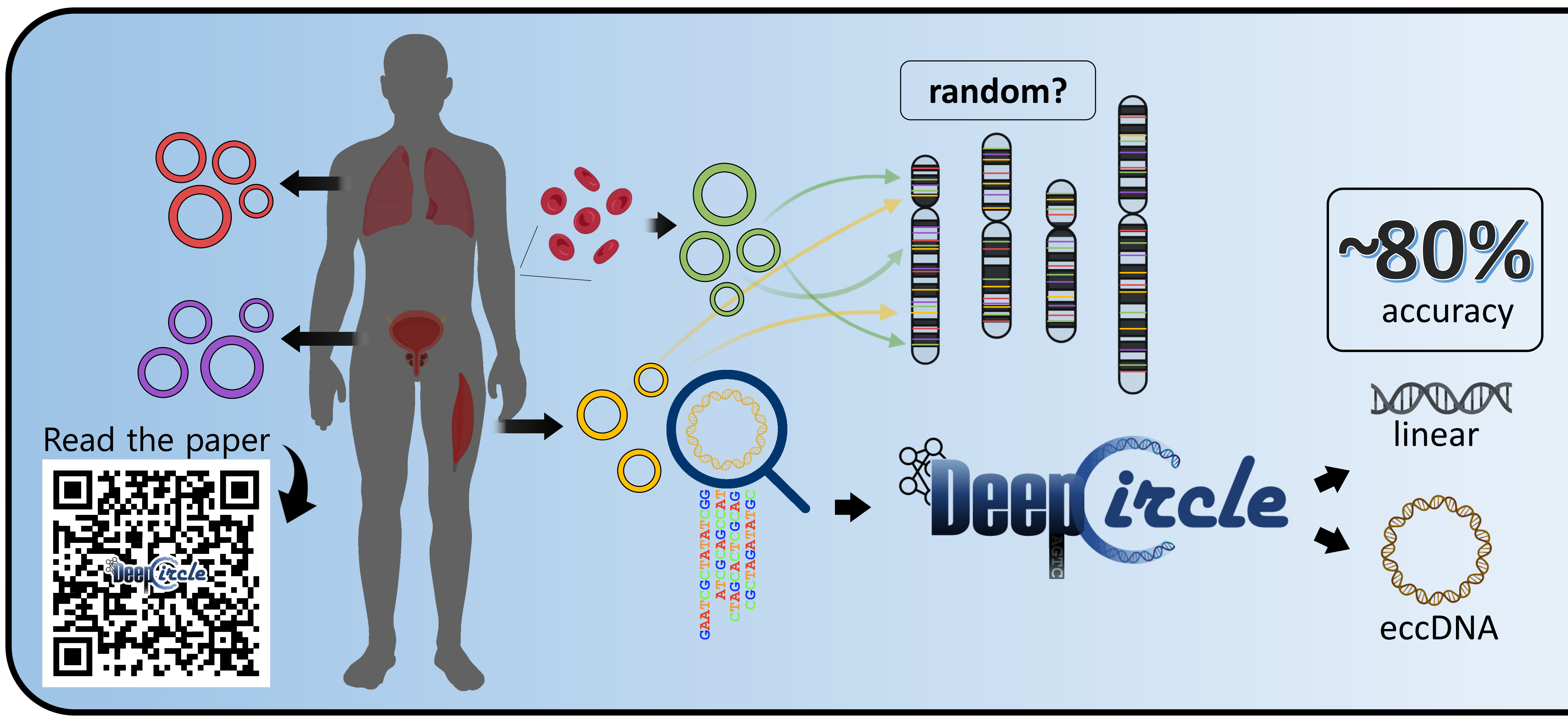
Motivation:
conflicting views on eccDNAs formation



① **Randomly produced:** apoptotic products, no specificity in genomic origins

② **Inducible production:** by physiological stress etc., hence useful as biomarkers

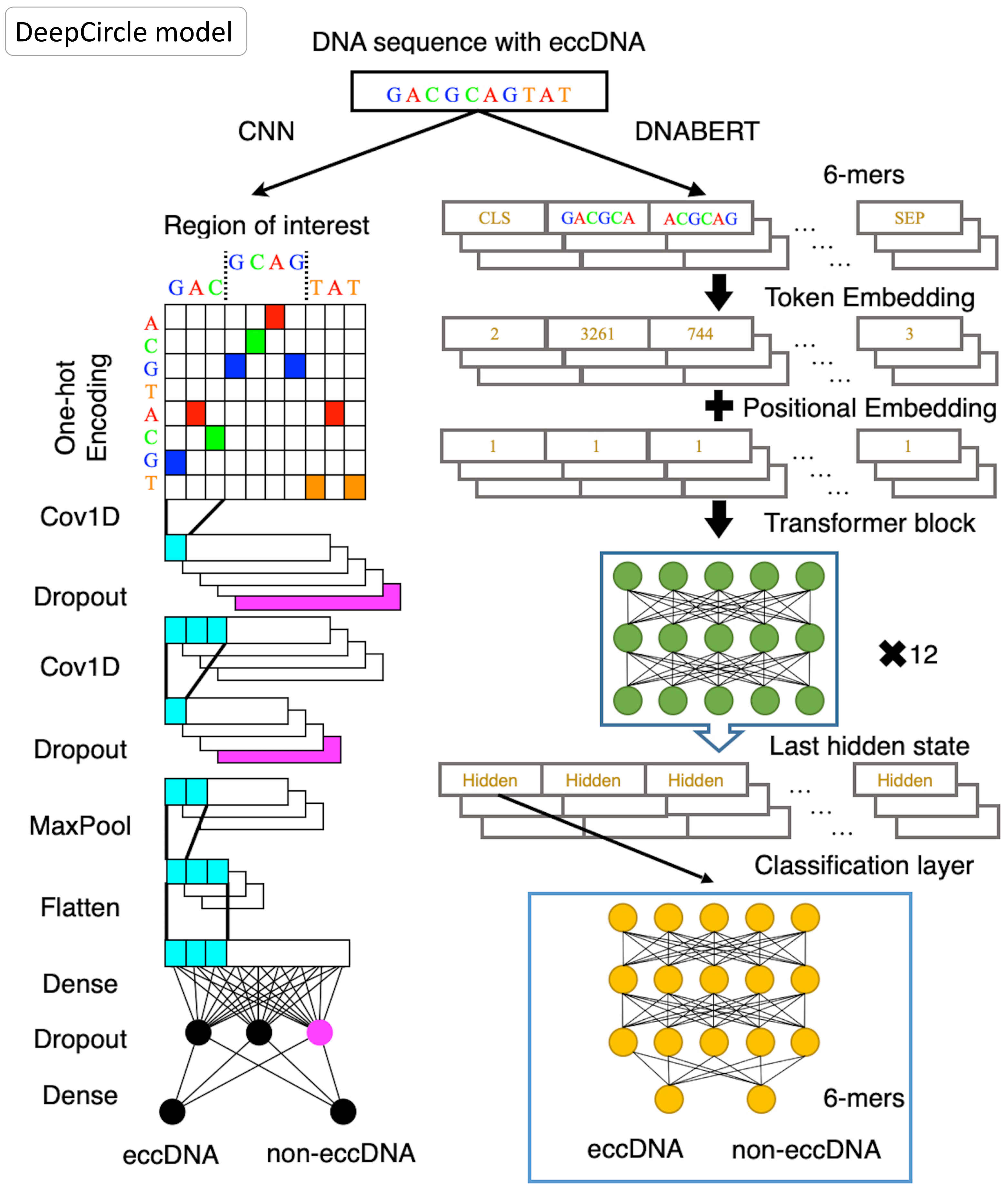
Low similarity of eccDNAs' genomic origins between tissues



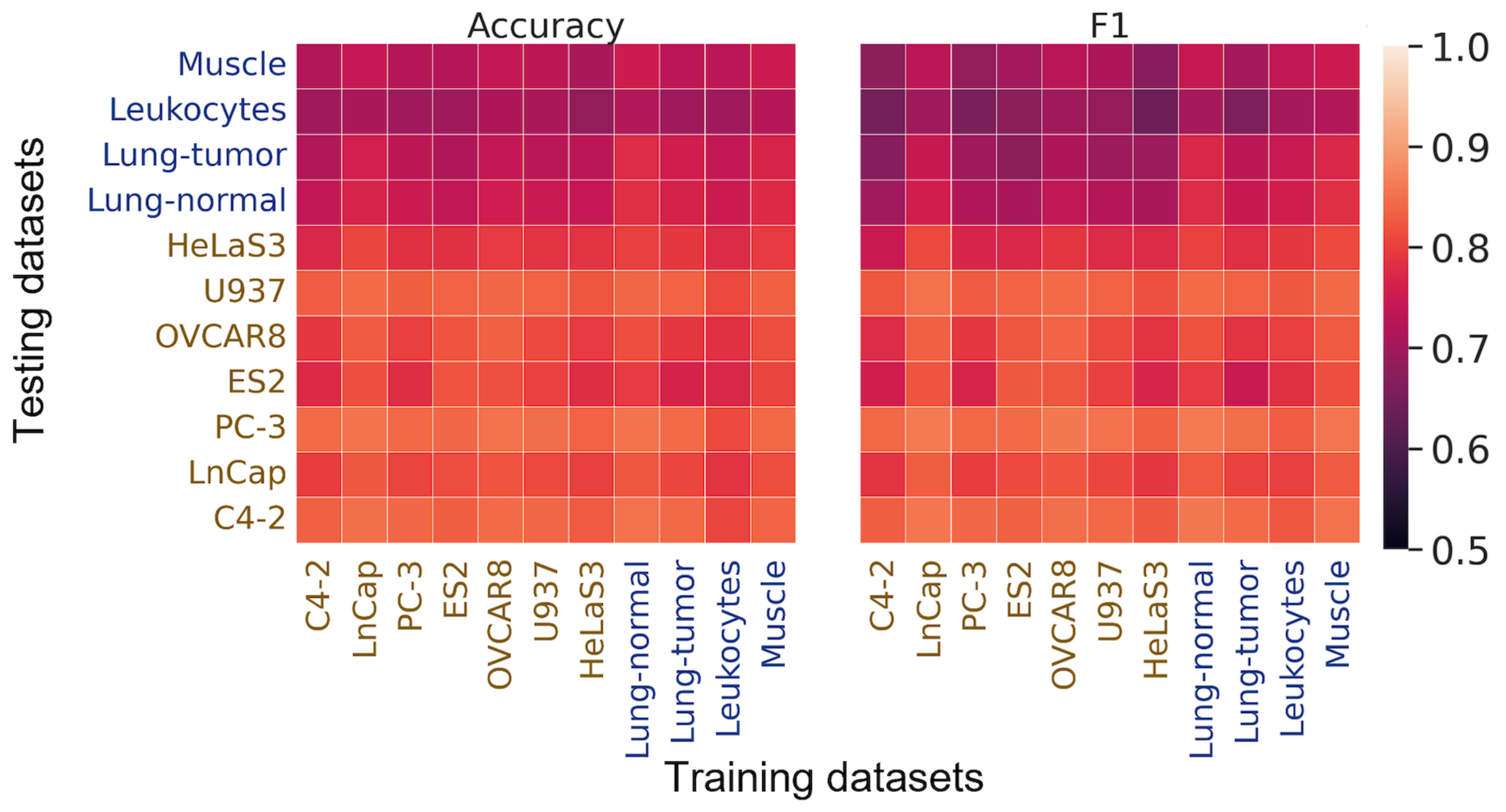
Conclusion
The ubiquitous eccDNAs lack specificity in genomic origins but are predictable from sequences

Significance
Deep learning helps re-assessing the perceived randomness of genomics elements

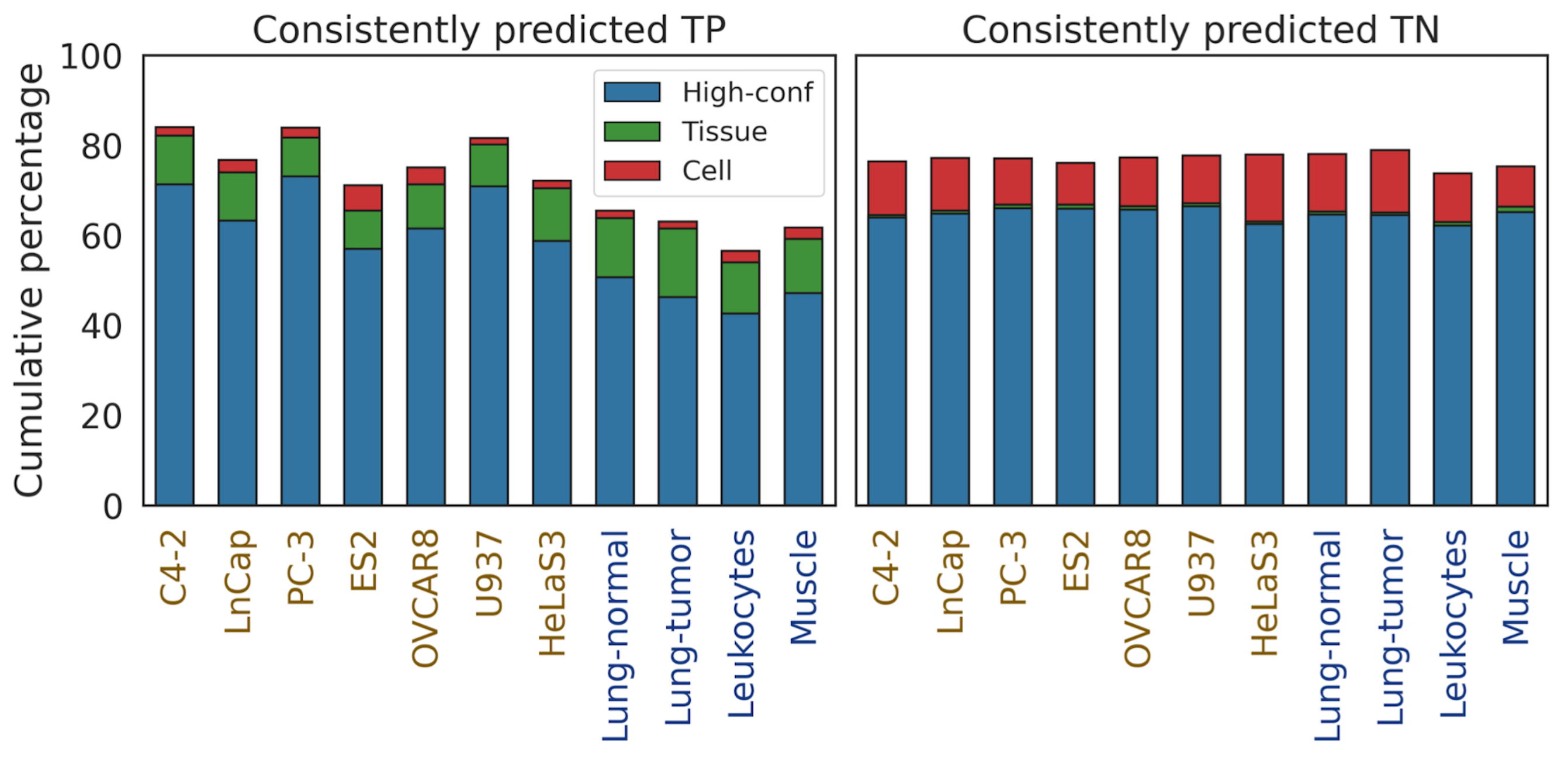
Method:
predict eccDNAs with CNN and DNABERT



Results:
robust prediction of short eccDNAs across diverse tissues suggesting shared features



Accurate predictions from models trained on one dataset and tested on others



High percentage of eccDNAs consistently predicted as true positives/ negatives by different models

Acknowledgements

National Science and Technology Council
(110-2221-E-001-013-MY3; 110-2811-E-001-502)
Academia Sinica Grand Challenge Program
(AS-GC-110-L15)

Visit us:
BITS
BIO-IT Station