

Xi Fu, PhD

Associate Research Scientist and Irving Cancer Early Scholar
Program of Mathematical Genomics, Columbia University Irving Medical Center
New York, NY 10032

xf2217@cumc.columbia.edu | github.com/fuxialexander | ORCID: 0000-0001-9835-5018

Research Profile

I develop biologically informed AI models, AI agents, and GPU-accelerated algorithms to study how gene regulation shapes cellular states in cancer. My work integrates regulatory genomics and structure-based analysis to investigate transcription-factor interactions, interpret noncoding variants, and support biological discovery.

Appointments

2025–present	Associate Research Scientist Program of Mathematical Genomics, Department of Systems Biology.	Columbia University
2025–present	Irving Cancer Early Scholar , Herbert Irving Comprehensive Cancer Center, Columbia University.	

Research Funding

2025–2027	Irving Cancer Early Scholar Award \$100,000 annually for two years. Development of cell-type-specific foundation models for transcription, agentic AI for automated biological discovery, and interpretation of noncoding variants in cancer.	Columbia University
-----------	---	---------------------

Education

2020–2026	PhD, Biomedical Informatics Advisor: Raul Rabadan. Thesis: <i>A foundation model of transcription regulation and application to cancer</i> . Defense: November 12, 2025; diploma: February 11, 2026. MPhil (2023) and MA (2022), Biomedical Informatics.	Columbia University
2016–2019	MPhil, Computer Science and Engineering Thesis: <i>Systematic Identification and Prioritization of Noncoding Variants in Hirschsprung’s Disease</i> .	The Chinese University of Hong Kong
2012–2016	BSc, Cell and Molecular Biology , The Chinese University of Hong Kong.	

Selected Research Contributions

2022–2024	GET: General Expression Transformer Initiated the study and brought together collaborators; developed the model and performed regulatory and structural interpretation. GET predicts expression across 213 human cell types and identified a transcription-factor interaction implicated in inherited leukemia risk. Co-first and co-corresponding author, <i>Nature</i> (2025).	
2026	p53–ER regulation in breast cancer and chromatin changes in muscle aging Developed Motiverse for GPU-accelerated motif-spacing analysis. Identified primate-specific Alu-derived p53–ER motif spacing of approximately 73 bp, supporting potential cooperation on the same nucleosome in breast cancer. Identified regulatory elements with age-associated accessibility changes in muscle cell types and tenocytes. First and corresponding author; unpublished manuscript (2026). Chromnitron: protein–DNA interaction modeling Co-designed the data curation and quality-control pipeline, model architecture, training paradigm, and mechanistic interpretation analyses of the protein encoder. The study identified and experimentally validated regulators of T-cell exhaustion. Co-first author; <i>bioRxiv</i> (2025; revised 2026).	

Published as Fu, X. and Fu, A. X. *Equal contribution; †corresponding author.

1. **Fu, X.**^{*†}, Mo, S.^{*}, Buendia, A.^{*}, Laurent, A. P., Shao, A., Alvarez-Torres, M. d. M., Yu, T., Tan, J., Su, J., Sagatelian, R., Ferrando, A. A., Ciccio, A., Lan, Y., Owens, D. M., Palomero, T., Xing, E. P.[†], and Rabadan, R.[†]. A foundation model of transcription across human cell types. *Nature* 637, 965–973 (2025). doi:10.1038/s41586-024-08391-z.
2. **Fu, A. X.**, Lui, K. N.-C., Tang, C. S.-M., Ng, R. K., Lai, F. P.-L., Lau, S.-T., Li, Z., Garcia-Barcelo, M.-M., Sham, P. C., Tam, P. K.-H., Ngan, E. S.-W., and Yip, K. Y. Whole-genome analysis of noncoding genetic variations identifies multiscale regulatory element perturbations associated with Hirschsprung disease. *Genome Research* 30, 1618–1632 (2020). doi:10.1101/gr.264473.120.
3. Su, J., Reynier, J.-B., **Fu, X.**, Zhong, G., Jiang, J., Escalante, R. S., Wang, Y., Aparicio, L., Izar, B., Knowles, D. A., and Rabadan, R. Smoother: a unified and modular framework for incorporating structural dependency in spatial omics data. *Genome Biology* 24, 291 (2023). doi:10.1186/s13059-023-03138-x.
4. Au, T. Y. K., Yip, R. K. H., Wynn, S. L., Tan, T. Y., **Fu, A. X.**, Geng, Y. H., Szeto, I. Y. Y., Niu, B., Yip, K. Y., Cheung, M. C. H., Lovell-Badge, R., and Cheah, K. S. E. Hypomorphic and dominant-negative impact of truncated SOX9 dysregulates Hedgehog-Wnt signaling, causing campomelia. *PNAS* 120, e2208623119 (2023).
5. Zhang, Z., Li, H., Weng, H., Zhou, G., Chen, H., Yang, G., Zhang, P., Zhang, X., Ji, Y., Ying, K., Liu, B., Xu, Q., Tang, Y., Zhu, G., Liu, Z., Xia, S., Yang, X., Dong, L., Zhu, L., Zeng, M., Yuan, Y., Yang, Y., Zhang, N., Xu, X., Pang, W., Zhang, M., Zhang, Y., Zhen, K., Wang, D., Lei, J., Wu, S., Shu, S., Zhang, Y., Zhang, S., Gao, Q., Huang, Q., Deng, C., **Fu, X.**, Chen, G., Duan, W., Wan, J., Xie, W., Zhang, P., Wang, S., Yang, P., Zuo, X., Zhai, Z., and Wang, C. Genome-wide association analyses identified novel susceptibility loci for pulmonary embolism among Han Chinese population. *BMC Medicine* 21, 153 (2023).
6. Cao, Q., Zhang, Z., **Fu, A. X.**, Wu, Q., Lee, T.-L., Lo, E., Cheng, A. S. L., Cheng, C., Leung, D., and Yip, K. Y. A unified framework for integrative study of heterogeneous gene regulatory mechanisms. *Nature Machine Intelligence* 2, 447–456 (2020).
7. Tang, C. S.-M., Li, P., Lai, F. P.-L., **Fu, A. X.**, Lau, S.-T., So, M. T., Lui, K. N.-C., Li, Z., Zhuang, X., Yu, M., et al. Identification of genes associated with Hirschsprung disease, based on whole-genome sequence analysis, and potential effects on enteric nervous system development. *Gastroenterology* 155, 1908–1922.e5 (2018).
8. Gao, C., Zhuang, X., Cui, Y., **Fu, X.**, He, Y., Zhao, Q., Zeng, Y., Shen, J., Luo, M., and Jiang, L. Dual roles of an Arabidopsis ESCRT component FREE1 in regulating vacuolar protein transport and autophagic degradation. *PNAS* 112, 1886–1891 (2015).

Reviews and Perspectives

1. Alvarez-Torres, M. d. M., **Fu, X.**, and Rabadan, R. Illuminating the noncoding genome in cancer using artificial intelligence. *Cancer Research* 85, 2368–2375 (2025). doi:10.1158/0008-5472.CAN-25-0482.
2. **Fu, X.** and Rabadan, R. Understanding variants of unknown significance: the computational frontier. *The Oncologist* 29, 653–657 (2024). doi:10.1093/oncolo/oyae103.

Preprints and Unpublished Manuscripts

1. Hayward, S. B., Vaitsiankova, A., Lama-Diaz, T., Chou, J., Taglialatela, A., Huang, J.-W., Wijesekarathanthi, Y., Heyza, J. R., Leuzzi, G., Chen, C., Wong, N., Lhakhang, T., **Fu, X.**, Buendia, A. L., Gheorghe, V., Esmaeili Anvar, N., Schmidt, J. C., Nussenzweig, A., Rabadan, R., Costanzo, V., Guérois, R., Hart, T., and Ciccio, A. Mapping the genetic landscape of the DNA damage response with Cas12a-based combinatorial knockout screens. Accepted in principle, *Nature Biotechnology*. Preprint: *bioRxiv* (2026). doi:10.64898/2026.06.07.728858.
2. Tan, J. *, **Fu, X.** *, Ling, X. *, Mo, S., Bai, J., Rabadan, R., Fenyo, D., Boeke, J. D., Tsigos, A., and Xia, B. Decoding the gene regulatory landscape through multimodal learning of protein–DNA interactions. *bioRxiv* (2025; revised January 18, 2026). doi:10.1101/2025.08.17.670761.
3. **Fu, X.**, Reglero, C., Swamy, V., Loh, J. W., Khiabani, H., Alberio, R., Forouhar, F., AlQuraishi, M., Ferrando, A. A., and Rabadan, R. Computational structure prediction methods enable the systematic identification of oncogenic mutations. *bioRxiv* (2022). doi:10.1101/2022.11.18.517015.
4. **Fu, X.**[†], Rabadan, R., and Levine, A. J. Alu-derived p53–nuclear-receptor grammar routes Cyclin D control in breast and prostate cancer. Unpublished manuscript (2026). [Manuscript PDF](#).

Workshop Contributions

1. Mo, S., **Fu, X.**, Hong, C., Chen, Y., Zheng, Y., Tang, X., Lan, Y., Shen, Z., and Xing, E. Multi-modal self-supervised pre-training for large-scale genome data. *NeurIPS AI for Science Workshop* (2021).

Additional Research Contributions

2021–2023	Structure-informed interpretation of cancer mutations Used AlphaFold2-based structural modeling, molecular dynamics, and evolutionary features to identify and interpret oncogenic hotspot mutations in tumor suppressors and oncogenes. First author, <i>bioRxiv</i> (2022). Advisor: Raul Rabadan.
2025–present	Eureka: AI-assisted genomic discovery Developed a platform integrating literature, genetic, regulatory, and structural evidence for variant interpretation and therapeutic target prioritization. Built benchmarks with controls for information leakage to evaluate target prioritization and enrichment for clinical success.
2017–2020	Disease-associated noncoding risk variants Developed a multiscale framework to prioritize regulatory elements, genes, pathways, and transcription-factor programs from whole-genome sequencing. Applied it to Hirschsprung disease; collaborators experimentally validated candidate mechanisms. The Chinese University of Hong Kong; advisor: Kevin Yip.
2018–2019	Network embedding for gene regulation Developed and interpreted a semi-supervised framework integrating physical and regulatory interactions in heterogeneous biological networks. The Chinese University of Hong Kong; advisor: Kevin Yip.
2020	Venous thromboembolism GWAS Performed genome-wide association analyses of susceptibility variants in a Han Chinese population. Shenzhen Zaozhidao Company.

Research Software

GET: Released open software for model training, regulatory interpretation, and analysis.

Astro (2026–present): Built and launched a browser-based system to coordinate AI research agents across local, cloud, and HPC environments, with reproducible task records. astroanywhere.com.

Invited Talks

- Nov. 2025, Columbia University, AI at VP&S Workshop: Foundation Models Across Scales.
- July 2025, Google Genomics.
- June 2025, Cancer Convergence Education Network, University of Chicago.
- Mar. 2025, New York University.
- Feb. 2025, Stanford University.
- Feb. 2025, Genentech.
- Feb. 2025, Tsinghua University.
- Jan. 2025, EMBL Heidelberg.
- Jan. 2025, Sanford Burnham Prebys Medical Discovery Institute.
- Nov. 2024, Cancer Convergence Education Network, Spanish National Cancer Research Centre.
- May 2024, Cancer Convergence Education Network, University of Rome.
- Dec. 2023, Cancer Convergence Education Network, Universitat Politecnica de Catalunya.
- Oct. 2023, Cancer Convergence Education Network, Keio University.

Mentoring

- Mentoring two PhD students on foundation-model training, including extensions of GET.
- Previously mentored one research assistant; one mentee is now a PhD student at Stanford University.

Teaching

- Guest lecturer, *Introduction to Computational Biomedicine and Health* (BINF 4001 / COMS 4560), Columbia University: “Single Cell Foundation Models” (2025); invited to return with “Introduction to Biological Foundation Models” (2026, upcoming).
- Teaching assistant for a full term of Introduction to Database Systems (undergraduate), The Chinese University of Hong Kong. Responsible for weekly computational labs and grading.
- Tutored biological and medical collaborators in AI and machine learning, including pretraining and fine-tuning biological foundation models.

Professional Service

Reviewer: *Nature* (2026); *Science* (2025); MLCB (2023, 2024); SPIGM workshop at ICML (2026); AI4Science workshop at ICML (2024); NeurIPS workshops.

Grant Proposal Contributions

Contributed to writing R01, ARPA-H, Evans Foundation, and DoD Idea grant proposals.

Honors and Awards

- | | |
|------|---|
| 2021 | Champion, DeeCamp Artificial Intelligence Bootcamp , Sinovation Ventures. |
| 2017 | Certificate of Merit Award for Teaching Assistant , The Chinese University of Hong Kong. |

Last updated: September 2026.