Fangping Wan Email : fangping.wan@pennmedicine.upenn.edu
Google Scholar, Homepage, Github

RESEARCH INTERESTS

AI for Science, Deep Learning, Generative Modeling for Biomolecular Design, Large Language Models, Graph Neural Networks, Drug Discovery, Computational Biology, Antibiotic Discovery

RESEARCH SUMMARY

Fangping Wan's research lies broadly in the field of **AI for Science**, focusing on developing and applying advanced machine learning methodologies to address key challenges in biomolecular modeling and therapeutic discovery. His work integrates **large language models**, **graph neural networks**, and **generative AI frameworks** to represent and design protein sequences, molecular structures, and biological interactions.
By combining **predictive modeling** with **generative design**, Fangping explores new paradigms for understanding molecular function and accelerating the identification of effective antimicrobial and therapeutic candidates. Positioned at the intersection of **computational biology**, **artificial intelligence**, and **biomedical engineering**, his research bridges methodological innovation with biological application, contributing to a unified vision of how AI can transform discovery in the life sciences and chemistry.

EDUCATION

- **Tsinghua University** Beijing, China
  Ph.D. in Computer Science and Technology; Advisor: Jianyang Zeng *September 2015 - July 2020*
- **Tianjin University** Tianjin, China
  B.S. in Computer Science and Technology; GPA: 3.85 *September 2011 - July 2015*

EXPERIENCE

- **Perelman School of Medicine at University of Pennsylvania** Philadelphia, PA USA
  Computational Scientist; Advisor: César de la Fuente *August 2024 - Now*
- **Perelman School of Medicine at University of Pennsylvania** Philadelphia, PA USA
  Postdoctoral Researcher; Advisor: César de la Fuente *August 2021 - August 2024*
- **Silexon AI Technology** Beijing, China
  Machine learning engineer *July 2020 - July 2021*

FUNDING EXPERIENCE

- Co-PI, ACCESS Allocations Award CIS250020, "Computational mining of novel antibiotics from biological dark matter." 1.5M ACCESS credits (GPU/CPU resources), Jun 2025 – Jun 2026. PI: César de la Fuente.
- Co-PI, ACCESS Allocations Award CIS230342, "Computationally mining antibiotics from genomes and proteomes of organisms." Indiana Jetstream2 and SDSC Expanse allocation (1.43M GPU SUs + 100K CPU hours + 10 TB storage), Dec 2023 – Dec 2024. PI: César de la Fuente.
- Contributor (AI/ML component), Defense Threat Reduction Agency (DTRA) Award HDTRA1-14-24-FRCWMD-BAA, "Non-negative matrix factorization approach to identify host–pathogen interactions and develop medical countermeasures to stem outbreaks." Funded 2022. PI: César de la Fuente.

PUBLICATIONS

First or co-first author papers

- Tianang Leng*, **Fangping Wan***, Marcelo D. T. Torres, Cesar de la Fuente. Predicting and generating antibiotics against future pathogens with ApexOracle. *NeurIPS 2025 2nd Workshop FM4LS*, 2025. (Journal version in preparation)
- Marcelo D. T. Torres*, **Fangping Wan***, Cesar de la Fuente. Deep learning reveals antibiotics in the archaeal proteome. *Nature Microbiology*, 2025.
- Angela Cesaro*, **Fangping Wan***, Haoyuan Shi*, Kaiyang Wang, C Mark Maupin, Matt L Barker, Jiqian Liu, Stephen J Fox, Jingjie Yeo, Cesar de la Fuente-Nunez. Antiviral discovery using sparse datasets by integrating experiments, molecular simulations, and machine learning. *Cell Reports Physical Science*, 2025.

- **Fangping Wan***, Marcelo D. T. Torres*, Changge Guan, Cesar de la Fuente-Nunez. Tutorial: guidelines for the use of machine learning methods to mine genomes and proteomes for antibiotic discovery. *Nature Protocols*, 2025.
- **Fangping Wan***, Marcelo D. T. Torres*, Jacqueline Peng, Cesar de la Fuente. Deep-learning-enabled antibiotic discovery through molecular de-extinction. *Nature Biomedical Engineering*, 2024.
- **Fangping Wan***, Felix Wong*, James J. Collins, Cesar de la Fuente. Machine Learning for Antimicrobial Peptide Identification and Design. *Nature Reviews Bioengineering*, 2024.
- **Fangping Wan***, Daphne Kontogiorgos-Heintz*, Cesar de la Fuente-Nunez. Deep generative models for peptide design. *Digital Discovery*, 2022.
- Yiyue Ge*, Tingzhong Tian*, Sulin Huang*, **Fangping Wan***, Jingxin Li, Shuya Li, Hui Yang, Lixiang Hong, Nian Wu, Enming Yuan, Lili Cheng, Yipin Lei, Hantao Shu, Xiaolong Feng, Ziyuan Jiang, Ying Chi, Xiling Guo, Lunbiao Cui, Liang Xiao, Zeng Li, Chunhao Yang, Zehong Miao, Haidong Tang, Ligong Chen, Hainian Zeng, Dan Zhao, Fengcai Zhu, Xiaokun Shen, Jianyang Zeng. An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *Signal Transduction and Targeted Therapy*, 2021.
- **Fangping Wan***, Shuya Li*, Tingzhong Tian, Yipin Lei, Dan Zhao, Jianyang Zeng. EXP2SL: A Machine Learning Framework for Cell-Line-Specific Synthetic Lethality Prediction. *Frontiers in Pharmacology*, 2020.
- Shuya Li*, **Fangping Wan***, Hantao Shu, Tao Jiang, Dan Zhao, Jianyang Zeng. MONN: A Multi-objective Neural Network for Predicting Compound-Protein Interactions and Affinities. *24th International Conference on Research in Computational Molecular Biology (RECOMB 2020)*, also in *Cell Systems*, 2020.
- **Fangping Wan***, Yue Zhu*, Hailin Hu*, Antao Dai, Xiaoqing Cai, Ligong Chen, Haipeng Gong, Tian Xia, Dehua Yang, Ming-Wei Wang, Jianyang Zeng. DeepCPI: A Deep Learning-based Framework for Large-scale *in silico* Drug Screening. *Genomics, Proteomics & Bioinformatics*, 2019.
- **Fangping Wan**, Lixiang Hong, An Xiao, Tao Jiang, Jianyang Zeng. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics*, 2019.

Co-author papers

- Natalie Maus, Kyurae Kim, Yimeng Zeng, Haydn Thomas Jones, **Fangping Wan**, Marcelo D. T. Torres, Cesar de la Fuente-Nunez, Jacob R. Gardner. Covering Multiple Objectives with a Small Set of Solutions Using Bayesian Optimization. *NeurIPS*, 2025.
- Marcelo D. T. Torres, Leo Tianlai Chen, **Fangping Wan**, Pranam Chatterjee, Cesar de la Fuente-Nunez. Generative latent diffusion language modeling yields anti-infective synthetic peptides. *Cell Biomaterials*, 2025.
- Khalid Usman, **Fangping Wan**, Dan Zhao, Jian Peng, Jianyang Zeng. Analyzing Large-Scale Single-Cell RNA-Seq Data Using Coreset. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2024.
- Angela Cesaro, Mojtaba Bagheri, Marcelo D. T. Torres, **Fangping Wan**, Cesar de la Fuente-Nunez. Deep learning tools to accelerate antibiotic discovery. *Expert Opinion on Drug Discovery*, 2023.
- Yiming Wang, **Fangping Wan**, Zhangtao Chen, Jonathan Nukpezah, Cesar de la Fuente-Nunez, Ravi Radhakrishnan. Structure-Based Prediction of Kinase Activation amidst a Varied Mutational Landscape Using Privileged Learning. *AIChE Annual Meeting*, 2023.
- Han Li, Xinyi Zhao, Shuya Li, **Fangping Wan**, Dan Zhao, Jianyang Zeng. Improving molecular property prediction through a task similarity enhanced transfer learning strategy. *Iscience*, 2022.
- Eugene F Douglass Jr., Robert J Allaway, Bence Szalai, Wenyu Wang, Tingzhong Tian, Adrià Fernández-Torras, Ron Realubit, Charles Karan, Shuyu Zheng, Alberto Pessia, Ziaurrehman Tanoli, Mohieddin Jafari, **Fangping Wan**, Shuya Li, Yuanpeng Xiong, Miquel Duran-Frigola, Martino Bertoni, Pau Badia-i-Mompel, Lídia Mateo, Oriol Guitart-Pla, Verena Chung, Jing Tang, Jianyang Zeng, Patrick Aloy, Julio Saez-Rodriguez, Justin Guinney, Daniela S Gerhard, Andrea Califano. A community challenge for a pancancer drug mechanism of action inference from perturbational profile data. *Cell Reports Medicine* 2022.
- Adi L Tarca, Balint Armin Pataki, Roberto Romero, Marina Sirota, Yuanfang Guan, Rintu Kutum, Nardhy Gomez-Lopez, Bogdan Done, Gaurav Bhatti, Thomas Yu, Gaia Andreoletti, Tinnakorn Chaiworapongsa, **The DREAM Preterm Birth Prediction Challenge Consortium**, Sonia S Hassan, Chaur-Dong Hsu, Nima Aghaeepour, Gustavo Stolovitzky, Istvan Csabai, James C Costello. Crowdsourcing assessment of maternal blood multi-omics for predicting gestational age and preterm birth. *Cell Reports Medicine*, 2021.
- Yipin Lei, Shuya Li, Ziyi Liu, **Fangping Wan**, Tingzhong Tian, Shao Li, Dan Zhao, Jianyang Zeng. A deep-learning framework for multi-level peptide–protein interaction prediction. *Nature communications*, 2021.
- Anna Cichonska, Balaguru Ravikumar, Robert J Allaway, Sungjoon Park, **Fangping Wan**, Olexandr Isayev, Shuya Li, Michael J Mason, Andrew Lamb, Minji Jeon, Sunkyu Kim, Mariya Popova, Jianyang Zeng, Kristen Dang, Gregory Koytiger, Jaewoo Kang, Carrow I Wells, Timothy M Willson, Tudor I Oprea, Avner Schlessinger, David H Drewry, Gustavo A Stolovitzky, Krister Wennerberg, Justin Guinney, Tero Aittokallio. Crowdsourced mapping of unexplored target space of kinase inhibitors. *Nature Communications*, 2021.

- Peiyuan Feng, An Xiao, Meng Fang, **Fangping Wan**, Shuya Li, Peng Lang, Dan Zhao, and Jianyang Zeng. A novel machine learning based framework for modeling transcription elongation. *Proceedings of the National Academy of Sciences (PNAS)*, 2021.
- Han Li, Xinyi Zhao, Shuya Li, **Fangping Wan**, Dan Zhao, and Jianyang Zeng. MoTSE: an interpretable task similarity estimator for small molecular property prediction tasks. *RECOMB*, 2021.
- Lixiang Hong, Jinjian Lin, Shuya Li, **Fangping Wan**, Hui Yang, Tao Jiang, Dan Zhao, Jianyang Zeng. A novel machine learning framework for automated biomedical relation extraction from large-scale literature repositories. *Nature Machine Intelligence*, 2020.
- Rong Ma, Yi Li, Chenxing Li, **Fangping Wan**, Hailin Hu, Wei Xu, Jianyang Zeng. Secure multiparty computation for privacy-preserving drug discovery. *Bioinformatics*, 2020.
- Yan Hu, Ziqiang Wang, Hailin Hu, **Fangping Wan**, Lin Chen, Yuanpeng Xiong, Xiaoxia Wang, Dan Zhao, Weiren Huang, Jianyang Zeng. ACME: pan-specific peptide–MHC class I binding prediction through attention-based deep neural networks. *Bioinformatics*, 2019.

In preparation

- Yimeng Zeng, Natalie Maus, Haydn Thomas Jones, Jeffrey Tao, **Fangping Wan**, Marcelo D. T. Torres, Cesar de la Fuente-Nunez, Ryan Marcus, Osbert Bastani, Jacob R. Gardner. Large Scale Multi-Task Bayesian Optimization with Large Language Models.
- Angela Cesaro, **Fangping Wan**, Marcelo D. T. Torres, Cesar de la Fuente-Nunez. Design of multimodal antibiotics against intracellular infections using deep learning.
- Marcelo D. T. Torres*, Yimeng Zeng*, **Fangping Wan***, Natalie Maus, Jacob R. Gardner, Cesar de la Fuente-Nunez. A generative artificial intelligence approach for antibiotic optimization.
- Changge Guan*, **Fangping Wan***, Marcelo D. T. Torres, Cesar de la Fuente-Nunez. Improving functional protein generation via foundation model-derived latent space likelihood optimization.

Miscellaneous

- **Fangping Wan***, Cesar de la Fuente-Nunez. Mining for antimicrobial peptides in sequence space. *Nature Biomedical Engineering (news & views)*, 2023.

CONFERENCE PRESENTATIONS

- **Invited Talks**:

  Fangping Wan, et al. (November, 2019).Best Performer Talk I: An ensemble learning based approach for the IDG-DREAM Drug-Kinase Binding Prediction Challenge. *The 12th annual RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges.* New York, USA.

  Fangping Wan, et al. (November, 2020). A multitask neural network approach for predicting drug targets from chemogenomics data. *The 13th annual RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges.* Virtually held.

AWARDS

Competitions

- Winner of CTD^2 Pancancer Drug Activity DREAM Challenge, 2020
- Winner of IDG-DREAM Drug-Kinase Binding Prediction Challenge, 2019

Scholarships

- Toyota Scholarship, 2019
- Baidu Future Star Award for Excellent Student, 2018
- BaoSteel Excellent Student, 2014
- CCF Excellent College Student, 2014
- National Scholarship, 2013
- TianJin Municipal People's Government Scholarship, 2012

MENTORING & SUPERVISION

- Supervised **Edward Sun** (Undergraduate, University of Pennsylvania, Fall 2024 – Summer 2025) on mining antimicrobial peptides (AMPs) from genomes using machine learning and developing classifier-free guidance flow-matching models for AMP generation.
- Supervised **Pia Geismar** (Undergraduate, University of Pennsylvania, Fall 2024 – Spring 2025) on mining AMPs from phage genomes using machine learning approaches.
- Supervised **Abdullah Ali** (Undergraduate, Swarthmore College, Summer 2024) on identifying AMPs from ancient mitochondrial DNA using machine learning frameworks.

SERVICES

Conference Reviewer:

- International Conference on Research in Computational Molecular Biology (RECOMB, 2023–2025)
- Intelligent Systems for Molecular Biology (ISMB, 2020, 2023, 2024)
- ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB, 2021, 2025)
- International Conference on Intelligent Computing (ICIC, 2018, 2019)

Journal Reviewer:

- Nature Communications
- Bioinformatics
- IEEE/ACM Transactions on Computational Biology and Bioinformatics
- Computational and Structural Biotechnology Journal
- Frontiers in Molecular Biosciences
- Frontiers in Pharmacology
- Frontiers in Medicine
- Frontiers in Chemistry
- Frontiers in Genetics
- Frontiers in Drug Safety and Regulation
- Frontiers in Microbiology
- Journal of Computer-Aided Molecular Design
- Scientific Reports
- BMC Bioinformatics
- International Journal of Peptide Research and Therapeutics

Teaching Assistant:

- Advanced Applications of Machine Learning (40470224-0), Tsinghua University, Fall 2016
- Computational Biology (30470093-0), Tsinghua University, Spring 2016

PROGRAMMING SKILLS

- **Languages**: Python, MATLAB, C++