Jitao David Zhang

A Computational Biologist in Drug Discovery

F. Hoffmann-La Roche AG
4070 Basel, Switzerland
jitao_david.zhang@roche.com
jdzhang.me

Key facts

- Computational biologist with 10+ years' industrial research experience
- Critical contributions to drug candidates and marketed products
- Author of patents, open-source software, and 40+ peer-reviewed publications
- Inventing and implementing proprietary technologies and workflows
- Teaching courses and supervising postdocs, students, and apprentices
- Experience in establishing research group and leading teams

Working experience

2022– **Expert scientist**, *F. Hoffmann-La Roche AG*, Basel, Switzerland
Devising and implementing the concept of Iterative Quantitative Benefit-Risk Evaluation (IQ-BRE) for drug-discovery projects. Highlights include critical support for a Phase 1 clinical trial of the drug candidate Selnoflast, causal inference in drug discovery and development, and Kinex, a new workflow characterizing kinase-associated safety and efficacy profiles of drug candidates.

2021– **Computer-science apprenticeship trainer and exam expert**, Switzerland
Training vocational apprentices specializing in application development and system programming. I co-host qualification examinations as a certified expert.

2018– **Lecturer**, *University Basel of Basel*, Basel, Switzerland
Teaching graduate- and post-graduate-level courses at Department of Mathematics and Computer Science, and contributing to lecture series running in other departments.

2018-2022 **Senior principal scientist**, *F. Hoffmann-La Roche AG*
Multiscale modelling of drug mechanism and safety, application of PACE (Pathway Annotated Chemical Ensemble, formerly known as SPARK) library for screening, development of machine-learning empowered toxicity screening assays. Highlights include the TeraTox assay, the identification of RepSox as a tool compound for ophthalmology, de-risking molecules in development, and informing key decisions for several projects.

2016-2018 **Principal scientist**, *F. Hoffmann-La Roche AG*
Developing the molecular phenotyping platform, developing the chemogenomic library SPARK, and supporting projects in the area of infectious diseases. Highlights include a pilot study of molecular phentoyping for phenotypic drug discovery, identification of EGF update as an early readout for nephrotoxicity, and development of the BioQC software.

2013-2016 **Senior scientist**, *F. Hoffmann-La Roche AG*
Developing the molecular phenotyping platform, supporting Phase IV study of Oseltamivir/Tamiflu (the IRIS study), and applying multi-omics data analysis for disease modelling. Highlights include identification of the time-dependent critical role TGF-beta signaling in breast cancer metastasis, and genomic analysis of the molecular neuropathology of tuberous sclerosis using a human stem cell model.

2011-2013 **Scientist**, *F. Hoffmann-La Roche AG*
Profiling compound efficacy and safety with omics methods, data mining, and machine learning. Highlights include the identification of a predictive gene network for drug safety by analyzing the TG-GATEs database, identification of the causal role of the JAK/STAT3 pathway in turning white adipose to brown, and informing decisions in multiple drug-discovery projects.

2006–2008 **Research associate**, *German Cancer Research Center*, Heidelberg, Germany
Bioinformatics software and algorithm development in the HUSAR Bioinformatics group

2005–2006 **Part-time translator, journalist, and associate editor**, *Kicker China*, Beijing, China
Translating German articles on European football leagues

Education

2008–2011 **Dr.rer.nat. Bioinformatics**, *German Cancer Research Center/ Universität Heidelberg*
Computational and statistical approaches to study gene networks, supervised by Dr. Stefan Wiemann

2007–2007 **Marie-Curie Fellow**, *Huber Group, European Institute of Bioinformatics*, Cambridge, UK

2006–2008 **M.Sc. Bioinformatics**, *Universität Heidelberg*, Germany

2002–2006 **B.Sc. Biology, hon.**, *Peking University*, Beijing, China

Patents

2020 Oligonucleotides for modulating RTEL1 expression

2018 Pyrrolo[2,3-b]pyrazine compounds as cccDNA inhibitors for the treatment of Hepatitis B Virus (HBV) infection

Book chapters

2016 Applied Biclustering Methods for Big and High-Dimensional Data Using R, Chapman & Hall/CRC Biostatistics Series, edited by Adetayo Kasim, S. H., Ziv Shkedy, Sebastian Kaiser & Talloen, W.

Selected open-source software

BESCA Single-cell omics data analysis pipeline, built together with BEDA colleagues

BioQC Detecting tissue heterogeneity in high-throughput expression data

ddCt A pipeline to collect, analyse and visualize qRT-PCR results

HBVouroboros A Snakemake-based workflow for hepatitis B virus (HBV) drug discovery

KEGGgraph Data mining and network analysis of biological pathways as graphs

Kinex Inference of causal serine/threonine kinases from phosphoproteomics data

ribios A collection of R packages for computational biology tasks in drug discovery

RTCA Analysis and visualization of data from Roche(R) xCELLigence RTCA systems

Publications with (co-)first or correspondence authorship

[1] Tom Michoel and Jitao David Zhang. “Causal inference in drug discovery and development”. In: *Drug Discovery Today* 28.10 (2023), p. 103737.

[2] Li Wang et al. “Discovery of a first-in-class orally available HBV cccDNA inhibitor”. In: *Journal of Hepatology* 78.4 (2023), pp. 742–753.

[3] Manuela Jaklin et al. “Optimization of the TeraTox assay for preclinical teratogenicity assessment”. In: *Toxicological Sciences* 188.1 (2022), pp. 17–33.

[4] Gregor Sturm, Markus List, and Jitao David Zhang. “Tissue heterogeneity is prevalent in gene expression studies”. In: *NAR Genomics and Bioinformatics* 3.3 (2021), lqab077.

[5] Jitao David Zhang. “Ten simple rules for doing a postdoc in pharma”. In: *PLOS Computational Biology* 17.6 (2021), e1008989.

[6] Filip Roudnicky et al. “Inducers of the endothelial cell barrier identified through chemogenomic screening in genome-edited hPSC-endothelial cells”. In: *Proceedings of the National Academy of Sciences* 117.33 (2020), pp. 19854–19865.

[7] Jitao David Zhang et al. “Multiscale modelling of drug mechanism and safety”. In: *Drug Discovery Today* 25.3 (2020), pp. 519–534.

[8] Faye Marie Drawnel et al. “Molecular Phenotyping Combines Molecular Information, Biological Relevance, and Patient Data to Improve Productivity of Early Drug Discovery”. In: *Cell Chemical Biology* 24.5 (2017), pp. 624–634.

[9] Jitao David Zhang et al. “Detect tissue heterogeneity in gene expression data with BioQC”. In: *BMC Genomics* 18 (2017), pp. 1–9.

[10] Nils Grabole et al. “Genomic analysis of the molecular neuropathology of tuberous sclerosis using a human stem cell model”. In: *Genome Medicine* 8.1 (2016), pp. 1–14.

[11] Jitao David Zhang et al. “Pathway reporter genes define molecular phenotypes of human cells”. In: *BMC Genomics* 16.1 (2015), pp. 1–10.

[12] JD Zhang et al. “Data mining reveals a network of early-response genes as a consensus signature of drug-induced in vitro and in vivo toxicity”. In: *The Pharmacogenomics Journal* 14.3 (2014), pp. 208–216.

[13] Jitao David Zhang et al. “Highly sensitive amplicon-based transcript quantification by semiconductor sequencing”. In: *BMC Genomics* 15.1 (2014), pp. 1–8.

[14] Jitao David Zhang et al. “Time-resolved human kinome RNAi screen identifies a network regulating mitotic-events as early regulators of cell proliferation”. In: *PloS ONE* 6.7 (2011), e22176.

[15] S Uhlmann et al. “miR-200bc/429 cluster targets PLCγ1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer”. In: *Oncogene* 29.30 (2010), pp. 4297–4306.

[16] Jitao David Zhang and Stefan Wiemann. “KEGGgraph: a graph approach to KEGG PATHWAY in R and Bioconductor”. In: *Bioinformatics* 25.11 (2009), pp. 1470–1471.

Other peer-reviewed publications

[17] Barbara Klughammer et al. “A randomized, double-blind phase 1b study evaluating the safety, tolerability, pharmacokinetics and pharmacodynamics of the NLRP3 inhibitor selnoflast in patients with moderate to severe active ulcerative colitis”. In: *Clinical and Translational Medicine* 13.11 (2023), e1471.

[18] Dongdong Chen et al. “Discovery of Novel cccDNA Reducers toward the Cure of Hepatitis B Virus Infection”. In: *Journal of Medicinal Chemistry* 65.16 (2022), pp. 10938–10955.

[19] Lorenzo Gatti et al. “Cross-reactive immunity potentially drives global oscillation and opposed alternation patterns of seasonal influenza A viruses”. In: *Scientific Reports* 12.1 (2022), p. 8883.

[20] Sophia Clara Mädler et al. “BESCA, a single-cell transcriptomics analysis toolkit to accelerate translational research”. In: *NAR Genomics and Bioinformatics* 3.4 (2021), lqab102.

[21] Marvin Reich et al. “Alzheimer’s Risk Gene TREM2 Determines Functional Properties of New Type of Human iPSC-Derived Microglia”. In: *Frontiers in Immunology* 11 (2021), p. 617860.

[22] Solveig Badillo et al. “An introduction to machine learning”. In: *Clinical Pharmacology & Therapeutics* 107.4 (2020), pp. 871–885.

[23] Simon Gutbier et al. “Large-Scale Production of Human iPSC-Derived Macrophages for Drug Screening”. In: *International Journal of Molecular Sciences* 21.13 (2020), p. 4808.

[24] Manuela Jaklin et al. “Focus on germ-layer markers: A human stem cell-based model for in vitro teratogenicity testing”. In: *Reproductive Toxicology* 98 (2020), pp. 286–298.

[25] Mohamad Zaidan et al. “Signaling pathways predisposing to chronic kidney disease progression”. In: *JCI Insight* 5.9 (2020).

[26] Sarvenaz Choobdar et al. “Assessment of network module identification across complex diseases”. In: *Nature Methods* 16.9 (2019), pp. 843–852.

[27] Filip Roudnicky et al. “Modeling the Effects of Severe Metabolic Disease by Genome Editing of hPSC-Derived Endothelial Cells Reveals an Inflammatory Phenotype”. In: *International Journal of Molecular Sciences* 20.24 (2019), p. 6201.

[28] Gregor Sturm et al. “Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology”. In: *Bioinformatics* 35.14 (2019), pp. i436–i445.

[29] Henrik Mueller et al. “A novel orally available small molecule that inhibits hepatitis B virus expression”. In: *Journal of Hepatology* 68.3 (2018), pp. 412–420.

[30] Franziska Boess et al. “Use of early phenotypic in vivo markers to assess human relevance of an unusual rodent non-genotoxic carcinogen in vitro”. In: *Toxicology* 379 (2017), pp. 48–61.

[31] Annie Moisan et al. “Inhibition of EGF Uptake by Nephrotoxic Antisense Drugs In Vitro and Implications for Preclinical Safety Profiling”. In: *Molecular Therapy-Nucleic Acids* 6 (2017), pp. 89–105.

[32] U Raza et al. “The miR-644a/CTBP1/p53 axis suppresses drug resistance by simultaneous inhibition of cell survival and epithelial-mesenchymal transition in breast cancer.” In: *Oncotarget* (2016).

[33] Erhard van der Vries et al. “Outcomes and susceptibility to neuraminidase inhibitors in individuals infected with different influenza B lineages: the influenza resistance information study”. In: *The Journal of Infectious Diseases* 213.2 (2016), pp. 183–190.

[34] Florian Haller et al. “Combined DNA methylation and gene expression profiling in gastrointestinal stromal tumors reveals hypomethylation of SPP1 as an independent prognostic factor”. In: *International Journal of Cancer* 136.5 (2015), pp. 1013–1023.

[35] Nicole Lenz et al. “Antiviral Innate Immune Activation in HIV-Infected Adults Negatively Affects H1/IC31-Induced Vaccine-Specific Memory CD4+ T Cells”. In: *Clinical and Vaccine Immunology* 22.7 (2015), pp. 688–696.

[36] Annie Moisan et al. “White-to-brown metabolic conversion of human adipocytes by JAK inhibition”. In: *Nature Cell Biology* 17.1 (2015), pp. 57–67.

[37] Jia Xu et al. “14-3-3ζ turns TGF-β’s function from tumor suppressor to metastasis promoter in breast cancer by contextual changes of Smad partners from p53 to Gli2”. In: *Cancer Cell* 27.2 (2015), pp. 177–192.

[38] Stefan Aigner et al. “Human pluripotent stem cell models of autism spectrum disorder: emerging frontiers, opportunities, and challenges towards neuronal networks in a dish”. In: *Psychopharmacology* 231 (2014), pp. 1089–1104.

[39] Umar Raza, Jitao David Zhang, and Özgür Şahin. “MicroRNAs: master regulators of drug resistance, stemness, and metastasis”. In: *Journal of Molecular Medicine* 92 (2014), pp. 321–336.

[40] Sumaiyah K Rehman et al. “14-3-3ζ Orchestrates Mammary Tumor Onset and Progression via miR-221–Mediated Cell Proliferation”. In: *Cancer Research* 74.1 (2014), pp. 363–373.

[41] Inn H Yuk et al. “Effects of copper on CHO cells: insights from gene expression analyses”. In: *Biotechnology progress* 30.2 (2014), pp. 429–442.

[42] Liana Adam et al. “Plasma microRNA profiles for bladder cancer detection”. In: *Urologic Oncology: Seminars and Original Investigations*. Vol. 31. 8. Elsevier. 2013, pp. 1701–1708.

[43] Emőke-Ágnes Horvát et al. “A network-based method to assess the statistical significance of mild co-regulation effects”. In: *PloS ONE* 8.9 (2013), e73413.

[44] Aoife Ward et al. “Re-expression of microRNA-375 reverses both tamoxifen resistance and accompanying EMT-like properties in breast cancer”. In: *Oncogene* 32.9 (2013), pp. 1173–1182.

[45] Sarah Jurmeister et al. “MicroRNA-200c represses migration and invasion of breast cancer cells by targeting actin-regulatory proteins FHOD1 and PPM1F”. In: *Molecular and cellular biology* 32.3 (2012), pp. 633–651.

[46] I Keklikoglou et al. “MicroRNA-520/373 family functions as a tumor suppressor in estrogen receptor negative breast cancer by targeting NF-κB and TGF-β signaling pathways”. In: *Oncogene* 31.37 (2012), p. 4150.

[47] S Uhlmann et al. “Global miRNA regulation of a local protein network: Case study with the EGFR-driven cell cycle network in breast cancer”. In: *Molecular Systems Biology* 8 (2012), p. 570.

[48] Florian Haller et al. “Localization-and mutation-dependent microRNA (miRNA) expression signatures in gastrointestinal stromal tumours (GISTs), with a cluster of co-expressed miRNAs located at 14q32. 31”. In: *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* 220.1 (2010), pp. 71–86.

Preprints (not peer-reviewed)

2023 Courzet *et al.* G-PLIP: Knowledge graph neural network for structure-free protein-ligand affinity prediction

2023 Valeanu *et al.* Kinex infers causal kinases from phosphoproteomics data

2023 Rot *et al.* splicekit: a comprehensive toolkit for splicing analysis from short-read RNA-seq

2022 Geser and Zhang. Gene symbol recognition with GeneOCR

2021 Zhang *et al.* A Novel, Anatomy-Similar in Vitro Model of 3D Airway Epithelial for Anti-Coronavirus Drug Discovery

2019 Fang *et al.* Gene-set enrichment with regularized regression

Selected invited talks

2023 *Finding hope in a hopeless time - How Predictive Modeling and Data Analytics shifts our perspectives about antimicrobial discovery*, workshop on the promise of artificial intelligence to antibacterial drug discovery, 7th AMR Conference, Basel, Switzerland

2022 *Towards causal modelling of drug-induced toxicity for preclinical to clinical translation*, the Third In Silico Toxicology meeting, online

2021 *Optimization of the TeraTox assay for preclinical teratogenicity assessment*, co-presentation with Manuela Jaklin, OpenTox Virtual Conference, online

2019 *Bioinformatics and exploratory data analysis in drug discovery: an industrial perspective*, ISMB/ECCB, Basel, Switzerland

2018 *Mathematics in drug discovery: a practitioner's view*, Perlen-Kolloquium, University of Basel, Switzerland

Regular teaching activities

2020– Courses on causal inference and computational toxicology for the PSI (Statisticians in the Pharmaceutical Industry) community, once a year

2019– Lecture series *Mathematical and Computational Biology in Drug Discovery* at the University of Basel, 2 hours per week in the spring semester

2018– Lecture series *Introduction to Applied Mathematics and Informatics In Drug Discovery* at the University of Basel, 2 hours per week in the fall semester

2018– Lecture series *From Novel Targets To Novel Therapeutic Modalities*, Master Programme Drug Sciences, University of Basel, one lecture per semester

2012– Leading company-internal courses, seminars and workshops about computational biology in drug discovery, on average 2 hours per week

Engagement in academic and vocational-training activities

2023– Invited member of selection committees for tenure track positions in universities

2022– Co-organizer of Roche PMDA (Predictive Modeling and Data Analytics) Summer Schools for PhD students, a one-week event taking place annually

2021–	Certified vocational trainer (German: *Lehrmeister*) and exam expert for vocational apprenticeship in computer sciences
2020–	Certified Software and Data Carpentry instructor, hosting company internal courses on programming and data analysis
2014–	Invited reviewer for research proposals by national funding agencies
2012–	Reviewer of scientific manuscripts for journals including *Bioinformatics*, *PLOS Comp. Biol.*, *NAR Genom. and Bioinform.*, *etc.*

Supervised postdocs

Dr. Davide Bassani	2023- (co-supervision with Dr. Neil Parrott and Dr. Nenad Manevski), computational approaches to preclinical pharmacokinetic property prediction
Dr. Milad Adibi	2020-2022 (co-supervision with Dr. Ekaterina Breous-Nystrom), multiscale modeling of drug-induced liver toxicity, currently senior computational biologist at University of Zurich
Dr. Simon Gutbier	2018-2020 (co-supervision with Dr. Christoph Patsch and Dr. Markus Britschgi), immune pathway characterization with tool-compound screening for Parkinson's Disease, currently principal scientist at Roche

Supervised master students and interns

Alexandra Valeanu	2023, internship working on inferring causal kinases with phosphoproteomics data
Simon Crouzet	2022, internship working on predicting protein-ligand interaction with graph neural networks, currently PhD student in Dal Peraro's group at EPFL
Anja Lieberherr	2022-2023 (co-supervision with Prof. Niko Beerenwinkel at ETH), master thesis on applying graph neural networks in drug discovery, currently junior technology consultant at BearingPoint, Zürich
Sarah Morillo	2021-2022, internship working on proteomics data and protein half life, currently PhD student in Erik van Nimwegen's group at Biozentrum, University of Basel
Andreea Ciuprina	2020 (co-supervision with Prof. Niko Beerenwinkel), internship and master thesis on computational inference of immune-cell contribution to drug-induced liver toxicity, currently data engineer and scientist at Endress+Hauser Flowtec
Rudolf Biczok	2018-2019 (co-supervision with Prof. Alexandros Stamatakis, KIT), internship and master thesis on building a gene-expression database and evaluating gene-set comparison metrics, currently full-stack Java architect and Azure cloud specialist at Bank for International Settlements (BIS)
Moaraj Hasan	2017-2018, internship working on automating gene-expression data analysis
Tao Fang	2017-2018 (co-supervision with Prof. Mark Robinson, UZH), internship and master thesis on histopathology prediction with deep neural networks, currently postdoctoral fellow in von Mering group at University of Zurich
Gregor Sturm	2016-2017, internship working on gene signatures and the BioQC software, currently clinical bioinformatics scientist, Boehringer Ingelheim

Trained vocational apprentices in computer sciences

Giulia Ferraina	2023-2024, currently co-developing a software system for drug discovery research
Jannick Lippuner	2023-2024, currently co-developing a software system for drug discovery research
Yannik Benndorf	2022-2023, currently bachelor of computer sciences in TU München
Paul Geser	2022-2023, currently local IT manager volunteer at MAF Madagascar
Leonardo Seminatore	2021-2022, currently bachelor of science in geomatics, FHNW

Experience in establishing research group and leading teams

2022–	I co-founded and lead a new research group with six permanent positions, two postdocs, and interns. The team consists of computational scientists with training or working experience in toxicology, chemistry, biology, and statistics. We perform applied research on computational safety and toxicology to inform decisions in drug discovery and development. I led the interviews and the on-boarding of new employees. I am responsible of integrating the team in the working environment and in the process of drug discovery and development. As the same time, I am accountable for defining the team's vision, strategy, and scientific focus.
2022–2023	I led a cross-functional team of about 30 people to improve the productivity and efficiency with regard to data, insight, and knowledge management. Stakeholders and team members gave positive feedback to the organization and the product of the team. The outcome has led to changes in how data and computational models are captured and used.