




# CHUSHU TAN

Computational Genomics Researcher | Multi-Omics Bioinformatician | Machine Learning Specialist

 chushuta@gmail.com |  +86-18008428969 |  Guangzhou, Guangdong, China

---

## RESEARCH PROFILE & COMPUTATIONAL EXPERTISE

Innovative computational biologist and data scientist with 4+ years of professional experience in high-throughput genomics, multi-omics integration, and machine learning for biomedical applications. Strategic expertise bridges three complementary research domains:

**Cancer Immunology & Disease Mechanisms** – Multi-omics characterization of tumor microenvironments using transcriptomics, immunogenomics, and machine learning to predict immunotherapy response; current focus on EBV-driven immune polarization in nasopharyngeal carcinoma.

**Computational Genomics & Genetics** – Proficiency in GWAS, WGS, WES, variant calling, and population genetics; experience integrating large-scale genomic datasets for biomarker discovery and complex trait analysis.

**Computational Psychology & Cognitive Neuroscience** – Background in biopsychology and clinical psychology with growing expertise in applying computational/statistical methods to large-scale behavioral, neurological, and psychiatric genomic data; interest in evolutionary psychology and cognitive evolution.

**Data Science & Machine Learning** – Advanced proficiency in statistical genomics, survival analysis, and machine learning (XGBoost, deep learning, model interpretation via SHAP); skilled in managing big data pipelines and reproducible research workflows.

---

## EDUCATION

**China University of Political Science and Law** | Beijing, China Master of Applied Psychology (part-time) | Jan 2021 – Present Coursework: Clinical Psychology, Family Therapy, Developmental Psychology, Psychotherapy Methods

**Monash University** | Melbourne, Australia Master of Information Technology | Jul 2018 – Jul 2019 Coursework: Software Development, Machine Learning, Database Systems, Network Systems, Distributed Computing

**University at Buffalo – SUNY** | Buffalo, USA Bachelor of Arts in Biological Sciences & Psychology | Jan 2011 – Jun 2017 Coursework: Evolutionary Genetics, Molecular Phylogeny,

Biopsychology, Neurobiology, Genetics, BiochemistryIndependent Research: Directed study in computational genomics and evolutionary analysis

---

## RESEARCH EXPERIENCE & PUBLICATIONS

First-Author (Submitted): "Viral Load Determines Immune Phenotype and Immunotherapy Response in Nasopharyngeal Carcinoma" – Multi-omics study (22 NPC + 4 cohorts) demonstrating EBV viral burden creates distinct immune phenotypes predictive of checkpoint inhibitor response; achieved AUC=1.0 classification using machine learning and identified actionable biomarkers (CD247, CXCL1). *Methods*: scRNA-seq, immune deconvolution, survival prediction (XGBoost, SHAP).

Co-Author (Published): "From Viral Infection to Malignancy: The Dual Threat of Viral Exposure and Environmental Factors in Cancer Development" – Meta-analysis integrating viral and environmental cancer risk factors with heterogeneity/bias assessment across cancer types.

---

## PROFESSIONAL EXPERIENCE

### FREELANCE BIOINFORMATICS RESEARCHER

Remote | Oct 2023 – Present

Advanced Immunogenomics & Single-Cell Analysis: - Performed cutting-edge scRNA-seq analysis (Scanpy, Palantir) mapping immune cell states, trajectories, and functional/exhaustion phenotypes across complex disease datasets - Conducted comprehensive immune profiling using immune deconvolution algorithms (CIBERSORT, xCell, quantiseq) to characterize tumor microenvironment composition - Generated publication-quality visualizations and statistical summaries for peer review

Viral-Host Interaction & Immunotherapy Response: - Designed and executed viral load analysis protocols and virus-host immune interaction studies in cancer contexts - Integrated viral genomic data with host immune transcriptomics and clinical outcomes - Validated computational predictions with literature-based and mechanistic validation approaches

Survival Prediction & Translational Modeling: - Built machine learning survival prediction models using multiple algorithms (XGBoost, Elastic Net, Cox Proportional Hazards) - Implemented model interpretation frameworks (SHAP, permutation importance) for clinical actionability - Developed frameworks translating computational predictions to clinical decision-making

Systematic Data Integration & Meta-Analysis: - Conducted comprehensive meta-analyses of therapeutic outcomes across multiple published studies - Assessed heterogeneity (Q-test,  $I^2$  statistics), publication bias (Egger's test), and sensitivity analyses - Integrated conflicting results into cohesive biological narratives

Deep Learning for Genomic Data: - Self-directed study of neural network architectures (CNNs for image data, RNNs for sequential data) - Implemented transfer learning approaches for genomic data classification - Explored multi-omics pattern recognition using deep learning

## **CLOUDNA BIOTECH – BIOINFORMATICS ENGINEER**

Changsha, China | Nov 2021 – Oct 2023

Genomic Pipeline Development & Clinical Application: - Led development of comprehensive somatic mutation and WES/WGS analysis pipelines using Common Workflow Language (CWL) - Integrated industry-standard tools: GATK variant calling, SAMtools, VEP annotation for clinical-grade genomic interpretation - Implemented quality control frameworks ensuring pipeline reproducibility and clinical validity - Optimized pipeline performance; reduced runtimes by 40% through algorithmic and infrastructure optimization

Machine Learning Platform Architecture: - Architected and led development of scalable ML platform incorporating multiple algorithms (XGBoost, SVM, Lasso/Ridge regression, Random Forest) - Designed user-friendly interface for non-computational researchers enabling biomarker discovery - Deployed models for treatment response prediction and patient stratification - Validated ML predictions against clinical outcomes

Bioinformatics Database & Knowledge Infrastructure: - Designed and maintained custom antibiotic resistance database (DbABC) integrating RGI and literature data - Participated in drug-disease knowledge graph research; optimized database queries and performance - Managed 50,000+ compound-target-disease relationships

## **SOUTHERN MEDICAL UNIVERSITY – RESEARCH VOLUNTEER**

Guangzhou, China | 2024 | Dr. Xiaoming Lyu's Lab

Cancer Immunology & Viral Research: - Participated in RNA-seq analysis for cancer immunology and virology research projects - Contributed to clinical bioinformatics pipeline development - Performed quality control and data preprocessing workflows - Collaborated with clinical teams on biomarker validation

## **HSBC – DATA ANALYST (CONTRACT)**

Guangzhou, China | Mar 2024 – Present

Enterprise Big Data Pipeline Optimization: - Optimized large-scale financial data pipelines using Apache Spark, Hive, SQL for 100+ GB daily data volumes - Designed and deployed Airflow orchestration workflows for cloud integration (Alibaba Cloud infrastructure) - Implemented data quality monitoring and automated error handling

Automated Data Collection & API Integration: - Designed and deployed automated web crawling systems for market data collection - Implemented API pipelines integrating internal AI/NLP models for text analysis - Reduced manual data processing time by 60%

## **ADDITIONAL PROFESSIONAL EXPERIENCE**

**Hunan Xiansai Biotechnology Co. – Information Engineer** | Jul 2021 – Oct 2021 - Applied ML models (SVM, Random Forest) for biomarker identification in Alzheimer's Disease research - Constructed MySQL-based patient databases and healthcare information systems

**San Ji Biotech – Bioinformatics Engineer** | Jun 2020 – Nov 2020 - Developed NGS analysis pipeline for Thalassemia variant detection project - Performed BCL file processing, sequence QC, alignment, and variant annotation

**Freelance Python Developer** | Dec 2020 – Feb 2021 - Developed plugin extensions and feature additions for Anki flashcard system

---

## TECHNICAL SKILLS

### PROGRAMMING & DATA SCIENCE

**Core:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), R (ggplot2, dplyr, Bioconductor), SQL, Bash/Shell, Git, Docker, Jupyter

**Machine Learning & AI:** PyTorch, TensorFlow, XGBoost, LightGBM, Random Forest, SVM, CNNs, RNNs, LSTMs, Transformers; Hugging Face, BERT, GPT models; NLP, prompt engineering, LLM fine-tuning; SHAP, LIME interpretation

**Statistics & Modeling:** Linear/Logistic regression, Cox survival analysis, t-tests, ANOVA, Bayesian methods, power analysis, effect size estimation

**Big Data:** Apache Spark, Hive, Presto, Airflow; Alibaba Cloud, AWS; MySQL, PostgreSQL, NoSQL, knowledge graphs

### COMPUTATIONAL GENOMICS & BIOINFORMATICS

**Multi-Omics:** Scanpy, Seurat, Palantir (single-cell); DESeq2, edgeR, GSEA (transcriptomics); GATK, SAMtools, VEP (genomics); minfi, ChAMP (epigenomics); MSstats, MSnbase (proteomics); CIBERSORT, xCell (immunogenomics); FastQC, STAR, BWA (NGS QC/alignment)

**Databases & Tools:** NCBI (GEO, SRA), Ensembl, UCSC Genome Browser, Galaxy, KNIME, Bioconductor (15+ packages)

### MOLECULAR BIOLOGY & LABORATORY TECHNIQUES

**Genomics:** qPCR/RT-qPCR, NGS, WES/WGS

**Protein Methods:** Western blot, ELISA, Flow cytometry, Immunofluorescence, Mass spectrometry (LC-MS/MS, DDA/DIA), Protein quantification (BCA, Bradford)

**Cell Work:** Cell culture (mammalian, immune cells), basic molecular cloning

**Quality Control:** Experimental design, controls, batch management, data validation

### COGNITIVE & CLINICAL PSYCHOLOGY

**Assessment:** DSM-5, clinical interviews, cognitive batteries, psychometric validation, behavioral analysis

**Domains:** Psychopathology, psychotherapy (CBT, family therapy), cognitive neuroscience, neuroplasticity, brain-behavior relationships

**Research Integration:** Large-scale phenotypic analysis, GWAS (cognitive/psychiatric traits), genomic-behavioral correlation

---

## RESEARCH INTERESTS & FUTURE DIRECTIONS

**Cancer Immunotherapy & Precision Oncology** – Integrating multi-omics to predict immunotherapy response; understanding tumor microenvironment heterogeneity; viral oncology

**Computational Genomics & Complex Traits** – Using machine learning to dissect genetic architecture of complex diseases; GWAS interpretation; polygenic risk scoring

**Evolutionary Genomics & Computational Biology** – Applying bioinformatics to understand evolutionary basis of human traits; population genetics; phylogenetic analysis

**Computational Psychology & Cognitive Neuroscience** – Bridging psychology, genomics, and machine learning to understand behavioral phenotypes; psychiatric genetics; cognitive evolution; computational modeling of cognition

**Methodological Focus** - Multi-omics integration and systems biology approaches - Machine learning for biomarker discovery and clinical prediction - Reproducible research and open-source tool development - Statistical rigor and biological validation of computational findings