

Siwen ZHANG

(31)0626049080

aristazhang521@outlook.com

Amsterdam, The Netherlands

[GitHub](#)

[ResearchGate](#)

[Personal website](#)

Data Scientist and Bioinformatics Engineer with 7+ years of experience in translational medical research and clinical bioinformatics. Specialized in incorporating machine learning and bioinformatics for data-driven insights in cancer diagnostics, cell therapy and pathogen surveillance. Proficient in Python, R, SQL, and visualization tools. Actively transitioning into broader AI-related healthcare data analytics and digital health roles.

Work and Research Experience

Middle Bioinformatics Engineer – R&D Dept **Precision Scientific (China)** **Sept 2021 – Sept 2024**

- Led end-to-end clinical and research bioinformatics workflows from study design to statistical reporting and visualization.
- Built MRD detection pipelines; undertook 4 gynecologic oncology studies; result presented at SGO 2024.
- Developed and supported cell & gene therapy projects: viral integration detection, AAV mutation detection, CRISPR off-target analysis.
- Developed ML classifiers on platelet RNA-seq to distinguish uterine sarcoma vs. leiomyoma (Scikit-learn).
- Performed low-pass WGS (~1×) WGD analysis; standardized definitions across 4 studies; contributed to national HRD reference standards.

Middle Bioinformatics Engineer – R&D Dept **GrandOmics Biosciences (China)** **Apr 2019 – May 2021**

- Developed molecular assays and analytical workflows for clinical sequencing applications; built user-friendly analysis systems and performance evaluation pipelines.
- Led development of China's first targeted capture product for third-generation sequencing, improving SV/SNV/InDel detection accuracy over gold-standard methods.
- Built Galaxy-based pathogen surveillance platform for COVID-19 (mutation tracking, lineage analysis, visualization)
- Designed targeted amplification panels for 30 respiratory pathogens and implemented end-to-end identification and reporting pipelines, collocation with Galaxy or Linux platform.
- Benchmarked basecalling accuracy across Nanopore (R9/R10), PacBio, and Illumina platforms.

Bioinformatics Engineer – R&D Dept **Geneis Beijing (China)** **Feb 2017 – Apr 2019**

- Developed a targeted amplification panel for lung cancer biomarkers; established baseline CNV thresholds using MAD and visualized MSI profiles.
- Conducted transcriptomic and methylation analysis of mouse oocytes (in vitro vs. in vivo, GV and MII stages); identified key transcription factors and maturation pathways.
- Conducted retrospective gastric cancer genomics profiling (SNV/InDel/CNV/MSI/TMB) and evaluated clinical associations and survival using Kaplan–Meier analysis.

Education

MSc Data Science (Computer Science)	Leiden University, The Netherlands	Sept 2025 - June 2027
MSc Bioinformatics	The University of Edinburgh, UK	Sept 2015 - Aug 2016
BEng Biomedical Engineering	Northeastern University, China	Sept 2011 - June 2015

Technical Skills

Data Science & Programming:

- Languages: Python, R, Shell scripting, C (basic), Java (basic)
- Machine learning: Sci-kit Learn, TensorFlow, Keras, HuggingFace
- Databases: SQL (MySQL)
- Visualization: matplotlib, seaborn, ggplot2, plotly, folium, Tableau (basic)
- Others: Git, Docker, WDL/Cromwell, Galaxy

Bioinformatics & Genomics:

- Omics: genome (WGS, panel), transcriptome (RNA-seq) and methylome (bisulphite-seq).
- Sequencing Platforms: NGS and TGS (Illumina, MGI, Nanopore, PacBio)

Certifications: [IBM Data Science Professional Certificate \(Coursera\)](#)

Publications and Patents

Publications:

1. Shu T, Liang Y, **Zhang S**, et al. The prognostic value of tumor-informed minimal residual disease detection using circulating tumor DNA in first-line treatment of ovarian cancer. *Gynecol Oncol*. Published online November 22, 2024. doi:10.1016/j.ygyno.2024.11.002
2. Geng C[†], Tong Y[†], **Zhang S**, et al. Sequence and Structure Characteristics of 22 Deletion Breakpoints in Intron 44 of the DMD Gene Based on Long-Read Sequencing. *Front Genet*. 2021;12:638220. Published 2021 Apr 30. doi:10.3389/fgene.2021.638220
3. Yi Jiang, Lihua Wu, Shasha Huang, Pidong Li, Bo Gao, Yongyi Yuan, **Siwen Zhang**, Guoliang Yu, Yong Gao, Hao Wu, Pu Dai; Study of complex structural variations of X-linked deafness-2 based on single-molecule sequencing. *Biosci Rep* 28 May 2021; 41 (6): BSR20203740. doi: <https://doi.org/10.1042/BSR20203740>
4. Xie, Z., Sun, C., **Zhang, S.**, Liu, Y., Yu, M., Zheng, Y., Meng, L., Acharya, A., Cornejo-Sanchez, D. M., Wang, G., Zhang, W., Schrauwen, I., Leal, S. M., Wang, Z., & Yuan, Y. Long-read whole-genome sequencing for the genetic diagnosis of dystrophinopathies. *Annals of clinical and translational neurology* (2020), 7(10), 2041–2046. doi: 10.1002/acn3.51201
5. Zheng Jiang, Hui Liu, **Siwen Zhang**, Jia Liu, Weitao Wang, Guoliang Zang, Bo Meng, Huixin Lin, Jichuan Quan, Shuangmei Zou, Dawei Yuan*, Xishan Wang, Geng Tian and Jidong Lang, “A Novel Method for Microsatellite Instability Detection by Liquid Biopsy Based on Next-Generation Sequencing”, *Current Bioinformatics* (2020) 15: 1. doi: 10.2174/1574893615666200324133451
6. Xu, H.[†]; Wu, X.[†]; Sun, D.[†]; Li, S.[†]; **Zhang, S.[†]**; Teng, M.; Bu, J.; Zhang, X.; Meng, B.; Wang, W.; Tian, G.; Lin, H.; Yuan, D.; Lang, J.; Xu, S. SEGF: A Novel Method for Gene Fusion Detection from Single-End Next-Generation Sequencing Data. *Genes* 2018, 9, 331. doi: 10.3390/genes9070331

[†] These authors have contributed equally to this work and share first authorship.

Patents: CN108304693B (2022), CN108949757B (2021), CN112593015B (2023)