

个人自述

宋捷（2026 年 1 月版）

q2516581@126.com · GitHub

方向：面向临床科研团队的数据分析与科研支持——围绕人群队列/组学与临床影像数据开展可复现分析、风险预测建模与论文产出支撑。

一、基本信息

- 姓名：宋捷（1994 年 10 月生）
- 电子邮箱：q2516581@126.com
- GitHub：github.com/lovivi

二、教育经历

时间	学校	专业/方向	学历
2012.09 – 2016.06	西北农林科技大学	生物技术	本科
2016.09 – 2019.06	西北农林科技大学	生物信息学（旱区作物逆境生物学国家重点实验室，导师：马闯）	硕士

硕士期间主要成果：研究方向为机器学习/深度学习在 RNA 甲基化预测中的应用，主导发表 SCI 论文 2 篇（一作 1 篇，二作 1 篇），参与发表 3 篇。

三、工作经历

时间	单位	职位
2019.07 – 2020.08	深圳碳云智能科技有限公司（iCarbonX）	数据挖掘工程师
2020.09 – 2021.12	中山大学附属第七医院大数据中心	实习研究员
2022.01 – 2024.12	深圳寻竹生物科技有限公司（Ailurus）	生信及 IT 主管

四、科研工作及主要成果

4.1 学术论文

#	论文题目	期刊	作者贡献
1	SynVectorDB: embedding-based retrieval system for synthetic biology parts	Database	通讯作者
2	Transcriptome-Wide Annotation of m5C RNA Modifications Using Machine Learning	Frontiers in Plant Science	第一作者
3	Predicting early-onset COPD risk in adults aged 20-50 using electronic health records and machine learning	PeerJ	通讯作者
4	PEA: an integrated R toolkit for plant epitranscriptome analysis	Bioinformatics	第二作者
5	deepEA: a containerized web server for interactive analysis of epitranscriptome sequencing data	Plant Physiology	第二作者
6	Evolution of the RNA N6-methyladenosine methylome mediated by genomic duplication	Plant Physiology	第四作者

#	论文题目	期刊	作者贡献
7	A Deep Convolutional Neural Network Approach for Predicting Phenotypes from Genotypes	Planta	第三作者
8	miRLocator: A python package and web server for predicting miRNAs from pre-miRNA sequences	Methods in Molecular Biology	第五作者

4.2 主要科研项目

- 1. UK Biobank 大规模人群队列数据挖掘** - 针对 50 万人全基因组及 2000+ 表型数据，开发自动化 GWAS 分析流程（数据清洗、关联分析、风险建模）
- 基于该数据完成早发型 COPD 风险预测模型构建并发表论文；完成 AD/MCI 关联分析及深圳市自然科学基金本子撰写。
- 创新性地结合大语言模型开发了自动化分析报告生成工具。
- 2. 合成生物学工具与数据库开发** - 开发模块化通路设计工具，基于 EcoCyc 构建 673 个可供实验的模块化通路簇。
- 开发载体设计平台配套模块：质粒拆分优化、菌落识别等，并构建本地元件库。
- 3. 免疫芯片数据挖掘与短肽疫苗设计** - 从零开发 ImmunoSignature 免疫芯片预处理及分析管线。
- 新冠疫情初期预测抗体响应区域，筛选的候选肽段。
- 4. 肿瘤靶向用药自动化报告系统** - 构建从测序下机数据到临床报告的全自动化管线，匹配知识库生成肿瘤分型及用药建议，已应用于临床。

本页由 2026 年 1 月版个人自述重构生成；联系方式以邮箱为准。