

宋捷

您好,我是宋捷。生信及衍生IT相关五年工作经验,期间我在创业团队就职时间较长,技能树相对全面,有管理和部门间协调以及完整的产品开发经验。有多个项目主导和参与经验,发表论文6篇。目前主要感兴趣的领域为**生物信息,合成生物学,医疗/生物数据建模与挖掘**,以及相关的产品开发,项目管理,统筹等相关工作。

如果您对我的背景和技能感兴趣,请随时与我联系。



教育经历

- 2012 | 2016 ● **西北农林科技大学 (本科)**
生命科学学院 生物技术 陕西
- 2016 | 2019 ● **西北农林科技大学 (硕士)**
旱区作物逆境生物学国家重点实验室(马闯实验室) 生物信息学 陕西
 - 机器/深度学习在RNA甲基化预测上的应用
 - 主导发表SCI论文2篇 (1作1篇, 2作1篇)
 - 生信分析管线构建及可视化
 - 参与发表SCI论文3篇

工作经历及职能

- 2019 | 2020 ● **数据挖掘工程师**
深圳碳云智能科技有限公司 (iCarbonX) 深圳
 - 开发基于深度学习对表达谱的降噪算法
 - IST免疫芯片数据的处理流程开发,基于此技术的短肽.疫苗的设计工作
- 2020 | 2021 ● **实习研究员**
大数据中心 深圳市中山大学第七附属医院 深圳
 - 生信相关: UKB数据的专项处理
 - IT和运维: 大数据中心协助指定管理细则,用户手册,并协助管理。
- 2022 | 2023 ● **生信及IT主管**
深圳寻竹生物科技有限公司 (Ailurus Ltd.) 深圳
 - 负责处理NGS和芯片数据,上下游分析和可视化工作,期间撰写SCI论文1篇(CA),专利1份。
 - 负责相关数据库的管理和维护,开发了自有元件库和可移植通路库。
 - 领导网站开发,负责客户服务部署,实现产品报告的自动化,以及内部IT工具的开发与维护。
 - 蛋白结构模型和IIm模型的测试及部署, finetune, 产品设计等,确保团队内部的模型使用。
 - 承担proposal设计和任务分配的责任,负责硬件资源的调配,确保团队高效运作。

主要项目经验

CONTACT INFO

✉ q2516581@126.com
🌐 lovisongjie.top
☎ songjie9075
📅 1994.10.24

SKILLS

项目技能: 常规/特殊生信流程构建,报告自动化处理,大模型应用开发,常规网站/网页开发。

运维及管理: 生信团队的管理,软硬件以及云服务的运行维护,资源调配等。

技术栈: 分析处理的是 R,python,bash,Snakemake,gala等, web和可视化使用R shiny latex flask bootstrap等, 数据库使用mongodb及sqlite。

常用数据库: 常用的核酸库 ncbi-refseq等, 病案库ukbiobank, TCGA等, 元件库snapgene, IGEN-part, addgene, VectorBuilder等, 通路库biocyc等, 文献库openalex,crossref。

Last updated on 2024-02-03 by R.

2023

● 模块化通路设计

构建了一套工具，可将关注通路所涉及的多个基因簇合并为一个模块化通路簇，用于底盘工具物种的移植。最终在基于ecocyc构建了673个可供实验的模块化通路簇。

· 本项目难点：1.尽可能使用常用菌株，调用基因簇，保留调控序列。2.整合基因簇需要算法优化。3.多个大型数据库间需要关连，同时设计多处比对，需要考虑效率问题。4.需要同时完成oligo拼接方案设计。

📍 深圳

2023

● 载体设计平台配套工具开发及初始元件库构建

配套设计平台开发了多个生信模块，质粒注释模块，质粒转义模块（基于客户祖传质粒推荐最接近的本地质粒和同义化路线），高通量比对模块，序列拆分oligo及接头优化，平板菌落坐标识别(CV)，分子对接模块配套(docklite,af2)，密码子优化模块，危险序列识别（病原）。

同时基于多数据库元件信息，构建了本地初版元件库。对市面多家元件库数据较为熟悉。

📍 深圳

2023

● 超高通量载体测试平台配套

RD设计了一套超高通量载体测试的实验方案，需配套统计和分析工具。接手后发现了早期代码的设计错误，重新基于实验设计写了一套完整的分析和报告系统，完善了该产品，工具独立申请了软著。

· 本项目难点：1.实验设计较创新，无可比对方案，需自行编写一套分析工具。2.需要基于结果迭代实验方案。

📍 深圳 Ailurus

2022

● 癌症靶向用药自动化报告项目

构建了一个自动化分析用户肿瘤下机测序数据，提取患者突变信息并匹配知识库来分析肿瘤分型和靶向用药信息，生成交付报告。

· 该项目难点：1.单份模板生成多份样式，存在大量latex表格；2.使用部门无计算机背景，且审核后需直接交付用户；3.数据库需持续维护

📍 医院项目

2021-2022

● ukbiobank数据的处理及组学分析管线

ukbiobank 数据（50w 人+80w 遗传位点+2000表型），结合galaxy系统开发并优化了自动化分析流程，包含数据提取，预处理，清洗，分组，关联分析，建模分析等。构建了一套ukb数据的表型-基因型关联工具。对AD和MCI进行了关联分析，构建了风险预测模型和评价指标，结合院内数据写了深自然本子（未中）。对早期型COPD进行关联分析，并撰写论文（已接收，peerj，CA）。

· 此项目期间通过对结果的结构化处理，结合gpt3.5-api和promot工程，开发了一套AI Agent，提供初步的文字稿，并自动校订成可供快速处理的latex论文初稿。同时编写了大量相关技术及数据说明文档。

📍 医院项目

2020

● 免疫芯片数据的挖掘和疫苗开发

ImmunoSignature 免疫芯片数据,小组内对该引进技术进行了从零开始的预处理管线和分析管线工作的开发工作。本人主要负责关联医院的临床数据和实验的下机数据，清洗，质控，和矫正工作。在新冠疫情初期，个人负责了探索抗体对病毒蛋白响应区域的预测工作，在小鼠实验阶段，筛选出的肽段能有效产生中和抗体。

📍 深圳 iCarbonX

 管理经验

2022
|
2023

● 生信及IT部门管理

Ailurus任职期间，管理生信及IT部门。除个人项目外，同时负责任务规划及分配，系统资源调配，技能培训，与其他部门和业务公司沟通等职能。

📍 深圳 Ailurus

2021

● 数据中心的组建及管理

21年6月受朋友委托完成区块链chia项目调研并运行(总投资400w)。快速组建创业团队并在15天内完成全项目调研、硬件测试、采购、项目选址、电路改造、管理软件编撰等，快速部署了两处数据中心。该项目由于政策风险快速终止。

📍 深圳 洛未信息



学术文章

● Transcriptome-Wide Annotation of m5C RNA Modifications Using Machine Learning.

Frontiers in Plant Science. Song J(1st)

● Predicting early-onset COPD risk in adults aged 20-50 using electronic health records and machine learning

PeerJ. Song J(CA).

● PEA: an integrated R toolkit for plant epitranscriptome analysis

Bioinformatics. Song J(2nd).

● deepEA: a containerized web server for interactive analysis of epitranscriptome sequencing data

Plant Physiology. Song J(2nd).

● Evolution of the RNA N6-methyladenosine methylome mediated by genomic duplication

Plant Physiology. Song J(4th)

● A Deep Convolutional Neural Network Approach for Predicting Phenotypes from Genotypes

Planta. Song J(3rd).

● miRLocator: A python package and web server for predicting miRNAs from pre-miRNA sequences

Methods in Molecular Biology. Song J(4th).