CygnusX BioOps

面向生命科学的受控 Agent 与可信计算交付基础设施

让科研问题转化为可验证、可复现的计算实验

汇报人：张建

26.10.31



科学研究是千百次迭代的长链路，而通用 Agent 专为短对话设计

单步 95% 的准确率，经 50 步黑盒传递后成功率不足 8%，链条越长，通用模型的幻觉累积越致命



科学研究 • 严谨长链路

数据发现 → 质控 → 比对 → 定量 → 标准化

差异分析 → 富集 → 注释 → 可视化 → 机制解读

高频参数调优 · 失败断点重试 · 专家介入判断

千百步跨周期迭代，任一环节出错即全盘重来



通用 Agent • 即时短链路

用户提问 → 模型临时生成代码/回答 → 结束

擅长：**单点问答、单脚本生成、文本总结、单步推理**

短板：**缺长程规划、环境报错难自愈、过程无法审计**

短板：缺长程规划、环境报错难自愈、过程无法审计

指望通用大模型一步到位做科学，现阶段效果必然差

三个结构性原因：为什么通用模型做不好科学

缺乏真实轨迹数据、领域知识外挂昂贵、长链路容错率极低，构成通用 AI 难以跨越的工程鸿沟

01 训练数据受限

- 高质量科学计算轨迹极其稀缺
- 模型没见过几百步成功案例
- 不知道何时该停、何时重试、为什么这样选

没见过完整科研闭环

02 上下文税

- 每次调用都要重新教领域知识
- QTL / GWAS / 实验室惯例全部外挂
- token 成本高，窗口被背景占满，留给任务的很少

窗口被背景知识占满

03 失败成本高

- 每一步错误沿链路放大
- 算力、时间、样本成本远高于对话场景
- 错误常在运行后期才暴露

一步算错即全盘作废

结论：指望通用大模型一步到位做科学，不现实，需要另一条路

一句话认识 CygnusX

面向生命科学的受控计算与可信交付基础设施，推动组学分析从“个人手艺”迈向“工业级可信交付”

AI 只做受控调度，绝不黑盒生成流程

让科学探索的灵活性，建立在工业级确定性底座之上

科研意图拆解 → 契约化能力路由 → 沙箱确定性计算 → 可复算证据交付

智能规划层（懂科研）

18位领域 Agent 拆解复杂科学问题，零代码门槛交互

受控执行层（算得准）

FlowFrame 标准管线 + 597项契约 Skill 无写权限隔离

证据资产层（可复现）

以项目为单位，自动封存 环境快照、参数版本、运行日志

18 领域 Agent	597 Analysis Skill	30 min 增量响应闭环	10 min 照单核验复现

把分

从数据发

数据/文

科研寻宝

跨 GEO /S

ce

验

CygnusX 科研生态

连接专业组件，构建可追溯科研链路



科研寻宝鼠

数据发现与文献探索

- 公共组学数据检索
- 文献来源与关联证据



Polariseq

测序数据获取与传输

- 下载 · 续传 · 完整性校验



CygnusX BioOps

天鹅座智能科研平台

项目 · Agent · 数据 · 计算 · 交付

统一项目上下文，组织受控科研任务



FlowFrame

流程规范与标准化接入

Flow 流程家族

- RNAFlow
- ATACFlow
- AnnoFlow
- BSRFlow
- GWASFlow
- DNAFlow



BioReport

分析报告与交付

- 结果 · 图表 · 方法说明

CygnusX 生态科研链路

数据发现 → 科学问题 → Agent 编排 → 计算实验 → 可验证证据

FROM STARDUST, TO DISCOVERY

四种使用入口，覆盖从轻量工具到多角色协同

阶梯式产品体验：不强迫所有科研人员一上来就使用 AI 对话，兼顾轻量日常刚需与深度智能探索

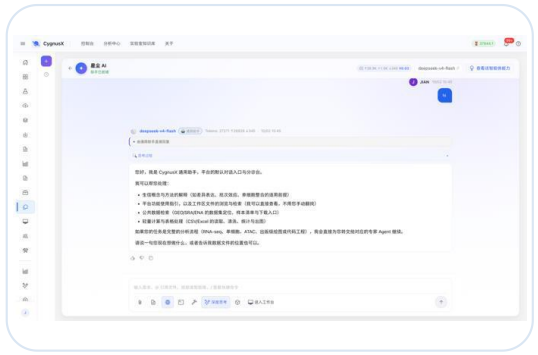


01 /工具箱

零 Token 自助分析

集成 FASTQ 质控、BLAST、火山图等 18 个常用生信工具

已全面上线

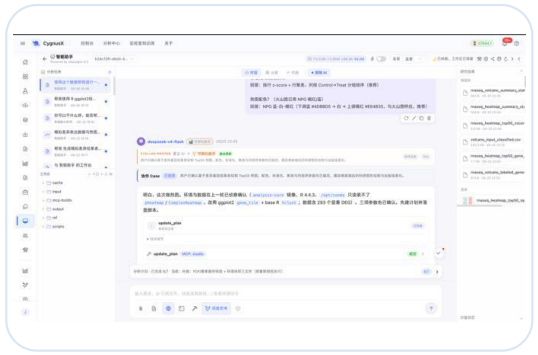


02 /AI 助手

自然语言任务入口

统一的对话交互中枢，支持 @ 文件与一键唤起高频增量任务

已全面上线

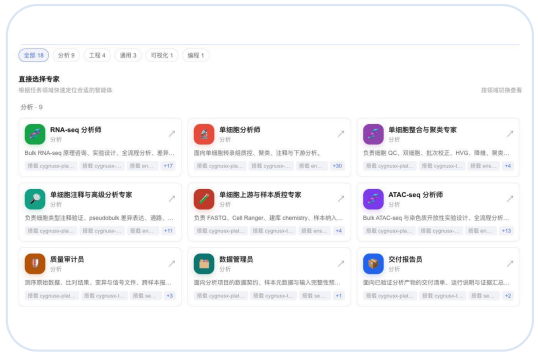


03 /AI 工作台

受控计划与运行

三栏受控界面，透明呈现沙箱挂载、代码写入与分步执行状态机

已全面上线



04 /多角色协同

专职 Agent 协作

探索性多 Agent 联动机制（含 Bridge、审批令牌与交付 Manifest）

实验性预览

真实案例：从一句研究问题，找到可用于分析的公开数据

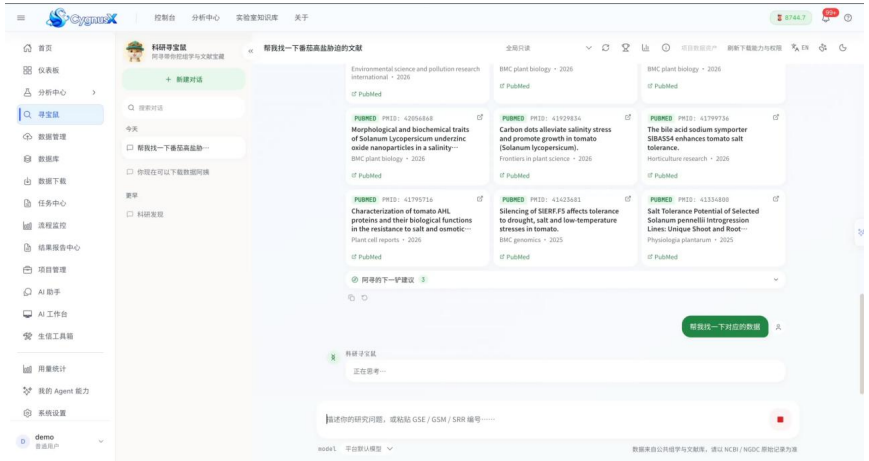
以公开转录组数据集 GSE148353 为例展示数据发现与下游分析路径

GSE148353

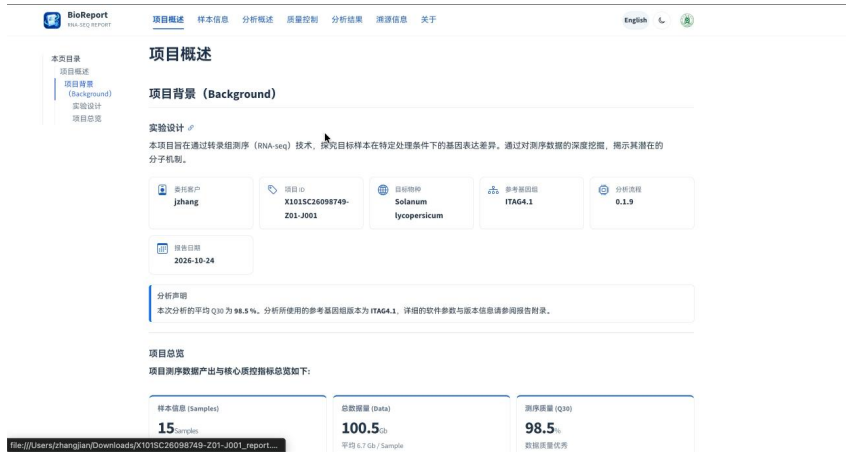
两份番茄材料盐胁迫对比转录组

18 个 RNA-seq 样本

2 个材料 × 3 个时间点 × 3 个重复，共 18 个样本



科研寻宝鼠数据/文献发现



可溯源报告

1 自然语言找数据 → 2 提取样本元数据 → 3 结构化实验设计 → 4 进入受控计算 → 5 归档报告与证据

真实案

以公开转录

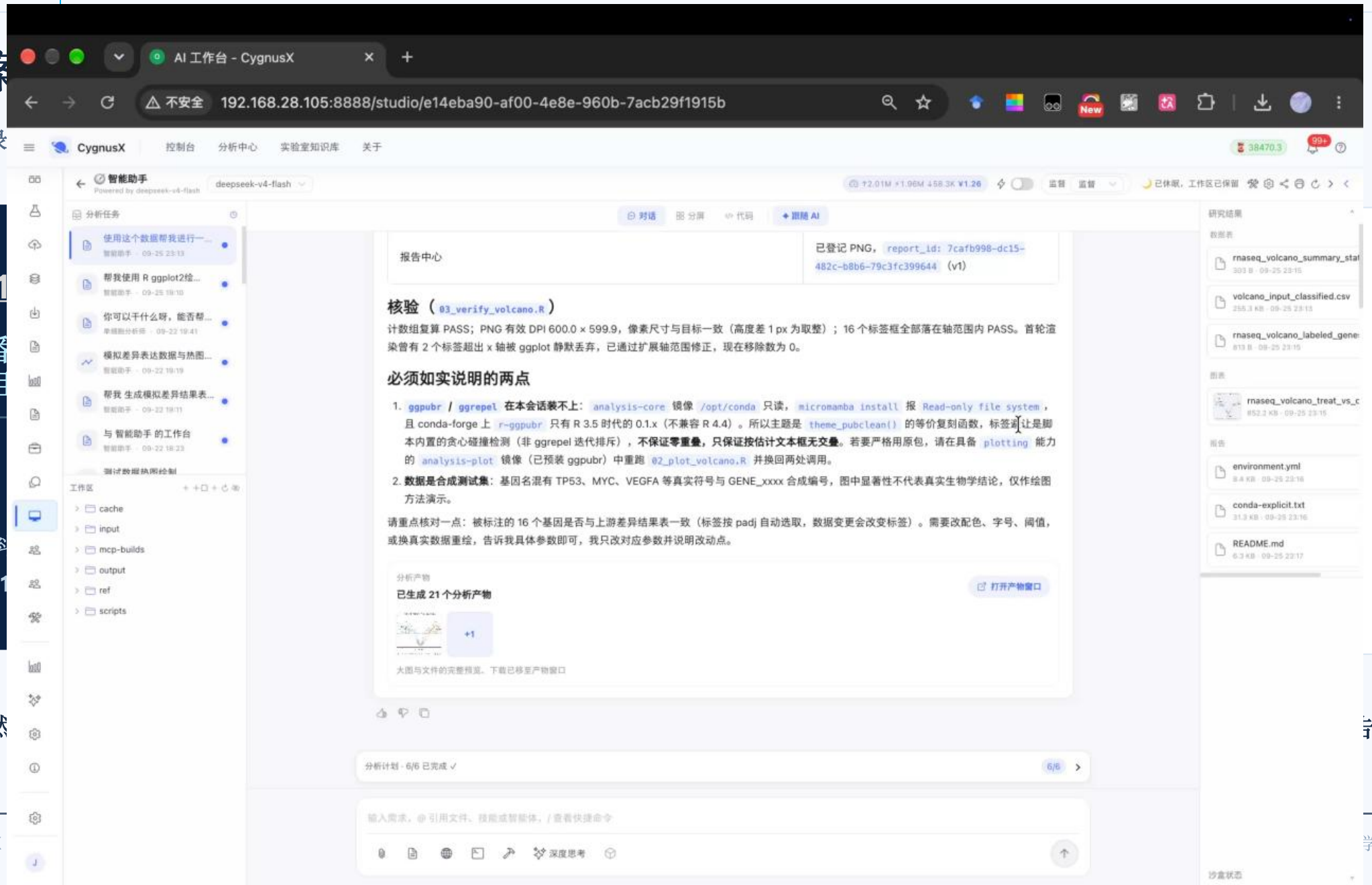
GSE1

两份番
转录组

18 个

2 个材料
复，共 1

1 自然



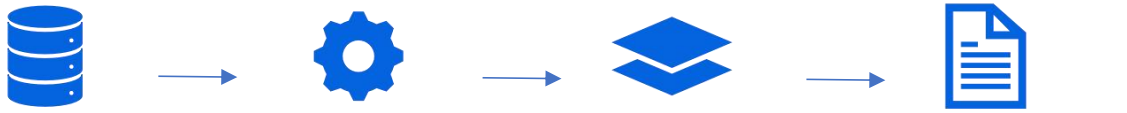
与证据

上游 标准分析

1 分析中心（确定性计算引擎）

数据获取→数据清洗→标准分析→生成标准化结果（矩阵+质量报告）

重资产·已成熟·版本化·可复现



数据获取	数据预处理	标准分析流程	标准化输出
下载与聚合样本解析	质控/去接头/比对	RNA-seq/ATAC-seq	表达矩阵/分析报告

成熟的工作流生态（示例）

snakemake nextflow

HISAT2 STAR Salmon DESeq2

标准化中间产物（Agent的输入）

表达矩阵+样本表+质量报告+元数据+版本与本源

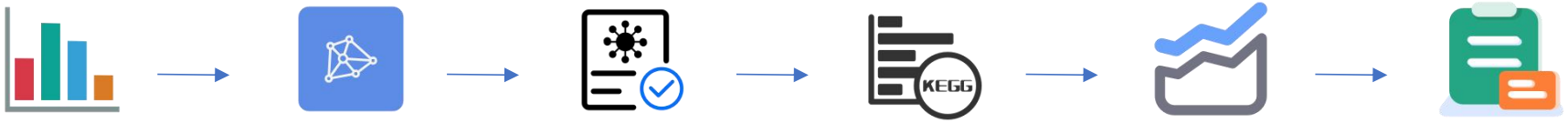
Matrix Metadata QC Report Provenance

下游 探索与高级分析

2 智能科研分析（Agent+专家共识）

基于标准化结果进行深入分析、可视化、注释与生物学解释

轻资产·探索性·灵活组合·解释与发现



差异分析	富集分析	细胞类型注释	机制与通路分析	高级可视化	文献证据与解读
DEG/DGE	GO KEGG GSEA	scRNA/空间组学	调控网络/关键基因	热图/UMAP/气泡图	九大文献源/证据链

计算结束不是终点：让每一次结果都有可追溯可复现依据

以项目与运行实例为单位记录过程，支持交接、核验与后续复算

项目上下文

Project /Session /Run

任务与环境

参数、软件版本、环境快照

计算过程

日志、任务状态、质量记录

产物清单

结果文件、**MD5**、报告说明

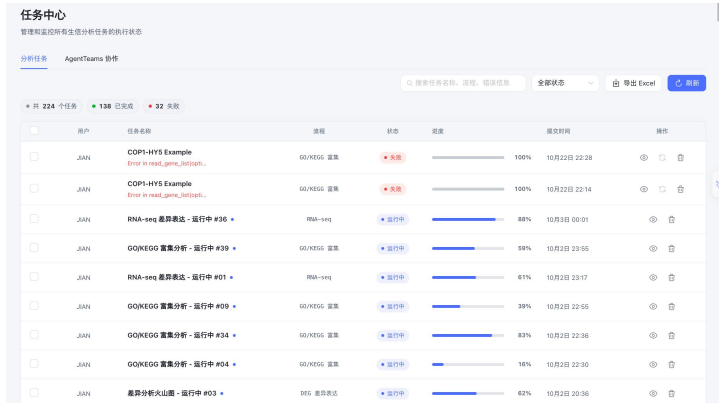
实际交付产物示意



可验证证据链：运行信息可查、分析产物可验、流程版本可溯。

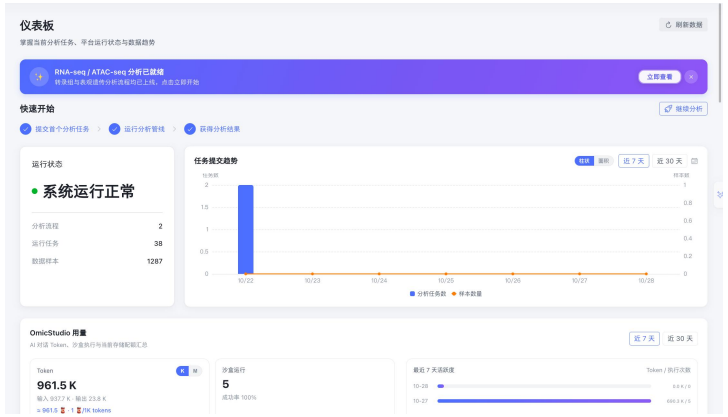
真实平台运行：任务提交、执行监控与结果管理可见

平台已在华中农业大学园艺林学学院部署使用



任务中心

执行状态、历史任务与操作记录



运行监控

资源使用、运行趋势及指标



报告与产物

分析结果与结构化输出

已实现：提交分析 → 查看运行记录 → 打开报告 /校验信息 → 基于Agent的定制分析

效率提升不仅来自 AI，还来自标准化与可复用的工程资产

以材料中的真实业务场景对比为依据，清楚呈现测量口径

标准分析提交	约 20 分钟	→ 约 5 分钟	约 75%	参数准备 /提交
增量需求方案	约 8 小时	→ ≤30 分钟	约 94%	Skill 覆盖范围
定制流程改造	约 5 人天	→ 约 1 人天	约 80%	模块替换场景
历史项目核验	传统方式以周计	→ 约 10 分钟	可核验	依赖证据归档条件

数据口径：为材料中的典型场景对比；主要反映人工准备/方案响应，不等于计算程序本身提速。

演进飞轮：依托华农 A+ 学科知识与真实轨迹，锻造园艺垂直领域模型

将平台沉淀的高质量计算轨迹与顶尖团队内部知识库内化进模型权重，从“外挂工具”进化为“真懂科研的内生智能体”

1 顶尖学科冷启动

部署华农园艺学院

覆盖约 2000 名硕博

预期 300+ 活跃用户

2 轨迹与知识库汇聚

互惠授权：换额度

安全红线：原始数据冻结

独家注入：内部知识库

3 专家审核与三层内化

筛选金子：专家审核

经验内化：写入权重

摆脱重税：参数运行时注入

4 垂直模型与飞轮

从能调到真懂,具备解读能力

模型越用越聪明

护城河正向循环



商业模式：从私有化部署切入，逐步积累维保与模板收入

市场测算与商业化目标来自商业计划书，均非已实现销售业绩

15 亿元
TAM · 国内可触达市场

2.4 亿元
SAM · 重点可服务市场

约 5000 万元
SOM · 五年累计销售目标

四种收入路径

个人按次分析

¥19–99/次起，阶梯定价（五年累计 3575 万）

课题组年度订阅

0.8–2.4 万元/年（五年累计 630 万）

高校/科研场景定制

3–10 万元/项（五年累计 270 万）

私有化 License + 维保

15–20 万/套 + 15–20%/年（五年累计 525 万）

核心竞争力来自生命科学实践与全栈工程能力的结合

懂科研需求，也能把复杂流程做成可部署、可维护的软件产品

创始人 /产品与技术负责人

张 建

华中农业大学博士

生命科学与组学分析研究背景

科研机构与生物技术企业工作经验

独立主导核心平台架构和全栈实现

首席 科学顾问

王 昕

华中农业大学园艺林学学院 教授 博士生导师

专家质量审核、科研需求定义、生信方法与流程审查

医学顾问

董 政

哈尔滨医科大学助理研究员

医学科研需求定义 · 生物医学场景适配 · 科研结果解读

CygnusX BioOps

让每一次科研计算，都成为可复现的证据

智能规划有边界 · 科学计算有标准 · 项目交付有证据

