



Notebooks Behind the Firewall: AI Agents and Real Patient Data in Jupyter

Introducing Biorouter, a secure AI-native biomedical research environment

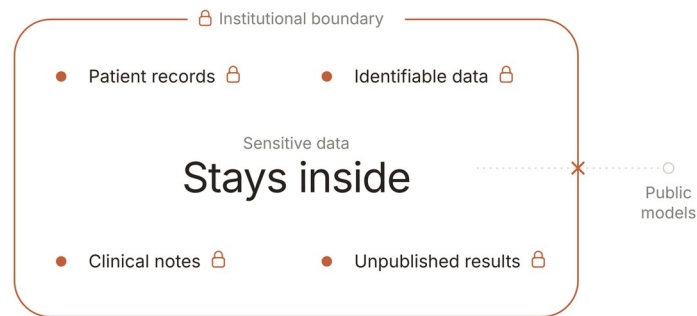
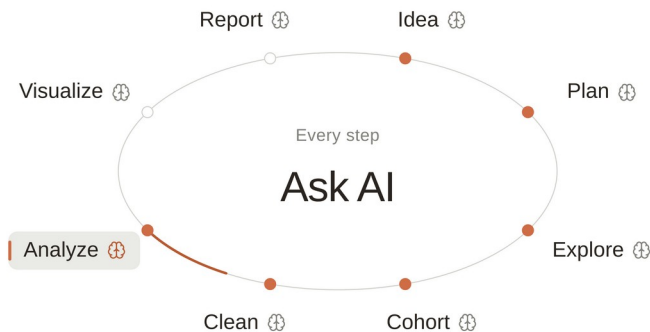
WanJun Gu

Baranzini Lab, UCSF

Jupyter Day 2026 · 10/19/2026

Two needs in biomedical research

An iterative workflow, and data that must stay protected.



01

Research is iterative

AI agents can help at each step, from idea to report.

02

Sensitive data must stay protected

Patient, identifiable and unpublished data has to remain inside a secure institutional environment.

What researchers need from AI tools

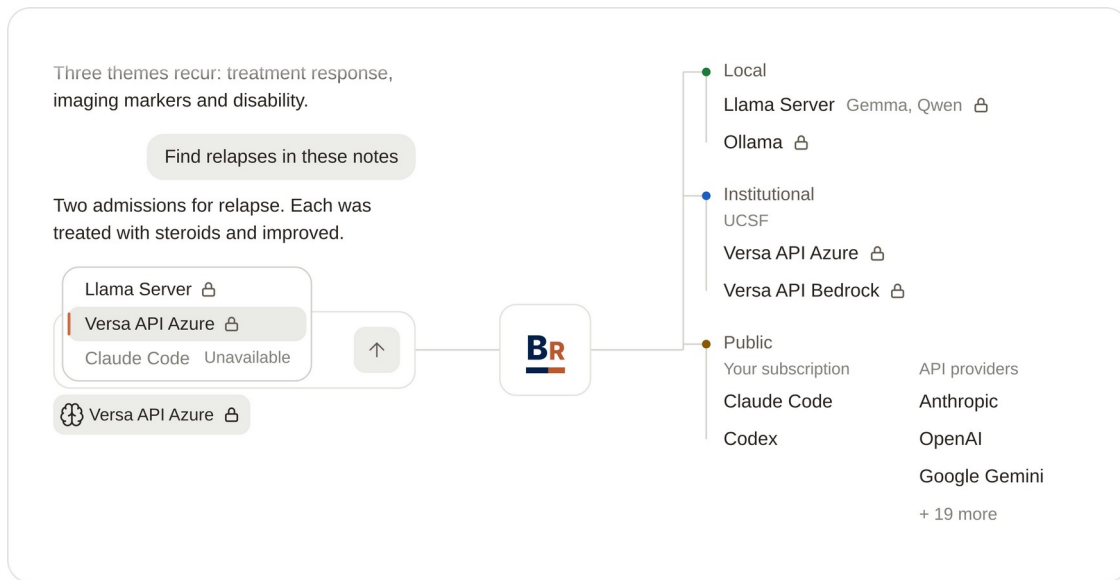
Existing options meet some of these needs, but not all.

	Agentic	Sensitive data	In-house tools
Commercial agents & APIs Claude Code, Codex and frontier model APIs	✓	✗	✓
Institutional chatbots Enterprise chat with institutional privacy	✗	✓	✗
Institutional LLM APIs Private models, but no harness: developers only	✗	✓	✗

Can one environment work across models,
and keep private work private?

UCSF Biorouter

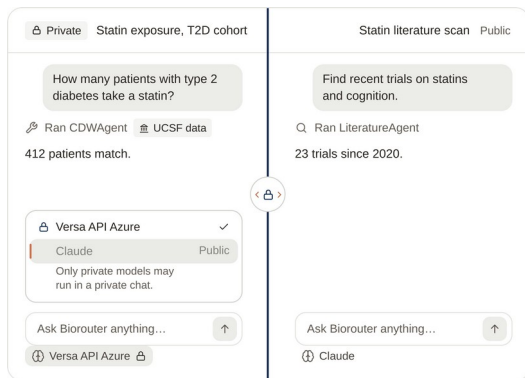
A single interface to local, institutional and public models.



- Agent harness
 - Workflows · Subagents
 - Scheduling · Local compute
- Extensions MCP
 - Health records · SPOKE · PubMed
 - 38 in the BAAM marketplace
- Skills
 - Replication · EHR · Model training
 - 129 in the BAAM marketplace

Three design principles

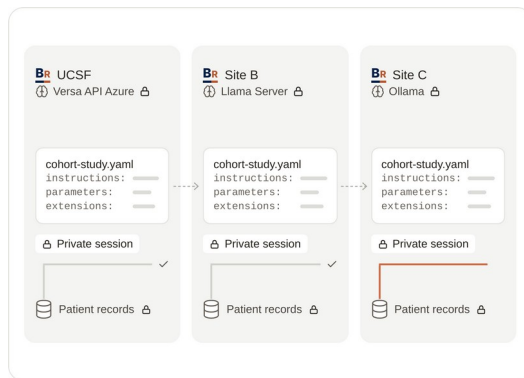
Privacy, reproducibility and domain connectors.



01

Private work stays private.

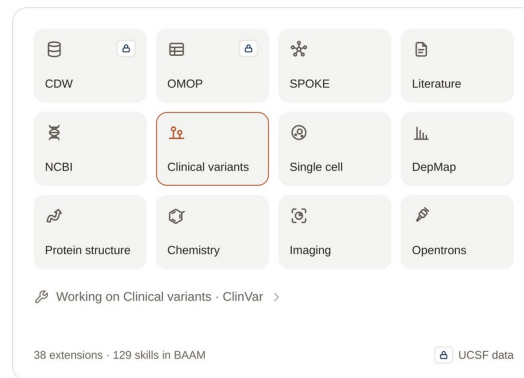
Models and connections determine whether a chat is private or public. Private chats are routed only to private models.



02

Share workflows, not data.

A workflow records models, extensions, prompts and skills, but not data. Each site reruns it on its own records.



03

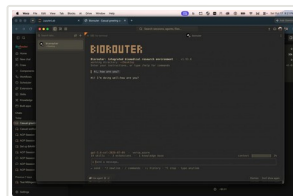
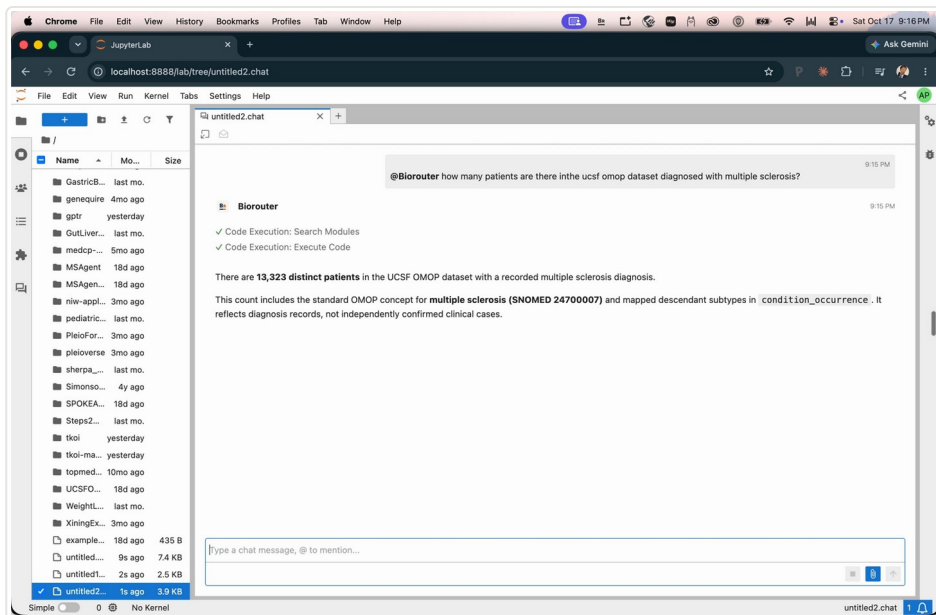
Biomedical connectors.

Connectors for EHR, SPOKE, literature, omics and lab systems, installed with a click or a sentence.

Four ways to use Biorouter

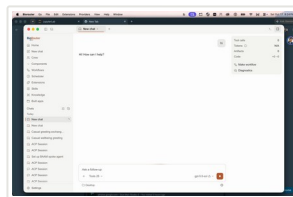
The same agents and privacy settings in each interface.

Jupyter Biorouter as an ACP agent in JupyterLab



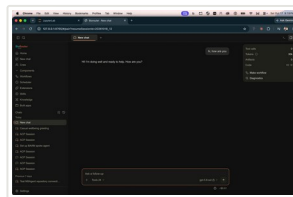
CLI

Any terminal, local or remote



Desktop app

Point-and-click GUI



Headless

Served from a terminal, opened in the browser

Benchmarking the agent harness

Similar on proprietary models, more accurate on Qwen 3.5.

3 agents

Codex · Claude Code · Biorouter

3 models

GPT-5.4 · Sonnet 4.6 · Qwen 3.5

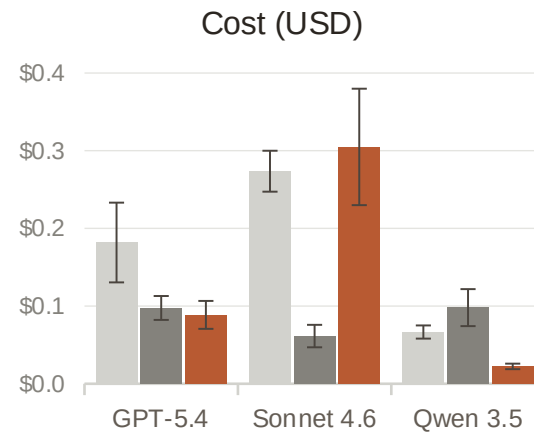
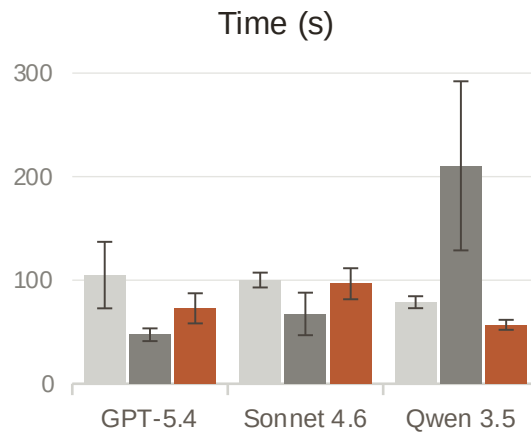
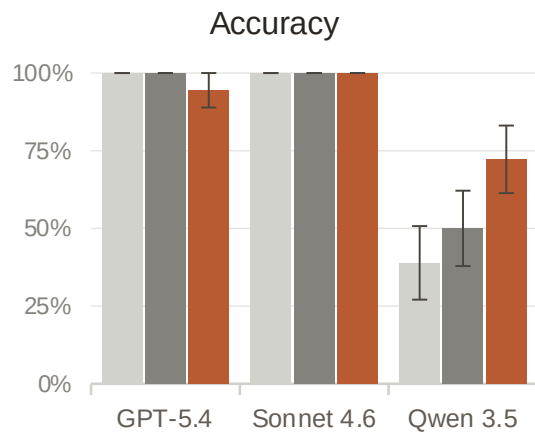
6 tasks

GWAS · scRNA-seq · SIR · PyMOL · iHS · PageRank

3 runs

each combination

■ Codex ■ Claude Code ■ Biorouter



Mean $\pm$ SE over 3 runs; all models via OpenRouter. Gu*, Bellucci* and Baranzini, under review at *Bioinformatics*.

Demo

Paper Replication

Our lab's published comparison of ocrelizumab and rituximab in MS, rerun on UCSF EHR and SPOKE.

6–8 months

Original study

< 30 minutes

Main results and figures

BioRouter

Home

New chat

Components

- Workflows
- Scheduler
- Extensions
- Skills
- Knowledge
- Built apps

RECENTS

See all

Today

- UCSF OMOP MS re...
- OMOP EHR query exam...
- SPOKE query session
- Modeling workflow specs
- Curate Nature paper dat...
- Set up SPOKE-FORGE
- Random regression plot
- Read Outlook email
- Mitochondria presentatio...

Yesterday

- Menopause neurodegen...
- Initial greeting 10

Verified:

- 9,129 bytes
- Built from the aggregate results and effect estimates
- Does **not** embed patient-level rows
- References the patient-level CSV metadata from 1,671 source rows

please also make some powerpoint slides summarizing the results

Ran Updating the work plan · 1 result ready >

Ran Replacing the task list · 1 result ready >

Ran Starting "Load PowerPoint handling context and inspect a..." · 1 result ready >

Ran Load PowerPoint handling context → Check local Python ... · 1 result ready >

Ran Marking "Load PowerPoint handling context and inspect a..." · 1 result ready >

Ran Starting "Generate a PowerPoint summary deck from the r..." · 1 result ready >

Ran Write a Python script to build the PowerPoint... → Run the... · 1 result ready >

Ran Marking "Generate a PowerPoint summary deck from the r..." · 1 result ready >

Working on the result · 2s

Desktop

⌘1/⌘4 | to navigate messages

HTML /Users/wanjun/Desktop/medcp_replication_ocrelizu... Copy Preview Raw

Rituximab vs ocrelizumab. Among patients with follow-up IgG.

Replication endpoints

Endpoint	Ocrelizumab	Rituximab	Effect estimate
Eligible MS cohort	1220	451	—
All-cause hospitalization	90/1220 (7.38%)	89/451 (19.73%)	RR 2.67 (2.04–3.51)
Hospitalization visit rate	11.69 / 100 PY	19.91 / 100 PY	Rate ratio 1.70 (1.54–1.88)
Common infection-related hospitalization	18/1220 (1.48%)	21/451 (4.66%)	RR 3.16 (1.70–5.87)
Hypogammaglobulinemia <565 mg/dL among IgG-tested	40/907 (4.41%)	34/348 (9.77%)	RR 2.22 (1.43–3.44)

Absolute event risks

All-cause hospitalization

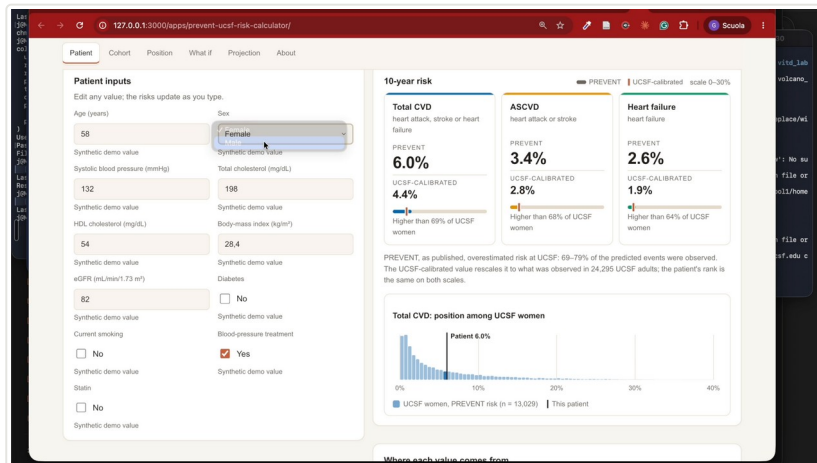
Ocrelizumab 7.38%

Rituximab 19.73%

 jupyterday 

Demos: clinical applications and knowledge graphs

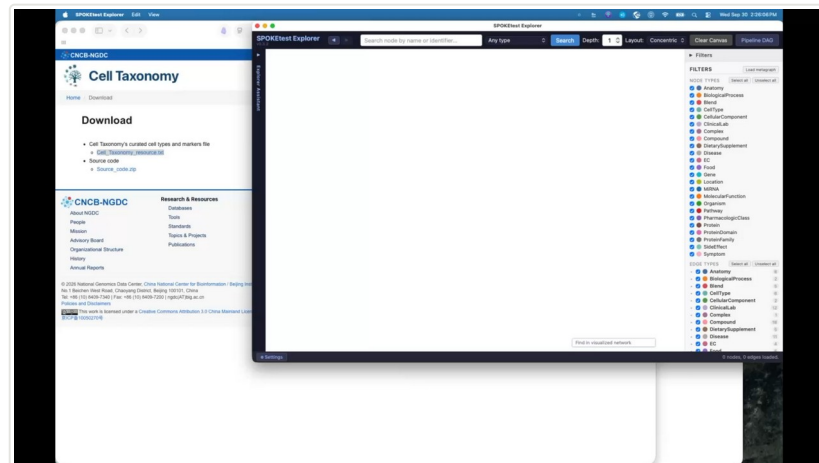
Using UCSF EHR data and the SPOKE knowledge graph.



Demo CVD Risk Application

Cohorts from UCSF EHR, AHA PREVENT risk from EHR data alone, and an interactive app built by the agent.

Agrees with preventr (R reference); under clinician review.

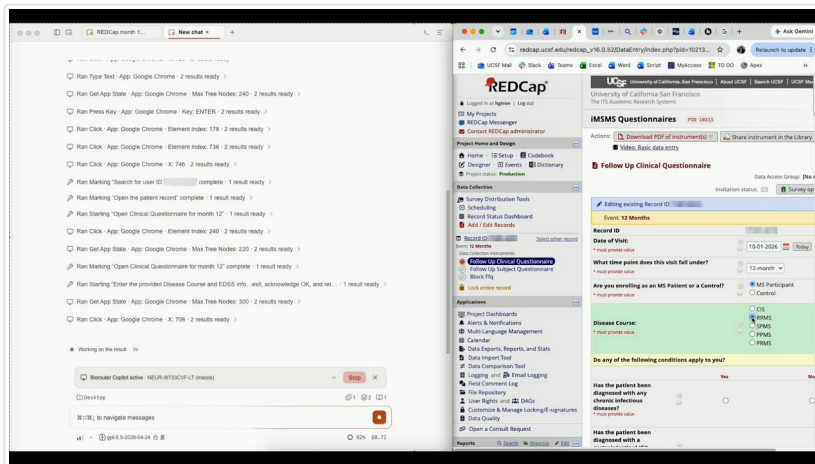


Demo Data Integration

Adds Cell Taxonomy markers to the SPOKE knowledge graph with SPOKE FORGE. A researcher sets the scope; the agent models and builds it.

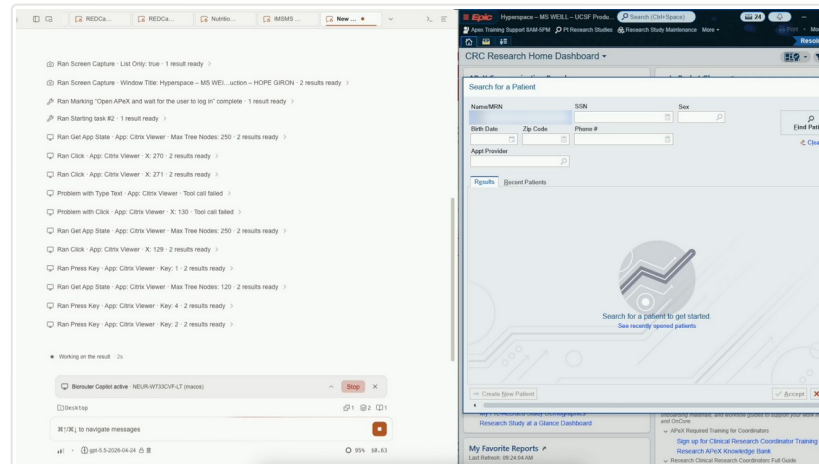
Biorouter Copilot

For data reachable only through a browser or clinical software.



Biorouter Copilot REDCap

Searches and sorts records, registers subjects, enters visit data.
Secure research surveys and databases, used through the browser.



Biorouter Copilot APEX

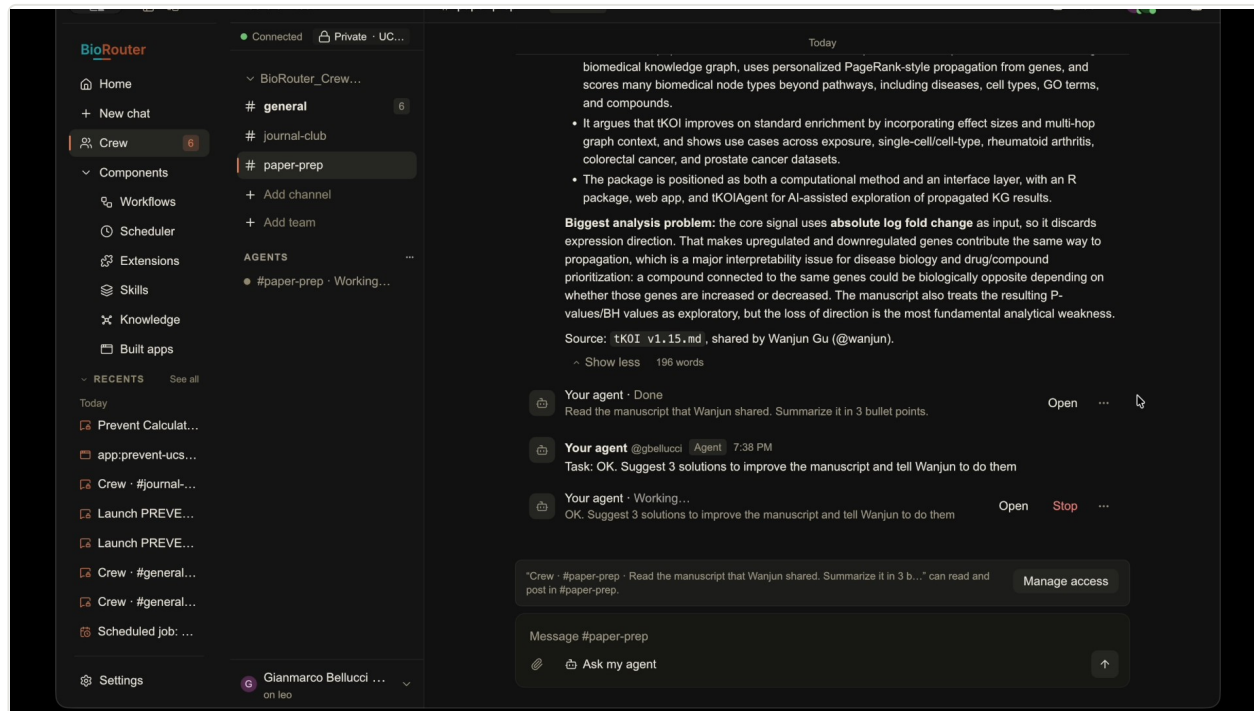
Opens APEX through Citrix, finds a patient by MRN, and reviews research studies and clinic visits.
UCSF's Epic EHR, the clinician-facing system of record.

Crew Beta

Lab channels with agents.

- Channels on the lab's own cluster
- Agents run tasks; people steer and share context
- Data stays on the cluster

Demo: editing a manuscript as a team, on the lab's Leo cluster. Works the same on Wynton.



Acknowledgements

Baranzini Lab

Gianmarco Bellucci, MD, PhD
Ilan Ladabaum, MS
Hope Giron
Braian Peetoom, MD
Amit Akula
Sergio Baranzini, PhD

Bakar Institute

Sharat Israni, PhD
Marina Sirota, PhD
Boris Oskotsky, PhD
Tomiko Oskotsky, MD

UCSF Academic Research Services

Evan Phelps, PhD
William Santo
Oksana Gologorskaya, MS
Anna Rubinsky, PhD
Rick Larsen, PhD

Jupyter Foundation, UC Berkeley

Min Ragan-Kelley
Sandeep Jayaprakash
Kirstie Whitaker, PhD
David L. Qiu
Fernando Pérez, PhD

MoonLAIT Lab · UCSF Psychiatry

Yang Yang Lab · UCSF Radiology

Individual Developers and Contributors

Yue Leng, PhD · Christina Kushnir, MS ·
Xing Xu, MD

Yang Yang, PhD · Zheren Zhu

Ethan Shen · Xi Zheng, MSc · James Xue, MSc · Jonathan Xue ·
Tianlu Zhu, MS · Vladimir Mitev



Questions and feedback

Scan the code to ask a question or leave feedback on this session.

Try Biorouter biorouter.ucsf.edu