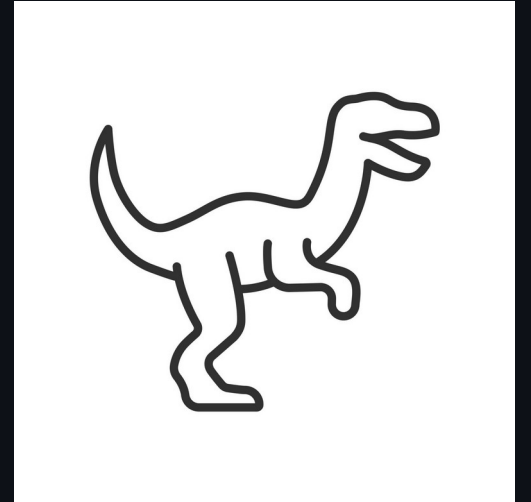


Validated skills are necessary but not sufficient for trustworthy agentic genomics

Science after pipelines

Manuel Corpas

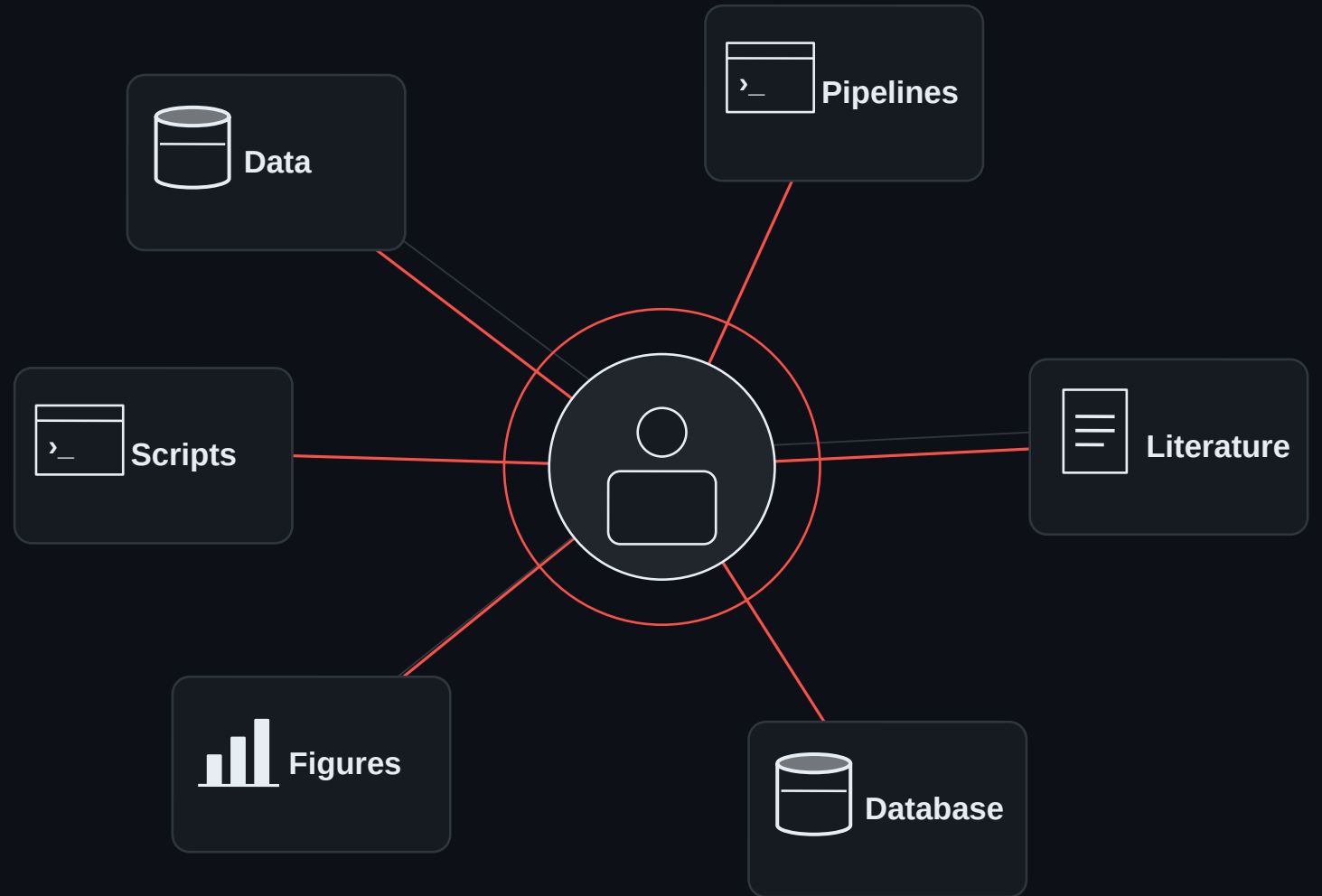
Francis Crick Institute





The human is the API.

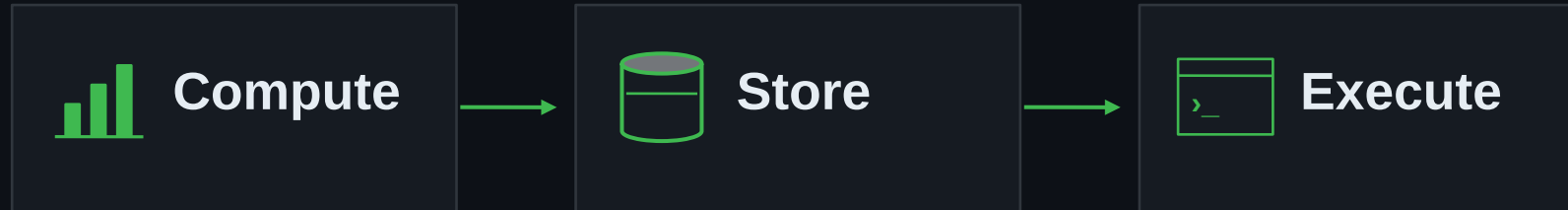
Expert attention
becomes glue.





Computation is automated. Coordination is not.

CAPABILITIES



HAND-OFFS

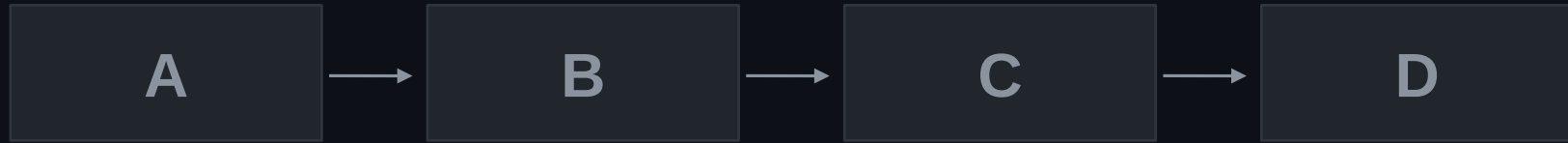


The scientist still connects the work.

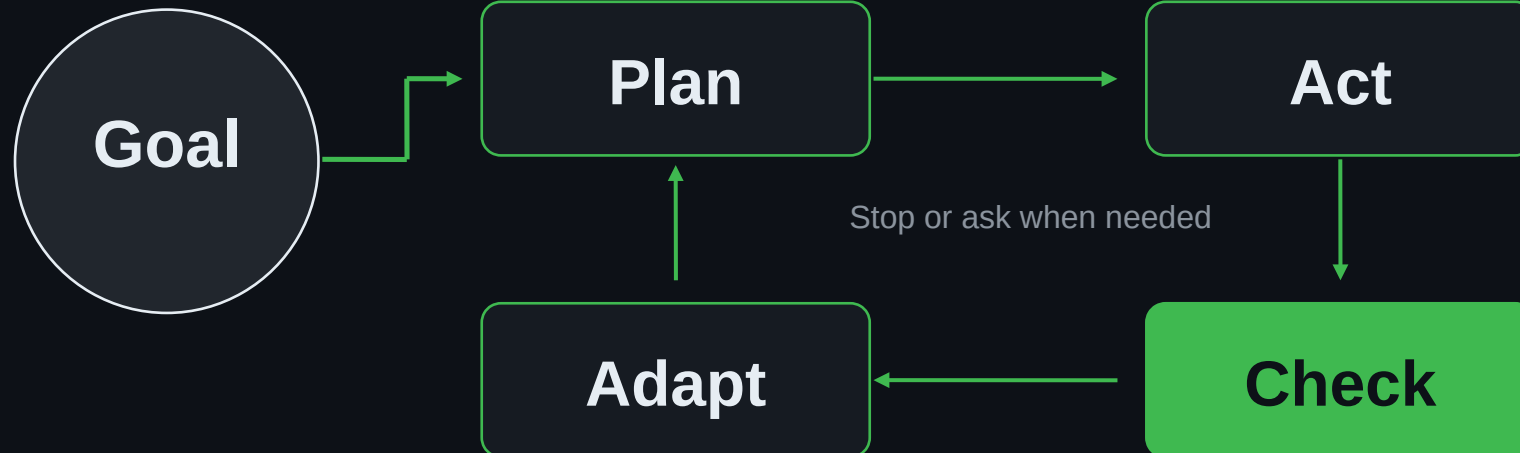


From commands to objectives.

COMMANDS



OBJECTIVE





A different morning in the lab.

Less operating.
More investigating.

OVERNIGHT ANALYSIS

SCHEMATIC

What explains the phenotype?

A	Hypothesis	Support
B	Alternative	Conflict
?	Unresolved	Unknown

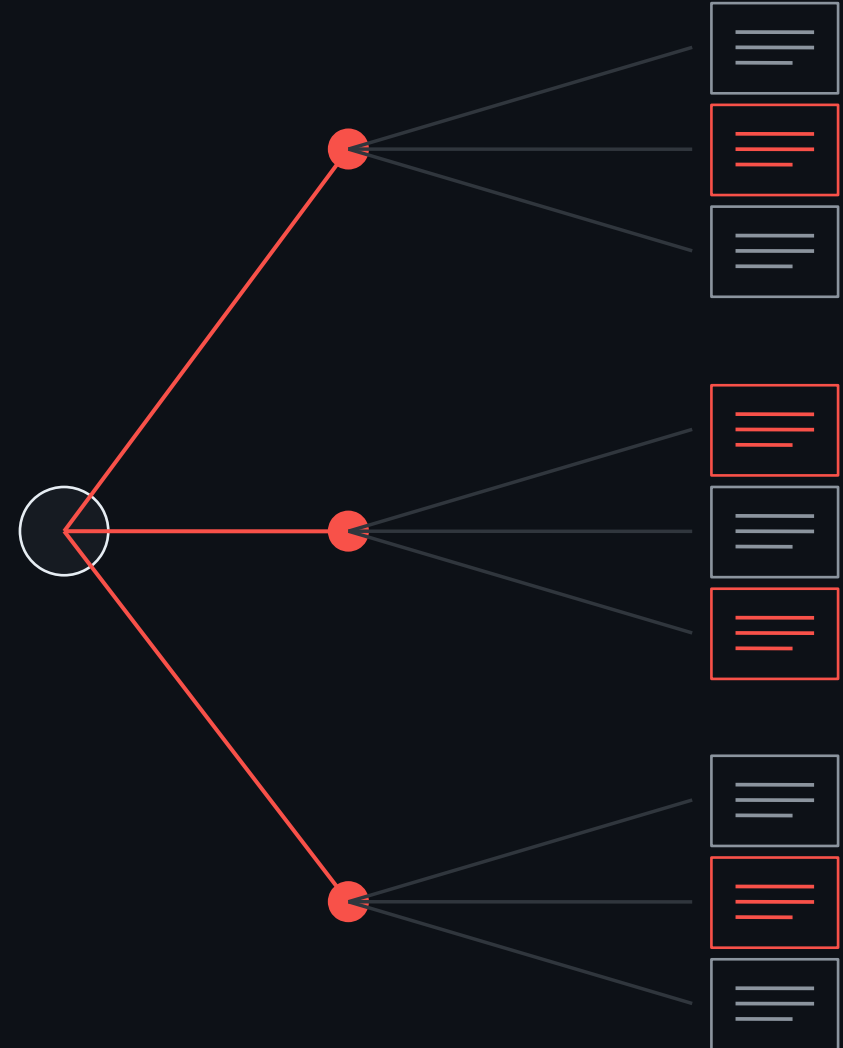
NEXT: a discriminating experiment



Bad science.

At machine speed.

Unconstrained choices.
Convincing conclusions.

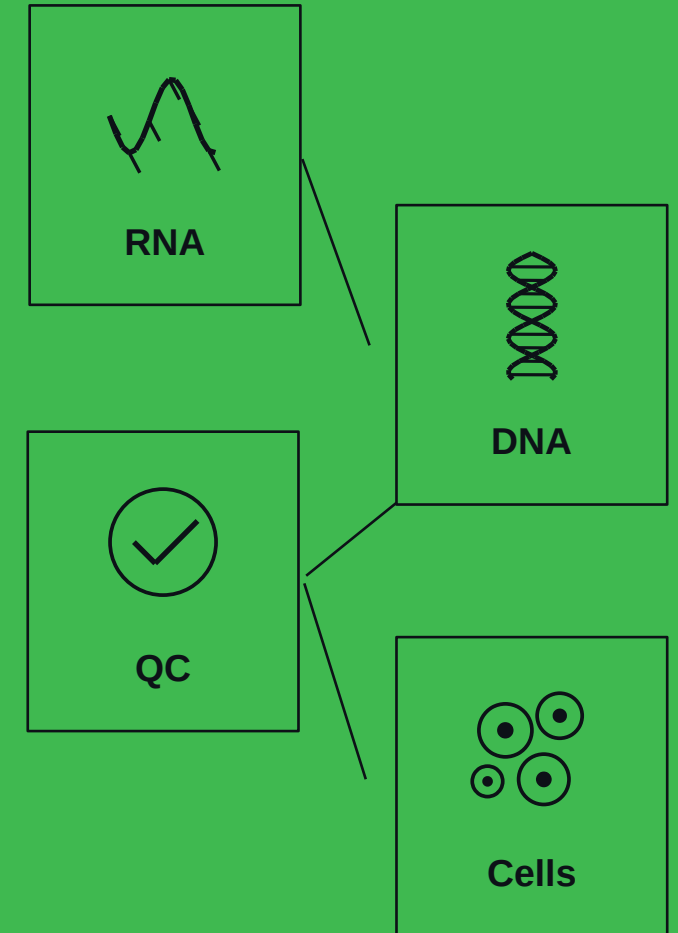




ClawBio

A skill library for agentic biology

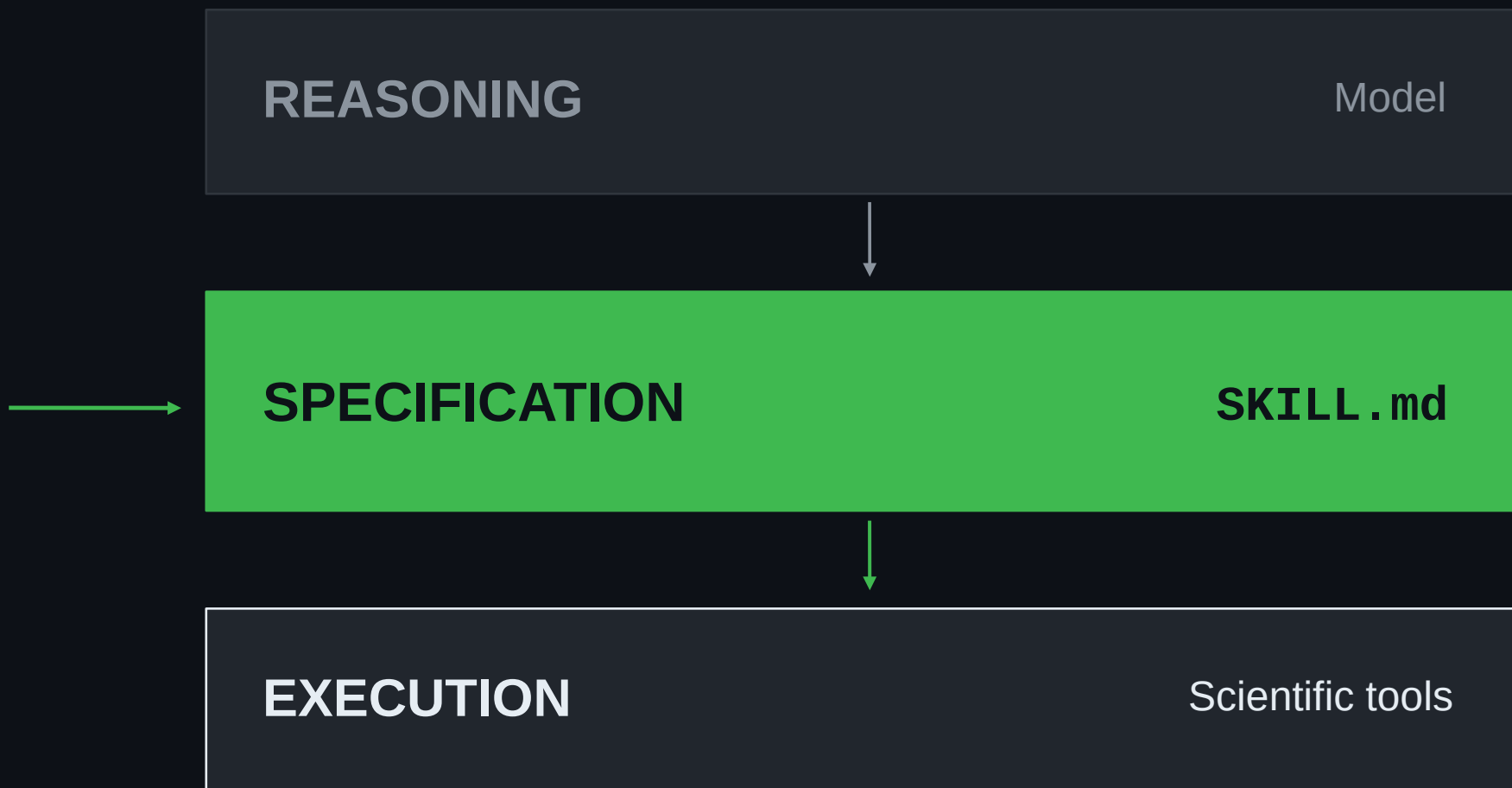
Agents reason.
Skills constrain.
Tools compute.





The missing layer.

Expertise lives here.



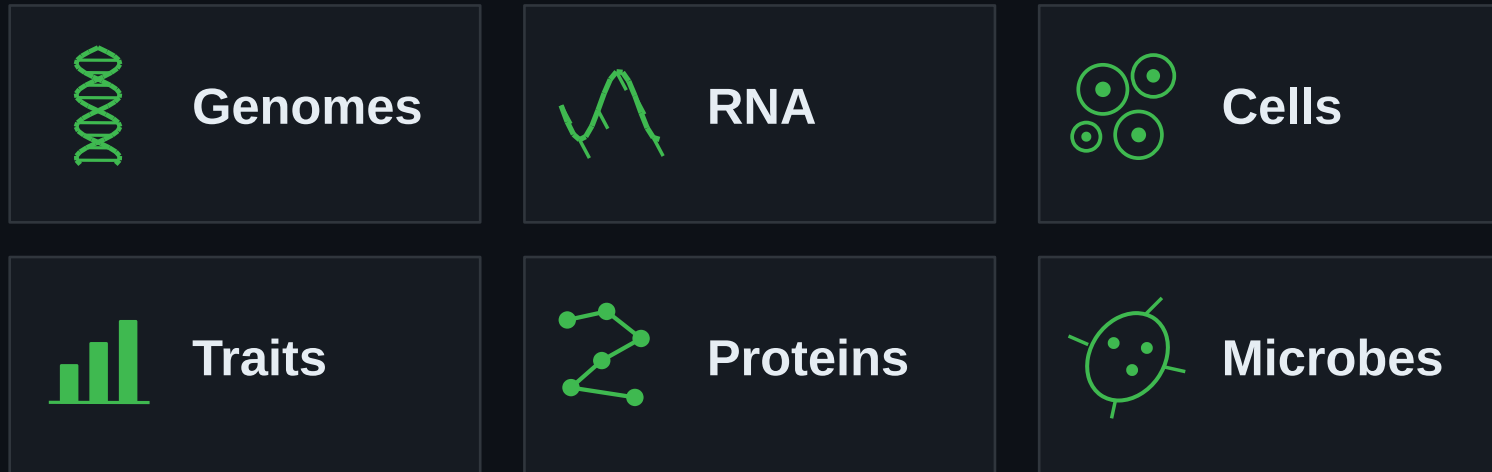


An executable methods section.





An ecosystem. Not another pipeline.



97
scientific skills

92
runnable demos

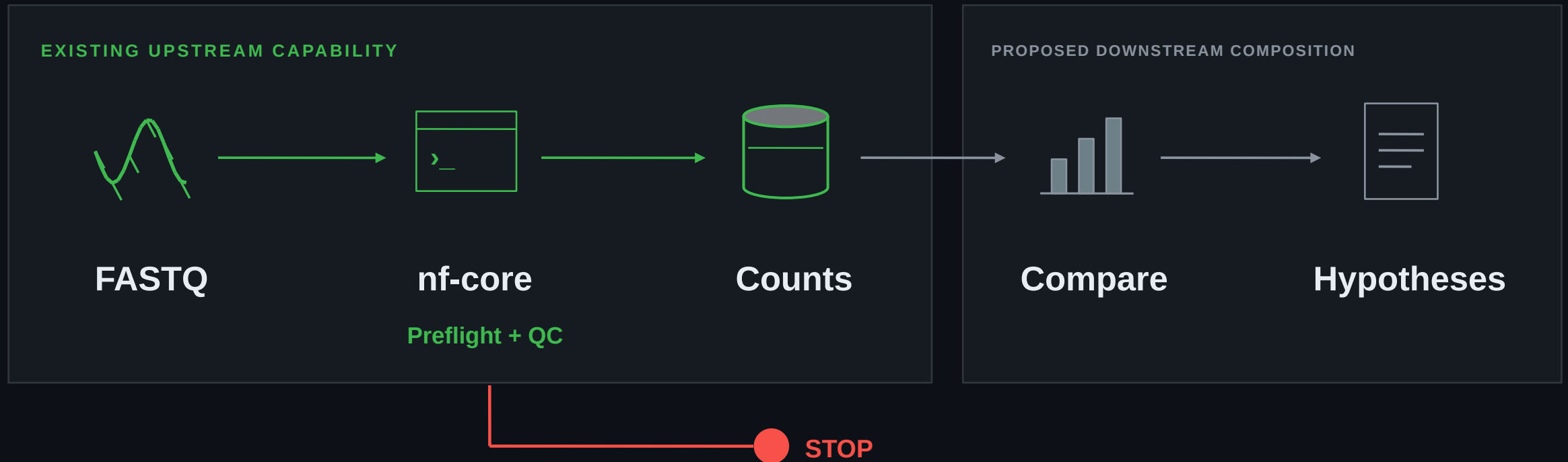
8,182
Galaxy tools

Catalogue breadth $\neq$ uniform validation



From RNA reads to a biological question.

Why does the perturbation rescue the phenotype?



Human judgement stays in the loop.



Every answer needs an evidence trail.



Inputs



Method



Execution



Outputs



EVIDENCE

- What ran
- What changed
- What failed

Replayable $\neq$ scientifically valid



Trust is measured.

161 / 182

reported checks passing

10 audited skills

Not clinical accuracy.
Not every skill.

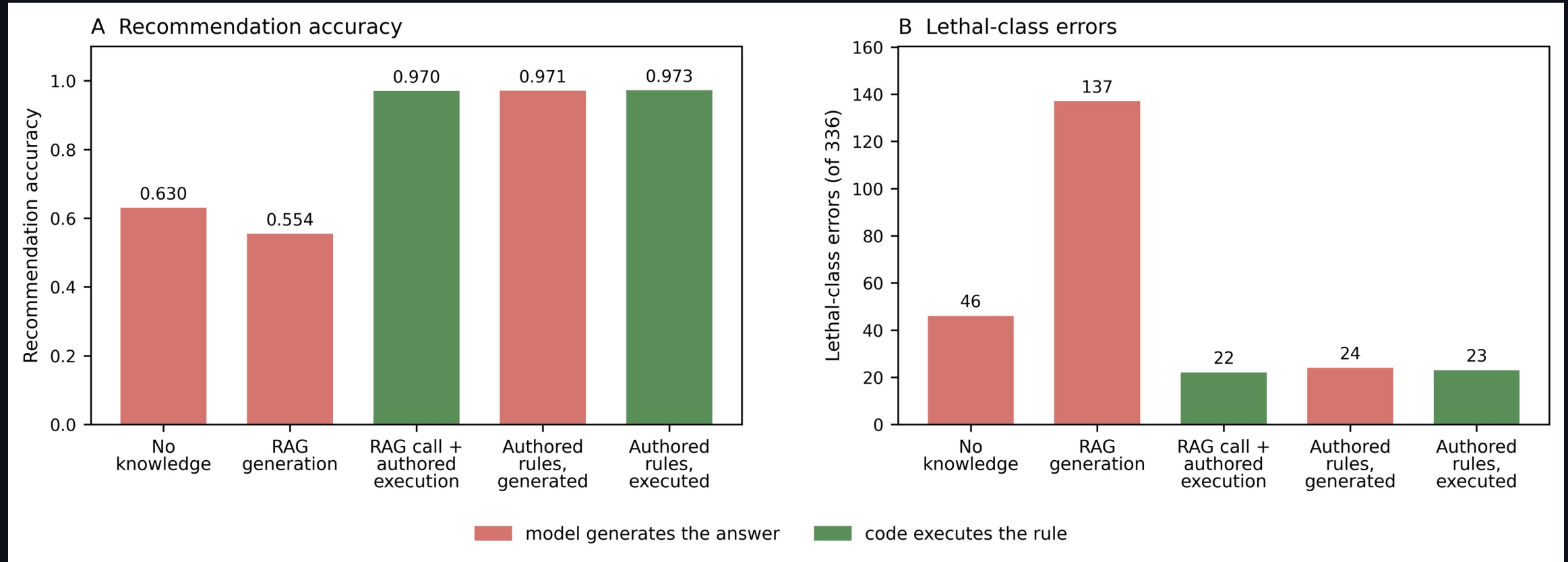


■ 161 passing

■ 21 remainder



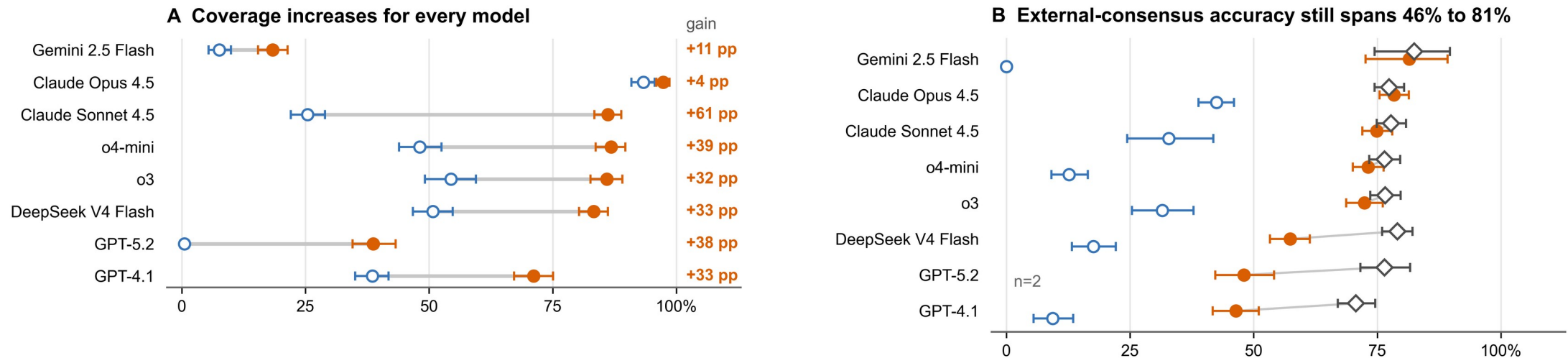
Validated rules change the result.





The input still needs validation.

○ No definitions ● Validated definitions supplied ◇ PyPGx on the same emitted pairs



Paired variant-call inputs: n=527 per model; identical units for all eight models.
Error bars: sample-cluster bootstrap 95% intervals; accuracy omitted below 10 calls.

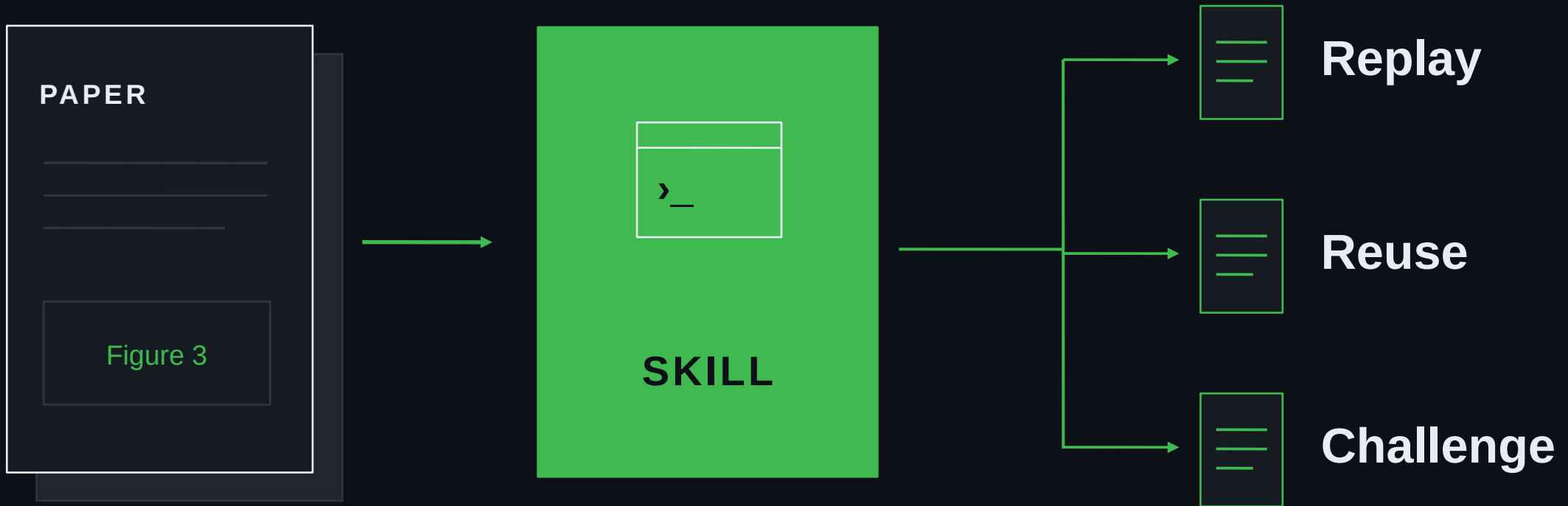


Build on the tools we already trust.





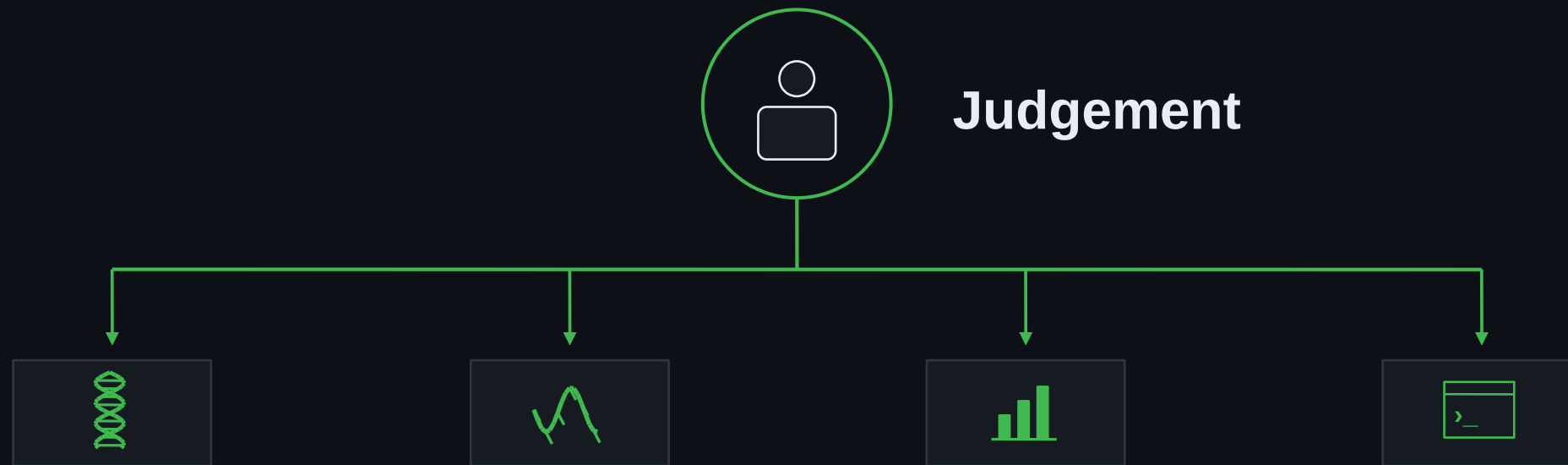
Publish methods that can execute.



A paper becomes an interface.



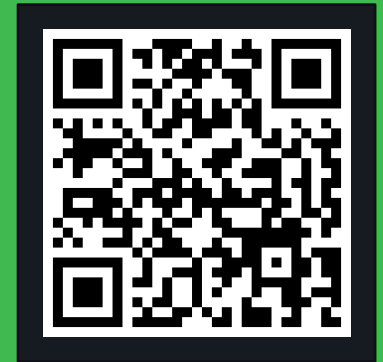
From operator to scientific director.





What would you trust an agent to do?

Which task?
What evidence?



github.com/ClawBio/ClawBio



Try a skill. Share what you learn.

01 Read the documentation and run a demo

docs.clawbio.ai

02 Follow the calendar for workshops and hackathons

luma.com/clawbio

03 Join the conversation in ClawBio Slack

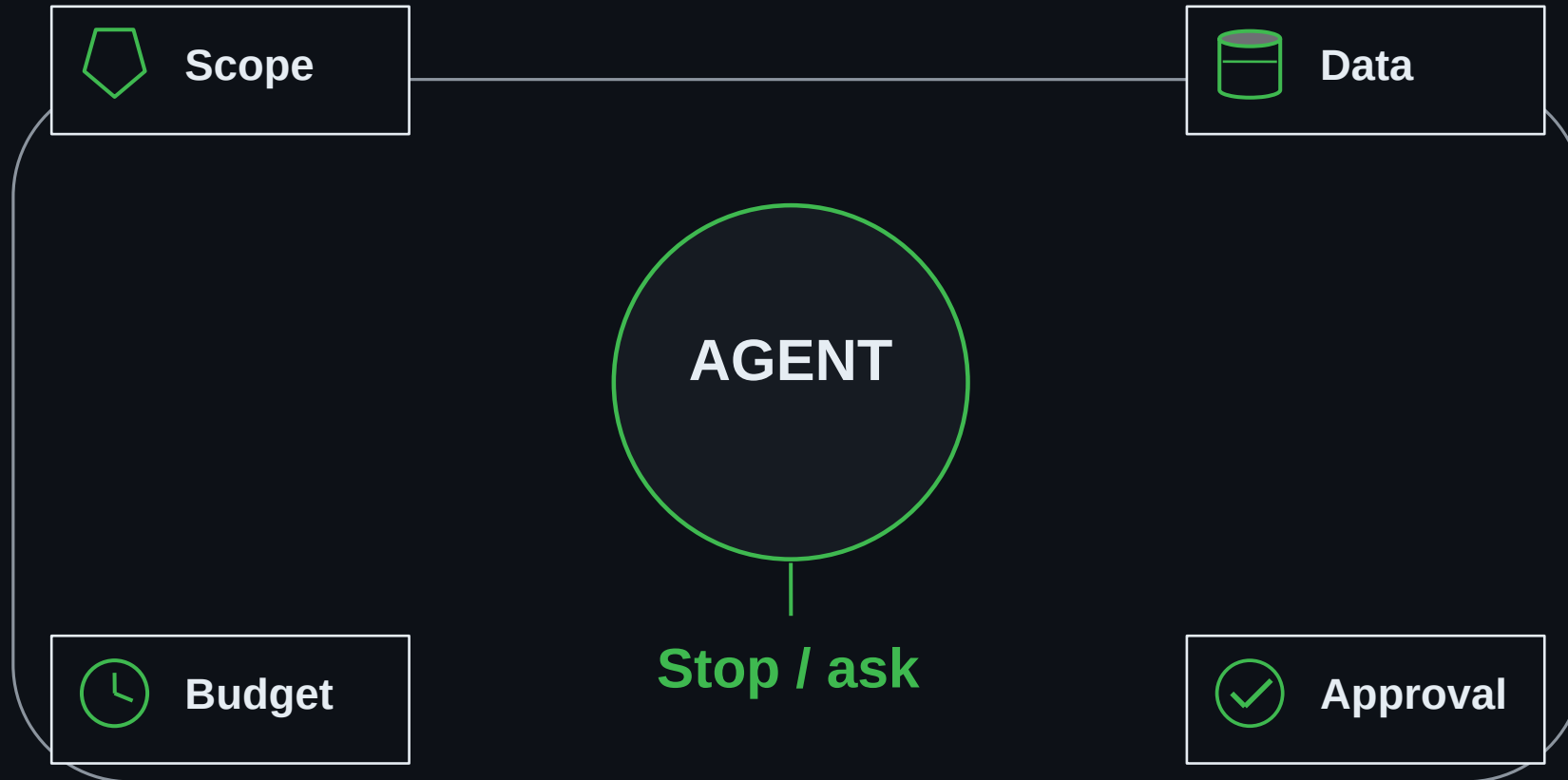
[Open the ClawBio Slack channel](#)

Choose one analysis. Define the evidence you would require.

clawbio.ai



Autonomy needs boundaries.



Local-first ≠ offline



Sources & scope

01 Current capabilities + reported counts

[README.md](#)

02 Pharmacogenomics: manuscript v90 + release v2.10

doi.org/10.5281/zenodo.22141182

03 Declared data destinations + network behaviour

[docs/data-handling.md](#)

Sources checked 12-13 September 2026. Repository counts were not independently rerun.

Component benchmark; no claim of end-to-end clinical validation.