Agentic genomics

Your science, made executable.

Skills, evals and equity for trustworthy execution

Manuel Corpas

University of Westminster / 7 September 2026

What would you test if you could direct the analysis yourself?

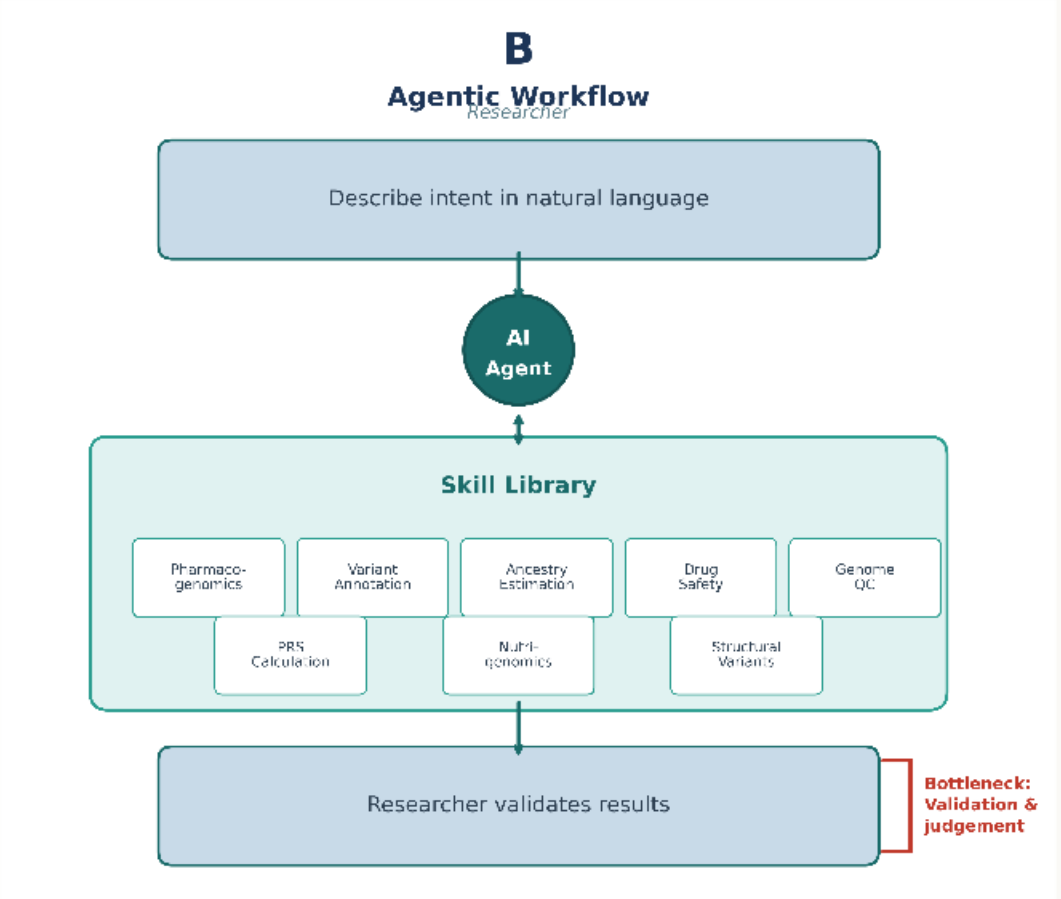
Inflammation. Ageing. Cancer. Cell identity.

Your expertise moves to the centre

Define the question.

Direct the execution.

Judge the evidence.

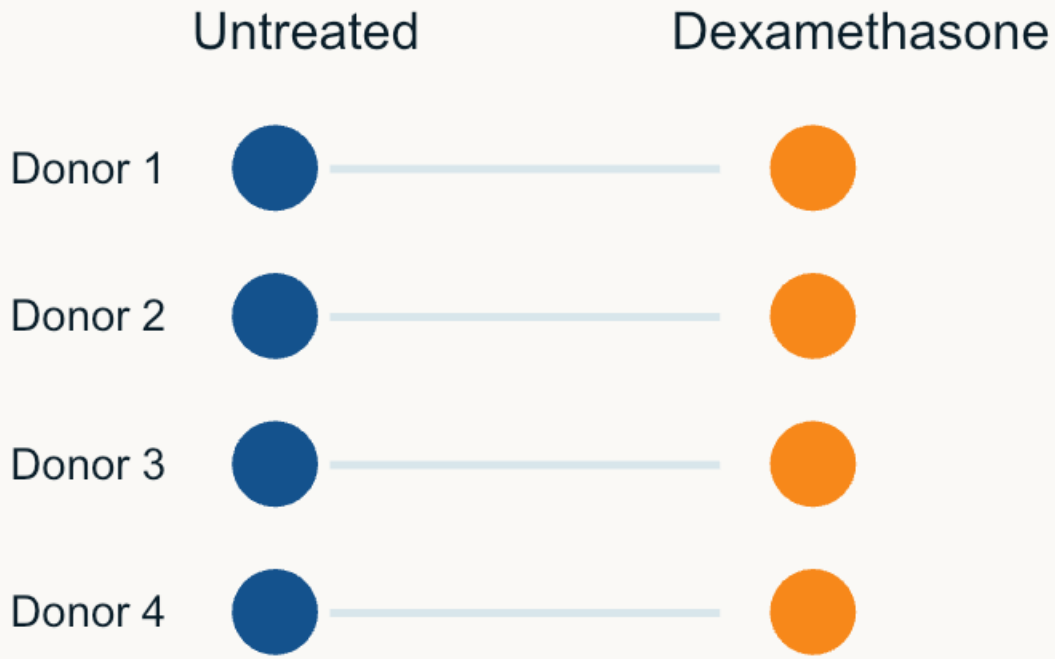


The bottleneck shifts towards validation and judgement.

Corpas, Guio & Fatumo. Cell Genomics (2026)

Which genes respond to dexamethasone?

Human airway smooth muscle
4 donor pairs
8 samples



Paired samples, not eight independent donors

Give the agent scientific context

Compare dexamethasone with untreated cells.

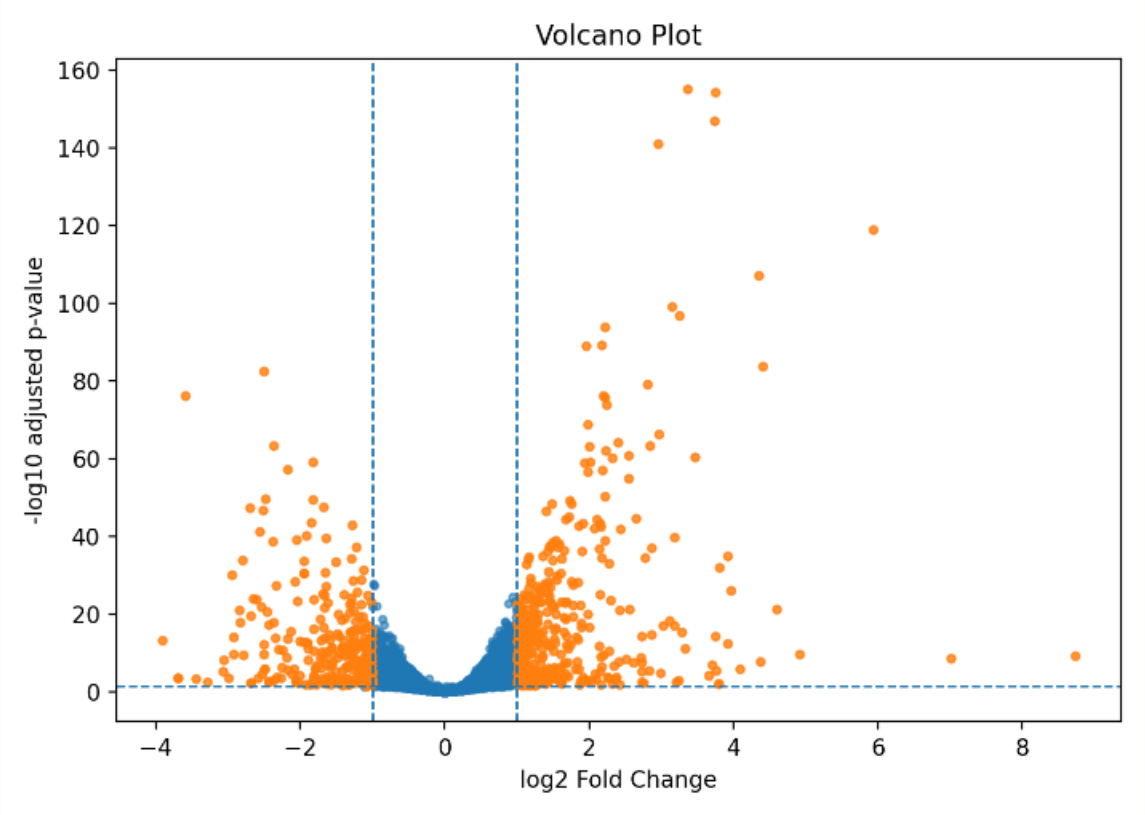
Keep the four donor pairs in the design.

Use rnaseq-de with PyDESeq2.

Stop if the requested control is absent.

Return QC, gene statistics and a reproducibility bundle.

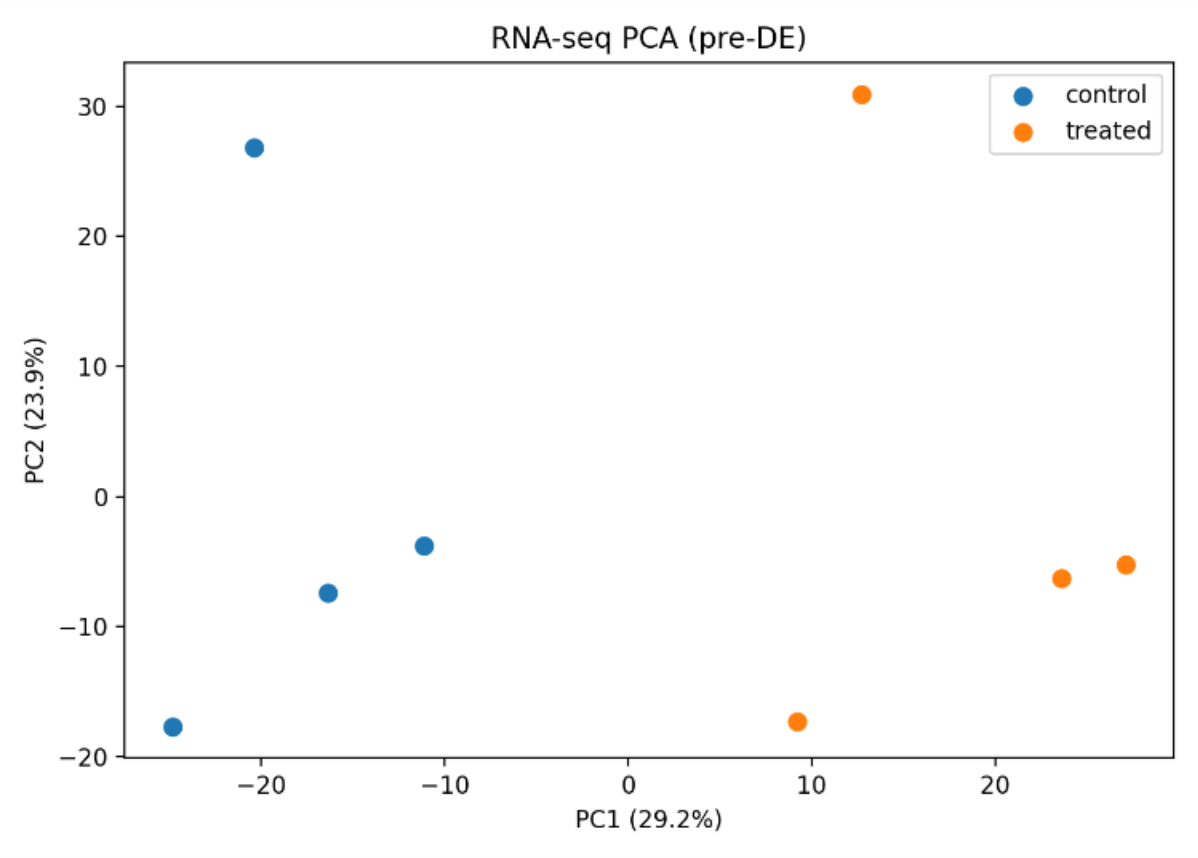
A question becomes a result you can inspect



677 genes meet the cut-offs

Adjusted p < 0.05
|log2 fold change| >= 1
16,172 genes tested

The design is part of the answer



Donors are not technical replicates.

The paired model:
~ celltype + dex

Run it again. Inspect the evidence.

Same input files. Same pinned operation.
Byte-identical gene-results table.

RUN 1

`a0c5bd3ad8005c61`

=

RUN 2

`a0c5bd3ad8005c61`

SHA-256 prefixes for de_results.csv. A fingerprint, not a validity claim.

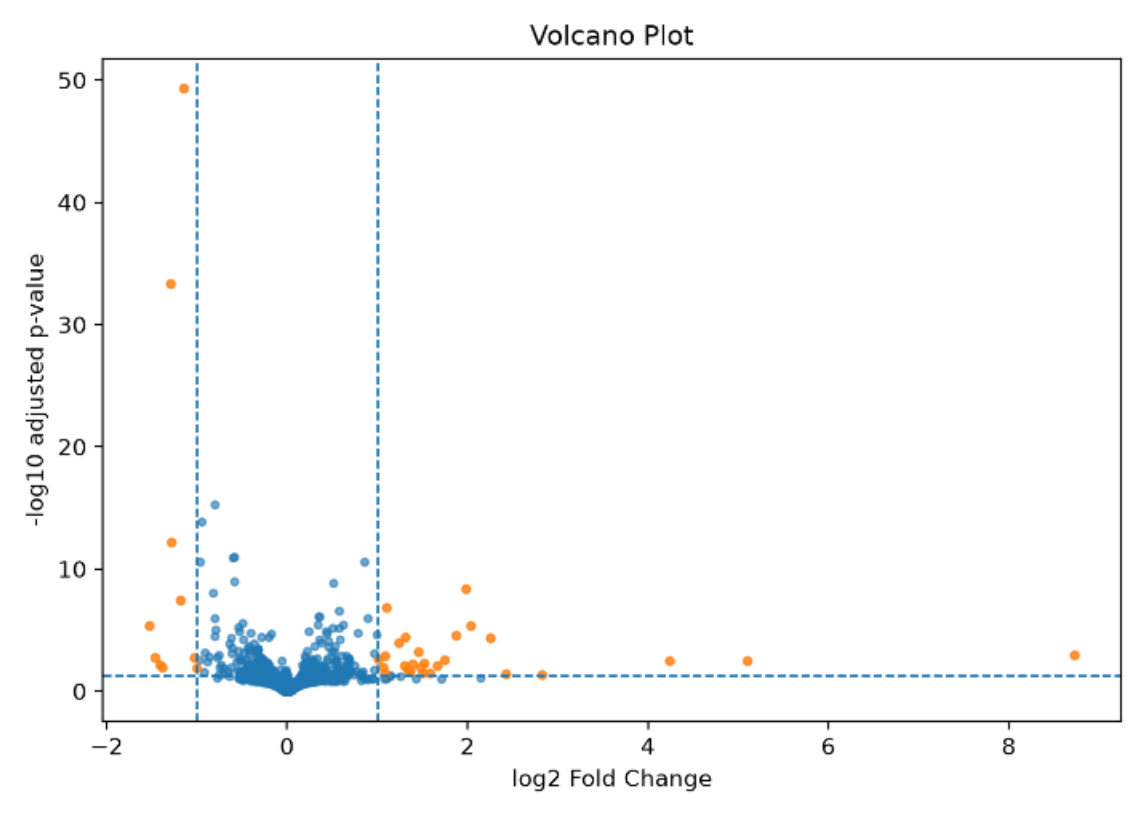
Now give it a control that does not exist

RECORDED ERROR

```
Contrast denominator 'unknown_control'
not present in dex
```

Exit code 1. No differential-expression result.

Recover a published biological signal



STMN2 falls after TDP-43 depletion

12 samples
log2 fold change: -1.15
adjusted p: about 4.9 x 10^-50

Klim et al. (2019), GSE121569. Original Crick demo figure

A skill is your method, made reusable

```
rnaseq-de/
  SKILL.md
  rnaseq_de.py
  tests/
  examples/
```

Scientific instructions

Executable operation

Tests and example data

Evidence from the run

Would you let this analysis proceed?

	BATCH A	BATCH B
Younger		No samples
Older	No samples	

Which effect is age, and which is batch?

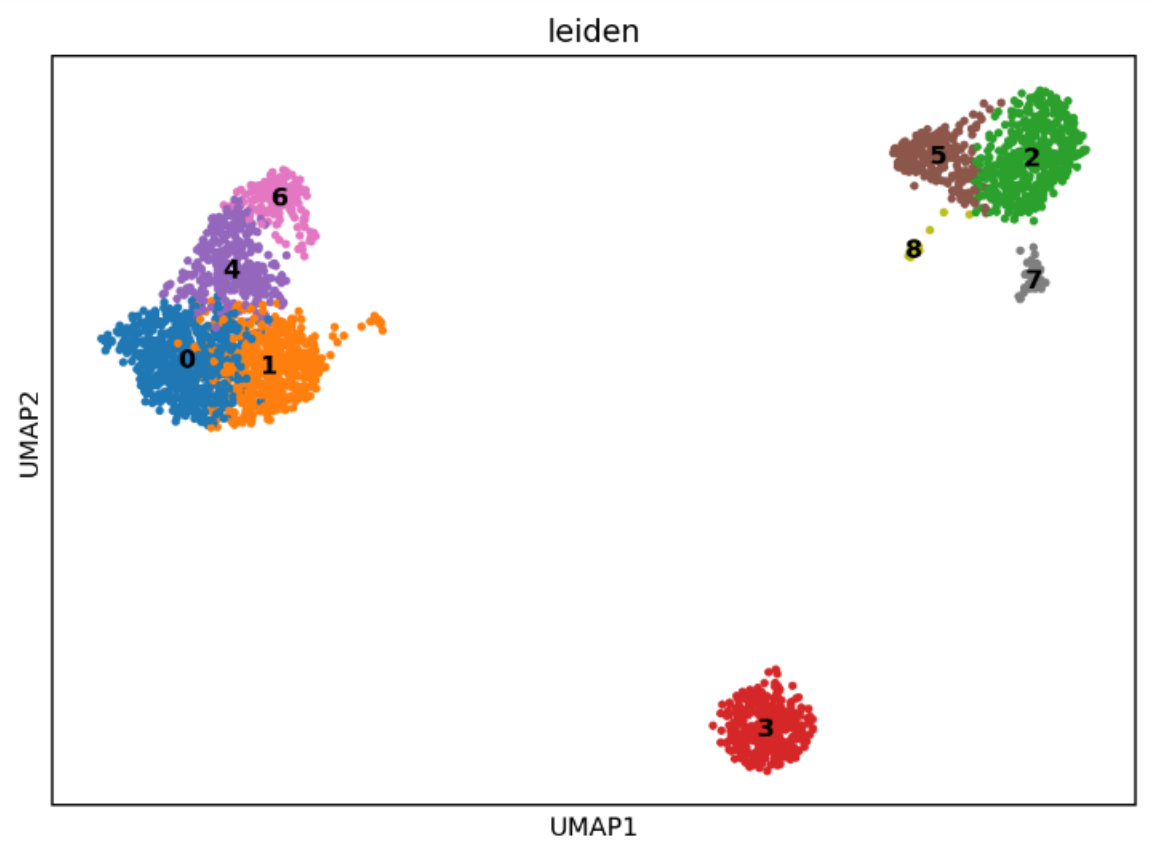
Illustrative design exercise

That is context engineering

1	2	3	4	5
Question	**Data**	**Design**	**Stop rules**	**Evidence**

You do not need to start by learning syntax.
You do need to make the science explicit.

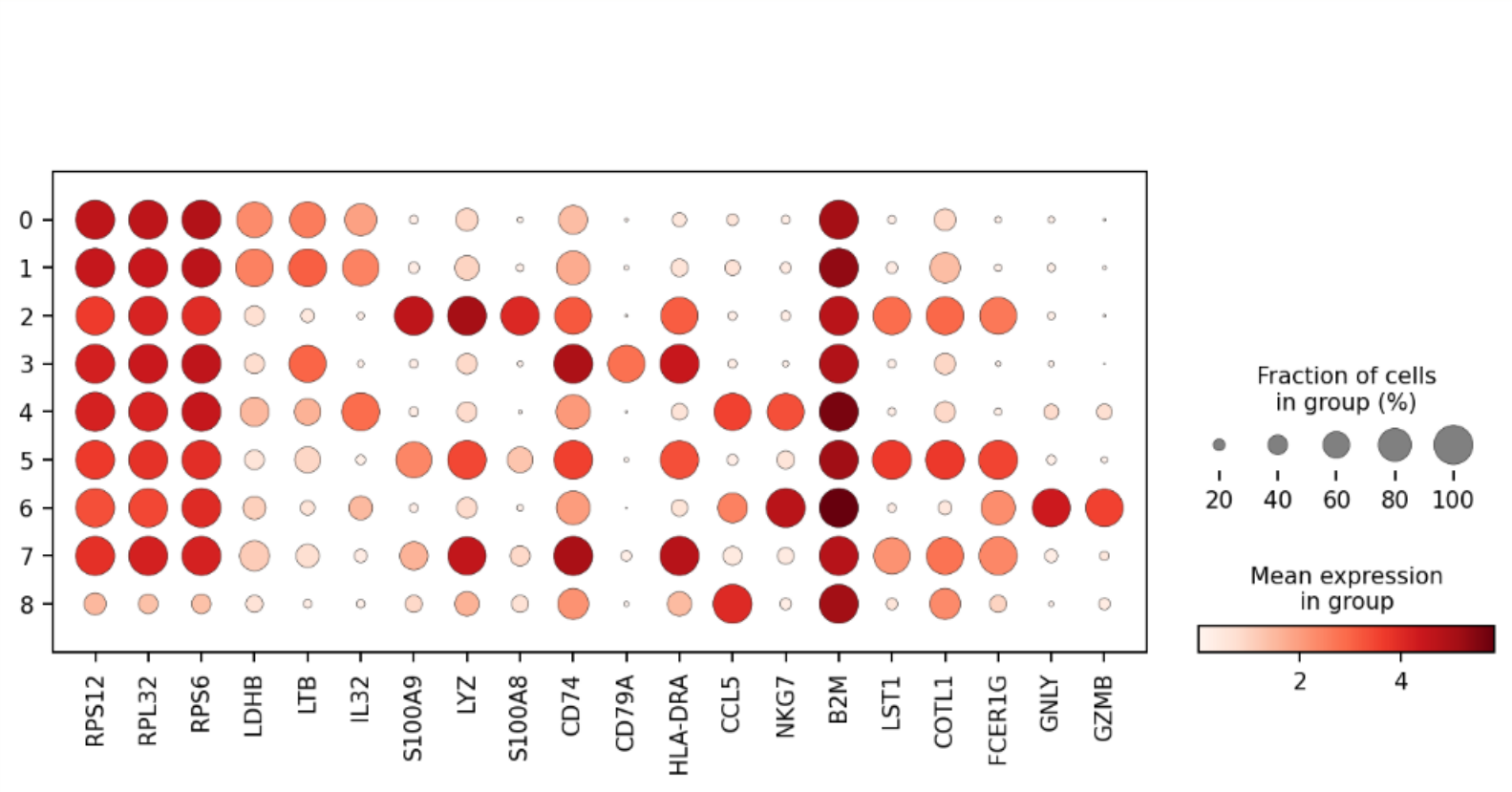
Which cells carry the signal?



A real immune-cell dataset

Public PBMC3k raw counts
QC, clustering, marker discovery

A cluster is a hypothesis, not an identity



Check marker evidence before naming the cells.

An ageing question needs more than an age estimate

`methylation-clock`

Does the assay match?

Are the required CpGs present?

Was the clock validated for this cohort?

A workflow to evaluate, not a clock result or clinical claim

Write your first context contract

Compare [group] with [control].

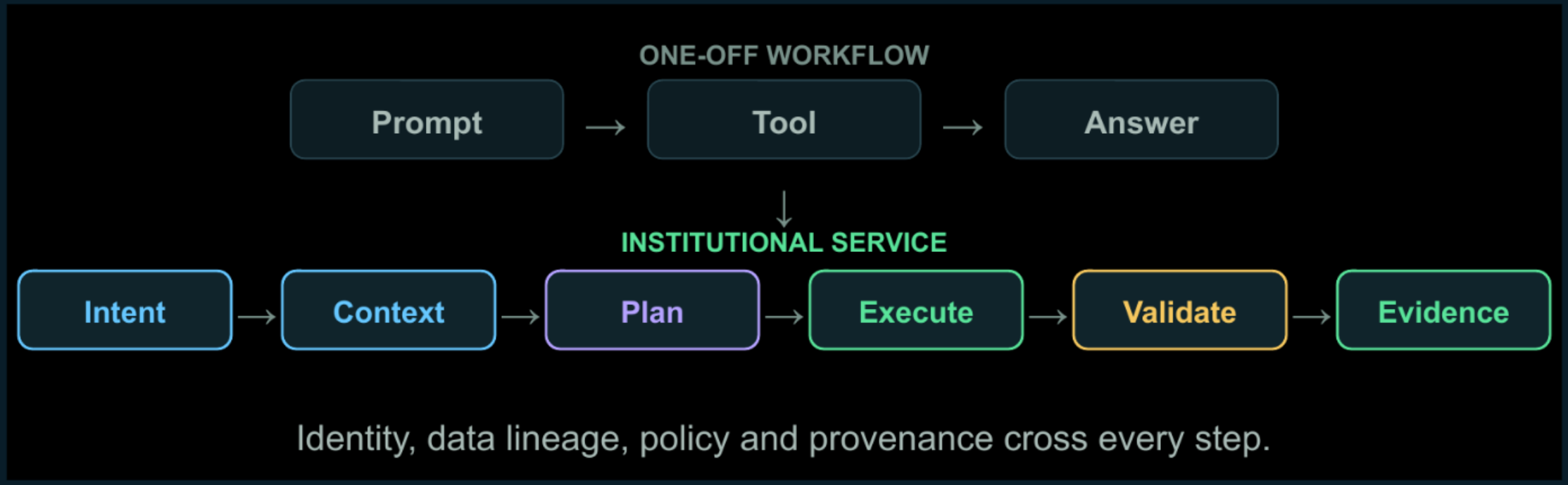
Account for [design or covariate].

Stop if [critical condition fails].

Return [evidence I can inspect].

Use the question you brought into the room.

Clinical-grade trust belongs to the whole process

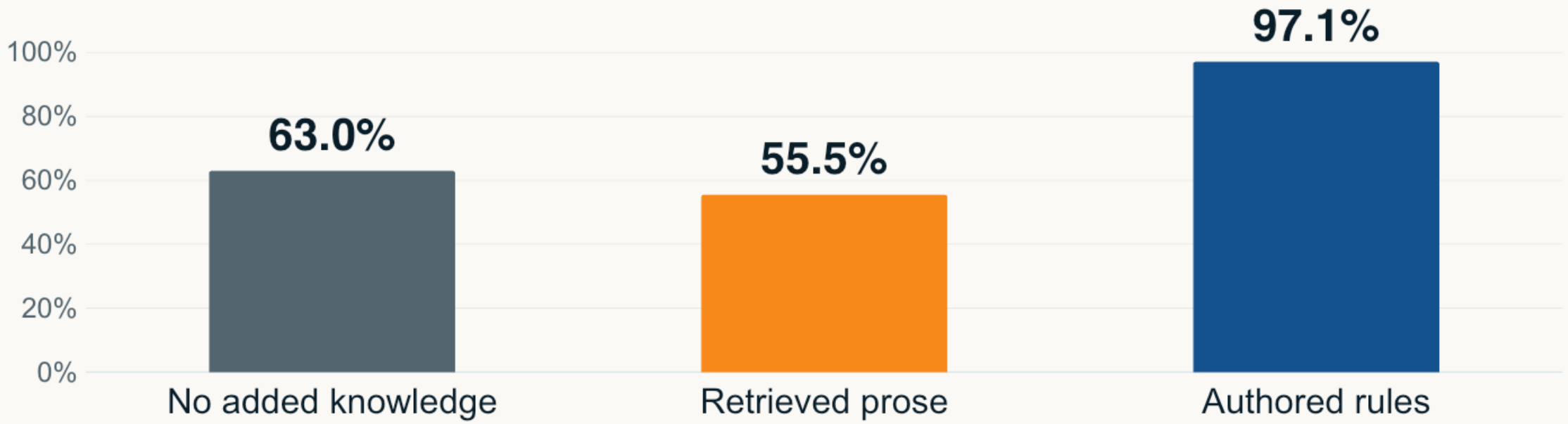


Deterministic execution. Traceable evidence. Validated use.
Skills execute. Evals challenge. Equity defines for whom.

Original visual: Manuel Corpas, Crick RNA Ideathon, 14 Jul 2026

Authored rules changed the score

Mean recommendation score (partial credit included)



2,640 attempts per arm. Authored logic and deterministic execution are distinct contributions.

A repeatable procedure can repeat a wrong rule

87 / 90

Responses echoed deliberately corrupted rules.
The other 3 hedged. None clearly rejected them.

Write the exam before trusting the answer

01 Known positive

02 Wrong control

03 Confounded design

04 Unseen cohort

What should pass? What must stop?

Whose science is missing from the validation?

4 donors

are not everyone.

Age. Sex. Ancestry. Tissue. Resources.

Count missing answers as well as wrong ones.
Limit claims to the settings actually tested.

Clinical grade is the destination

1

Research tool

2

Benchmarked workflow

3

Clinical-grade process

Independent validation. Defined use. Governance. Monitoring.

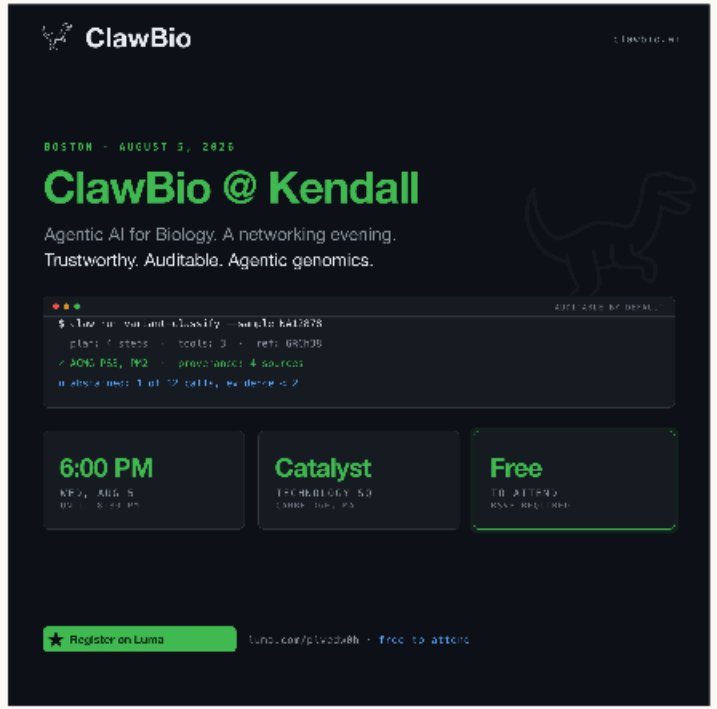
A SCIENTIST WHO STARTED

"Something clicked."

Cameron Lloyd
Scientist, bioinformatician, contributor

An idea became an MVP at a ClawBio hackathon.

People are making it their own



London. Madrid. Berlin. Cambridge, MA.

Original ClawBio event artwork. London, Berlin and Cambridge MA, 2026

A scientific commons you can inspect

Open source. Transparent. Academic roots.
Community built. Open to challenge.

Evidence before vendor loyalty.

1,126 stars / 48 commit-contributor accounts

ClawBio public repository. Metrics checked 6 Sep 2026

Your first contribution need not be code

A missing control.
A better reference case.
A population we overlooked.

A real community fix: verify that RNA-seq shrinkage matches the contrast the scientist requested.

ClawBio PR #378, krudo-taco, merged 31 Aug 2026

Three things to take home

1 **When, not if.**

Agentic execution is already possible.

2 **Start while the field is forming.**

Early adoption can change what you can do.

3 **Speak the language of context.**

Your scientific judgement is a way in.

YOUR BEGINNING

Your next question is where this begins.

Open one skill. Make one result inspectable.

docs.clawbio.ai

Skills. Evals. Equity.

Replay and live-demo parameters

AIRWAY: 8 samples, 4 paired donor-derived lines
~ celltype + dex; dex,treated,control
PyDESeq2; min count 10 in at least 2 samples
STMN2: 12 samples; ~ condition
condition,TDP43,control; PyDESeq2

All inputs, commands, outputs and checksums: DEMO-PACK

Single-cell and methylation boundaries

PBMC: public raw counts, not a disease cohort
Scanpy QC, UMAP, Leiden, marker discovery
No CellTypist labels requested
Methylation clock: contract example only
No age prediction or clinical claim shown

Keep the biological donor as the unit of replication.

The scientific foundations

Corpas, Guio & Fatumo. Cell Genomics (2026).
Agentic genomics: From pipeline automation to autonomous validation.
DOI: 10.1016/j.xgen.2026.101305

Corpas et al. Benchmark manuscript v90 + supplement v48.
Public data release: DOI 10.5281/zenodo.22141182

ClawBio Technical Note v5. Working manuscript.
Inspectable specification layer; evaluated 8 August 2026.

Datasets, figures and community credit

Himes et al. PLOS ONE (2014), GSE52778.
DOI: 10.1371/journal.pone.0099625
Klim et al. Nature Neuroscience (2019), GSE121569.
DOI: 10.1038/s41593-018-0300-4
PBMC3k: 10x Genomics / Scanpy public example.
Original figures: Manuel Corpas, Imperial and Crick talks.
Contributor story: Cameron Lloyd, April 2026.
ClawBio skills and events: docs.clawbio.ai