

10

AI BIOINFORMATICS STARTUP IDEAS

WORTH BUILDING IN 2026



1. CLINICAL VARIANT CURATION AND REPORTING COPILOT

ACMG/AMP EVIDENCE CATEGORIES

	PS	Pathogenic Strong	e.g., De novo (PS2), Functional study (PS3)
	PM	Pathogenic Moderate	e.g., Absent from controls (PM2), Mutational hot spot (PM1)
	PP	Pathogenic Supporting	e.g., Multiple lines of computational evidence (PP3)
	BA	Benign Strong	e.g., Allele frequency >5% (BA1), Well-established benign (BA2)
	BS	Benign Supporting	e.g., Computational evidence (BS1), Lack of segregation (BS4)
	BP	Benign Supporting	e.g., In silico predictors (BP4)

CLASSIFICATION: Likely Pathogenic

Evidence Aligned Lab Rules Applied Confidence High (0.92) ACMG/AMP v3.3



DRAFTS. NOT DECISIONS.

Assembles evidence, applies ACMG/AMP codes and lab rules, and drafts reports with citations—so clinicians review, not rebuild.



THE DATA FLYWHEEL

Every curated case enriches a proprietary database. Recurring variants auto-match prior annotations. Only novel variants need fresh curation.



REGULATORY MOAT

CLIA/CAP requirements, audit trails, and documentation standards create high switching costs and durable stickiness.

VARIANT REPORT DRAFT

Patient: 27-8492 Sample: WES

VARIANT SUMMARY

Gene: BRCA1
Variant: c.5266dupC
Protein: p.Gln1756Profs*74
Zygosity: Heterozygous
Classification: Likely Pathogenic
Confidence: High (0.92)

CITATIONS (5)

- PMID: 12345678
- PMID: 23456789
- PMID: 34567890
- PMID: 45678901
- PMID: 56789012

AUDIT TRAIL

- ✓ Evidence Assembled
- ✓ Rules Applied
- ✓ Reviewed by: Dr. A. Lee
- ✓ Date: 2025-05-20

CLASSIFICATION RATIONALE

ACMG/AMP Evidence Codes:

PS2 PM2 PP3 PM1 PPS

Population Data

Absent from gnomAD (PM2)

Computational Evidence

Multiple lines of computational evidence support a damaging effect on the gene or gene product (PP3)

Functional Evidence

Well-established in vitro or in vivo functional studies supportive of a damaging effect (PS3)

De Novo

De novo (PS2)

Lab Rule

Applied Lab Rule: BRCA1 LOF rule (version 2.1)

Conclusion

The evidence supports a classification of Likely Pathogenic.



SUPRA
MONO

2. VALIDATION AND REPRODUCIBILITY LAYER FOR CLINICAL NGS PIPELINES



- PIPELINE VALIDATION
- DRIFT DETECTION
- VERSION CONTROL
- ACCREDITATION DOCUMENTATION

„COMPLIANCE APPRECIATES AS AI PROLIFERATES.“



PIPELINE VERSION CONTROL

PIPELINE: Germline WES v2.3.4

COMMIT: a7f3c9d
Update BWA-MEM to v0.7.18
Ref: GRCh38 v109
Author: bioeng
Date: 2025-05-10

VERSIONS

- v2.3.4
- v2.3.3
- v2.3.2
- v2.3.1
- v2.2.9

DRIFT MONITORING

- Variant Concordance 99.41% ✓
- Coverage Stability 99.22% ✓
- QC Metrics Stable ✓

AUDIT TRAIL

- Input Data ✓
- Reference ✓
- Parameters ✓
- Dependencies ✓
- Environment ✓
- Output ✓

DOCUMENTATION

Machine-Readable JSON / RO-Crate

EXPORT



PIPELINE VALIDATION CHECKLIST

UNIT TESTS

- Parsers *PASS*
- Transformations *PASS*
- Variant Calling *PASS*

INTEGRATION TESTS

- Tools Interoperability *PASS*
- File Formats *PASS*

END-TO-END ACCURACY

- Truth Sets *PASS* GRCh38 v109
- Sensitivity / Specificity *PASS* 99.2% / 99.4%
- Reproducibility *PASS* 100% (3x runs)

VERSION CONTROL

- Pipeline Version v2.3.4
- Reference Genomes GRCh38 v109
- Dependencies Pinned

ACCREDITATION DOCUMENTATION

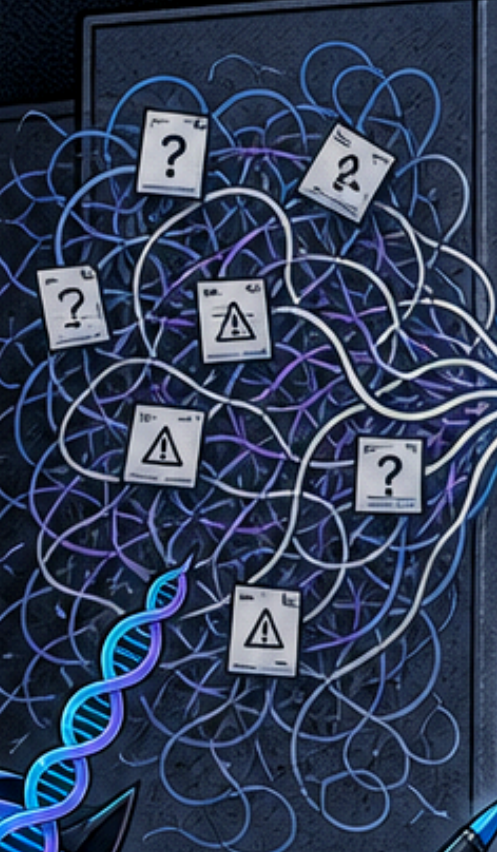
- CLIA/CAP Compliance *COMPLETE*
- ISO 15189 *COMPLETE*
- Audit Trail *COMPLETE*

VALIDATION STATUS: **VALIDATED**

3. BIOLOGY-NATIVE DATA HARMONIZATION INFRASTRUCTURE

THE BENCH PRODUCES DATA THAT NO ENTERPRISE DATA PLATFORM WAS DESIGNED TO HANDLE.

- FASTQ FILES → GZIP TXT
- BAM ALIGNMENTS → BAM BAI
- MASS SPEC OUTPUTS → RAW MGF
- ELN NOTES → PDF DOCX
- ASSAY METADATA → JSON CSV
- PLATE READER EXPORTS → XLSX



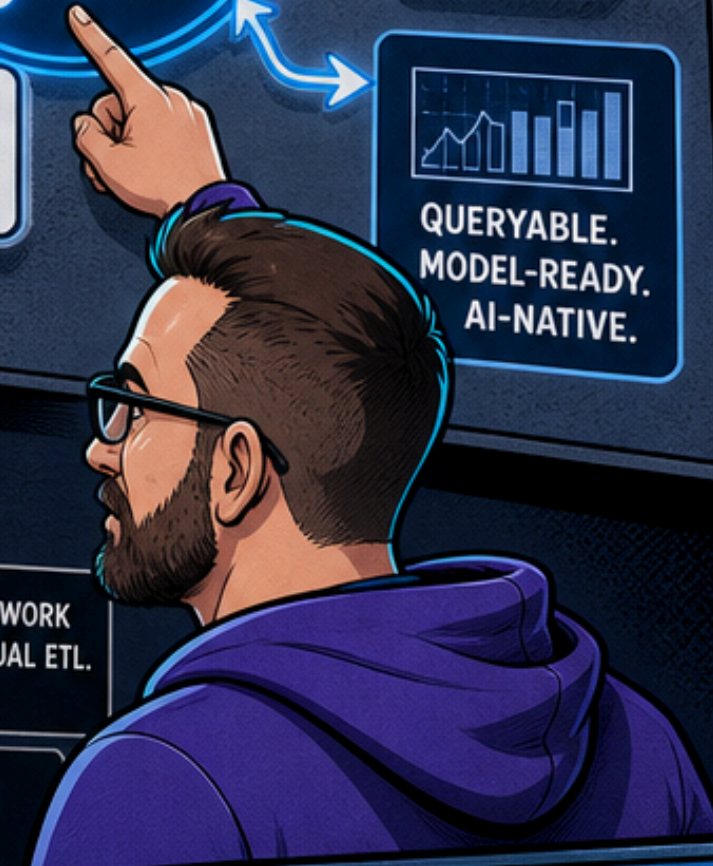
- FILE FORMATS (e.g., .fastq)
- INSTRUMENT PROVENANCE



- SAMPLE & EXPERIMENT RELATIONSHIPS

- ASSAY CONTEXT

- QUERYABLE. MODEL-READY. AI-NATIVE.



⌚ MOST BIOINFORMATICS & ML WORK STARTS WITH WEEKS OF MANUAL ETL. **CLEANING > MODELING.**

👤 WHEN A RESEARCHER LEAVES, THEIR DATA CLEANING LOGIC OFTEN LEAVES WITH THEM.

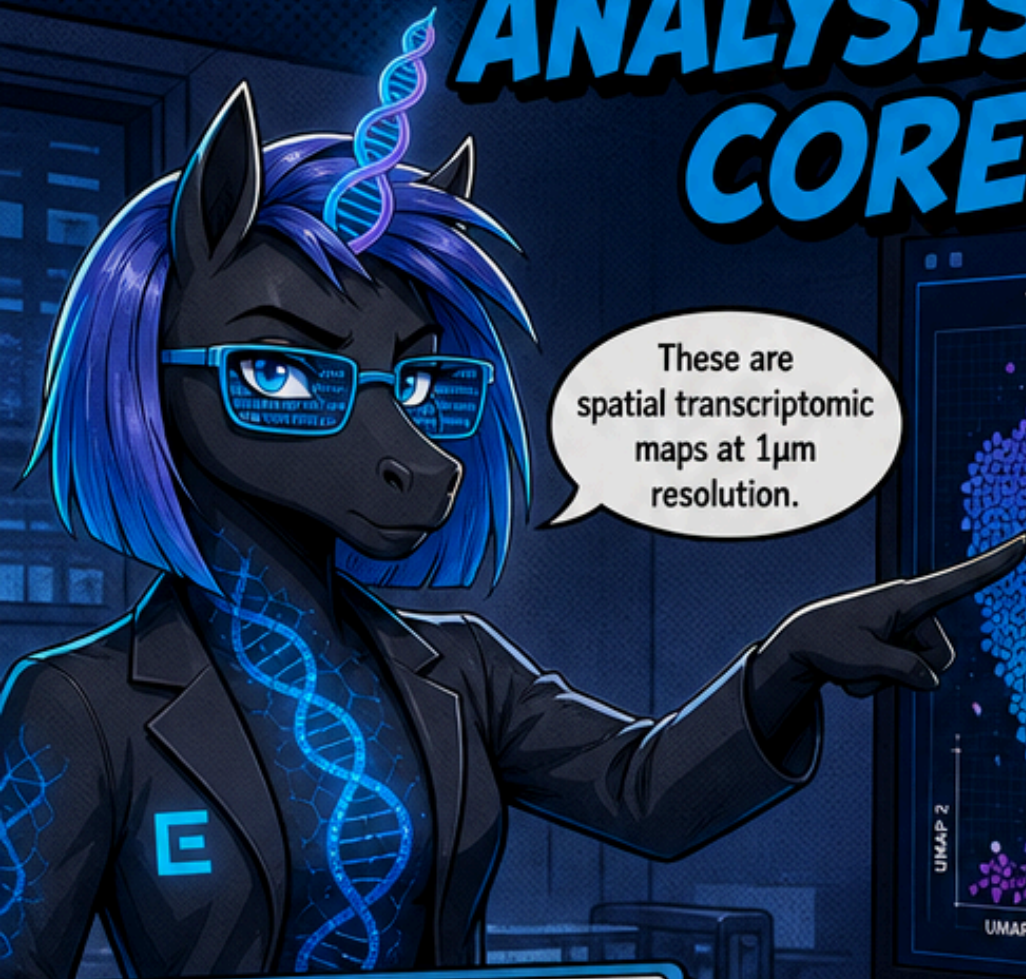
☁️ CLOUD BIOINFORMATICS IS GROWING. SCALABLE, SECURE, COLLABORATIVE ENVIRONMENTS ARE THE STANDARD. **SOURCE: 2-8, 2-9**

🛡️ THE AI-NATIVE ANSWER ISN'T A BETTER WAREHOUSE. IT'S A **BIOLOGY-LITERATE INGEST LAYER** THAT BUILDS A PER-CUSTOMER KNOWLEDGE GRAPH.

THE GENOMICS VERSION OF THE BROWNFIELD PLAY: THE HARD, UNGLAMOROUS ETL NOBODY WANTS TO DO.

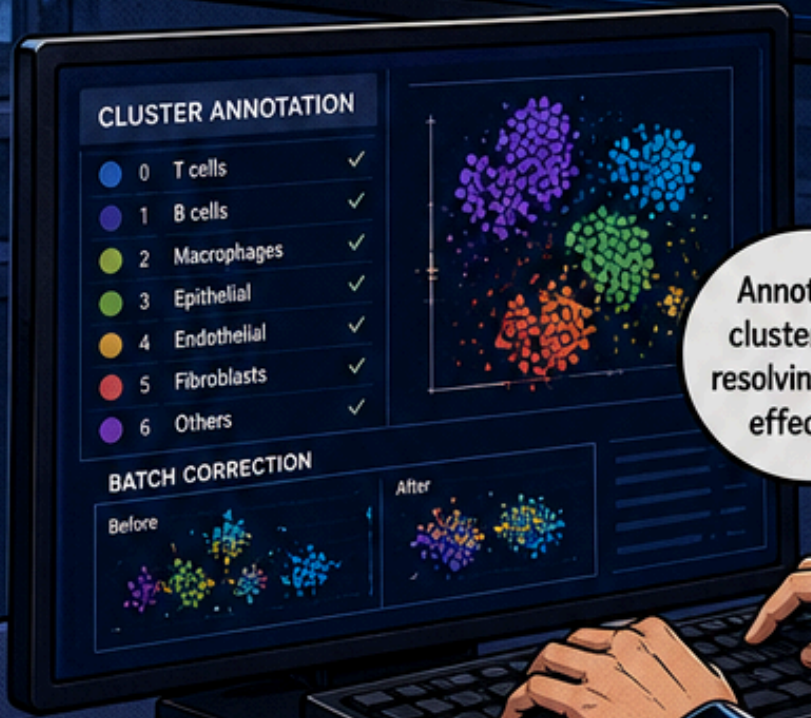
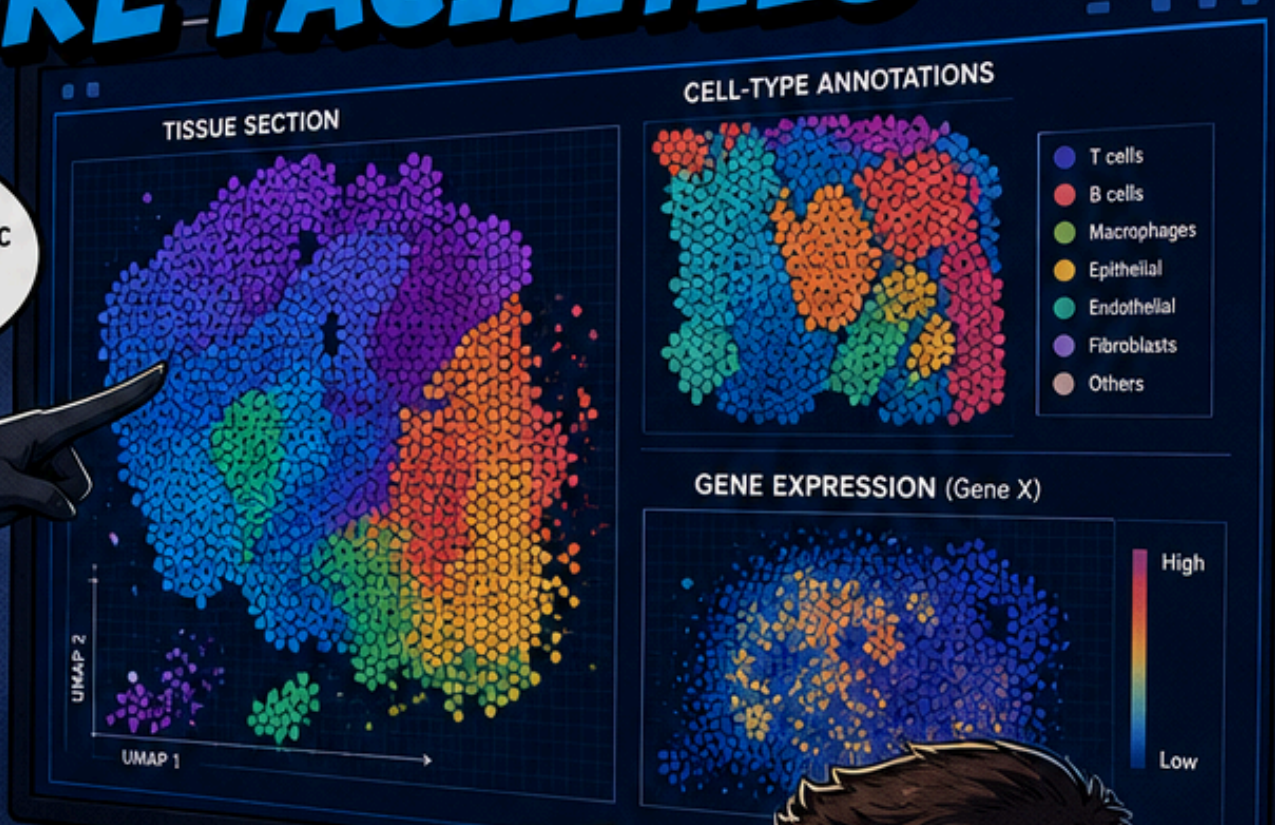
- ✓ PROPRIETARY DATA GRAPH BUILT FOR EACH CUSTOMER BECOMES THE MOAT.
- ✓ MORE DATA FLOWS IN. RICHER THE GRAPH.
- ✓ HARDER TO RECONSTRUCT ELSEWHERE.

4. SPATIAL AND SINGLE-CELL ANALYSIS COPILOT FOR CORE FACILITIES



These are spatial transcriptomic maps at 1µm resolution.

Single-cell and spatial transcriptomics are generating data **faster** than the analysis capacity **to make sense of it.**



Annotating clusters and resolving batch effects...



MASSIVE DATA INFLUX
Cell atlas initiatives and spatial platforms are outpacing analysis capacity.

CORE FACILITY BOTTLENECK
No bioinformatics bandwidth to keep up. Expert-gated tasks slow everyone down.

VENDORS IGNORE THEM
Pharma software skips core facilities. Smaller deals, ignored buyer.

FOUNDATION MODELS
scGPT, Geneformer, scFoundation enable flexible single-cell and spatial analysis.

AGENTIC COPILOT BUILT FOR CORE FACILITIES
Operator-first. Underserved. High willingness to pay. Durable moat.

5. MULTI-OMICS INTEGRATION AND INTERPRETATION FOR TRANSNATIONAL TEAMS

GENOMICS PATHWAY



PROTEOMICS PATHWAY



SHARED PATHWAY IMPLICATED
PI3K → AKT → mTOR

OUTCOME LINKED:
OVERALL SURVIVAL

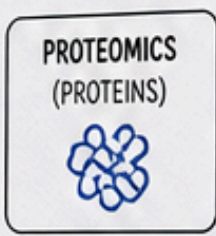
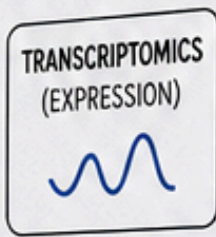


SIGNAL STRONGEST IN
SUBGROUP: PTEN LOSS

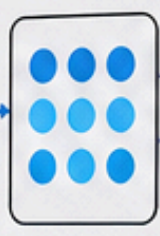
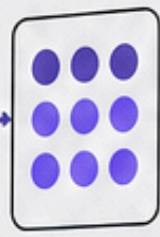


MULTI-OMICS INTEGRATION

DATA SOURCES



FEATURES

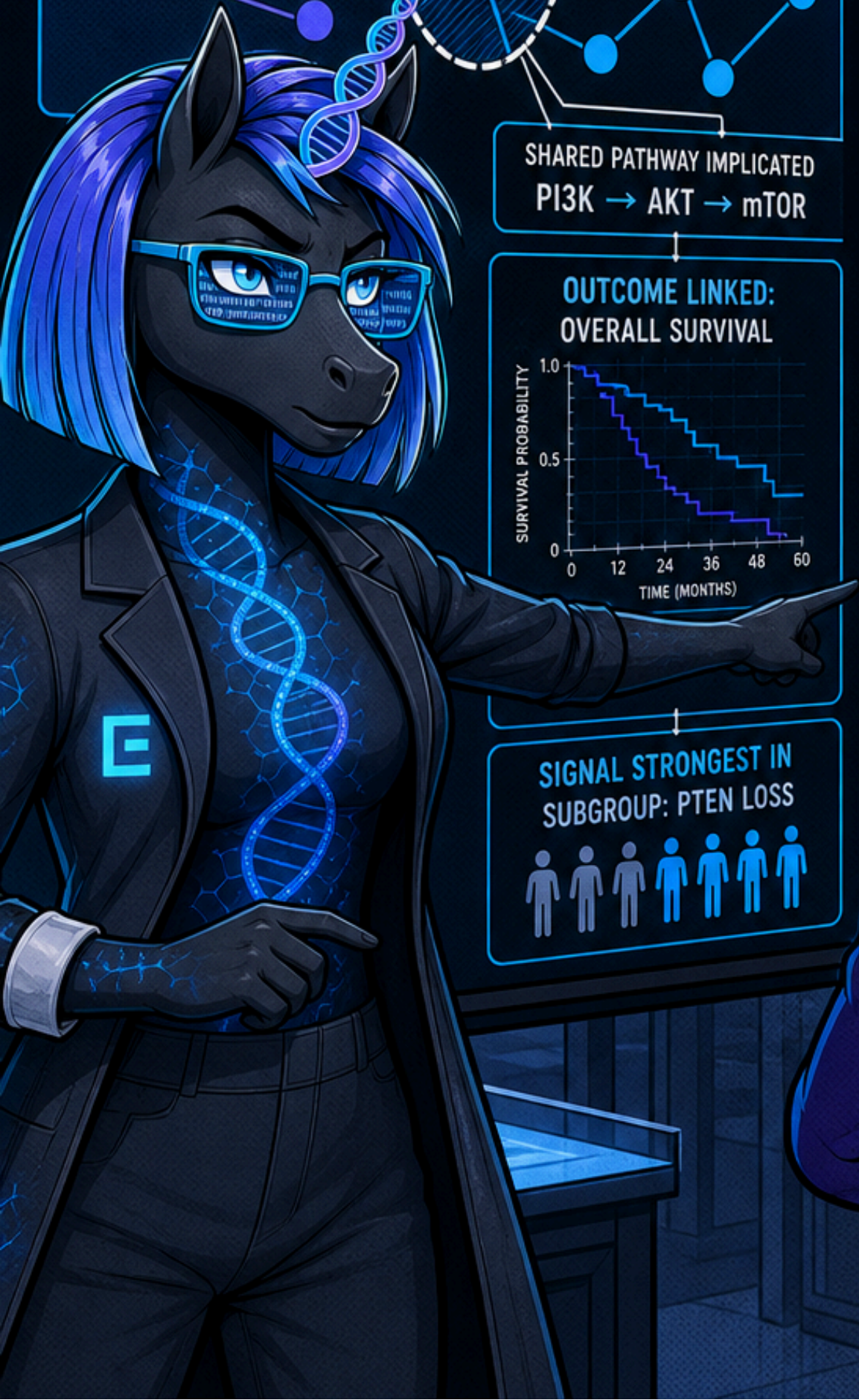


INTEGRATED INSIGHTS



INTERPRETABLE HYPOTHESES

- PATHWAY
- SUBGROUP
- COVARIATE
- OUTCOMES



6. PHARMACOGENOMICS CLINICAL DECISION SUPPORT AT THE POINT OF CARE

Genotype-guided prescribing has been clinically validated for years. The gap is **last-mile deployment**: getting the right recommendation in front of the right clinician at the moment of prescribing.



EHR

PATIENT: JAMES T.
MRN: 12345678
DOB: 07/14/1966



PHARMACOGENOMIC
PROFILE
CYP2C19 *2/*2
SLCO1B1 *5/*5
TPMT *1/*3A



**GENE-DRUG INTERACTION
HIGH RISK**

CLOPIDOGREL

CYP2C19 *2/*2 – POOR METABOLIZER
REDUCED DRUG EFFICACY

RECOMMENDATION

CONSIDER ALTERNATIVE:



TICAGRELOR
STANDARD DOSE

Guideline: CPIC 2023 | Level of Evidence: A

PHARMACY INTERFACE

NEW PRESCRIPTION

MEDICATION

TICAGRELOR ▾

DOSE

90 mg BID ▾

ROUTE

Oral ▾

DURATION

30 days ▾

✓ NO SIGNIFICANT
GENE-DRUG
INTERACTIONS
DETECTED

SUBMIT RX

GENE-DRUG
INTERACTION
ALERT.
ACTION
RECOMMENDED.

SWITCHING
TO TICAGRELOR.
GUIDELINE-CONCORDANT.
PATIENT-SAFE.

7. Regulatory Submission Copilot for Genomic and AI Diagnostics

SUBMISSION PACKAGE CHECKLIST

- Analytical Validation Studies
- Clinical Validation Package
- Software Documentation
- Post-Market Surveillance Plan
- Risk Management File
- Labeling & Instructions
- Predicate Device Comparison
- Biocompatibility / Usability
- Cybersecurity
- Quality System Documentation

REGULATORY PATHWAY

FDA 510(k)

De Novo

EU-IVDR

STATUS

- Complete
- Complete
- In Progress
- In Progress
- Pending
- Pending
- Pending
- Pending
- Pending
- Pending
- Pending

GUIDANCE REFERENCE

- FDA 510(k) Guidance
- FDA De Novo Guidance
- FDA SaMD Guidance
- FDA Postmarket Guidance
- ISO 14971
- 21 CFR 801
- FDA 510(k) Guidance
- ISO 10993 / IEC 62366
- FDA Cybersecurity Guidance
- 21 CFR Part 820



AI COPILOT: Generate → Track → Maintain
Stay aligned. Stay audit-ready. Stay ahead.

FDA GUIDANCE LIBRARY

- 510(k) Guidance [PDF](#)
- De Novo Guidance [PDF](#)
- SaMD Guidance [PDF](#)
- Postmarket Guidance [PDF](#)
- AI/ML Action Plan [PDF](#)
- Cybersecurity Guidance [PDF](#)

FDA 510(k) GUIDANCE

General Principles

Device Description

- SUBMISSION TEMPLATES
- PREDICATE DATABASE
- APPROVAL TRACK RECORDS
- REGULATORY INTELLIGENCE

- Traceability
- Validation Evidence
- Transparency
- Safety
- Effectiveness

8. APPLIED METAGENOMICS FOR CLINICAL MICRO, AMR, AND SURVEILLANCE



CLINICAL AMR REPORT METAGENOMIC SEQUENCING ANALYSIS

SAMPLE SUMMARY

Sample ID: CLIN-24-1187
Sample Type: BAL
Collection Date: 2025-05-07
Sequencing Depth: 45.2 M reads

DETECTED PATHOGENS

Organism	Relative Abundance
<i>Klebsiella pneumoniae</i>	32.7%
<i>Pseudomonas aeruginosa</i>	18.4%
<i>Acinetobacter baumannii</i>	11.2%
<i>Enterococcus faecium</i>	4.3%
<i>Staphylococcus aureus</i>	2.1%

MICROBIAL PROFILE

- 32.7%
- 18.4%
- 11.2%
- 4.3%
- 2.1%
- 31.3% Other

ANTIMICROBIAL RESISTANCE PREDICTIONS

Gene / Mechanism	Associated Drug Class	Prediction	Confidence
blaKPC-2	Carbapenems	RESISTANT	High
blaTEM-1B	Penicillins	RESISTANT	High
aac(6)-Ib-cr	Aminoglycosides	RESISTANT	Moderate
oqxAB	Fluoroquinolones	RESISTANT	Moderate
mecA	Beta-lactams (MRSA)	NOT DETECTED	High

THERAPY INSIGHTS

Active options predicted:

- ✓ Ceftazidime-avibactam
- ✓ Polymyxin B
- ✓ Tigecycline

⚠ High-risk AMR detected. Confirm with phenotypic testing.

REFERENCE DATABASE

Clinical Bacteria | AMR Genes | Virulence Factors | Plasmids | Foodborne Pathogens

Klebsiella pneumoniae (Clinical Isolate)

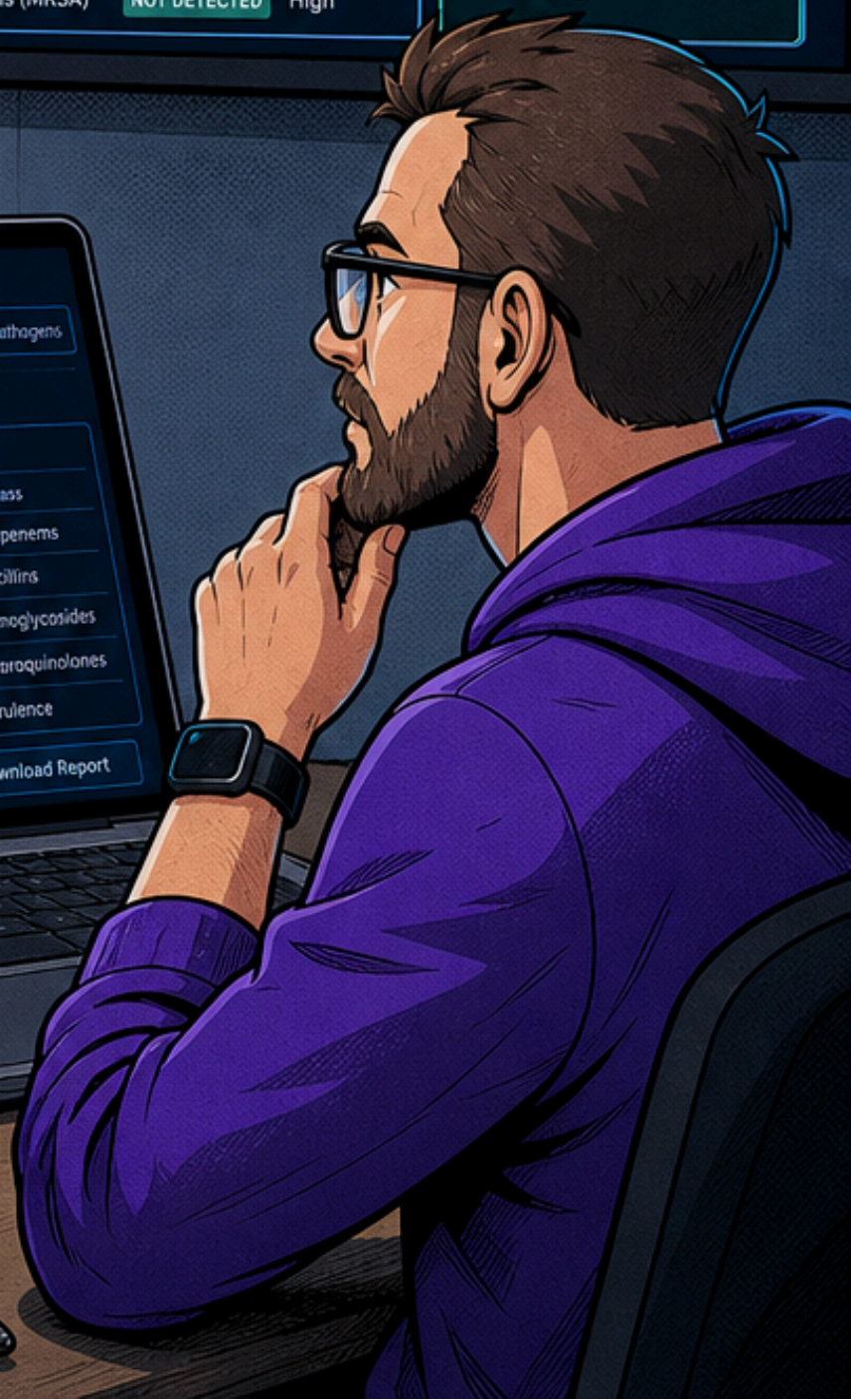
OVERVIEW

Strain ID: KP-CLIN-00571
Source: Blood
Geography: Global
Isolation Year: 2023
Taxonomy: *K. pneumoniae sensu stricto*

ANNOTATED GENES

Gene	Function	AMR Class
blaKPC-2	Carbapenemase	Carbapenems
blaTEM-1B	Beta-lactamase	Penicillins
aac(6)-Ib-cr	Aminoglycoside acetyltransferase	Aminoglycosides
oqxA	Efflux pump	Fluoroquinolones
rmpA	Regulator of mucoidy	Virulence

Alignment | Phylogeny | Resistance Profile | Download Report



APPLIED METAGENOMICS


- ✓ CLINICAL IMPACT
- ✓ REGULATORY READY
- ✓ PURPOSE-BUILT DBS
- ✓ REAL-WORLD BUYERS

9. IMMUNOINFORMATICS TOOLING FOR THE LONG TAIL

The immunoinformatics space has a specific structural problem: the well-funded players are building **therapeutics, not tools**.

They're using their neoantigen prediction models, TCR/BCR repertoire analysis pipelines, and vaccine design platforms internally, on their own pipeline, and they have **no commercial incentive** to productise them for external users.

That leaves academic immunology labs, small biotechs, and contract research organisations using a patchwork of open-source tools, custom scripts, and occasional consulting engagements. None of these buyers are poorly motivated. They're **just poorly served**.

 Investments support the development of advanced bioinformatics tools for drug discovery, personalized medicine, and disease modeling²⁻⁴, and immunotherapy is one of the faster-growing areas within that.

A focused tooling platform for neoantigen prediction, repertoire analysis, and vaccine design, targeted at academic and small-biotech teams, doesn't need to compete with Genentech's internal tools. It needs to be **substantially better** than what this buyer currently uses.

The moat is **proprietary immune datasets** accumulated through customer engagements and the depth of the domain knowledge encoded in the workflows.

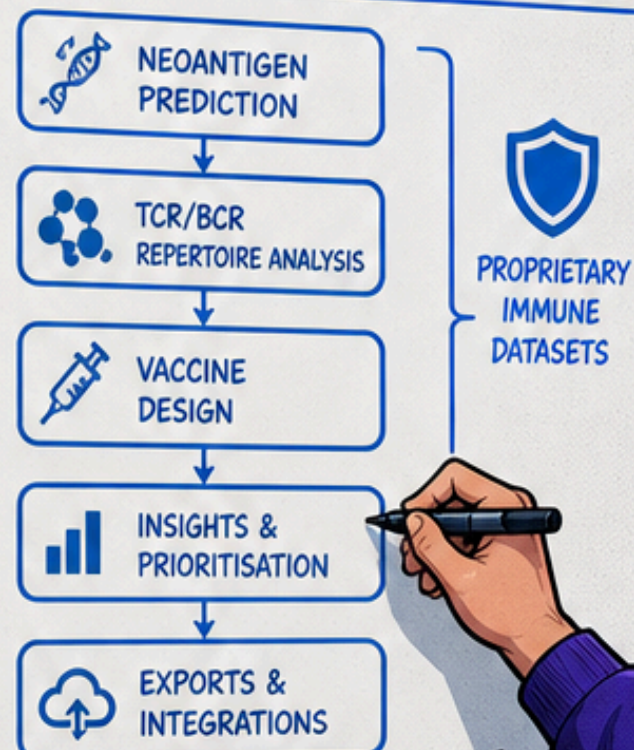
These are not general ML problems. They require a team that understands HLA binding, clonotype analysis, and immunodominance, and that **domain depth is genuinely hard to replicate**.

**PATCHWORK TOOLS.
DISCONNECTED.
HARD TO REPRODUCE.
HARDER TO TRUST.**

TODAY: FRAGMENTED & MANUAL



TOMORROW: UNIFIED & PURPOSE-BUILT



**NOT A MOTIVATION
PROBLEM.
A TOOLING GAP.**

**ONE PLATFORM.
ONE WORKFLOW.
DEEP IMMUNOLOGY.
REAL ADVANTAGE.**

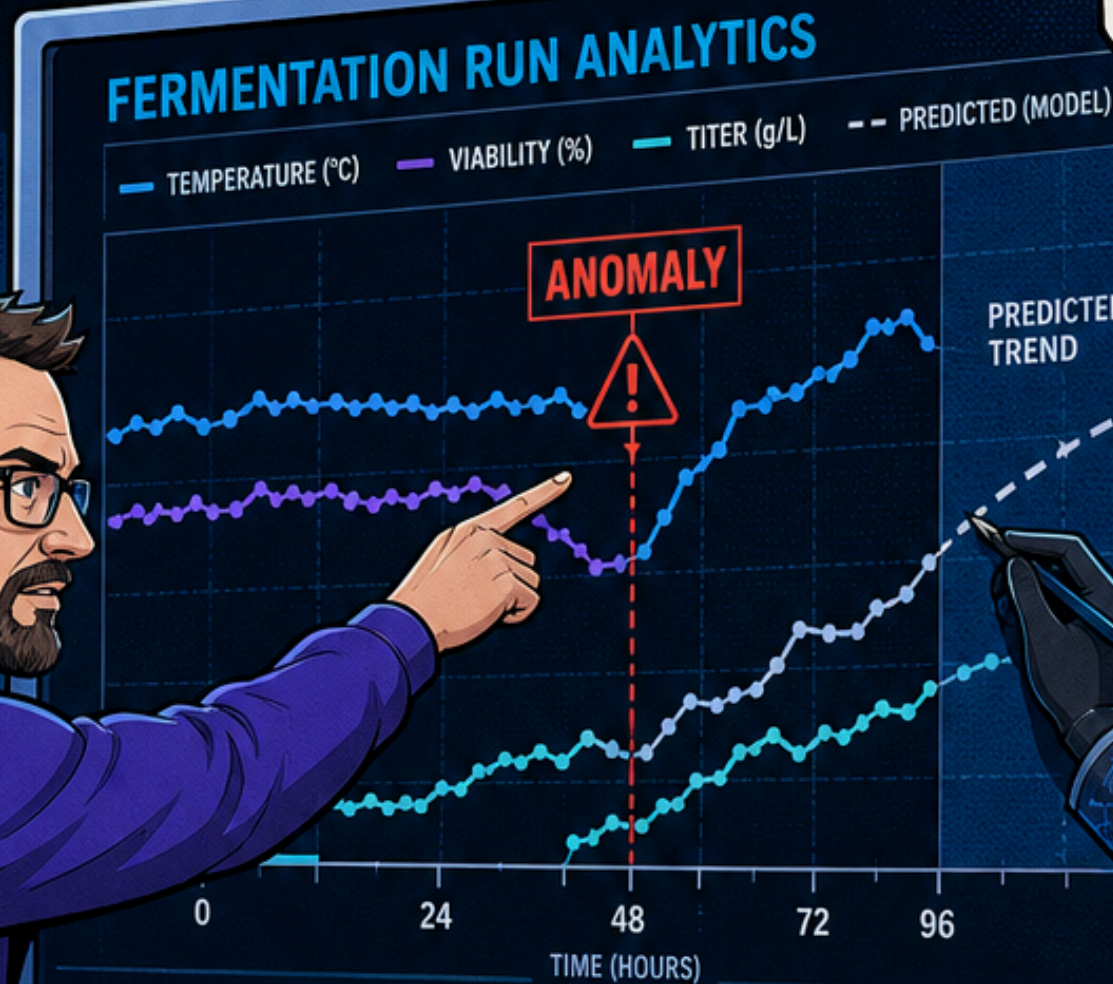
OUR EDGE:

- ✓ IMMUNE DOMAIN EXPERTISE
- ✓ PROPRIETARY DATA
- ✓ DEPTH IN WORKFLOWS
- HARD TO REPLICATE**

10. BIOPROCESS OPTIMIZATION FOR CDMOS AND BIOLOGICS MANUFACTURING

Using your facility's own run data to build predictive models that cut development timelines, reduce batch failures, and flag process drift before it causes a deviation.

ANOMALY DETECTED.
TEMP SPIKE CAUSED YIELD DEVIATION IN RUN 9F37.



OUR MODEL PREDICTED THE DRIFT 6 HOURS BEFORE IT CROSSED THE DEVIATION LIMIT.

RUN SUMMARY

TOTAL RUNS 1,248
BATCH FAILURES 7.3% ↓
AVG. DEVELOPMENT TIME -18.6% ↓
PROCESS DRIFT ALERTS 23 ↑

KEY PROCESS PARAMETERS

TEMP.  
pH  
DO  
AGITATION  

PREDICTIVE INSIGHTS

- ✓ EARLY DRIFT DETECTION
- ✓ OPTIMAL FEED STRATEGY
- ✓ YIELD IMPROVEMENT
- ✓ RISK OF DEVIATION: LOW



YOUR DATA. YOUR MODELS. BETTER OUTCOMES. LESS RISK.

WHAT THESE TEN IDEAS HAVE IN COMMON

*Data compounds.
Moats widen.*

*Regulation
appreciates.
Moats deepen.*



DATA MOATS

- proprietary data
- compounds over time
- harder to replicate



AI + REGULATORY INTERSECTION

- compliance creates real barriers
- switching is painful

**DEFENSIBLE
SCALABLE
DURABLE
IMPACTFUL**



TECHNICAL READINESS

- AI advancing fast
- foundation models maturing
- enterprise adoption accelerating

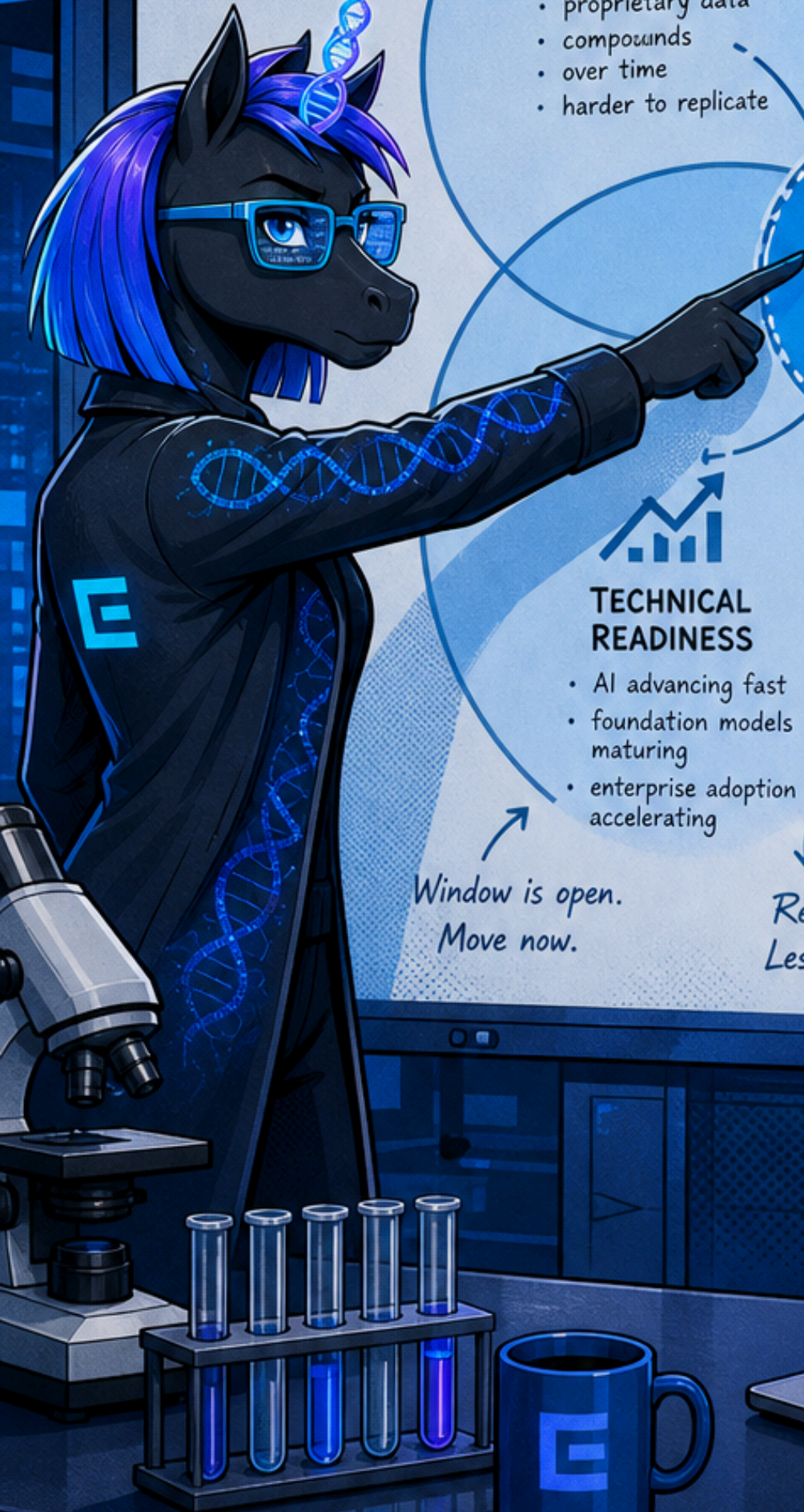


UNDERSERVED SEGMENTS

- academic labs
- mid-tier CDMOs
- hospital micro labs
- small biotech teams

*Window is open.
Move now.*

*Real needs.
Less competition.*



EVOTRON STUDIO
AGENCY AGENT SYSTEM

```
vo = {  
  "agentic_workhorse",  
  "always_on",  
  ts: 12,  
  ise: "invisible tech",  
  it: "visible results"
```

```
ship(value) {  
  ders.focus();  
  execute();  
  n results.delivered();
```

TYPE. JUST OUTCOMES.
EVOTRON STUDIO



<https://evotronstudio.co.nz>