Getting Started with AI in Molecular and Genomic Pathology

Path-iQ Getting Started Series · No. 17 of 20 · A practical primer for the practising pathologist

July 2026

1. The subspecialty that has used machine learning longest without calling it that

Molecular pathology has been running machine learning in production for years. Variant callers use statistical models. Copy-number analysis uses segmentation algorithms. Read alignment, base calling on nanopore platforms, and structural variant detection are all learned models. The bioinformatics pipeline in every molecular laboratory is, in substantial part, machine learning infrastructure that predates the current AI conversation.

What has changed is the layer above: interpretation. Calling a variant is largely solved. Deciding what it means is not, and that is where the current wave of tools is landing.

2. Variant interpretation: the real bottleneck

The ACMG/AMP framework for sequence variant classification assigns evidence codes and combines them into a five-tier classification. It is a rule-based expert system, and it was designed to bring consistency to a process that was previously ad hoc. It has improved reproducibility substantially, but interlaboratory discordance persists — particularly in the assignment of variants of uncertain significance, which remain the dominant practical problem in clinical genomics.

Machine learning enters here in several ways.

In silico pathogenicity prediction. Computational predictors of variant effect contribute a specific evidence code within the ACMG framework. The current generation of deep learning predictors — including protein language models and structure-based methods trained on evolutionary and structural data — substantially outperforms earlier tools, and calibration studies have prompted revision of how much evidential weight computational predictions should carry. This is the most direct and best-established application.

Literature and database curation. Natural language processing to extract variant-phenotype evidence from published literature. Curation is the rate-limiting step in variant classification, and it scales poorly with human effort.

Splicing prediction. Deep learning models predicting splice-altering effects from sequence have materially improved on earlier approaches and now inform routine classification of intronic and synonymous variants.

Automated draft classification. Systems that assemble evidence codes and propose a classification for scientist review. This is the workflow application with the largest efficiency potential and the one requiring the most careful governance.

3. Cytogenetics and genomic profiling

Karyotype analysis. Metaphase chromosome analysis is a visual pattern-recognition task performed by highly trained cytogeneticists in shrinking supply. Automated metaphase finding and karyotype assembly are established technologies, with the human confirming rather than performing the initial classification.

FISH signal enumeration. Counting fluorescent signals across hundreds of nuclei is repetitive, quantitative, and subject to fatigue. Automated enumeration is a natural fit.

Copy-number and optical genome mapping interpretation. Genome-wide copy-number analysis produces complex profiles requiring pattern interpretation, and machine learning approaches to classify these are advancing.

Methylation-based tumour classification. The most clinically consequential machine learning application in molecular pathology. Random-forest classification of DNA methylation profiles now underpins CNS tumour diagnosis worldwide — the Heidelberg classifier v12.8, trained on 7,495 profiles, recognises 184 subclasses at 95% subclass-level accuracy with well-calibrated probability estimates. The same approach has been extended to sarcoma, meningioma, sinonasal tumours and intrahepatic bile duct cancer, demonstrating that the method generalises well beyond its origin.

Morphology-to-molecular inference. The reverse direction: predicting molecular status from H&E. Covered throughout this series; from the molecular pathologist's perspective the appropriate framing is triage — deciding which test to order first — not replacement.

4. Your first 90 days

Weeks 1–3: audit your VUS burden and your turnaround. How many variants per case are classified as uncertain, and how much scientist time goes into classification? This defines where automation would help.

Weeks 4–8: evaluate computational predictor calibration in your own pipeline. Which predictors do you use, at what thresholds, and have you reviewed them against current calibration guidance? This costs nothing and frequently reveals that thresholds in use are inherited rather than justified.

Weeks 9–12: pilot automated evidence assembly, not automated classification. A system that gathers population frequency, computational predictions, literature references and prior classifications into a structured summary for the scientist saves substantial time while leaving the classification decision entirely human. This is the highest-value, lowest-risk entry point in the subspecialty.

5. Validation specifics

Reference set composition determines classifier boundaries. A methylation classifier can only assign classes present in its reference set, and version differences change class structure. Always record classifier version in the report.

Population representation is a first-order concern in variant interpretation. Allele frequency thresholds depend on the reference population, and ancestries under-represented in genomic databases generate excess variants of uncertain significance. This is a documented health equity problem, not a technical footnote.

Circularity is a real methodological trap. Computational predictors trained on ClinVar classifications and then used as evidence for ClinVar-style classification create feedback loops. Understand what your predictors were trained on.

Tumour purity, ploidy and specimen quality constrain every downstream molecular inference regardless of algorithm quality.

6. Failure modes to anticipate

Over-weighting computational evidence. The ACMG framework assigns computational prediction limited weight for good reason. Tools that let a strong in silico prediction drive a classification are misusing the framework.

Automation of the wrong step. Automating classification is tempting and risky; automating evidence gathering is safe and delivers most of the benefit.

Silent reference database updates. Population frequency databases and clinical variant databases update continuously, so a variant's classification can change without anyone re-examining the case. Systematic reanalysis policies matter more than any single algorithm.

Classifier over-confidence at low purity. Methylation classifiers on low-purity specimens can return plausible but unreliable calls. Respect the calibrated confidence threshold.

7. What to watch next

Large language models applied to variant curation and report generation are arriving quickly, and molecular pathology is unusually exposed because its output is structured text derived from literature — precisely what these systems handle well and precisely where fabricated citations would be most damaging. Governance here should precede deployment. On the classification side, expect methylation-based and multi-omic classifiers to continue extending into new tumour families, and expect professional bodies to issue guidance on how machine-learning-derived classifications should be worded, versioned and caveated in diagnostic reports.

Vendor due-diligence checklist

Use this before any procurement conversation. The first seven questions apply to every AI tool in pathology; the last three are specific to this subspecialty. A vendor unable or unwilling to answer them in writing has told you something important.

Applies to any pathology AI tool

What is the intended use statement, and what is explicitly excluded from it?

What is the regulatory status in my jurisdiction — cleared, CE-marked, research use only?

Can I see performance on an independent external cohort, not only internal validation?

What is the prevalence of the target condition in the validation cohort versus my practice?

What is the failure mode when the model is uncertain — does it abstain, or always answer?

What monitoring does the vendor provide for performance drift after deployment?

Who is liable if an algorithmic output contributes to a diagnostic error?

Specific to this subspecialty

What data were the computational predictors trained on, and is there circularity with the classification databases I use?

How is reference database updating handled, and what triggers systematic reanalysis?

What is the ancestry composition of the reference population underlying frequency thresholds?

Ask for answers in writing, and keep them. When the tool underperforms in your laboratory — and at some point it will — the gap between what was claimed and what was validated is where the conversation starts.

8. Key references

#

Source

Note

1

Richards et al., ACMG/AMP variant classification standards

Foundational interpretation framework

2

Heidelberg CNS Tumour Methylation Classifier v12.8, *Cancer Cell*, 2025

7,495 profiles, 184 subclasses, Brier 0.028

3

Capper et al., *Nature*, 2018

Original methylation classification approach

4

Koelsche et al.; Jurmeister et al., *Nat Commun*, 2022;13:7148

Extension to sarcoma and sinonasal tumours

5

Calibration of computational variant effect predictors

Evidence weighting within ACMG framework

Glossary of terms used in this primer

Whole-slide image (WSI) — a digitised glass slide, typically gigapixel in scale, produced by a scanner and viewed on screen rather than under a microscope.

Foundation model — a large model pretrained on vast unlabelled image data to learn general-purpose visual features, then adapted to specific tasks with comparatively little labelled data.

Weakly supervised learning — training where only a slide-level label is available (for example “contains cancer”) rather than region-by-region annotation. Most clinical pathology AI is built this way, because exhaustive annotation is unaffordable.

AUROC — area under the receiver operating characteristic curve, a single figure summarising discrimination across all thresholds. 0.5 is chance, 1.0 is perfect. It says nothing about calibration or about performance at the threshold you would actually use.

Calibration — whether a model’s stated probability matches observed frequency. A well-calibrated model that says “80% confident” is right about 80% of the time. Calibration matters more than accuracy when a clinician acts on the number.

Abstention — the model declining to answer when confidence falls below a threshold, rather than returning its best guess. Essential wherever rare entities are in play.

Batch effect — systematic differences arising from scanner, stain lot, laboratory or site rather than from biology. The leading cause of models performing worse elsewhere than where they were built.

Automation bias — the human tendency to defer to a system that is usually correct, and so to stop checking independently. A property of the workflow, not of the algorithm.

Predetermined Change Control Plan (PCCP) — a regulator-agreed specification of changes a manufacturer may make to an authorised device without a new submission. Its existence acknowledges that these systems are not static.

Shadow deployment — running an algorithm in parallel with routine practice, without acting on its output, to characterise its behaviour on your own material before it touches a live workflow.

Path-iQ — Global Pathology AI Review. Prepared for practising pathologists beginning AI adoption.