Bridging genomics and tissue structure with AI/ML

Laura Acqualagna, Director of AI/ML Engineering, AI/ML R&D, GSK

Exploring Applications of AI in Genomics and Precision Health: A Workshop
October 28, 2025

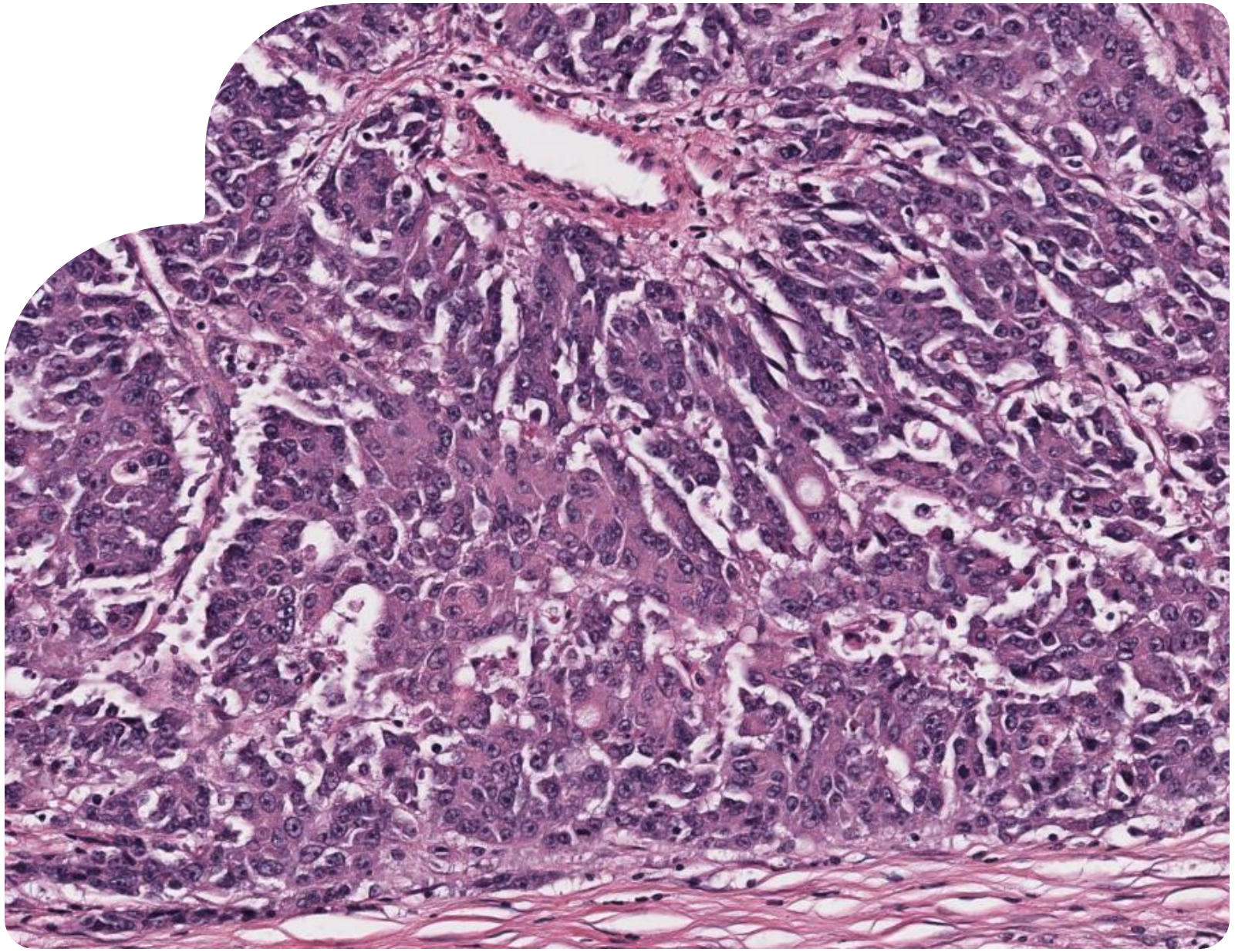
Disclosure

Laura Acqualagna is a full-time employee at GSK.

This is a histology image from a tumour resection of a patient with colorectal cancer. The pathologist can see the tumour architecture, cell morphology, and immune infiltration.

But the critical question the oncologist needs answered is: *Does this tumour have a KRAS mutation? Will it respond to anti-EGFR therapies?*

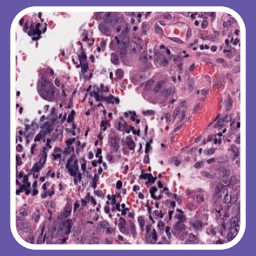
To answer this today, we need expensive molecular testing, additional tissue, and 2-3 weeks of waiting. **What if we could answer this question directly from this H&E slide?**



Images shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Bridging Traditional Pathology and Molecular Oncology

What Pathologists See (H&E):



- Tissue architecture
- Cell morphology
- Nuclear features
- Immune infiltration
- Stromal patterns

What Oncologists Need (Molecular Data):

- Driver mutations (KRAS, TP53, EGFR, etc.)
- Gene expression signatures
- Pathway activation status
- Predictive biomarkers for therapy
- Prognostic indicators

The Problem:

Cost:

- NGS panel: $1,000-$5,000 per patient
- RNA-seq: $500-$1,500
- H&E staining: $50-$100

Time:

- Molecular testing: 2-4 weeks turnaround
- H&E diagnosis: 2-3 days
- Treatment decisions delayed

Tissue Availability:

- Small biopsies may be exhausted by routine testing
- Molecular tests are destructive
- Can't perform comprehensive profiling on limited samples

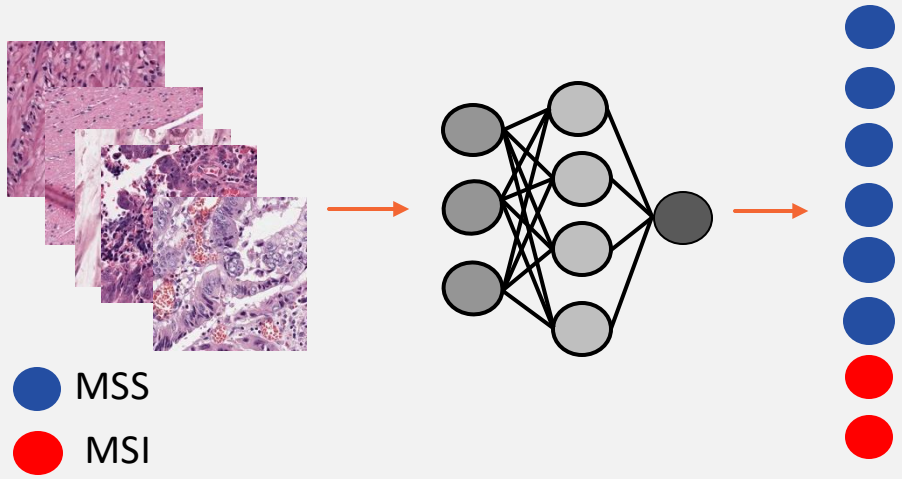
Malapelle, U., et al (2025). Costs of biomarker testing in advanced non-small cell lung cancer: a global study comparing next-generation sequencing and single-gene testing. *The journal of pathology. Clinical research*, *11*(2), e70018. https://doi.org/10.1002/2056-4538.70018
Fleming, K. E., et al (2024). Biomarker Turnaround Times and Impact on Treatment Decisions in Patients with Advanced Non-Small Cell Lung Carcinoma at a Large Canadian Community Hospital with an Affiliated Regional Cancer Centre. *Current oncology (Toronto, Ont.)*, *31*(3), 1515–1528. https://doi.org/10.3390/curroncol31030115

The Hidden Code: What If H&E Already Contains the Answers?

- **Slide-level molecular predictions** - Identify mutations, subtypes, and biomarkers from whole slides
- **Cost-effective screening** - Flag patients who need confirmatory molecular testing at a fraction of the cost
- **Emerging spatial methods** - New approaches predict gene expression patterns at cellular resolution (spatial transcriptomics from H&E)

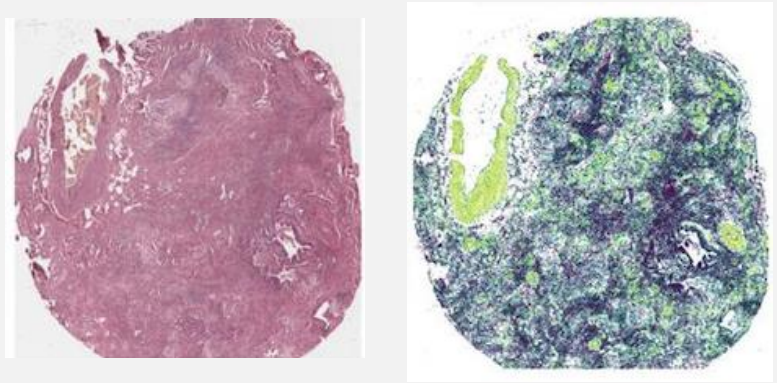
Slide-Level Predictions *Clinical Translation Path*

E.g. ”MSI mutated (F1 score: 0.78)"



- ✓ Predicts aggregate molecular features for entire tumour
- ✓ Validated in multiple studies
- ✓ AUCs 0.7-0.9 across cancer types
- ✓ Ready for screening applications

Spatial Omics Prediction *Research Frontier (2024-2025)*



- ✓ Predicts gene expression at cellular/spatial resolution
- ✓ Generates spatially-resolved molecular maps.
- ✓ **Examples:**
  - Gene expression heatmaps across tissue
  - Cell-type specific transcriptional states
  - Tumour microenvironment mapping
  - Regional heterogeneity analysis

Nonchev, Kalin, et al. "DeepSpot: Leveraging Spatial Context for Enhanced Spatial Transcriptomics Prediction from H&E Images." *medRxiv* (2025): 2025-02.

Multiple-Instance Learning for H&E WSI level prediction of genetic signatures

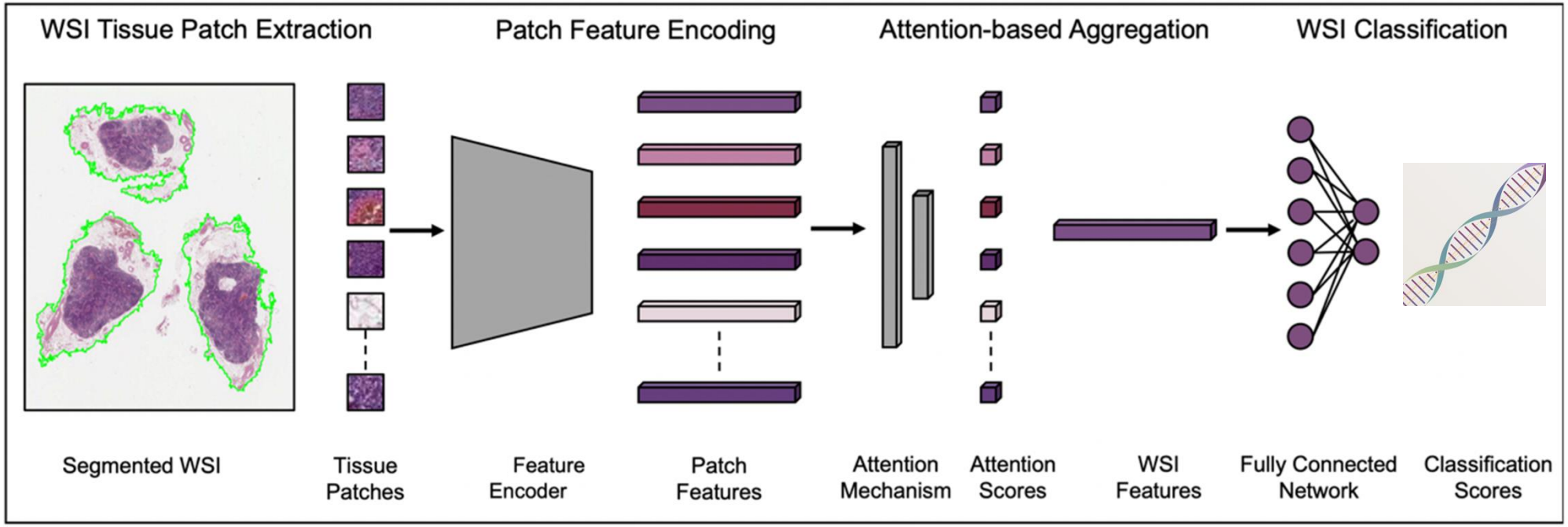


Figure adapted from: Allen, K.E.; Breen, J.; Hall, G.; Mappa, G.; Zucker, K.; Ravikumar, N.; Orsi, N.M. Multiple Instance Learning for the Detection of Lymph Node and Omental Metastases in Carcinoma of the Ovaries, Fallopian Tubes and Peritoneum. *Cancers* **2025**, *17*, 1789. https://doi.org/10.3390/cancers17111789

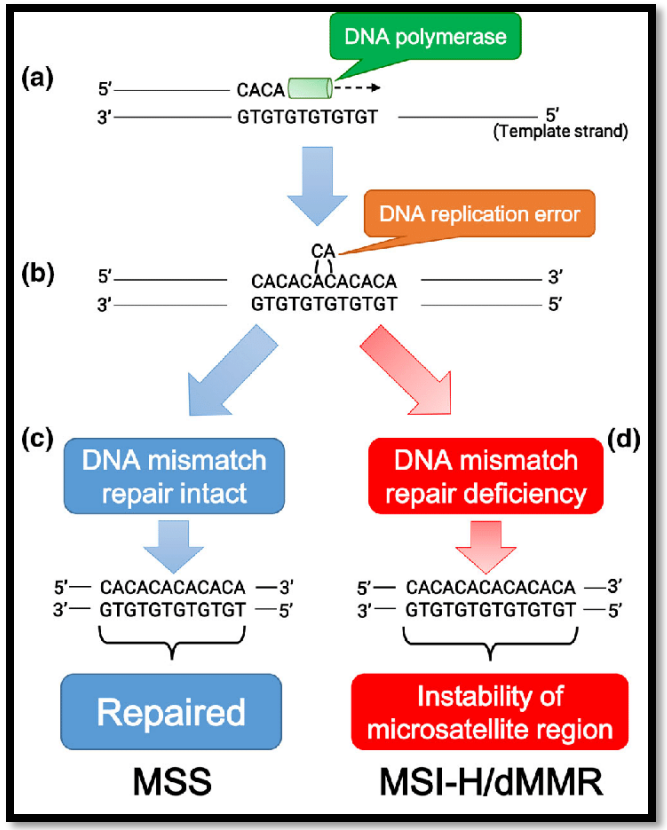
Interpretability: extract and quantify heatmaps indicative of areas attended by the model, by overlaying cells and tissue characterization

Use case – Colorectal Cancer (CRC)

- Immunotherapies are approved for unresectable or metastatic MSI-H/dMMR CRC.
- **3.5% to 5%** of metastatic CRC patients have dMMR/MSI-H tumors
- AUC for MSI-H/MSS prediction between **0.8 to 0.95**, with sensitivity within the 70–95% range[1]

- Some **MSS** patients CRC may respond to immunotherapy, particularly those with specific genomic mutations or high levels of immune cell infiltration, and often when immunotherapy is used in combination with other treatments

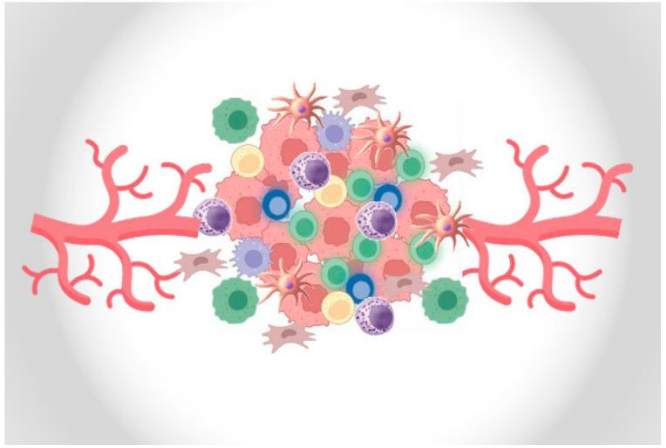
- Identify alternative genetics/omics signatures indicative of response and bridge those to tissue structure in H&E



Tumor-Immune Interactions

Immune State

Immune cell composition



Immune Infiltration & Engagement

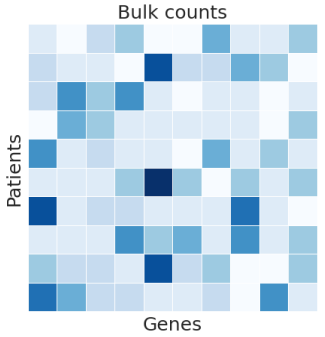
Tumor State

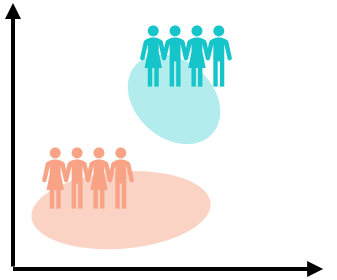
Mutational burden (TMB)

Micro-satellite Stability

PDL1 expression

etc..



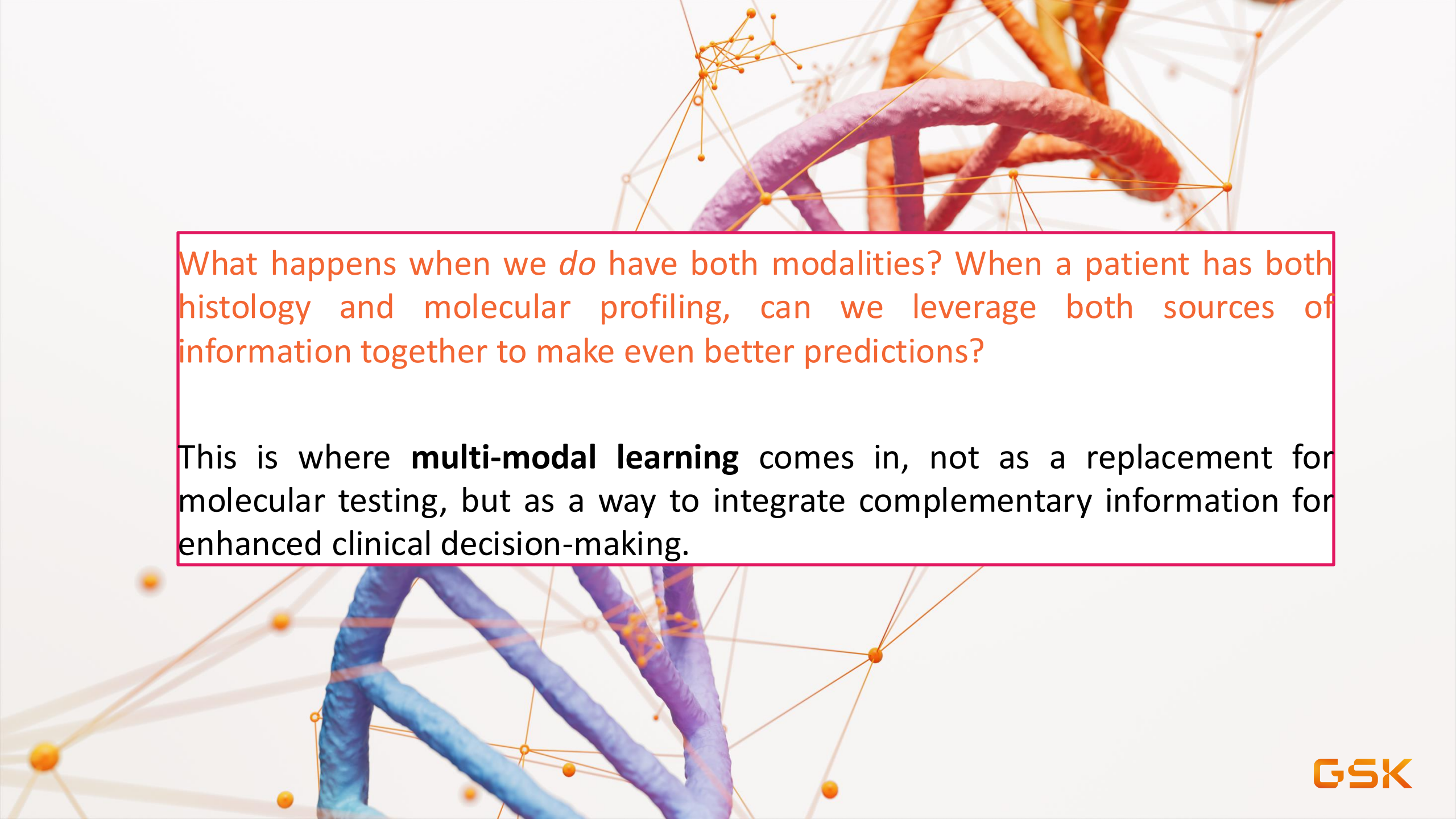


Patient Response & Survival

Fig2: Eso, Yuji, et al. "Microsatellite instability and immune checkpoint inhibitors: toward precision medicine against gastrointestinal and hepatobiliary cancers." *Journal of gastroenterology* 55.1 (2020): 15-26.

Fig 1: Jeong, S.-R.; Kang, M. Exploring Tumor–Immune Interactions in Co-Culture Models of T Cells and Tumor Organoids Derived from Patients. *Int. J. Mol. Sci.* **2023**, *24*, 14609. https://doi.org/10.3390/ijms241914609

[1] Li, H., et al. (2025). Systematic review and meta-analysis of deep learning for MSI-H in colorectal cancer whole slide images. *NPJ digital medicine*, *8*(1), 456. https://doi.org/10.1038/s41746-025-01848-z



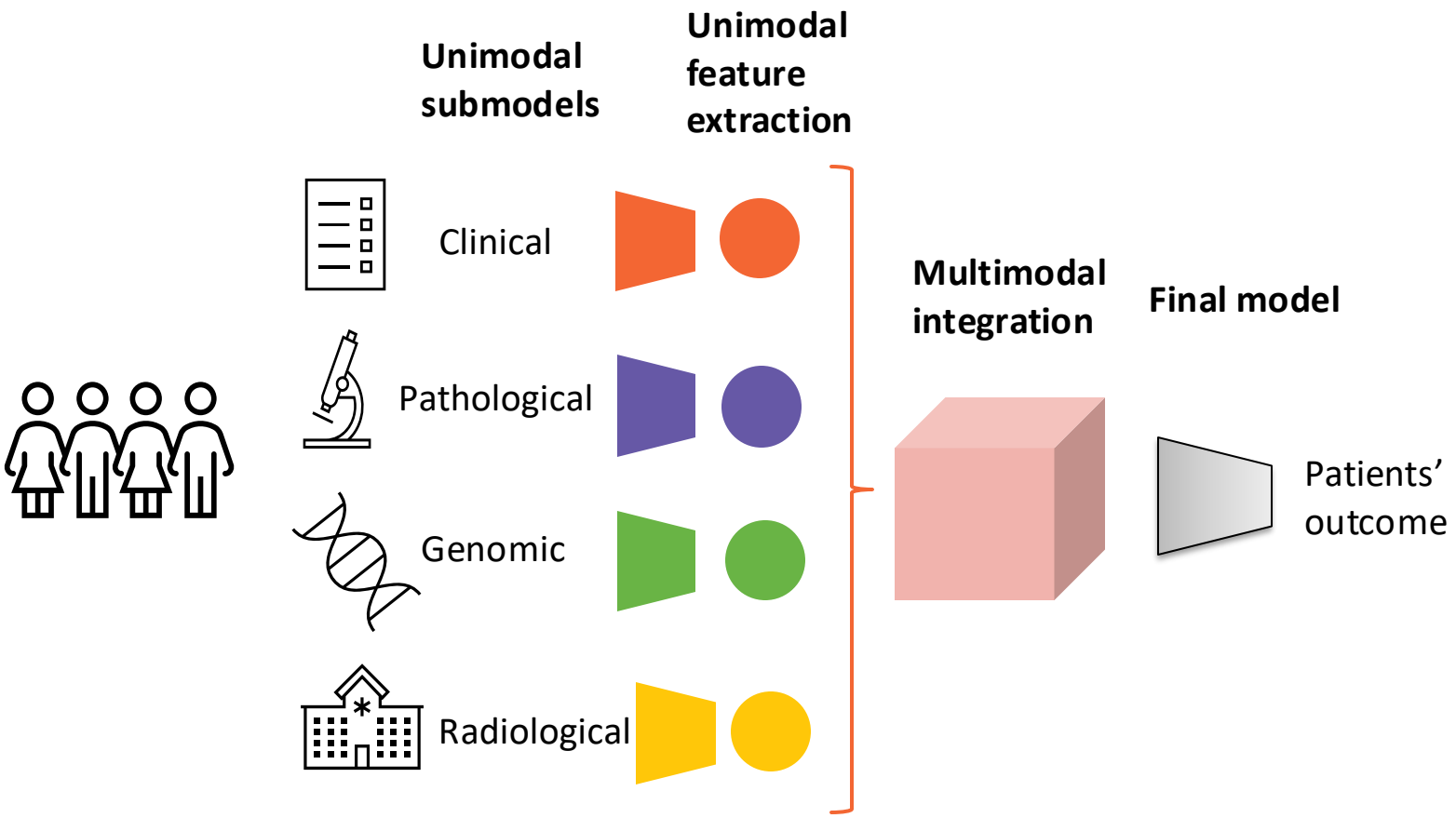
What happens when we *do* have both modalities? When a patient has both histology and molecular profiling, can we leverage both sources of information together to make even better predictions?

This is where **multi-modal learning** comes in, not as a replacement for molecular testing, but as a way to integrate complementary information for enhanced clinical decision-making.

GSK

AI/ML for maximizing the utility of multimodal data

- Unimodal models are the building blocks for Multimodal models
- Multimodal models integrate features across modalities and borrow strengths/ infer correlation across different modalities
- Foundation models contribute by providing rich representations of unimodal or multimodal data



Integration strategies:

- Late fusion
- Early fusion
- Intermediate fusion

Challenges in AI and multi-modal learning for translational medicine

Data integration

Harmonizing diverse data sources (e.g. genomics, imaging, EHRs) and ensuring interoperability

Lack of standardization

Working with incomplete or biased datasets

Model validation

Importance of validating AI/ML models using diverse, high-quality datasets

Reproducibility and bias, particularly in underrepresented patient populations

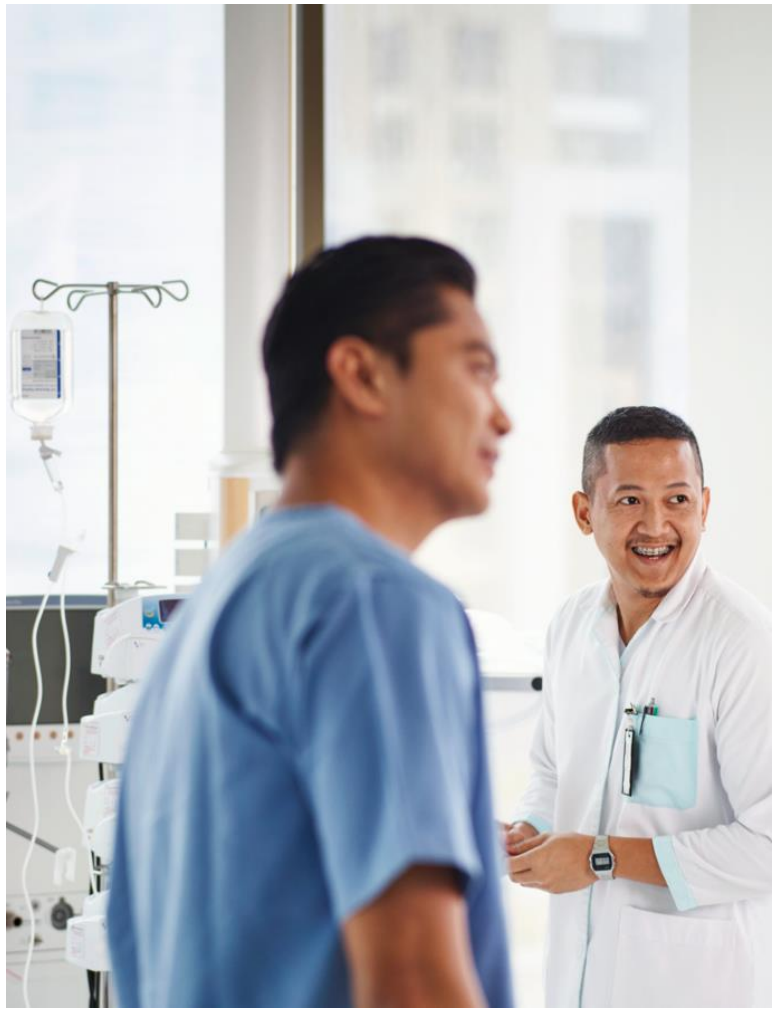
Clinical translation

Overcoming the barriers to deploying AI models in real-world clinical settings, such as regulatory approval, physician training and patient trust

Privacy risks and need for transparency in AI-assisted decision-making

Conclusions

- **Current Limitations:** Molecular testing is costly, time-consuming, and destructive, with limited tissue availability for comprehensive profiling.
- **Cross-modality learning:** H&E slides can predict molecular features like mutations and biomarkers, enabling cost-effective and faster screening.
- **AI/ML Integration:** Multimodal learning combines histology and molecular data for enhanced clinical decision-making.
- **Immunotherapy Insights:** MSI-H/dMMR CRC patients benefit from immunotherapy, and some MSS CRC patients may respond under specific conditions.
- **Challenges for AI in Medicine:** Key barriers include data integration, standardization, validation, reproducibility, and clinical translation.



Acknowledgments

AI/ML Clinical Development Leadership:

Danielle Belgrave

Kim Branson

AI/ML Clinical Decision Support – computational pathology team:

Adam Marcus, Ali Varamesh, Auke Van der Shaar, Ben Rosenfeld-Schreiber, Eleanor Barr, Eleni Chiou, Gonçalo Figueira, Kaiyan Xiao, Yi Zhao, Wolf Byttner

AI/ML Complex System Biology team:

Ching-Hao Wang, Jason Wagoner, Richard Li, Stefan Groha

All our collaborators across Oncology R&D.

Thanks for your attention !

Q&A