

# H&E → ORION Virtual Multiplexing Briefing

Gentles Lab

5 November 2025

## Clinical Motivation

---

- ▶ Accelerate multiplex immunofluorescence (Orion, 20 channels) by predicting from routine H&E.
- ▶ Reduce wet-lab turnaround for macrophage-rich TA118-like cohort; prioritise atypical regions.
- ▶ Provide spatial biomarker estimates to support therapy stratification and rapid QC.

# Workflow Overview

---

- ▶ VALIS rigid & non-rigid registration (Bio-Formats reader).
- ▶ Tissue segmentation via Laplacian + Otsu, morphological cleanup.
- ▶ Crop aligned slides into  $2048 \times 2048$  cores centred on tissue.
- ▶ Global quantile scaler (train-set only) for Orion intensities.
- ▶ Stratified sampling to favour rare/speckled markers.
- ▶ Distributed (DDP) training with mixed precision and channel-aware losses.

## Dataset Snapshot

---

- ▶ 319 paired cores exported as float32 NPY ('core\_\*\*\*').
- ▶ Orion cube consistently 20 channels; H&E normalised to [0,1].
- ▶ Train/val split: deterministic stratification (seeded shuffle).
- ▶ Quantile scaler persisted to 'orion\_scaler.json' and broadcast to all ranks.

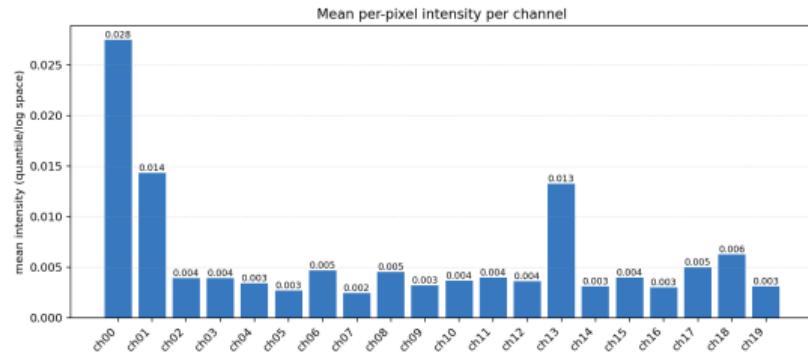
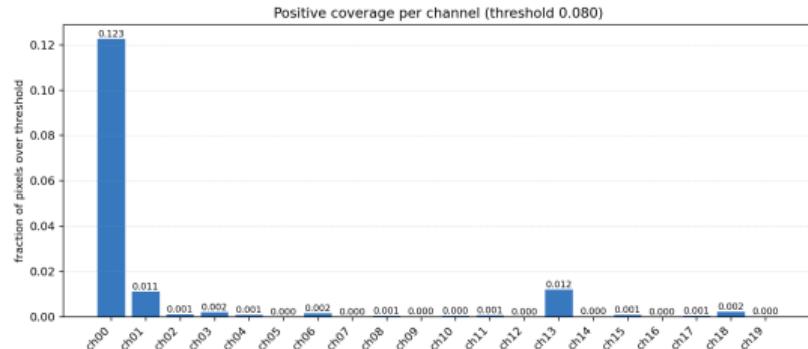
## Marker Panel Highlights

---

- ▶ Macrophage subsets: SPP1, FOLR2, NLRP3, LYVE1, IL4I1.
- ▶ Immune context: HLA-DR (DC), CD3 $\varepsilon$ , CD8 $\alpha$ , FOXP3, CD15.
- ▶ Tumor/stroma: Pan-CK, SMA, FAP, GFPT2.
- ▶ Supports spatial mapping of immune suppression, fibrosis, vascular niches.

# Quantitative Insights

- ▶ Coverage at threshold 0.08 spans  $\sim 12\%$  (ch00) to  $< 0.01\%$  (ch05/ch16).
- ▶ Mean intensities highly skewed; motivates per-channel scaling and sampling.
- ▶ Speckle-heavy markers (ch02, ch13, ch11) require regularisation.



# Pipeline Diagram

---

INSERT PIPELINE SCHEMATIC  
(Registration → Tissue detection → Core export → Training)

# Training Configuration

---

- ▶ Input patches:  $224 \times 224$ , augmentations (flip, rotate, color jitter, resize).
- ▶ Oversample positives:  $p = 0.65$  with channel-aware probabilities from coverage stats.
- ▶ Optimiser: AdamW ( $\text{lr } 3 \times 10^{-4}$ ), cosine decay  $\pm$  warmup, grad clip 1.0.
- ▶ Loss blend: center-weighted MSE + coverage penalty + MS-SSIM (optional) + TV + presence head.
- ▶ Multi-GPU torchrun (DDP) with AMP; best checkpoints tracked via val loss.

# Quantile Scaling & Log Transform

For Orion intensity  $x_c(u, v)$  in channel  $c$ :

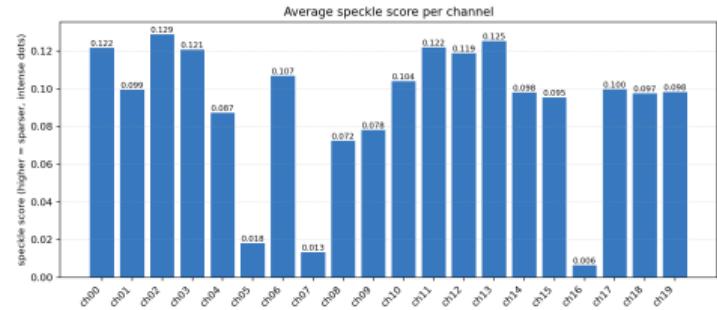
$$\tilde{x}_c(u, v) = \frac{x_c(u, v) - q_c^{\text{low}}}{q_c^{\text{high}} - q_c^{\text{low}} + 10^{-6}}$$

$$z_c(u, v) = \log(1 + \max(\tilde{x}_c(u, v), 0))$$

- ▶  $q_c^{\text{low}}, q_c^{\text{high}}$ : global train-set quantiles (1%, 99.5%).
- ▶ Stabilises dynamic range; prevents bright outliers from dominating gradients.
- ▶ Log transform makes low-intensity “dot” markers resolvable and aligns with Poisson-like noise.

## Why pathologists care

- ▶ Preserves relative marker ordering across cores.
- ▶ Enables consistent back-transformation for ORION review ( $\exp(z) - 1$ ).
- ▶ Avoids over-saturation of stromal-rich or hemorrhagic regions.



# Swin-UNet Architecture

---

- ▶ Encoder: 'swin\_tiny\_patch4\_window7\_224' (windowed self-attention).
- ▶ Feature pyramid with  $1 \times 1$  lateral projections to 192 channels.
- ▶ Bilinear upsample + skip summation; decoder halves width before Softplus output.
- ▶ Excels at long-range morphology patterns, robust to sparse signals.

INSERT SWIN-UNET DIAGRAM

# ConvNeXt-UNet Variant

---

- ▶ Encoder: ‘convnext\_tiny’ (hierarchical depthwise conv blocks).
- ▶ Shares decoder topology; emphasises local texture cues.
- ▶ Faster convergence on abundant markers; more variance on speckled channels.
- ▶ Lower computational footprint, candidate for edge deployment.

INSERT CONVNEXT-UNET DIAGRAM

# Loss Design and Sampling

## Composite loss

$$\begin{aligned}\mathcal{L} = & \lambda_{\text{mse}} \mathcal{L}_{\text{center}} + \lambda_{\text{cov}} \mathcal{L}_{\text{cov}} + \lambda_{\text{SSIM}} \mathcal{L}_{\text{MS-SSIM}} \\ & + \lambda_{\text{TV}} \mathcal{L}_{\text{TV}} + \lambda_{\text{pres}} \mathcal{L}_{\text{presence}}\end{aligned}$$

$$\mathcal{L}_{\text{center}} = \frac{1}{|\Omega|} \sum_{(u,v) \in \Omega} w_c(u,v) (y_c(u,v) - \hat{y}_c(u,v))^2$$

$$w_c(u,v) = 1 + \gamma \mathbb{1}\{\hat{y}_c(u,v) > \tau\}$$

- ▶  $\Omega$ :  $12 \times 12$  central window; emphasises diagnostically annotated zone.
- ▶  $\gamma = 3$  boosts pixels above log-threshold  $\tau = 0.10$ —stops “washing out” rare spots.

## Auxiliary terms

$$\mathcal{L}_{\text{cov}} = \|\bar{y}_c - \hat{y}_c\|_1 \quad (\text{per-channel mean matching})$$

$$\mathcal{L}_{\text{MS-SSIM}} = 1 - \text{MS-SSIM}(y, \hat{y})$$

$$\mathcal{L}_{\text{TV}} = \|\nabla_h \hat{y}\|_1 + \|\nabla_v \hat{y}\|_1$$

$$\mathcal{L}_{\text{presence}} = \text{BCEWithLogits}(\cdot, \mathbb{1}\{\hat{y}_c^{\max} > \tau\})$$

- ▶  $\bar{y}_c$ : spatial average intensity—keeps global abundance realistic.
- ▶ TV discourages hallucinated speckle noise while preserving sharp edges.
- ▶ Presence head gives on/off probability per marker, aiding triage dashboards.

# Adaptive Sampling Strategy

## Per-channel probability

$$p_c \propto \left( \frac{1}{\text{coverage}_c + 10^{-4}} \right)^\alpha (\text{mean}_c + 10^{-4})$$

$$\alpha = 1.0, p_c = \frac{p_c}{\sum_k p_k}$$

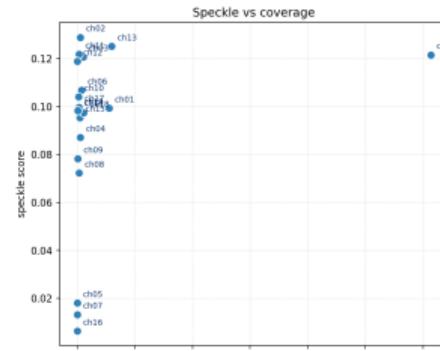
- ▶ Encourages sampling of channels with scarce positive pixels.
- ▶ Speckle-heavy top- $k$  (2, 13, 11, 0) double resample budget to catch rare dots.

## Patch acceptance

accept if  $\exists c \in \{c^*, \text{all}\} : \frac{|\{(u, v) : y_c(u, v) > \tau\}|}{ps^2} > \epsilon_c$

## Clinical rationale

- ▶ Guarantees each mini-batch contains macrophage-rich examples.
- ▶ Aligns model focus with markers flagged by pathologists for TA118.
- ▶ Sampling metadata ( $c^*$ ) logged for auditability and error analysis.

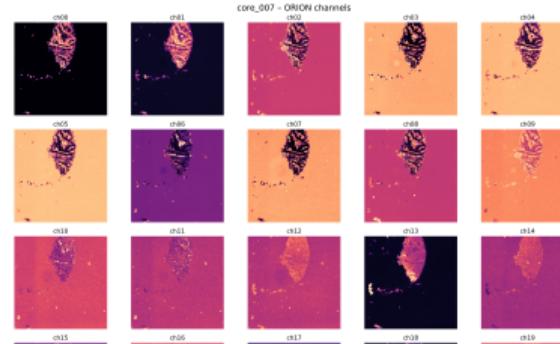
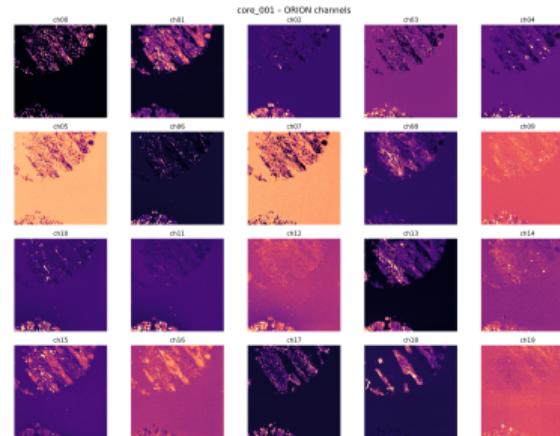
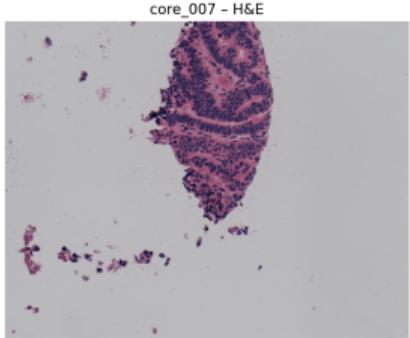
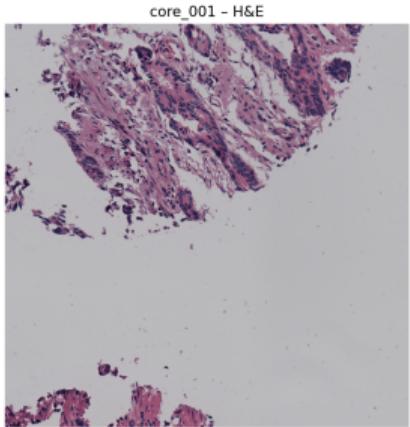


# Training Metrics

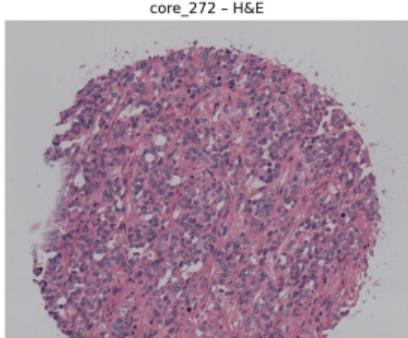
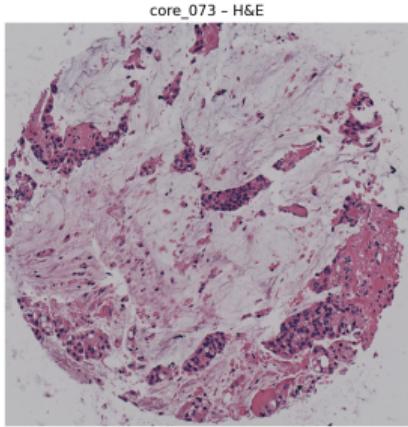
---

INSERT TRAIN/VAL LOSS CURVES  
(e.g., metrics from 'runs\_\*' once exported)

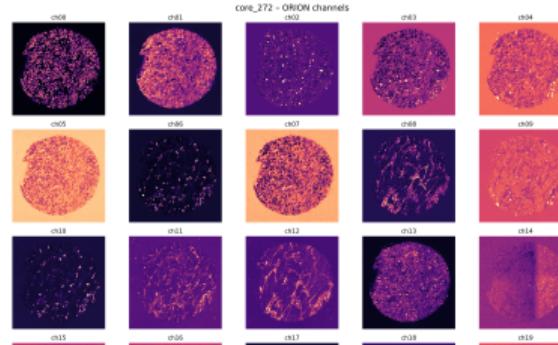
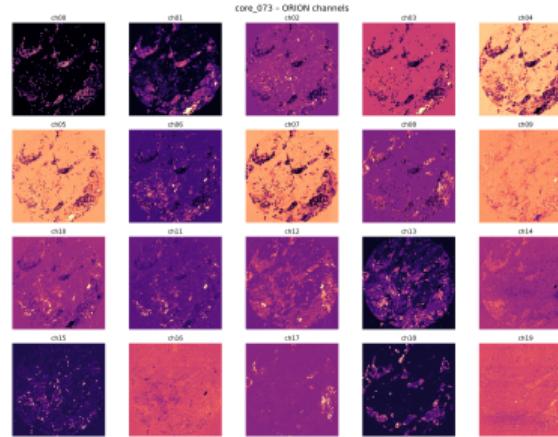
# Qualitative Data QC



# Extended QC Gallery



Gentles Lab



H&E → ORION Virtual Multiplexing

5 November 2025

16 / 23

# Model vs Ground Truth Placeholders

---

INSERT SWIN-UNET vs GT (core\_001)  
per-channel montage

INSERT CONVNEXT vs GT (core\_001)  
per-channel montage

INSERT CHANNEL TRIPTYCH (FOLR2 / CD163 / SPP1)

## Comparative Observations

---

- ▶ Swin-UNet maintains morphology on sparse macrophage markers; reduced halo artefacts.
- ▶ ConvNeXt-UNet sharper on abundant epithelial markers (Pan-CK, SMA) but noisier on low coverage.
- ▶ Channel-aware sampling benefits both; evaluate cross-site generalisation in next phase.
- ▶ Future: ensemble or knowledge distillation to balance accuracy vs efficiency.

## Novelty vs Literature

---

- ▶ Full 20-channel Orion prediction with global quantile scaling and speckle-aware sampling.
- ▶ Presence-aware auxiliary head uncommon in prior H&E → IF translation studies.
- ▶ End-to-end pipeline: registration, QC, DDP training for large cohorts.
- ▶ Extends beyond prior single-marker virtual staining and GAN-based approaches.

## Related Work

---

- ▶ Rivenson et al., *PNAS* 2019: Virtual staining of auto-fluorescence (single channel).
- ▶ Fu et al., *Nat. Biomed. Eng.* 2020: Hyperspectral marker imputation without channel-aware sampling.
- ▶ Lu et al., *Med. Image Anal.* 2022: GAN-based H&E → IF translation, limited scaling.

ADDITIONAL REFERENCES / DOI LINKS AS NEEDED

## Pathologist-Focused Takeaways

---

- ▶ Rapid in silico multiplexing to prioritise cores for lab validation.
- ▶ Channel coverage maps and presence logits offer interpretability hooks.
- ▶ Feedback requested: critical markers, acceptable error ranges, integration needs.

## Next Steps

---

- ▶ Export quantitative metrics (PSNR/SSIM per channel); align with human review.
- ▶ Automate registration QA alerts for misaligned cores.
- ▶ Extend to additional TA cohorts; collect clinical endpoints for outcome modelling.
- ▶ Investigate uncertainty estimates and active learning for rare marker discovery.

# Discussion

---

Questions, feedback, and marker priorities welcome.