Migrating Algorithms Beyond the Development Sandbox

Eric Wirch
Chief Technology Officer
Corista LLC

June 27th, 2018

Algorithmic Development Lifecycle

Work with development partner to:

- Establish a goal and initial criteria
- Gather annotated or labelled data
- Iterative development and experimentation
- Validate the algorithm
- Review of utility for that partner
- Ensure the algorithm's utility for a broad audience

Under the right circumstances, AI has the potential to improve productivity & accuracy. Algorithmic development can be labor intensive, for both humans and computers.

The challenge: design efficient, user-friendly tools around creating AI workflows.

Is the Algorithm Ready for a Broad Audience?

Algorithm may either:

- Require standardization
  - Tissue processing
  - Stain standardization (no longer each institution's choice)
  - Specific special stains or protocols
- Be tolerant of inter-lab differences
  - Additional training and testing
  - Initial institution-specific validation

Factors for Consideration Regarding Tolerance

- Differences in slide production, reproducibility
  - Tissue thickness, placement
  - Human factors
- Institution-specific staining best practices
- Differences in stains per-manufacturer
- Differences in standard scanning / WSI generation
  - Compression
  - Manufacturer
  - Resolution

corista

Migrating from the Sandbox

Engage additional research partners which:

- Introduces variations in preparation
- Measure algorithm's tolerance to inter-lab differences
- Requires review of results by multiple partners
- Reduces the bias of a single organization
- Requires lots of data and expertise

Case Study: Glomerulus Detection

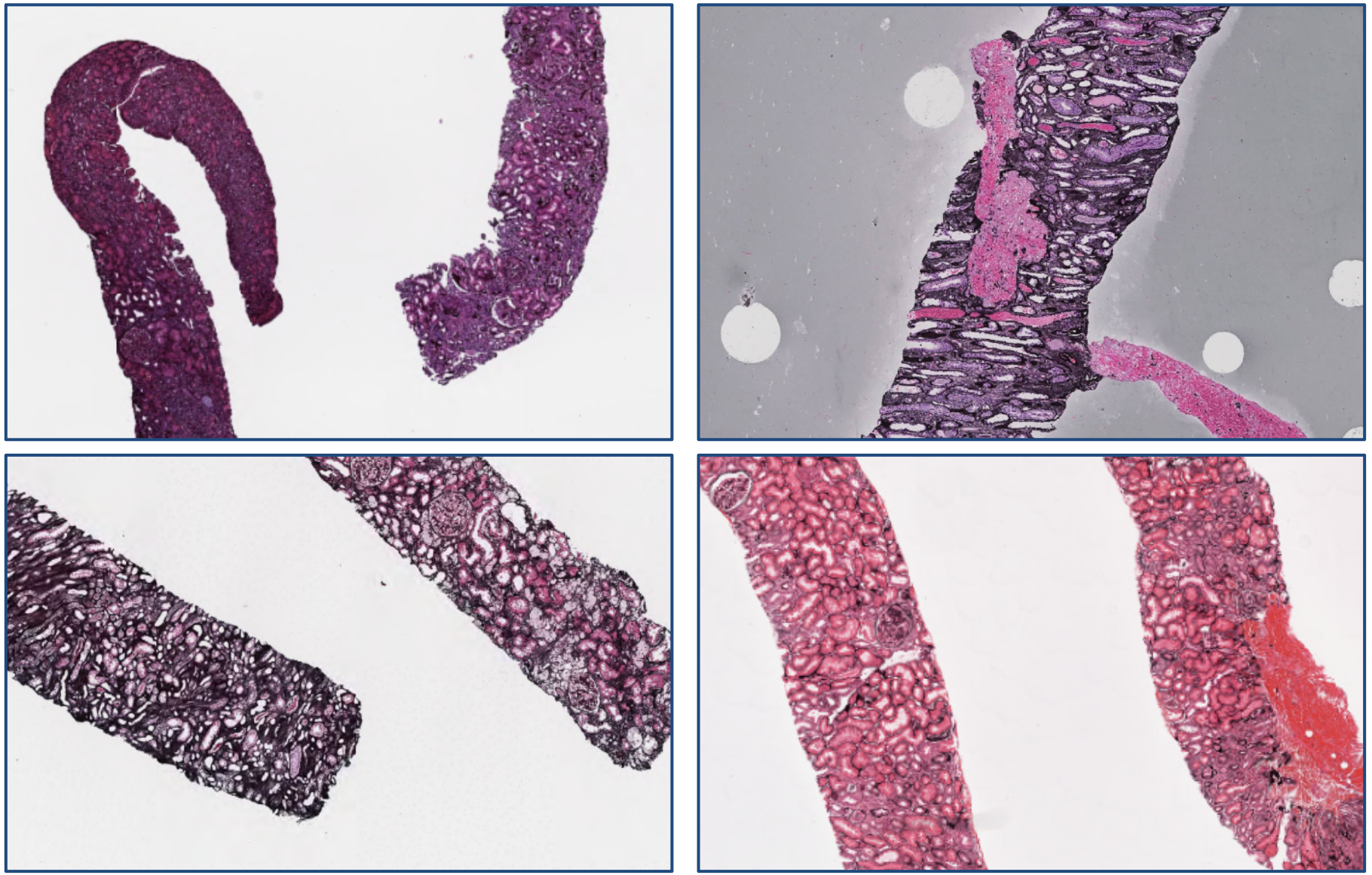
Parameters:

- Four stain renal biopsy (H&E, Trichrome, Basement Membrane, PAS)
- Stain Detection and CNN selection
- Detection (and eventual classification) of glomeruli w/ CNN

Goals:

- Simultaneous, registered navigation of all slides at once
  - Ease of annotation by collaborators
  - Clickable detection results

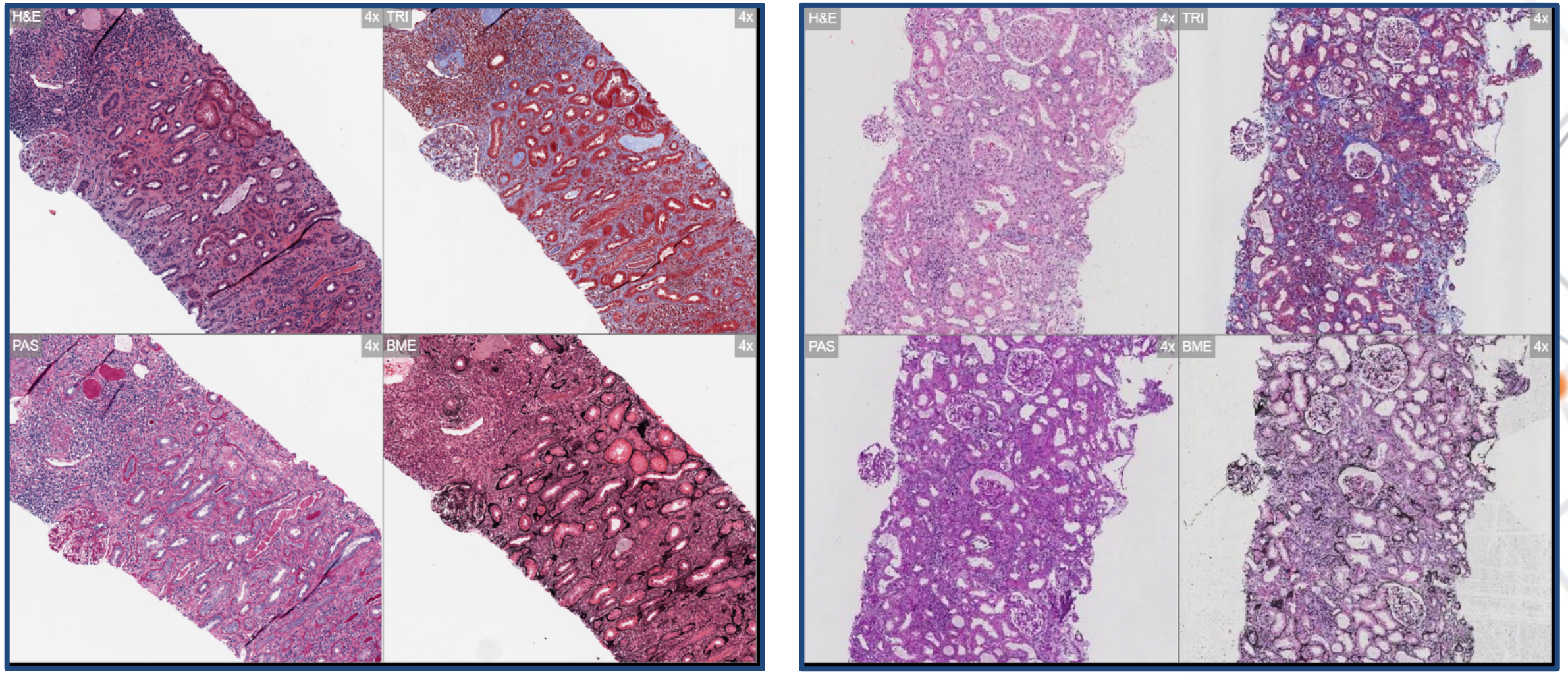
Inter-lab WSI Differences



BME Stains from 4 Labs

corista

Inter-lab WSI Differences

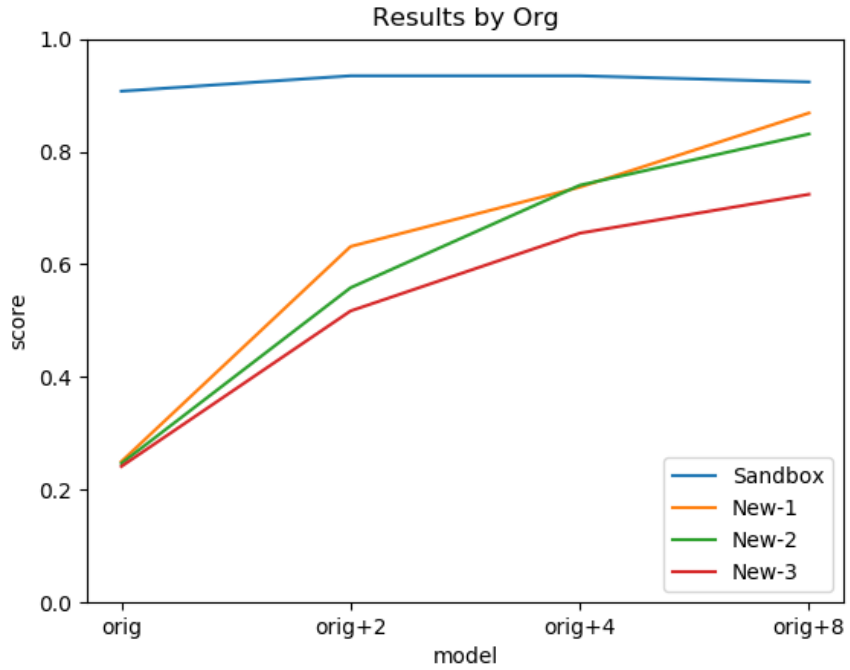


4 Stain Panel from 2 Labs

corista

Stain Classification Results

	Results by Org			
	Sandbox	New-1	New-2	New-3
Orig. Model	90.71%	25.00%	24.68%	24.14%
Orig. + 2/Org	93.44%	63.16%	55.84%	51.72%
Orig. + 4/Org	93.44%	73.68%	74.03%	65.52%
Orig. + 8/Org	92.35%	86.84%	83.12%	72.41%

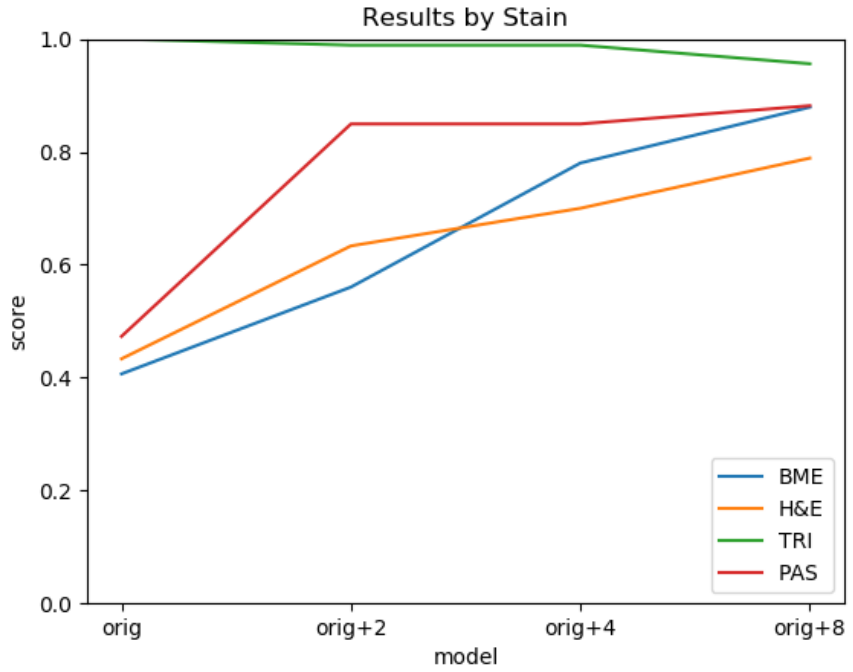


- Original model represents 24 Sandbox cases
- Significant improvements for new Org's when a minimal amount of data is included
- Sandbox Org sees minimal impact from including additional data

corista

Stain Classification Results

	Results by Stain			
	BME	**H&E**	**TRI**	**PAS**
Orig. Model	**40.66%**	**43.33%**	**100.00%**	**47.31%**
Orig. + 2/Org	**56.04%**	**63.33%**	**98.90%**	**84.95%**
Orig. + 4/Org	**78.02%**	**70.00%**	**98.90%**	**84.95%**
Orig. + 8/Org	**87.91%**	**78.89%**	**95.60%**	**88.17%**



- Numbers are sensitivity only
- Significant improvements when a minimal amount of data is included
- 87% average with unbalanced, minimal inclusion training

corista

CNN Detection Comparison

Model	Test Data	Detection Accuracy	
		POS %	NEG %
Sandbox Model	Out-of-Sample Sandbox	94.6	93.3
Sandbox Model	Out-of-Sample 4 Orgs	10.8	36.8
Sandbox Model	Out-of-Sample New-Org1	8.7	25.0
New-Org1 Model	Out-of-Sample New-Org1	92.0	71.5

Model Statistics

- Identical CNN Structure
- Sandbox Model trained on 25 Cases (6.4M images)
- New-Org1 Model trained on 3 Cases (1.6M images)

Evaluation Metrics

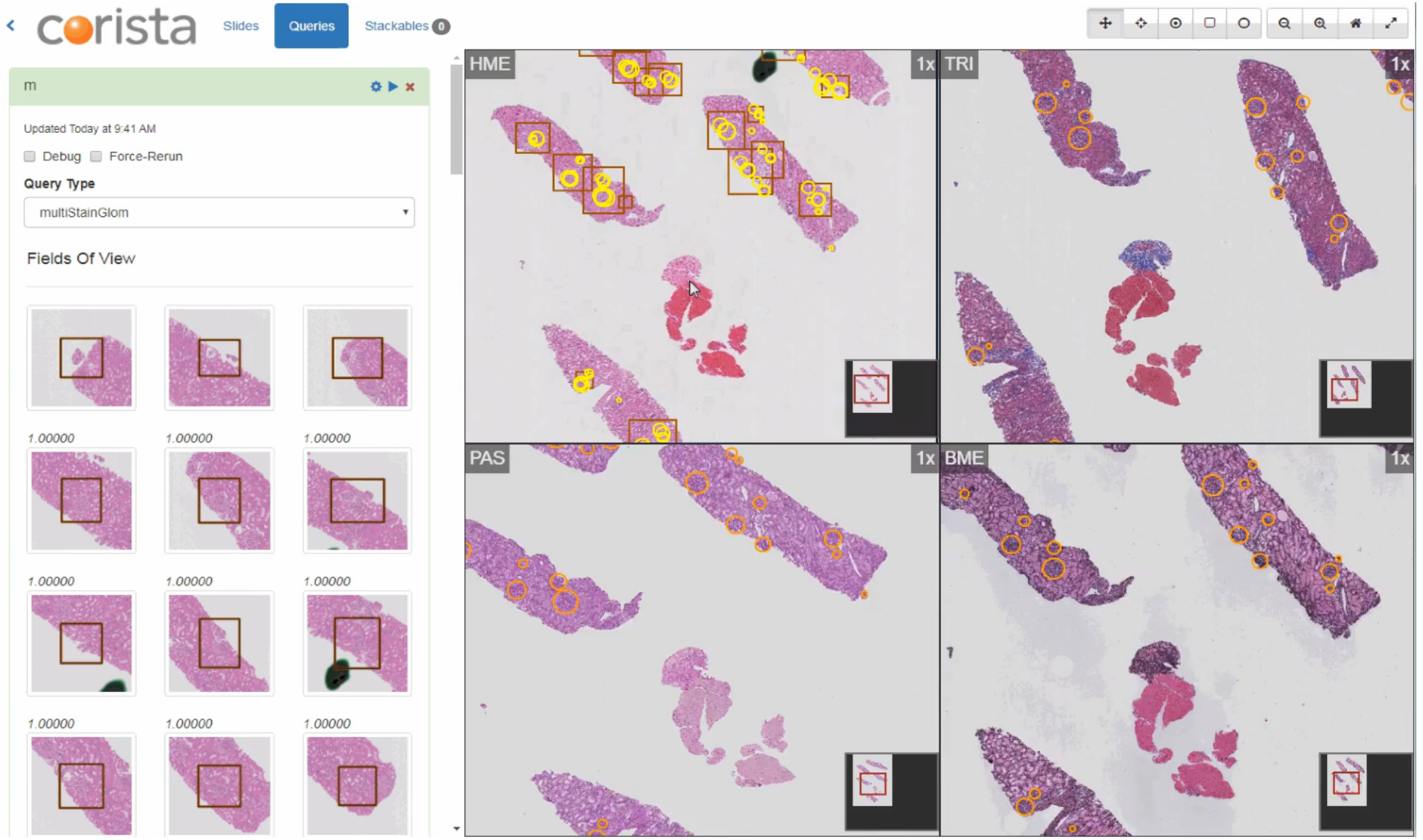
POS % (Sensitivity) =
(# Correct Detects) / (# Annotated Glomeruli)

NEG % (Specificity) =
1- [(# False Detects) / (# Total Detects)]

Strongly suggests that the methodology holds for new data sources, if we can handle the challenges.

corista

Registered Results Review



corista

Challenges to the Sandbox Algorithm

- Digitizer differences
  - Vendor
  - Light source & image capture
  - Resolution
  - Compression
- Staining / Handling differences
  - Histology processes
  - Stain manufacturer
  - Age
  - Saturation

Techniques to Mitigate Differences

- Multi-resolution, multi-technique image registration
- Color normalization
- Model retraining
  - Balanced retraining from scratch
  - Incremental retraining of existing model
- Multi-expert review of input data and results
  - Labor intensive unless the tool facilitates this

Multi-resolution Image Registration

The WSIs require us to leverage multiple registration techniques:

- Feature-based registration
- Intensity-based registration

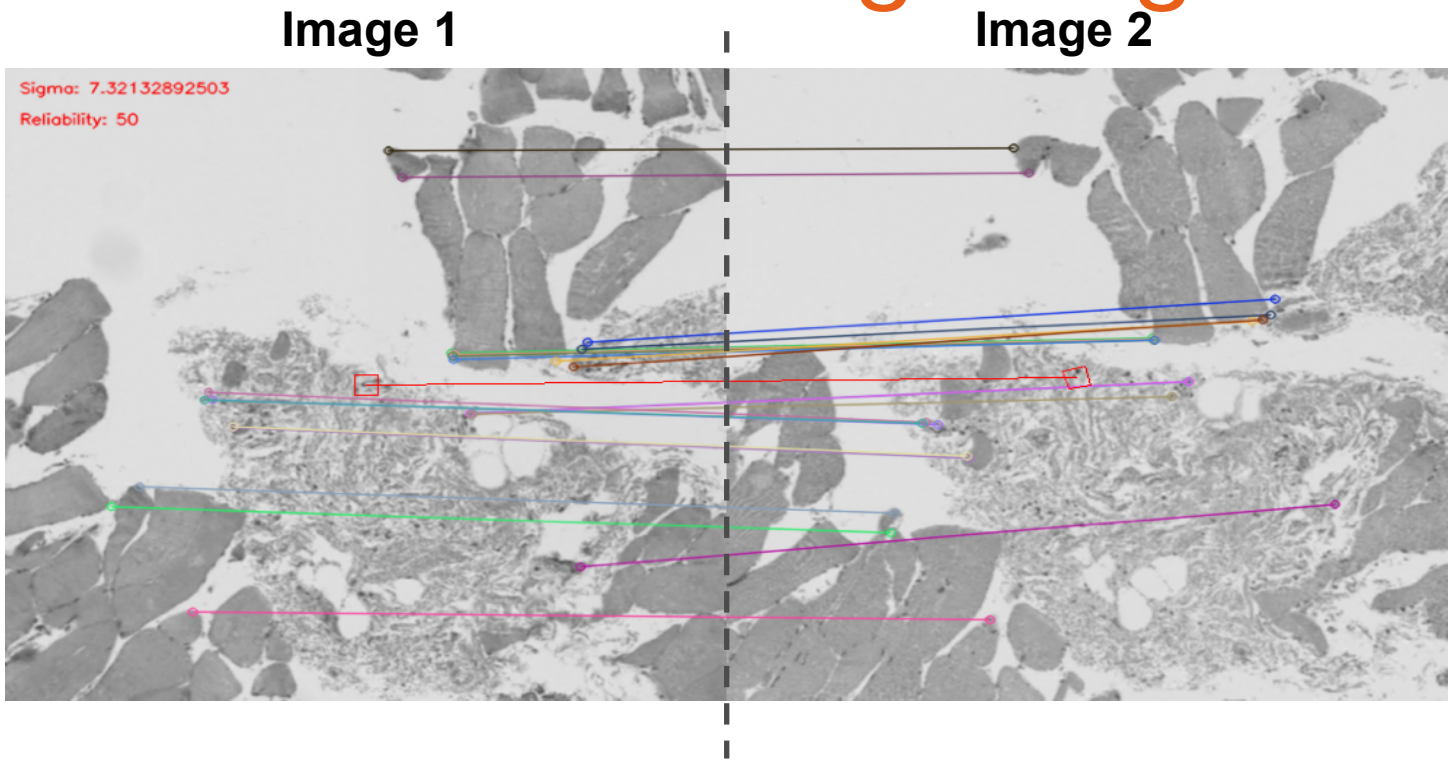
Once coarse registration (overview-level) is achieved, we iteratively register at higher resolutions to refine the results.

Feature-based Image Registration

- Feature-based registration leverages a feature detector like SURF, SIFT, ORB
- Finding invariant features
- Features are then matched based on similarity and “distance”
- Outliers are removed to create a general, then more specific mapping (ATM)

corista

Feature-based Image Registration



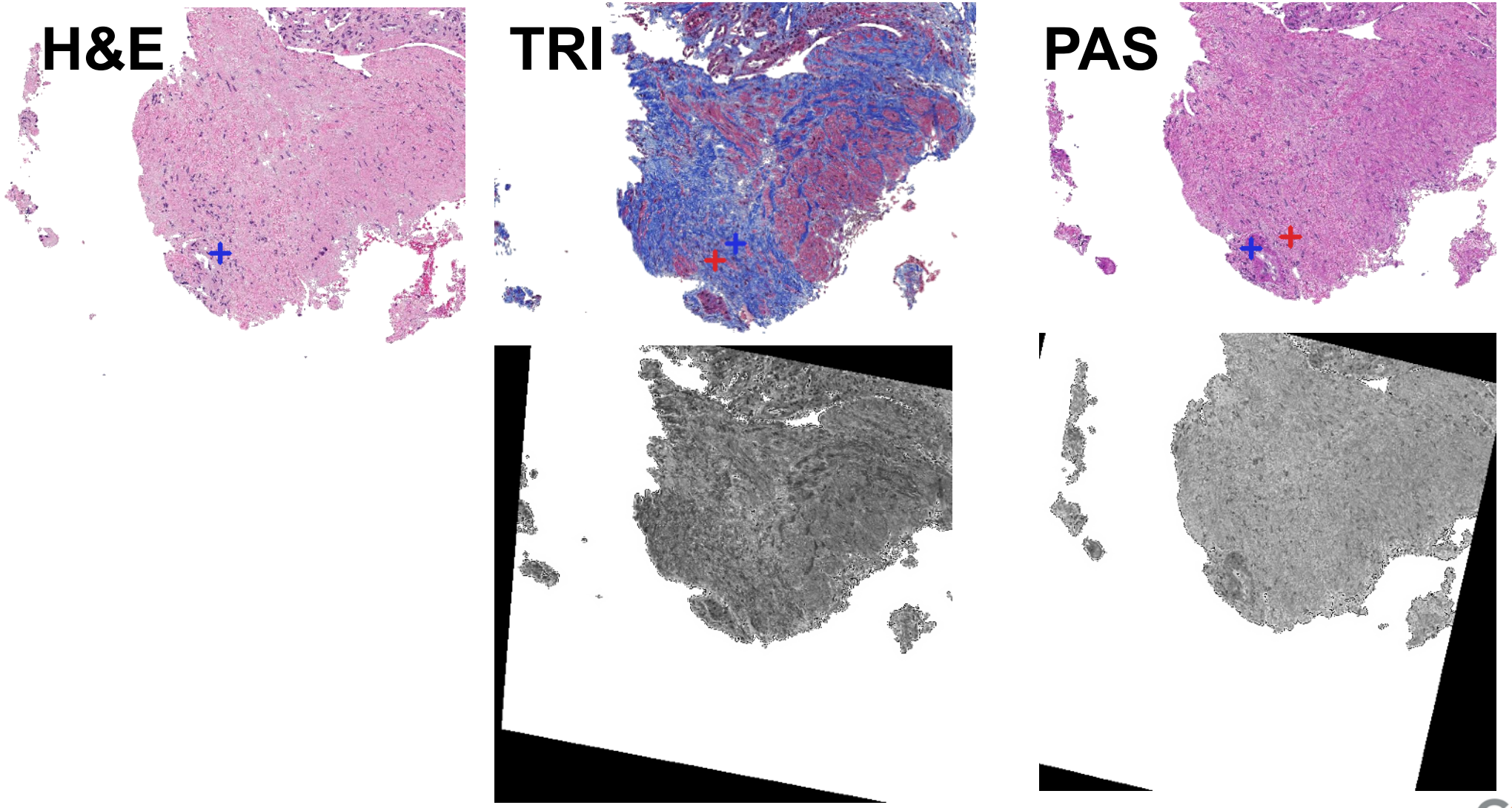
- Image Registration generates a transform matrix, mapping arbitrary points from one image to another image
- The transform matrix is used to align detections and align Fields of View when looking at the results across stain sets

corista

Intensity-based Image Registration

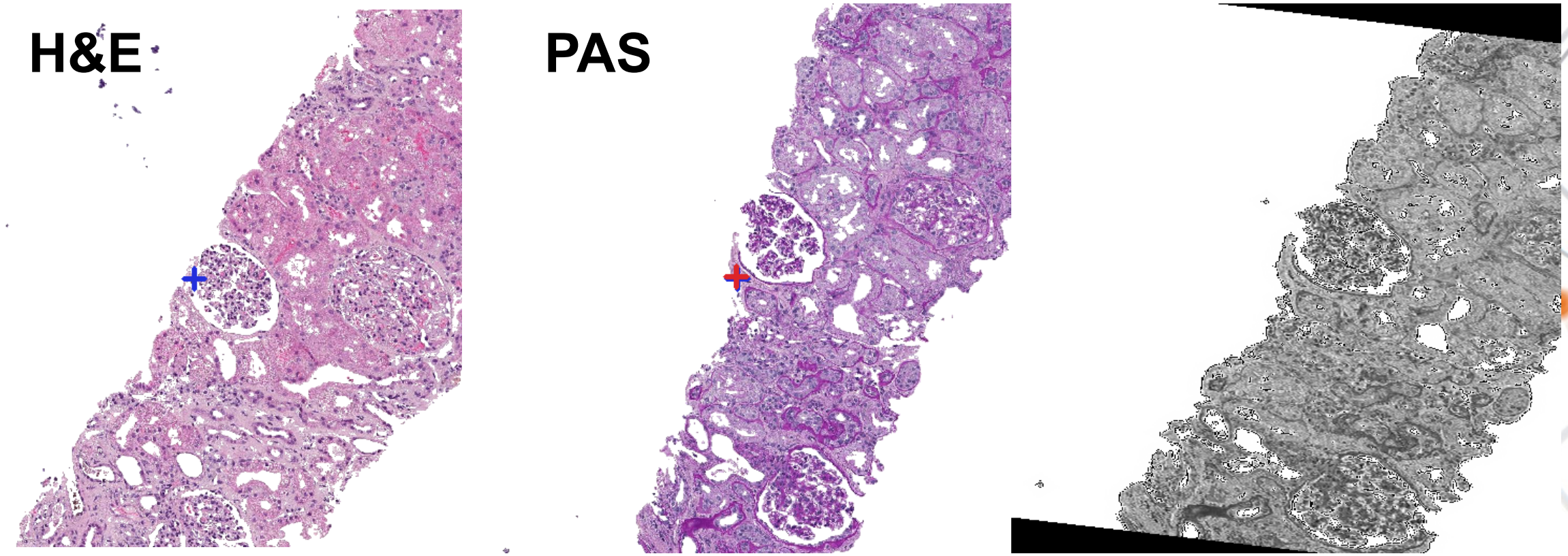
- Intensity-based image registration first involves isolating the foreground and background
- The foreground structures (the tissue) are converted to grayscale (intensity)
- Intensity image is smoothed (blurred)
- Intensities are matched and evaluated, eliminating outliers
- Outliers are removed to generate the increasingly more specific mapping

Intensity-based Image Registration



corista

Intensity-based Image Registration



corista

Color Normalization

- Matches all input slides to a ‘target’ histogram (color profile)
- Involves segmentation of foreground from background
- Target histogram can be from a single WSI or an aggregate
- Can be implemented in multiple color spaces
- Matches the Cumulative Distribution Function (CDF) of the WSI to the target histogram

corista

Color Normalization

Example normalization in the RGB color space

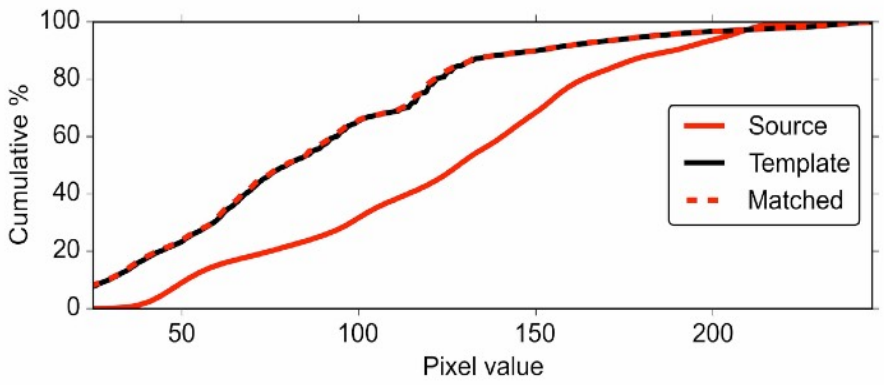


Image A

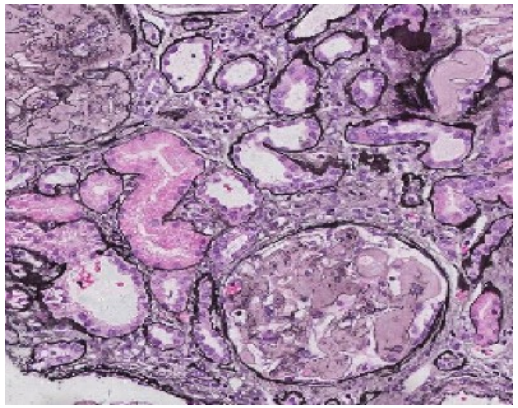
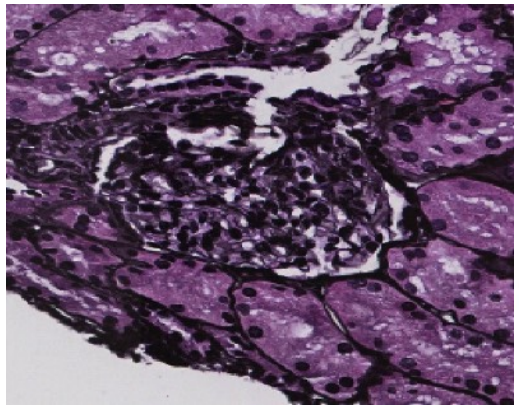
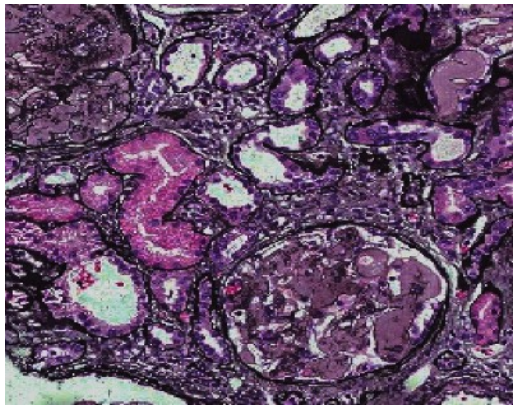


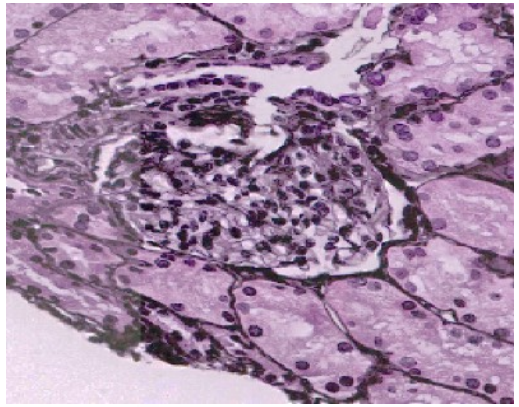
Image B



A→B



B→A



corista

Model Retraining Options

- Balanced retraining from scratch with old + new data
  - More balanced results, needs as much time as original training
- Incremental training using existing model and a mix of data
  - Uses proportionally less time for retraining
  - Tendency to shift CNN to favor the new data over the existing model
- Retraining of classifier layer with old + new data
  - Significantly less time than retraining from scratch (~ ⅕ time)
  - Smaller shift from previous CNN, improvement on new data