## About OMICS International

OMICS International through its Open Access Initiative is committed to make genuine and reliable contributions to the scientific community. OMICS International hosts over 700 leading-edge peer-reviewed Open Access Journals and organizes over 1000+ International Conferences annually all over the world. OMICS International journals have over 3 million readers and the fame and success of the same can be attributed to the strong editorial board which contains over 50000 eminent personalities that ensure a rapid, quality and quick review process. OMICS International signed an agreement with more than 1000 International Societies to make healthcare information Open Access. OMICS International Conferences make the perfect platform for global networking as it brings together renowned speakers and scientists across the globe to a most exciting and memorable scientific event filled with much enlightening interactive sessions, world class exhibitions and poster presentations.

# Automatic Image Processing for Estimation of Prostate Cancer Tumour Regions and Patient Outcomes

Patrick Jackman, William Gallagher, William Watson

School of Medicine and Medical Science &
School of Biomolecular and Biomedical Science

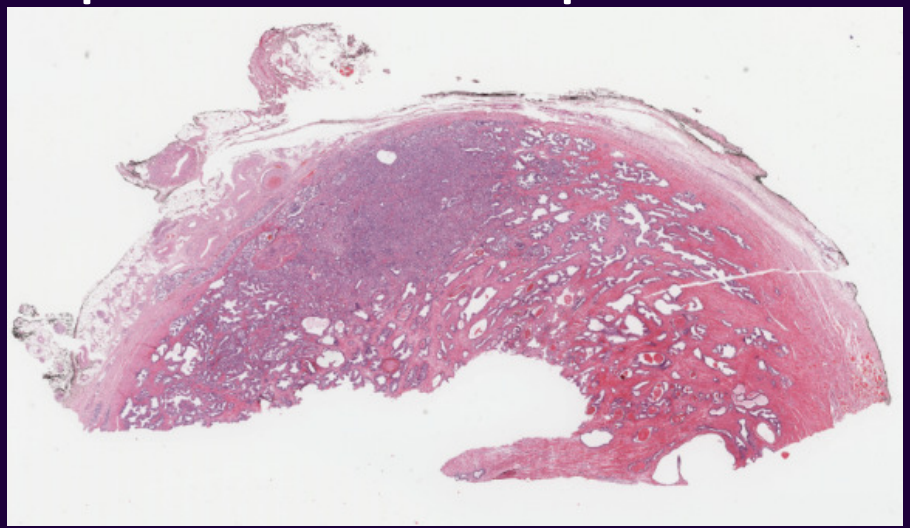
Conway Institute, University College Dublin, Ireland

# Why Automated Image Processing

- Manual examination of tissue - sampling errors
- Reproducibility and repeatability issues
- In contrast Automatic image processing can:
  - extract all potential features of interest
  - Use statistical and predictive modelling
  - Integrate with electronic databanks and repositories

# Recent Advances of Automated Image Processing

- Bright field imaging with extremely fine granularity
- Normally used after staining e.g. H&E etc.
- Images contain billions of pixels
- Processing these datasets is challenging
- Images broken up into small squares or ‘tiles’



# Recent applications of Automated Image Processing

- Challenge is isolation of objects of interest
- Traditional methods can be unworkably slow,
- Speed versus precision compromise required
- Vehicle for generating ergonomic ‘heatmaps’

- Reduce the workload of pathologists

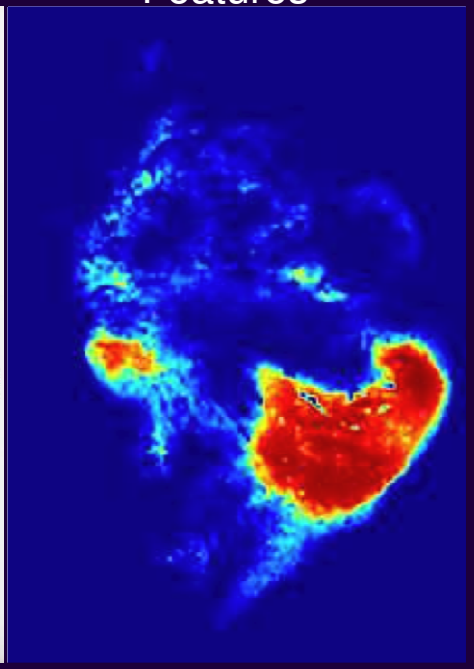
# Building an Automated Image Processing Solution

- Use full tissue sections in algorithm development
- Algorithm must be applicable to biopsies
- Histopathologist provide ground truth tumour data
- Algorithm can identify tissue regions where tissue features are ‘hot’ or ‘cold’

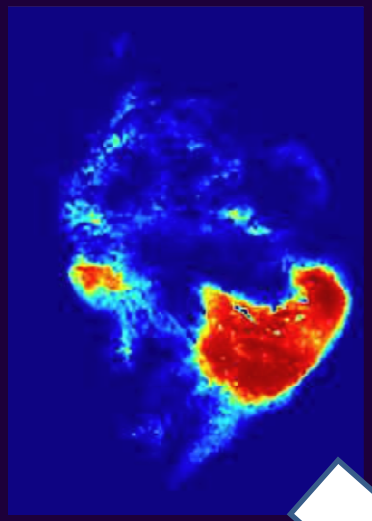
Histopathologist Annotated

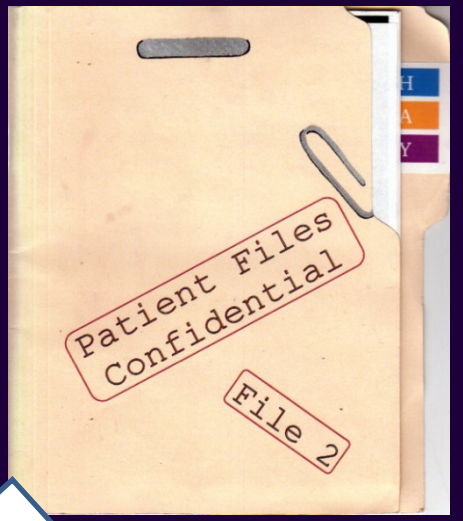


‘Hot’ and ‘Cold’ Tissue Image Features



# Building an Automated Image Processing Solution








Aggressive

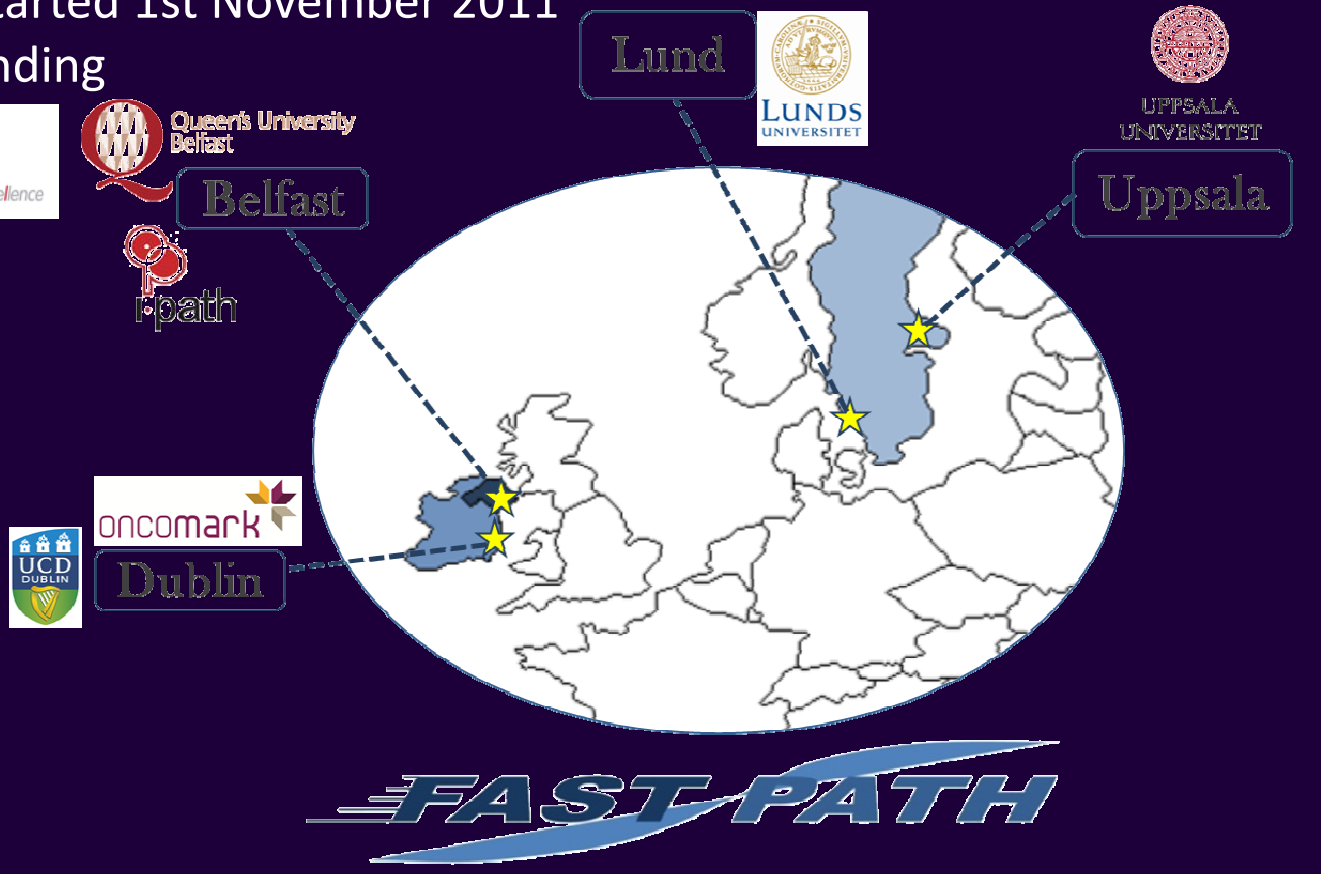
Significant

Indolent

# FASTPATH Project:
# A novel approach to Digital Pathology being implemented by a new research consortium

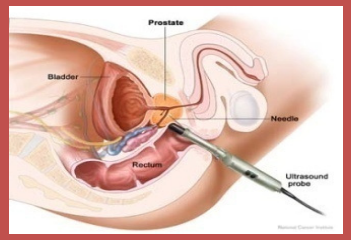
**Marie Curie Industry-Academia Partnership and Pathways (IAPP) programme**

4 academic partners, 2 SMEs
4 years, started 1st November 2011
€1.9M funding



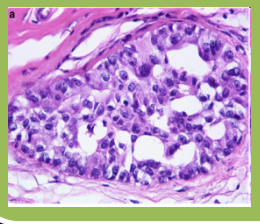

MARIE CURIE ACTIONS

# FASTPATH Project

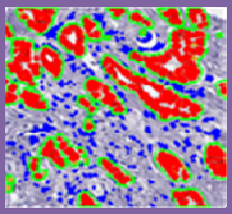



**Prostate cancer biopsies**

**Discrimination of morphological subtypes**



**Quantitation of prognostic biomarkers**





**Solutions for high-performance and high-throughput image analysis**

**Online image library and search engine**

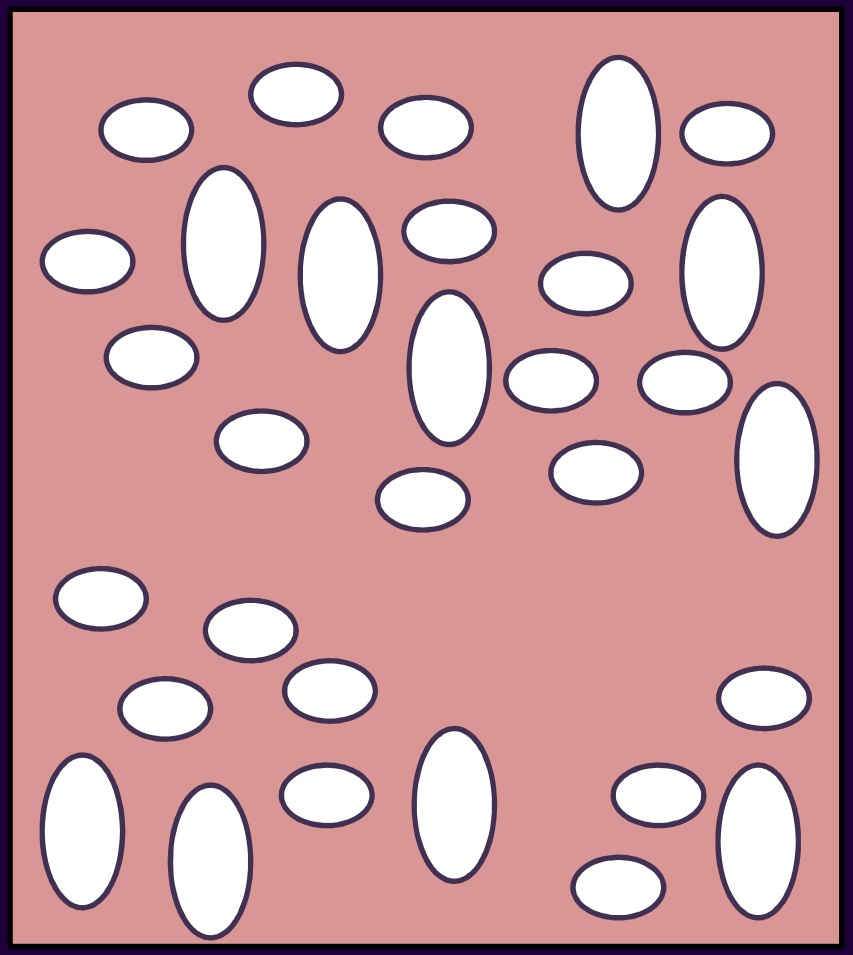


# Work package

- Seek new image features that can identify Indolent, Significant and Aggressive cancer
- Features must be clinically relevant
- Features choice validated by histopathologists

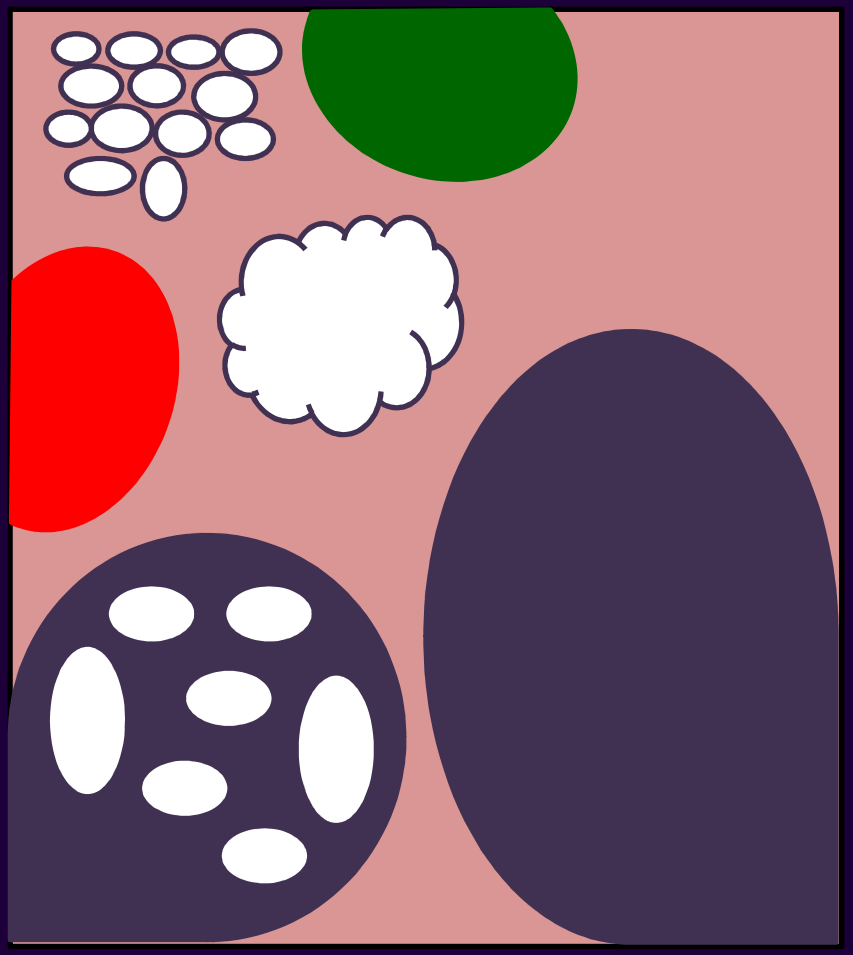
- ‘Digital Pathologists Rulebook’ was created

# Signs of Healthy Prostate Tissue at Low Magnification



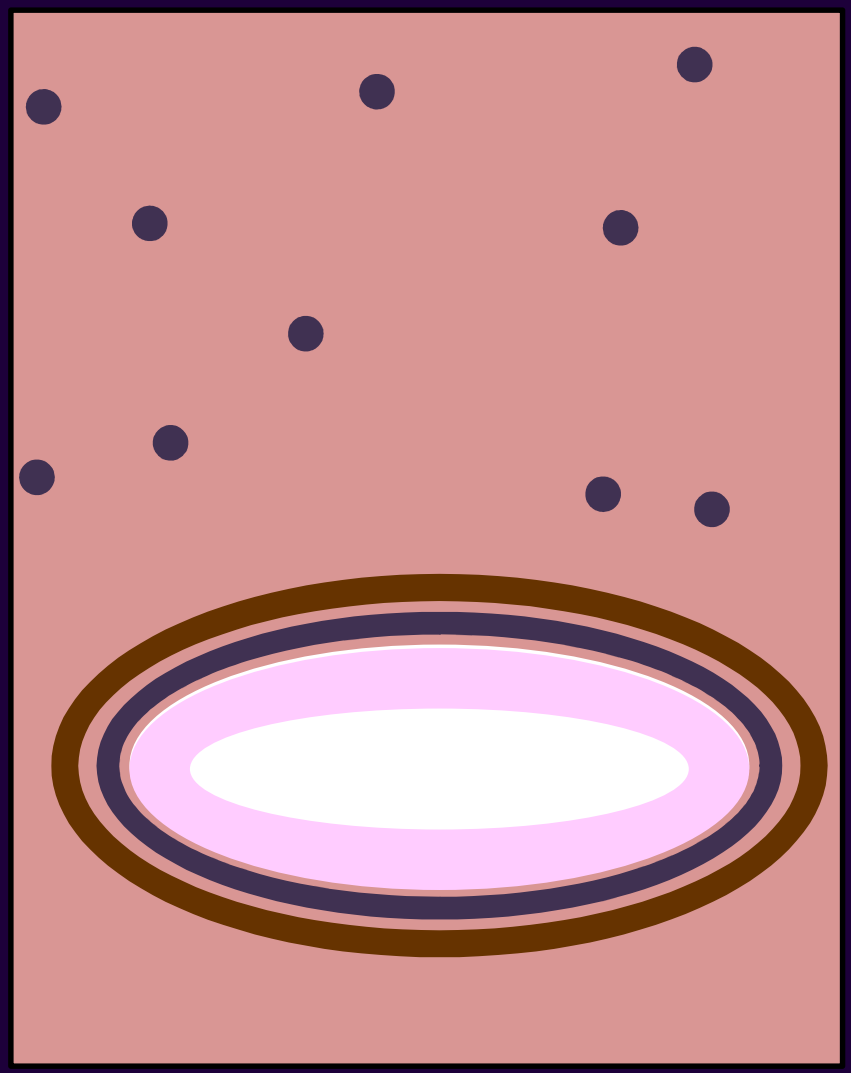
- Low magnification view:
- L1: No solid purple patches
- L2: No purple patches encasing white lumen
- L3: White lumen are separated by pink tissue
- L4: No small white lumen
- L5:No very large white lumen
- L6: Some variation in lumen size but low size diversity
- L7: No red or green patches from Racemase staining

# Signs of Unhealthy Prostate Tissue at Low Magnification



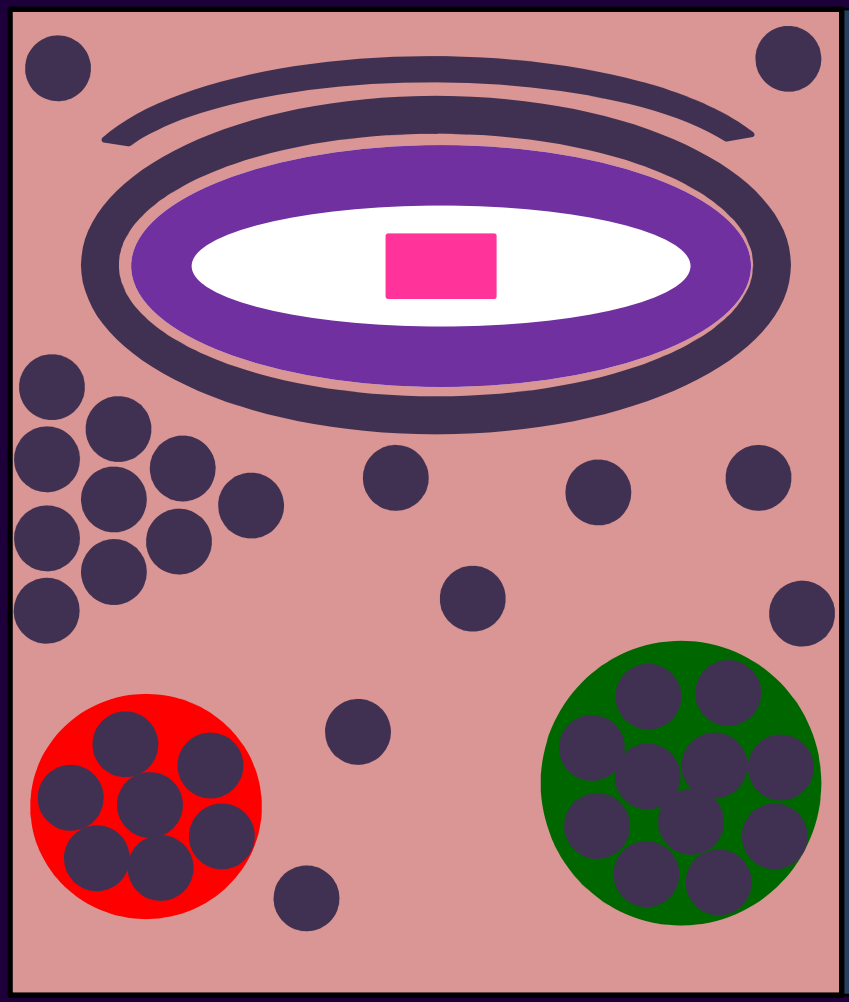
- Low magnification view:
- L1: Solid purple patches
- L2: Purple patches encasing white lumen
- L3&L4: Groups of uniform small white lumen not well separated by pink tissue
- L5: Huge fused white lumen
- L6: High lumen size diversity
- L7: Red or green patches from Racemase staining

# Signs of Healthy Prostate Tissue at High Magnification



- High magnification view:
- H1: Two intact purple rings of normal sized cell nuclei around the lumen
- H2: Outer purple ring turns brown with P63 stain
- H3: No red/green spots from Racemase staining
- H4: Faded pink lumen fringe
- H5: White centre to the lumen
- H6: Pink background has a low density of normal sized nuclei

# Signs of Unhealthy Prostate Tissue at High Magnification



- High magnification view:
- H1: Nuclei larger than normal
- H1: Second purple ring missing
- H2: Brown P63 stain fails
- H3: Red/Green spots from Racemase staining
- H4: Sharp purple lumen fringe
- H5: Sharp pink centre to lumen
- H6: Pink background has higher density of large cell nuclei and some cell clustering

# Digital Pathologists Rulebook

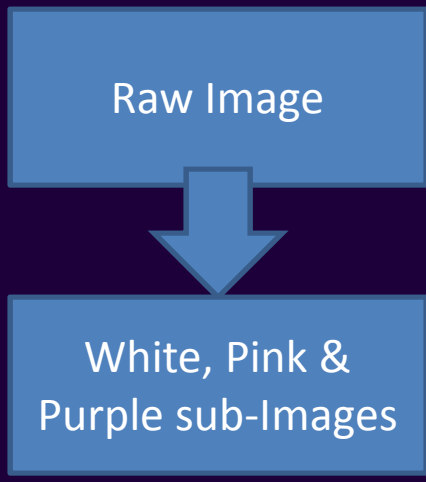
- Consensus amongst the pathologists about high magnification features
- Substantial disagreement at low magnification
- Rulebook encoded in Matlab with a view to finalisation as C++

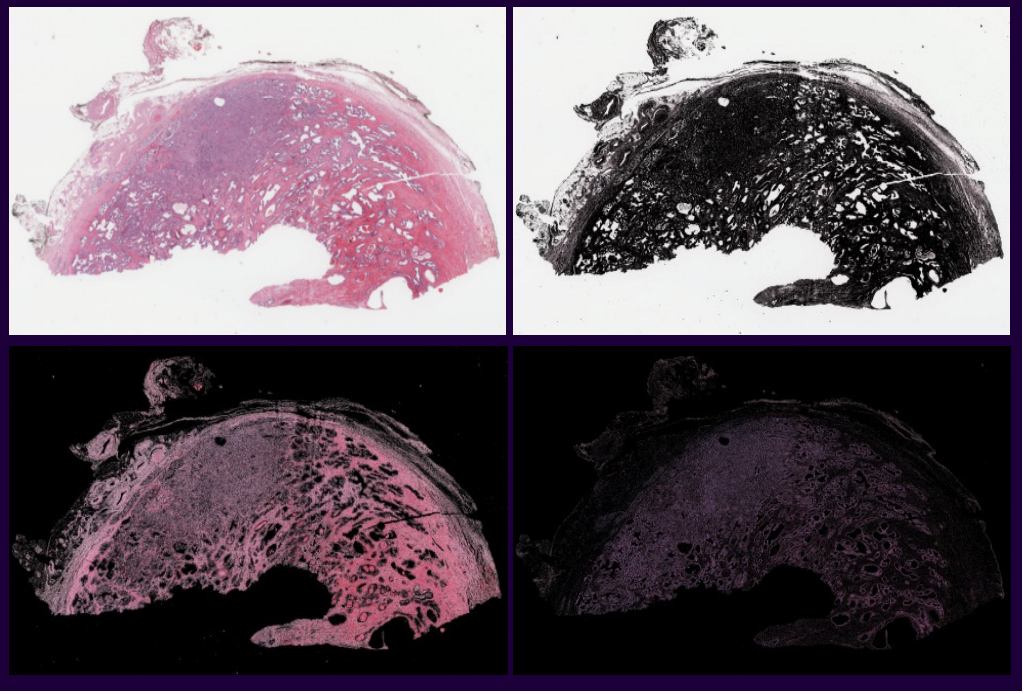
# Implementing Digital Pathologists Rulebook

- Cohort of Irish Prostate Cancer patients were used for digital imaging post H&E staining
- The corresponding clinical histories (e.g. PSA, DRE etc.) were available
- Full tissue sections were digitally scanned
- Sections for 140 patients were analysed via customised programs in the Software Matlab

# Implementing Digital Pathologists Rulebook

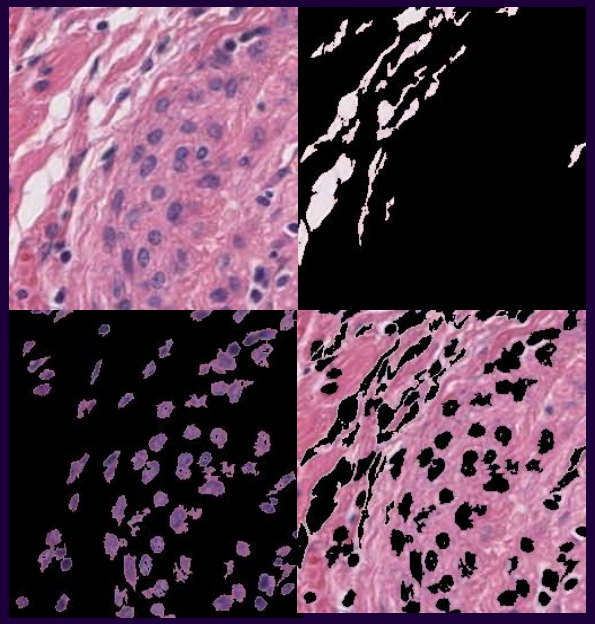
- Step 1 is to define boundaries between White, Pink and Purple
- A Fuzzy-C-Means algorithm performs the segmentation
- Very dark pink regions proved troublesome



# Implementing Digital Pathologists Rulebook

- Step 2 is to call each pixel as White, Pink or Purple
- Each 256 x 256 pixel ‘tile’ within the image is analysed
- Very small objects are dismissed as noise or artifacts



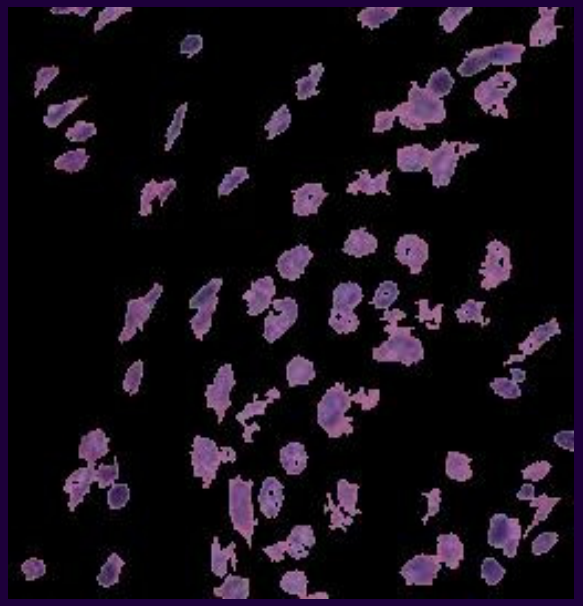
Raw Image

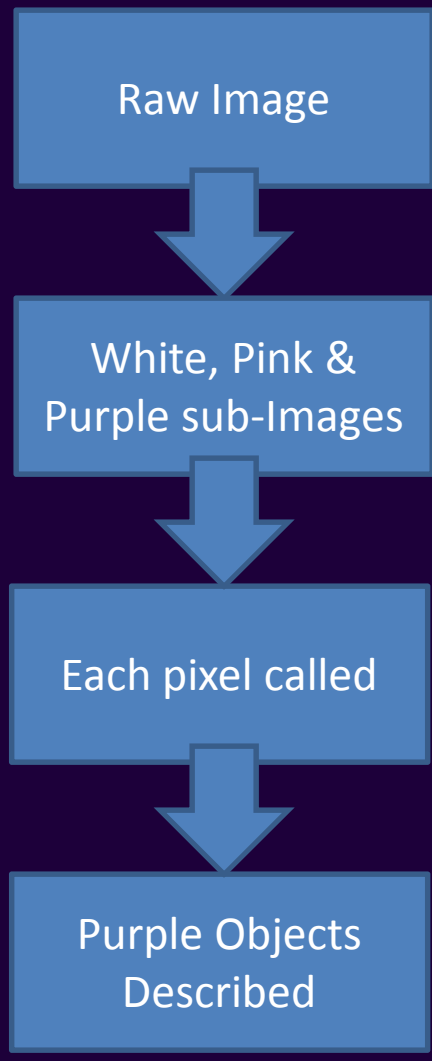
White, Pink & Purple sub-Images

Each pixel called

# Implementing Digital Pathologists Rulebook

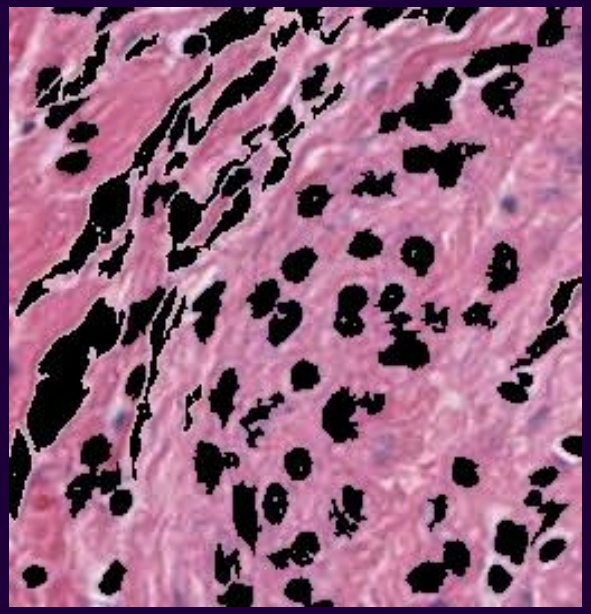
- Step 3 is to describe purple objects
- Purple objects within each ‘tile’ are analysed
- Cell nuclei features from rules H1 & H6

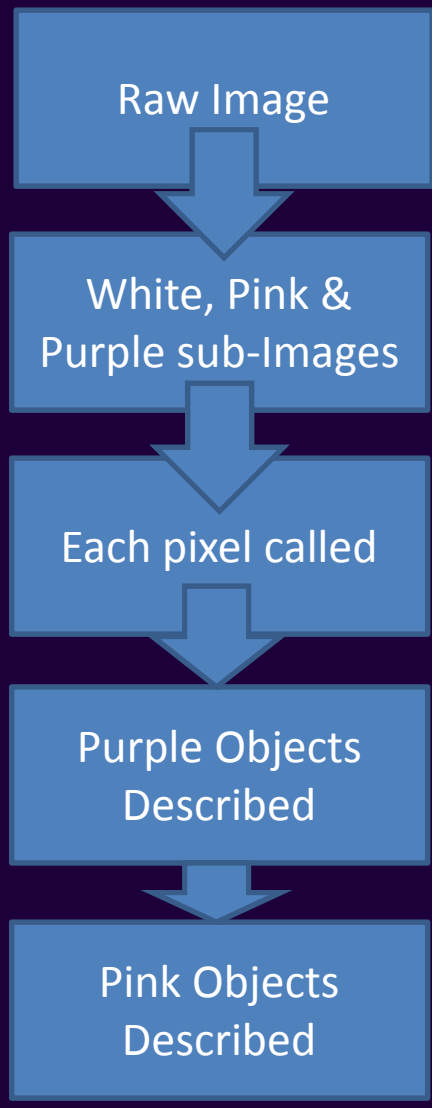




# Implementing Digital Pathologists Rulebook

- Step 4 is to describe pink objects
- Pink objects within each ‘tile’ are analysed
- Stromal features from rules L3 & H6

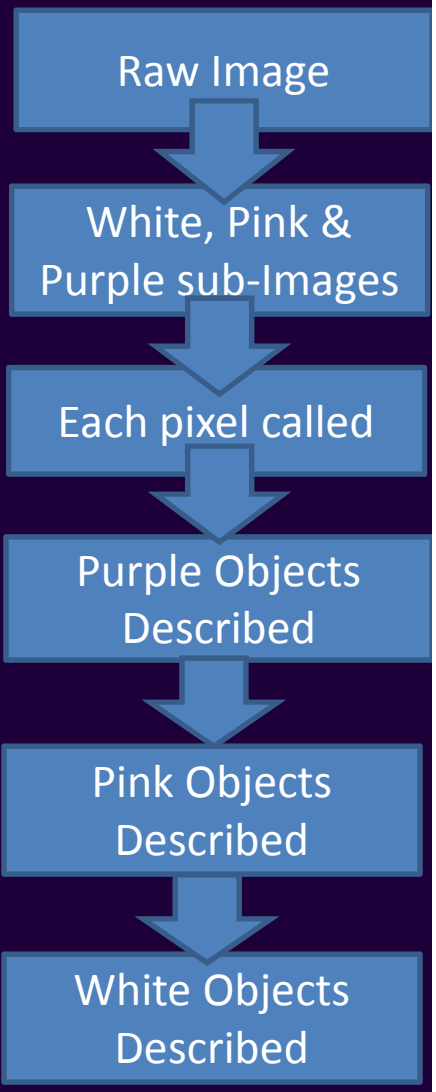




# Implementing Digital Pathologists Rulebook

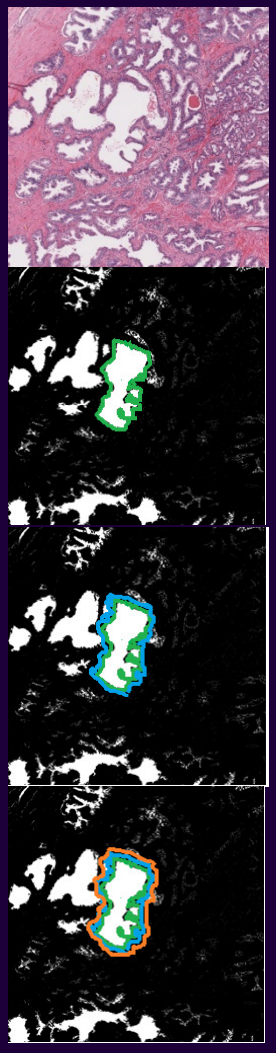
- Step 5 is to describe white objects
- White objects within each ‘tile’ are analysed
- Luminal features from rule L3
- As lumina cross ‘tile’ boundaries rules L2,L4,L5 & L6 are difficult to implement





# Implementing the Digital Pathologists Rulebook

- Step 6 is to quantify loss of integrity of the cellular rings around the gland lumen
- The boundary of the white lumen is examined to search for purple pixels forming an epithelial ring from rule H1
- The boundary of the white lumen is further examined to search for a basal ring from rule H1



Raw Image

White, Pink & Purple sub-Images

Each pixel called

Purple Objects Described

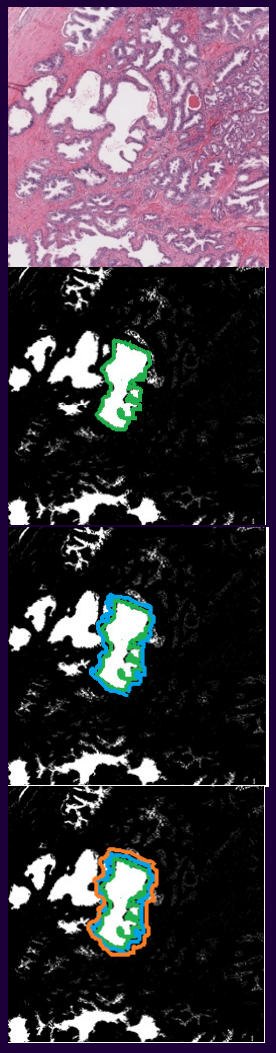
Pink Objects Described

White Objects Described

Gland Boundaries Described

# Implementing the Digital Pathologists Rulebook

- Extremely challenging to ensure that only genuine lumen are identified



Raw Image

White, Pink & Purple sub-Images

Each pixel called

Purple Objects Described

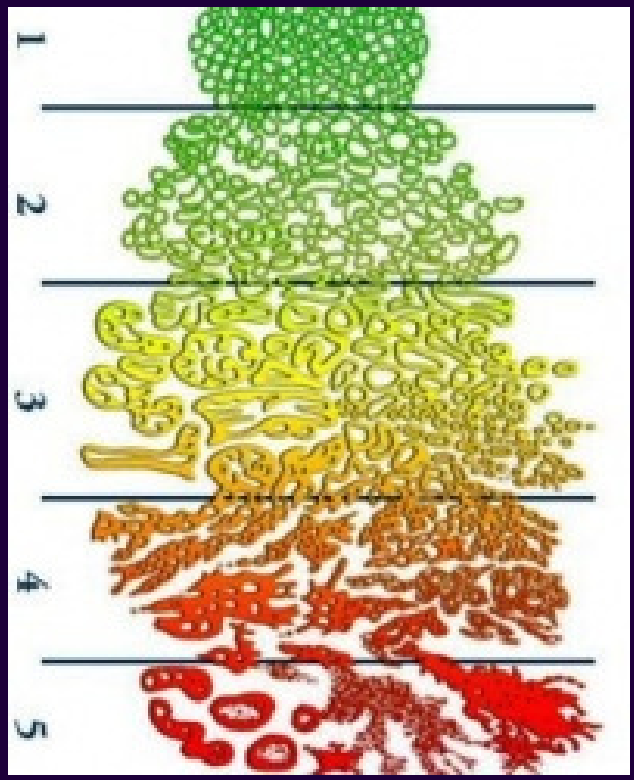
Pink Objects Described

White Objects Described

Gland Boundaries Described

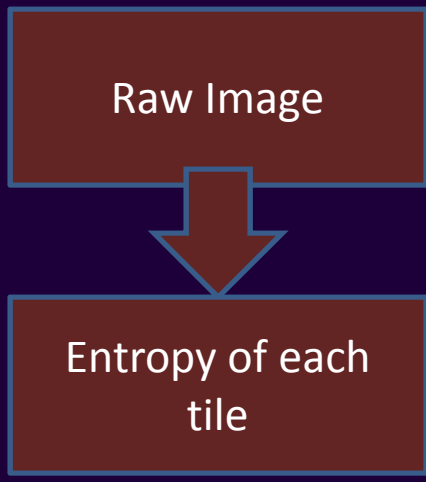
# Novel Approaches in Tissue Characterisation

- Original concept of Gleason et al. was to express how normal tissue structures are replaced
- Ergonomic for a visual assessment by highly skilled pathologists
- Problems of reproducibility, repeatability and lack of granularity
- NOT suited to automatic image processing



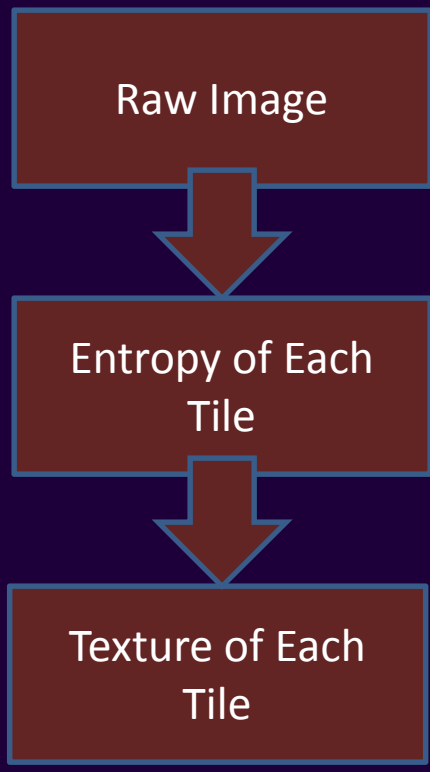
# Novel Approaches in Tissue Characterisation

- Need to express degradation of tissue structures in a reproducible, repeatable and quantifiable way
- First concept of degradation can be drawn from the 2nd law of Thermodynamics
- Entropy of each tile leads to identification of degraded regions and tissue summary features



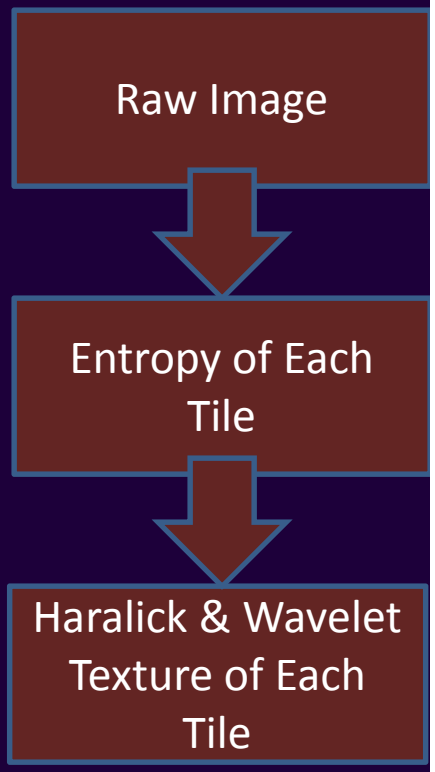
# Novel Approaches in Tissue Characterisation

- Second concept of degradation comes from image texture which will alter as the normal glandular structure is replaced
- Quantification of texture is not suited to the human eye to brain function but is ideal for computerised solutions



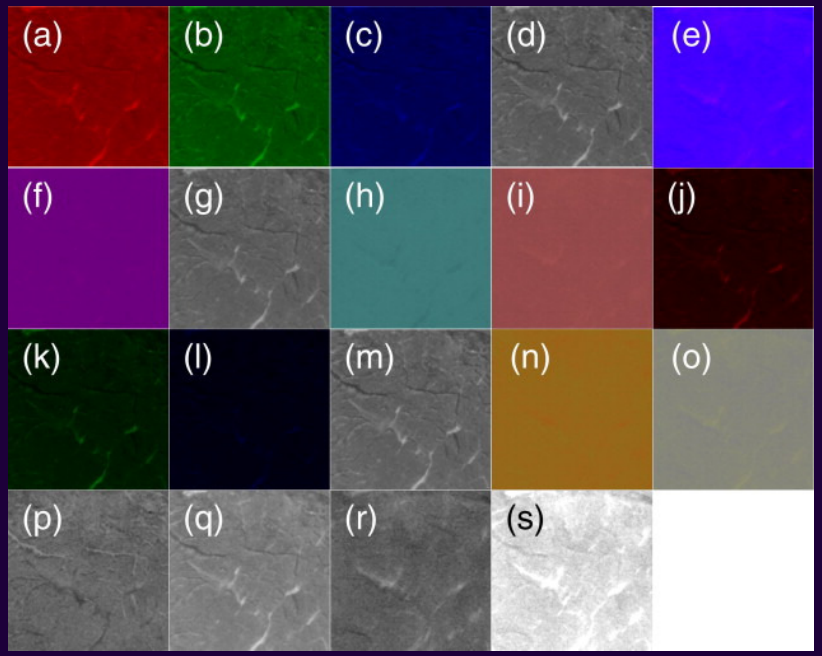
# Novel Approaches in Tissue Characterisation

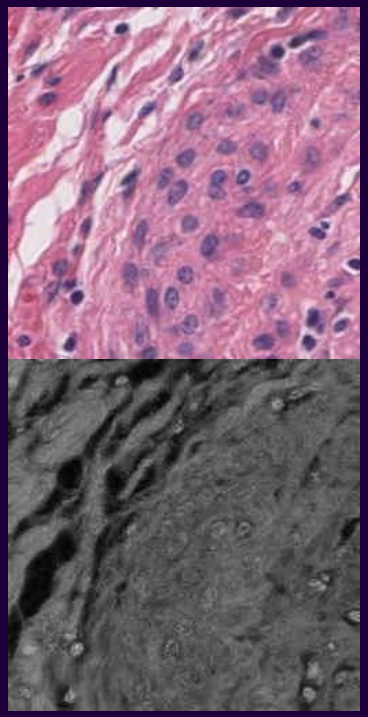
- Texture can be expressed according to the formulae proposed by the Computer Scientist Haralick in 1973
- Alternatively by formulae called Wavelet transforms
- Texture features for each tile leads to identification of degraded regions
- Summary features across the whole section can also be calculated

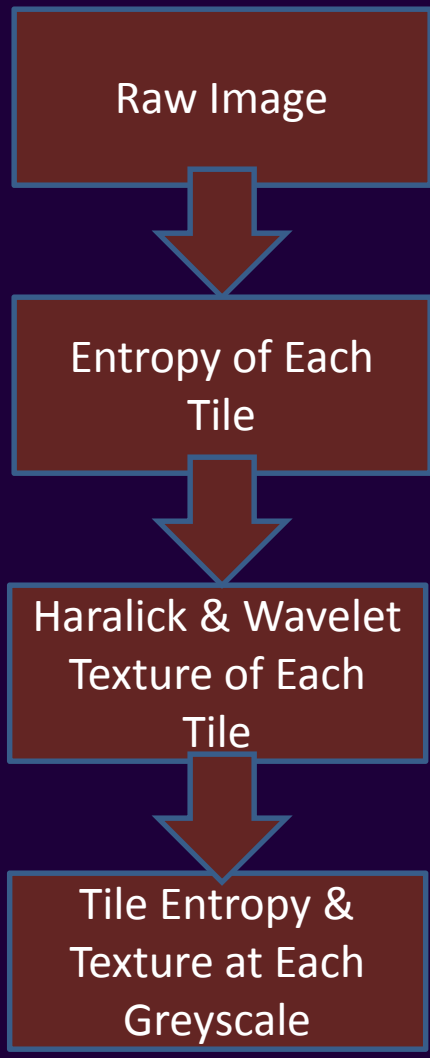


# Novel Approaches in Tissue Characterisation

- 18 common greyscales were used to search for the most effective entropy and texture features

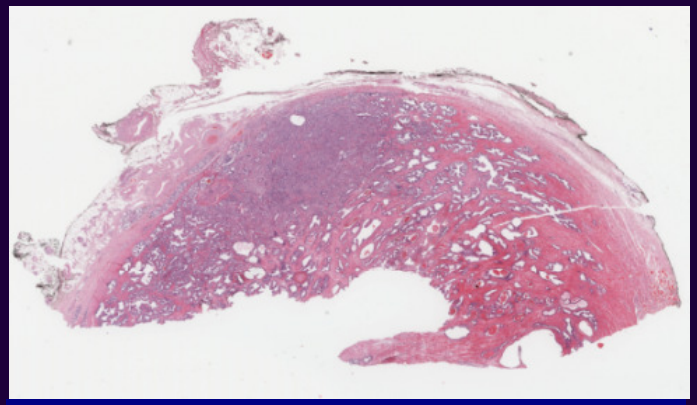


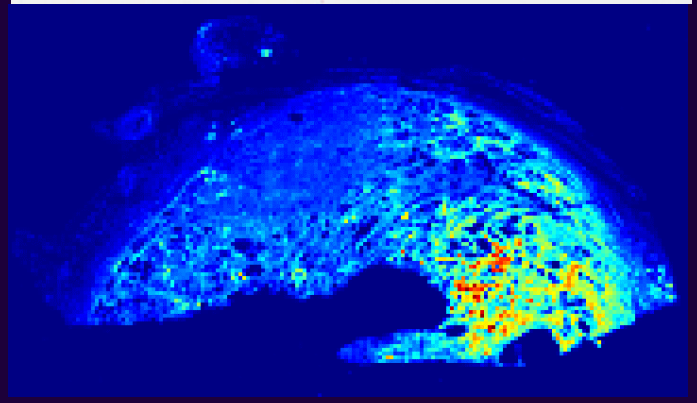





# Highlighting Regions of Interest

- Each tile average of a feature leads to a ‘heatmap’
- Hundreds of heatmaps can be generated
- Histopathologists ground truth the tumour regions of 9 sections





Purple Objects Described by Tile

Pink Objects Described by Tile

White Objects Described by Tile

Gland Boundaries Described by Tile

Tile Entropy & Texture at Each Greyscale

# Novel Approaches in Estimating Tissue Regions of Interest

- Combining useful heatmaps leads to tumour identification
- ‘Riskmap’ thus generated by observing which heatmaps correlate with the ground truth data
- The Riskmap can be safely used if it is representative, robust, precise and accurate

# Novel Approaches in Estimating Patient Outcomes

- True power in Automated Image analysis is when it is used to make a prediction of ultimate outcomes
- Image features can be quantified and used to build predictive models
- Outcome is validated from an expert assessment of radical prostatectomy for all 140 patients
- The model can be safely used if it is representative, robust, precise and accurate

# Limitations in Data Analysis

- Automatic Image analysis also contains dangers due to the multiple comparisons or ‘coin tossing’ problem

- Thus robust statistical modelling is required for validation

# Data Pre-Processing

- Number of image features is too great so the volume of the data must be reduced
- Redundant features need to be removed by statistical techniques such as:
  - Principal Component Analysis (PCA)
  - Partial Least Squares Regression (PLSR)
  - Global Optimisation Algorithms (GOA)

# Highlighting Regions of Interest

- Using the reduced datasets and the ground truth data, tissue classification models can be built with methods such as:
  - Linear Correlation
  - Partial Least Squares Regression (PLSR)
  - Discriminant Analysis (DA)
  - Support Vector Machines (SVM)
  - Neural Networks (NN)
  - Fuzzy Logic (FL)
- Performance based on correct classification rates or correlation rather than Area under the Curve (AUC)

# Predicting Patient Outcomes

- Each of the features can be summarised over the whole section with a similar need to remove redundant features
- The outcome for each patient is known
- Predictive models can be constructed

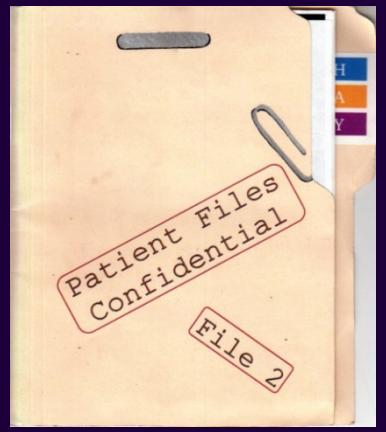
Purple Objects Described Overall

Pink Objects Described Overall

White Objects Described Overall

Gland Boundaries Described Overall

Overall Entropy & Texture at Each Greyscale



# Predicting Patient Outcomes: Model Construction

- Using the reduced dataset of features and the ground truth data, patient outcome models can be built with analytical methods such as:
  - Discriminant Analysis (DA)
  - Partial Least Squares Regression (PLSR)
  - Neural Networks (NN)
  - Fuzzy Logic (FL)

# Predicting Patient Outcomes

- Each of the features can be summarised over the whole section with a similar need to remove redundant features
- The outcome for each patient is known
- Predictive models can be constructed
- Model performance can be quantified as correct classification rate

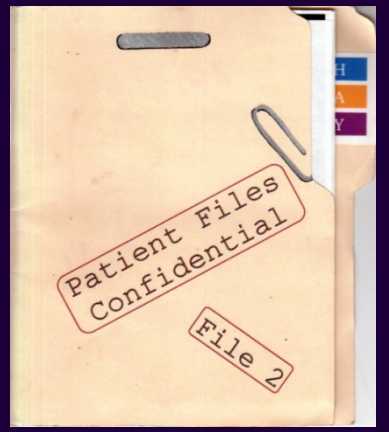
Purple Objects Described Overall

Pink Objects Described Overall

White Objects Described Overall

Gland Boundaries Described Overall

Overall Entropy & Texture at Each Greyscale



# Interpreting Misclassifications

- Traditionally False Positives and False Negatives are not distinguished
- This leads to a variation on 'Pascal's Wager'

| Believe Treatment Needed: | Treatment Needed: YES | NO |
|---|---|---|
| YES | No Early and Painful Death but Significant Health Consequences | Unnecessary Significant Health Consequences |
| NO | Early and Painful Death | No Early and Painful Death nor Significant Health Consequences |

# Interpreting Misclassifications

- Lord William Blackstone considered such dilemmas when balancing the risk of hanging an innocent man versus releasing a murderer back into the community
- "*It is better that ten guilty persons escape than that one innocent suffer*“



# Interpreting Misclassifications

- Similar dilemma is faced by crop farmers
- Do they spray their crops with pesticide?
- They typically balance false negatives against false positive at 5:1

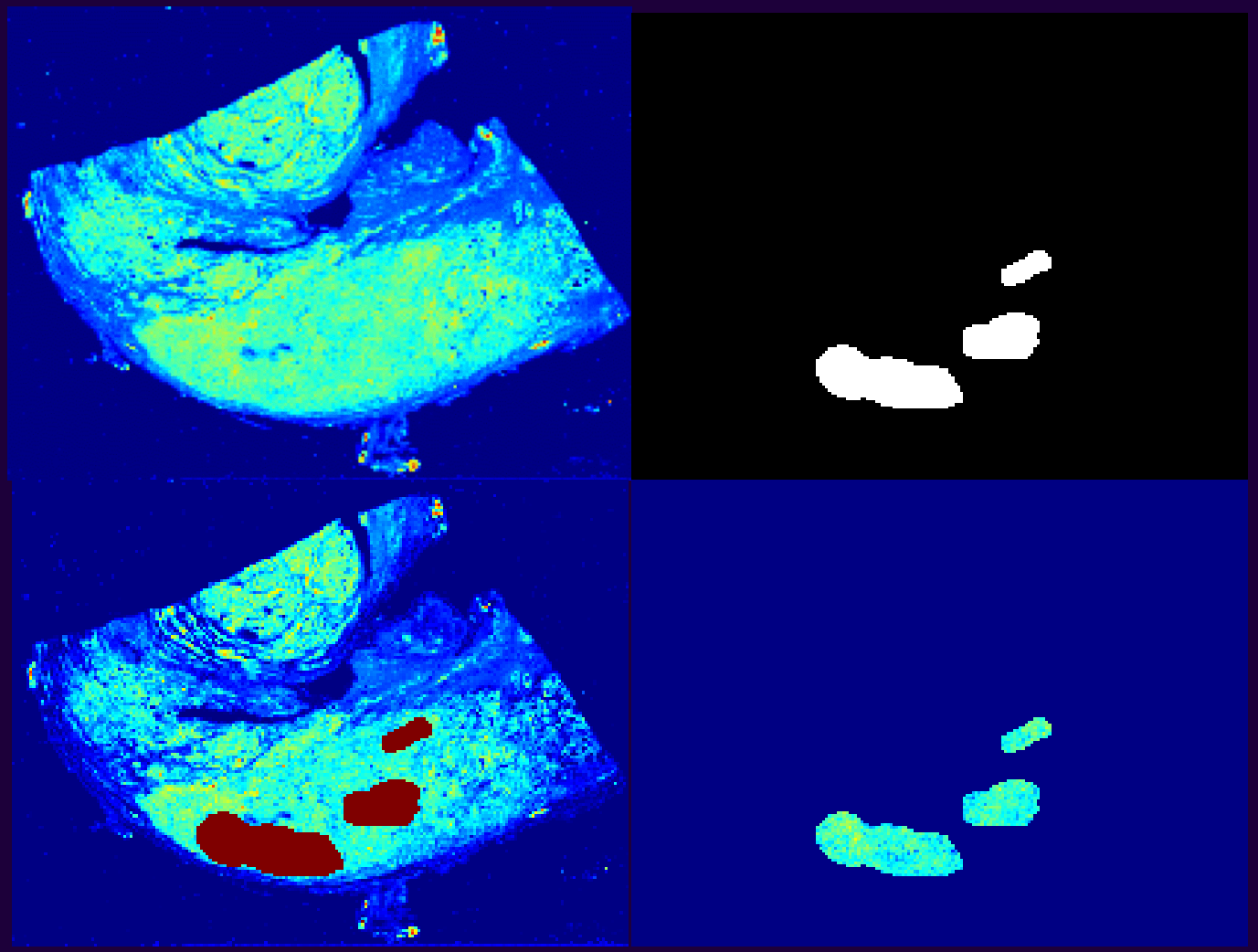


| Believe Treatment Needed: | Treatment Needed: YES | NO |
|---|---|---|
| YES | Crops saved from wipeout by disease at cost of spraying | Unnecessary costs of spraying incurred |
| NO | Crops wiped out and no cost of spraying incurred | Crops safe and no cost of spraying |

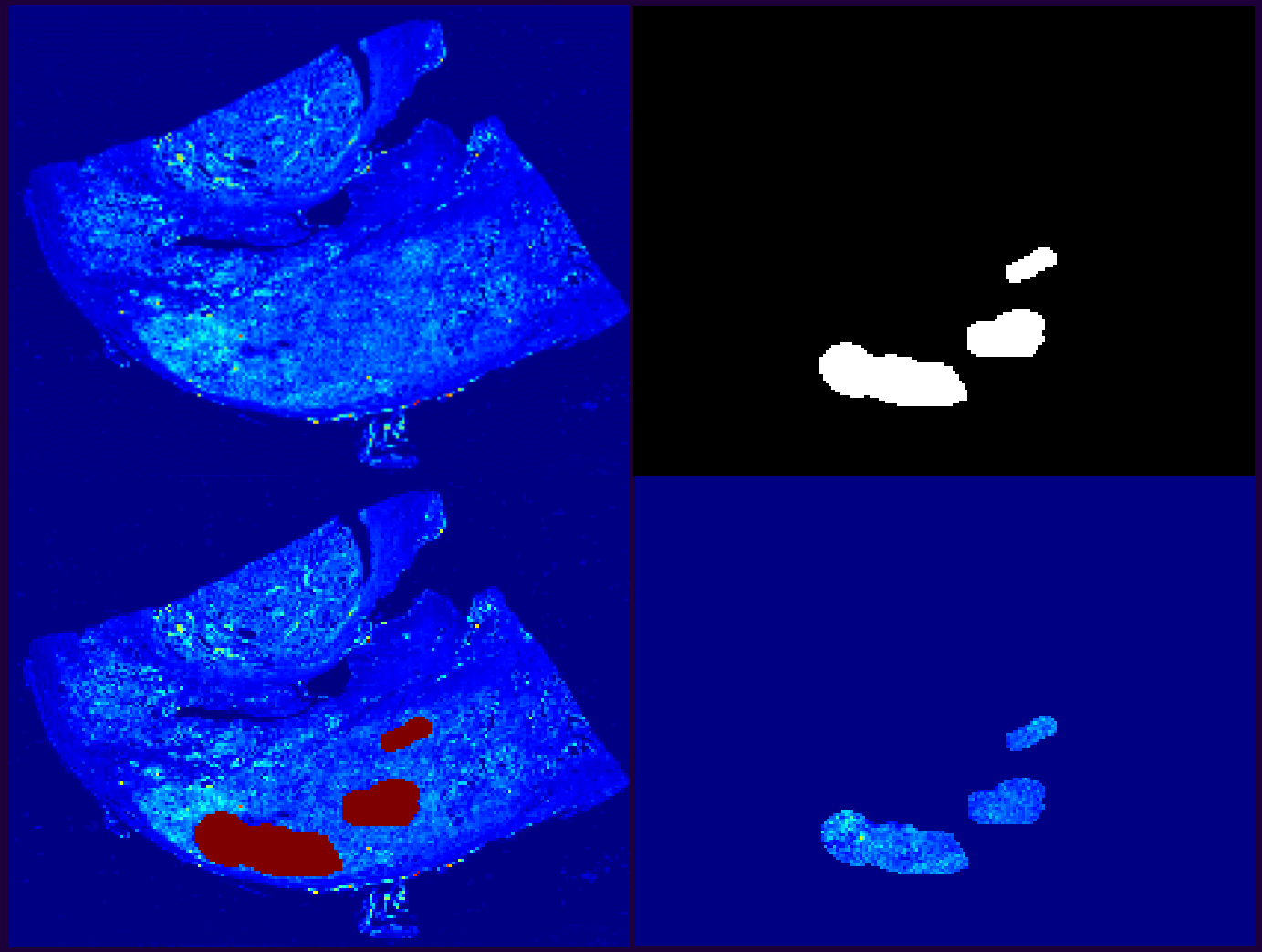
# Tissue Classification Procedures

- Heatmaps were compared to manual tissue annotations of three pathologists
- First reduction of data volume is to measure linear correlation with manual annotation in each heatmap
- Best features retained for further visual and statistical analysis

# Tissue Classification Results



# Tissue Classification Results



# Tissue Classification Results

- No Individual heatmap had a linear correlation greater than 0.25
- Partial Least Squares Regression (PLSR) searches for vectors that correlates with ground truth data
- PLSR on all heatmaps created a Riskmap with increased correlation of 0.35
- Small amount of annotated tumour area impedes identification of strong correlations
- Strongly uneven datasets (Tumour / Not-Tumour) increases risk of trivial solutions

# Improving Tissue Classification Results

- One conservative pathologist was swaying the consensus
- Additional pathologist input with a ‘minus one’ consensus would annotate larger areas
- Annotation of additional samples with emphasis on Aggressive patients would lead to more even datasets
- Quantitative classification of tumours also possible with additional samples

# Patient Stratification Procedures

- The number of potential predictive features is too great so an initial screening step was applied
- 10-fold cross validated PLSR finds the most useful feature vectors and these are used for Discriminant Analysis with full cross validation
- False Negatives and False Positives balanced 10:1 and results are adjusted accordingly

# Patient Stratification Results

- Three way classification (Ind, Sig, Agg)
  - Raw correct classification rate = 73%
  - Adjusted correct classification rate = 74%
- Two way classification (Ind, not Ind)
  - Raw correct classification rate = 87%
  - Adjusted correct classification rate = 78%
- Two way classification (not Agg, Agg)
  - Raw correct classification rate = 83%
  - Adjusted correct classification rate = 89%

# Further Patient Stratification Procedures

- Partial Least Squares Regression (PLSR) with Discriminant Analysis only absorbs linear variability
- Much of the variability could be non-linear and Neural Networks can absorb non-linear trends
- Neural Networks can also absorb noise so a double validation step is used as a safeguard
- Neural Networks can fail to converge

# Patient Stratification Results

- Three way classification (Ind, Sig, Agg)
  - Raw correct classification rate = 77%
  - Adjusted correct classification rate = 78%
- Two way classification (Ind, not Ind)
  - Raw correct classification rate = Failed to Converge
  - Adjusted correct classification rate = Failed to Converge
- Two way classification (not Agg, Agg)
  - Raw correct classification rate = Failed to Converge
  - Adjusted correct classification rate = Failed to Converge

# OVERALL SUMMARY

- Automatic solutions are preferable to Manual solutions
- Pathology rules can be automated and implemented
- Image features can be successfully applied to a clinical cohort
- Models built to identify regions of interest and stratify patients
- Model performances can be objectively quantified

# CONCLUSION

- Integration of Digital Pathology and Statistical Methodology can offer a viable and ergonomic tool to reduce the burden on histopathologists

# Let Us Meet Again

We welcome you all to our future conferences of OMICS International

Please Visit:

http://www.omicsonline.org/

www.conferenceseries.com

http://www.conferenceseries.com/pathology-meetings