

## V6 - Digital pathology and MRI diagnostics

**Pathology** (from the Greek roots of *pathos* (πάθος), meaning "experience" or "suffering", and *-logia* (-λογία), "study of") is a significant component of the causal study of disease and a major field in modern medicine and diagnosis.

**Digital pathology (DP)** includes all aspects of

- acquisition,
- process management, and
- data interpretation

to yield pathology information from a digitized pathology sample's image.

[www.wikipedia.org](http://www.wikipedia.org)

Bhargava, Madabhushi

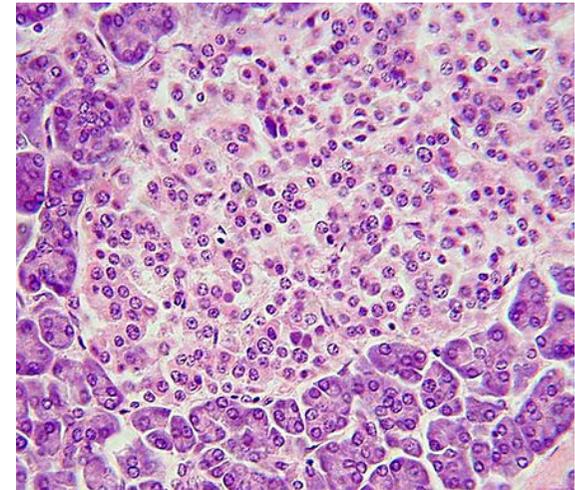
Annu. Rev. Biomed. Eng. 2016. 18:387–412

## Include biological or chemical markers or tissues

Staining with **hematoxylin and eosin** (H&E) involves application of hemalum, a complex formed from aluminum ions and hematein.

Hemalum colors nuclei of cells (and a few other objects) **blue**. The nuclear staining is followed by counterstaining with an aqueous or alcoholic solution of **eosin Y**.

This solution colors eosinophilic structures in various shades of **red**, **pink** and **orange**.



Also used are:

- Immunohistochemical (IHC) imaging
- label-free methods for pathology use spectral imaging.
- Direct recording of chemical composition. This eliminates the need for dyes or stains

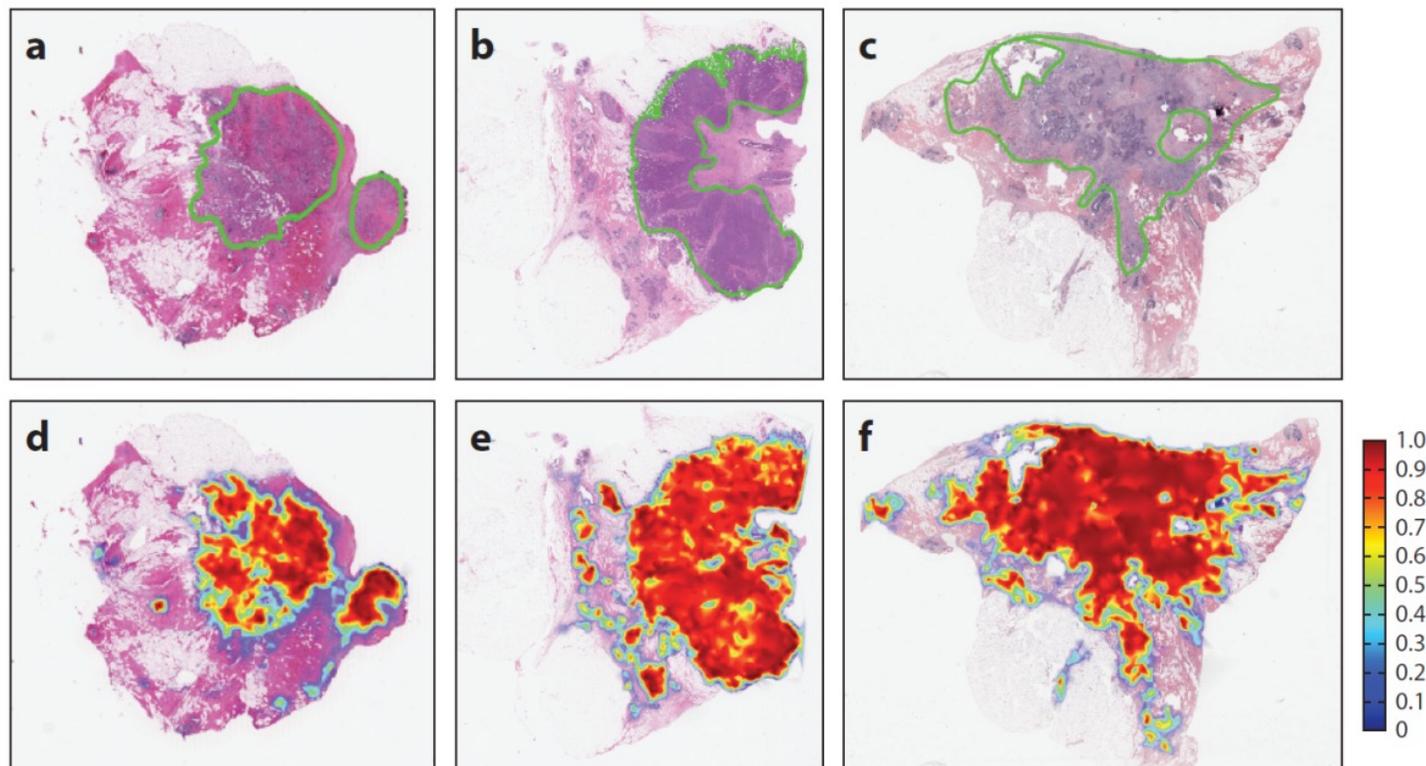
[www.wikipedia.org](http://www.wikipedia.org)

Bhargava, Madabhushi

Annu. Rev. Biomed. Eng. 2016. 18:387–412

## Digital pathology

(*Top*) Probabilistic output of a deep learning classifier for regions of invasion.



(*Bottom*) Corresponding hematoxylin and eosin images with a pathologist's markup of the extent of cancer extent. Note the concordance between the two rows.

Bhargava, Madabhushi  
Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Quantitative histomorphometry

Quantitative histomorphometry (QH) involves computerized image analysis tools for quantitatively assessing cancer tissue and non-cancer tissue morphology and architecture.

QH measurements can be divided broadly into three groups:

- architectural,
- shape, and
- texture based.

# Architectural QH measurements

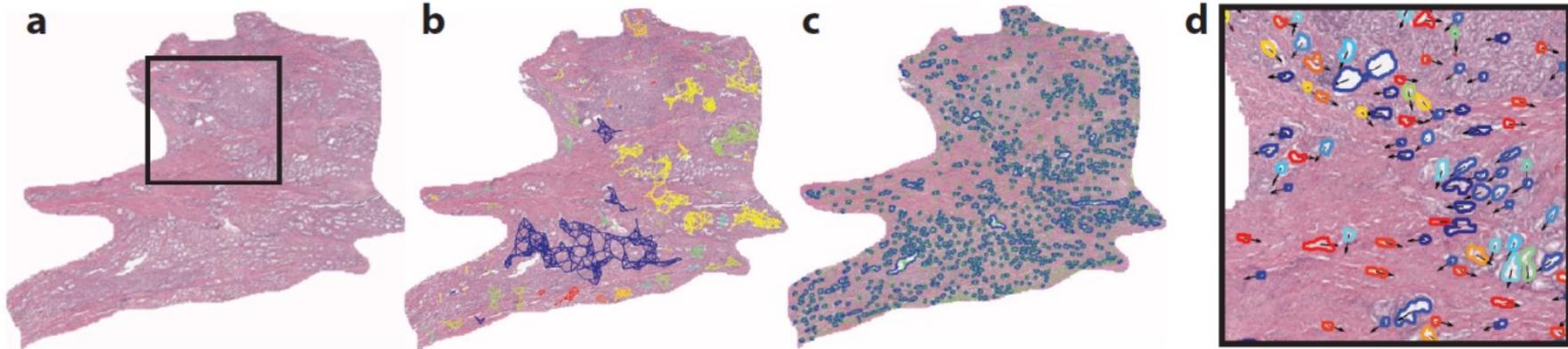
Architectural features capture the arrangement and **spatial topology** of histologic primitives such as individual nuclei, tubules, mitoses, and lymphocytes.

The spatial location of a particular primitive is considered to be a node in a graph.

The nodes are then connected using graph construction algorithms [e.g., Voronoi, Delaunay, minimum spanning tree].

Quantitative measurements (e.g., internode distance, clustering coefficient of the nodes = density of links between the neighbors of a node) can quantitatively characterize the graph and, hence, the image.

## Global and cell cluster graphs



(a) Prostate cancer tumor region. The region of interest (ROI) is outlined in blue.

(b) Cluster graphs establish localized gland networks.

(c) Delaunay triangulation reveals a global graph which traverses stromal and epithelial boundaries, whereas co-occurring gland tensors compute localized features from the gland networks.

(d) The ROI from panel a. The color map of the gland orientations ( $0^\circ$ ,  $180^\circ$ ) demonstrates the variation in local gland orientation. Gland orientations are architecturally differently arranged in tissue from patients with and without disease recurrence.

Bhargava, Madabhushi

Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Shape QH measurements

The **shape** of individual histologic primitives can indicate the presence of disease.

Shape features such as

- fractal dimension: ratio comparing how a detail in a pattern changes with the scale at which it is measured
- angularity, size, and
- smoothness of the boundary

differ between nuclei and glands in high and low grades of prostate and breast cancers.



As the length of the measuring stick is scaled smaller and smaller, the total length of the coastline measured increases (-> fractal dimension)

Also, the disorder (or entropy) in the orientation of nuclei and glands in prostate tissue is related to the tumor recurrence in patients with prostate cancer.

[www.wikipedia.org](http://www.wikipedia.org)

## Texture-based QH measurements

Texture refers to quantitative measures of spatial neighborhood interactions between pixel intensities within local neighborhoods in an image.

These could include first-order spatial intensity interactions (e.g., mean, standard deviation, median, variance) within local neighborhoods and second-order interactions (e.g., co-occurrence features).

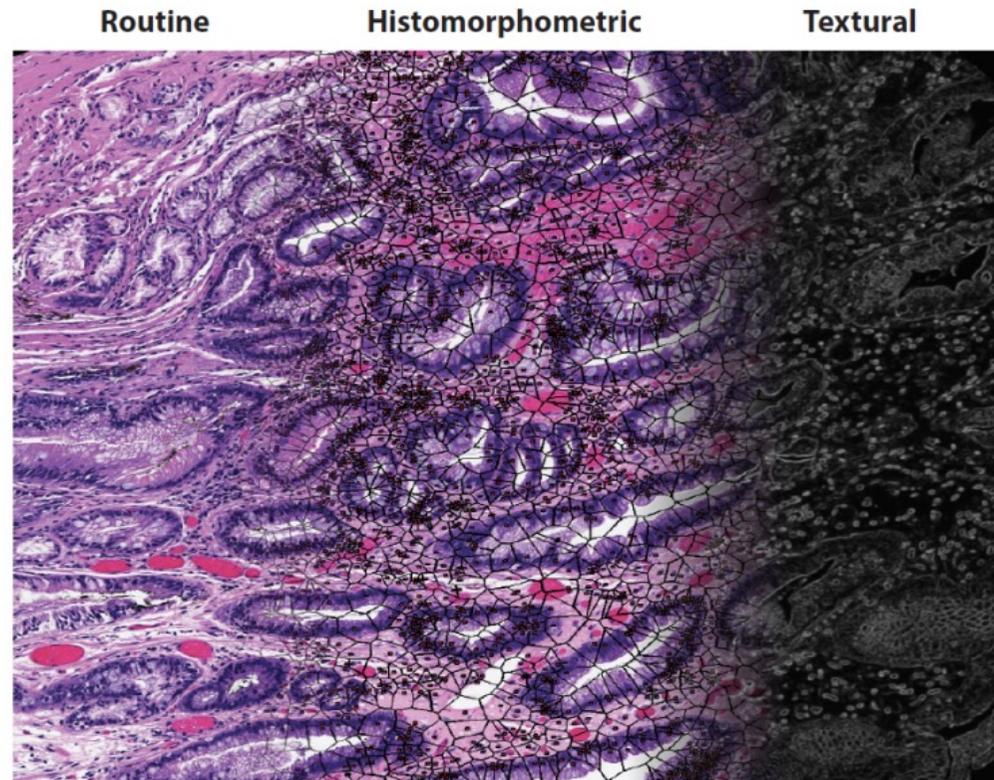
More complex textural features can also be extracted; these include steerable and multiscale gradient features via mathematical operators such as Gabor filters, local binary patterns, and Laws filters.

The shape and texture of nuclei within the stroma are significantly correlated with disease recurrence and patient outcome in breast, prostate, and oropharyngeal cancers.

## A digital stain

(Left) A routine hematoxylin and eosin tissue image.

The left image can be converted into a histomorphometric representation comprising nuclear architecture (*middle*) and textural measurements (*right*).



The figure shows the digital stain representation of a routine H&E image, with overlays of nuclear architecture networks and capture of stromal and epithelial textural variations.

Bhargava, Madabhushi

Annu. Rev. Biomed. Eng. 2016. 18:387–412

## Analysis of digitized images

A typical analysis pipeline involves a **machine learning classifier** that takes as input a series of manually or computer-extracted features and employs those features to render a prediction.

In the context of digital pathology, **predictions** might involve

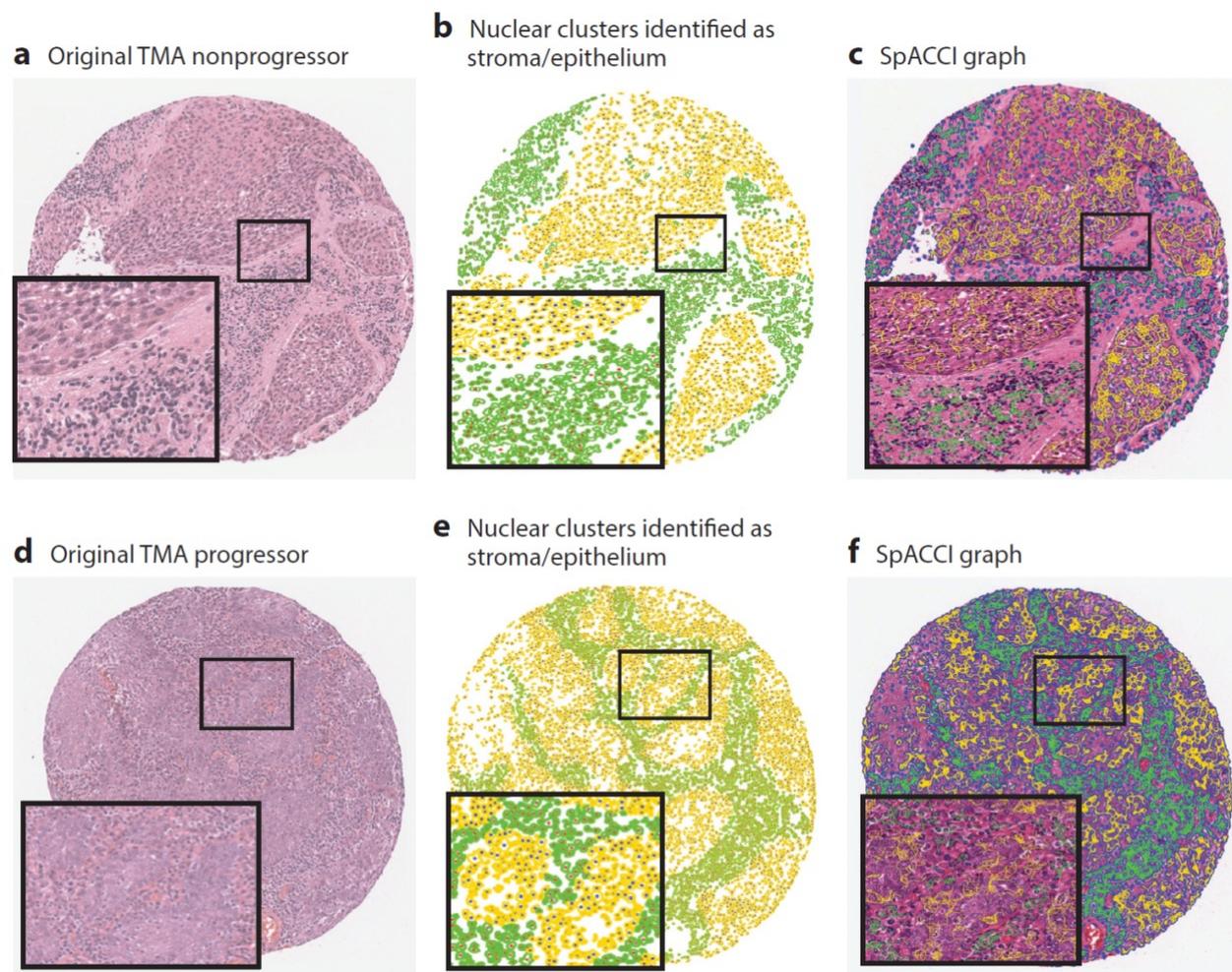
- a low-level recognition (e.g., Is the primitive a nucleus or not?),
- a diagnostic decision (e.g., Is the tissue region of interest cancerous or not?),
- or a prognostication (e.g., Will the patient have early or distant disease recurrence?).

# Digital pathology

Nuclear architecture features can be extracted within the stromal and epithelial compartments within hematoxylin and eosin tissue sections.

The combination of stromal and epithelial nuclear architecture features, referred to as a **SpaCCI graph**, enables improved prediction of which p16+ oropharyngeal cancers will and will not progress.

TMA: tissue microarray.



Bhargava, Madabhushi  
Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Principles of chemical imaging

IR imaging provides high image contrast, fast data recording, and high molecular sensitivity.

Vibrational frequencies within molecules directly resonate with optical frequencies in the mid-IR spectral region.

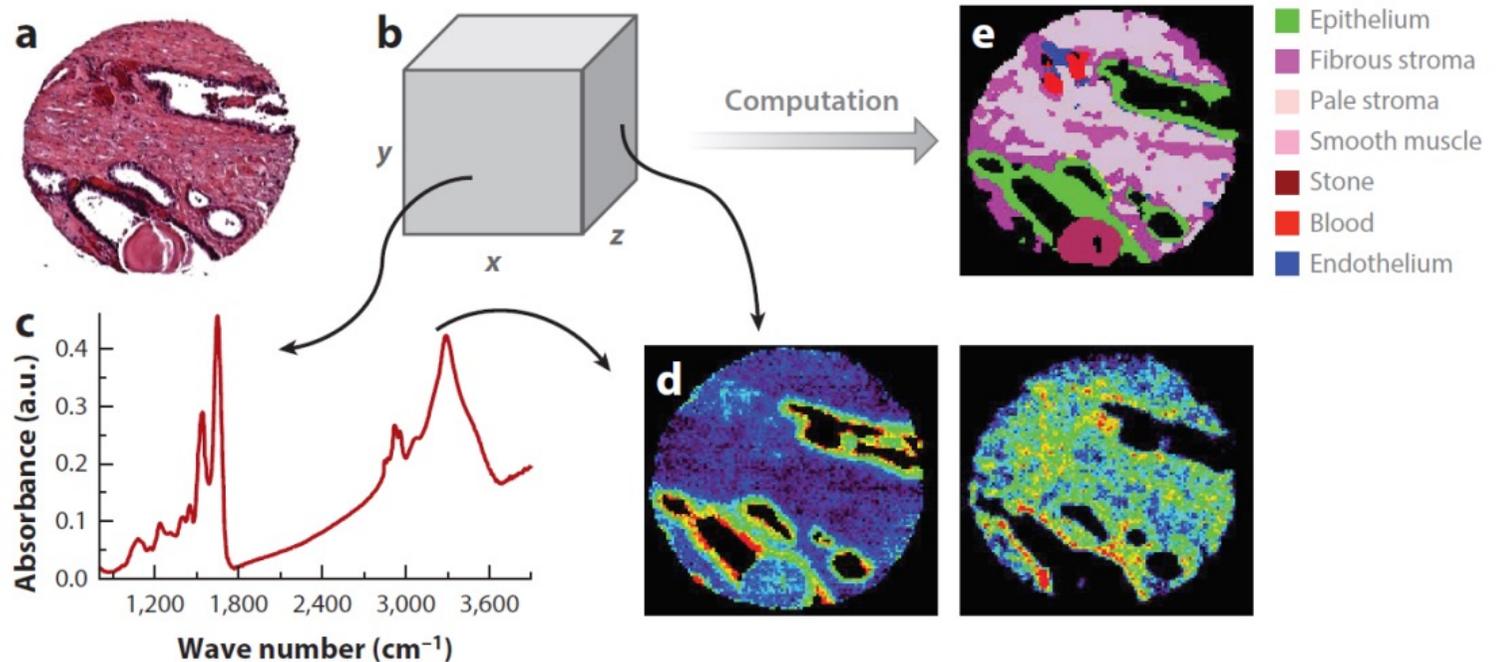
Thus, light absorption provides a quantitative molecular fingerprint of the material, providing ample molecular biomarkers.

No dyes or stains are needed to visualize molecular content, so data can be recorded from a variety of samples without prior knowledge of the type or composition of the sample.

Thereafter, informatics techniques are used to extract the desired information or to discover new information.

# Chemical imaging

(a) Conventional imaging in pathology requires eyes and a human to recognize cells.



(b) In chemical imaging data, both (c) a spectrum at any pixel and the spatial distribution of any spectral feature can be observed, as in (d, left) nucleic acids (at  $\sim 1,080 \text{ cm}^{-1}$ ) and (right) collagen (at  $\sim 1,245 \text{ cm}^{-1}$ ).

(e) Computational tools can then translate the chemical imaging data into knowledge used in pathology.

Bhargava, Madabhushi  
Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Comparison of molecular and chemical imaging

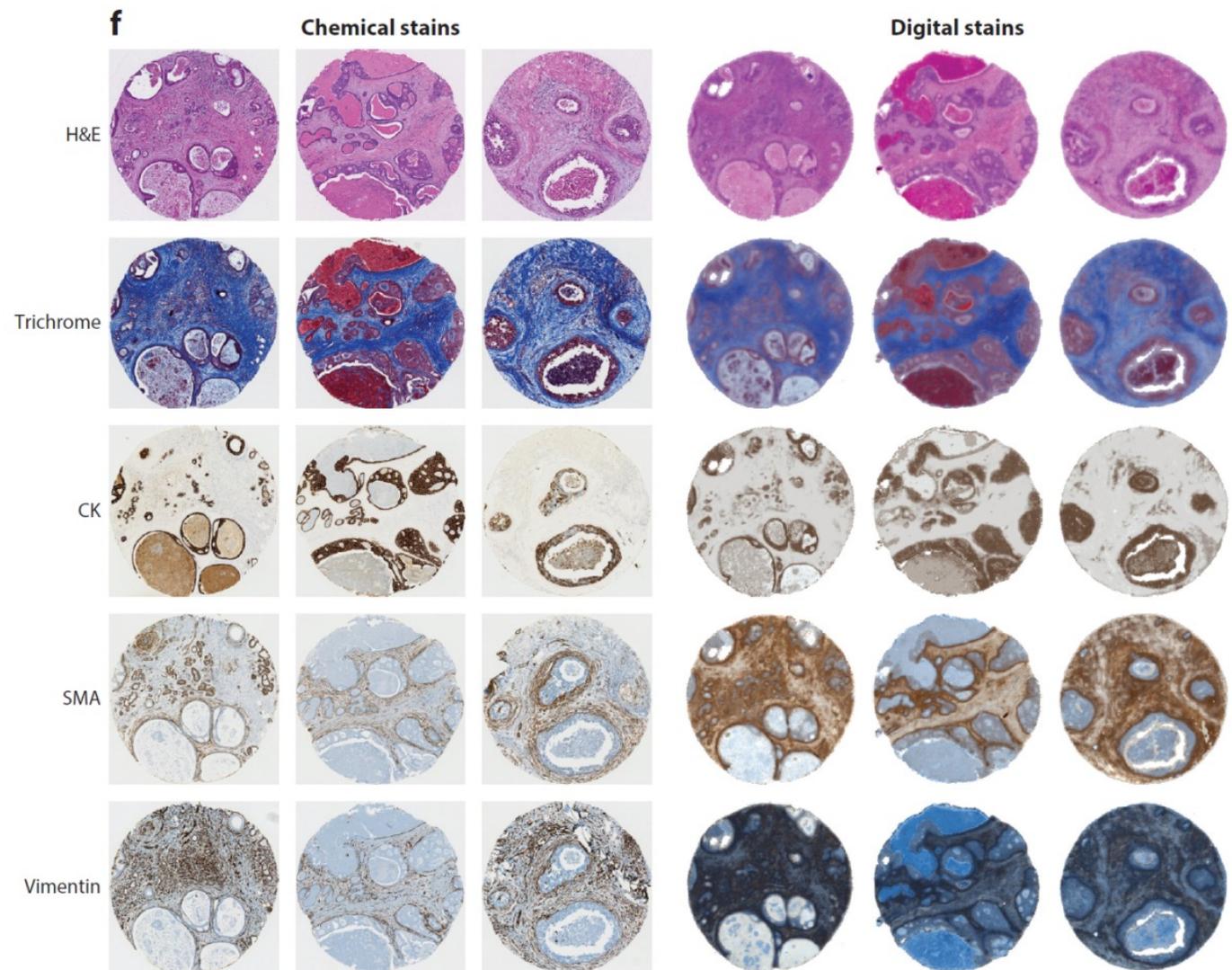
Molecular imaging (*left*)  
can be reproduced by  
chemical imaging (*right*).

Abbreviations:

CK, cytokeratin;

H&E, hematoxylin and  
eosin;

SMA, smooth muscle  $\alpha$ -  
actin.

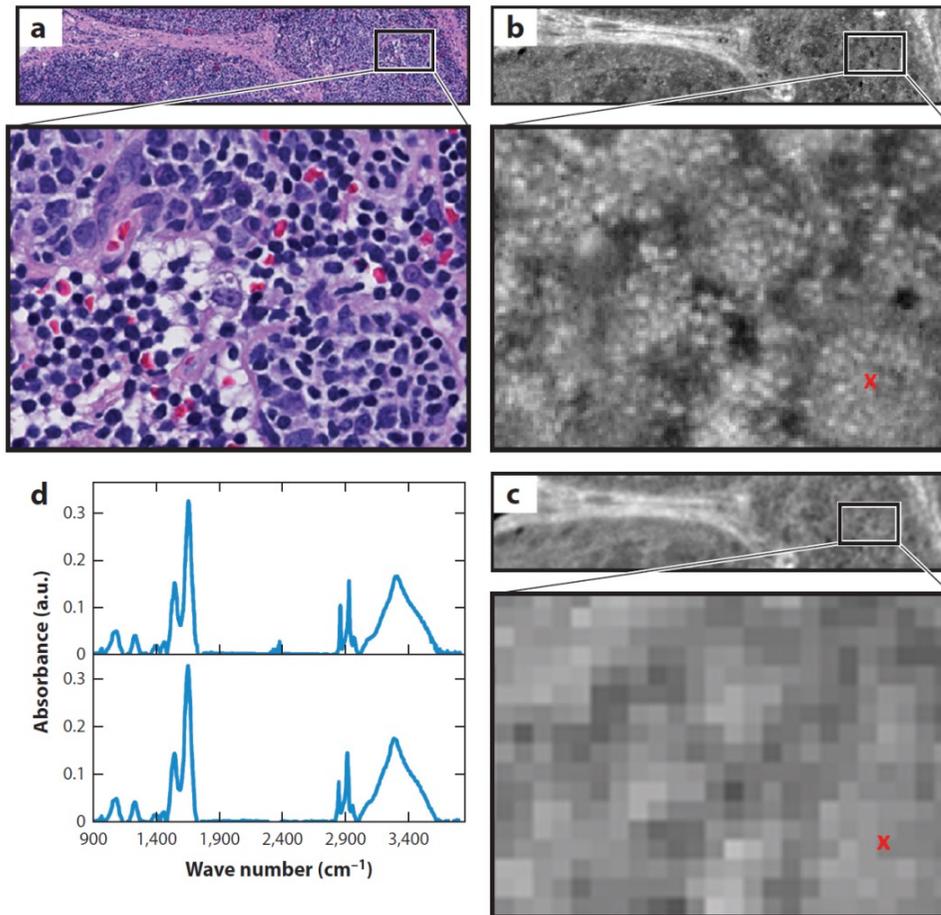


Bhargava, Madabhushi

Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Comparison of H&E stain and IR imaging

Comparison of hematoxylin and eosin (H&E)-stained optical microscopy and infrared (IR) images of lymph node tissue.



(a) An H&E-stained image from a healthy lymph node biopsy.

(b) A high-definition IR image of a serial section of the lymphoid tissue.

(c) The same region imaged with a lower-resolution Fourier transform IR (FT-IR) imaging spectrometer. The IR images show the absorbance at 3,075 cm<sup>-1</sup> after baseline correction.

(d) Sample spectra plotted from the pixel marked with a red x in panel c.

There is a slight discordance between the H&E and IR images because they are on different tissue sections.

Bhargava, Madabhushi  
Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Combine multiple data sources

Tumors with similar morphologic phenotypes may have significantly different behaviors and outcomes.

Combination of multiple, independent sources of clinical, molecular, and pathological data can provide more predictive power,

**Table 1** Open-source and commercial image analysis platforms for image analysis and biomarker quantification from digital pathology and multiplexed immunohistochemical (IHC) images

Digital pathology software	Capabilities and features
Precision (Aperio)	Stain separation, cell counting, separation of broad tissue classes
inForm (Caliper)	Separation of broad tissue classes
Virtuoso (Ventana)	Quantification of IHC staining
Halo (Indica)	Segmentation, analysis
Visiopharm	Segmentation, feature analysis of angiogenesis, bone growth, neural structure; cancer detection
TissueStudio	Image analysis of IHC staining
Developer XD (Definiens)	Separation of broad classes and features on the basis of shape, morphology, color of segmented objects
FARSIGHT	Optical microscopy applications, focusing on fluorescence
ImageJ	Java image analysis package designed with an open architecture to provide extensibility via plugins; recent push to integrate digital pathology images

Bhargava, Madabhushi

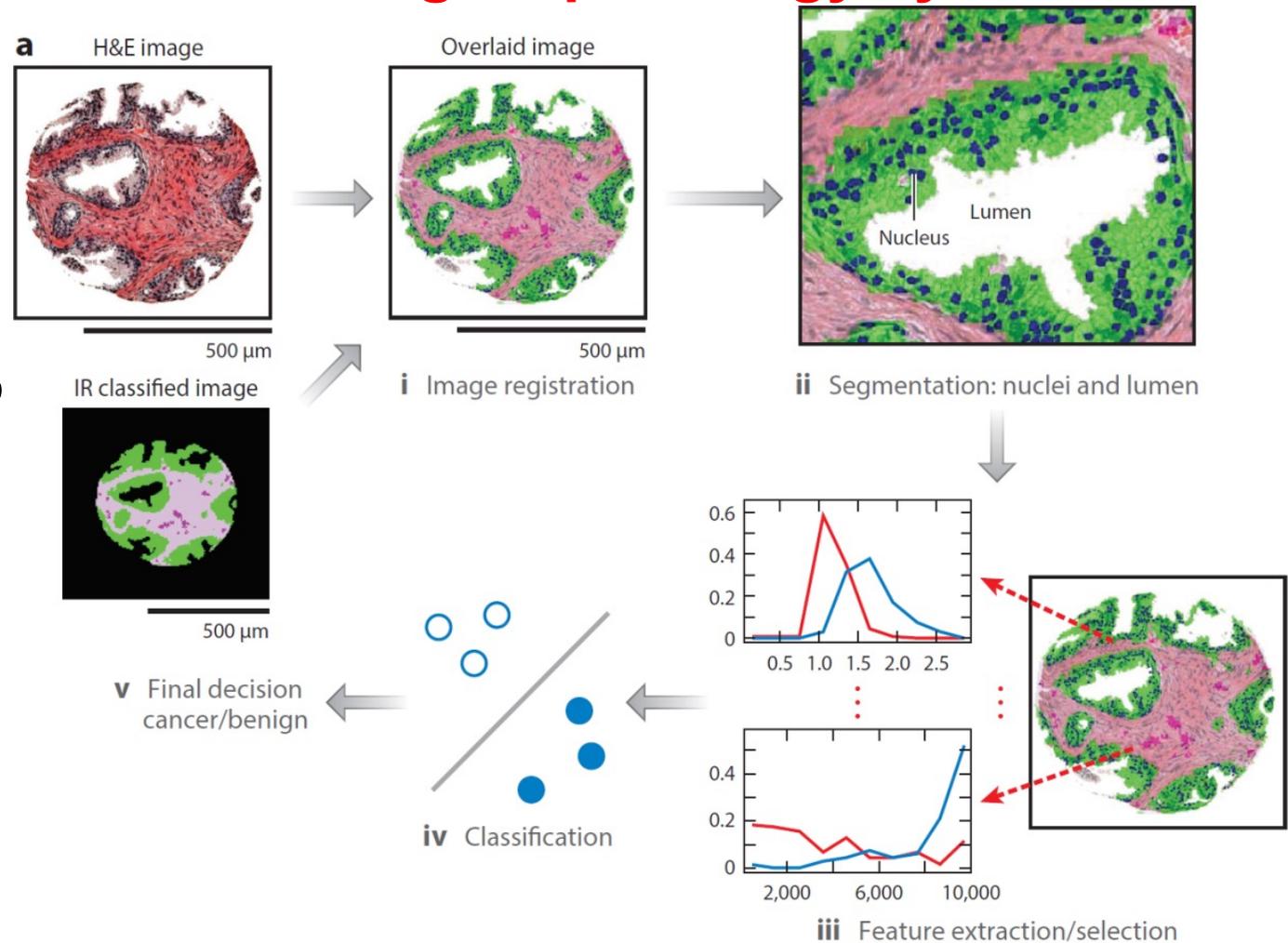
Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Overview of multimodal digital pathology system

(i) A Fourier transform infrared spectroscopy data-based cell type classification is overlaid on a hematoxylin & eosin-stained image, leading to

(ii) segmentation of nuclei and lumens in a tissue sample.

(iii) Features are extracted and selected, then (iv) used by the classifier to (v) predict whether the sample is cancerous or benign.

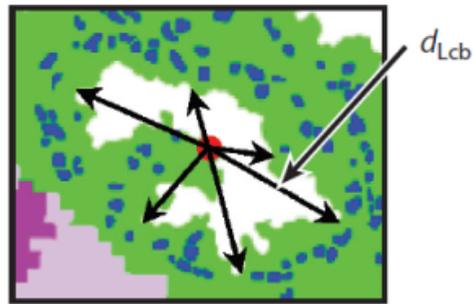


Bhargava, Madabhushi  
Annu. Rev. Biomed. Eng. 2016. 18:387–412

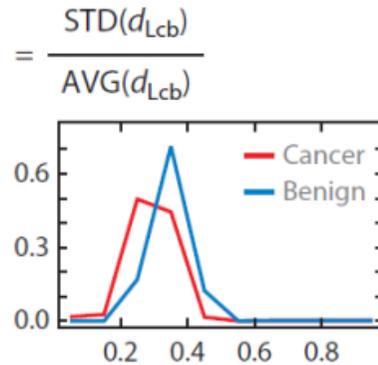
# Example features

Large ratio: tumor fills bounding circle better

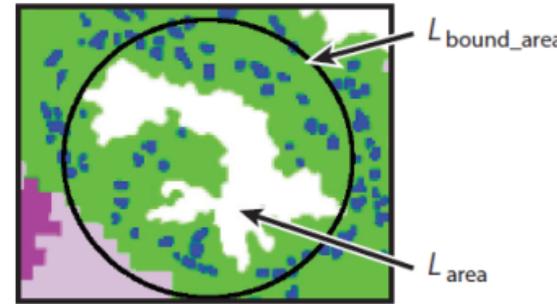
**b i** Lumen distortion



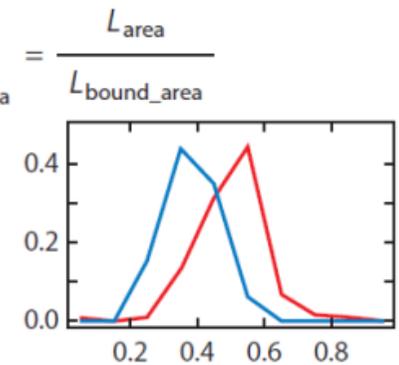
**distortion(lumen)**



**ii** Minimum bounding circle ratio

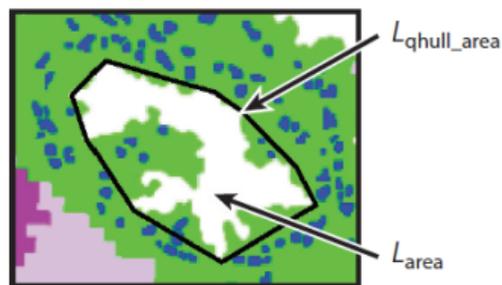


**bound\_ratio(lumen)**

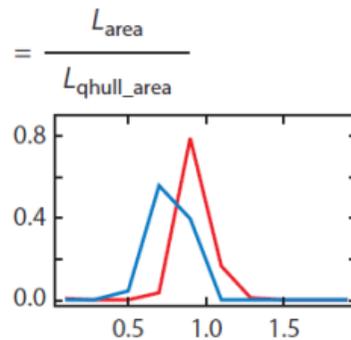


Each panel shows one feature, along with the distributions of the feature's values for cancer (*red*) and benign (*blue*) classes.

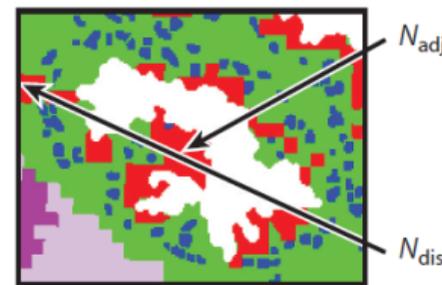
**iii** Convex hull ratio



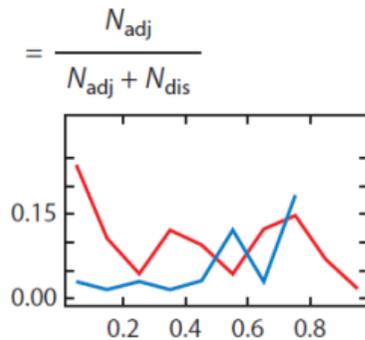
**qhull\_ratio(lumen)**



**iv** Spatial association: lumen and cytoplasm



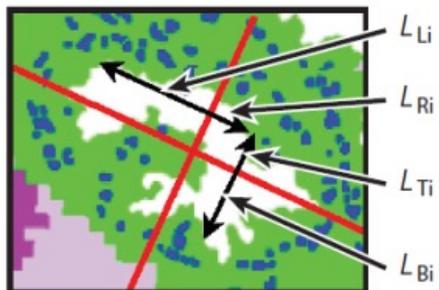
**assoc(lumen,cytoplasm)**



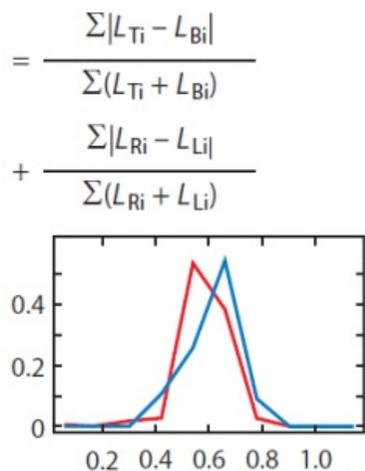
Bhargava, Madabhushi  
Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Digital pathology

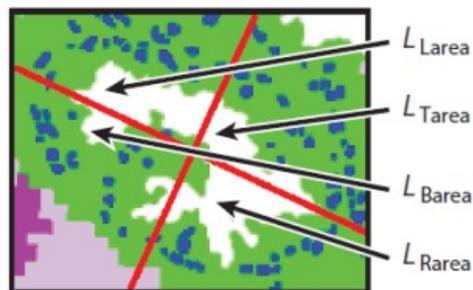
v Symmetric index of lumen boundary



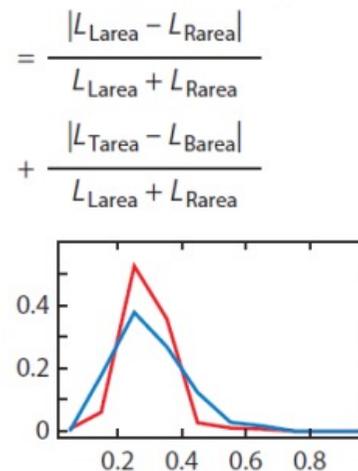
boundary\_sym(lumen)



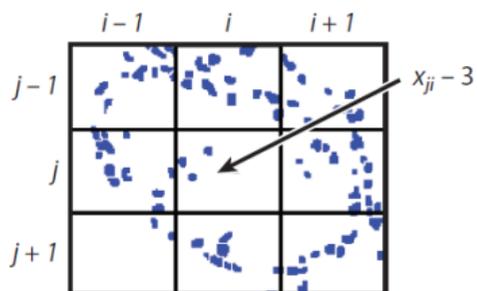
vi Symmetric index of lumen area



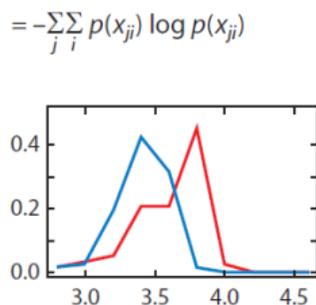
area\_sym(lumen)



vii Entropy of nuclei distribution



H(nuclei)

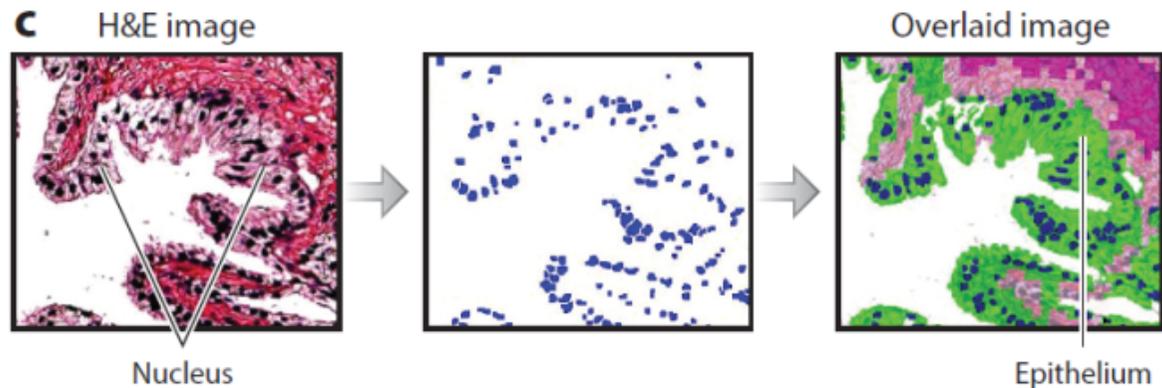


Bhargava, Madabhushi  
Annu. Rev. Biomed. Eng. 2016. 18:387–412

# Classification of tumor tissue

IR and H&E images can be overlaid with an automated alignment algorithm.

The features allow better classification of cancer than does H&E staining alone.



AUC, area under the curve;

AVG, average;

STD, standard deviation

	10 CV (Data 1)		10 CV (Data 2)		Validation	
	AUC		AUC		AUC	
	AVG	STD	AVG	STD	AVG	STD
<b>IR and H&amp;E</b>	0.982	0.0030	0.974	0.0145	0.956	0.0089
<b>H&amp;E only</b>	0.968	0.0052	0.880	0.0175	0.918	0.0100

10 CV: 10-fold cross validation

## Example: Wilms tumor

**Wilms tumor**, also known as **nephroblastoma**, is a cancer of the kidneys that typically occurs in children, rarely in adults.

It is named after Dr. Max Wilms, a German surgeon (1867–1918) who first described it.

Approximately 500 cases are diagnosed in the U.S. annually (rare tumor).

The majority (75%) occur in otherwise normal children; a minority (25%) are associated with other developmental abnormalities.

Wilms tumor is highly responsive to treatment, with about 90% of patients surviving at least five years.

Diagnose tumor e.g. with MRI scan:

This is a sort of NMR experiment.

Measure T1 and T2 spin relaxation times of tissues.

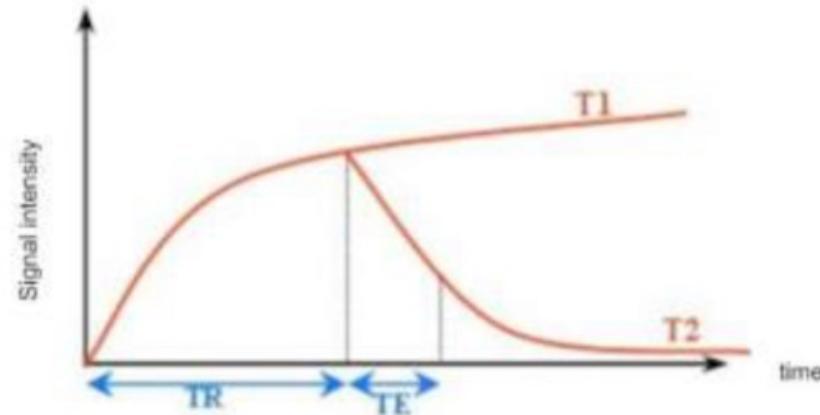


Figure 1: T1 and T2 weighting times [R21]

## Non-invasive MRI diagnostics: data sets

Vera Bazhenova (MSc Comp Sci UoS 2014) analyzed vertical cross section MRI sets of scans for patients with nephroblastoma tumor.

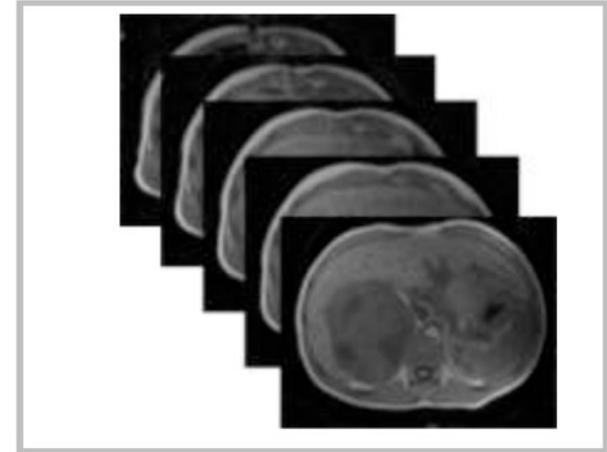
Each set contains 20 and 50 scans.

The following part of this lecture was taken from her MSc thesis.

### **Aim of this project:**

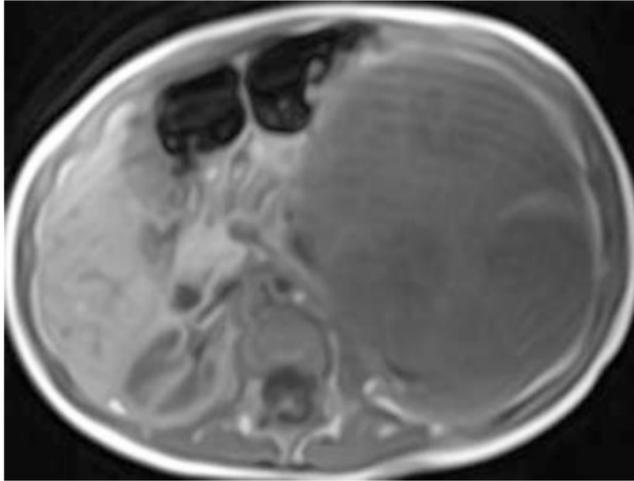
Identify precise **location** of the tumor.

This can be basis for surgery (where to operate?) or be used for diagnostic purposes (follow tumor growth).

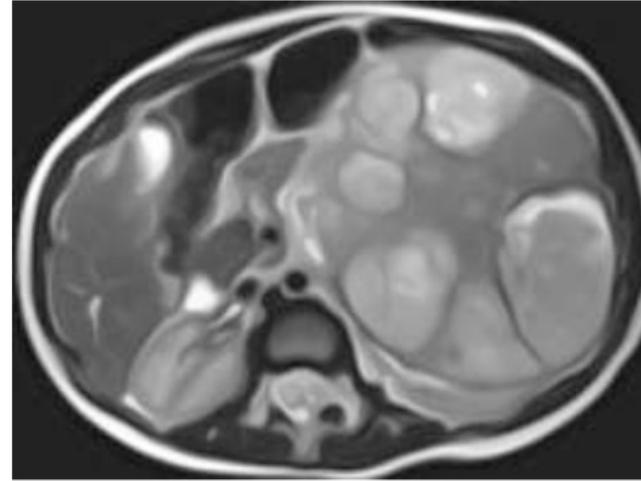


## Input data

T1-weighted scan



T2-weighted scan



T1-weighted scans appear more suitable for digital analysis since the tumor region has a more homogeneous contrast.

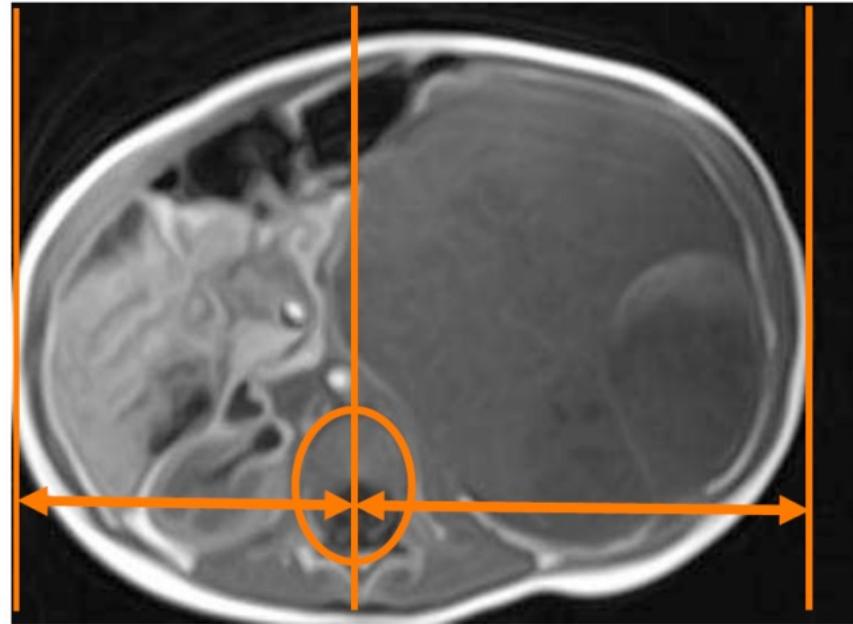
The body contours are well visible and can be easily distinguished from the background.

## Use spine location to detect asymmetry

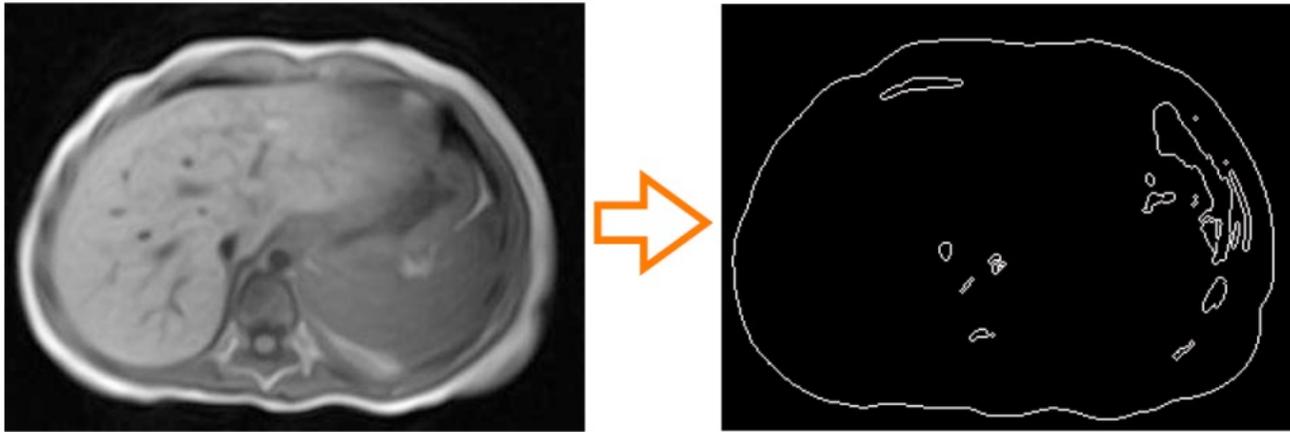
The Nephroblastoma tumor affects in more than 95% of the cases only one kidney of the patient.

In healthy individuals, the spine is located in the center of body cross section.

When the affected kidney grows abnormally, the spine appears shifted either to the left or to the right side.



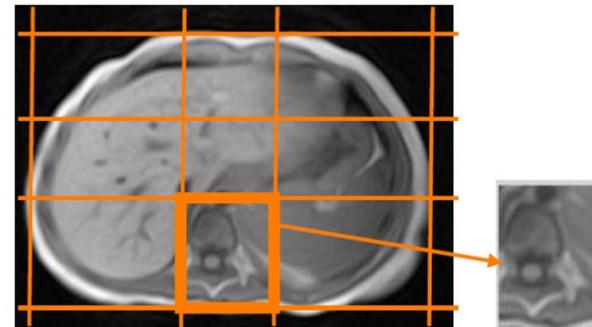
## Determine perimeter



To locate the spine region, the body boundary is detected using a **perimeter detection function** applied to a binary image.

A pixel is considered as a part of the perimeter if it has a nonzero brightness and it is connected to at least one zero-valued pixel.

The “region of interest” for the spine is vertically located in the middle third and horizontally in the lower third of the body.

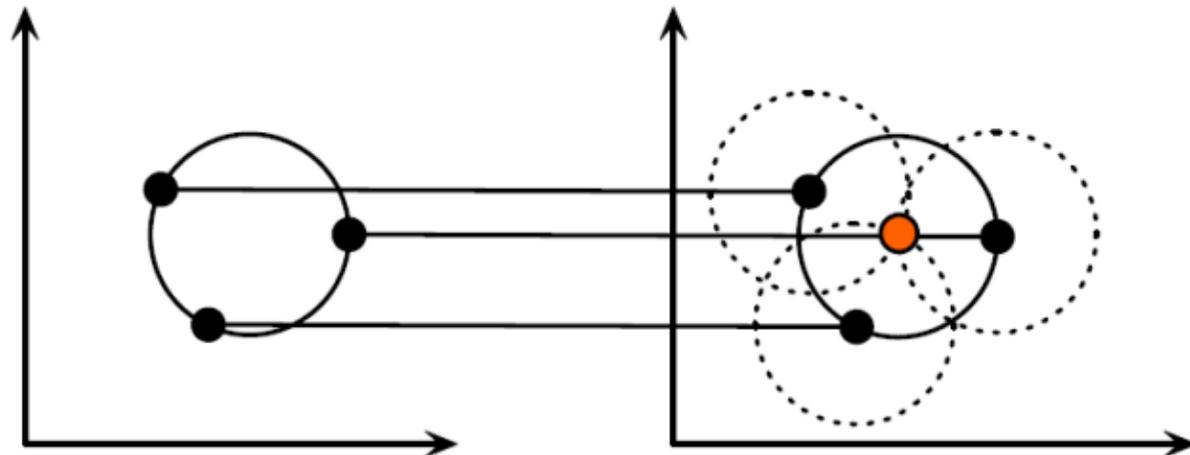
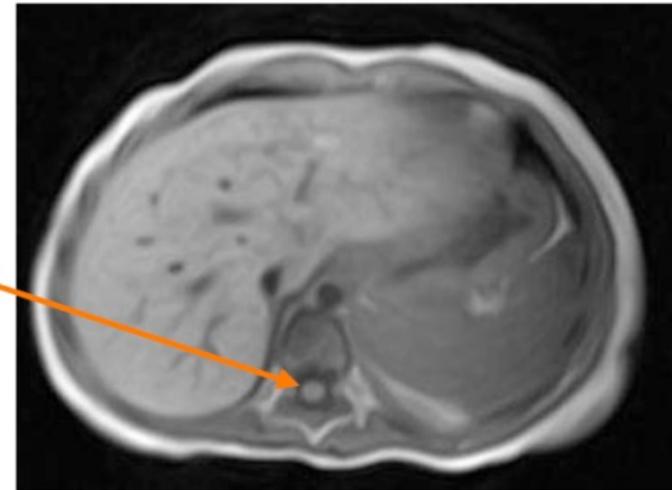


## Task: automatic detection of spine

In a T1-weighted MRI scan, the middle of the spine  $M_s$  appears as a white circle at the level of the liver.

→ Apply the **circular Hough transform** to the first scans of a series until a spine center is detected.

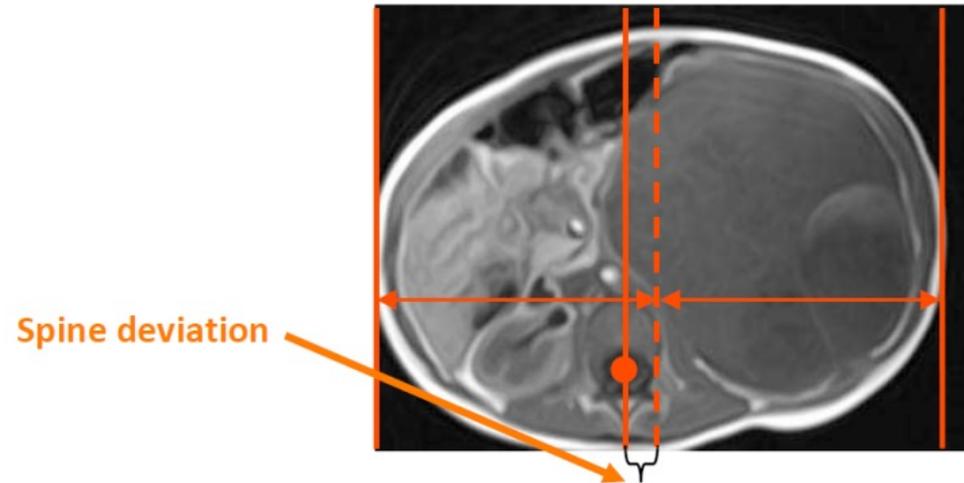
Circular shape  
to be detected  
automatically



Each point in geometric space (left) generates a circle in parameter space (right). The circles in parameter space intersect at the  $(a, b)$  that is the center in geometric space.

[https://www.cis.rit.edu/class/simg782/lectures/lecture\\_10/lec782\\_05\\_10.pdf](https://www.cis.rit.edu/class/simg782/lectures/lecture_10/lec782_05_10.pdf)

## Spine position



Detect the spine middle in all scans of the MRI series.

A patient who shows a significant **deviation** of the spine from the center is flagged as candidate to have a certain class of diseases including a Nephroblastoma tumor.

The **direction** of the deviation indicates to us which side of the body is likely affected by this disease.

## Masked scan

If a disease is present, we prepare a **body mask** that hides

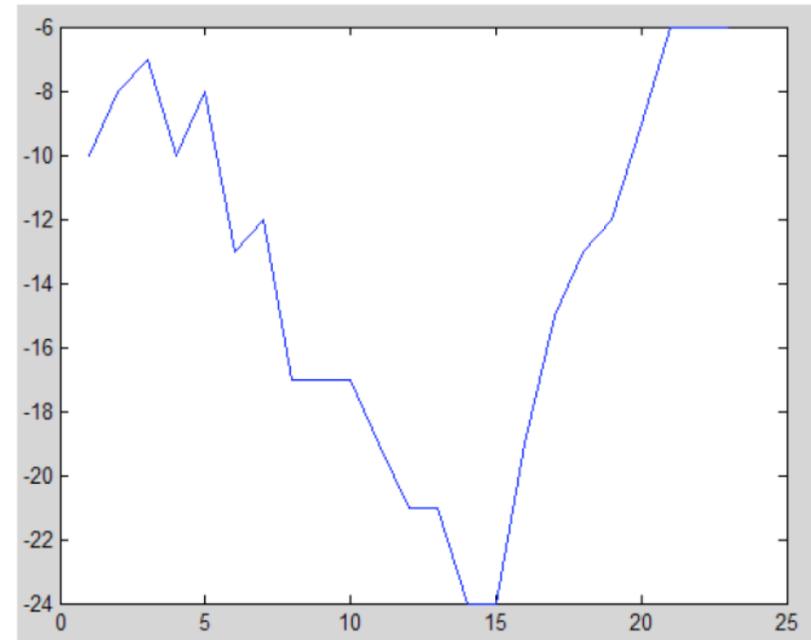
- the spine (1),
- the region below the spine (2),
- the body perimeter (3) and
- the side which presumably does not contain a tumor.



## Spine deviation curve

Another output of the spine detection algorithm is the index of the scan with the **maximum deviation** of the spine from the center.

This index is used in order to extract the gray value range of the tumor in order to enhance the accuracy of the tumor recognition algorithm.



The figure shows the spine deviation curve for a real MRI scan.

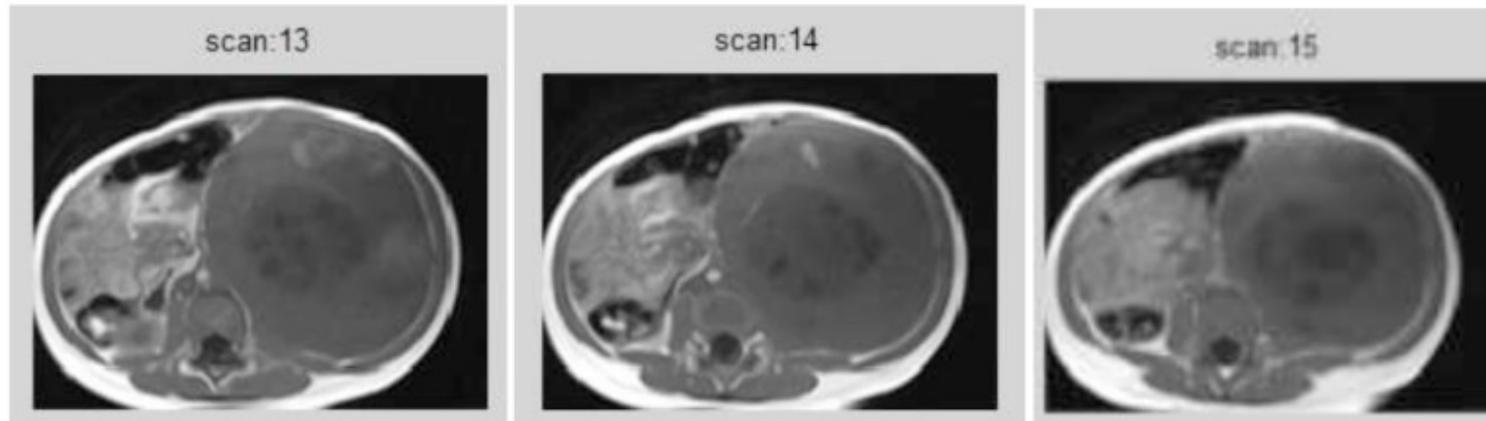
According to the coordinate system adopted here, a negative deviation means that a disease occurs on the right side of the body

# Tumor detection

Detection of the tumor is performed in two main steps.

In the first step, the tumor gray value range is determined.

In the second step, the precise region of the tumor is detected.



- Use the scan with the largest deviation
- Identify the largest blob. Even if the liver is on the same side as the tumor, the tumor is likely already larger than the liver.

## Image denoising

The delivered MRI scan series are usually quite noisy and need to be pre-processed in order to be suitable for detecting the tumor.

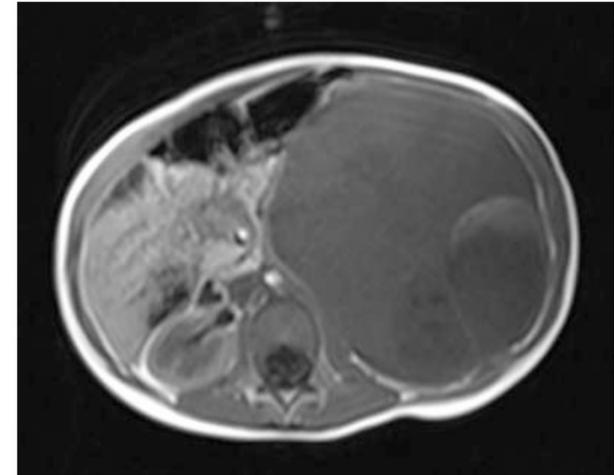
For this, diffusion filtering is used. This denoising algorithm removes noise while it preserves edges.

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -\nabla(-D \nabla \rho(\vec{r}, t)) = D \Delta \rho(\vec{r}, t)$$

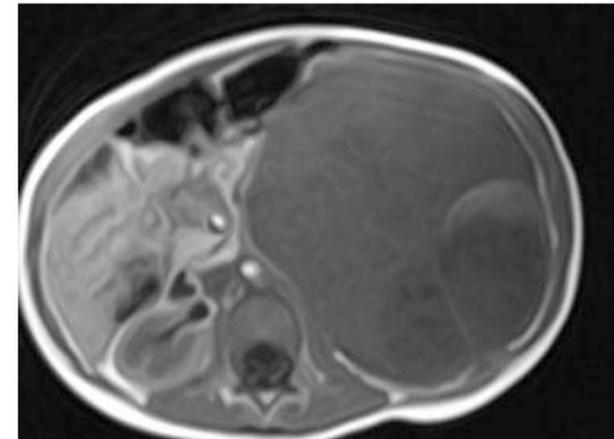
Illustrate diffusion equation.

The above “diffusion equation” is applied iteratively to an input image until the output becomes smooth enough and reaches the wished noise elimination.

In addition, other filters are applied, e.g. the median filter



a. Original MRI scan



b. Filtered MRI scan

# Denoising: median filter

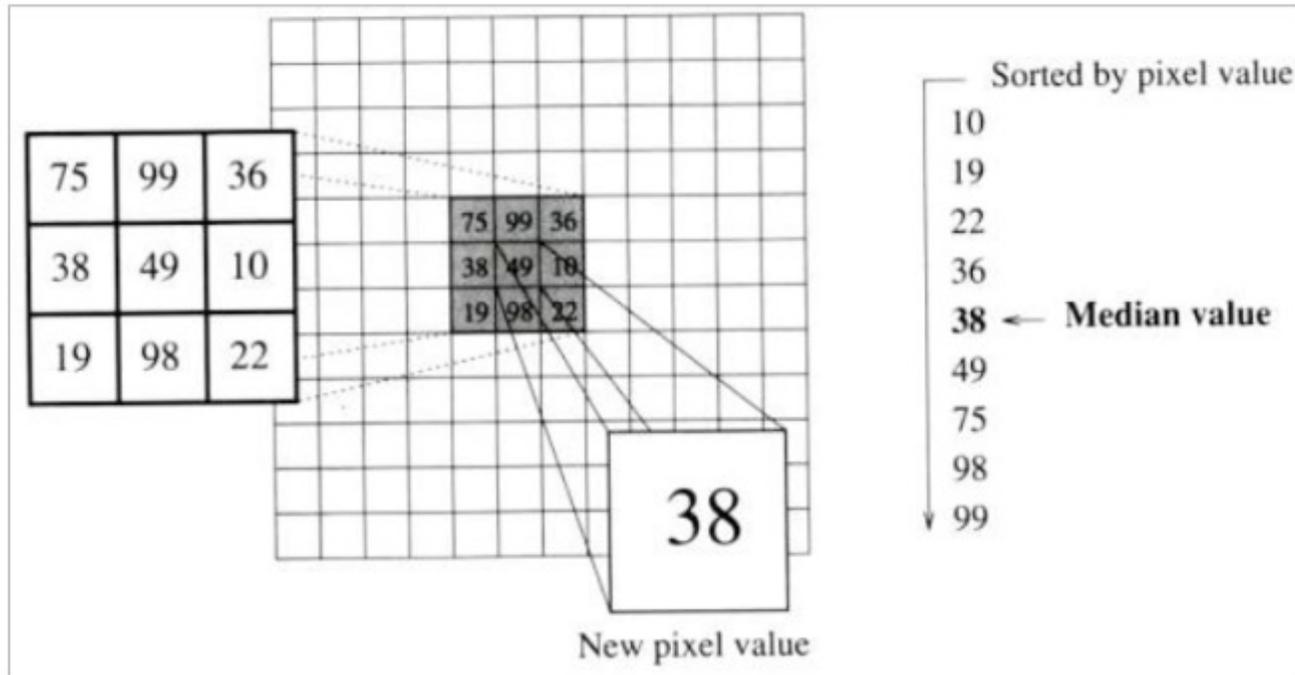


Figure 2: Median filter [R24]

## Determine gray levels

Apply edge enhancement filter.

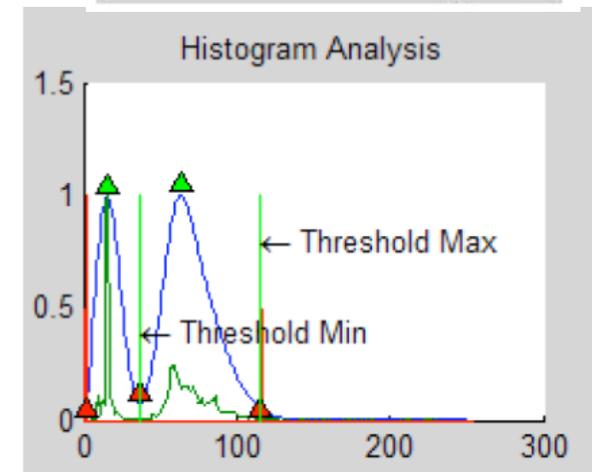
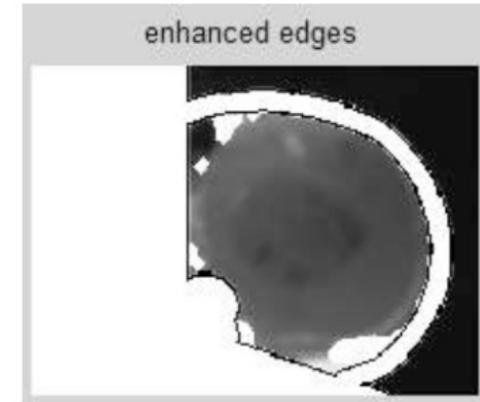
Then analyze the **histogram** of the resulting image.

Extract minima and maxima in order to separate data clusters by applying the optimal thresholding.

Data clusters are then defined as maxima surrounded by minima.

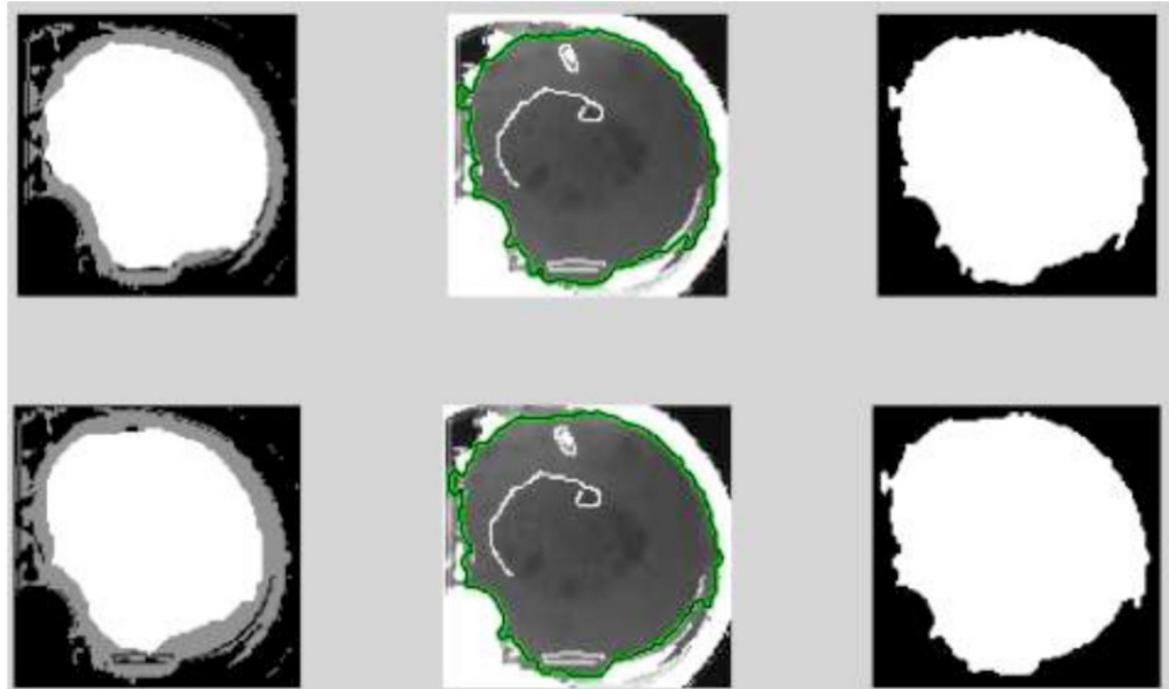
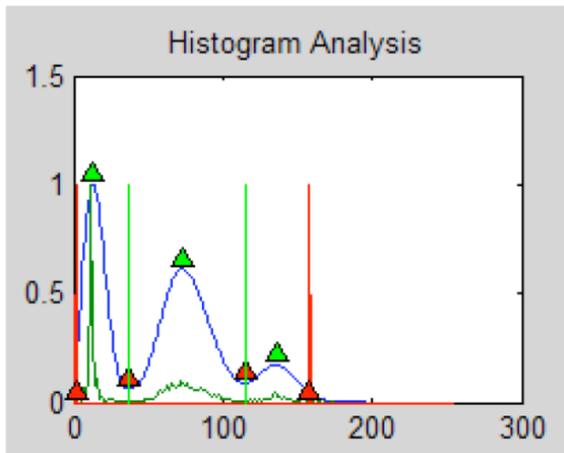
The first cluster always represents the noise and the image background. The second cluster usually represents the tumor.

Hence the indices of the minima of the tumor cluster should represent the gray value range of the tumor

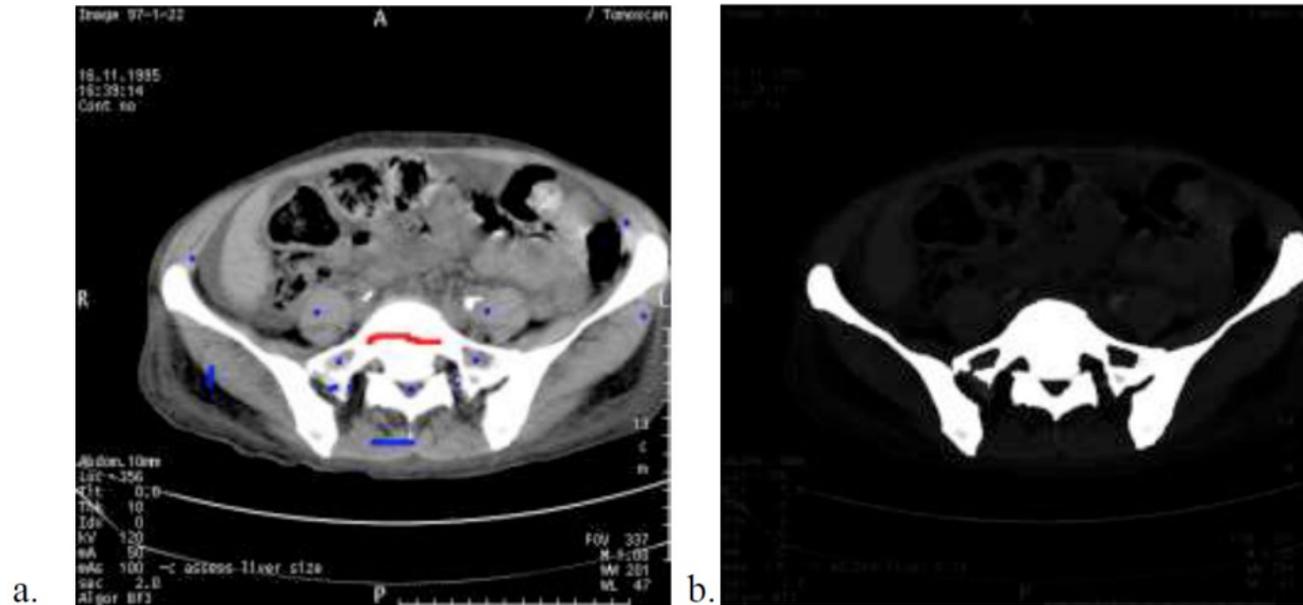


## Fine detection of tumor blob

- (1) Apply double thresholding using the just calculated threshold min and max gray values in order to extract the tumor blob.
- (2) fill the resulting image in order to get a mask.
- (3) Subtracting this mask from the thresholded image gives us the body segmentation.
- (4) Apply GrowCut on the extracted blob.
- (5) Recompute histogram for this region.



# Grow Cut

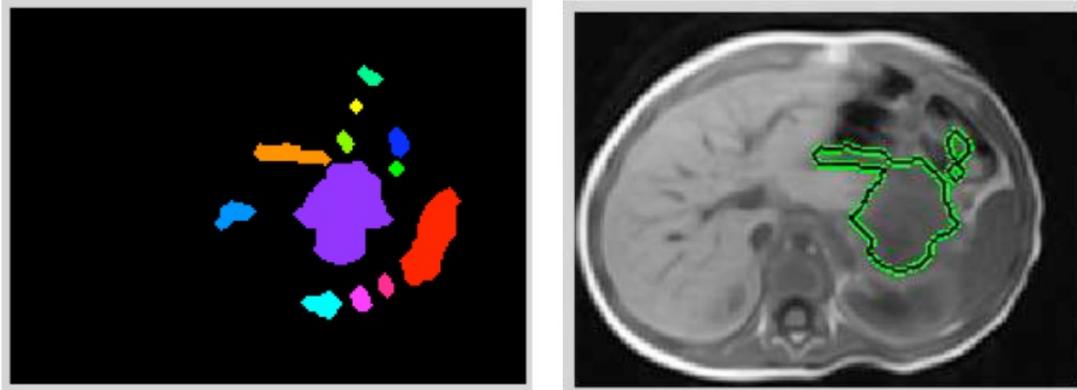


Grow Cut is initialized by selecting seed-points.

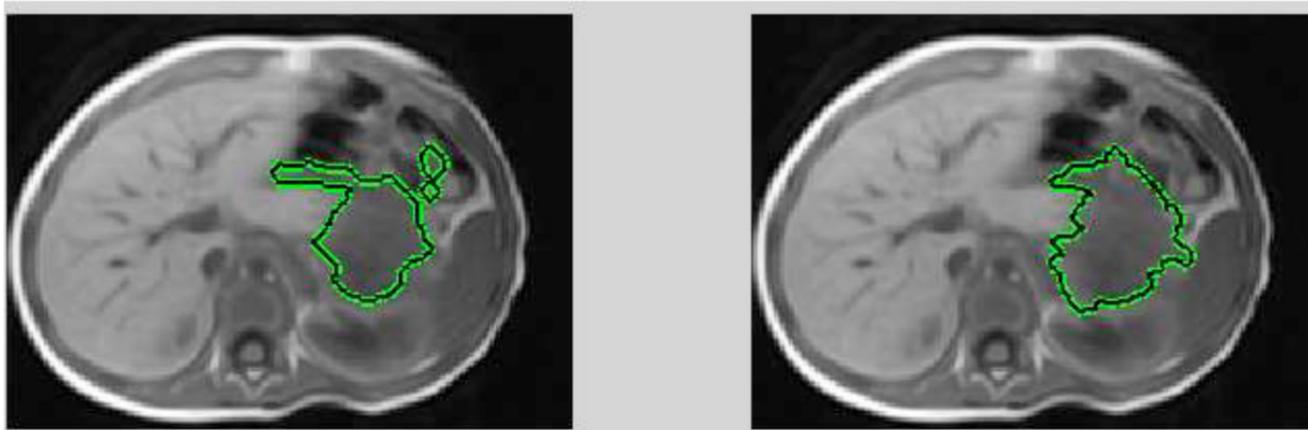
Usually, two types of seed-points are used.

One type of seed-points is used as foreground and the second one for background.

## Blob recognition: tumor detection

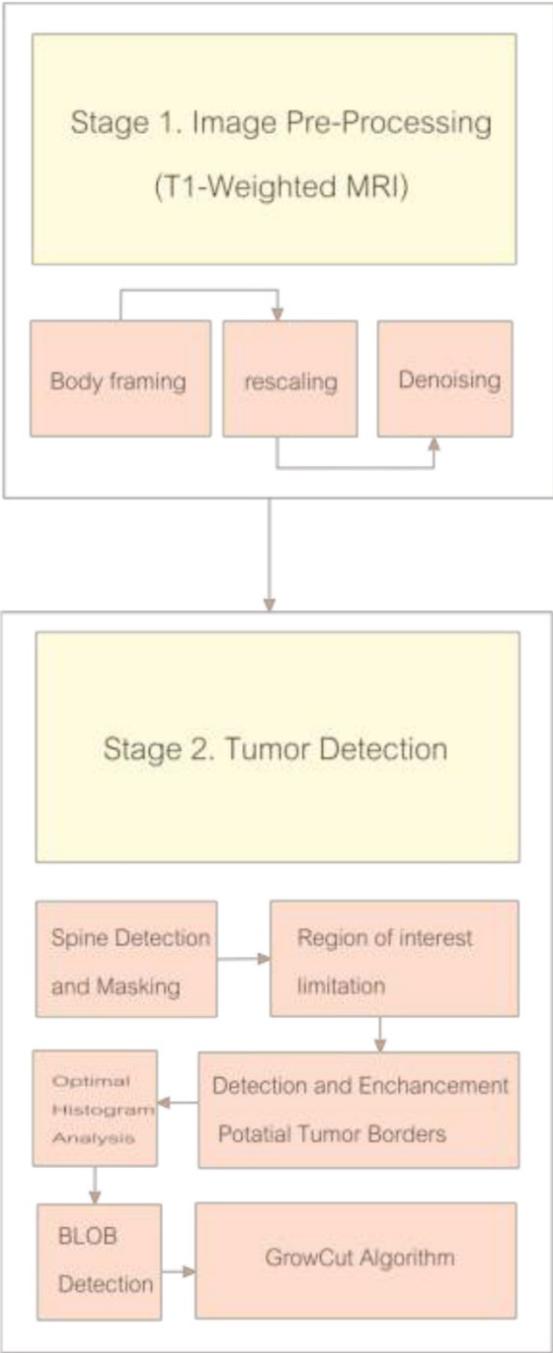


Apply some further hokus-pokus, e.g. blob detection



End result of automated tumor detection.

# Algorithm overview

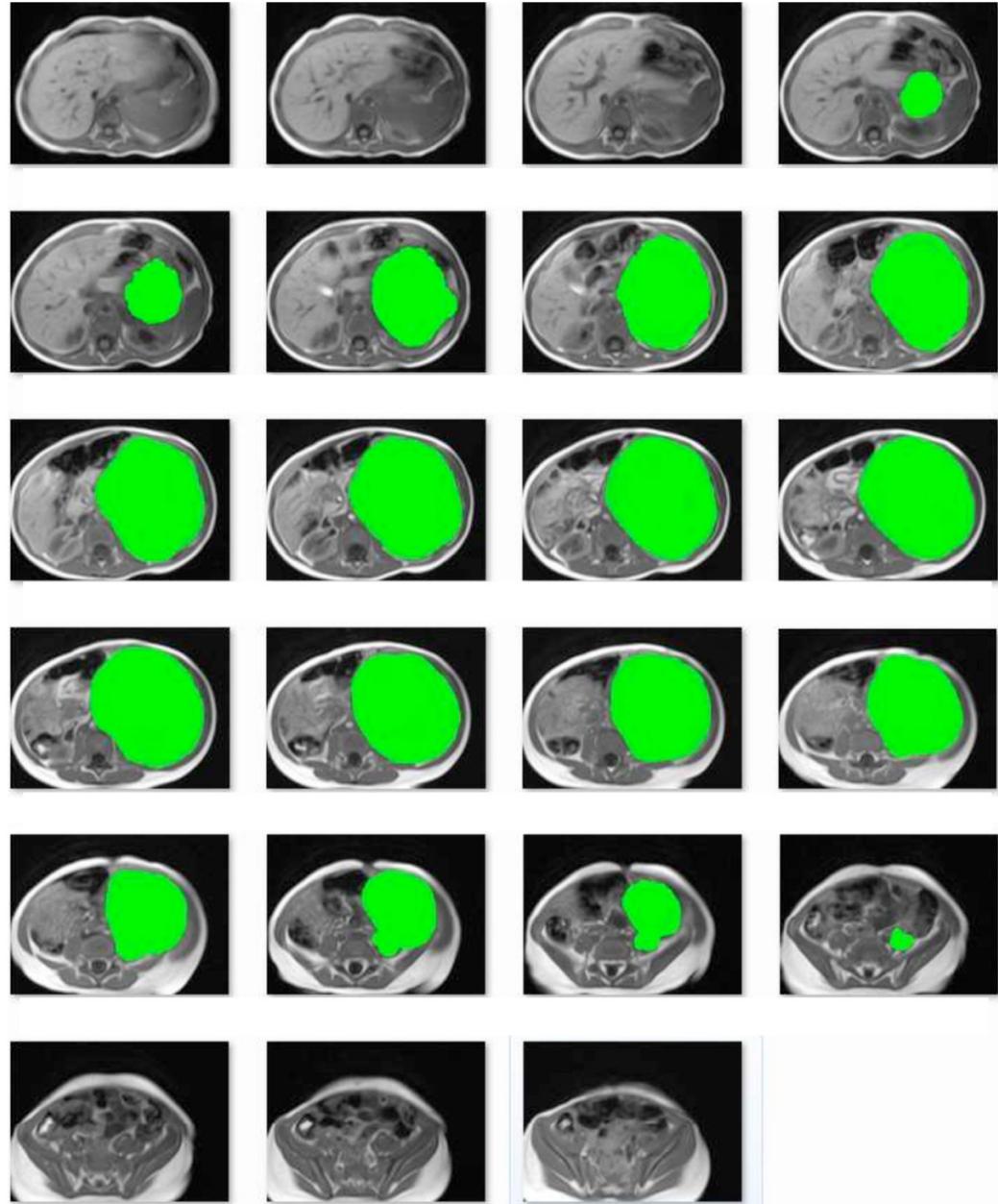


## Gold standard

Gold standard:

Manually marked scans of series  
ID 2 from 1 till 20.

These are horizontal slices  
through the body at different  
levels from top to bottom.



## Dimensions of tumor

Scan ID	Mean Gray Level	Detected Area (mm <sup>2</sup> )	True Area	Perimeter	Centroid X	Centroid Y
4	72.95	1499.00	993.00	215.38	124.86	75.48
5	65.25	3142.00	1871.00	316.74	124.95	77.43
6	63.38	5120.00	4281.00	381.50	128.46	74.18
7	63.64	5876.00	5772.00	313.32	125.40	71.73
8	77.66	6511.00	6104.00	358.53	125.54	70.90
9	78.36	7509.00	6697.00	385.40	126.02	70.11
10	69.85	7887.00	6974.00	412.48	124.95	70.36
11	68.17	7782.00	6758.00	348.19	123.54	68.86
12	70.27	7922.00	6929.00	378.43	122.87	69.37
13	68.05	7779.00	7036.00	362.19	122.65	69.95
14	67.12	7253.00	6565.00	363.71	121.55	70.73
15	63.84	6518.00	5561.00	346.63	119.99	70.71
16	80.47	6258.00	5379.00	354.98	122.57	70.92
17	76.70	5001.00	4091.00	318.63	118.17	70.35
18	59.95	3490.00	3280.00	273.56	116.20	68.83
19	61.33	2516.00	2059.00	270.49	108.91	71.66
20	71.76	411.00	222.00	86.91	107.26	92.24
<b>Average=</b>	69.34	5439.65	4739.53	322.77	121.40	72.58
	<b>Tumor Volume (cm3):</b>	<b>924.74</b>	<b>805.72</b>			

**Table 3: Scan Series ID 2 statistics - 1**

# Accuracy of automated detection

**True positive TP:** Pixels correctly recognized as tumor tissue

**False positive FP:** Pixels of healthy tissue incorrectly identified as tumor

**True negative TN:** Pixels of healthy tissue correctly identified as healthy

**False negative FN:** Pixels of tumor tissue incorrectly identified as healthy

**Sensitivity** (True Positive rate) =  $TP / (TP + FN)$  It measures the proportion of actual positives which are correctly identified as such.

**Specificity** (True Negative rate) =  $TN / (TN + FP)$  It measures the proportion of negatives which are correctly identified as such.

# Accuracy of automated detection

Scan ID	Sensitivity	Specificity	Precision	NPV	fall_out	FDR	Accuracy	F1
6	0.54	1.00	1.00	0.97	0.00	0.00	0.97	0.71
7	0.63	1.00	1.00	0.97	0.00	0.00	0.98	0.77
8	0.74	1.00	1.00	0.98	0.00	0.00	0.98	0.85
9	0.56	1.00	0.99	0.95	0.00	0.01	0.95	0.72
10	0.53	1.00	0.99	0.95	0.00	0.01	0.95	0.69
11	0.56	1.00	1.00	0.94	0.00	0.00	0.95	0.72
12	0.73	1.00	0.99	0.96	0.00	0.01	0.97	0.84
13	0.85	0.99	0.95	0.98	0.01	0.05	0.97	0.89
14	0.89	0.99	0.93	0.98	0.01	0.07	0.98	0.91
15	0.90	0.99	0.92	0.99	0.01	0.08	0.98	0.91
16	0.80	1.00	0.96	0.98	0.00	0.04	0.98	0.87
17	0.72	1.00	1.00	0.97	0.00	0.00	0.97	0.84
18	0.87	1.00	0.91	0.99	0.00	0.09	0.99	0.89
19	0.20	1.00	1.00	0.97	0.00	0.00	0.97	0.34
20	0.58	1.00	0.98	0.99	0.00	0.02	0.99	0.73
<b>Average=</b>	0.67	1.00	0.97	0.97	0.00	0.03	0.97	0.78

The specificity is very high = healthy tissue correctly detected.

The sensitivity has still room for improvement = not all tumor-affected regions are detected.