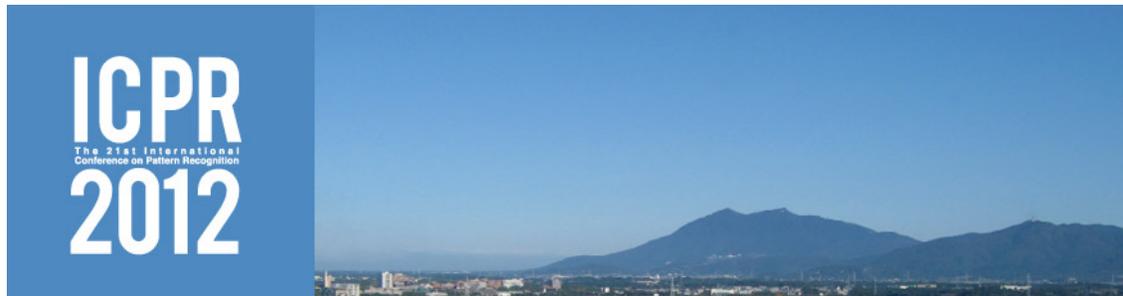


Mitosis Detection in Breast Cancer Histological Images

An ICPR 2012 Contest



February 8, 2012

1 Organizers

- IPAL (UMI CNRS, I²R/A*STAR, Université Joseph Fourier), Singapore
 - Ludovic Roux
 - Daniel Racoceanu
 - Nicolas Loménie
 - Maria Kulikova
 - Humayun Irshad
- TRIBVN Company, Châtillon, France
 - Jacques Klossa
- Pitié-Salpêtrière Hospital, Paris, France
 - Prof. Frédérique Capron
 - Dr. Catherine Genestie
 - Gilles Le Naour
- The Ohio State University, College of Medicine, School of Biomedical Science, USA
 - Metin Gurcan

2 General Description of the Problem

Breast cancer is a major cause of death in the world. According to figures published by the World Health Organization, the three most common cancers in the world in 2008 in terms of incidence were lung (1.52 million cases), breast (1.29 million) and colorectal (1.15 million) [1]. For women, breast cancer is by far the most frequent cancer worldwide with an estimated 715,000 new cases diagnosed in more developed regions (26.5% of the total), and 577,000 in less developed countries (18.8%).

To diagnose breast cancer, samples of tissue are removed from breast. From these samples, histology slides are produced. These slides are stained generally with H&E. Pathologists analyse breast biopsy slides and give them a grade according to one of two international grading systems:

- The Scarff, Bloom and Richardson grading system.

- The Elston and Ellis grading system [2]. This grading system is a modification of the Scarff, Bloom and Richardson grading system. It is also known as the Nottingham grading system.

The Elston and Ellis grading system is recommended by the World Health Organization [3]. It is derived from the assessment of three morphological features: tubule formation, nuclear pleomorphism and mitotic count. Here is how Elston and Ellis describe mitotic counts [2]:

“Mitotic activity is best assessed at the periphery of the tumour where active growth is most likely. A minimum of 10 consecutive fields is assessed. Strict criteria for the identification of mitotic figures must be employed, and only nuclei in which clear morphological feature of metaphase, anaphase and telophase are counted. Hyperchromatic¹ and apoptotic² nuclei are ignored and care is taken to avoid mistaking lymphocytes within a tumour for mitoses.”

Several studies on automatic tools to process digitized slides have been reported focusing mainly on nuclei or tubule detection. Mitosis detection is a challenging problem and has not been addressed well in the literature.

Mitotic count is an important parameter in breast cancer grading as it gives an evaluation of the aggressiveness of the tumour. Detection of mitosis is a very challenging task since in images, they appear as small objects which have a large variety of shapes. The four main phases of a mitosis are prophase, metaphase, anaphase and telophase. Figure 1 shows examples and description of the mitosis phases. The shape of the nucleus is very different depending on the phase of the mitosis. On its last stage, the telophase, a mitosis has two distinct nuclei, but they are not yet full individual cells. A mitosis in telophase must be counted as one single mitosis, it should not be miscounted as two mitosis. Artefacts are also common and should not be confused with mitosis.

The objective of the contest is to detect mitosis on H&E stained histological images obtained using the biopsy slides of different patients.

3 Dataset

Prof. Frédérique Capron and Dr. Catherine Genestie, two experienced pathologists of Pitié-Salpêtrière Hospital in Paris, France, have provided a set of 5 breast cancer biopsy slides. The slides are stained with hematein and eosin (H&E). In each slide, the pathologists selected 10 high power fields (HPF) at 40X magnification. An HPF has a size of $512 \times 512 \mu\text{m}^2$ (that is an area of 0.262 mm^2), which is the equivalent of a microscope field diameter of 0.58 mm. These 50 HPFs contain more than 300 mitosis in total. As there are several possible shapes for mitosis, it is necessary to have a large dataset to be able to cover all the cases. Two-thirds of the HPFs will be used for training purpose, the other third being for testing.

The slides have been scanned by three different equipments:

- an Aperio XT scanner (scanner A);
- a Hamamatsu NanoZoomer scanner (scanner H);
- and a 10 bands multi-spectral microscope M. The spectral bands are all in the visible spectrum. In addition, for each spectral band, the digitization has been performed at 17 different focus planes (17 layers Z-stack), each plane being separated from the other by 500 nm.

Scanner A has a resolution of $0.2456 \mu\text{m}$ per pixel. Scanner H has a slightly better resolution of $0.2273 \mu\text{m}$ (horizontal) and $0.22753 \mu\text{m}$ (vertical) per pixel, so a pixel of scanner H is not exactly a square. At last, multispectral microscope M has the best resolution of $0.185 \mu\text{m}$ per pixel. Table 1 shows the resolutions of the different scanners and the microscope. For example, a mitosis having an area of $30 \mu\text{m}^2$ will cover about 500 pixels of the image produced by scanner A, about 580 pixels of the image produced by scanner H, and about 880 pixels of the image produced by multispectral microscope M.

For each slide, there is one RGB image produced by scanner A, one RGB image produced by scanner H, and 170 grey scale images for the multi-spectral microscope M (10 spectral bands and 17 layers Z-stack for each spectral band).

The pathologists have annotated mitosis manually. They made the annotations in each selected HPF on the images generated by the scanner A, the scanner H and the multi-spectral microscope M (see Figure 2).

¹A hyperchromatic nucleus has elevated chromatin, that is there is an abundance of DNA that stains darkly when stained for histological viewing.

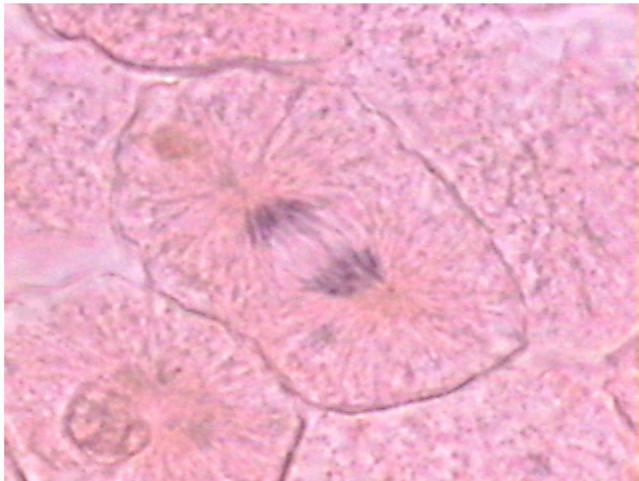
²Apoptosis is the process of programmed cell death. A cell undergoing apoptosis shows a characteristic morphology, including cell shrinkage and rounding, and chromatin undergoes condensation into compact patches.



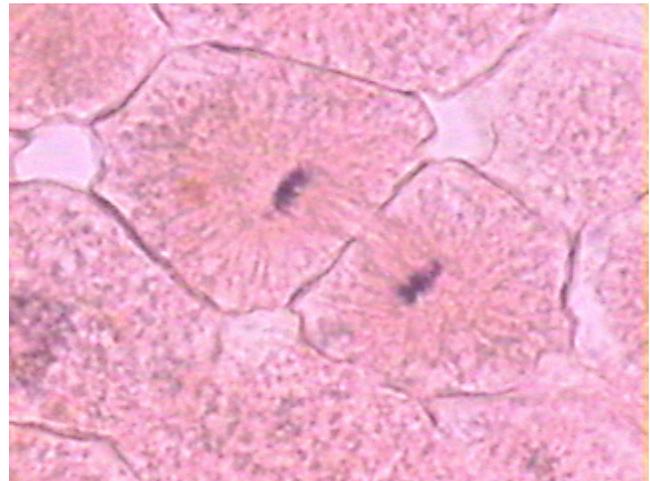
(a) **Prophase** is a stage of mitosis in which the chromatin condenses (it becomes shorter and fatter) into a highly ordered structure called a chromosome in which the chromatin becomes visible.



(b) **Metaphase** is a stage of mitosis in the eukaryotic cell cycle in which condensed & highly coiled chromosomes, carrying genetic information, align in the middle of the cell before being separated into each of the two daughter cells.



(c) **Anaphase** is the stage of mitosis when chromosomes separate in an eukaryotic cell. Each chromatid moves to opposite poles of the cell, the opposite ends of the mitotic spindle, near the microtubule organizing centers.



(d) **Telophase** is a stage of mitosis in a eukaryotic cell in which the effects of prophase and prometaphase events are reversed. Two daughter nuclei form in the cell. The nuclear envelopes of the daughter cells are formed from the fragments of the nuclear envelope of the parent cell. As the nuclear envelope forms around each pair of chromatids, the nucleoli reappear.

Figure 1: Four mitosis phases.

Equipment	Resolution	Dimension of HPF to cover an area of $512 \times 512 \mu\text{m}^2$
Scanner A	$0.2456 \mu\text{m}$ per pixel	2084×2084 pixels
Scanner H	$0.2273 \mu\text{m}$ horizontal and $0.22753 \mu\text{m}$ vertical per pixel	2252×2250 pixels
Multispectral microscope M	$0.185 \mu\text{m}$ per pixel	2767×2767 pixels

Table 1: Resolution of the scanners A and H and the multispectral microscope M.

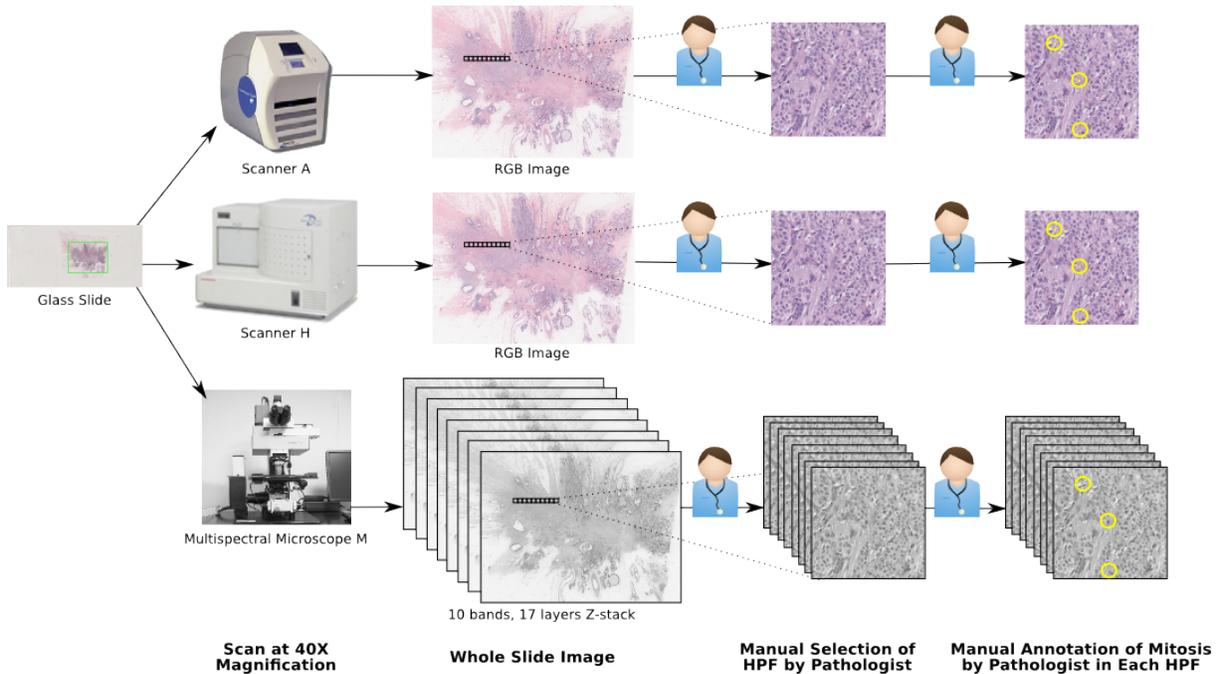


Figure 2: Generation of whole slide images, selection of high power fields and annotation of mitoses.

3.1 Training Data Set

Each HPF comes with a ground truth text file (CSV type), which indicates the coordinates of each single mitosis regions within the image. Organisation of a mitosis coordinates CSV text file (see Figure 3):

- The file has no header
- One line gives the coordinates of all the pixels which belong to the same mitosis
- The coordinates of a pixel are: x location, y location
- The origin of the image is its top left corner (coordinates 0,0)

3.2 Scanner A

Images of scanner A are RGB images stored in bitmap format and are named as `Ass.ff.bmp` where

- `ss` is slide number (00 to 04)
- `ff` is HPF number (00 to 09)

To cover an area of $512 \times 512 \mu\text{m}^2$, each image has a size of 2084×2084 pixels.

For each image, a CSV file `Ass.ff.csv` contains the list of coordinates of pixels of all the mitosis in this image, and a JPEG file `Ass.ff.jpg` shows in yellow colour the location of all the mitosis in the image.

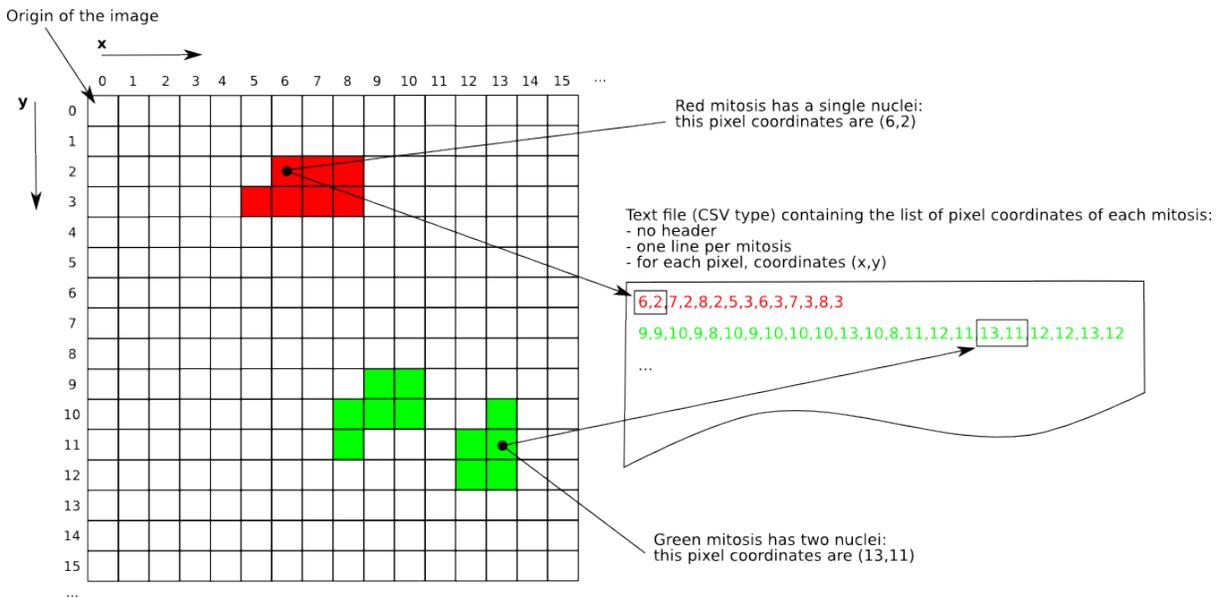


Figure 3: Organisation of a text file containing the list of pixel coordinates for the mitosis located in an HPF.

3.3 Scanner H

Images of scanner H are RGB images stored in bitmap format and are named as Hss.ff.bmp where

- ss is slide number (00 to 04)
- ff is HPF number (00 to 09)

To cover an area of $512 \times 512 \mu\text{m}^2$, each image has a size of 2252×2250 pixels. Because pixels of scanner H are not square, the images obtained from scanner H are not squares, they have more pixels on x axis than on y axis.

For each image, a CSV file Hss.ff.csv contains the list of coordinates of pixels of all the mitosis in this image, and a JPEG file Hss.ff.jpg shows in yellow colour the location of all the mitosis in the image.

3.4 Multispectral microscope M

The camera attached on top of the microscope generates images of 1360×1360 pixels. However, to cover an area of $512 \times 512 \mu\text{m}^2$, we need 2767×2767 pixels. Therefore, we will use 4 images to cover the same area as the two scanners. But these 4 images do not cover completely the $512 \times 512 \mu\text{m}^2$ area, 47 pixels are missing in width and in height to cover fully the area.

Each image, covering a quarter of a scanner image, is labelled a, b, c or d depending on its position in the scanner image. Figure 4 shows the location of each quarter a, b, c, d. As the quarters do not cover completely the $512 \times 512 \mu\text{m}^2$ area, compared to the scanner images, there is a small gap on the borders, and also a small gap between quarters a, b, c and d.

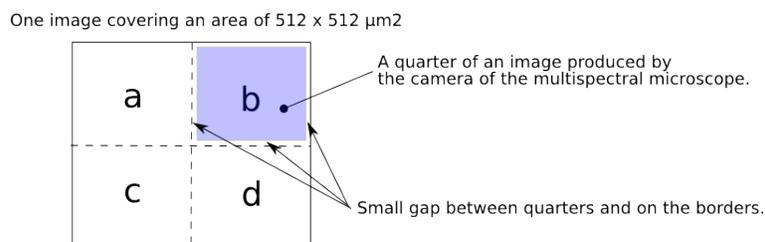


Figure 4: Location of quarters a, b, c and d of multispectral microscope in scanner image.

Images of multispectral microscope M are stored in bitmap format and are named as Mss.ffq.bbzz.bmp where

- ss is slide number (00 to 04)

- ff is HPF number (00 to 09)
- q is a letter code for a quarter of a scanner image (a, b, c or d)
- bb is spectral band number (00 to 09)
- zz is z-stack number (00 to 16)

Table 2 and figure 5 describe the 10 spectral bands of the multispectral microscope. They are all in the visible spectrum.

Band number	Wavelength (nm)	Main visible colour
0	665-710	red
1	410-750	white
2	410-450	dark blue
3	430-490	blue
4	460-510	cyan
5	480-540	green
6	525-575	apple green
7	580-595	yellow
8	590-640	orange
9	625-670	purple red

Table 2: Spectral bands of the multispectral microscope.

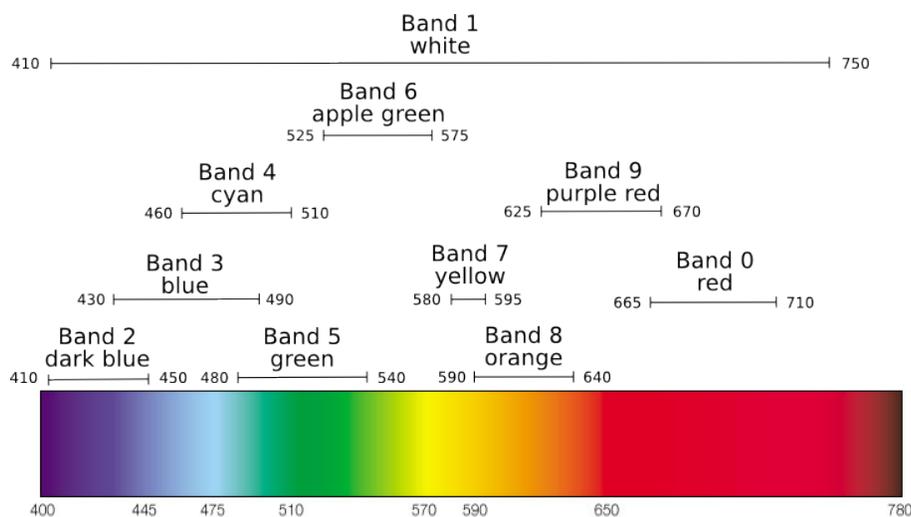


Figure 5: Spectral bands of the multispectral microscope.

For each image, a CSV file `Mss_ffq_bbzz.csv` contains the list of coordinates of pixels of all the mitosis in this image, and a JPEG file `Mss_ffq_bbzz.jpg` shows in yellow colour the location of all the mitosis in the image.

4 Description of the Competition Tasks

The competition consists in being able to detect mitosis on an image. Different types of images are provided. The contestants can take advantage of using the information of some of the spectral bands which may offer better discrimination for mitosis detection, or to concentrate only on RGB images.

The result of the detection should be, for each image, a text file (CSV type) containing the list of pixel coordinates of all the pixels covering the detected mitosis areas (the origin point of an image is at the top left corner of the image). One line of the file contains the list of all pixels coordinates of a single mitosis. This is exactly the same format as CSV files provided with training data set (see Figure 3).

5 Evaluation Metrics

The main goal is to be able to tell what is the mitotic count on each slide. A segmented mitosis would be counted as correctly detected if its centroid is localised within a range of $5\mu\text{m}$ of the centroid of a ground truth mitosis.

5.1 Centroid-based Measures

The centroid-based measures deal with the number of objects detected. As the main goal of the contest is to give the correct mitotic count, the centroid-based measures are the most important ones. These measures are defined as follows:

- D = number of mitosis detected.
- TP = number of True Positives, that is the number of mitosis that are ground truth mitosis among the D mitosis detected
- FP = number of False Positives, that is the number of mitosis that are not ground truth mitosis among the D mitosis detected
- FN = number of False Negatives, that is the number of ground truth mitosis that have not been detected
- recall (sensitivity) = $\frac{\text{TP}}{\text{TP}+\text{FN}}$
- precision (positive predictive value) = $\frac{\text{TP}}{\text{TP}+\text{FP}}$
- F-measure = $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$
- localisation quality: Mean and standard deviation of the distance between detected and ground truth centroids

5.2 Region-based Measures

The region-based measures deal with the surface of the objects. Only segmented mitosis considered as correctly detected (that is having their centroid localised within a range of $5\mu\text{m}$ of the centroid of a ground truth mitosis) are considered for region-based measures. These measures are defined as follows:

- overlap = $\frac{|\mathcal{A}(S) \cap \mathcal{A}(G)|}{|\mathcal{A}(S) \cup \mathcal{A}(G)|}$
- recall (sensitivity) = $\frac{|\mathcal{A}(S) \cap \mathcal{A}(G)|}{|\mathcal{A}(G)|}$
- specificity = $\frac{|N - \mathcal{A}(S) \cup \mathcal{A}(G)|}{|N - \mathcal{A}(G)|}$
- precision (positive predictive value) = $\frac{|\mathcal{A}(S) \cap \mathcal{A}(G)|}{|\mathcal{A}(S)|}$
- F-measure = $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$

where N is the total number of pixels in a HPF, $\mathcal{A}(S)$ is the area of the segmented mitosis, $\mathcal{A}(G)$ is the area of the ground truth mitosis, $|\mathcal{A}(S)|$ is the number of pixels of the area of segmented mitosis, $|\mathcal{A}(G)|$ is the number of pixels of the area of ground truth mitosis.

For the evaluation of the algorithms of the participants, we will provide a set of test images on August 1st, 2012. The participants will have one month to run their algorithms on test images, and to provide their results to the organizers of the contest. The participants have to provide their results (list of detected mitosis) in the same format as the training data, that is one CSV text file for each HPF (CSV text file format must be as described in Figure 3).

6 Ranking of the Contestants

The contestants will be ranked according to the following 3 centroid-based measures:

1. recall (sensitivity)
2. precision (positive predictive value)
3. F-measure

For each measure, the 3 best contestants are kept and given a medal (gold, silver or bronze). The contestants will be ranked according to the number of medals they collected.

If there are contestants with equal number and value of medals, the localisation quality (mean and standard deviation of the distance between detected and ground truth centroids) will make the difference between them.

Acknowledgement

This contest is supported by the French National Research Agency ANR, project MICO under reference ANR-10-TECS-015.

References

- [1] Peter Boyle and Bernard Levin, editors. *World Cancer Report 2008*. International Agency for Research on Cancer, 2008.
- [2] C.W. Elston and I.O. Ellis. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology*, 19:403–410, 1991.
- [3] Fattaneh A. Tavassoli and Peter Devilee, editors. *Tumours of the Breast and Female Genital Organs*. World Health Organization Classification of Tumours. Pathology & Genetics. International Agency for Research on Cancer, September 2003.