

# An Objective View of Cancer

There's considerable research showing what should come as no surprise: Pathologists are human. Differences between malignant and benign tissue can be subtle. Interpretations are often subjective. A recent study at Johns Hopkins found that 2 to 3 percent of diagnoses are either wrong—misreading malignant tissue as benign or vice-versa—or could have been more accurate, potentially leading to unnecessary or inappropriate treatment.

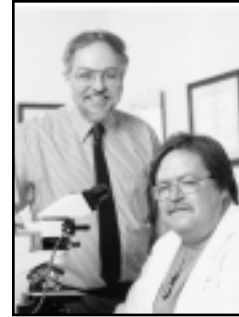
What if you could build a database of accurately diagnosed tissue and do a computerized search with undiagnosed samples? Several years ago, Dr. Michael Becich, a pathologist at the University of Pittsburgh Medical Center took

this question to the Pittsburgh Supercomputing Center. The result is Computer-Based Image Retrieval, a still-evolving software method that does essentially what pathologists do when they flip pages in a reference book, but objectively, more comprehensively and in a matter of seconds.

“We’ve developed a tool,” says Becich, director of genitourinary pathology and informatics at UPMC, “that classifies microscopic images by image content. Instead of using text, where you’d go to a database and say show me your images of prostate cancer, for instance, CBIR uses computerized image-classification to create image signatures. For an unknown sample, the image signature acts as the search key.”

The aim, emphasizes PSC scientist Art Wetzel, who developed the signature-matching software, is not to replace the skill of doctors, but to provide more information in a more objective manner: “A book can’t cover the range we build into a database. We’re trying to save search time and broaden the potential choices. There are many situations where a dollar of computing can save much more than a dollar of pathologist’s time.”

The PSC-UPMC team, which includes pathologists John Gilbertson and Rebecca Crowley and grad student Lei Zheng, has built an archive of about 120,000 images, representing a range of tissue categories—such as muscular, glandular, and gastrointestinal. Using this data, CBIR has shown impressive ability at matching unknown samples with the correct category. Going beyond classification to diagnosis, the team has focused on prostate tissue, Becich’s specialty.



▲ Art Wetzel, Pittsburgh Supercomputing Center, and Michael Becich, University of Pittsburgh Medical Center.

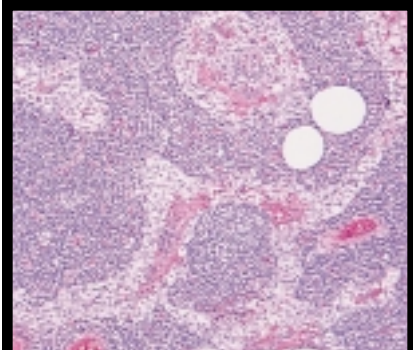
## Toward Consistent, Objective Diagnosis

Prostate cancer is the most frequently occurring non-skin cancer in American men. Recent data indicates more than 200,000 new cases each year and more than 40,000 deaths, second only to lung cancer for cancer-related mortality in U.S. males.

The best available method for evaluating long-term prognosis is Gleason grading. Pathologists examine the tissue under a microscope and assign a number—from one through five, one being least severe—according to the structural patterns they observe. By adding the two most predominant Gleason grades in a sample, the tissue gets a score between two and 10. This score is a major factor in treatment planning. The problem, however, is that research shows a 20 to 40 percent variation when different pathologists score the same samples.

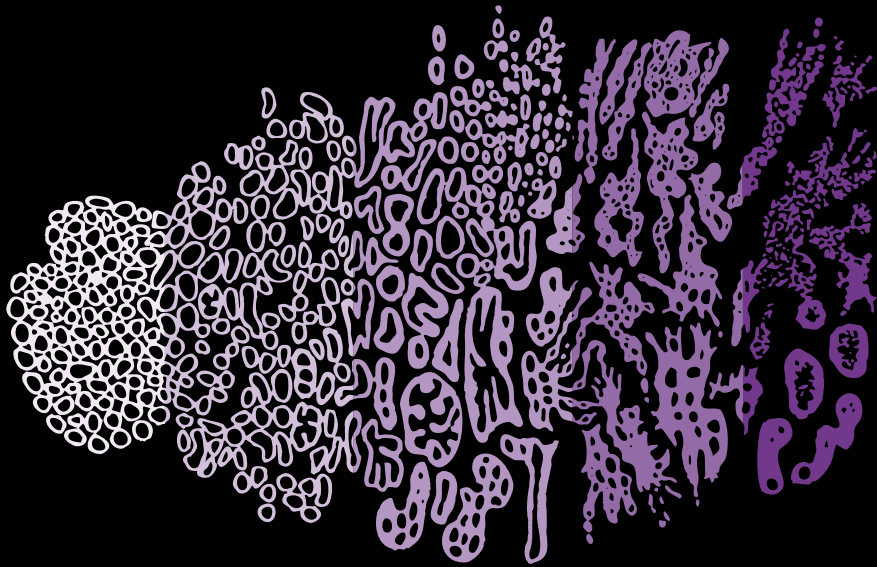
To create image signatures, Wetzel implemented a series of image-classification algorithms, some based on coloration, others on mathematical methods, which allow the computer to discern features that often aren’t visually apparent. For prostate grading, he turned to a method called “spanning trees” by which, in effect, the software draws lines connecting cell nuclei. The distribution of lengths and angles of these lines correlates well with Gleason grades.

On samples where pathologists have identified a region of interest, CBIR achieves 80 percent agreement with Gleason grading by UPMC pathologists,



**Lymph Node Tissue**  
An example of glandular tissue, which CBIR can differentiate from other tissue categories, such as muscular and gastrointestinal, on the basis of image features.

[Bringing Consistency, Comprehensiveness and Cost-Saving to Diagnosis]

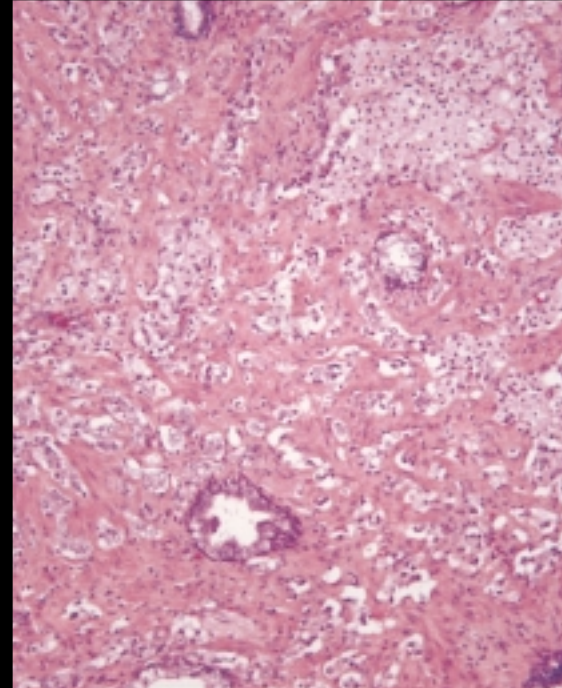


1 2 3 4 5

The original of this graphic is courtesy of Lippincott, Williams & Wilkins.

### Gleason Grading

At less severe grades of prostate cancer (left), glandular structure is relatively organized and differentiated from the background. At higher grades, structure breaks down. The large image (below) is an example of a Gleason grade 4 prostate.



and most disagreements are confined to one grade level. It's difficult to gauge, says Becich, what this means with respect to accurate diagnosis: "It's hard to know if we're doing better or worse, until we develop a new, objective scale, but it's clear—since pathologists have this 20 percent variability—that CBIR can do as well as humans."

"Part of what CBIR offers," says Wetzel, "is putting a stake in the ground and getting an objective evaluation of where the boundaries are between grades. It's not so much that there's always a right or wrong answer; it's that we need a consistent answer."

### What's Ahead

CBIR's ability at Gleason grading is so far confined to cases in which a human picks out the region of interest. Though relatively easy for pathologists, it's a

challenging software problem: "How do you know which parts of a slide are important," says Wetzel, "and ignore artifacts like torn edges, poor staining? Things human vision can do automatically are hard to program. We need to move into some of these areas."

With prostate grading, Becich is tracking the results of cases in which CBIR picks out matches using signatures unrelated to the structural patterns of Gleason grading. "We're currently looking at a large number of cases in which we have clinical follow-up to see if CBIR picks out poor prognosis tumors better than human grading."

The PSC-UPMC team sees CBIR as a tool that will eventually work with high-performance networking to provide "telemicroscopy" for pathologists at remote locations, to save time and improve diagnoses. "When you broaden the range of choice," says Becich, "you improve the ability to arrive at correct diagnosis and appropriate treatment."

### More information:

<http://www.psc.edu/science/wetzel.html>

### Spanning Trees

Through digital manipulation, CBIR differentiates cell nuclei, which form roughly circular structural patterns evaluated in Gleason grading, by their dark coloration. The software then connects the nuclei positions, taking account of the characteristics of the space between nuclei, to arrive at a weighted-length tree structure that correlates well with Gleason grading by pathologists.

