面向癌症计算机辅助诊断与预后的组织病理图像分析

徐 军

江苏省大数据分析技术重点实验室

南京信息工程大学

2016年5月18日

临床合作单位

CASE WESTERN RESERVE UNIVERSITY

VANDERBILT UNIVERSITY MEDICAL CENTER

NASHNP

Penn UNIVERSITY of PENNSYLVANIA

RUTGERS Cancer Institute of New Jersey

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center

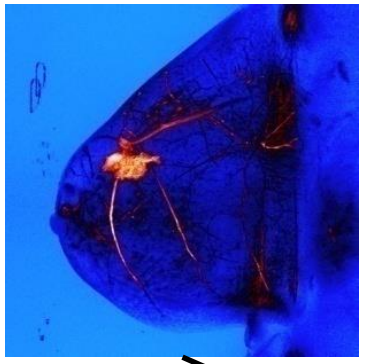
Cleveland Clinic

江苏省肿瘤医院 JIANGSU CANCER HOSPITAL

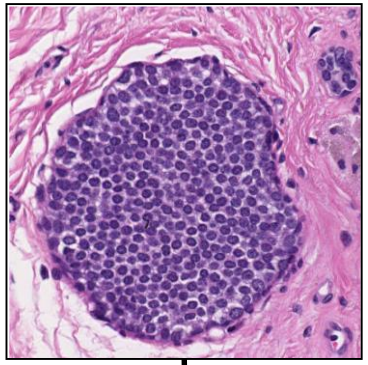
计算机辅助系统(CAD/CAP)

宏观 ⟷ 微观

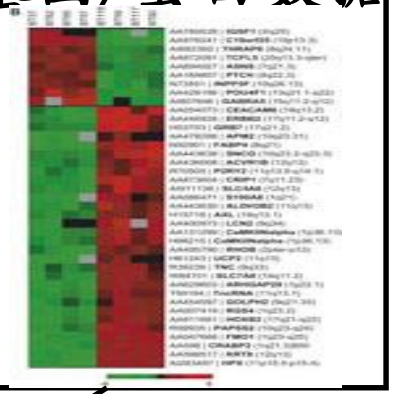
放射图像



组织病理图像



基因/蛋白数据



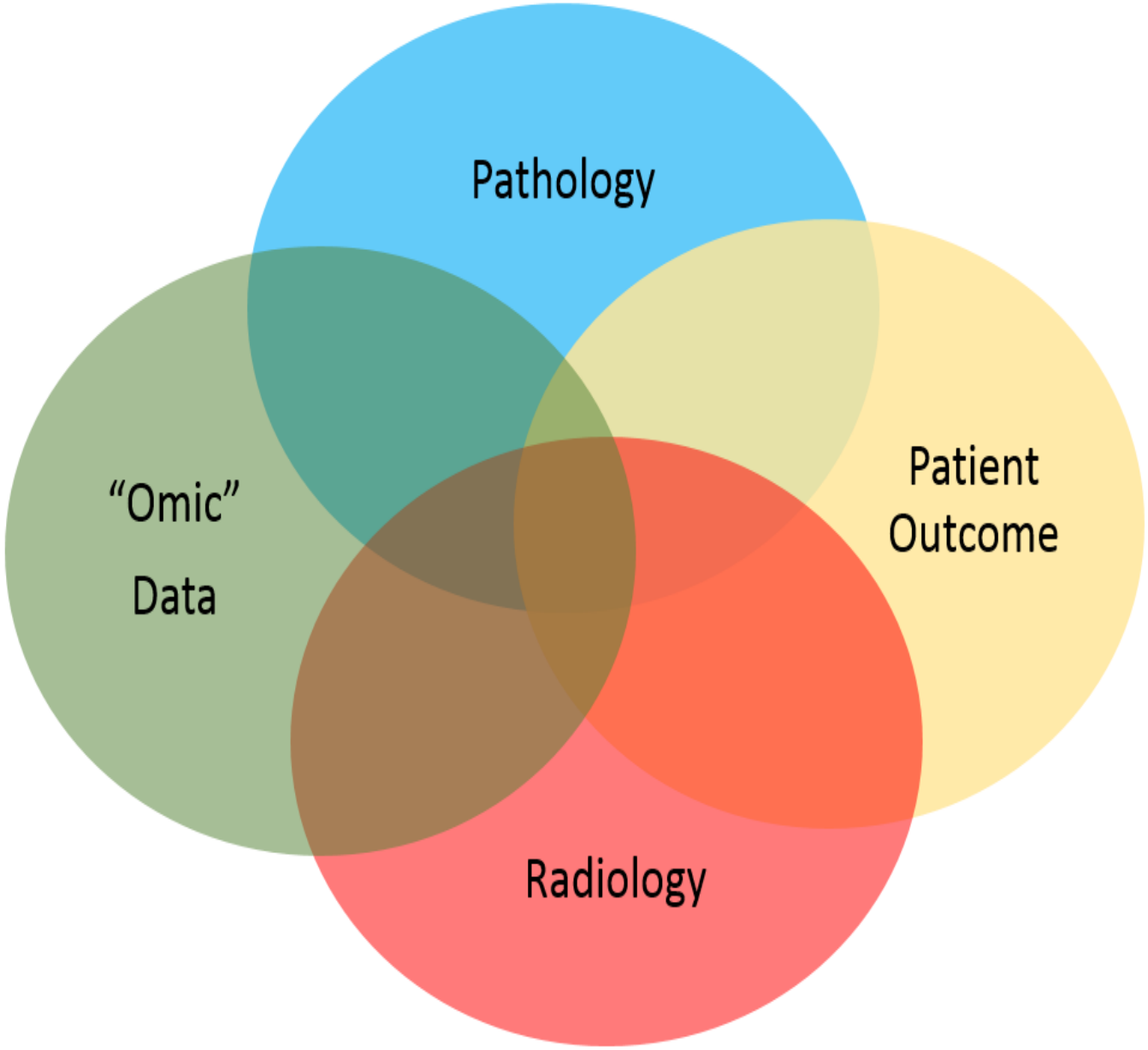
计算机辅助检测、诊断及预后(CAD/CAP)

乳腺癌

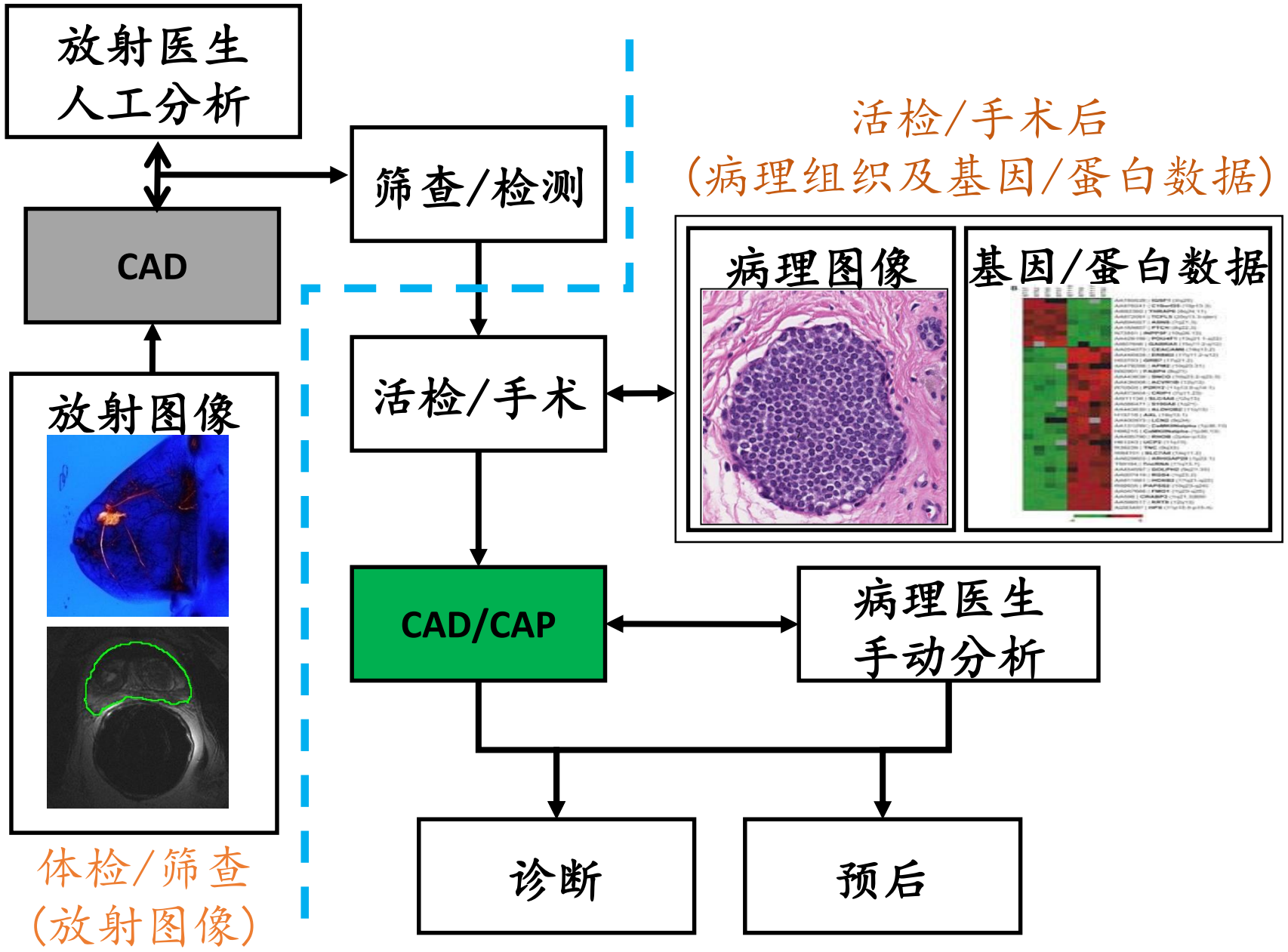
前列腺癌

其他类型癌症/疾病

计算机辅助系统(CAD/CAP)



计算机辅助系统临床应用流程



提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

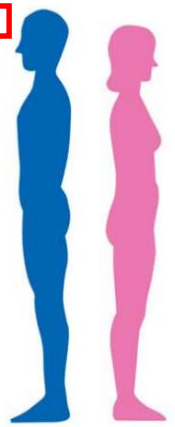
提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

2015中美癌症统计数据

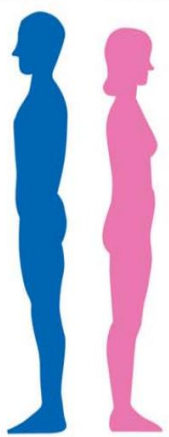
Estimated New Cases

	Males		Females	
Prostate	220,800	26%	Breast	231,840	29%
Lung & bronchus	115,610	14%	Lung & bronchus	105,590	13%
Colon & rectum	69,090	8%	Colon & rectum	63,610	8%
Urinary bladder	56,320	7%	Uterine corpus	54,870	7%
Melanoma of the skin	42,670	5%	Thyroid	47,230	6%
Non-Hodgkin lymphoma	39,850	5%	Non-Hodgkin lymphoma	32,000	4%
Kidney & renal pelvis	38,270	5%	Melanoma of the skin	31,200	4%
Oral cavity & pharynx	32,670	4%	Pancreas	24,120	3%
Leukemia	30,900	4%	Leukemia	23,370	3%
Liver & intrahepatic bile duct	25,510	3%	Kidney & renal pelvis	23,290	3%
All Sites	**848,200**	**100%**	**All Sites**	**810,170**	**100%**



Estimated Deaths

	Males		Females	
Lung & bronchus	86,380	28%	Lung & bronchus	71,660	26%
Prostate	27,540	9%	Breast	40,290	15%
Colon & rectum	26,100	8%	Colon & rectum	23,600	9%
Pancreas	20,710	7%	Pancreas	19,850	7%
Liver & intrahepatic bile duct	17,030	5%	Ovary	14,180	5%
Leukemia	14,210	5%	Leukemia	10,240	4%
Esophagus	12,600	4%	Uterine corpus	10,170	4%
Urinary bladder	11,510	4%	Non-Hodgkin lymphoma	8,310	3%
Non-Hodgkin lymphoma	11,480	4%	Liver & intrahepatic bile duct	7,520	3%
Kidney & renal pelvis	9,070	3%	Brain & other nervous system	6,380	2%
All Sites	**312,150**	**100%**	**All Sites**	**277,280**	**100%**



2015年全球新增癌症病例1400万，　　死亡800万；

我国人口13.6亿占全球人口的19%

我国　　429万(30%)，死亡281.4万(35%，66%)

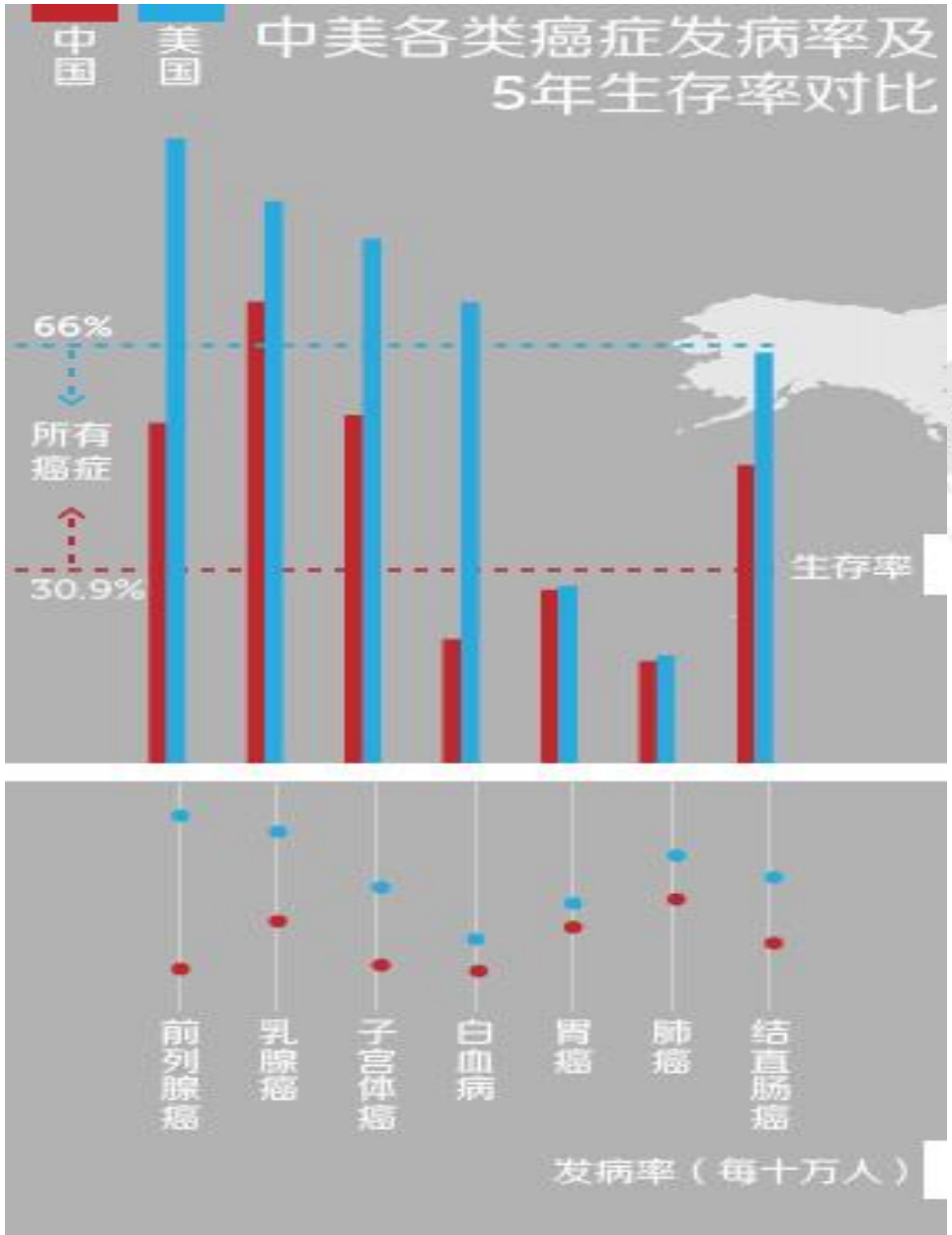
美国　　166万(12%)，死亡 59万　(7%，36%)

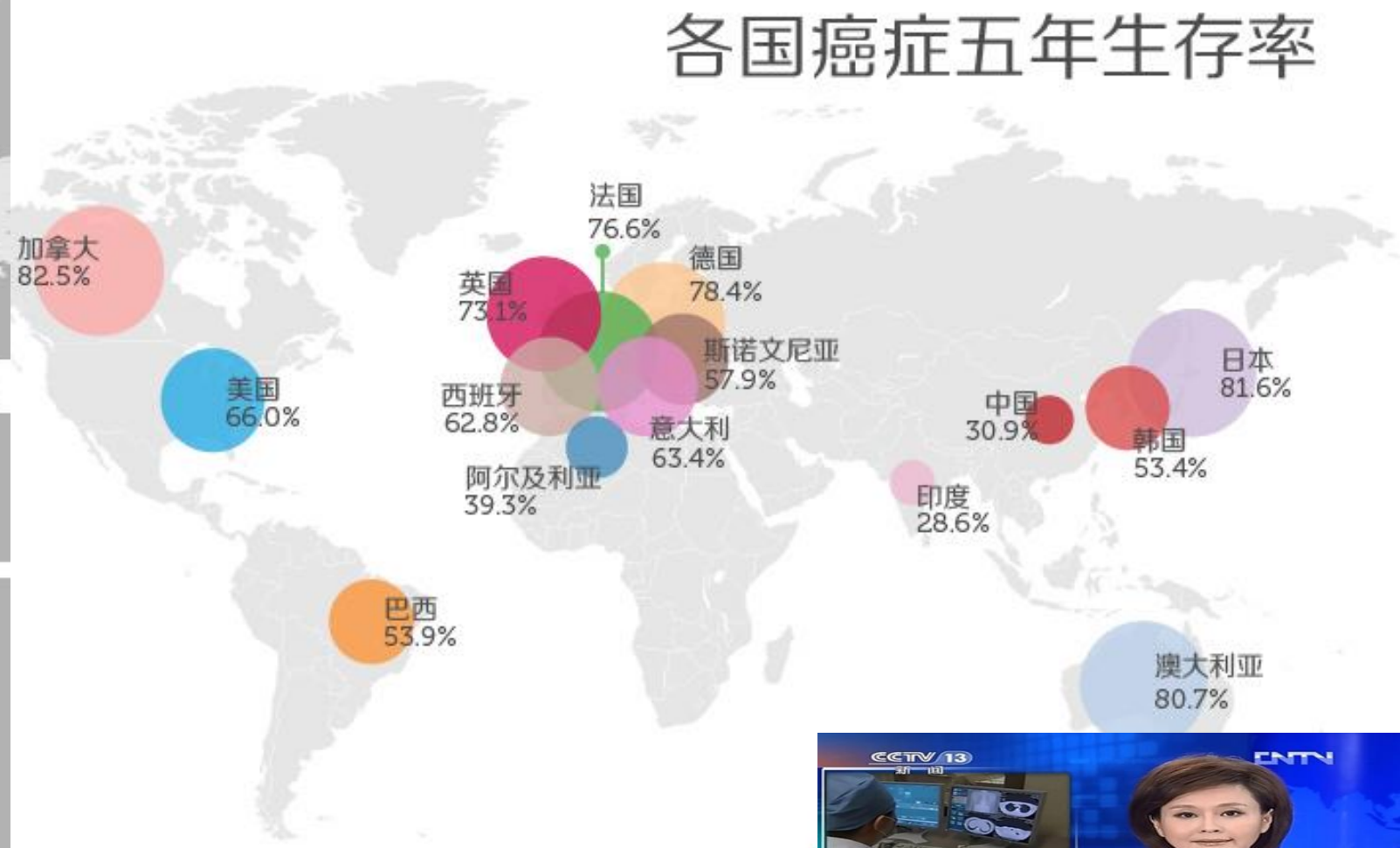
许多的癌症发病和死亡病例可以通过减少危险因素、增加临床护理的功效来防止。尤其是农村地区和弱势群体的健康保健问题。

Chen, W., et al., Cancer statistics in China, 2015. CA: A Cancer Journal for Clinicians, 2016.

Siegel, R.L., K.D. Miller, and A. Jemal, Cancer statistics, 2015. CA: A Cancer Journal for Clinicians, 2015. 65(1): p. 5-29.

中国癌症调查：五年存活率远低于发达国家



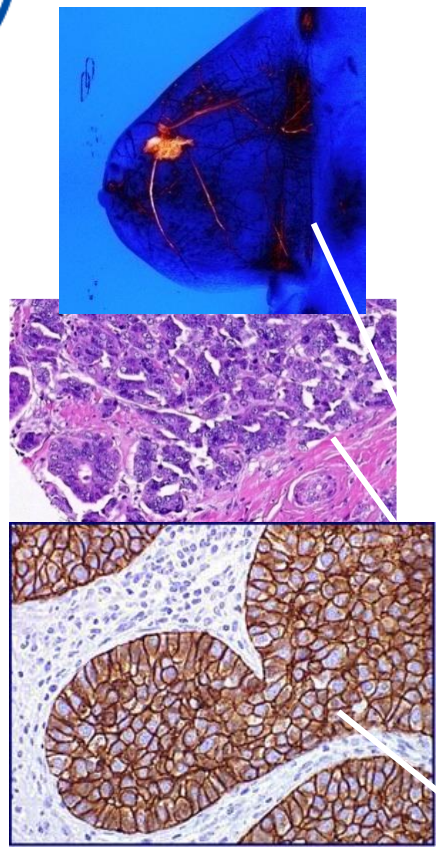


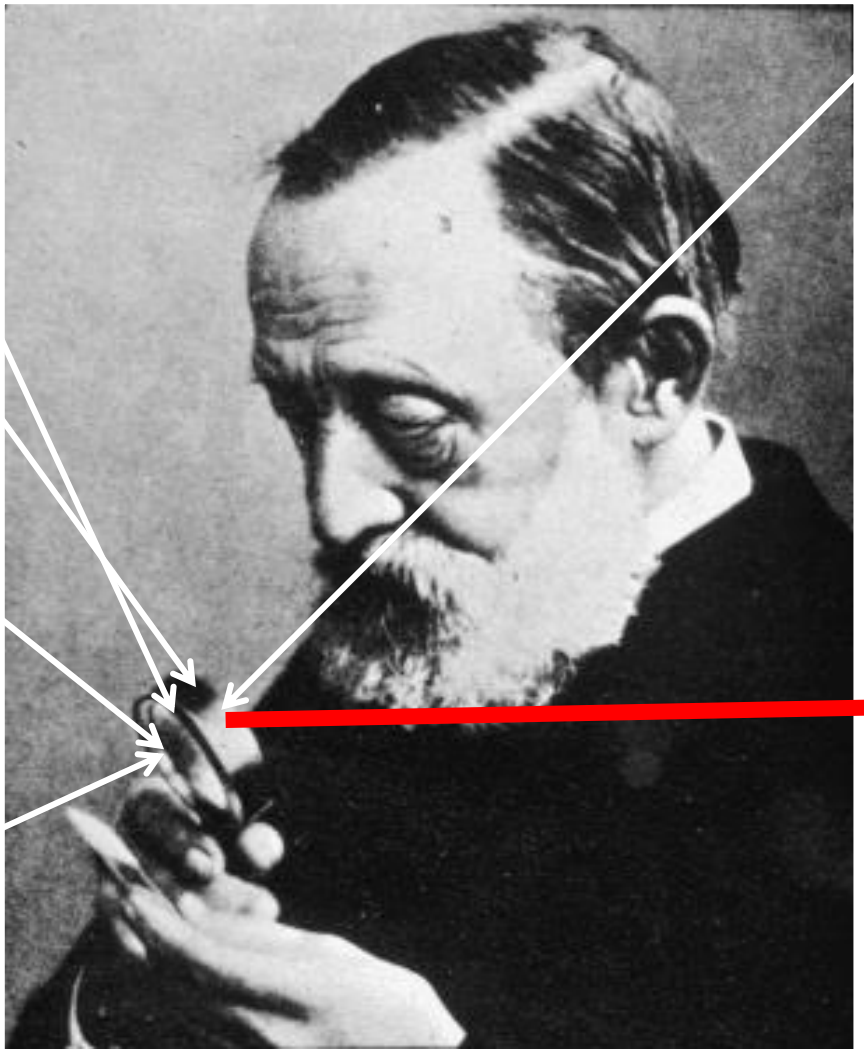
提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

组织病理图像分析是癌症诊断的“金标准”

它融合了多种模态的数据信息





信息化诊断以指导**精准医学**研究和临床护理

Courtesy to Dr. Andrew Beck

组织病理图像分析：我国现状

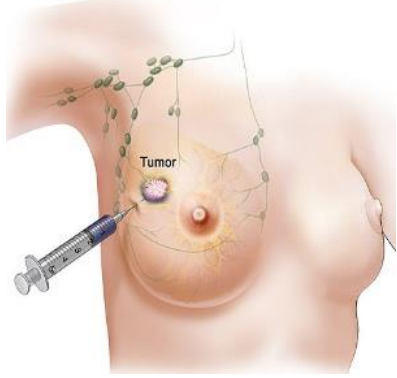
南方周末-病理科：“医学之本”竟成“科室之末”（2015.7.16）

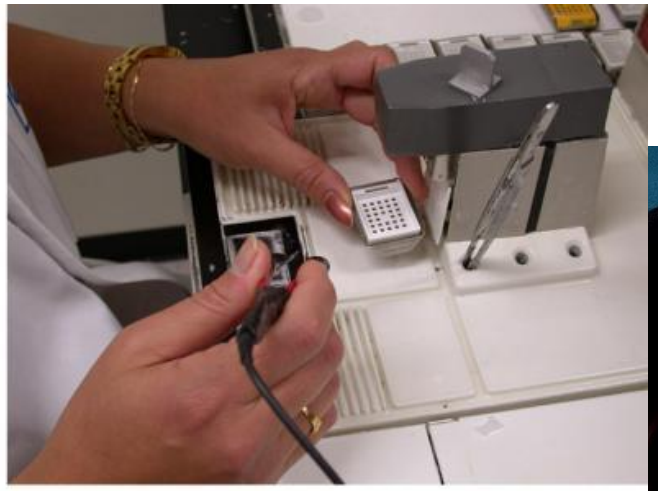
- 付丽教授每天都会看上数百张切片，而根据切片的复杂程度，每个病例需花费10～40分钟不等的时间。腰椎病、视力下降、慢性神经损害等几乎成了病理医生标配的“职业病”。
- 全国有执照的病理医生仅9000余人，按照每100张病床配备1名-2名病理科医师计算，缺口高达4万-9万人。
- 尽管影像学和各种检查技术飞速发展，但“病理诊断仍然是肿瘤各种检查方法中最可靠的金标准，也是疾病的最终诊断”
- 被“现代医学之父”威廉·奥斯勒对病理的评价“As is our pathology, so is our medicine（病理学为医学之本）”，“doctor's doctor（医师的医师）”
- 全国病理科医生不足万人，患者成为了病理科困境下最大的受害者。（导致严重的“过度治疗”和“治疗不当”）
- “没人干、没人会、没学生学”，这是中国病理界的现状。
- 地市级和基层医院的对病理会诊的需要很迫切，确实跟自身能力欠缺有关。2011年，原卫生部开展远程病理会诊平台，从申请会诊医院和专家会诊结果看，二级甲等医院初诊意见与专家会诊意见的符合率仅为35%，市级医院诊断的符合率仅37%，而县级医院诊断的符合率只有26%。

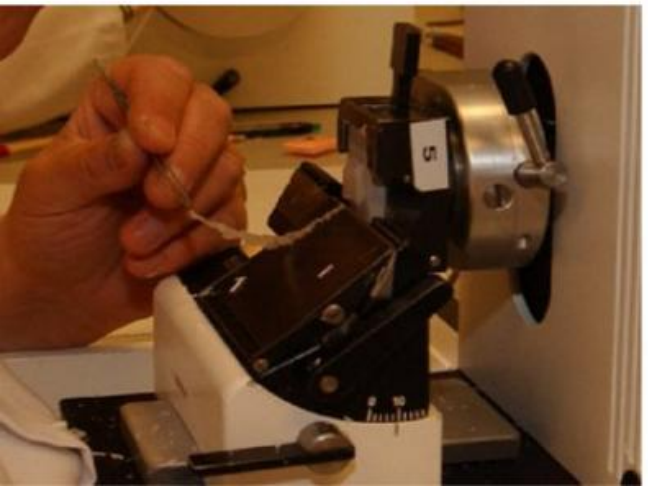
提 纲

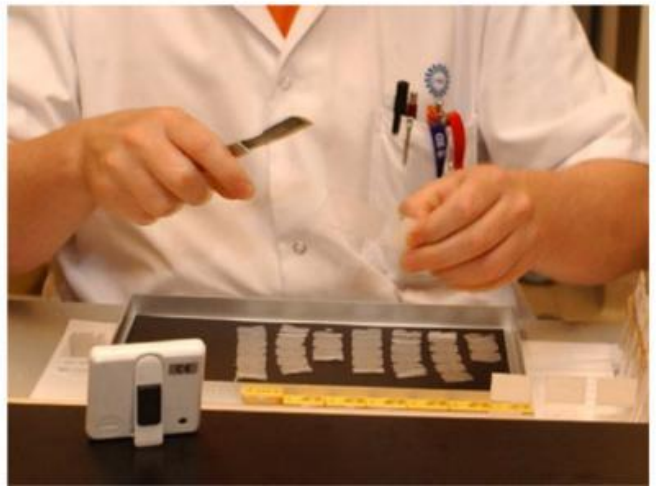
- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

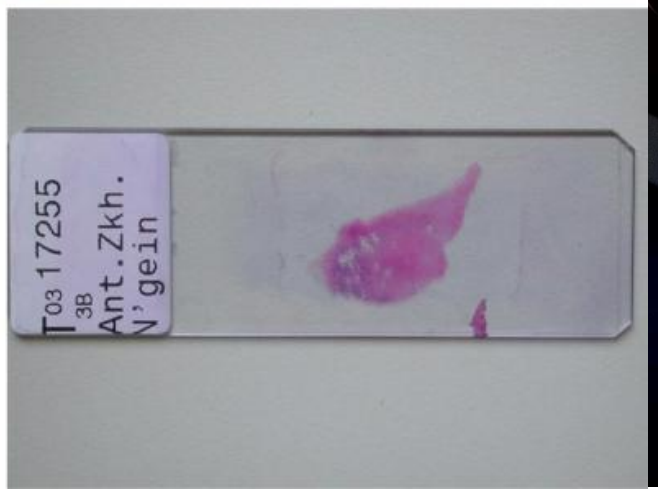
组织病理图像：组织切片的制作流程











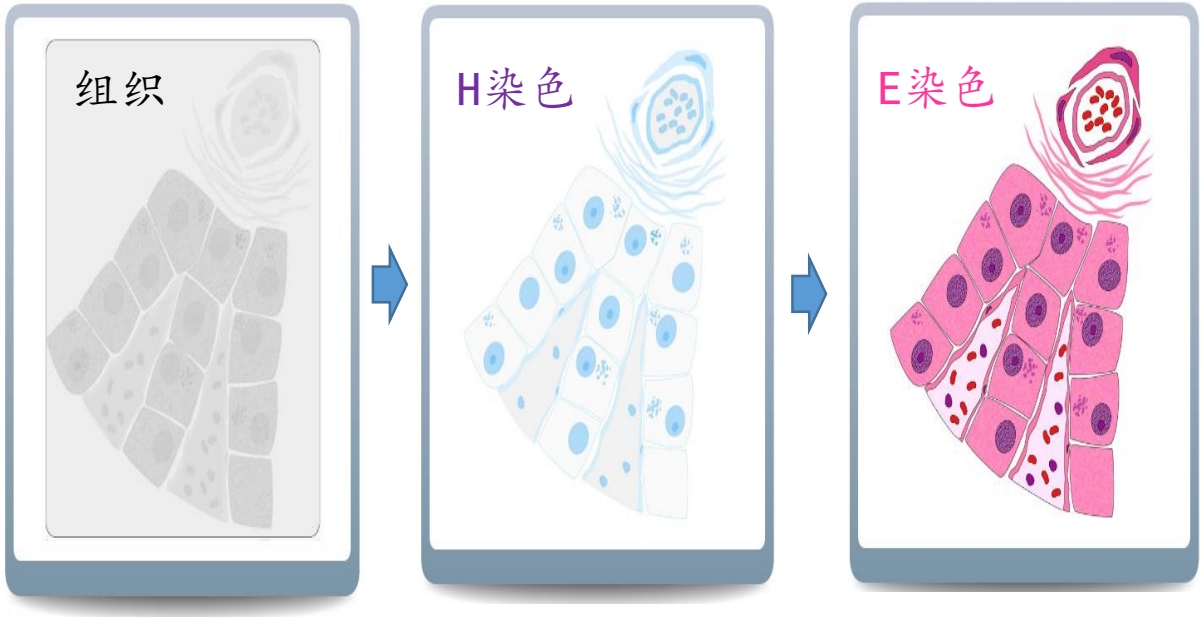
从左上到右下：1）把手术中得到的组织切成小块；2-3）把每个小组织块放到小盒并用福尔马林浸泡和石蜡固化；4）切片切割机把石蜡固定的组织切成非常薄的薄片；5)每个薄片放置到玻璃片上染色；6）H&E 染色剂染色以后的玻璃片

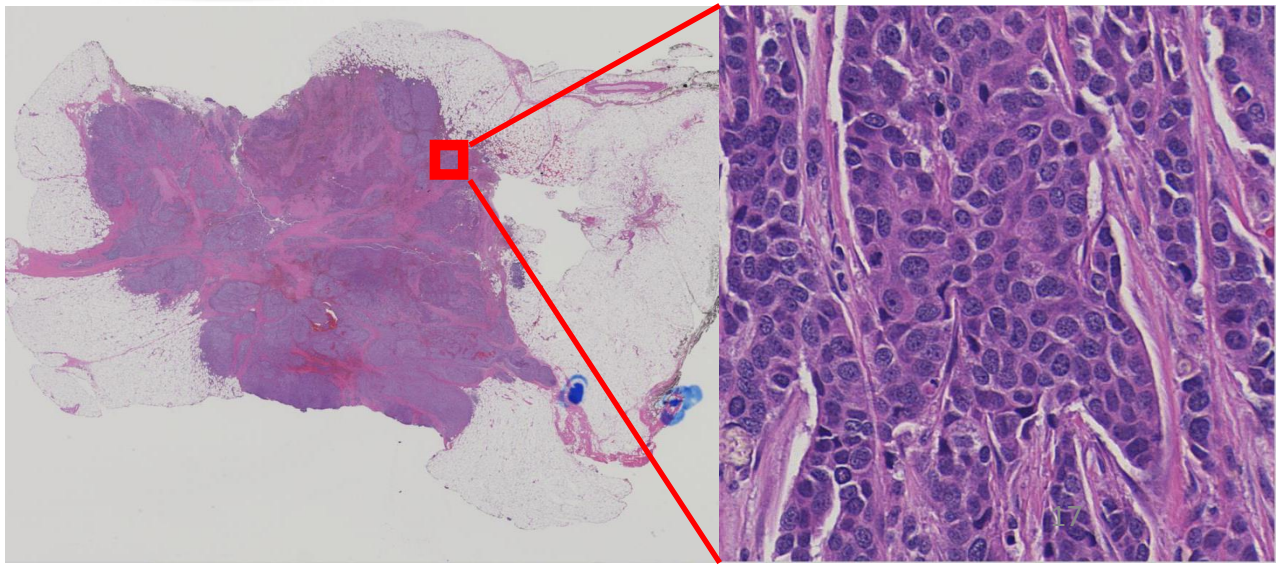
组织病理图像：组织切片的制作流程

组织病理图像:病理医生的人工分析

组织病理图像:H&E染色原理

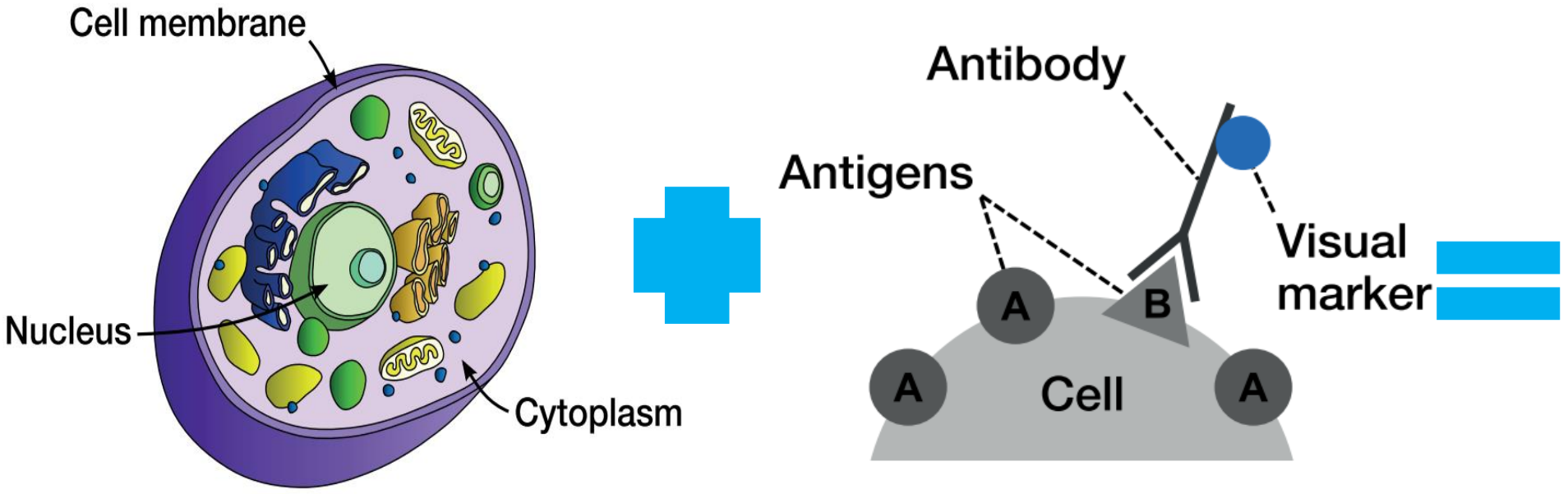
- **H&E(苏木精&曙红) 染色:**
- 这种染色是组织学和组织病理学使用最广泛的一种染色，它在临床医学诊断中被广泛采用，并且通常是金标准。比如当病理医生分析一张疑似癌症的切片的时候，通常都是分析H&E染色的切片。
- 这种染色方法的基础是组织结构对不同染料的结合程度不同。细胞的酸性和碱性成分分别和碱性染料和酸性染料结合。
- 苏木精（嗜碱性结构）染色细胞的酸性部分（核），能够展现良好的内部核细节，核染色。
- 曙红（嗜酸性结构）充当酸性染色，胞浆(细胞质)和纤维结缔组织被染成颜色深浅不一的粉红色、桔黄色、红色。

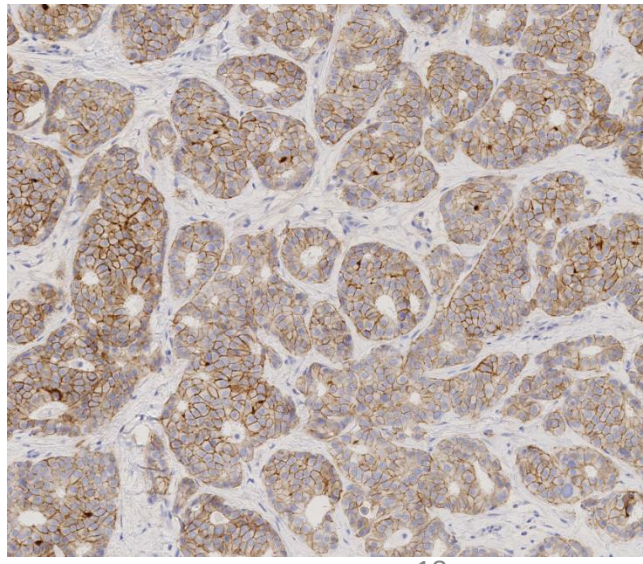




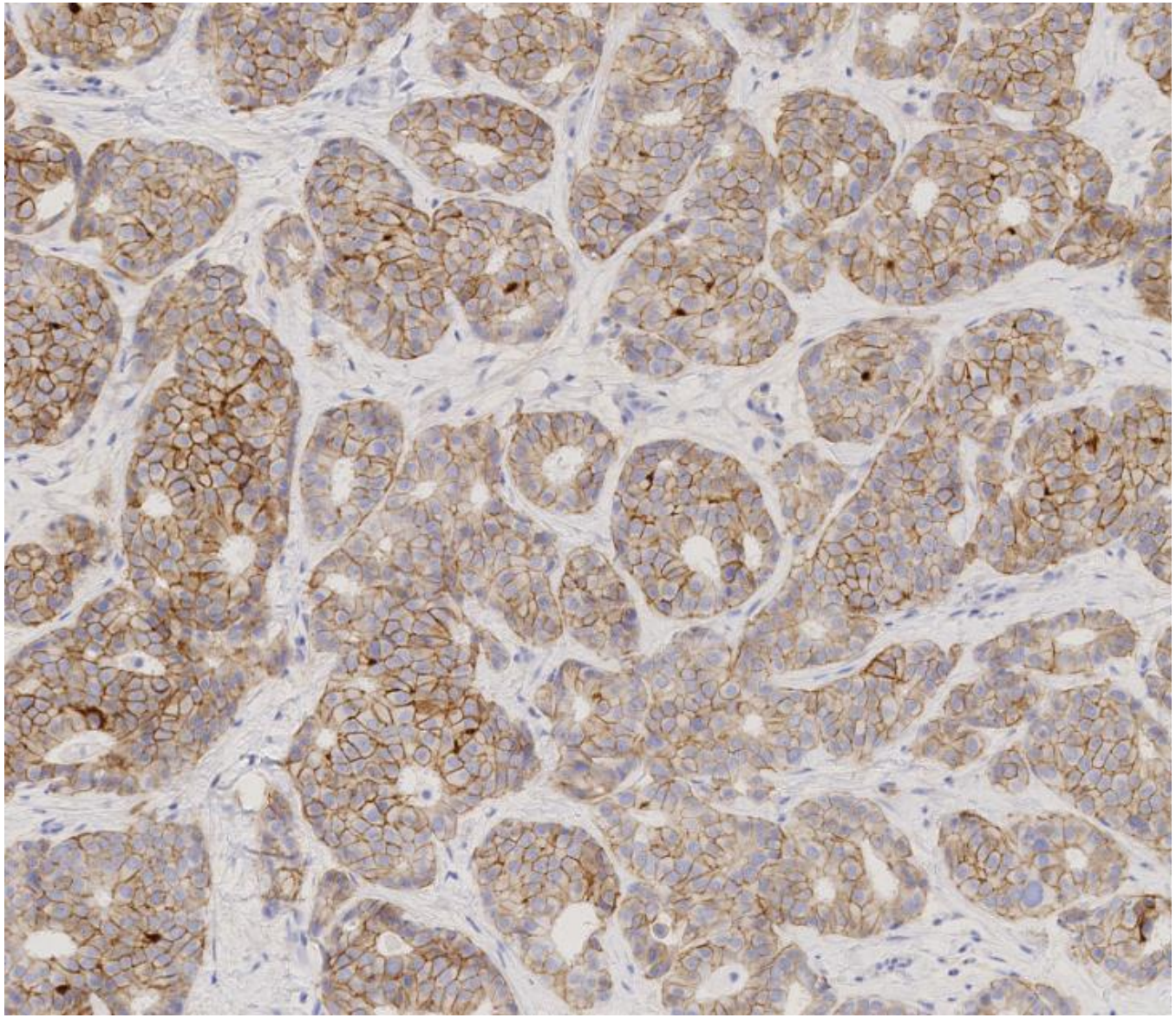
组织病理图像：IHC 染色原理

- **IHC(免疫组织化学)染色**
- IHC = immunology +histology + chemistry
- 免疫组化运用可见标记方式，通过特定的抗体、抗原的反应来识别特定的组织成分。

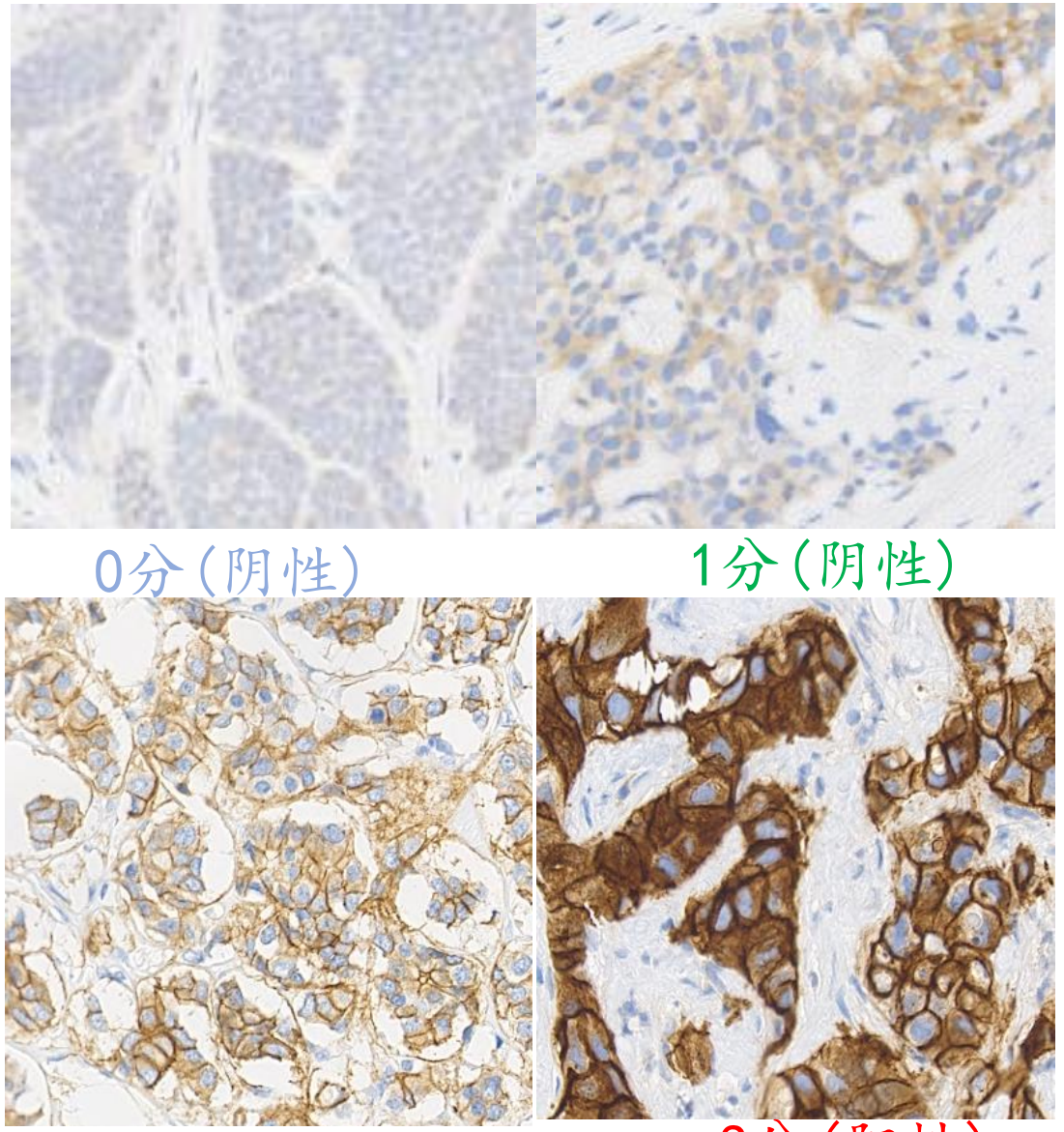




组织病理图像：HER2染色与恶性程度评分

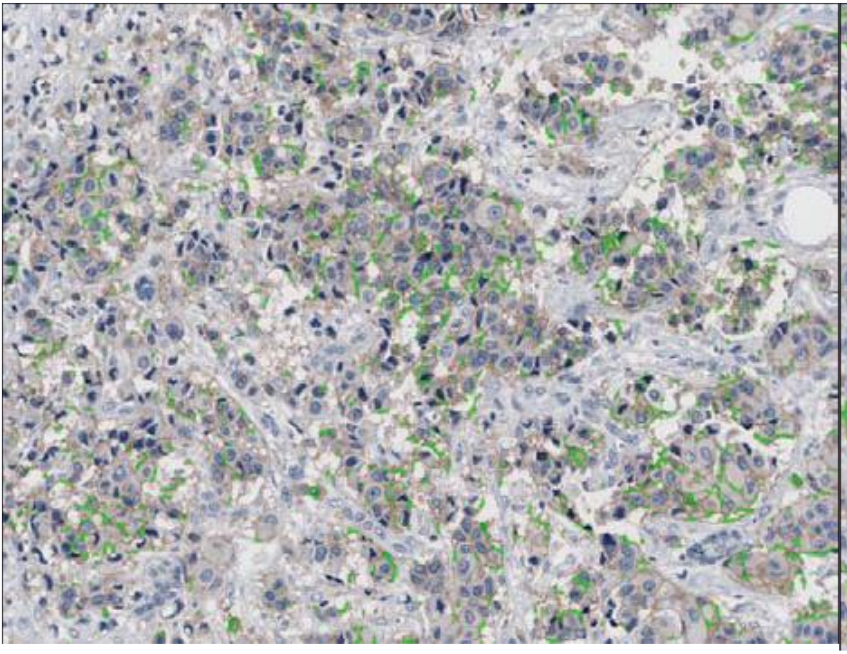


人表皮生长因子受体2(HER2)是主要影响恶性上皮细胞的生长的蛋白质，它是预测乳腺癌恶性程度高低最为重要的预后因子之一。

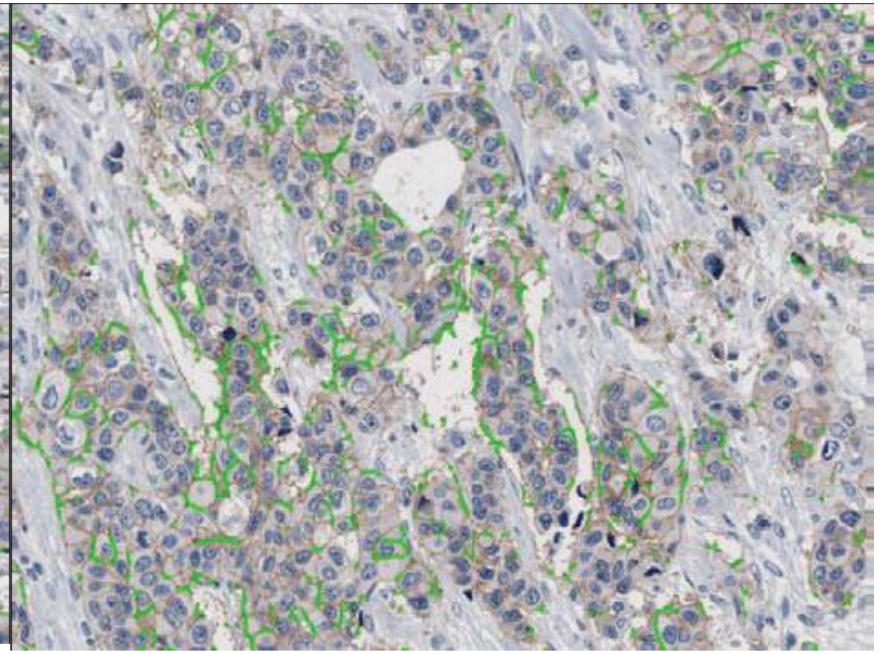


HER2蛋白0-3等级样例

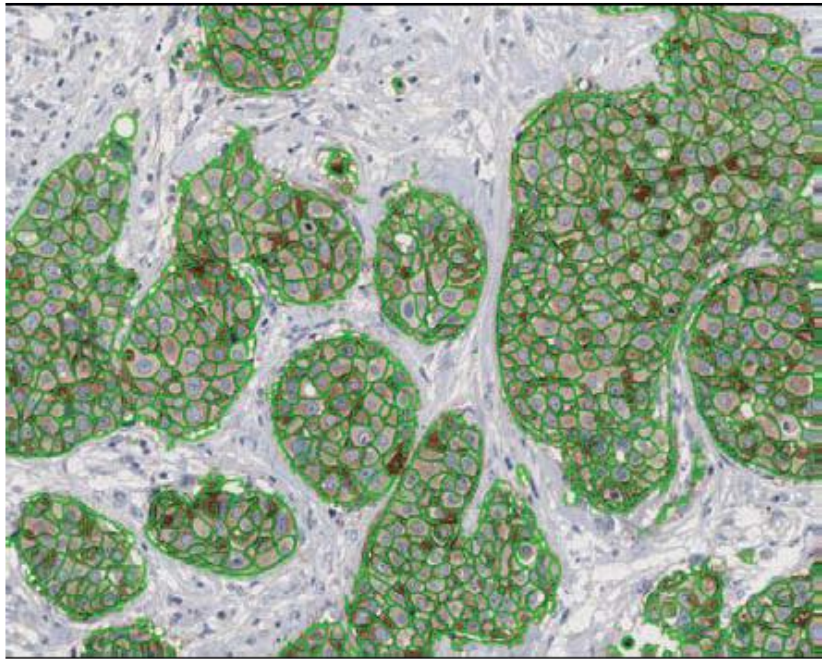
组织病理图像：HER2染色与恶性程度评分



0/1+（阴性）

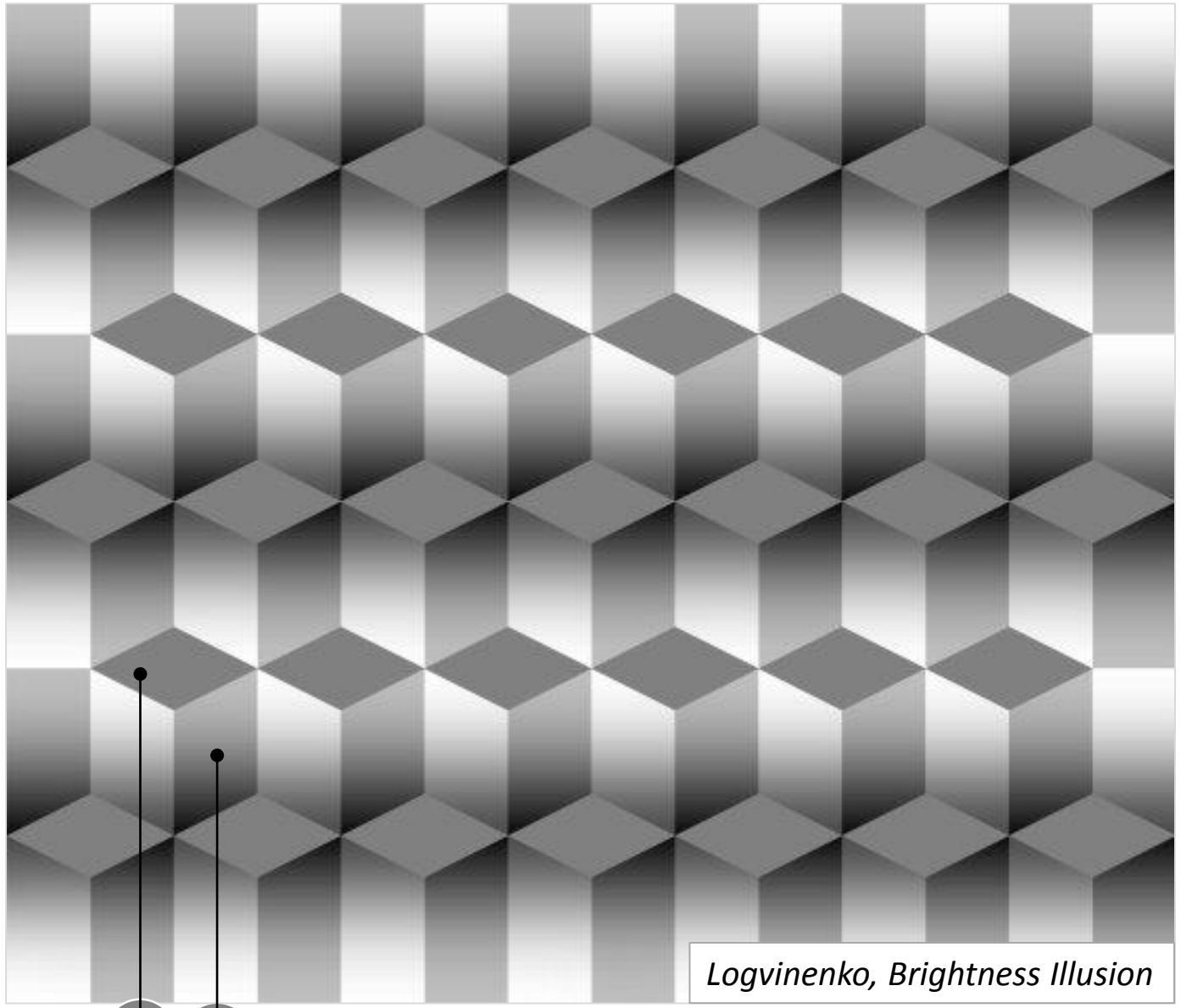


2+（中性）



3+（阳性）

视觉的亮度感应现象

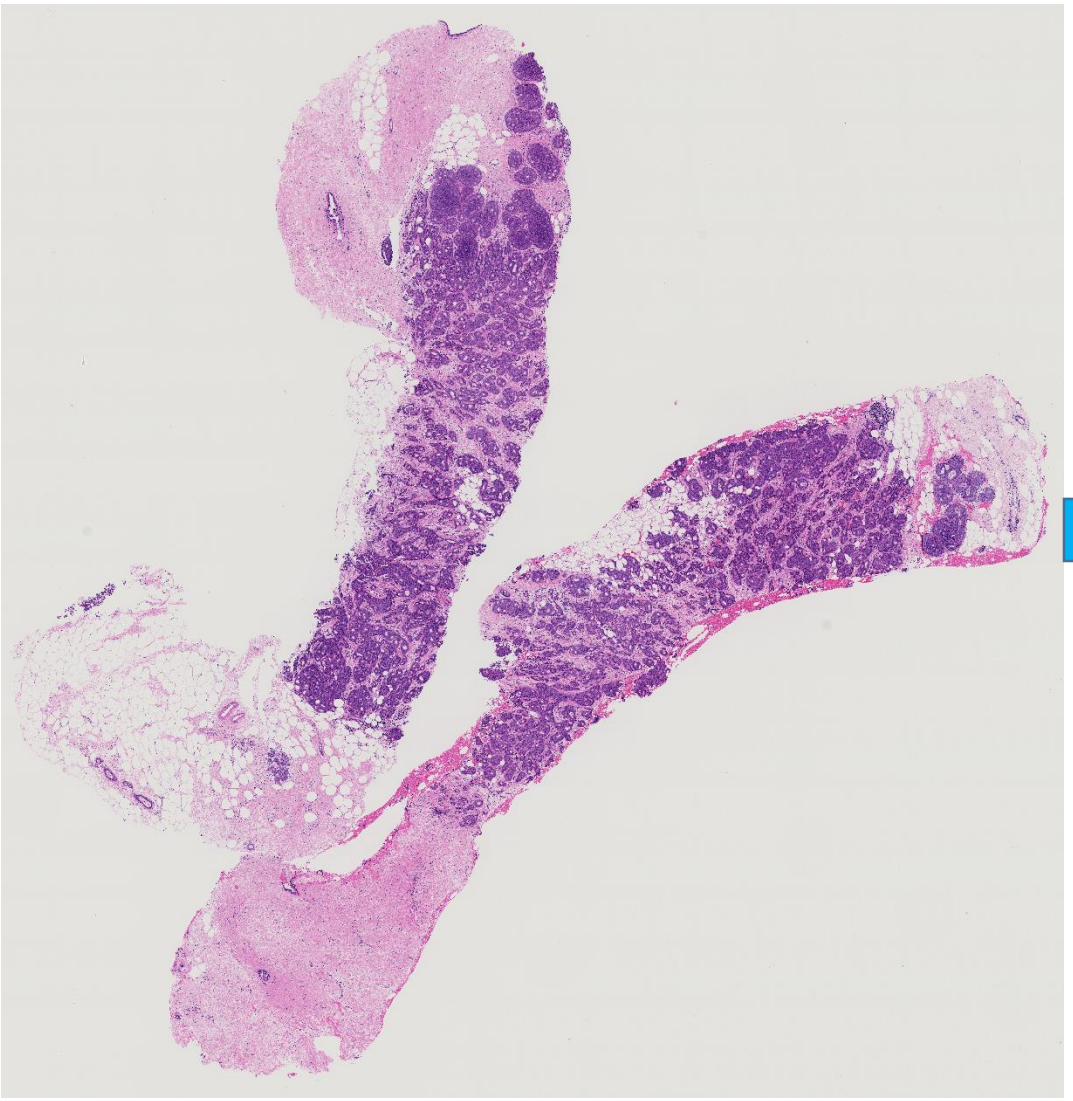


Logvinenko, Brightness Illusion

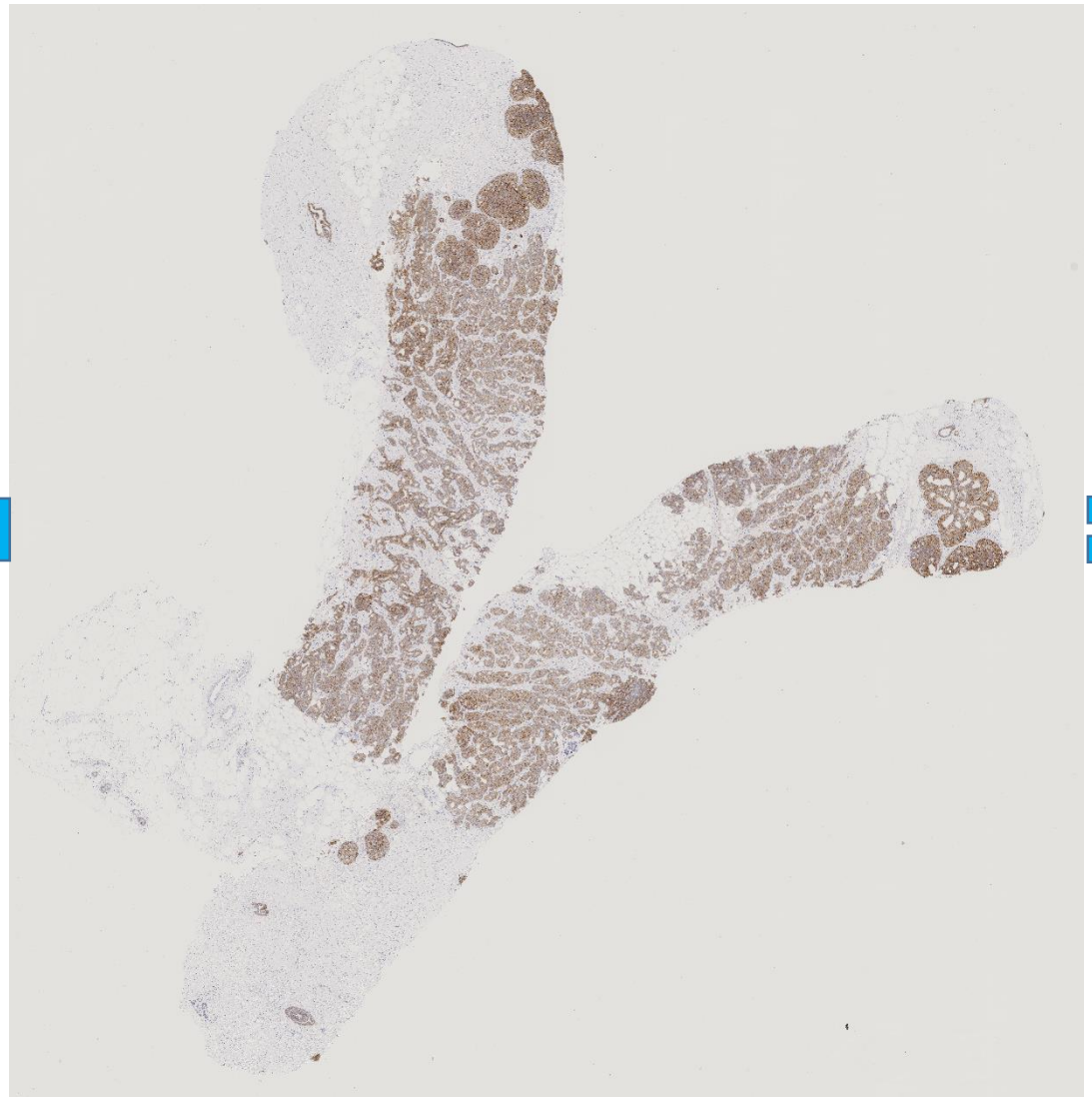
R: 132
G: 132
B: 132

(Heinemann, 1972; Yund & Armington, 1975).

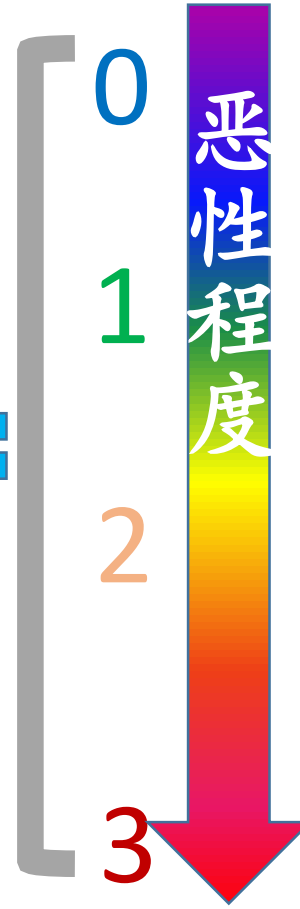
组织病理图像：H&E+IHC 染色



H&E



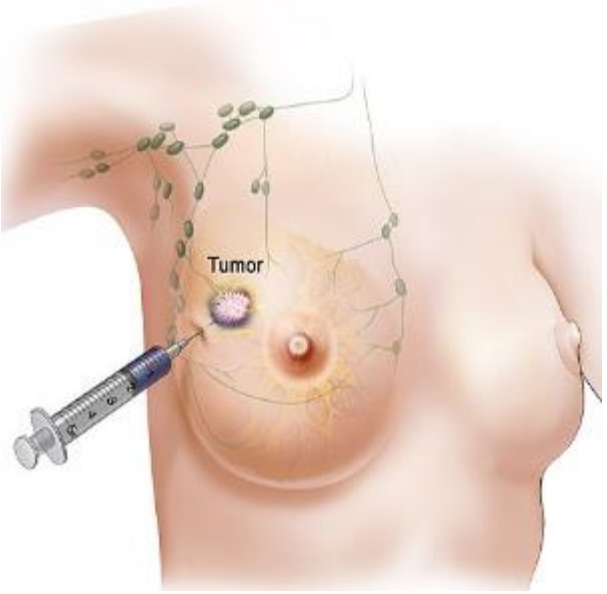
IHC



提 纲

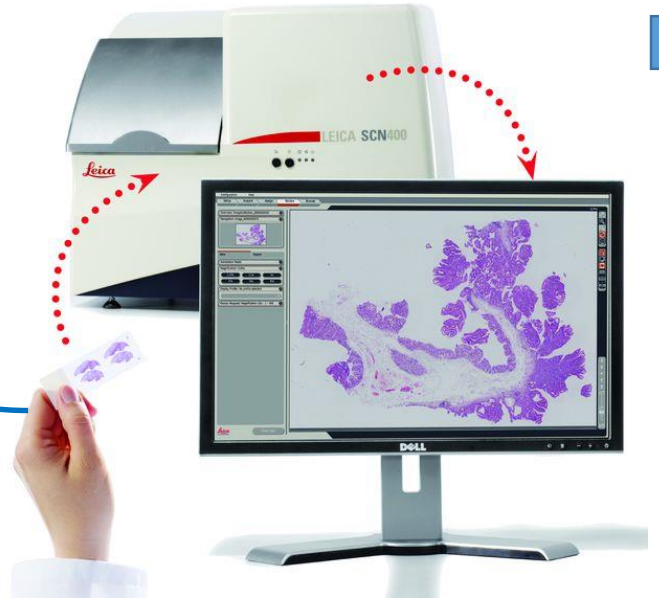
- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

组织病理图像数字化：玻璃片转化为图像

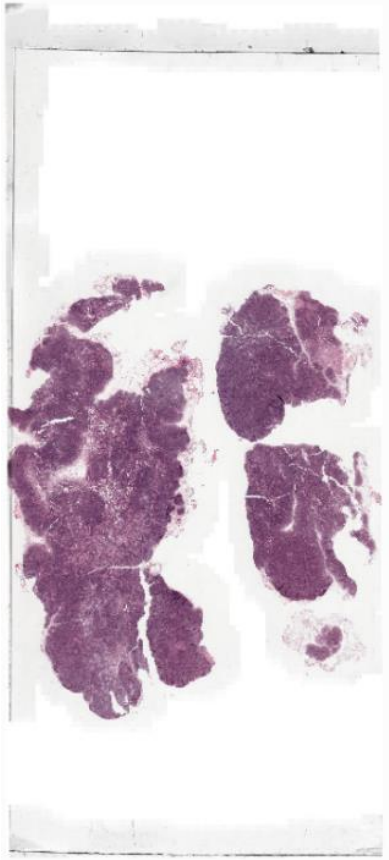


手术/活检

染色、玻璃切片制作

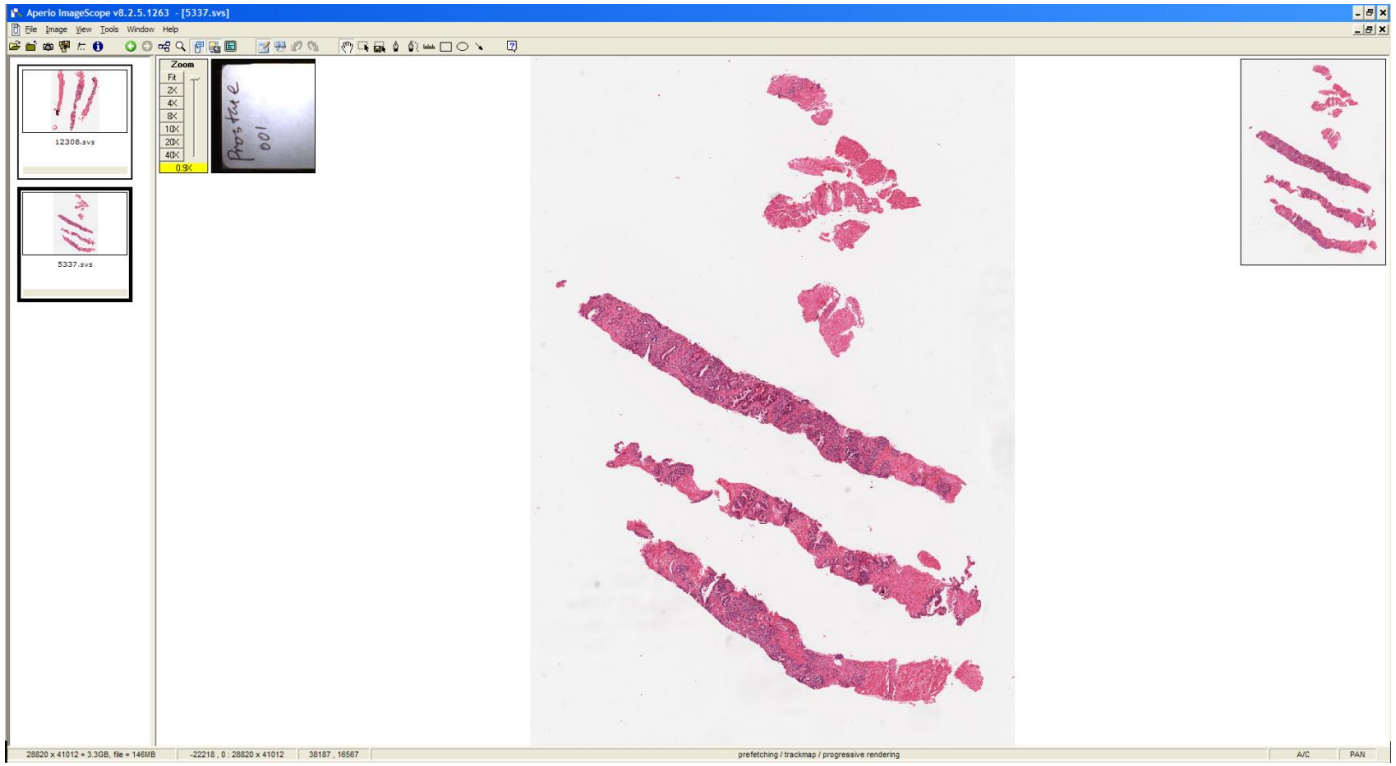


切片数字化



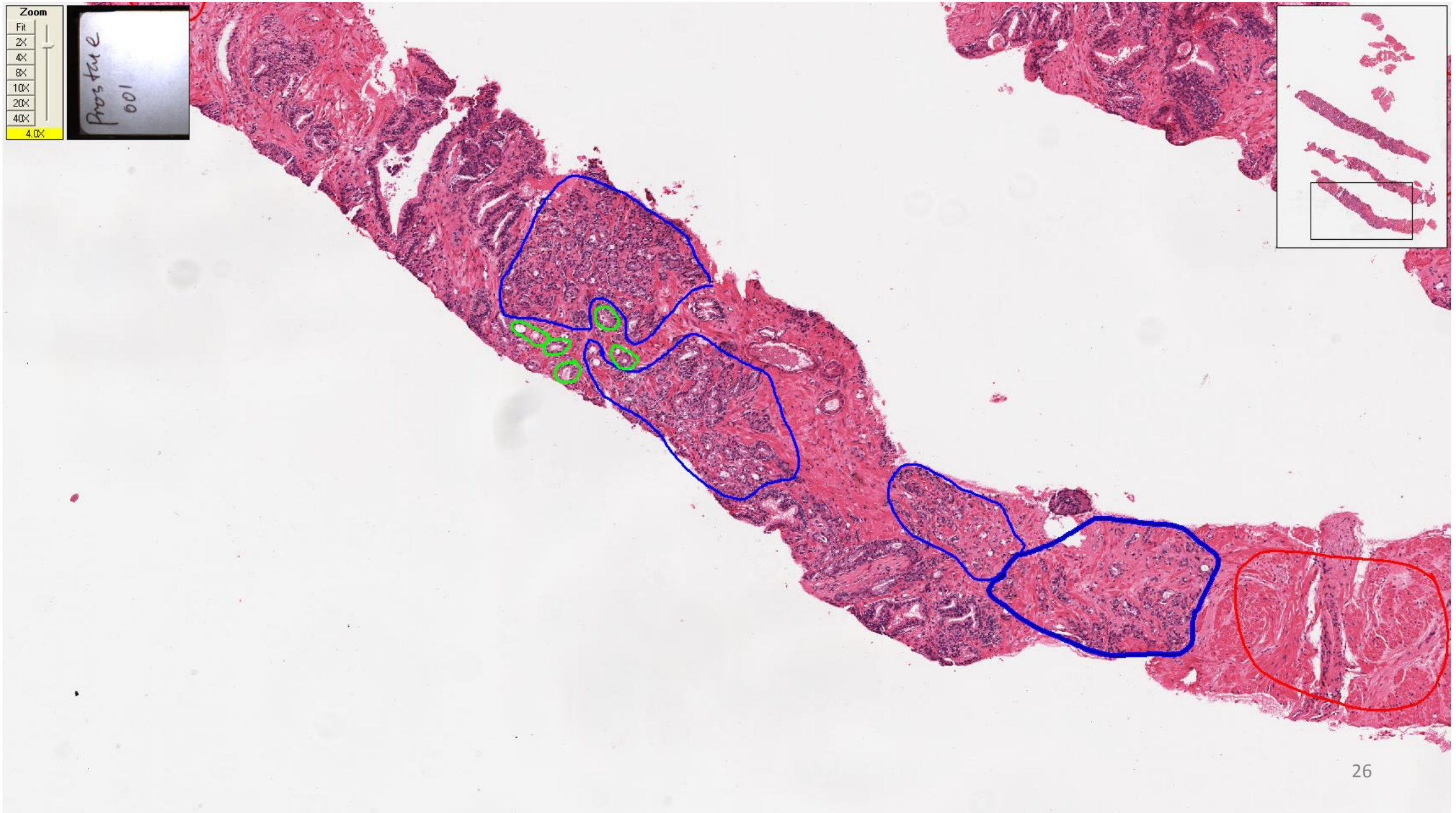
数字化图像

组织病理图像数字化：玻璃片转化为图像



- Aperio ImageScope 把玻璃片扫描成全扫描图像
- 软件便于浏览和标注数字病理图像
- 可以把单独标注的内容保存（存为XML文件）
- 可以使用任何一种编程软件读取图像（比如用MATLAB）

病理医生标注的感兴趣区域



组织病理切片数字化的意义

INNOVATIONS

A Better Lens on
DISEASE

Computerized pathology slides may help doctors make faster and more accurate diagnoses • BY MIKE MAY

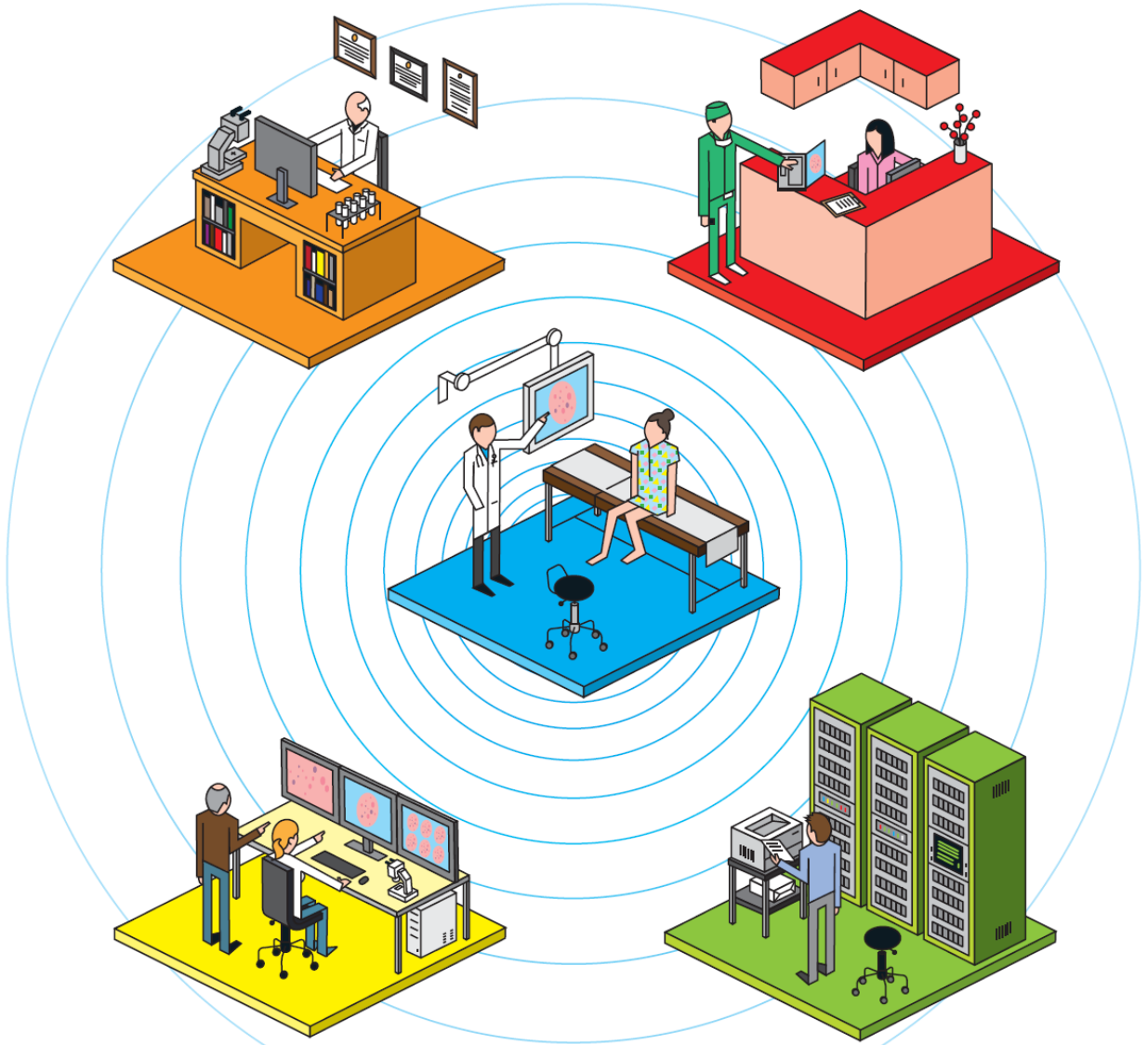
In the late 1990s Dirk G. Soenksen imagined a new future for pathology. At the time, pathologists often sat on telephone books to get a good view through their microscopes, yet Soenksen's children viewed high-resolution monitors when merely playing Nintendo. "Why can't microscopists look at computer monitors, too?" he wondered.

That question sent Soenksen on an extended journey, beginning in his garage. After 18 months of intense laboring, he emerged as the head of a newly created digital-pathology company called Aperio, which he now runs in Vista, Calif. Beyond merely moving images of diseased tissues from microscopes to computers, his technology—as well as that of other start-ups and even established health care companies—promises to make anatomical pathology, which involves the interpretation of biopsies, far more quantitative. This advance should, in turn, enhance the accuracy of diagnosing diseases and help physicians likely to be able to inspect a sample as a
file. In general, today's pathologists lack th
ity to make or obtain digitized slides, and
of such slides is approved by the U.S. Foo
Drug Administration for only a few medic
plications, all related to breast cancer.

For now, the hundreds of millions of p
ogy slides prepared annually get handled a
have for more than 100 years. A tissue s
gets cut into paper-thin, or thinner, section
a stain brings out specific features. Then,
thologist puts the glass slide under a micro
In a breast cancer biopsy, for example, a p
ogist looks for a range of features in the
including the number of abnormal cells
section and the tumor grade, the latter de
ing on features such as cell structure. "No
is done by eyes over the microscope, look
every little point," says George K. Michal
los, chair of the department of pathology
University of Pittsburgh.

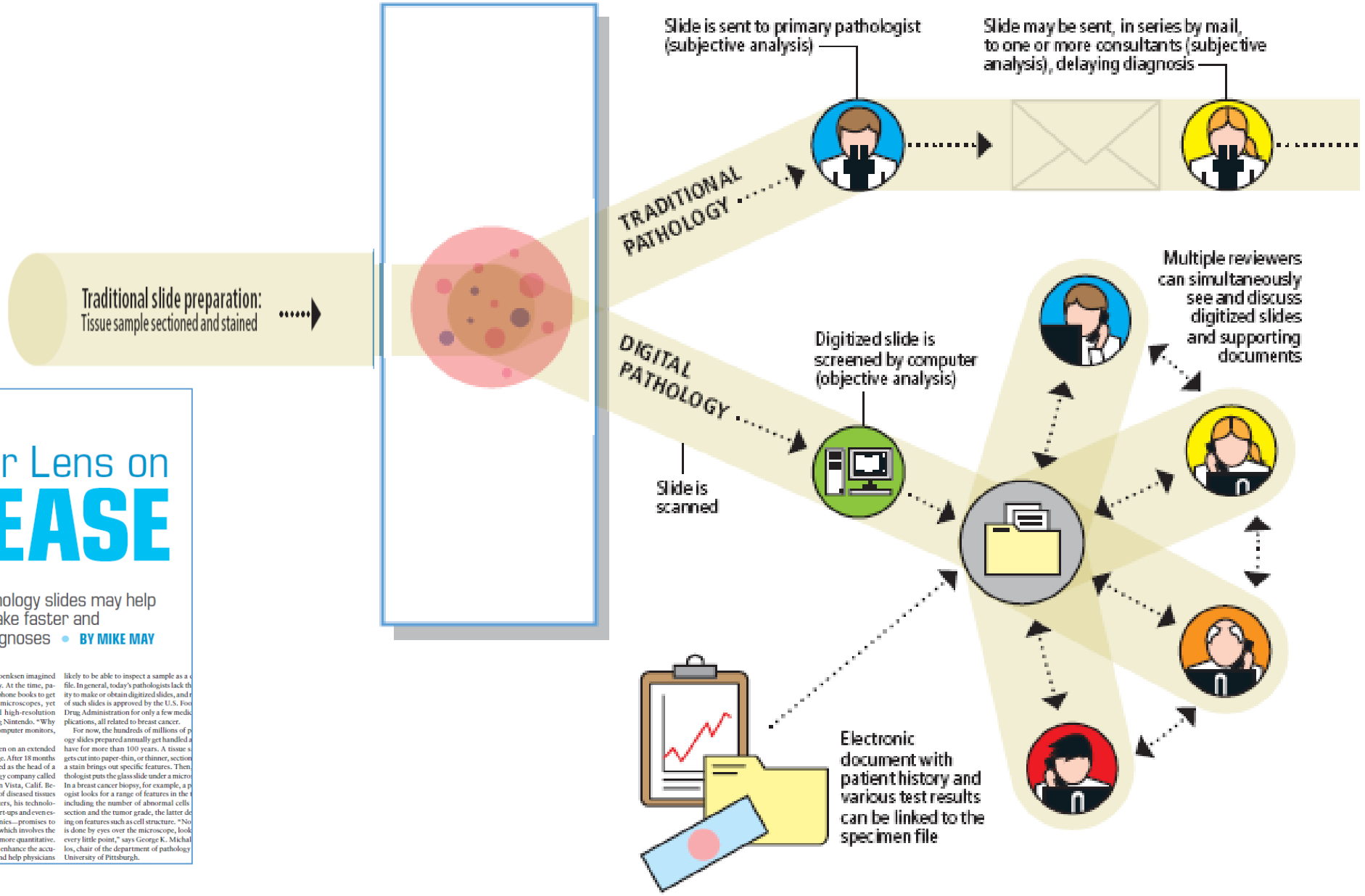
KEY CONCEPTS

- A remake of pathology, a profession that has processed samples the same way for more than 100 years, is long overdue.
- Emerging techniques allow computerized images of biopsies to be manipulated in novel ways.
- Ultimately, digital pathol-



Mike May, A better lens on disease, *Scientific American,* 2010, 302(5):74-77.

组织病理切片数字化的意义



INNOVATIONS

A Better Lens on
DISEASE

Computerized pathology slides may help doctors make faster and more accurate diagnoses • BY MIKE MAY

KEY CONCEPTS

- A remake of pathology, a profession that has processed samples the same way for more than 100 years, is long overdue.
- Emerging techniques allow computerized images of biopsies to be manipulated in novel ways.
- Ultimately, digital pathol-

In the late 1990s Dirk G. Soenksen imagined a new future for pathology. At the time, pathologists often sat on telephone books to get a good view through their microscopes, yet Soenksen's children viewed high-resolution monitors when merely playing Nintendo. "Why can't microscopists look at computer monitors, too?" he wondered.

That question sent Soenksen on an extended journey, beginning in his garage. After 18 months of intense laboring, he emerged as the head of a newly created digital-pathology company called Aperio, which he now runs in Vista, Calif. Beyond merely moving images of diseased tissues from microscopes to computers, his technology—as well as that of other start-ups and even established health care companies—promises to make anatomical pathology, which involves the interpretation of biopsies, far more quantitative. This advance should, in turn, enhance the accuracy of diagnosing diseases and help physicians likely to be able to inspect a sample as a c file. In general, today's pathologists lack th ity to make or obtain digitized slides, and r of such slides is approved by the U.S. Foo Drug Administration for only a few medic plications, all related to breast cancer.

For now, the hundreds of millions of p ogy slides prepared annually get handled a have for more than 100 years. A tissue s gets cut into paper-thin, or thinner, section a stain brings out specific features. Then, thologist puts the glass slide under a micro In a breast cancer biopsy, for example, a p ogist looks for a range of features in the including the number of abnormal cells section and the tumor grade, the latter de ing on features such as cell structure. "No is done by eyes over the microscope, look every little point," says George K. Michal los, chair of the department of pathology University of Pittsburgh.

Mike May, A better lens on disease, *Scientific American,* 2010, 302(5):74-77.

提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

组织病理图像分析：研究机遇

Scientific American 5/2010

INNOVATIONS

A Better Lens on DISEASE

Computerized pathology slides may help doctors make faster and more accurate diagnoses • BY MIKE MAY

In the late 1990s Dirk G. Soenksen imagined a new future for pathology. At the time, pathologists often sat on telephone books to get a good view through their microscopes, yet Soenksen's children viewed high-resolution monitors when merely playing Nintendo. "Why can't microscopists look at computer monitors, too?" he wondered.

That question sent Soenksen on an extended journey, beginning in his garage. After 18 months of intense laboring, he emerged as the head of a newly created digital-pathology company called Aperio, which he now runs in Vista, Calif. Beyond merely moving images of diseased tissues from microscopes to computers, his technology—as well as that of other start-ups and even established health care companies—promises to make anatomical pathology, which involves the interpretation of biopsies, far more quantitative. This advance should, in turn, enhance the accuracy of diagnosing diseases and help physicians likely to be able to inspect a sample as a
file. In general, today's pathologists lack th
ity to make or obtain digitized slides, and
of such slides is approved by the U.S. Foo
Drug Administration for only a few medic
plications, all related to breast cancer.

For now, the hundreds of millions of p
ogy slides prepared annually get handled a
have for more than 100 years. A tissue s
gets cut into paper-thin, or thinner, section
a stain brings out specific features. Then
thologist puts the glass slide under a micro
In a breast cancer biopsy, for example, a p
ogist looks for a range of features in the
including the number of abnormal cells
section and the tumor grade, the latter de
ing on features such as cell structure. "No
is done by eyes over the microscope, look
every little point," says George K. Michal
los, chair of the department of pathology
University of Pittsburgh.

KEY CONCEPTS

- A remake of pathology, a profession that has processed samples the same way for more than 100 years, is long overdue.
- Emerging techniques allow computerized images of biopsies to be manipulated in novel ways.
- Ultimately, digital pathol-

Science Translational Medicine 11/2011

RESEARCH ARTICLE

IMAGING

Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival

Andrew H. Beck,[1,2]* Ankur R. Sangoi,[1,3] Samuel Leung,[4] Robert J. Marinelli,[5] Torsten O. Nielsen,[4] Marc J. van de Vijver,[6] Robert B. West,[1] Matt van de Rijn,[1] Daphne Koller[7]†

The morphological interpretation of histologic sections forms the basis of diagnosis and pro
cancer. In the diagnosis of carcinomas, pathologists perform a semiquantitative analysis of a
phological features to determine the cancer's histologic grade. Physicians use histologic gradi
assessment of a carcinoma's aggressiveness and a patient's prognosis. Nevertheless, the determ
in breast cancer examines only a small set of morphological features of breast cancer epithelia
been largely unchanged since the 1920s. A comprehensive analysis of automatically quantitate
features could identify characteristics of prognostic relevance and provide an accurate and rep
for assessing prognosis from microscopic image data. We developed the C-Path (Computatio
system to measure a rich quantitative feature set from the breast cancer epithelium and stroma
including both standard morphometric descriptors of image objects and higher-level contextu
global image features. These measurements were used to construct a prognostic model. We ap
system to microscopic images from two independent cohorts of breast cancer patients [from
Cancer Institute (NKI) cohort, $n = 248$, and the Vancouver General Hospital (VGH) cohort, $n = 328$
model score generated by our system was strongly associated with overall survival in both the
cohorts (both log-rank $P \leq 0.001$). This association was independent of clinical, pathological, and
Three stromal features were significantly associated with survival, and this association was s
association of survival with epithelial characteristics in the model. These findings implicate stromal morpho-
logic structure as a previously unrecognized prognostic determinant for breast cancer.

Andrew Beck @ Harvard Medical School

INTRODUCTION

In the mid-19th century, it was first appreciated that the process of carcinogenesis produces characteristic morphologic changes in cancer cells (1). Patey and Scarff showed in 1928 (2) that three histologic features—tubule formation, epithelial nuclear atypia, and epithelial mitotic activity—could each be scored qualitatively, and the assessments could be combined to stratify breast cancer patients into three groups that showed significant survival differences. This semiquantitative morphological scoring scheme has been refined over the years (3–5) but still remains the standard technique for histologic grading in invasive breast cancer. Although considerable effort has been devoted recently to molecular profiling for assessment of prognosis and prediction of treatment response in cancer (6, 7), microscopic image assessment is still the most commonly available (and in some places in the world, the only) tool that is financially and logistically feasible.

Although the three epithelial features scored in current grading systems are useful in assessing cancer prognosis, valuable prognostic information can also be derived from other factors, including properties of the cancer stroma such as its molecular characteristics (8–15) and morphological features [such as stromal fibrotic focus, a scar-like area in the center of a carcinoma (16)]. Thus, we sought to develop a high-accuracy, image-based predictor to identify new clinically predictive morphologic phenotypes of breast cancers, thereby providing new insights into the biological factors driving breast cancer progression.

The development of such a system could also address other problems relevant to the clinical treatment of breast cancer. A limitation to the current grading system is that there is considerable variability in histologic grading among pathologists (17), with potentially negative consequences for determining treatment. An automated system could provide an objective method for predicting patient prognosis directly from image data. Moreover, once established, this system could be used in breast cancer clinical trials to provide an accurate, objective means for assessing breast cancer morphologic characteristics, allowing objective stratification of breast cancer patients on the basis of morphologic criteria and facilitating the discovery of morphologic features associated with response to specific therapeutic agents.

[1]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Biomedical Informatics Training Program, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Department of Pathology, El Camino Hospital, Mountain View, CA 94040, USA. [4]Genetic Pathology Evaluation Centre, University of British Columbia, Vancouver, British Columbia V6H 3Z6, Canada. [5]Department of Biochemistry, Stanford University, Stanford, CA 94305, USA. [6]Department of Pathology, Academic Medical Center, Meibergdreef 9, 1105AZ Amsterdam, Netherlands. [7]Department of Computer Science, Stanford University, Stanford, CA 94305, USA.
*Present address: Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA.
†To whom correspondence should be addressed. E-mail: koller@cs.stanford.edu

RESULTS

Experimental design overview

We developed the Computational Pathologist (C-Path), a machine learning–based method for automatically analyzing cancer images and predicting prognosis. To construct and evaluate the model, we acquired hematoxylin and eosin (H&E)–stained histological images from breast cancer tissue microarrays (TMAs) (figs. S4 and S5). The

Science Translational Medicine 11/2011

FOCUS

COMPUTATIONAL MEDICINE

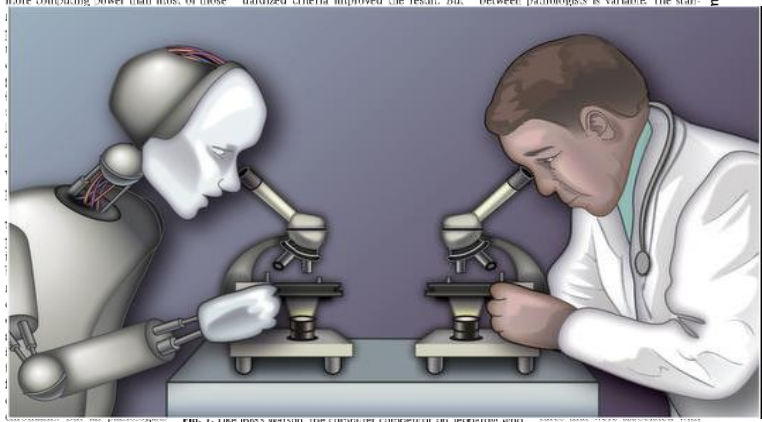
C-Path: A Watson-Like Visit to the Pathology Lab

David L. Rimm

Computer-based quantification of tumor morphology has arguably solved the problem of standardized tumor grading (Beck *et al.*, this issue).

MAN VERSUS MACHINE

Since computers were invented, technophiles have tried to generate a computer algorithm that could essentially "be" a pathologist, providing an objective readout for tumor grade. For the most part, early machines simply measured distances between features (such as the diameter of the nucleus in microns). Given that the average wristwatch has more computing power than most of those human pathologists have expertise to different degrees, which results in subjectivity in diagnosis. This fact was dramatically illustrated by Rosai when he collected reviews from a panel of expert pathologists to make a specific diagnosis of premalignancy (2). The agreement between the experts was surprisingly low. Scanitt and colleagues responded a year later (3), showing that standardized criteria improved the result. But



know that there are some features that are difficult to describe, hard to teach, and, in turn, hard to learn, often requiring more than 10 years before a pathologist can be considered "expert."

So, it is has been notoriously difficult to emulate that expertise with a machine. Even

Department of Pathology, Yale University School of Medicine, New Haven, CT 06520, USA.
E-mail: david.rimm@yale.edu

puter system at the time. By the early 1990s, automation first entered commercial anatomic pathology as PAPNET—a computer-aided screen for abnormal cells in Pap smears (5). Although this system failed to change patient care, it heralded a new age in anatomic pathology in which computers could assist

COMPUTATIONAL PATHOLOGY

One of the most critical subjective tumor-evaluation parameters is histologic grade. Although there are standardized criteria for grading different histologies, the agreement between pathologists is variable. The stan-

in diagnostic cytology. Now, around 20 years later, computer-assisted gynecologic cytology screening is broadly used in instruments such as the FocalPoint (Becton Dickinson) (6) and the ThinPrep (Hologic) (7) imaging systems. Similarly, there are computers in every lab that operate instruments, coordinate information systems, and manage and display images. But even now, 75 years after the invention of the computer, it has still not penetrated the realm of diagnostic anatomic pathology. As described by Beck and colleagues, C-Path could someday transform the use of computers in pathology and medicine (1) (Fig. 1).

Fig. 1. Like IBM's Watson, the computer competitor on Jeopardy! who defeated all human opponents, C-Path (1) could advance pathology beyond the subjective capabilities of its human counterparts.

After a series of computational validation steps, the authors were able to show that the resultant score was significantly associated with survival in the training set.

The exciting part of this study was how well the computer was able to predict survival on the basis of tissue samples from a

组织病理图像分析：研究机遇

南京信息工程大学 Nanjing University of Information Science & Technology

Nature 10/2013

OUTLOOK | MEDICAL IMAGING

Will machines be able to judge a patient's prognosis? This prototype microscope aims to do part of the job.

SOFTWARE

The computer will see you now

From image-analysis software to lens-free microscopes that fit on a mobile phone, new tools are providing pathologists with clearer and more informative images.

BY KATHERINE BOURZAC

In the seventeenth century, natural historians such as Galileo, Antonie van Leeuwenhoek and Robert Hooke learned to grind lenses and make the first microscopes, revealing the hidden landscapes of life. They saw for the first time the cells in cork, blood and other tissues, and van Leeuwenhoek found swimming 'animalcules' in dental plaque and observed the movement of sperm.

Physicists and engineers are now trying to bring about a similar shift in perspective for microscopy. In most pathology labs, doctors diagnose diseases by poring over tissue slices on glass microscope slides — classifying tumours, for example, based on subtle visual cues that are difficult to quantify. But this is starting to change. Just as lenses once revealed vistas that were previously invisible to the human eye, so software is opening up a new window on biology.

The latest digital tools make it possible to do a visual search in microscopy images, automate diagnosis, and sync image data with the genomic profiles of tumours. Some researchers are even doing away with lenses altogether, creating computational microscopes based on inexpensive hardware that could be used for point-of-care diagnostics, particularly in poor areas with few doctors.

BIG DATA

Pathology has remained stubbornly analogue and qualitative, however. The experienced pathologist's main tools are glass slides, a compound microscope whose design has hardly changed in more than 200 years, and eyes that have seen thousands of tumours. "Most of a pathologist's medical decisions are based on morphology," the structural details of cells and tissues revealed under a microscope, says David Rimm, a pathologist at the Yale School of Medicine in Connecticut.

Just because a method is old is no reason to abandon it, of course. But advocates of digital pathology worry about inconsistencies that can lead to false negatives and misdiagnoses. Experienced pathologists are better than younger ones at identifying rare tumours, but they often disagree with one another and even with their own assessment of the same sample from weeks before.

One hurdle to digitizing clinical microscopy is the size and complexity of the images, says Metin Gurcan, who specializes in biomedical informatics at Ohio State University and was an early advocate of digital pathology. First, a biopsy is sliced into sections and placed on multiple slides. A digital image of a single slide, magnified under the microscope, has about 10 billion pixels and requires about 30 gigabytes of memory. A typical prostate biopsy, for example, uses more than 20 slides and needs about 600 gigabytes.

That's a lot of information for pathologists to scan through — and a lot of data for software to sift. "The number and type of cells found

NEJM 02/2015

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

FEBRUARY 26, 2015

Francis S. Collins @ NIH

A New Initiative on Precision Medicine

Francis S. Collins, M.D., Ph.D., and Harold Varmus, M.D.

"Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier."

— President Barack Obama, State of the Union Address, January 20, 2015

President Obama has long expressed a strong conviction that science offers great potential for improving health. Now, the President has announced a research initiative that aims to accelerate progress toward a new era of precision medicine (www.whitehouse.gov/precisionmedicine). We believe that the time is right for this visionary initiative, and the National Institutes of Health (NIH) and other partners will work to achieve this vision.

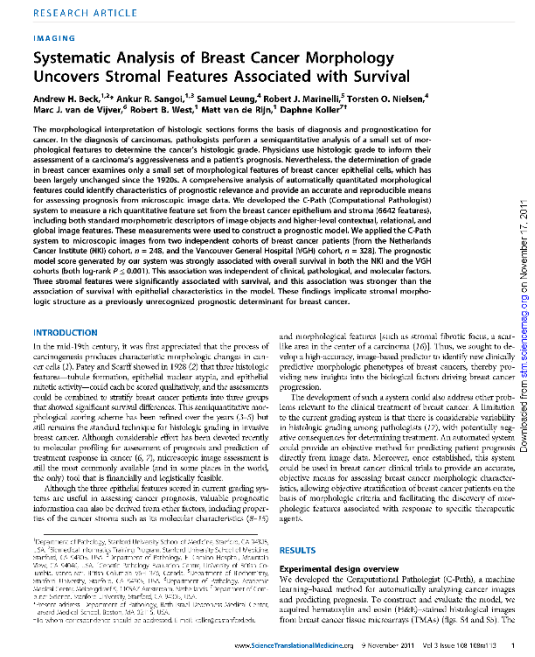
The concept of precision medicine — prevention and treatment strategies that take individual variability into account — is not new[1]; blood typing, for instance, has been used to guide blood transfusions for more than a century. But the prospect of applying this concept broadly has been dramatically improved by the recent development of large-scale biologic databases (such as the human genome sequence), powerful methods for characterizing patients (such as proteomics, metabolomics, genomics, diverse cellular assays, and even mobile health technology), and computational tools for analyzing large sets of data. What is needed now is a broad research program to encourage creative approaches to precision medicine, test them rigorously, and ultimately use them to build the evidence base needed to guide clinical practice.

The proposed initiative has two main components: a near-term focus on cancers and a longer-term aim to generate knowledge applicable to the whole range of health and disease. Both components are now within our reach because of advances in basic research, including molecular biology, genomics, and bioinformatics. Furthermore, the initiative taps into converging trends of increased connectivity, through social media and mobile devices, and Americans' growing desire to be active partners in medical research.

Oncology is the clear choice for enhancing the near-term impact of precision medicine. Can-

C-Path = Computerized Pathologist

Dr. Andrew Beck @
Harvard Medical School



RESEARCH ARTICLE

IMAGING

Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival

Andrew H. Beck,[1,2]* Ankur R. Sangoi,[1,3] Samuel Leung,[4] Robert J. Marinelli,[5] Torsten O. Nielsen,[4] Marc J. van de Vijver,[6] Robert B. West,[1] Matt van de Rijn,[1] Daphne Koller[7]†

The morphological interpretation of histologic sections forms the basis of diagnosis and prognostication for cancer. In the diagnosis of carcinomas, pathologists perform a semiquantitative analysis of a small set of morphological features to determine the cancer's histologic grade. Physicians use histologic grade to inform their assessment of a carcinoma's aggressiveness and a patient's prognosis. Nevertheless, the determination of grade in breast cancer examines only a small set of morphological features of breast cancer epithelial cells, which has been largely unchanged since the 1920s. A comprehensive analysis of automatically quantitated morphological features could identify characteristics of prognostic relevance and provide an accurate and reproducible means for assessing prognosis from microscopic image data. We developed the C-Path (Computational Pathologist) system to measure a rich quantitative feature set from the breast cancer epithelium and stroma (6642 features), including both standard morphometric descriptors of image objects and higher-level contextual, relational, and global image features. These measurements were used to construct a prognostic model. We applied the C-Path system to microscopic images from two independent cohorts of breast cancer patients [from the Netherlands Cancer Institute (NKI) cohort, $n = 248$, and the Vancouver General Hospital (VGH) cohort, $n = 328$]. The prognostic model score generated by our system was strongly associated with overall survival in both the NKI and the VGH cohorts (both log-rank $P < 0.001$). This association was independent of clinical, pathological, and molecular factors. Three stromal features were significantly associated with survival, and this association was stronger than the association of survival with epithelial characteristics in the model. These findings implicate stromal morphologic structure as a previously unrecognized prognostic determinant for breast cancer.

INTRODUCTION

In the mid-19th century, it was first appreciated that the process of carcinogenesis produces characteristic morphologic changes in cancer cells (1). Paley and Scarff showed in 1928 (2) that three histologic features—tubule formation, epithelial nuclear atypia, and epithelial mitotic activity—could each be scored qualitatively, and the assessments could be combined to stratify breast cancer patients into three groups that showed significant survival differences. This semiquantitative morphological scoring scheme has been refined over the years (3–5) but still remains the standard technique for histologic grading in invasive breast cancer. Although considerable effort has been devoted recently to molecular profiling for assessment of prognosis and prediction of treatment response in cancer (6, 7), microscopic image assessment is still the most commonly available (and in some places in the world, the only) tool that is financially and logistically feasible.

Although the three epithelial features scored in current grading systems are useful in assessing cancer prognosis, valuable prognostic information can also be derived from other factors, including properties of the cancer stroma such as its molecular characteristics (8–15) and morphological features [such as stromal fibrotic focus, a scar-like area in the center of a carcinoma (16)]. Thus, we sought to develop a high-accuracy, image-based predictor to identify new clinically predictive morphologic phenotypes of breast cancers, thereby providing new insights into the biological factors driving breast cancer progression.

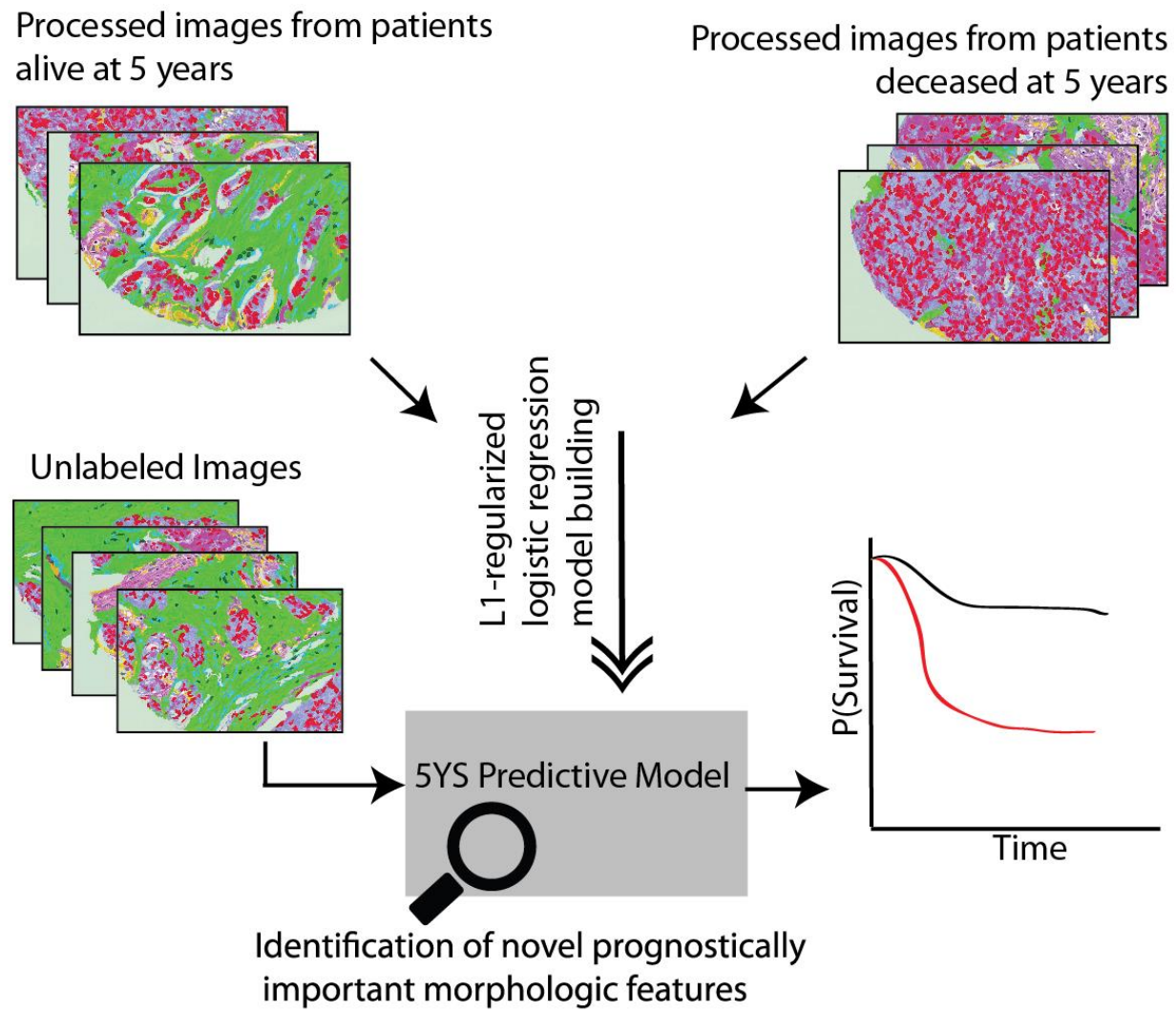
The development of such a system could also address other problems relevant to the clinical treatment of breast cancer. A limitation to the current grading system is that there is considerable variability in histologic grading among pathologists (17), with potentially negative consequences for determining treatment. An automated system could provide an objective method for predicting patient prognosis directly from image data. Moreover, once established, this system could be used in breast cancer clinical trials to provide an accurate, objective means for assessing breast cancer morphologic characteristics, allowing objective stratification of breast cancer patients on the basis of morphologic criteria and facilitating the discovery of morphologic features associated with response to specific therapeutic agents.

RESULTS

Experimental design overview

We developed the Computational Pathologist (C-Path), a machine learning–based method for automatically analyzing cancer images and predicting prognosis. To construct and evaluate the model, we acquired hematoxylin and eosin (H&E)–stained histological images from breast cancer tissue microarrays (TMAs) (figs. S4 and S5). The

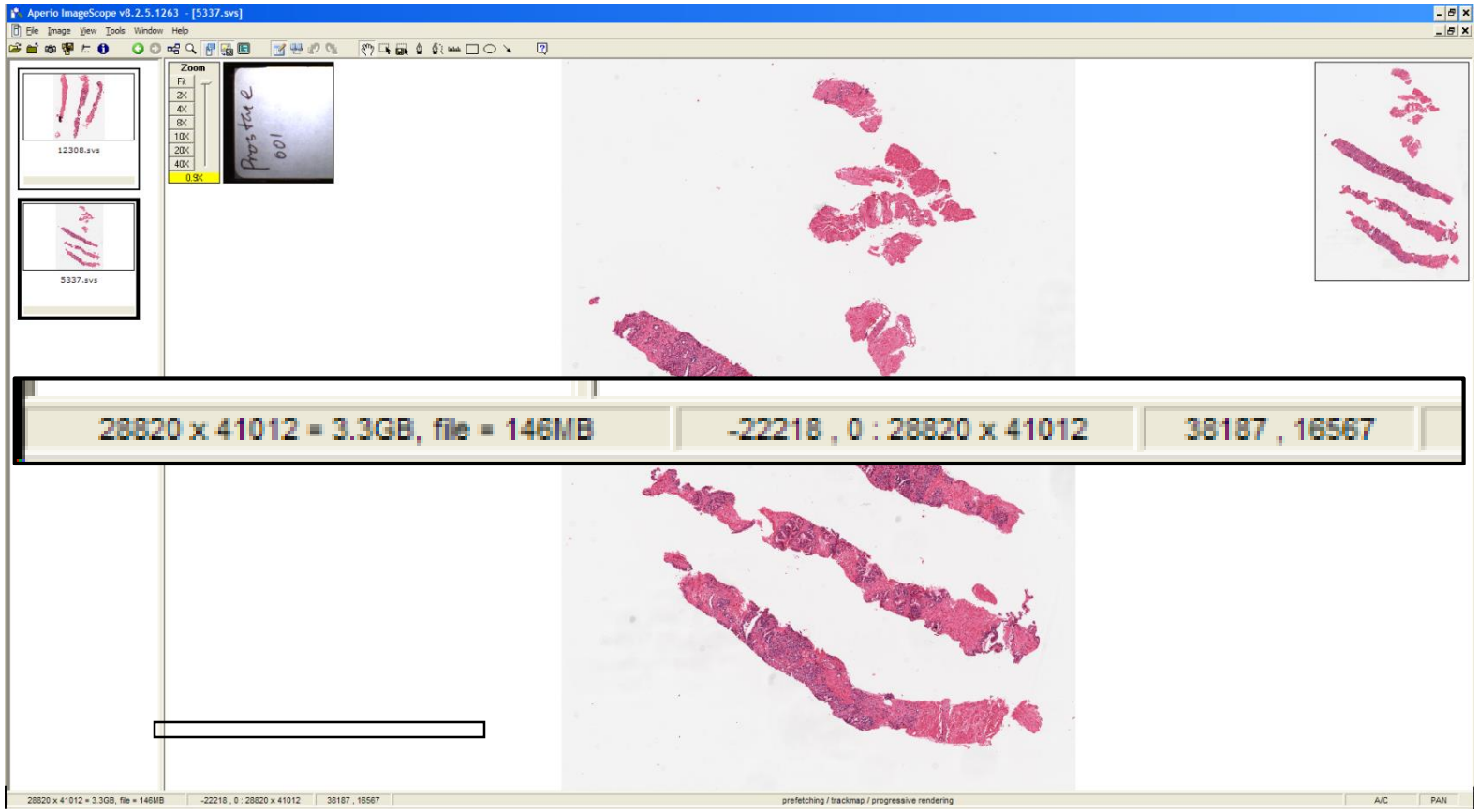
[1]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Biomedical Informatics Program, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA. [4]Genetic Pathology Evaluation Center, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [5]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [6]Department of Pathology, Academic Medical Center, Amsterdam 1105 AZ, Netherlands. [7]Department of Computer Science, Stanford University, Stanford, CA 94305, USA.
*To whom correspondence should be addressed. E-mail: andrew.beck@bidmc.harvard.edu



Beck, A.H., at el., 2011. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science Translational Medicine* 3, 108ra113.

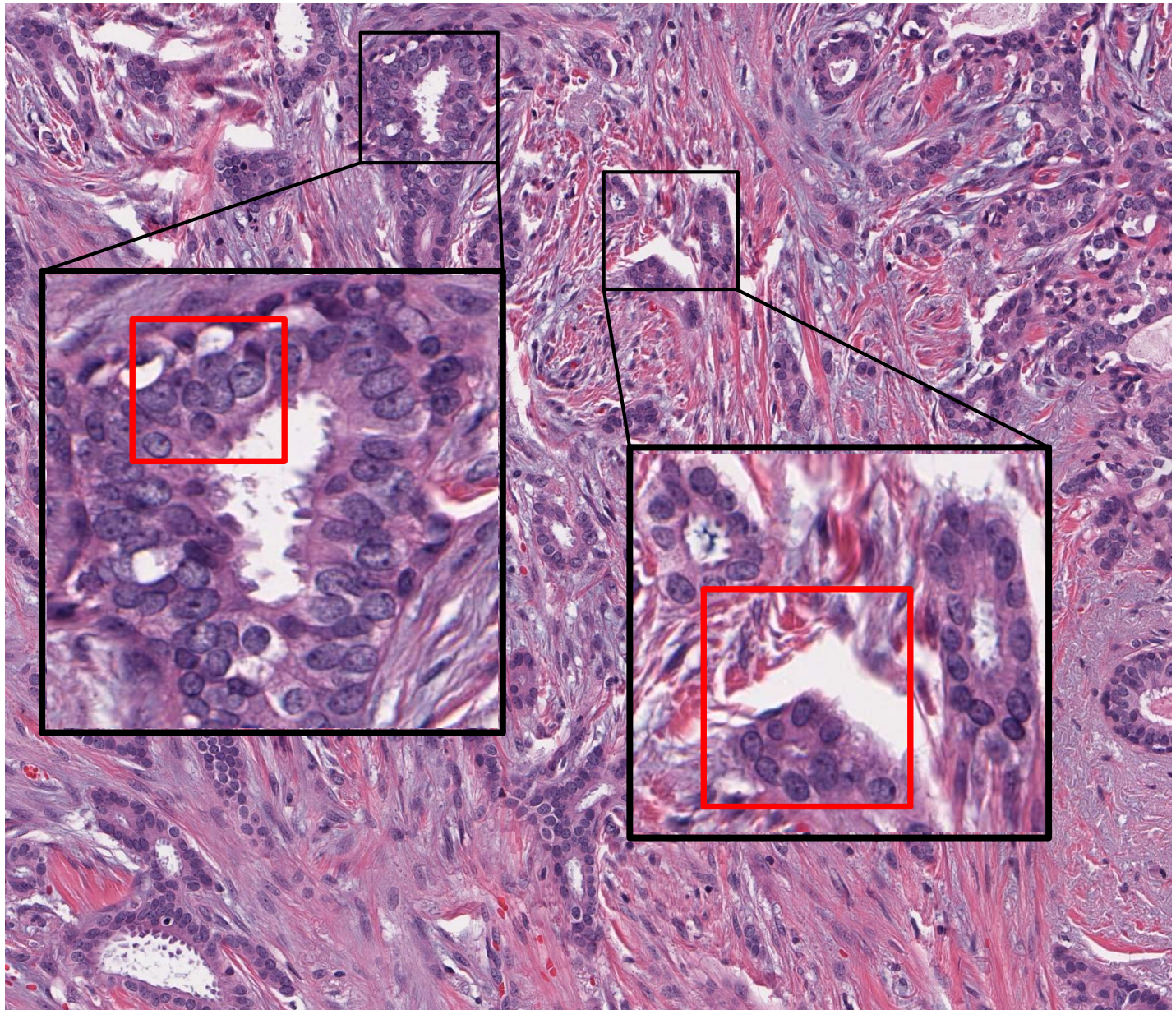
Courtesy to Dr. Andrew Beck

组织病理图像分析：面临的挑战



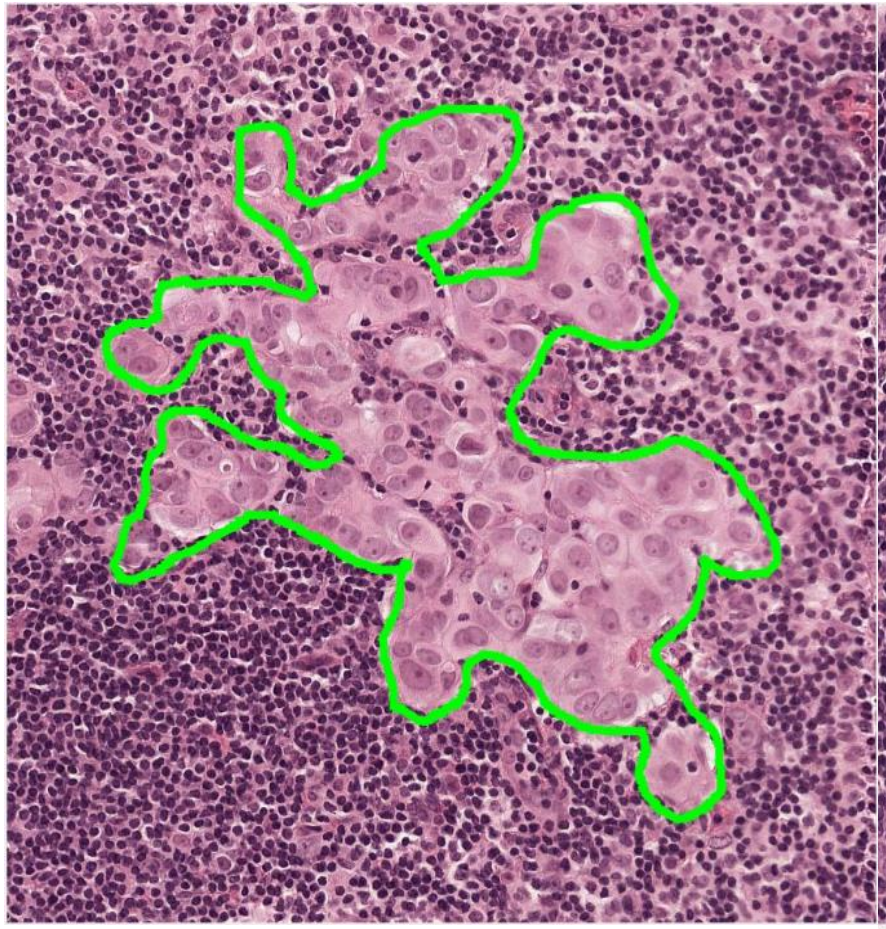
- 图像尺寸大，该全扫描图像压缩前的尺寸为3.3 G
- 28820 x 41012 = 12亿个像素(扫描分辨率0.2(40X)-0.5(20X)微米/像素，~3分钟)
- 需要用智能的方法才能检测图像中的病变区域

组织病理图像分析：面临的挑战

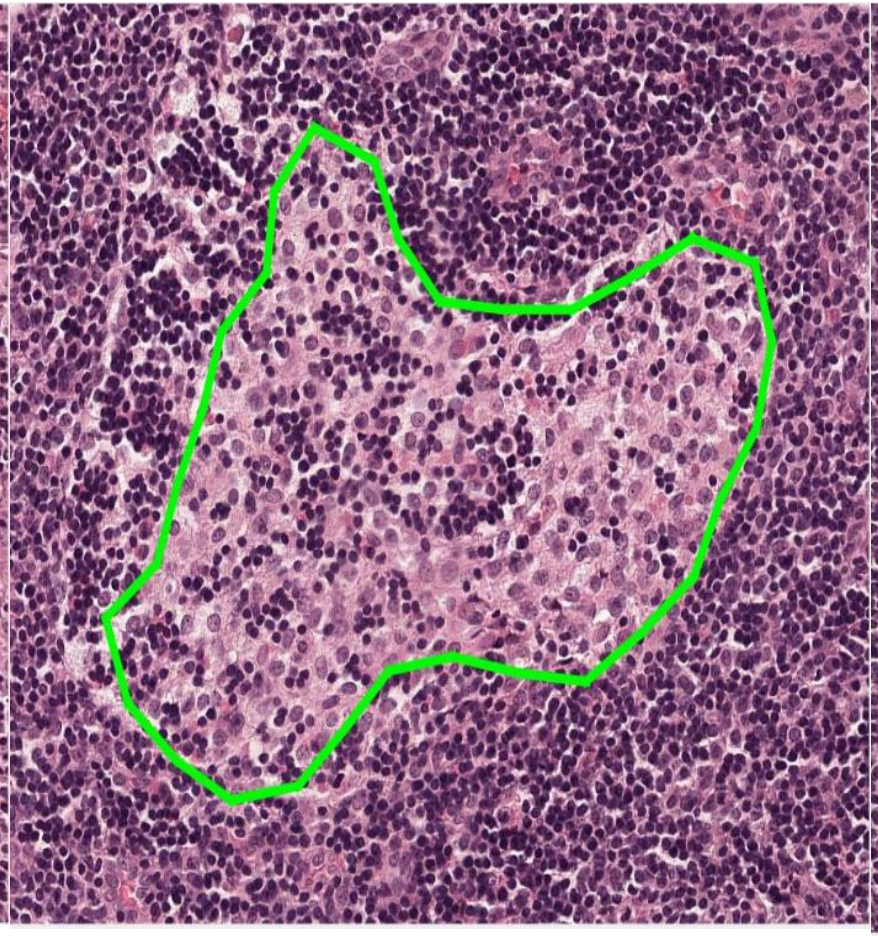


- 高分辨率全扫描图像的处理问题
- 大量的细胞和不同的组织结构交织
- 同步地检测与分割成千上万个细胞和不同的组织结构
- 细胞紧密地聚积
- 形状不规则、边界模糊、重叠
- 染色错误
- 噪声干扰

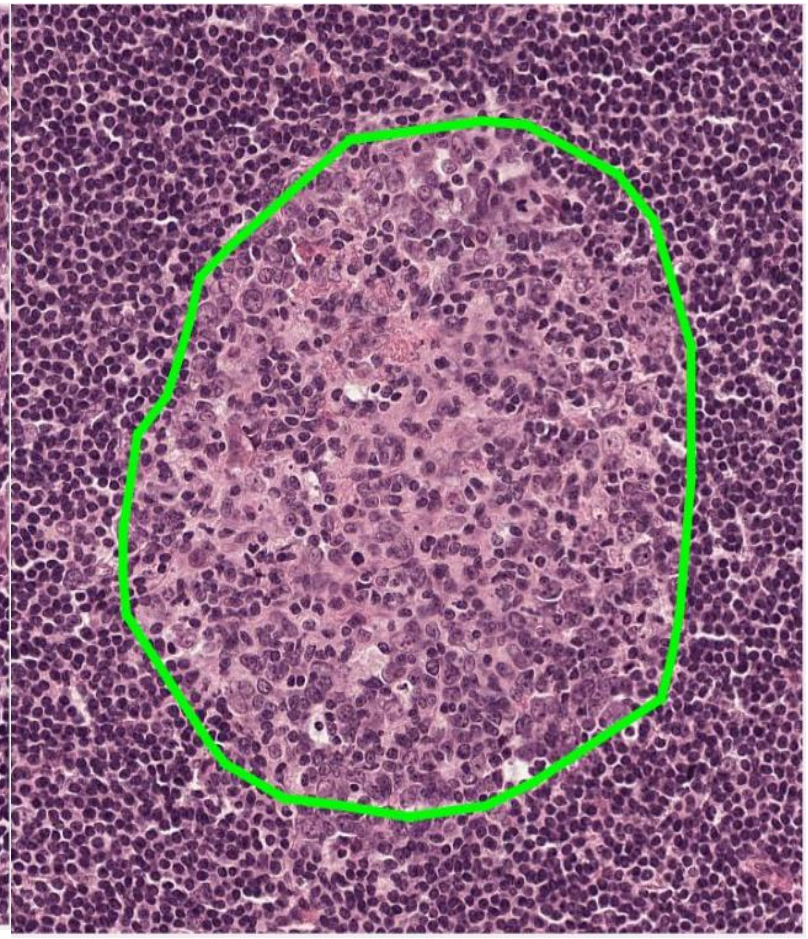
组织病理图像分析：面临的挑战



癌转移区域

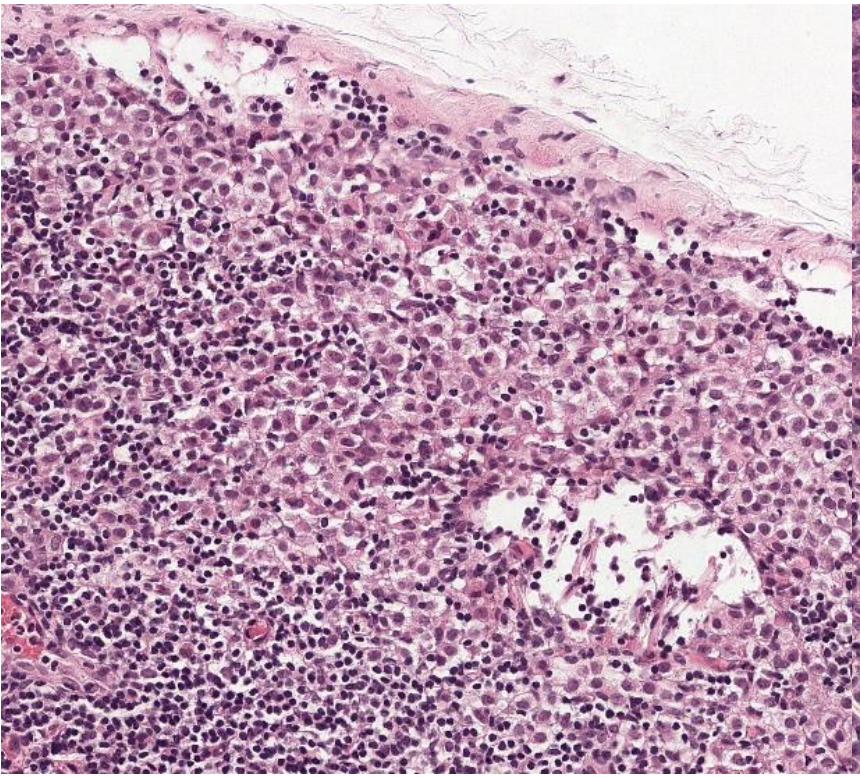


淋巴窦组织

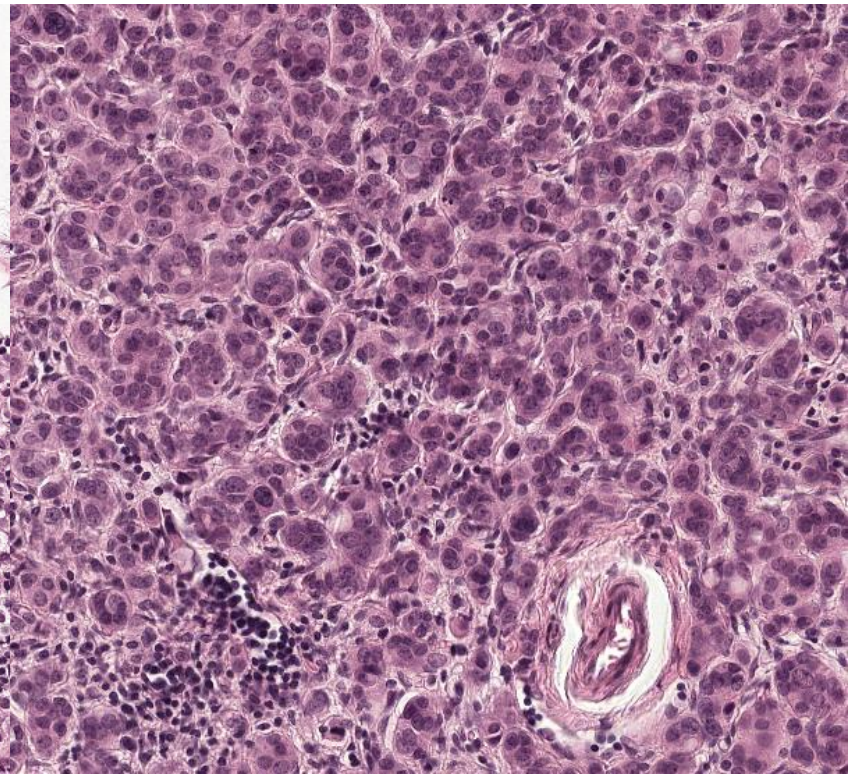


次级滤泡生发中心

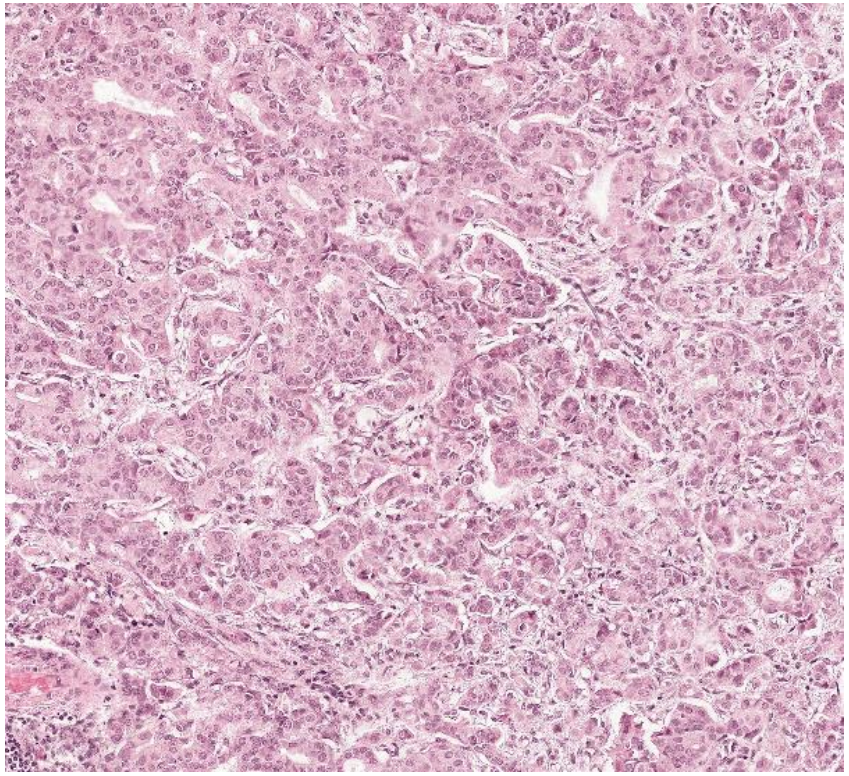
组织病理图像分析：面临的挑战



淋巴细胞和癌细胞混合



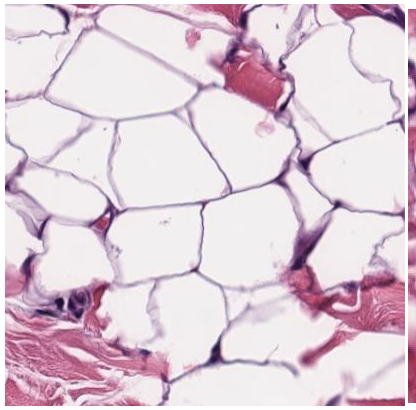
细胞的重叠和紧靠



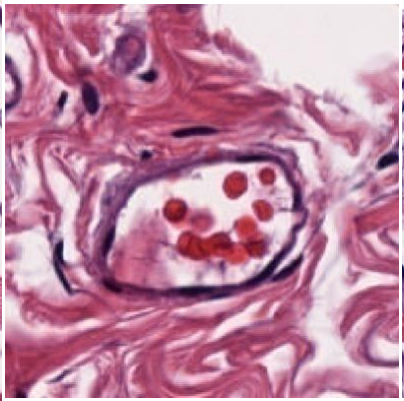
癌细胞形状和大小异质性高

组织病理图像分析：面临的挑战

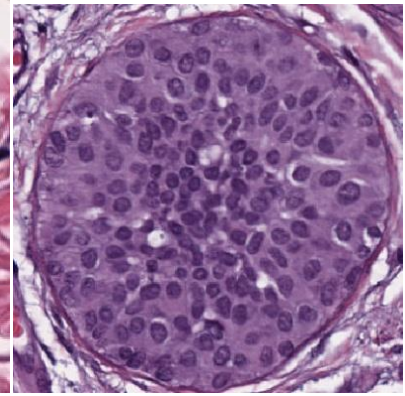
Area	Explanation	Tissue sample
Blood vessel	Either a small artery or vein. Arteries commonly have thicker walls than veins. Blood vessels are important infiltration routes for tumors.	
Duct	Lactiferous ducts are tubular milk transport structures in breasts. In histopathology slices lactiferous ducts look like holes or cylinders depending on the cut direction.	
Ductal carcinoma in situ (DCIS)	Ductal carcinoma in situ stands for the most common noninvasive breast cancer that originates from a lactiferous duct. DCIS is a preliminary stage of cancer where cells already seem malignant and contain genetic changes but the cells are still contained inside the duct and have not invaded the stroma.	
Fat	Fat looks like empty drops because the fat dissolves in the slice preparation process before the cutting and dyeing of paraffin blocks.	
Inflammatory cells	Inflammatory tissue contains lymphocytes, neutrophils and eosinophils. Lymphocytes look like dark, little and round cells and their dark nucleus fills almost the whole cell body.	
Invasive ductal carcinoma	Invasive ductal carcinoma is a malignant breast cancer which originates from a lactiferous duct and invades the stroma.	
Lobule	Lobule (terminal duct-lobular unit (benign)) is a unit in the end of lactiferous duct from which milk is secreted. Lobule consists of little glands, which form a round structure, and of the small distal part of the duct.	
Lymph vessel	Lymph vessels are part of the lymphatic system where lymph passes through lymph nodes and returns to bloodstream. Lymph vessels have a thin wall. Lymph vessels are important infiltration routes for tumors.	
Nerve	Nerves are important structures in tumor diagnostics because benign changes do not normally grow near the nerves. Invasion to a nerve surrounding traditionally means that the tumor is malign, even though there are exceptions.	
Red Blood cells	Red blood cells have a biconcave disc shape. They are red cells that do not have nuclei. Red blood cells have a diameter of 5 micrometers and they are usually found in the lumen of blood vessels.	
Stroma	Stroma consists of connective tissue surrounding and supporting biological tissues, cells and organs. Whereas parenchyma refers to the functional parts of the tissue (e.g. the actual mutant cells).	



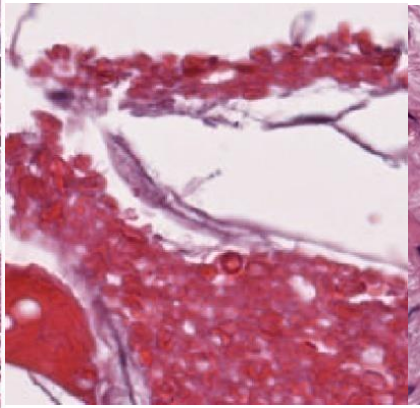
脂肪



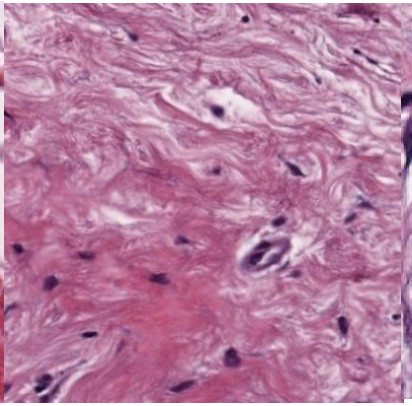
血管



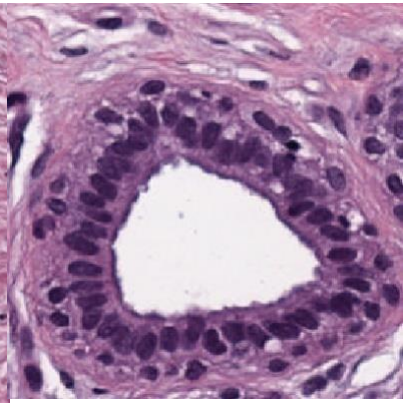
导管原位癌



红细胞



基质

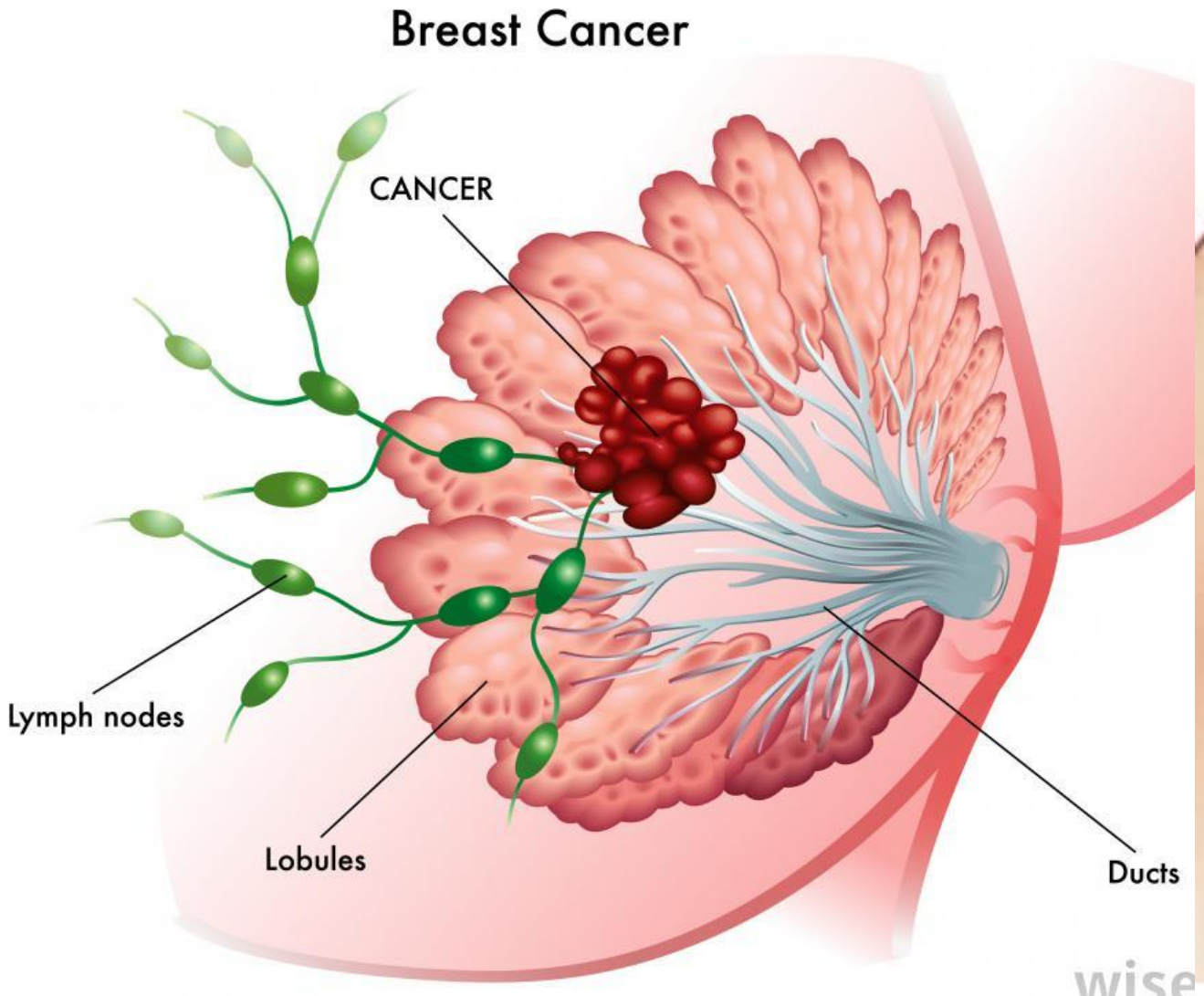


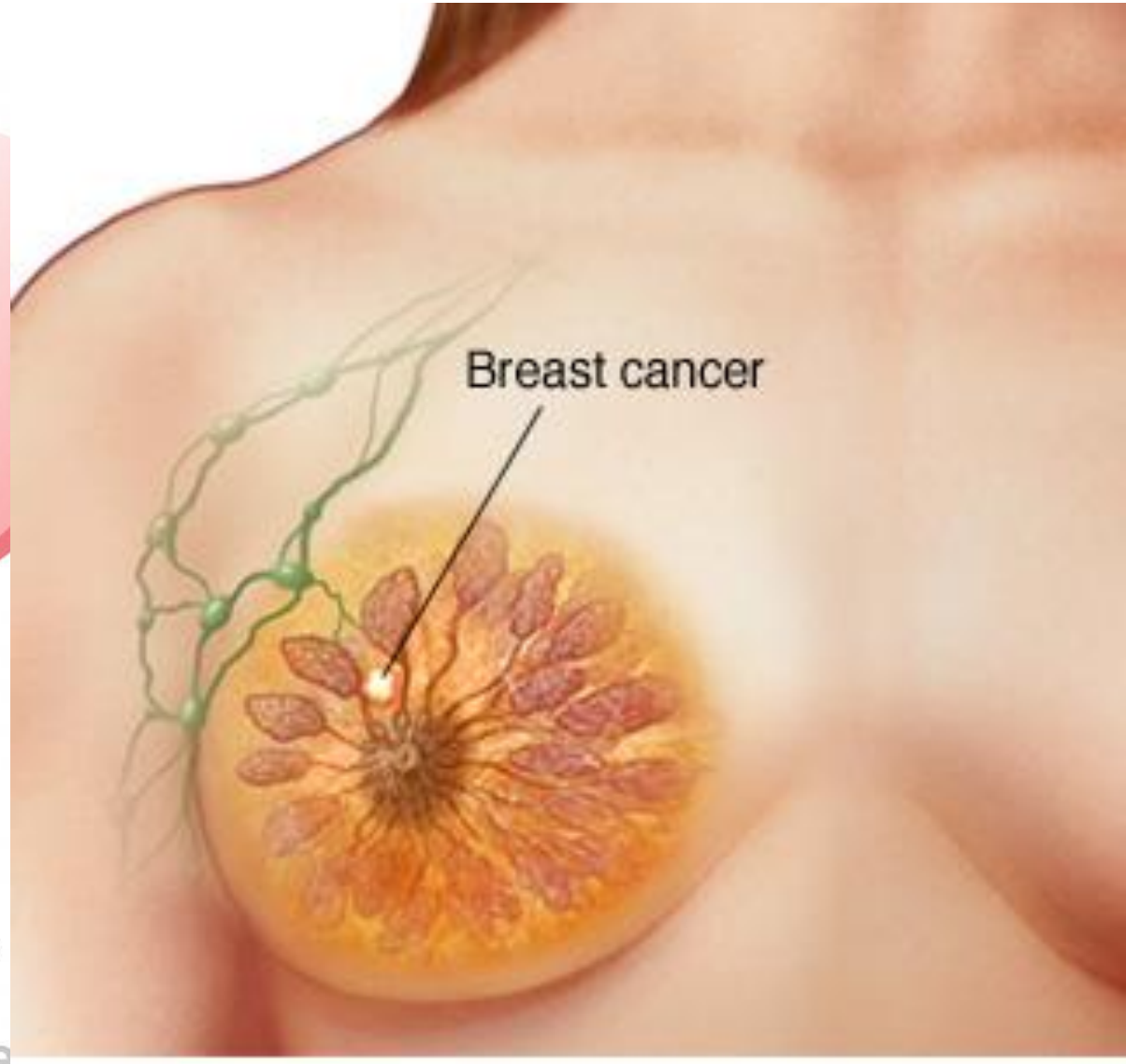
导管

提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

基于图像分析的乳腺癌诊断和预后





乳腺癌临床诊断与预后流程

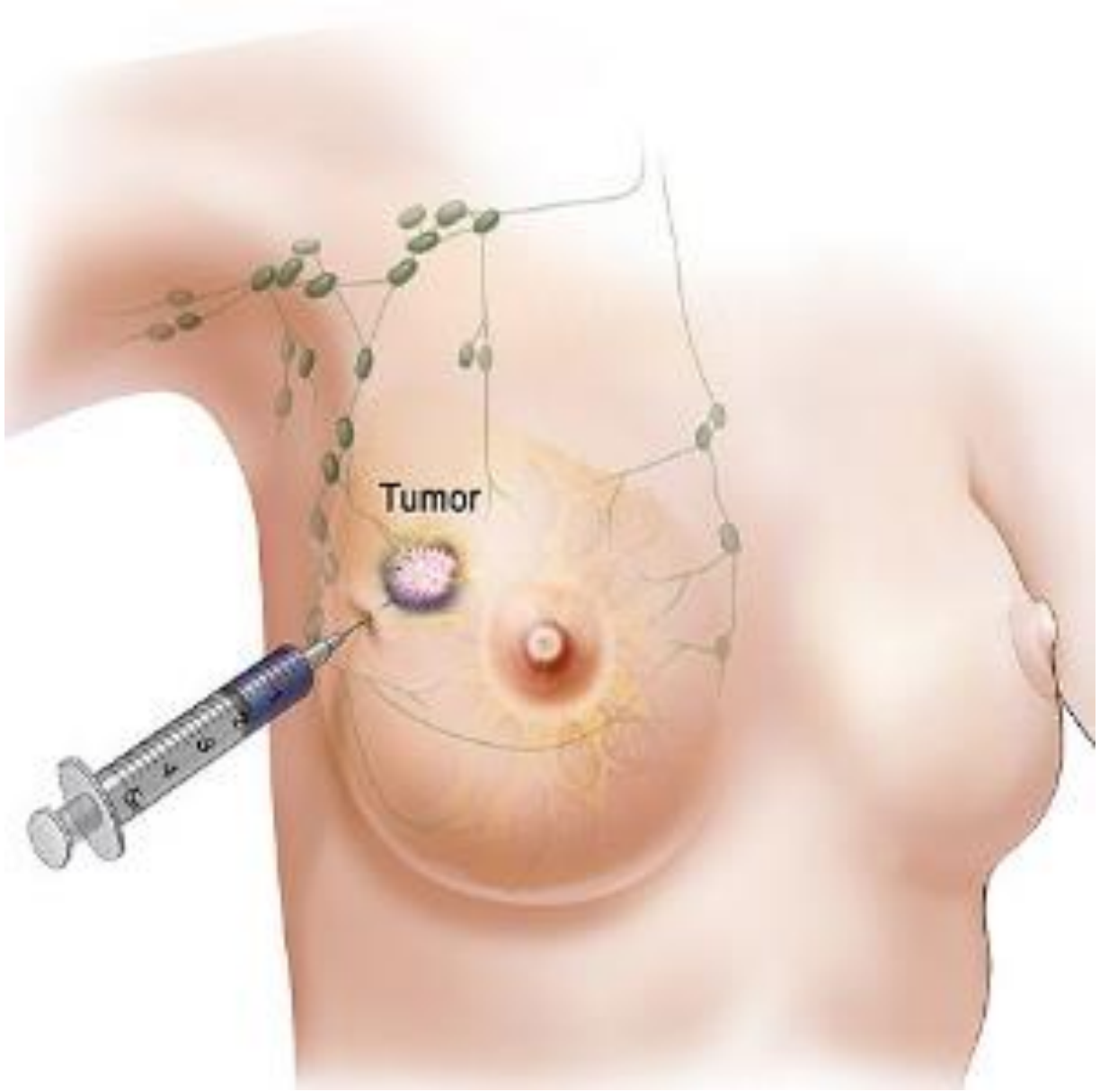
体检/筛查

乳腺癌临床诊断与预后流程

体检/筛查

↓

穿刺/手术



乳腺癌临床诊断与预后流程

体检/筛查

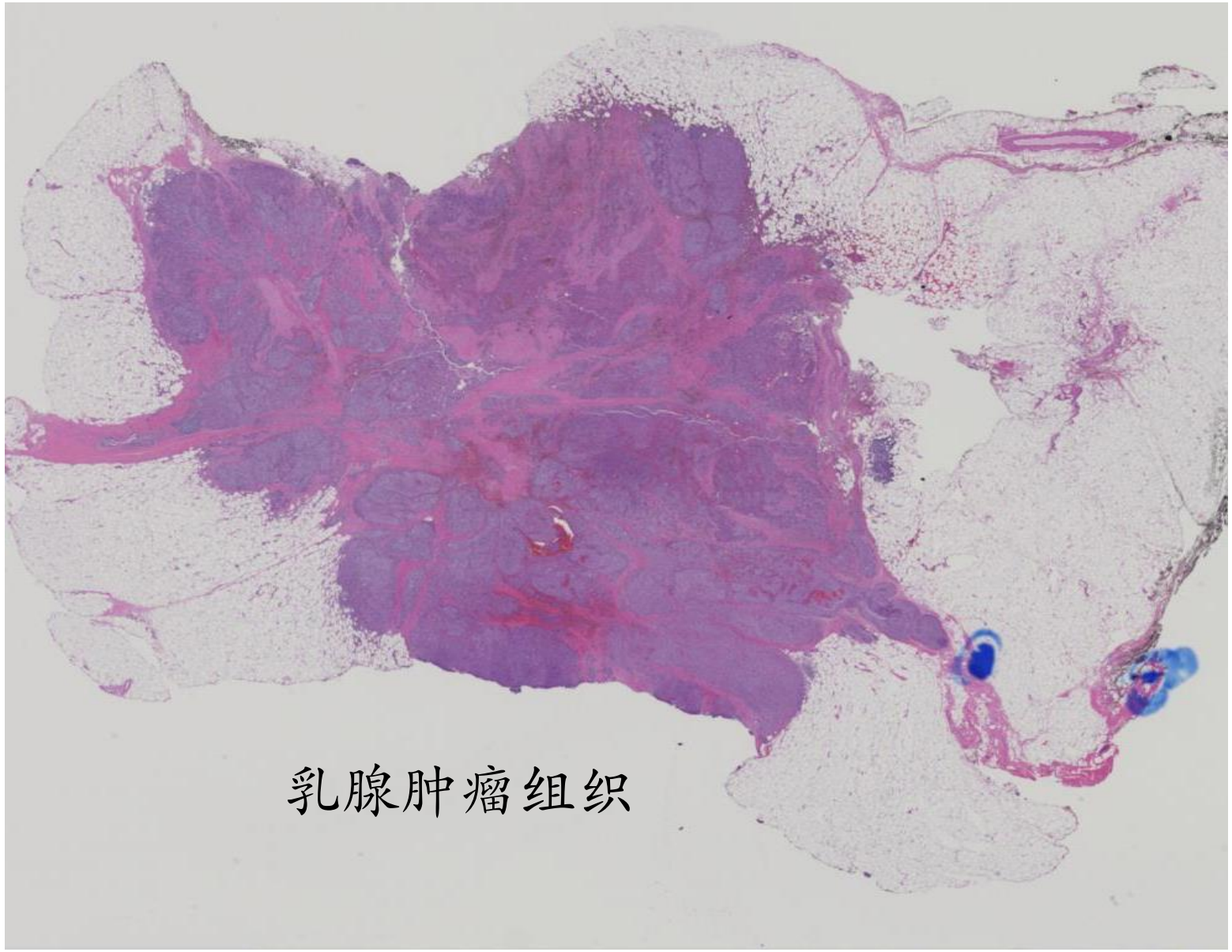
↓

穿刺/手术

↓

切片制作/染色

↓



乳腺肿瘤组织

乳腺癌临床诊断与预后流程

体检/筛查

↓

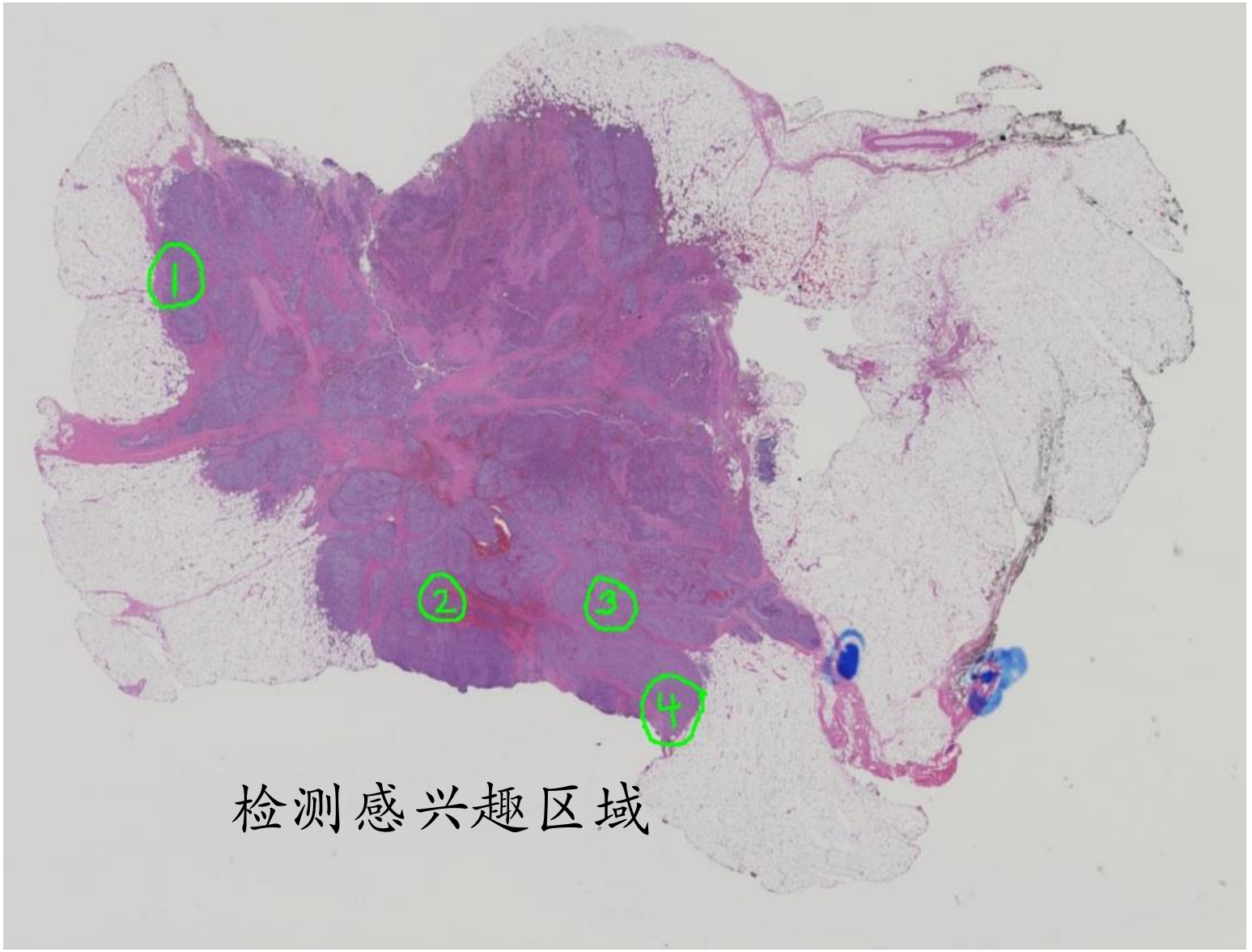
穿刺/手术

↓

切片制作/染色

↓

人工分析



检测感兴趣区域

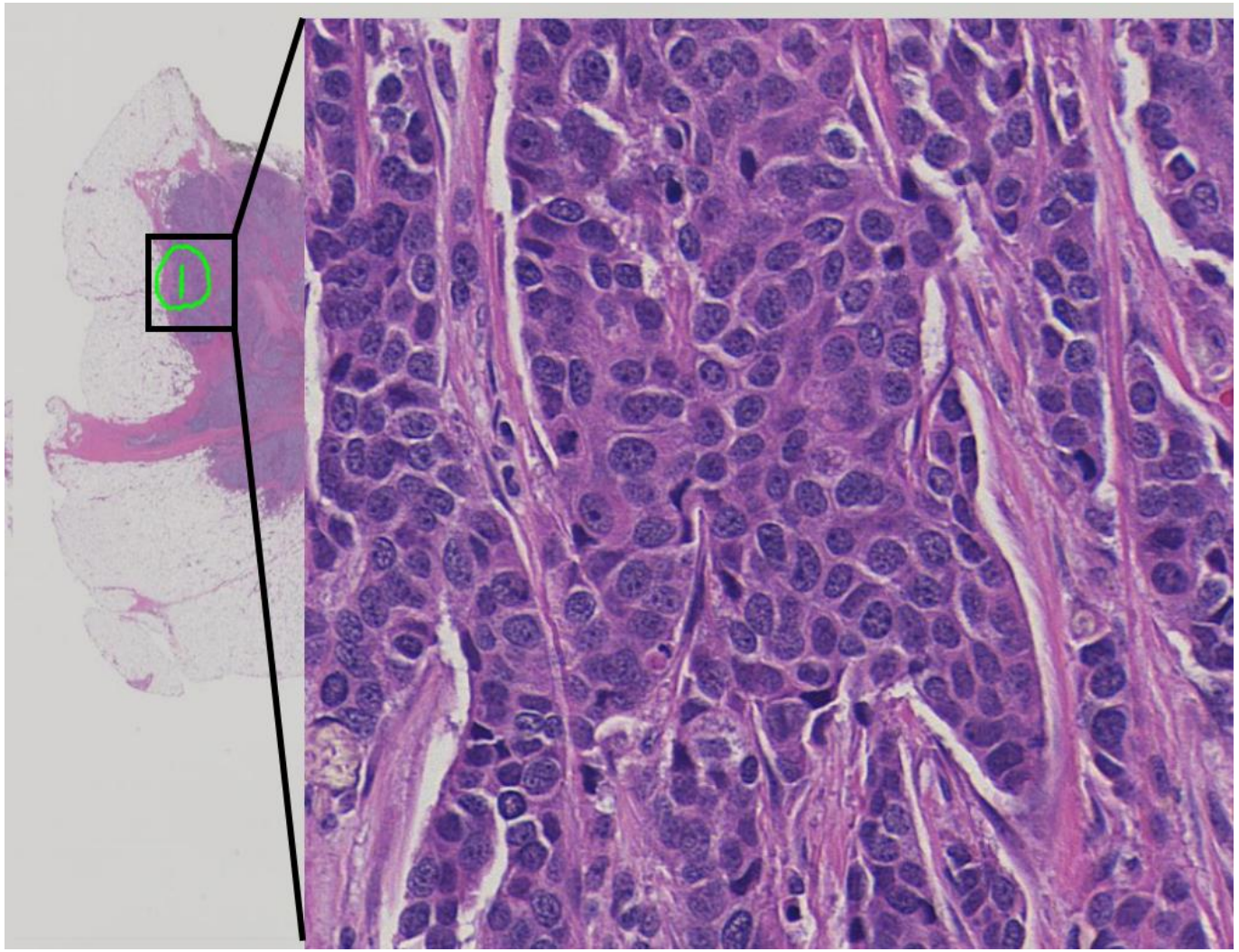
乳腺癌临床诊断与预后流程

体检/筛查

穿刺/手术

切片制作/染色

人工分析



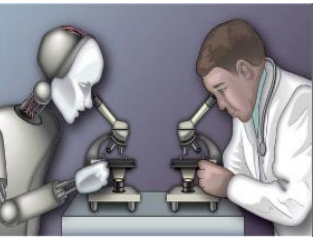
乳腺癌临床诊断与预后流程

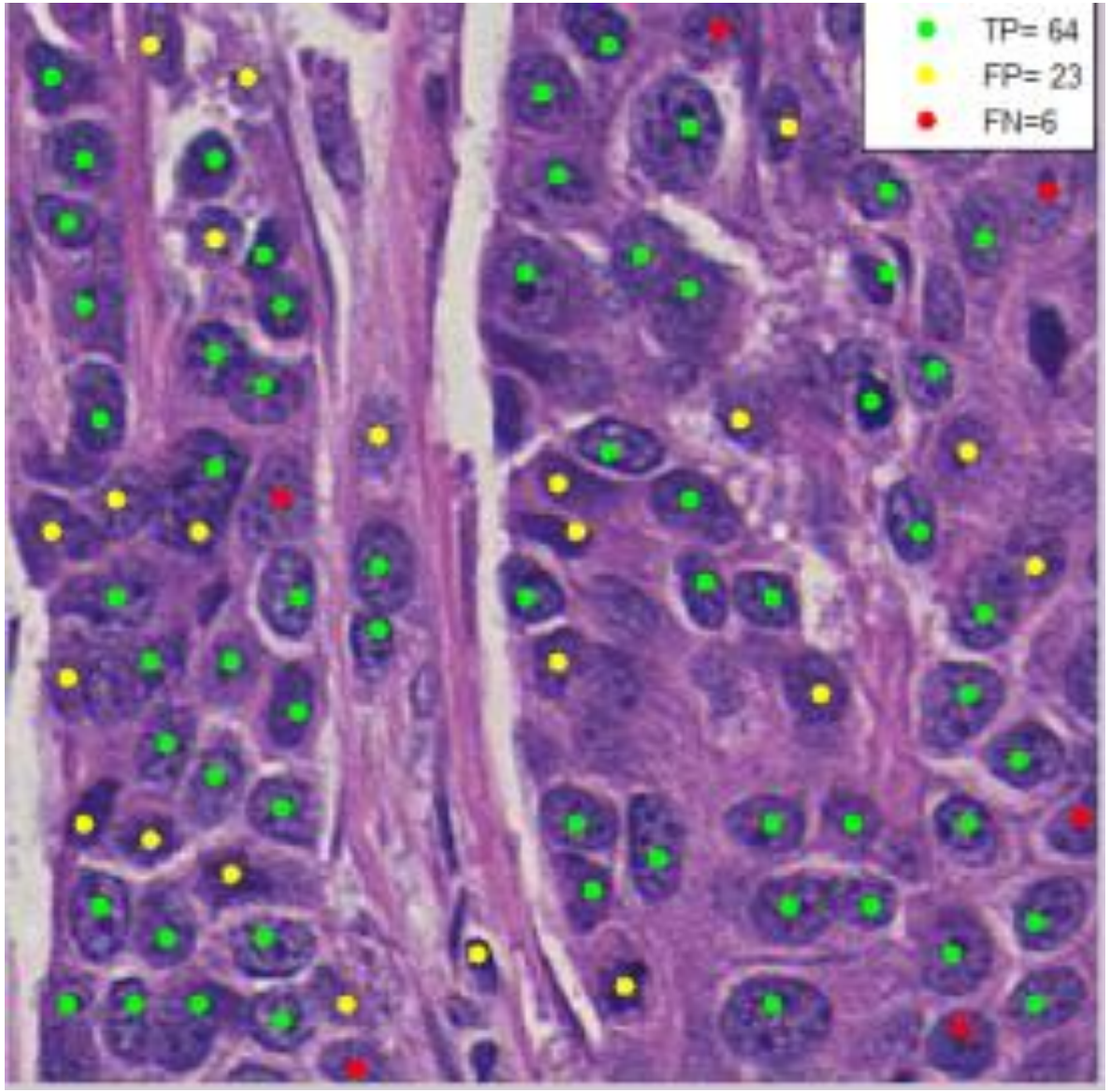
体检/筛查

穿刺/手术

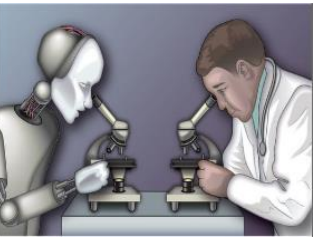
切片制作/染色

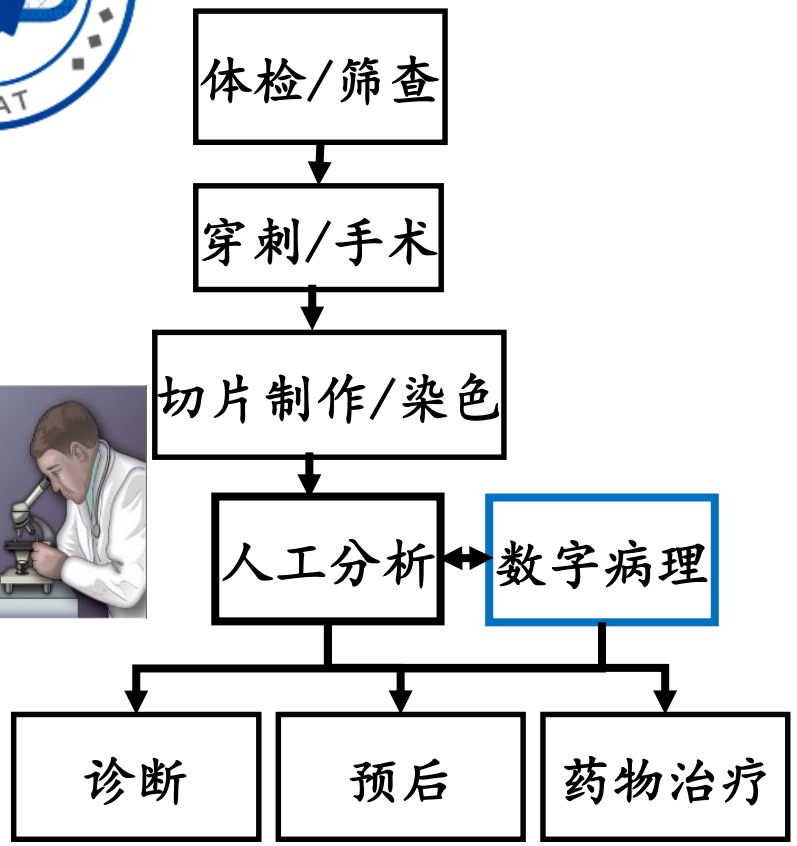
人工分析 ↔ 数字病理



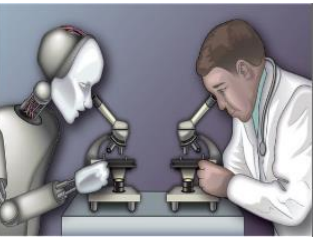


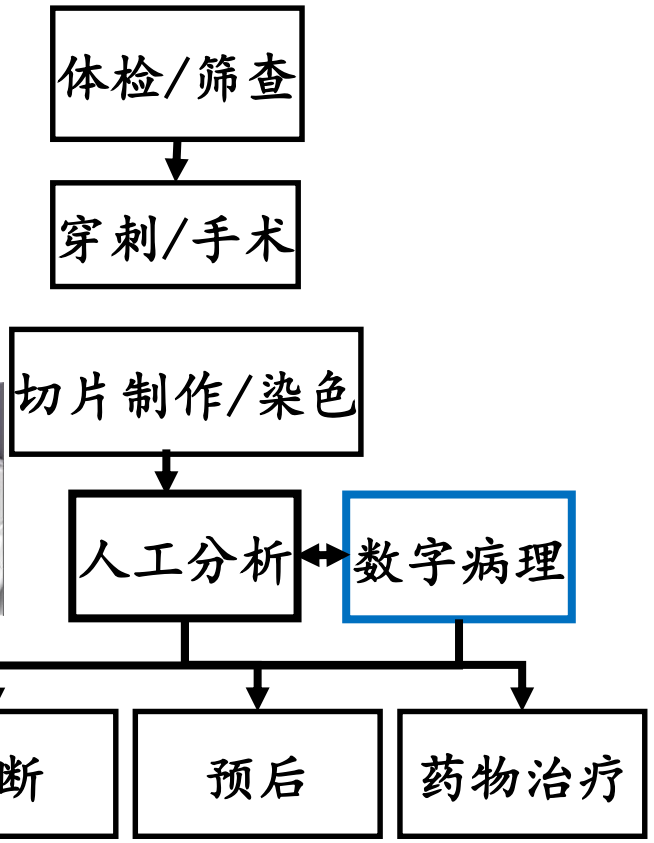
乳腺癌临床诊断与预后流程

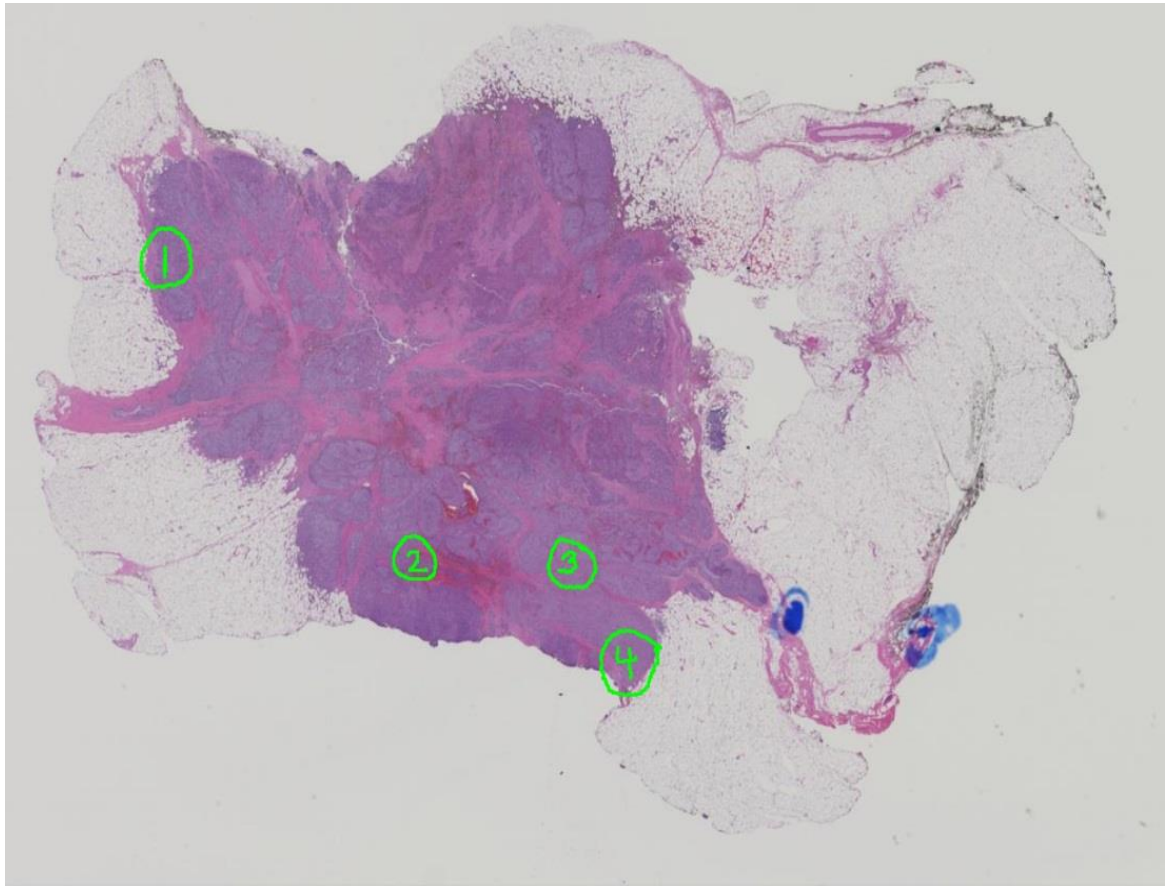




乳腺癌临床诊断与预后流程



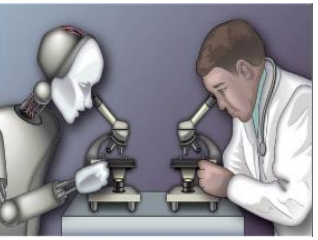


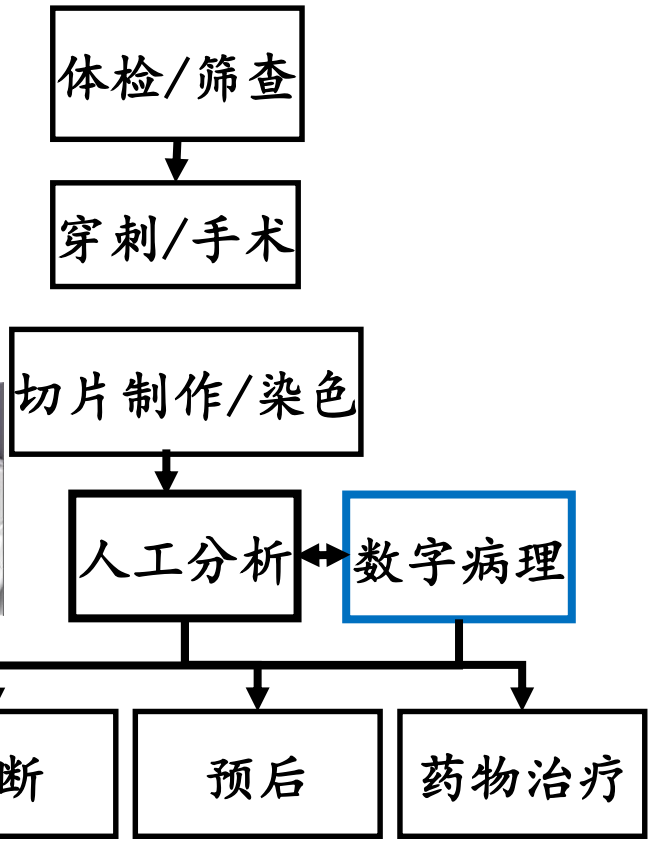


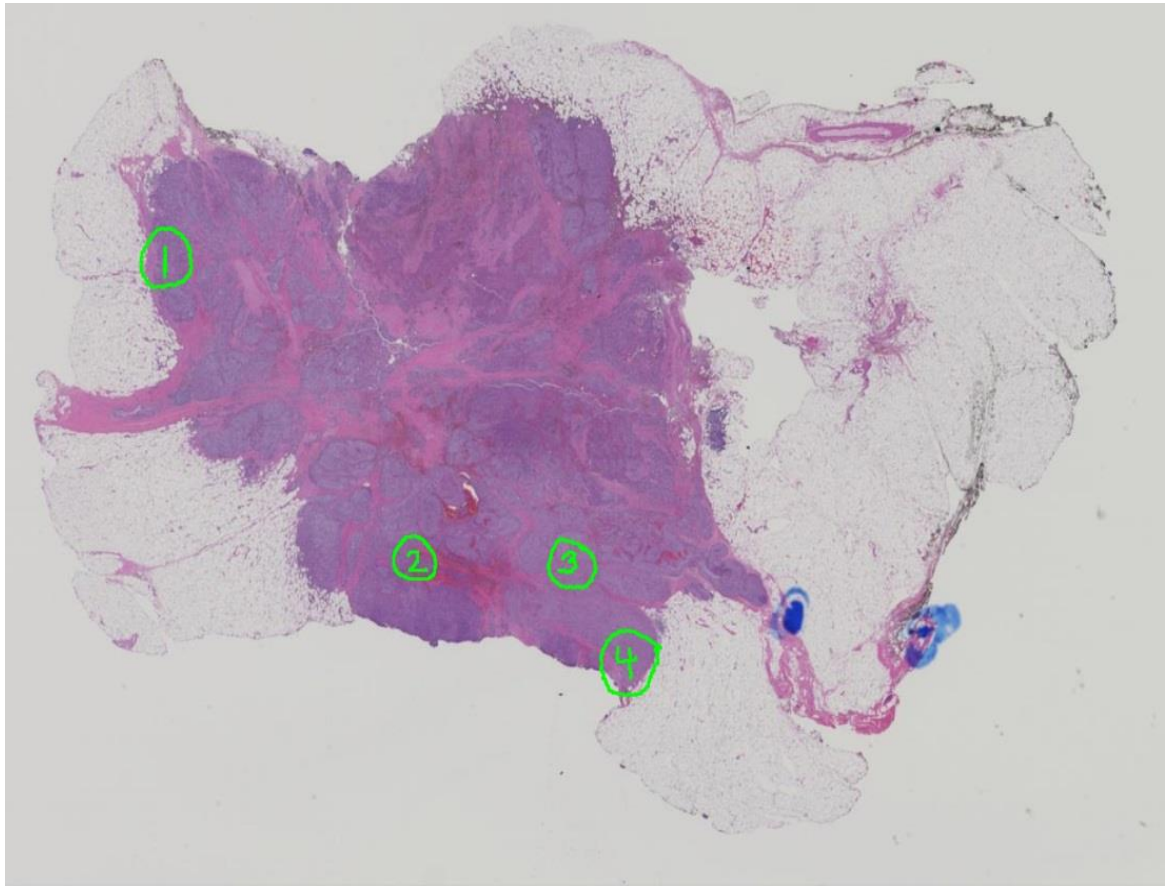
人工分析的缺陷：

- **主观性** - 分析的结论会随专家/单位的不同而不同
- **耗时** - 图像尺寸大，而且组织类型多（大多为良性/正常）
- **定量性**-无法使用定量的图像特征和度量方法
- **一致性** - 不同专家之间的一致性非常得低

乳腺癌临床诊断与预后流程



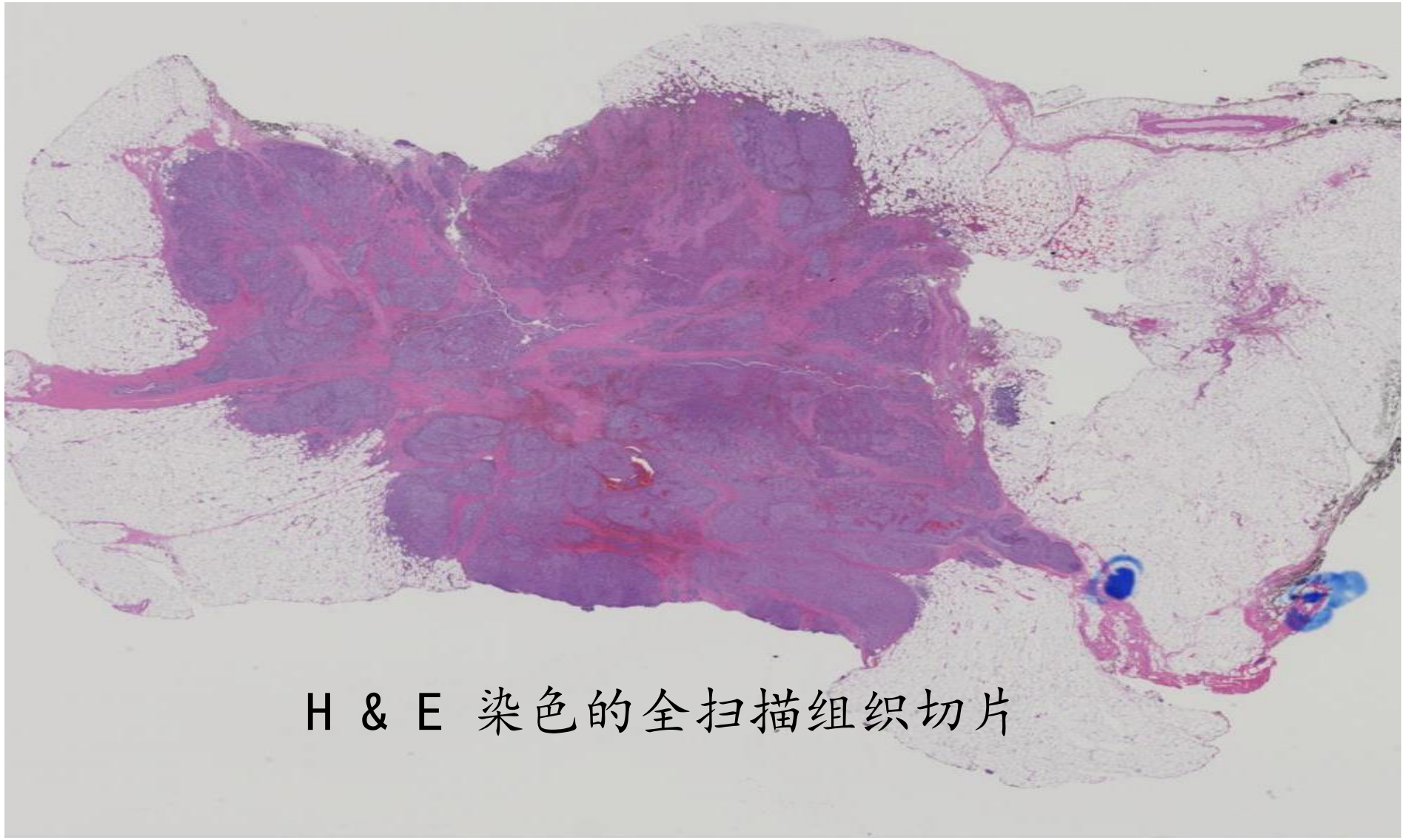




数字病理的优势:

- 能自动地分析、分割、检测感兴趣区域
- 能够定量地评估病变区域的恶性程度
- 结果具有可重复性

乳腺癌临床诊断与预后流程

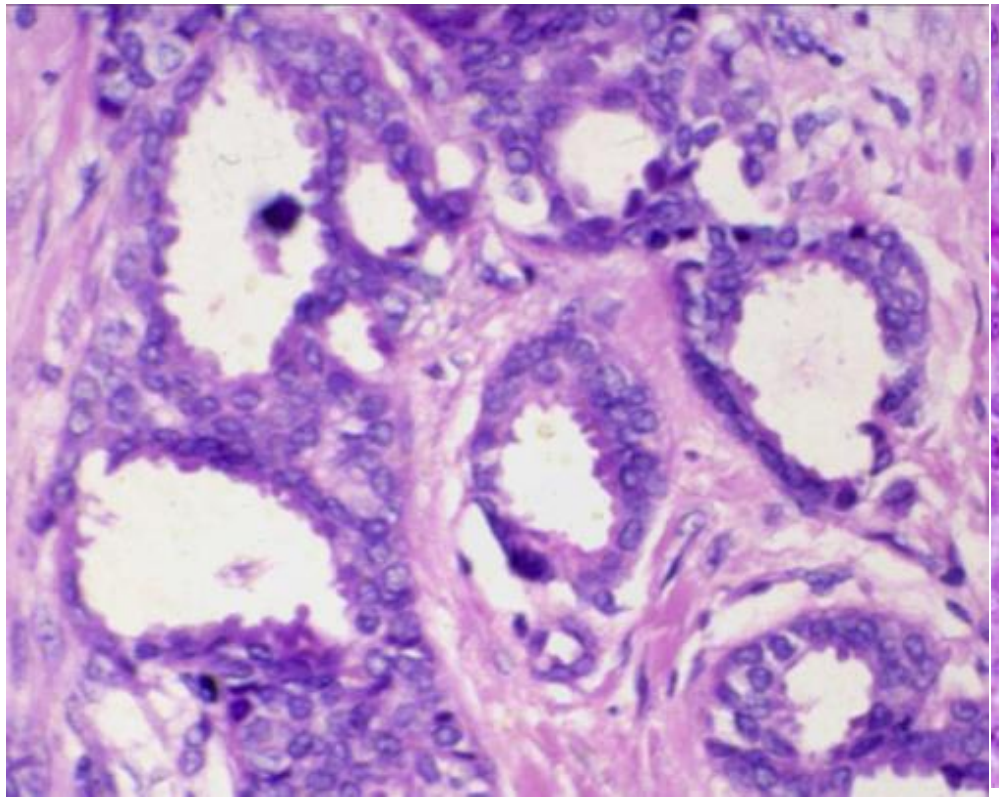


H & E 染色的全扫描组织切片

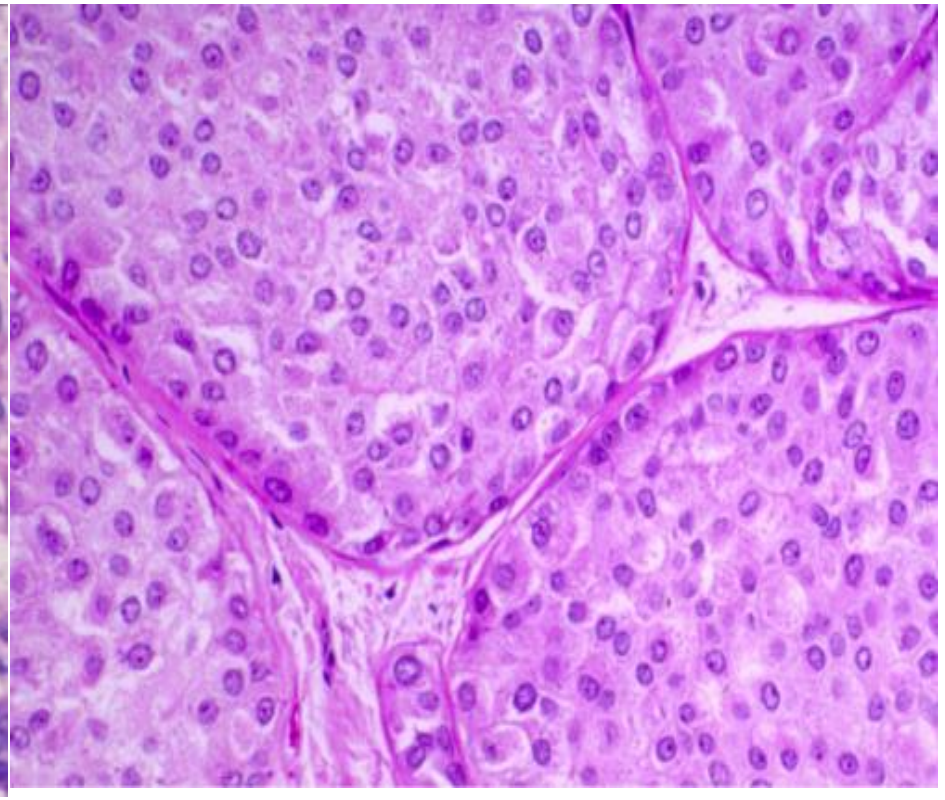
诺丁汉分级系统(NGS)

Elston, C. W. Ellis, I. O., “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, 1991.

乳腺癌临床诊断与预后流程



腺体结构规则(正常)



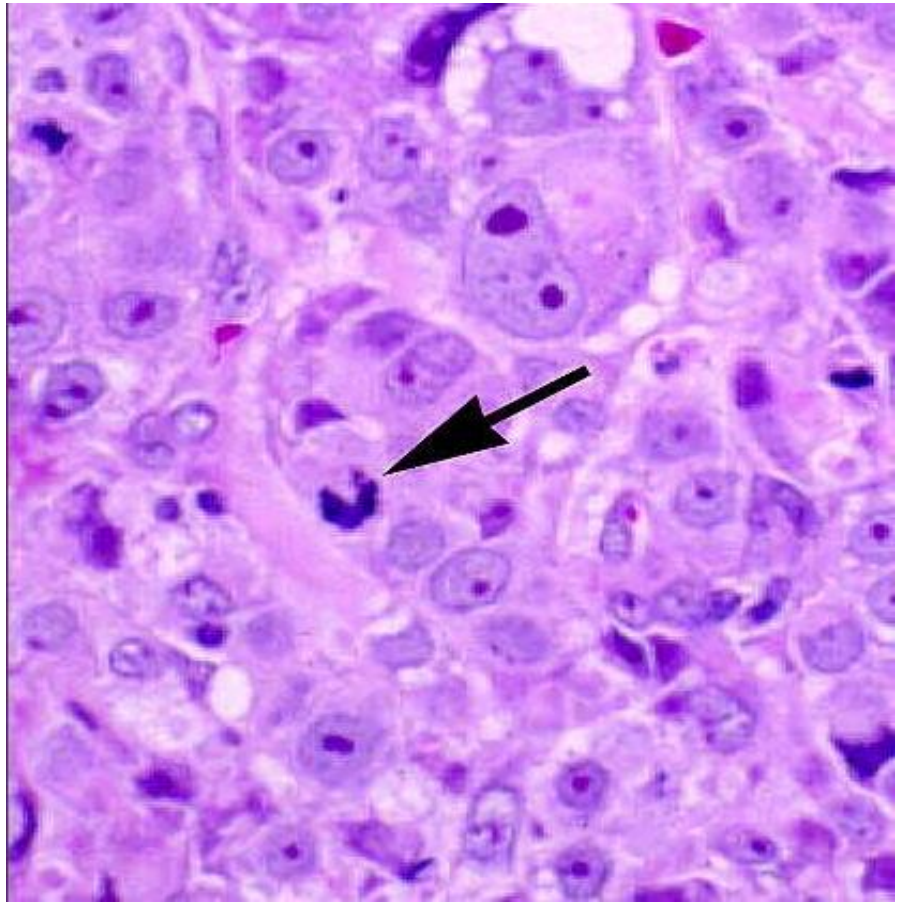
无规则的腺体结构(恶性)

1.腺体的形成程度

诺丁汉分级系统(NGS)

Elston, C. W. Ellis, I. O., "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, 1991.

乳腺癌临床诊断与预后流程

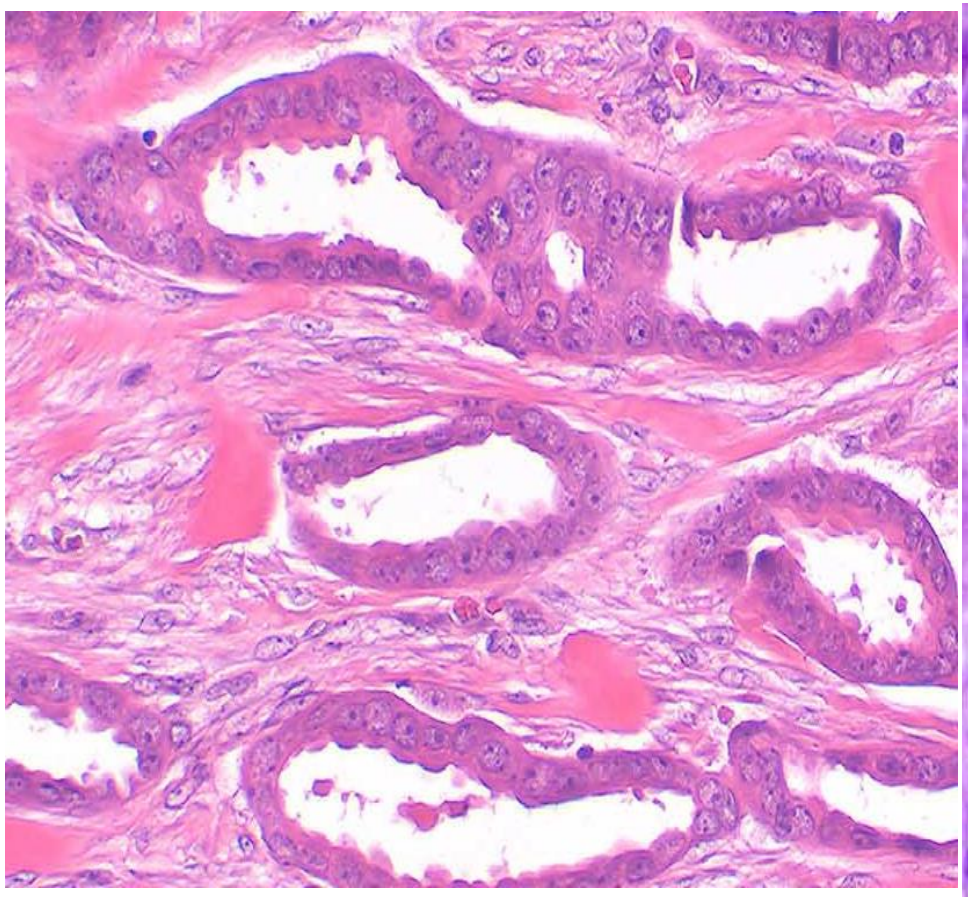


诺丁汉分级系统(NGS)

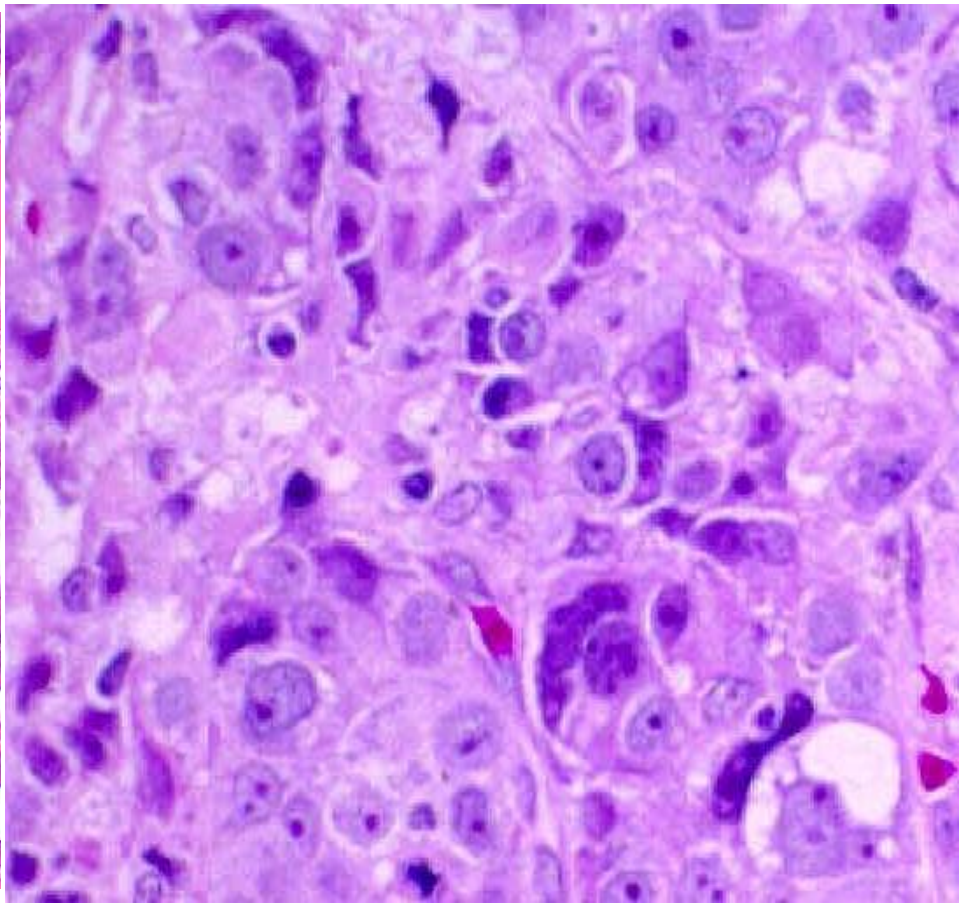
Elston, C. W. Ellis, I. O., "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up", Histopathology, 1991.

1. 腺体的形成程度
2. 肿瘤有丝分裂次数（细胞分裂速度）

乳腺癌临床诊断与预后流程



细胞的异形性小（正常）



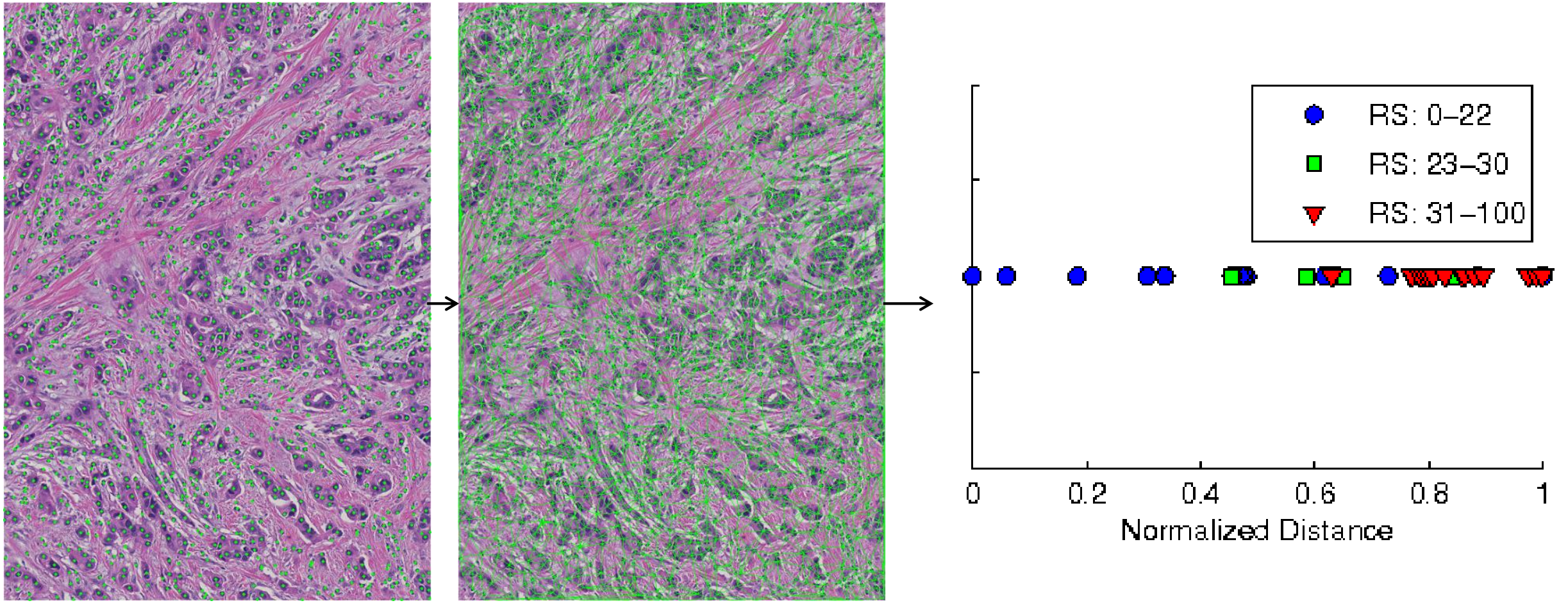
细胞的异形性高（恶性）

诺丁汉分级系统(NGS)

Elston, C. W. Ellis, I. O., “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, 1991.

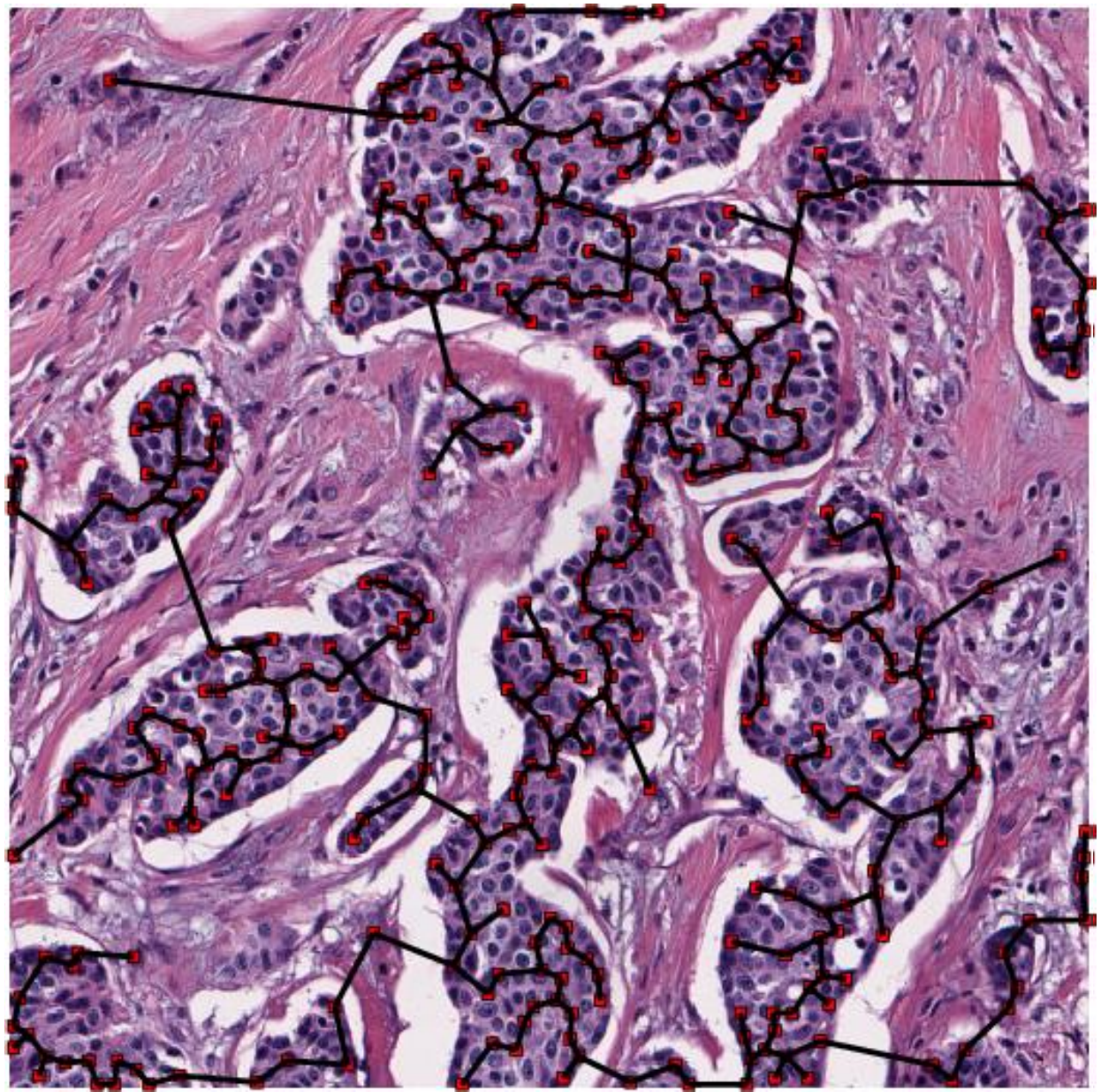
1. 腺体的形成程度
2. 肿瘤有丝分裂次数（细胞分裂速度）
3. 细胞的异质性(“多形性” 或者肿瘤细胞的“丑陋”程度)

雌激素阳性乳腺癌计算机辅助预后系统



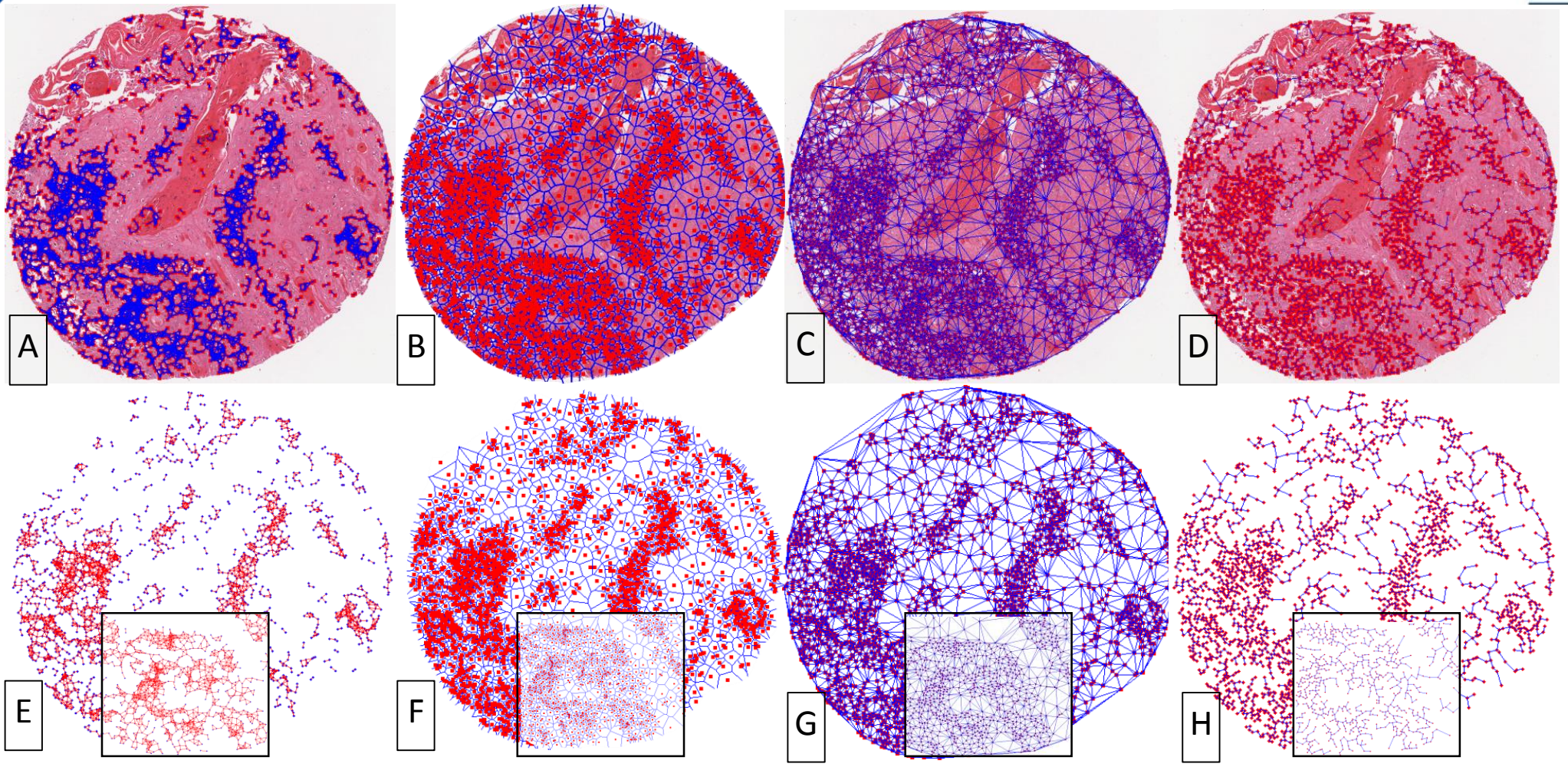
Basavanhally, A, **Xu J**, Ganesan S, Madabhushi A. 2009. Computer-aided Prognosis (CAP) of ER+ Breast Cancer Histopathology and Correlating Survival Outcome with Oncotype Dx Assay. *IEEE International Symposium on Biomedical Imaging. :851-854.*

基于细胞核的图特征：拓扑结构特征



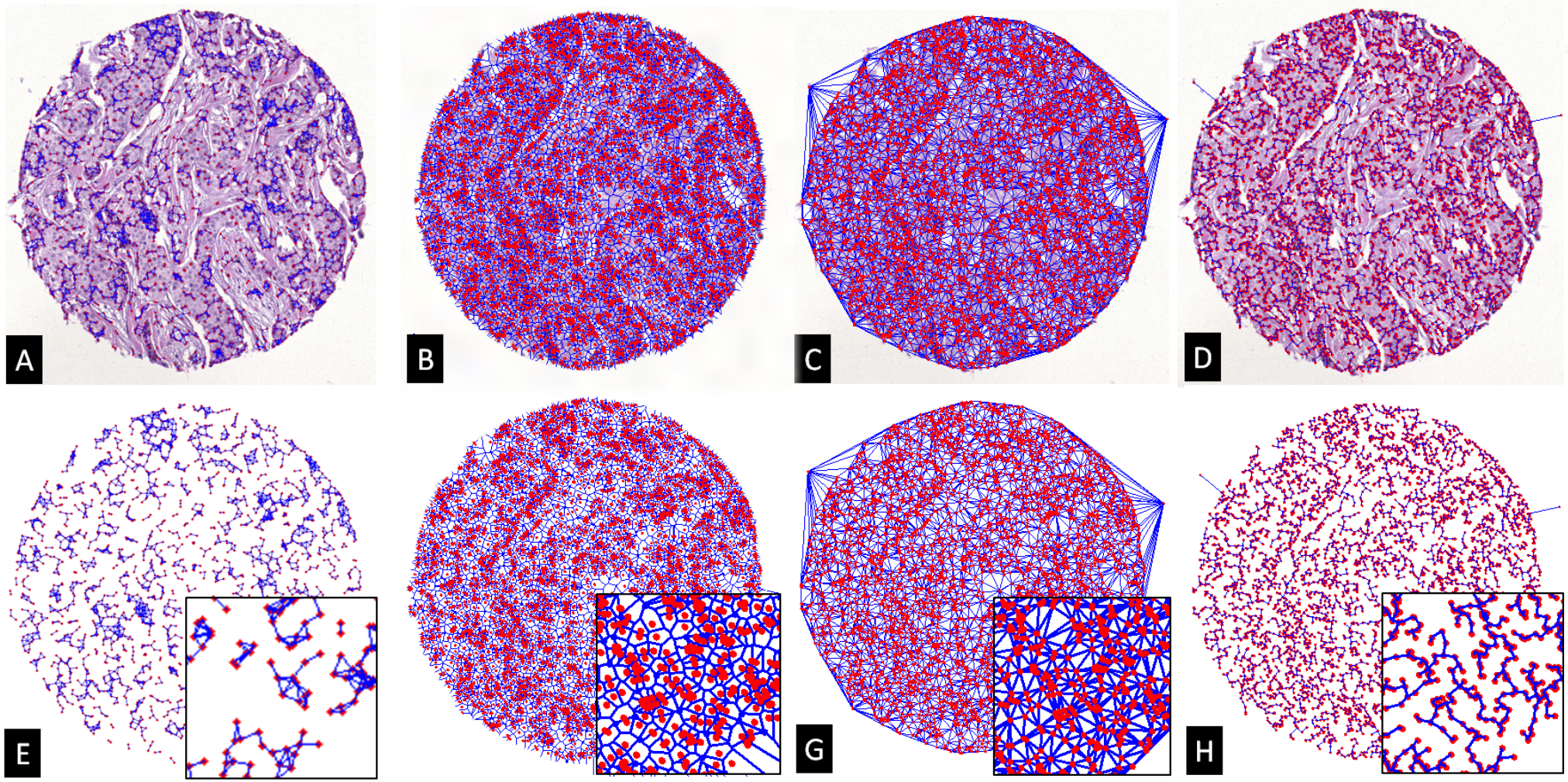
- Voronoi Diagram
  - Polygon area
  - Polygon perimeter
  - Polygon chord length
- Delaunay Triangulation
  - Triangle side length
  - Triangle area
- Minimum Spanning Tree
  - Edge length

Nuclear-graph: topological features (Oral cancer TMAs)



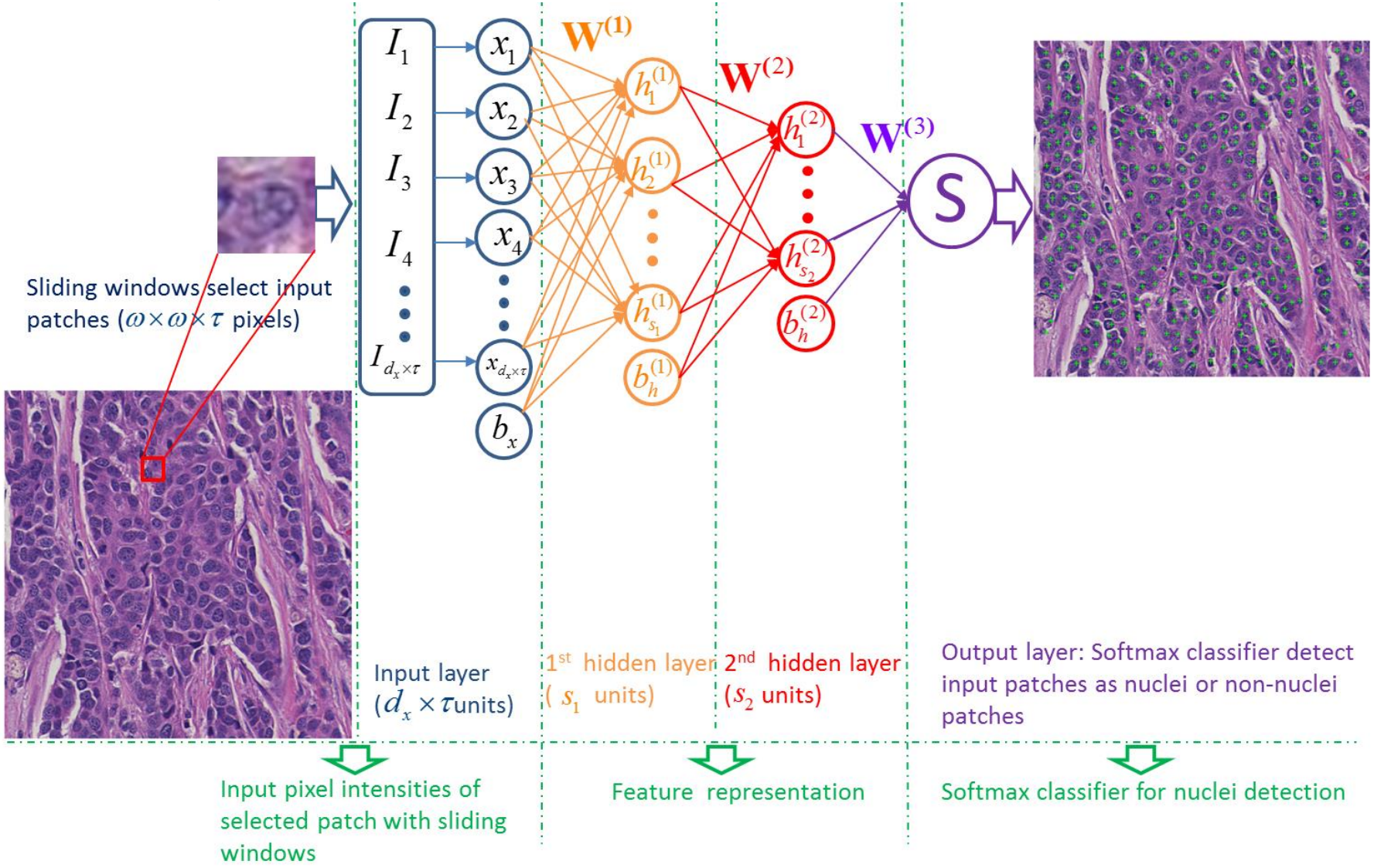
Cruz-Roa A, **Xu J**, Madabhushi A, "A note on the stability and discriminability of graph based features for classification problems in digital pathology", *Proc. SPIE 9287, 10th International Symposium on Medical Information Processing and Analysis, 2015.*

Nuclear-graph: topological features (Breast TMAs)



Cruz-Roa A, **Xu J**, Madabhushi A, "A note on the stability and discriminability of graph based features for classification problems in digital pathology", *Proc. SPIE 9287, 10th International Symposium on Medical Information Processing and Analysis, 2015.*

基于稀疏自编码器的细胞检测

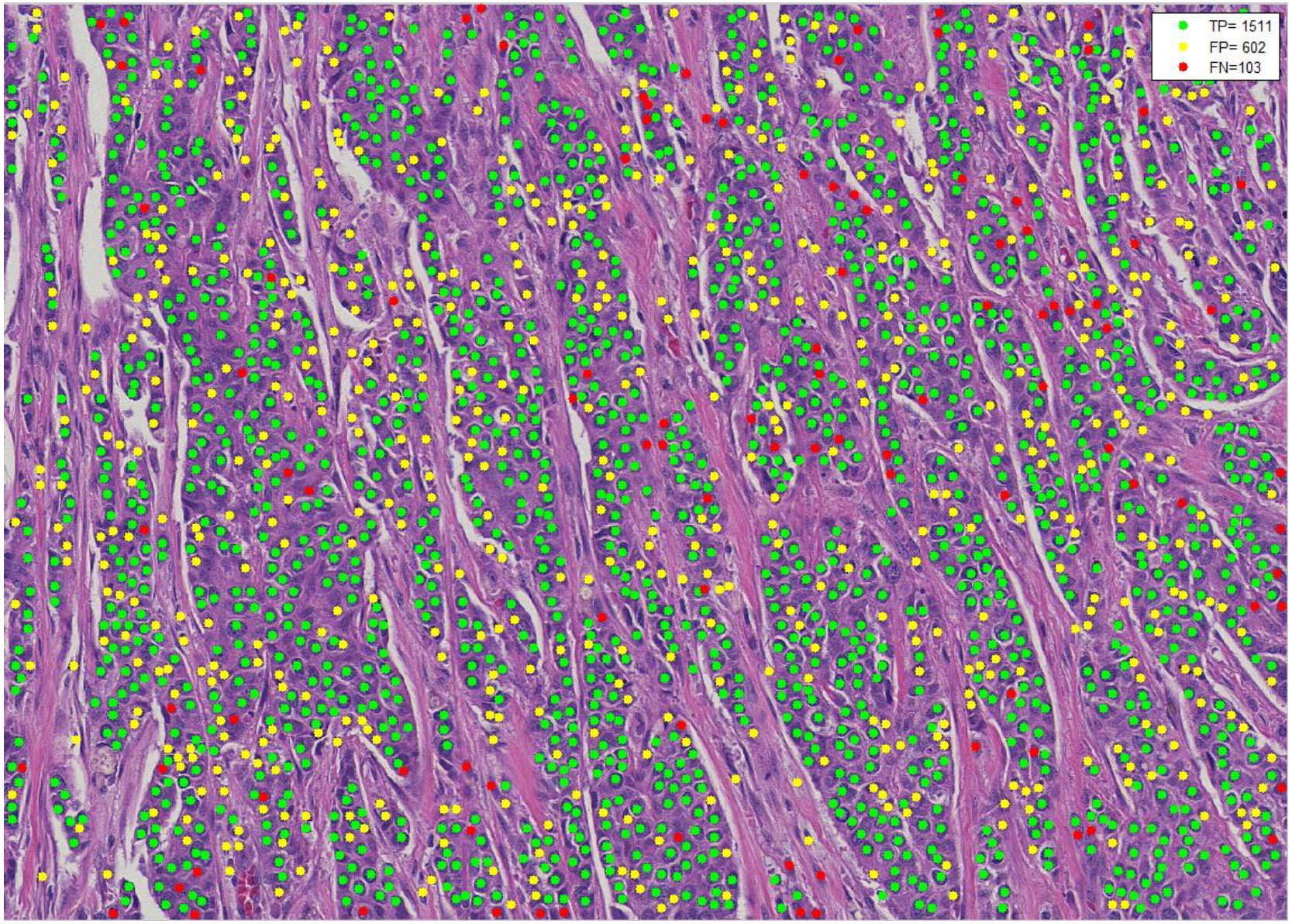


***Xu J** , et al. "*Stacked Sparse Autoencoder (SSAE) based Framework for Nuclei Patch Classification on Breast Cancer Histopathology"， ISBI2014.

***Xu J** , et al. "*Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology". *IEEE Trans. on Medical Imaging,* 2016

*Zhang X, Dou H, **Xu J,** Zhang S,* "Fusing Heterogeneous Features for the Image-Guided Diagnosis of Intraductal Breast Lesions", *IEEE Journal of Biomedical and Health Informatics,* 2016

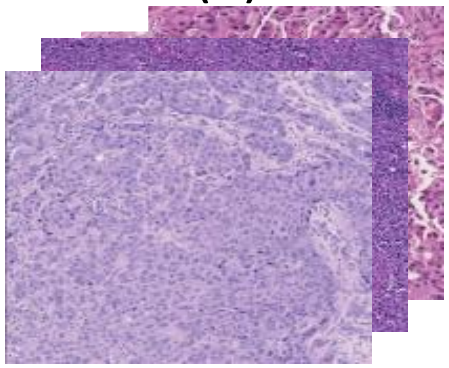
基于稀疏自编码器的细胞检测



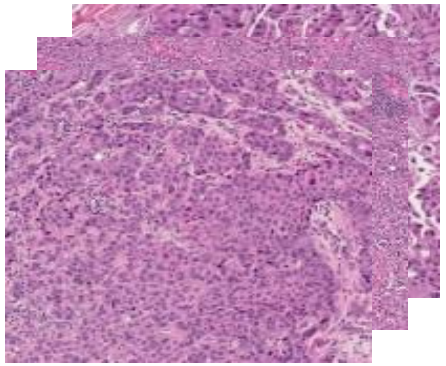
***Xu J** , et al. "*Stacked Sparse Autoencoder (SSAE) based Framework for Nuclei Patch Classification on Breast Cancer Histopathology"，ISBI2014.
***Xu J** , et al. "*Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology". *IEEE Trans. on Medical Imaging,* 2016

基于图像分析的乳腺癌恶性程度自动评分

(a)

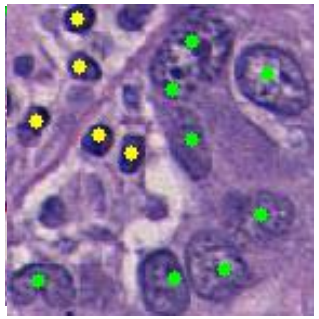


彩色标准化

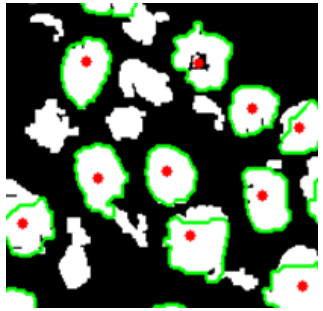


图像预处理

(b)



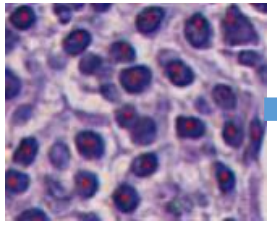
细胞检测

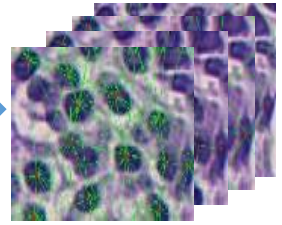


细胞分割

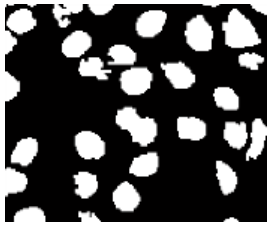
检测与分割

(c)

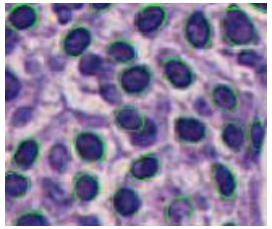


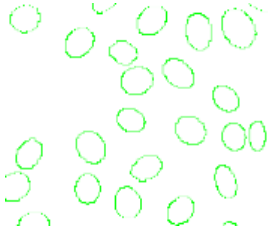


结构特征



纹理特征

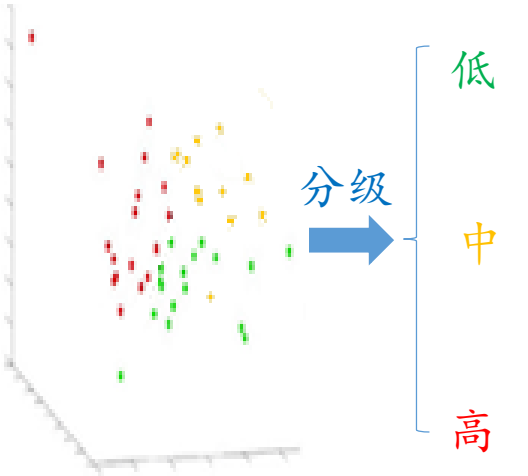




形态特征

特征提取

(d)



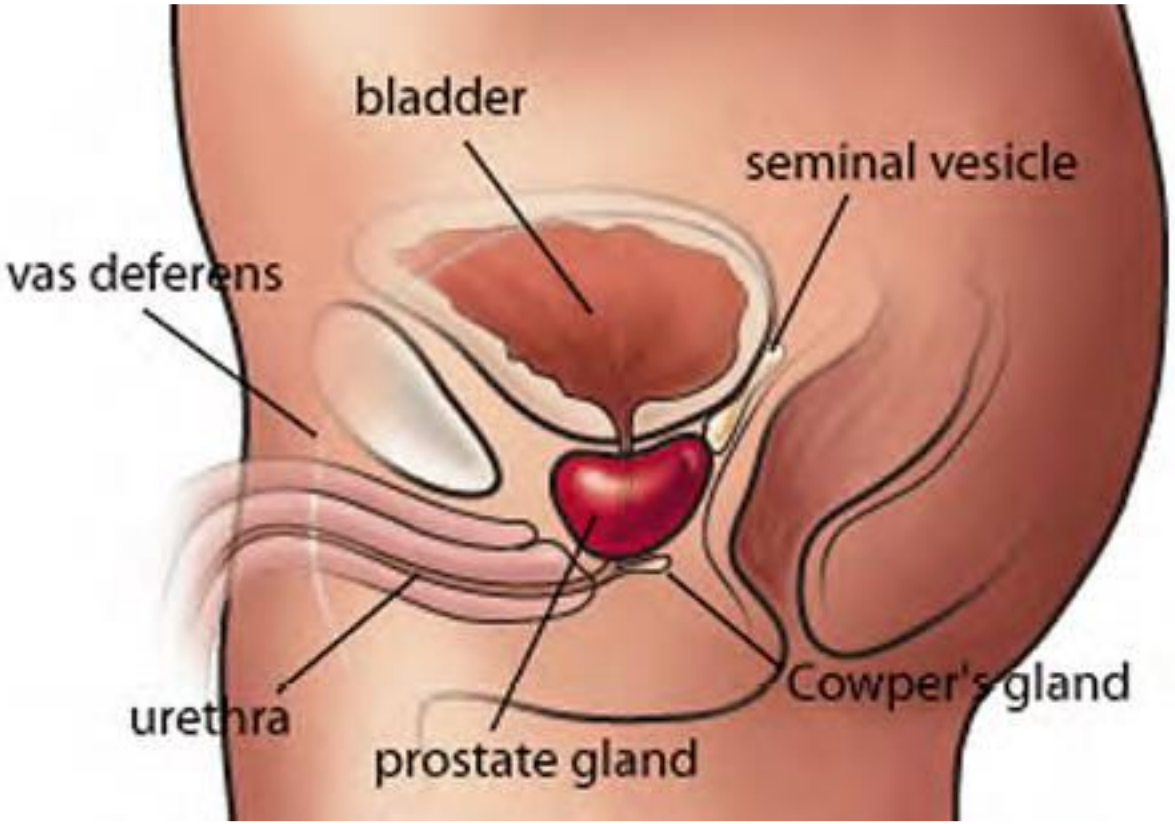
支持向量分类器

自动评分

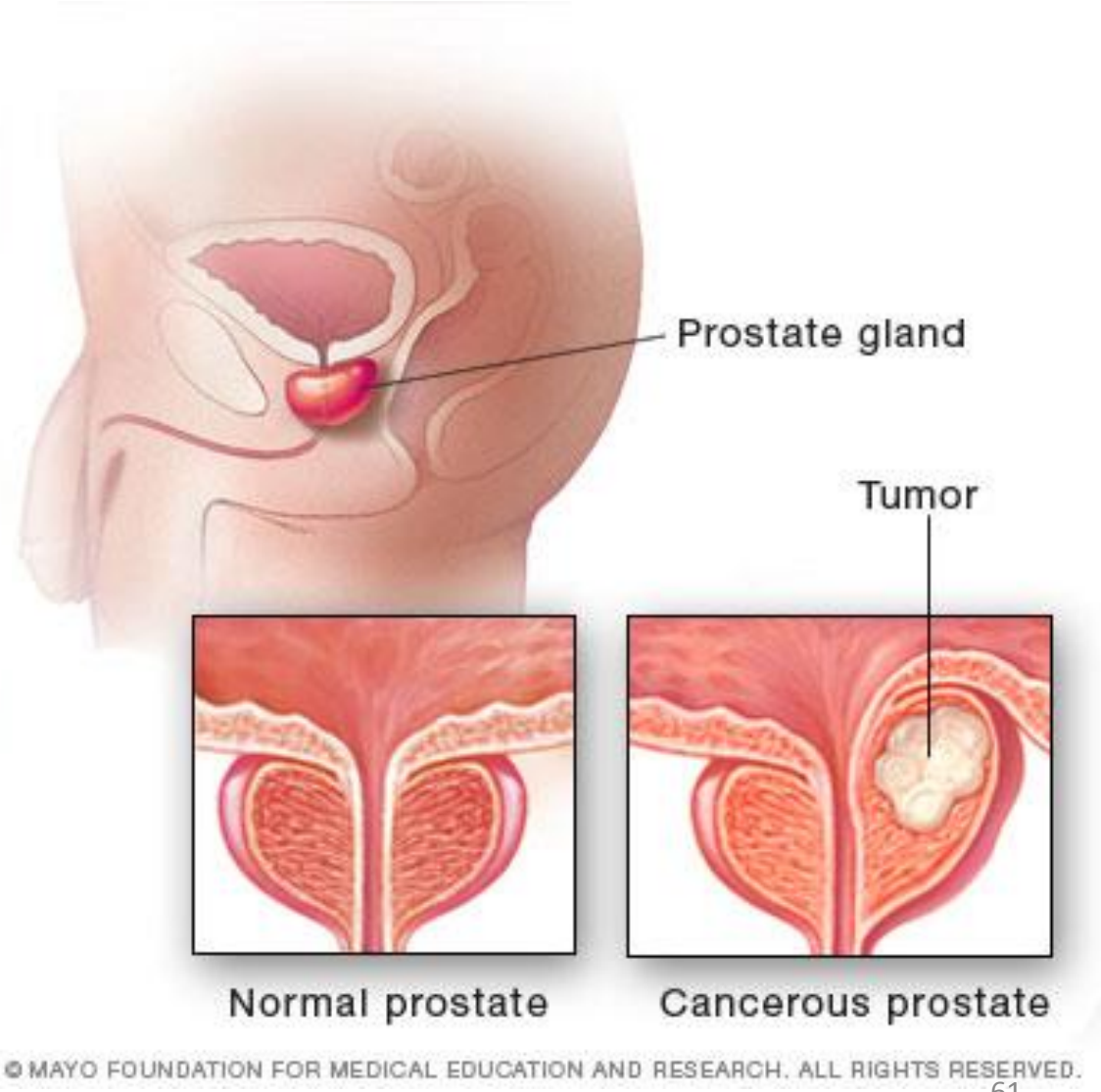
提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

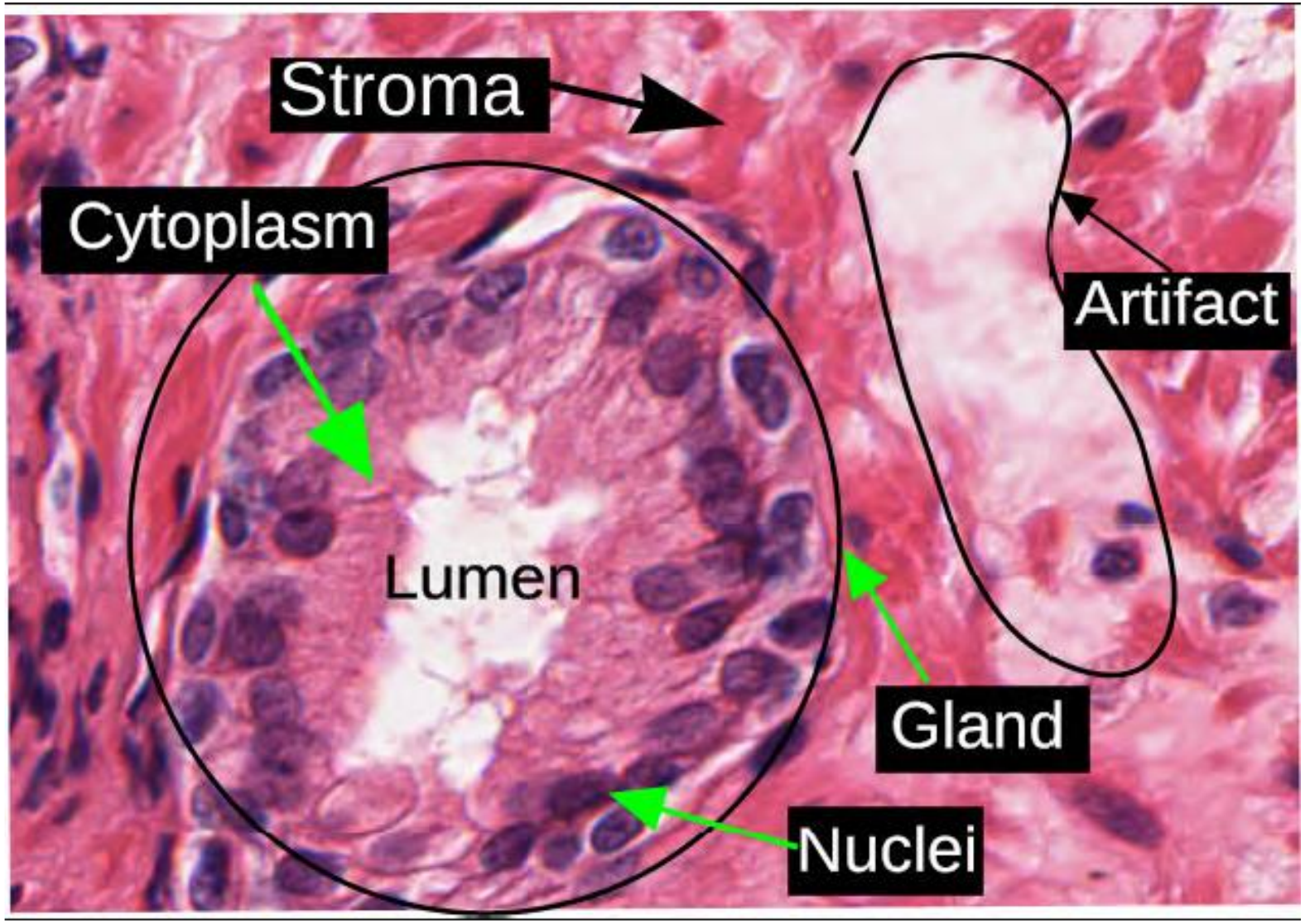
基于图像分析的前列腺癌诊断和预后



Prostate in the male body

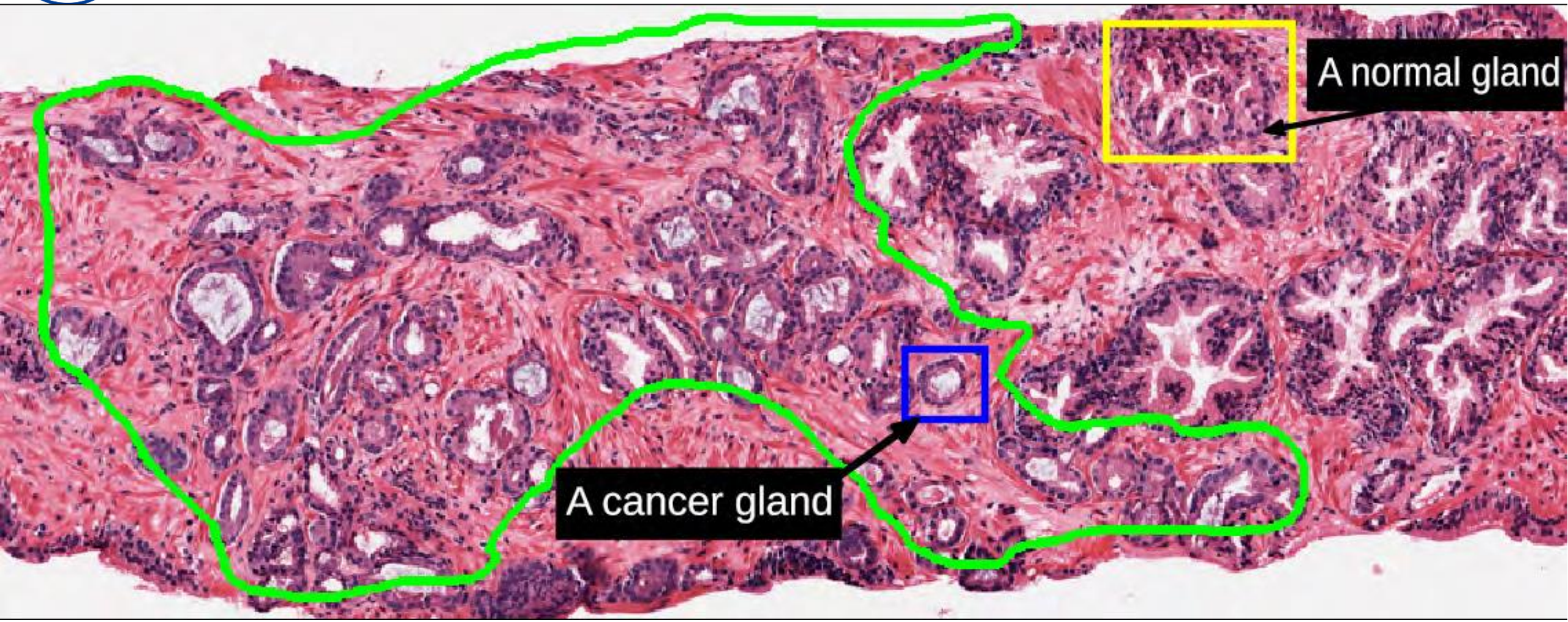


前列腺腺体的基本单元



A gland with basic components (nuclei, cytoplasm and lumen) and an artifact.

正常和癌变的腺体

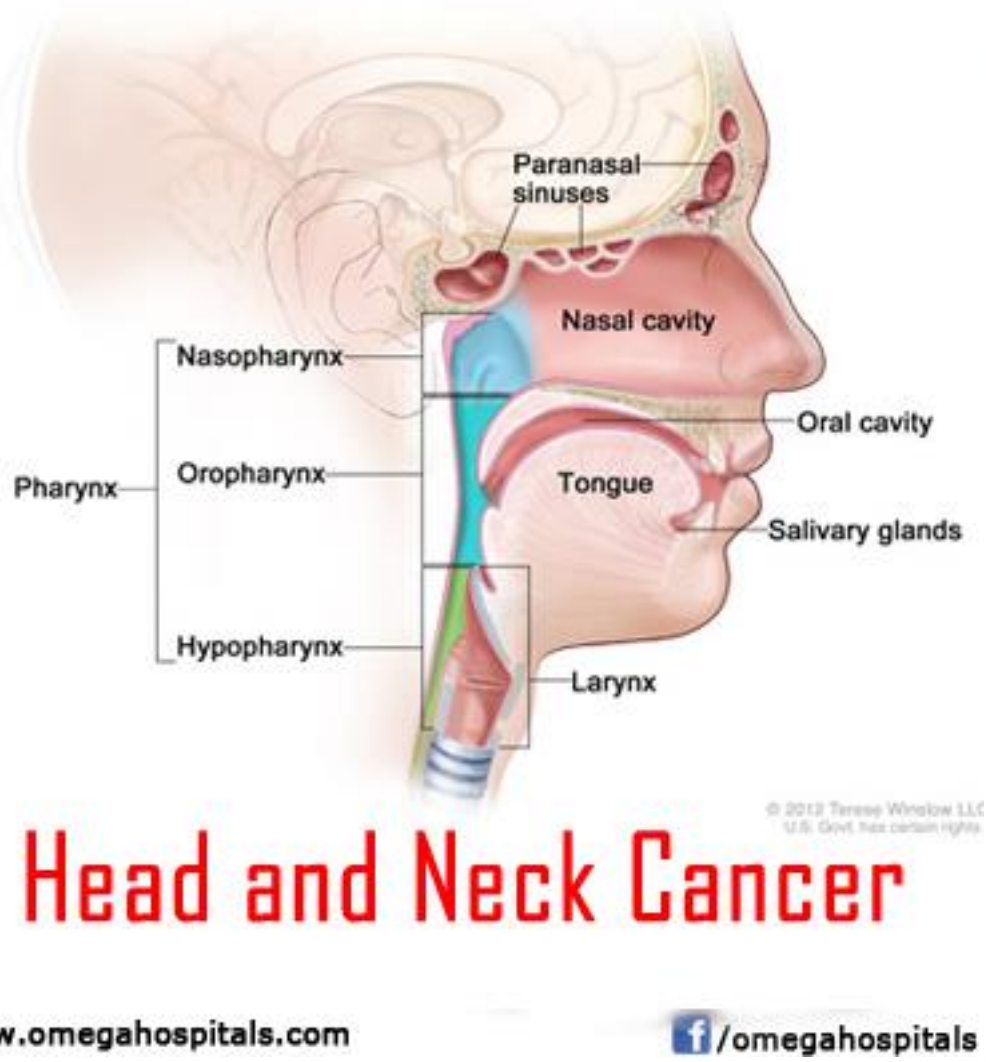


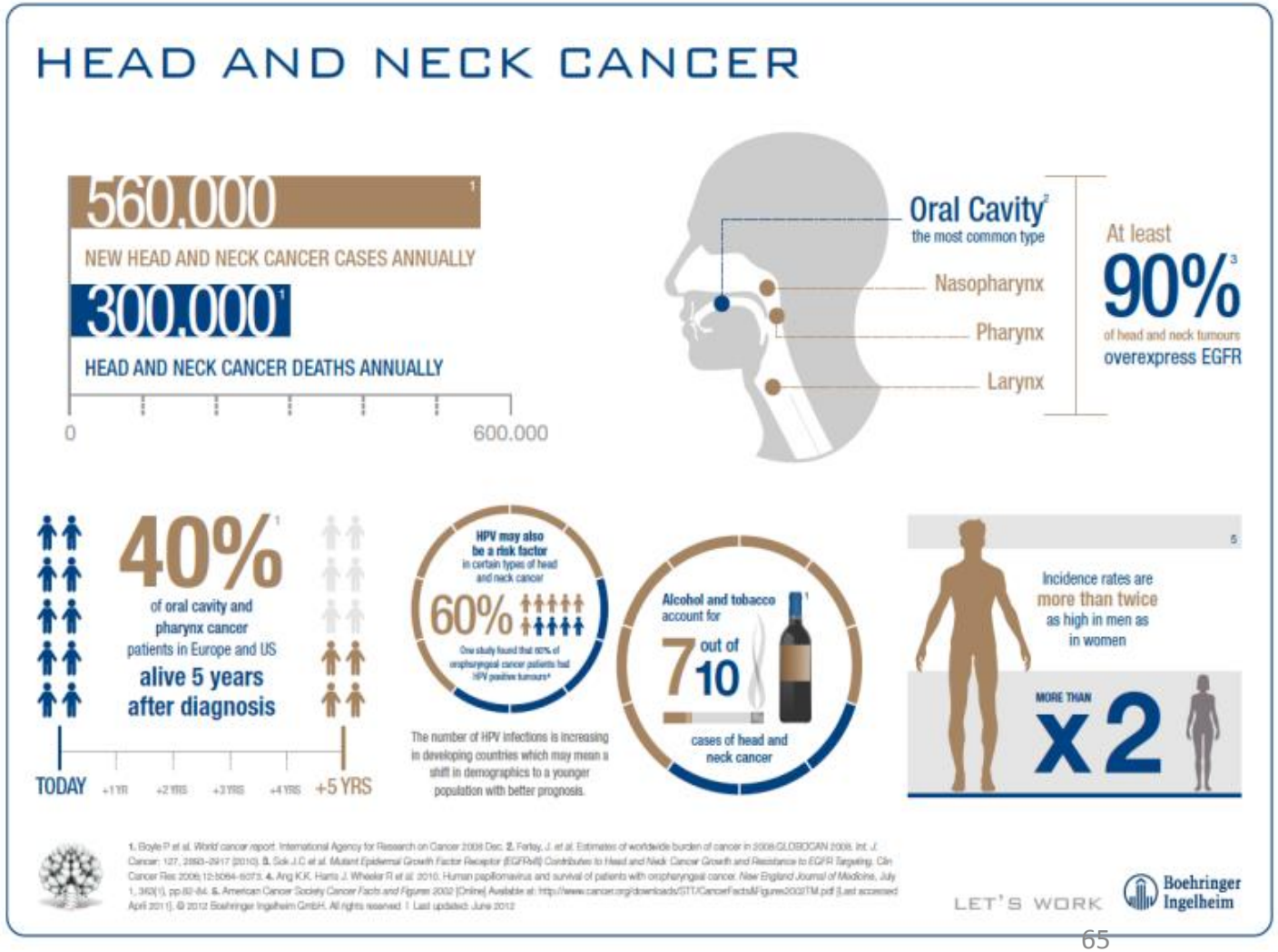
A tissue image showing the cancer glands in a cancer region annotated by a pathologist (green contour); normal glands are present in the region outside the green contour.

提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

基于图像分析的头颈部癌诊断和预后

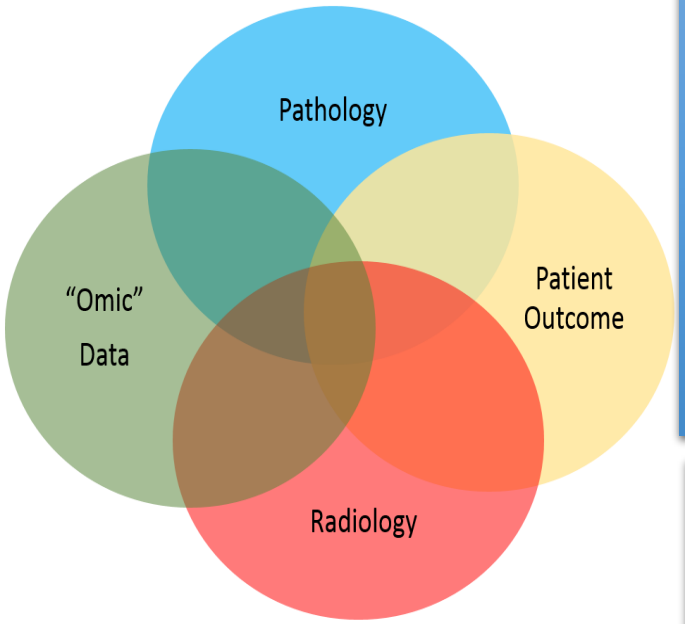


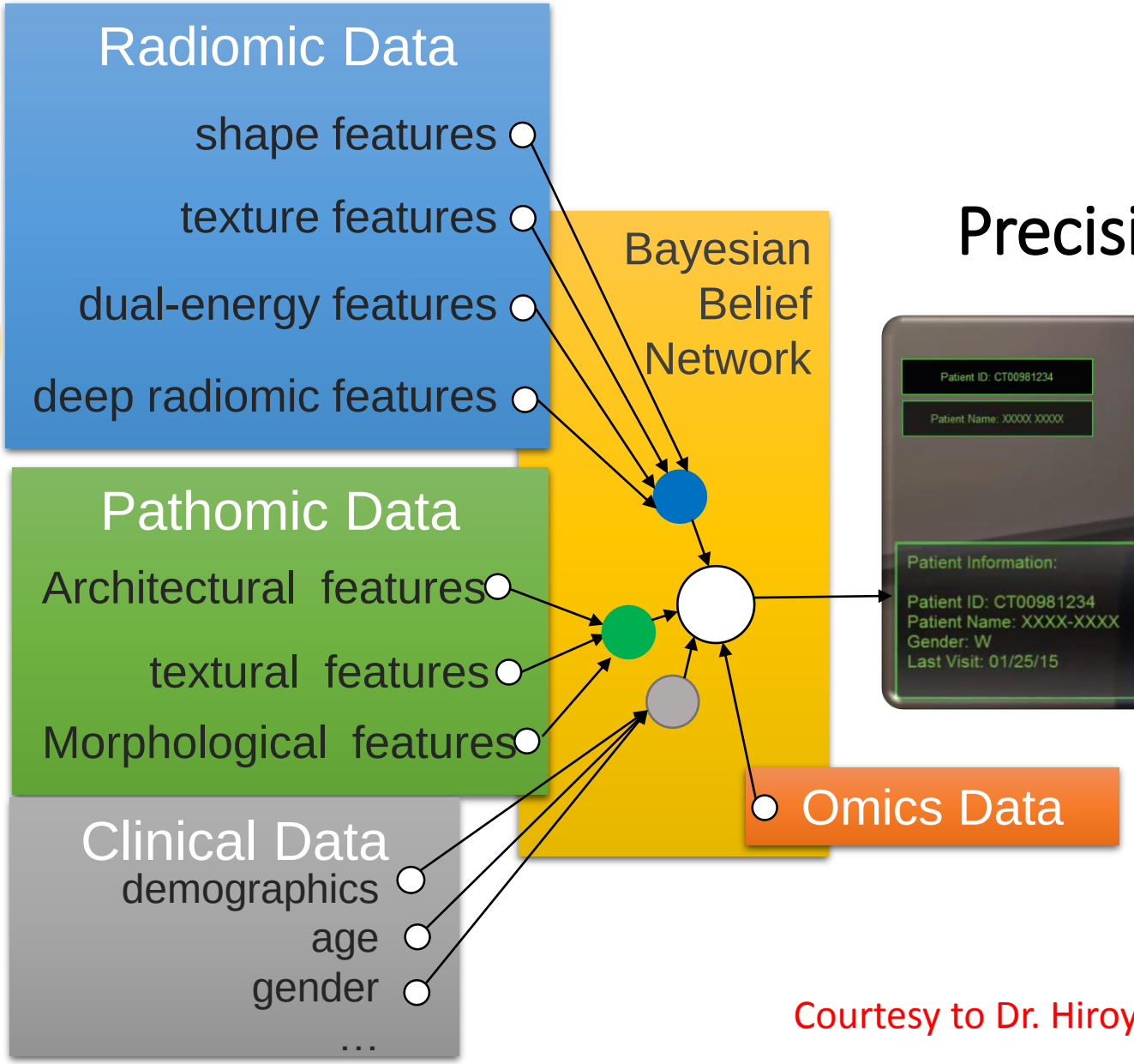


提 纲

- 中美癌症统计、主要类型癌症的五年生存期
- 组织病理分析在癌症诊断与预后中的地位和作用
- 从组织切片到组织病理图像
  - 组织切片的制作、H&E、IHC染色原理
  - 组织切片数字化
  - 病理图像分析的机遇与挑战
- 组织病理图像分析与癌症的计算机辅助诊断与预后
  - 乳腺癌
  - 前列腺癌
  - 头颈部癌
- 未来研究展望

未来的研究展望





Precision Diagnosis

Courtesy to Dr. Hiroyuki Yoshida

未来的研究展望

- Pathomics, Radiomics, Genomics的数据融合---精准医疗
- 与临床专家协同工作，根据医学专家的Domain Knowledge提高算法的有效性
- 开发更有效的高通量的针对不同病理组织结构的检测、分割、特征提取(结构、形态、纹理)算法
- 全扫描组织病理图像分析
- 非常有前景的研究领域，研究成果如何转化为临床应用

Any Questions or Comments?