Review

Diagnostic Pathology and Laboratory Medicine in the Age of “Omics”

A Paper from the 2006 William Beaumont Hospital Symposium on Molecular Pathology

William G. Finn, M.D.

From the Department of Pathology, University of Michigan, Ann Arbor, Michigan

Functional genomics and proteomics involve the simultaneous analysis of hundreds or thousands of expressed genes or proteins and have spawned the modern discipline of computational biology. Novel informatic applications, including sophisticated dimensionality reduction strategies and cancer outlier profile analysis, can distill clinically exploitable biomarkers from enormous experimental datasets. Diagnostic pathologists are now charged with translating the knowledge generated by the “omics” revolution into clinical practice. Food and Drug Administration-approved proprietary testing platforms based on microarray technologies already exist and will expand greatly in the coming years. However, for diagnostic pathology, the greatest promise of the “omics” age resides in the explosion in information technology (IT). IT applications allow for the digitization of histological slides, transforming them into minable data and enabling content-based searching and archiving of histological materials. IT will also allow for the optimization of existing (and often underused) clinical laboratory technologies such as flow cytometry and high-throughput core laboratory functions. The state of pathology practice does not always keep up with the pace of technological advancement. However, to use fully the potential of these emerging technologies for the benefit of patients, pathologists and clinical scientists must embrace the changes and transformational advances that will characterize this new era. (*J Mol Diagn* 2007, 9:431–436; DOI: 10.2353/jmoldx.2007.070023)

The evolution of molecular biology from its birth as a distinct field of science in the 1960s through the present has seen a major shift in emphasis from the creation of novel data-generating methods to the creation of novel informatic systems for the analysis of high-dimensional datasets. The derivation of data previously was the rate-limiting step in biomedical scientific advancement. In past decades of scientific discovery, careers were made by generating data, often through novel laboratory methods. Indeed, Nobel prizes have been awarded to the inventors or discoverers of novel technical laboratory methods, and the discovery and characterization of a single human gene (through traditional methods of linkage analysis, cloning, etc) once was enough to consume the span of an investigator's entire career.

Now, the sequence of the entire human genome resides in the public domain, and datasets derived from novel genomic and proteomic applications are being generated more quickly than they can be meaningfully analyzed. The linear process of hypothesis-driven discovery that characterized past decades of science has been replaced by the hypothesis-generating power of high-throughput, array-based technologies that provide vast and complex datasets that may be mined in various ways using emerging tools of computational biology.^{1–3} Different approaches to these high-throughput methodologies have spawned an assortment of “omics”: genomics,^{4,5} functional genomics (transcriptomics),¹ proteomics,^{4,5} immunomics,⁶ metabolomics,^{7,8} interactomics,⁹ interferomics,¹⁰ etc. Each of these emerging disciplines

This article is a result of material presented at the William Beaumont Hospital 15th Annual Symposium on Molecular Pathology: DNA Technology in the Clinical Laboratory. This symposium took place on September 13–15, 2006 in Troy, MI.

Accepted for publication May 16, 2007.

Address reprint requests to William G. Finn, M.D., University of Michigan Department of Pathology, Room M242 Medical Science I, 1301 Catherine Rd., Ann Arbor, MI 48109-0602. E-mail: wgfinn@umich.edu.

has in common the simultaneous characterization of dozens, hundreds, or thousands of genes (genomics), gene transcripts (transcriptomics), or proteins (proteomics) that, in aggregate and in parallel, reveal aspects of biological function that cannot be culled from traditional linear methods of discovery.

At various times in the history of our specialty, both anatomical and clinical pathology have undergone improvement either by incremental evolution (eg, improvement in existing diagnostic platforms, development of novel reagents, cytochemical stains, immunohistochemistry) or by more disruptive eras of *transformation* (eg, replacement of manual interpretive diagnostic tasks with high-throughput automation; changing paradigms of diagnosis from morphological to molecular definitions as is the case in monitoring of chronic myelogenous leukemia patients on imatinib therapy). During the more disruptive times of transformational change, there may be isolated areas in which pathologists switch from *diagnostic operators*, rendering diagnostic opinions case by case, to *designers of diagnostic operations*, overseeing the development of novel, high-throughput diagnostic platforms. As such, the impending revolution in molecular and computational biology is comparable with previous periods of sweeping change in diagnostic pathology.

The translation of basic science discovery to clinical laboratory testing has long been the purview of practicing pathologists. Traditionally, this task was undertaken via the validation of individual tests as they became available through hypothesis-driven translational research. Now, however, the practice of anatomical and clinical pathology is immersed in the same revolution impacting basic science: the ability to generate massive datasets in short time periods and unprecedented analytical capacity due to information systems advances. Although the adaptation of emerging technologies to clinical diagnostics may seem daunting in the information age, it really represents a continuation of the traditional role of the pathologist as the steward of translational application of new analytical discoveries to diagnostic medicine. But how will the shift in scientific discovery from linear hypothesis-driven validation to high-throughput, hypothesis-generating discovery affect pathology and laboratory medicine, and what can we as diagnostic pathologists learn from the “omics” revolution?

Direct Application of Array Technologies

The simplest answer to the challenge of applying genomic and proteomic methods to diagnostic pathology is the direct application of the high-throughput, array-based technologies that mark the “omics” revolution. The Amplichip CYP450 is often credited as the first commercially marketed clinical testing platform based on DNA microarray technology. This product is an oligonucleotide array jointly developed by Roche (Basel, Switzerland) and Affymetrix (Santa Clara, CA) for the determination of individual genotypes for the *CYP2D6* and *CYP2C19* genes involved in differential rates of drug metabolism by the cytochrome P450 system.^{11,12} The Amplichip system

detects certain mutations or polymorphisms of the *CYP2D6* gene and the *CYP2C19* gene and was approved by the Food and Drug Administration in late 2004 for the determination of genotype of these two genes and the resulting predicted phenotypes (poor metabolizer, intermediate metabolizer, extensive metabolizer, or ultrarapid metabolizer). Results may be used to predict efficacy or adverse reaction to a number of medications (antipsychotics, antidepressants, antiarrhythmics, anticoagulants, antimalarials, etc) that are principally metabolized via the cytochrome P450 system.

More recently, the Food and Drug Administration approved the use of the MammaPrint test, a microarray-based platform for determining the likelihood of breast cancer recurrence based on the gene expression profile of 70 genes within the tumor sample.¹³ MammaPrint uses DNA microarray technology, and the genes selected for analysis on this array were validated via European cohort studies that support the predictive value of gene expression profiles derived from this platform for women aged 60 and younger with node-negative stage I or II breast cancer (<http://www.fda.gov/bbs/topics/NEWS/2007/NEW01555.html>, last accessed February 9, 2007). As more tumor-specific gene expression profiles are generated, more examples of microarray-based technologies in day-to-day laboratory medicine practice will continue to emerge and evolve.

Bioinformatics and the Future of Diagnostic Pathology

Although the direct application of high-throughput, array-based technologies will no doubt advance diagnostic pathology and laboratory medicine, the greatest contribution of the genomic and proteomic revolution to laboratory testing will be in the continued development of the novel field of *computational biology* and in the application of informatics to laboratory medicine. Advances in information technology will not only allow the translation of genomic and proteomic discovery to the diagnostic testing platforms but will also allow us to transform existing diagnostic modalities such as the traditional histopathological examination and allow us to optimize more recent technologies such as clinical flow cytometry and molecular diagnostics.

Dimensionality Reduction: Making Sense of Enormous Datasets

The derivation of six or eight blood cell indices from a routine hematology analyzer can be reported directly to the patient record. The complex microanatomy of a liver biopsy can be subjectively interpreted by a qualified pathologist. What, however, does one do to accommodate the clinical application of a 50,000-gene array dataset or an extensive proteomics profile? With the development of high-throughput array technologies, much effort has been spent in the distillation of unmanageably large

datasets into relevant principal components via some form of factor analysis or *dimensionality reduction*.^{14–16}

The concept of dimensionality reduction involves the discovery of the *intrinsic dimensionality* of a given high-dimensional dataset. It is possible that the essence of a 50,000-gene array dataset can be captured in the distillation of those 50,000 dimensions into a few dozen that can be manipulated to represent the principal components of the high-throughput analysis.

A model frequently used by engineers is the *Swiss roll*.¹⁵ Imagine taking the piece of paper you are currently reading and rolling it up. The result is a true three-dimensional structure, with length, height, and width that can be measured as separate dimensions. However, the entire essence of that object can be observed by unrolling it into a two-dimensional structure (the original piece of paper). In other words, the three-dimensional roll has an *intrinsic dimensionality* of two dimensions. Put another way, the roll represents data mapped to a two-dimensional *manifold* embedded within three-dimensional space.

Analysis of high-dimensional datasets can be considerably enhanced through the discovery of the reduced manifold within which the high-dimensional dataset resides. Part of the revolution in computational biology has involved the discovery of novel dimensionality reduction strategies and manifold embedding techniques. Although a detailed review of specific dimensionality reduction methods is beyond the scope of this review, many groups have devised novel approaches to complement more traditional forms of factor or principal component analysis.^{15, 17–21}

Cancer Outlier Profile Analysis: Clinically Applicable Computational Biology

Dimensionality reduction strategies may suffer from an inability to extract the essence of recurring specific characteristics that may only be present on a subset of cases within a group. For instance, specific classes of cancer may show heterogeneous patterns of gene amplification, fusion, mutation, or deletion. Therefore, large-scale analysis of such cancer classes over many samples may fail to identify consistent patterns of gene overexpression because of this heterogeneity. Using existing datasets from gene expression profiling experiments, Tomlins et al²² applied a novel bioinformatics approach dubbed cancer outlier profile analysis as a means by which to identify recurring patterns of gene overexpression that may characterize distinct subsets of known cancer types, but that may not be detectable using traditional analysis methods (such as *t*-tests or signal-to-noise ratios) because of consistent overexpression of the target gene(s) in only a subset of cases. Using cancer outlier profile analysis, two members of the ETS family of transcription factors, ETV1 and ERG, were identified as outliers in prostate cancer. Further analysis revealed that outlier overexpression of these two transcription factors was mutually exclusive, implying that they may be functionally redundant in prostate cancer development and that they

may be involved in similar mechanisms of gene expression amplification. Additional analysis of cDNA transcripts of ERG and ETV1 in prostate cancer cell lines indicated fusion of the 5' untranslated region of *TM-PRSS2* (a prostate-specific, strongly androgen-regulated gene) to either ERG or ETV1. This discovery was validated by fluorescence *in situ* hybridization as corresponding to cytogenetic translocations involving the *TM-PRSS2* locus on chromosome 21q22.3 and the corresponding chromosomes harboring one of the ETS family genes.

Thus, purely computational manipulation and meta-analysis of existing high-throughput gene expression datasets eventually led to discovery of a novel group of recurring chromosomal translocations in prostate cancer.²³ This stunning discovery will have an extraordinary impact on prostate cancer diagnosis, with the ETS/*TM-PRSS2* fusions being easy targets for clinical fluorescence *in situ* hybridization analysis, and on prostate cancer management and treatment. This discovery also signals a new era in the rapid translation of new biological discovery to the clinical diagnostic laboratory.¹

The Impact of Information Systems Advances on Histopathology and Microscopy

Rapidly evolving information technologies and the doubling of computer speed and memory every 18 months or so ("Moore's law") are beginning to have a measurable impact on the practice of clinical histopathology. Histological sections can be scanned routinely and rapidly into digital "virtual microscopy" platforms.²⁴ Once this occurs, a subjective pattern of microanatomy becomes an objective set of *data*: data that can be mined, manipulated, and analyzed just as is the case currently with high-dimensional microarray datasets.

Once histopathological slides are transferable into digital datasets, the possibilities become virtually limitless. Currently, the digitization of histological slides is primarily applied to education, but it also enables remote diagnostics through distant manipulation of the virtual microscope platform, thereby enabling pathologists to provide microscopic services regardless of their distance to the site of acquisition of a biopsy or other surgical sample.²⁵ On a higher level, digitization of histopathology data allows for a transformation of the way we store, maintain, and retrieve histological materials.

Currently, glass slides are physically stored in laboratories and warehouses and linked to accession numbers and text reports. Diagnostic coding systems such as SNOMED (CAP, Northfield, IL) can then be used to search and retrieve examples of specific types of disease. Searches involve entering text strings such as "follicular lymphoma" or "glioblastoma" or other corresponding codes and then retrieving lists of slides linked to those search terms. Scanned images of histology slides, however, can be compressed via dimensionality reduction strategies such as vector quantization (a lossy compression method that divides images into local domains and extracts composite vectors from those domains and is

used, for instance, in the creation and storage of image or sound files in computer applications).²⁶ The stored composite vectors can be queried against an established library ("vocabulary") of established vectors to find the closest match. In other words, one can develop systems whereby searches are not based on *what we called* a given tumor but instead are based on *what the tumor actually looks like*. This content-based search capability will revolutionize histopathology and will pave the way to robust automated interpretation systems for anatomical pathology.

Many approaches to automated image recognition, archiving, and retrieval are being developed in addition to the vector quantization approach described above. Examples of such systems include innovative graph-embedding techniques,²⁷ novel approaches to shape recognition combined with data reduction for the development of automated cell identification as a decision-support tool,²⁸ and robotic telepathology combined with sophisticated color decomposition algorithms for the automated analysis of staining patterns on histological or immunohistochemical preparations,²⁹ among others.

From the standpoint of histopathology research, this type of data compression and storage allows for analysis of histological images as high-dimensional datasets directly analogous to the microarray datasets currently used in gene expression profiling experiments.³⁰ Imagine viewing the now-familiar red-and-green "heat maps" that are embedded in every gene expression profiling manuscript but with the heat map representing the digital signature of a histological slide, analyzable by hierarchical cluster analysis, principal component analysis, or any computational method currently applied to high-dimensional gene array datasets.

The evolution of the virtual microscope may also transform the job description of histotechnologists. Although clearly there will be continuing demand for talented histotechnologists in the preparation of histological materials, one can envision a time in which histotechnologists expand their expertise to include image scanning, storage, and retrieval. Histotechnologists would then become curators of vast image archives, with context-based rather than text-based search capability.

Although the capability to transform histopathology slides into high-dimensional datasets exists today and will improve rapidly with time, it remains to be seen just how this capability will be adopted into day-to-day clinical practice. Currently, these systems are predominantly used in education: medical school histopathology courses and regional and national continuing education programs are rapidly converting to virtual microscopy platforms.^{31–33} In the future, the content-based search capabilities of such digital platforms could have substantial impact on the retrieval of appropriate case material for clinical and translational research.

It is also tempting to speculate that such technology could evolve into automated histopathology diagnosis. However, this seems quite unlikely in the foreseeable future, since true medical diagnosis remains a highly complex interpretive task that will continue to require the

final interpretation and opinion of a licensed physician, and automated systems will never be licensed to practice medicine. More realistically, it would seem that the digital histology platform would evolve as a robust decision-support tool for the practicing pathologist. Ultimately, analysis paradigms will have to be derived by consensus among diagnostic pathologists. The power of the digital histology platform will be in the unsupervised discovery of differences and similarities among disease entities that may not have been readily apparent by traditional histological examination.

Optimizing Clinical Flow Cytometry through Information Technology

Flow cytometry is a prime example of a transforming technology that uses the power of modern technology combined with modern information systems to yield diagnostic data, usually for the diagnosis and classification of leukemias and lymphomas. Flow cytometry is also a prime example of a transforming technology whose potential is generally underused by clinical practitioners.

Through the 1980s and 1990s, clinical flow cytometry expanded the power of immunophenotyping by allowing for the simultaneous evaluation of surface antigen expression patterns at the level of the single cell. Early in the clinical application of flow cytometry, computer memory was quite expensive, and processor speed was a fraction of what it is today. As a result, the analysis of the multitude of data generated by routine flow cytometry was largely a real-time endeavor, with users having to "gate" on cell populations of interest and exclude others in the name of preserving precious computer memory. Data were then stored in the form of finished analyses rather than in the form of raw surface marker and light scatter characteristics for each cell analyzed ("list mode").

The technological and informatic limitations on flow cytometry led by necessity to a process of exclusionary gating of cell populations based on combinations of marker expression and laser light scatter characteristics and analyzing the gated populations for the presence or absence of certain markers or combinations of markers.^{34,35} As the technology evolved, however, the accelerated capabilities of computer processors and the cheap cost of computer memory allowed the saving of every bit of information acquired by the flow cytometric interrogation of every single event (cell) analyzed. This now allows for the dynamic, iterative, multifaceted analysis of surface antigen patterns through the virtual analysis of high-dimensional list mode datasets. Pathologists can therefore subjectively examine visual patterns on ungated flow cytometry histograms, rendered high-dimensional through the use of color-eventing (or "painting") to delineate distinct cell subtypes based on the now routine use of five or six-color (seven- or eight-parameter) clinical flow cytometers.^{36,37} Diseases can then be characterized by their overall patterns or shapes on such high-dimensional analyses rather than having to use the antiquated approach of gating single populations and numerically translating the magnitude of each individual

marker.^{38,39} High-dimensional flow cytometry datasets can also be reduced to lower dimensional manifolds, analogous to dimensionality reduction strategies for genomic or proteomic datasets.⁴⁰

Despite considerable advancements in the capability of even the most routine clinical flow cytometers, most clinical laboratories still choose the exclusionary processes of gating and analyzing predetermined cell populations that were mandated in an age when computers simply could not keep up with the data being acquired. This can lead to numerous pitfalls, including inadvertent exclusion of cell populations that may be of diagnostic interest.⁴¹ Further progress will require the full realization of the current potential of diagnostic flow cytometry.

Although clinical flow cytometry predated the "omics" revolution, it is really a form of proteomic analysis: the simultaneous evaluation of expression of several surface proteins in a single patient sample. A potential disadvantage of flow cytometry as a proteomic platform is the limited number of proteins analyzable in routine clinical practice (a few dozen) versus the hundreds or thousands of proteins detectable by other high-throughput platforms commonly used in research (such as mass spectrometry or fixed protein arrays). However, this disadvantage is compensated for by the fact that flow cytometry yields this proteomic information *at the level of the intact cell* rather than within cell homogenates or supernatants. This allows for the virtual (rather than physical) selection of cell populations within the same analysis, based on any number of user-defined antigen expression or light scatter criteria. The evaluation of protein expression patterns, metabolic pathway function, etc, at the level of the intact cell is the essence of the emerging field of *cytomics*.⁴² If the potential of clinical flow cytometry is to be realized, then it must be approached in the future as a cytomic platform, not simply a platform for the linear process of surface antigen detection or validation.

Newer approaches to flow cytometry as a cytomic platform are now emerging, which include the unsupervised hierarchical cluster analysis of flow cytometry data in the classification of disease, the elucidation of molecular pathways at the level of the individual cell, and the treatment of flow cytometry data as a form of high-dimensional space rather than a set of surface marker expression values.^{43–49} The acceleration in information systems technology will, I hope, finally fulfill the potential of flow cytometry as a clinical diagnosis and management tool.

Facing the Future

It is exciting to consider the potential that the revolution in genomics, proteomics, and computational biology will have on the future of diagnostic pathology and laboratory medicine. However, it is of more pressing importance to examine whether the potential of bioinformatics capabilities are being fully realized in the current practice of pathology and laboratory medicine. Currently, biophysicists, engineers, and computer scientists design highly sophisticated clinical testing platforms that use the latest in precision design, hardware, optics, and software.

Once created, these technologies are then sold to laboratories run by physicians, who are generally less expert in fundamentals of these technologies. The result is likely that currently available technologies may be underused in clinical laboratory medicine practice.⁵⁰

In the information age, it is important for laboratory supervisors and medical directors to team up with process engineers to allow for the optimization of technology and information systems in the clinical laboratory. Deployment of emerging technologies and information systems in clinical laboratories will require fostering a sense of innovation in diagnostic pathology and laboratory medicine, fields that have historically benefited from a culture of stability and process-oriented control. The future will probably be shaped by the extent to which we can balance the need for control in quality and safety issues with the need for innovation in new diagnostic approaches for the benefit of patients. As the pace of technological advancement accelerates, it behooves us to harness the potential of emerging informatic, genomic, and proteomic applications for the optimization of our specialty.

References

1. Rhodes DR, Chinnaiyan AM: DNA microarrays: implications for clinical medicine. *J Invest Surg* 2002, 15:275–279
2. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM: Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001, 412:822–826
3. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JJ, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000, 403:503–511
4. Conrads TP, Hood BL, Issaq HJ, Veenstra TD: Proteomic patterns as a diagnostic tool for early-stage cancer: a review of its progress to a clinically relevant tool. *Mol Diagn* 2004, 8:77–85
5. Fenselau C: A review of quantitative methods for proteomic studies. *J Chromatogr B Analyt Technol Biomed Life Sci* 2006, [Epub ahead of print]
6. Wang X, Yu J, Sreekumar A, Varambally S, Shen R, Giacherio D, Mehra R, Montie JE, Pienta KJ, Sanda MG, Kantoff PW, Rubin MA, Wei JT, Ghosh D, Chinnaiyan AM: Autoantibody signatures in prostate cancer. *N Engl J Med* 2005, 353:1224–1235
7. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L: HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007, 35:D521–D526
8. Larive CK: Metabonomics, metabolomics, and metabolic profiling. *Anal Bioanal Chem* 2007, 387:523
9. Ideker T, Valencia A: Bioinformatics in the human interactome project. *Bioinformatics* 2006, 22:2973–2974
10. Sachidanandam R: RNAi as a bioinformatics consumer. *Brief Bioinform* 2005, 6:146–162
11. de Leon J, Susce MT, Murray-Carmichael E: The AmpliChip CYP450 genotyping test: integrating a new clinical tool. *Mol Diagn Ther* 2006, 10:135–151
12. Heller T, Kirchheiner J, Armstrong VW, Luthe H, Tzvetkov M, Brockmoller J, Oellerich M: AmpliChip CYP450 GeneChip: a new gene chip that allows rapid and accurate CYP2D6 genotyping. *Ther Drug Monit* 2006, 28:673–677

13. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, Wessels LF, Van't Veer LJ: Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006, 7:278
14. Rollins DK, Zhai D, Joe AL, Guidarelli JW, Murarka A, Gonzalez R: A novel data mining method to identify assay-specific signatures in functional genomic studies. *BMC Bioinformatics* 2006, 7:377
15. Costa JA, Hero AO: Classification constrained dimensionality reduction. Presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005 March 19–23, Philadelphia, PA
16. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999, 96:2907–2912
17. Donoho DL, Grimes C: Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc Natl Acad Sci USA* 2003, 100:5591–5596
18. Yan S, Xu D, Zhang B, Zhang HJ, Yang Q, Lin S: Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans Pattern Anal Mach Intell* 2007, 29:40–51
19. Souvenir R, Pless R: Image distance functions for manifold learning. *Image Vis Comput* 2007, 25:365–373
20. Belkin M, Niyogi P: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 2003, 15:1373–1396
21. Agrafiotis DK, Xu H: A self-organizing principle for learning nonlinear manifolds. *Proc Natl Acad Sci USA* 2002, 99:15869–15872
22. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 2005, 310:644–648
23. Rubin MA, Chinnaiyan AM: Bioinformatics approach leads to the discovery of the TMPRSS2:ETS gene fusion in prostate cancer. *Lab Invest* 2006, 86:1099–1102
24. Catalyürek U, Beynon MD, Chang C, Kurc T, Sussman A, Saltz J: The virtual microscope. *IEEE Trans Inf Technol Biomed* 2003, 7:230–248
25. Helin H, Lundin M, Lundin J, Martikainen P, Tammela T, van der Kwast T, Isola J: Web-based virtual microscopy in teaching and standardizing Gleason grading. *Hum Pathol* 2005, 36:381–386
26. Balis UJ: Implementation of a region of interest-based query using vector quantization, generalized affine class-based vocabularies, and multimodal Chebyshev polynomial normalization to retrieve context-matched imagery from existing digital image repositories. Abstract. *Arch Pathol Lab Med* 2005, 129:811
27. Madabhushi A, Shi J, Rosen M, Tomaszewski JE, Feldman MD: Graph embedding to improve supervised classification and novel class detection: application to prostate cancer. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv* 2005, 8:729–737
28. Foran DJ, Comaniciu D, Meer P, Goodell LA: Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy. *IEEE Trans Inf Technol Biomed* 2000, 4:265–273
29. Chen W, Foran DJ, Reiss M: Unsupervised imaging, registration and archiving of tissue microarrays. *Proc AMIA Symp* 2002, 136–139
30. Kashgarian M: The contribution of quantitative techniques including morphometry to renal diagnosis. *Ultrastruct Pathol* 2006, 30:339–343
31. Kumar RK, Velan GM, Korell SO, Kandara M, Dee FR, Wakefield D: Virtual microscopy for learning and assessment in pathology. *J Pathol* 2004, 204:613–618
32. Dee FR, Lehman JM, Consoer D, Leaven T, Cohen MB: Implementation of virtual microscope slides in the annual pathobiology of cancer workshop laboratory. *Hum Pathol* 2003, 34:430–436
33. Krippendorf BB, Lough J: Complete and rapid switch from light microscopy to virtual microscopy for teaching medical histology. *Anat Rec B New Anat* 2005, 285:19–25
34. Thompson JM, Gralow JR, Levy R, Miller RA: The optimal application of forward and ninety-degree light scatter in flow cytometry for the gating of mononuclear cells. *Cytometry* 1985, 6:401–406
35. Borowitz MJ, Guenther KL, Shults KE, Stelzer GT: Immunophenotyping of acute leukemia by flow cytometric analysis. Use of CD45 and right-angle light scatter to gate on leukemic blasts in three-color analysis. *Am J Clin Pathol* 1993, 100:534–540
36. Braylan RC, Benson NA, Iturraspe J: Analysis of lymphomas by flow cytometry. Current and emerging strategies. *Ann NY Acad Sci* 1993, 677:364–378
37. Kussick SJ, Wood BL: Using 4-color flow cytometry to identify abnormal myeloid populations. *Arch Pathol Lab Med* 2003, 127:1140–1147
38. Kroft SH: Role of flow cytometry in pediatric hematopathology. *Am J Clin Pathol* 2004, 122(Suppl):S19–S32
39. McKenna RW, Washington LT, Aquino DB, Picker LJ, Kroft SH: Immunophenotypic analysis of hematogones (B-lymphocyte precursors) in 662 consecutive bone marrow specimens by 4-color flow cytometry. *Blood* 2001, 98:2498–2507
40. De Rosa SC, Brenchley JM, Roederer M: Beyond six colors: a new era in flow cytometry. *Nat Med* 2003, 9:112–117
41. Lagaay AM, Bosman CB, Van der Keur M, Ligthart GJ, Schuit HR, Tanke HJ, Hijmans W: Gating of the so-called 'lymphocytic' cell population for the quantification of natural killer cells (CD16⁺) by flow cytometry causes loss of CD16 positive cells. *J Immunol Methods* 1990, 133:235–244
42. Valet GK, Tarnok A: Cytomics in predictive medicine. *Cytometry B* 2003, 53:1–3
43. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP: Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 2005, 308:523–529
44. Perez OD, Nolan GP: Phospho-proteomic immune analysis by flow cytometry: from mechanism to translational medicine at the single-cell level. *Immunol Rev* 2006, 210:208–228
45. Habib LK, Finn WG: Unsupervised immunophenotypic profiling of chronic lymphocytic leukemia. *Cytometry B Clin Cytom* 2006, 70:124–135
46. Roederer M, Hardy RR: Frequency difference gating: a multivariate method for identifying subsets that differ between samples. *Cytometry* 2001, 45:56–64
47. Roederer M, Moore W, Treister A, Hardy RR, Herzenberg LA: Probability binning comparison: a metric for quantitating multivariate distribution differences. *Cytometry* 2001, 45:47–55
48. Toedling J, Rhein P, Ratei R, Karawajew L, Spang R: Automated in-silico detection of cell populations in flow cytometry readouts and its application to leukemia disease monitoring. *BMC Bioinformatics* 2006, 7:282
49. Petrusch U, Haley D, Miller W, Floyd K, Urba WJ, Walker E: Polychromatic flow cytometry: A rapid method for the reduction and analysis of complex multiparameter data. *Cytometry A* 2006, 69:1162–1173
50. Davis GM, Lantis KL, Finn WG: Laboratory hematology practice: present and future. *Cancer Treat Res* 2004, 121:167–179