

A Systems Approach to Model Metastatic Progression

Barry S. Taylor,^{1,3} Sooryanarayana Varambally,^{1,4} and Arul M. Chinnaiyan^{1,2,3,4}

Departments of ¹Pathology, ²Urology, and ³Bioinformatics Program and ⁴Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan

Abstract

Proteomic profiling of human disease has seen much early activity with the accessibility of the newest generation of high-throughput platforms and technologies. Nevertheless, the nature of the dynamic physiologic milieu and high dimensionality of the data has complicated major diagnostic and prognostic breakthroughs. Our recent article in *Cancer Cell* delineates an integrative model for culling a molecular signature of metastatic progression in prostate cancer from proteomic and transcriptomic analyses and shows its facility as a predictor of prognosis. The study leveraged direct proteomic analysis of tumor tissue extracts, differential feature selection characterizing the proteomic alterations of prostate cancer subclasses, and integration with public and study-derived genomic data to construct a multiplex gene signature representing progression of indolent cancer to aggressive disease. This further predicted clinical outcome in a variety of solid tumors. This review describes the context of the work, the framework for the analysis itself, and a look forward to the promise of this systems approach to human disease. (*Cancer Res* 2006; 66(11): 5537-9)

Introduction

Prostate carcinoma is the leading cancer diagnosis in American men and second only to lung cancer in mortality (1). However, the mechanics of its transition from clinically localized to aggressive disease is only vaguely understood. As our understanding of causal cancer genetics matures (2) and as molecular profiling of individual markers of progressive disease develops (3), so too must new approaches that incorporate a systems perspective (4). This will integrate profiling data from genomic and proteomic studies to fully dissect the biological networks and molecular programs that dictate metastatic progression, which complicates prognosis and confounds clinical intervention.

Proteomic Profiling of Human Tumors: A Discriminatory Challenge

In parallel with the extensive body of work on the cancer transcriptome in classification, signatures, and regulatory networks (5–7), the cancer research community devotes significant attention to the proteomic alterations implicated in cancer. This includes work in protein identification, relative abundance quantitation, localization, and modifications. Additionally, a variety of technologies and methods development efforts are committed to sample preparation, mixture complexity reduction, and computational

analyses in characterizing pathologically benign and disease tissues. These are well reviewed elsewhere (8), and our focus here is on the molecular pathology of prostate cancer progression and how whole-proteome analysis provides a powerful and complementary front in high-throughput disease science.

Cancer profiling efforts in cell lines and samples ranging from biofluids to tissue generate secretory proteomes, phosphoproteomes, partial interactomes, and myriad discoveries that dovetail with clinical proteomics research in diagnostic and prognostic biomarkers. Work in these domains include discriminating serologic fingerprints of prostate cancer (9), quantification of secreted and cell-surface proteins in prostate cancer cell line (10), etc. Despite this work and a panoply of candidate cancer biomarkers, few have been approved for, or implemented in, clinical practice (11–13). The challenge to the clinical and cancer proteomics communities is in overcoming the complications of single-molecule markers in favor of discriminatory multiplex markers. These will benefit from the combinatorial promise of high sensitivity, specificity, reproducibility, and robustness to physiologic variability.

Concomitant to this goal, our primary interest lies in profiling the differential proteomic alterations between stages of prostate cancer initially characterized by antibody-based protein expression and post-translational modifications. The possibilities for such a compendium stratified by disease stage are exciting for mining new biomarkers and molecular networks and generating hypotheses in understanding the possible molecular commonality in the biology of cancer progression.

An Analytic Framework

Interrogative efforts in both oncoproteomics and the cancer transcriptome using high-throughput experimental platforms and their highly dimensional data ushered in a 'systems' era that necessitates integrated approaches to analysis. A burgeoning need to extract richer molecular insight than can be provided by a single experimental platform or data type is driving integration of data from multiple sources in larger multivariate studies. This wider systems biology view proffers a dense range of possibilities and models, be it mathematical or molecular, of fundamental disease biology.

Examples abound of activity in integrating experimental data. These generally adhere to one of three nonexclusive models. The first is the integration of various data available in the public domain that does not necessarily include study-derived experimental data. These are often characterized by data-mining efforts and substantial bioinformatics as well as leveraging common standards of data representation and portability for which the community is generating significant momentum. The second model of integration is the supplementary model, where experimental data are generated and buttressed with other heterogeneous public data sources. Studies in *Saccharomyces cerevisiae* include gene function inference (14) and, importantly

Requests for reprints: Arul M. Chinnaiyan, Department of Pathology, University of Michigan Medical School, 1301 Catherine Road, MSI, Room 4337, Ann Arbor, MI 48109-0602. E-mail: arul@med.umich.edu.

©2006 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-06-0415

in the context of human cancer, computational and analytic methodologies for cancer signatures (7, 15). The third model for data integration enjoying significant activity is integrating multifaceted study-derived genomic and proteomic data in both mammalian and nonmammalian systems. This has showed varying success but profound power. One recurrent observation in this context is of relatively discordant expression between steady-state analyses of mRNA transcript levels and their cognate proteins. This is observed again in *S. cerevisiae*, with work on metabolic pathway perturbations (16), as well as in human cancers, including lung adenocarcinomas (17). Early efforts applying this third model seeded the genesis of a fully integrative framework for analysis of metastatic progression in prostate cancer (Fig. 1; ref. 18).

Power and Universality in Cancer Progression

The aforementioned study used direct proteomic analysis on pooled tissue extracts of grossly dissected tumor specimens from distinct patients to generate an initial estimation of the prostate cancer proteome. This included benign, clinically localized, and metastatic samples and yielded qualitative levels of hundreds of proteins and post-translational modifications. This analysis used an antibody-based high-throughput immunoblot approach. It confidently recapitulated known alterations as well as character-

ized novel associations. The study executed multiple validation steps that included traditional immunoblot on individual extracts for a subset of alterations as well as tissue microarray-based immunohistochemical staining to distinguish between stromally and epithelially expressed proteins. The result is a compendium of differential proteomic alterations that characterize prostate cancer progression from benign to clinically localized and ultimately to metastatic disease.

Correlation analysis with the prostate cancer transcriptome interrogated concordance with the qualitative proteomic alterations assessed by the high-throughput immunoblot assay. This correlation included both publicly available gene expression studies and internal profiling of the pooled and individual samples on Affymetrix (Santa Clara, CA) microarrays as validation. Mapping between genes and the differential counts of dysregulated proteins between clinically localized prostate cancer relative to benign and metastatic relative to clinically localized allowed for expression meta-analysis by standardized statistical inference procedures. Concordance in this context represents similar qualitative differential status in most microarray studies between mRNA transcript and proteomic alteration. The study witnessed only weak concordance, which is similar to results reported previously (17).

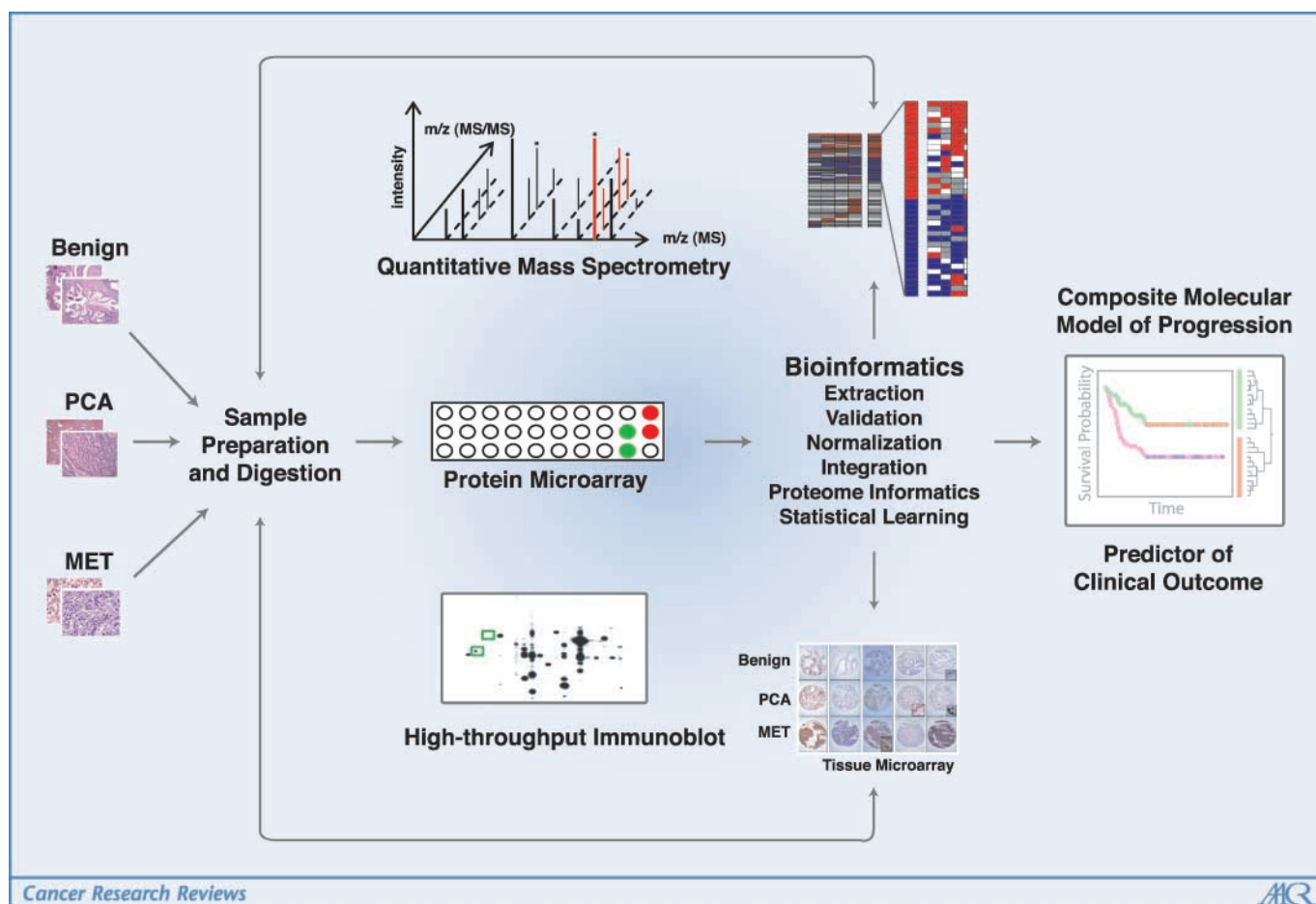


Figure 1. An integrative framework for the molecular profiling and analysis of cancer tissue. Proteomic profiles of benign, clinically localized, and metastatic tissues are generated from high-throughput experimental platforms, immunoblotting in the cited study, but intuitively extended to protein microarray platforms and quantitative mass spectrometry. These are integrated with transcriptomic data to generate a concordant signature of disease progression, which serves as a predictor of clinical outcome and may represent fundamental disease biology.

The study also conducted an enrichment analysis for the androgen hallmarks of refractory disease, as it focused on clinically localized to metastatic prostate cancer as a representation of disease progression. The loss of androgen dependence characterizes the transition from hormone-sensitive to refractory prostate cancer. There was statistically significant enrichment for androgen-regulated proteomic alterations; that is, a significant count of alterations correlated with genes that are functionally induced *in vivo* in response to androgen. Critically, this focus on the later transition in prostate cancer progression identified a 50-gene concordant signature of those entities overexpressed in metastatic samples at the mRNA and protein level. This ensemble contained individually known targets of invasive disease, including the transcriptional repressor EZH2 (3). This supports the intuitive conclusion that more aggressive genes are expressed in progressive prostate cancer and the hypothesis that this 50-gene ensemble may serve as a multiplex signature of prostate cancer progression when applied to primary disease.

Our group chose two prostate cancer gene expression studies to test this 50-gene ensemble (19, 20). These were selected as they were already annotated with recurrence information and met a level of analytic rigor in sample count and type I and II errors. One study was used to build a prediction model on which Kaplan-Meier survival analysis on clusters from unsupervised hierarchical clustering showed a significant clinical outcome distinction in time to recurrence. We used the second data set as an independent test of validity of the concordant signature. In a second Kaplan-Meier survival analysis on class predictions derived from the signature, it showed a significant difference in survival based on risk stratification from a k-nearest neighbor model developed using the first study as training. The results are equally compelling in a recapitulation of the analysis after reversing the data sets.

In a multivariate Cox proportional hazards regression analysis, the 50-gene concordant signature was a superior predictor of prostate cancer recurrence in a model that included several standard clinical variables. Additionally, a univariate model on an iteratively analyzed ranked gene set refined the signature of progression to nine genes that demarcated two primary clusters that differed significantly in Kaplan-Meier survival analysis. Taken together, these results suggested that mRNA transcripts that correlate with protein levels in metastatic disease could be used as

a multiplex gene predictor of progression in clinically localized prostate cancer. Furthermore, in a test of the universality of such a predictive ensemble, this can be extrapolated to other solid tumor types. Primary tumors of the breast, lung, and gliomas bearing the concordant signature were more likely to advance to metastasis than those in which it was lacking. Thus, the predictor may represent inherent biology in disease progression given the homogenous sample type from which it was derived and its use in other solid tumor types.

Perspective

Building a signature of progression from a homogenous cancer type that is generally extensible to other clinically localized solid tumors represents a unique approach with which to characterize general cancer progression. Furthermore, the signature suggests common molecular machinery in poorly differentiated aggressive neoplasms and is predictive of clinical outcome at the clinically localized level. This is a powerful validation of the integrative approach showing clinical import. This integration of high-throughput data can be extended to other proteomic platforms, whether they are expression from quantitative mass spectrometry, protein-protein interaction, or post-translational modification screens on newer array-based platforms (Fig. 1). Although the computational and statistical complexities of such studies are significant, it is an entirely tractable problem and the appropriate partitioning and then integration of such data may be the best model to absorb the heterogeneity of cancer progression. Most critically, the implications for a predictive model of disease progression are significant at the bedside. Tailoring therapeutic intervention from a priori knowledge of disease course is a powerful clinical tool that may result in improved outcome and reduced recurrence. Further, the marriage of such a model with those in early detection has the potential to manifest in overt survival benefit for patients.

Acknowledgments

Received 2/1/2006; revised 3/5/2006; accepted 3/14/2006.

Grant support: Department of Defense grant PC051081 (A.M. Chinnaiyan and S. Varambally), American Cancer Society grant RSG-02-179-MGO (A.M. Chinnaiyan), NIH Prostate Specialized Program of Research Excellence grant P50CA69568 (A.M. Chinnaiyan and S. Varambally) and Early Detection Research Network UO1 grant CA111275-01 (A.M. Chinnaiyan), and University of Michigan Cancer Center (UMCC) Support Grant 5P30 CA46592 funding of the UMCC Bioinformatics Core.

References

- Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. *CA Cancer J Clin* 2005;55:10–30.
- Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 2005;310:644–8.
- Varambally S, Dhanasekaran SM, Zhou M, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002;419:624–9.
- Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. *Science* 2004;306:640–3.
- Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.
- Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;33:49–54.
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM. Mining for regulatory programs in the cancer transcriptome. *Nat Genet* 2005;37:579–83.
- Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198–207.
- Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002;62:3609–14.
- Martin DB, Gifford DR, Wright ME, et al. Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. *Cancer Res* 2004;64:347–55.
- Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis, and treatment selection. *Nat Rev Cancer* 2005;5:845–56.
- Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA. Clinical proteomics: translating benchside promise into bedside reality. *Nat Rev Drug Discov* 2002;1:683–95.
- Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004;4:309–14.
- Troyanskaya OG, Dolinski K, Owen AB, Altman RB, Botstein D. A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc Natl Acad Sci U S A* 2003;100:8348–53.
- Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome. *Nat Genet* 2005;37 Suppl: S31–7.
- Ideker T, Thorsson V, Ranish JA, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 2001;292:929–34.
- Chen G, Gharib TG, Huang CC, et al. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* 2002;1:304–13.
- Varambally S, Yu J, Laxman B, et al. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* 2005;8:393–406.
- Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest* 2004; 113:913–23.
- Yu YP, Landsittel D, Jing L, et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J Clin Oncol* 2004;22:2790–9.