Published in final edited form as:

Stat Biosci. 2016 June ; 8(1): 1–7. doi:10.1007/s12561-014-9111-7.

Strategies for Genomic and Proteomic Profiling of Cancers

Gilbert S. Omenn

University of Michigan, Ann Arbor, MI 48109-2218, USA

Gilbert S. Omenn: gomenn@umich.edu

Abstract

Omics-based technology platforms have made new kinds of cancer profiling tests feasible. There are several valuable examples in clinical practice, and many more under development. A concerted, transparent process of discovery with lock-down of candidate assays and classifiers and clear specification of intended clinical use is essential. The Institute of Medicine has now proposed a three-stage scheme of confirming and validating analytical findings, validating performance on clinical specimens, and demonstrating explicit clinical utility for an approvable test (Micheel et al., Evolution of translational omics: lessons learned and path forward, 2012).

Keywords

Cancer proteomics; Biomarker discovery; Bioinformatics; Omics-based clinical assays; Splice isoforms; Her2/neu breast cancer

1 Introduction

Nature and Science on 15–16 February 2001 published special issues with the nearly-complete Human Genome Sequence from the competing public sector and private sector sequencing projects, along with numerous articles of annotation [3,16]. The ensuing decade has been rich with progress in genomics and biomedical science, more broadly. Translation to targeted therapies, better diagnostics, and prognostic tests has been much slower than projected. As shown in Fig. 1, integration of multi-level omics measurements and behavioral and environmental variables is essential to truly link genotypes and phenotypes.

In 1965 President Lyndon Johnson swooped down on the NIH campus in Bethesda, Maryland, by helicopter and called upon the research community to “unlock the secrets of the laboratory”; he reminded us that the Nation is impatiently waiting for results that yield better medicines and better health. Now NIH Director Francis Collins similarly is trying to accelerate translation with a new organization, the National Center for Advancing Translational Science [1].

Molecular analyses of cancers can reveal information about the mechanisms of initiation, progression, and metastasis and provide the foundation for clinical tests. Mutations and copy number (amplification or deletion) variation are key genomic measures, as in the

glioblastoma The Cancer Genome Atlas (TCGA) project [17]. Biomarkers and credible omics-based tests must meet rigorous standards in discovery, confirmation, analytical validation, clinical utility, clinical decision making, and FDA and reimbursement review, as laid out in a comprehensive report from the Institute of Medicine [7] (Fig. 2). Cancers are particularly challenging because they are very heterogeneous in causation, progression, response to therapies, and risks of metastasis and death.

2 Prostate Cancers

2.1 Gene Expression (Transcriptomic) Tests

Utilizing extensive datasets and embedded analytical tools in OncoPrint, Chinnaiyan and associates revealed a previously unsuspected causal pathway in the majority of prostate cancers, the fusion of the androgen-induced serine protease TMPRSS2 gene to one or another gene of the ETS transcription factor family [14]. The result is to put ERG or other transcription factor under androgen drive, leading to proliferation of prostate cancer cells. Bioinformatics was critical to this discovery, with an outlier analysis stimulating all kinds of confirmatory laboratory experiments. This was the first demonstration of highly recurrent gene fusion in solid tumors, a phenomenon thought to be restricted to leukemias, lymphoma, and sarcomas. Additional pathways critical to prostate cancers now include SPINK 1 and B-Raf, which can be treated with already approved targeted therapies that inhibit EGFR and Raf kinase, respectively. Instead of always looking for new drugs, we can seek out the patients for whom targeted drugs are already available.

2.2 Metabolomic Tests

The overriding patient-centric question about prostate cancers is, how can we tell apart the 10–15 % of cancers that will become invasive, metastatic, and lethal from the relatively innocuous cancer that look the same by histology and clinical staging when first diagnosed? This question about prognosis is much more important than earlier detection, especially with prostate cancers—and with ductal carcinoma in situ of the breast, as well.

Sreekumar et al. [12] combined transcriptomic and metabolomic analyses with molecular concept mapping to uncover a previously unsuspected role of amino acid and methylation pathways and to link the two through the methylated amino acid methylglycine (sarcosine). Sarcosine is barely detectable in normal prostate or prostate with benign hyperplasia; it is detectable but at a low level in localized prostate cancers; its level is greatly increased in metastatic prostate cancers. So we asked whether this is a statistical correlation (a biomarker candidate) or a potential biomediator. Experiments to increase sarcosine concentration in slowly growing prostate cancer cells by adding sarcosine or the precursor glycine to the medium or transfecting with N-methyltransferase gene or knocking down the degradation enzyme all transformed these cells into migratory, invasive cells that metastasized when put into nude mice. Conversely, rapidly growing, invasive prostate cancer cells were made innocuous by increasing degradation or inhibiting biosynthesis of sarcosine. Thus, sarcosine is an outstanding candidate as a mediator. Assay of this metabolite (and a panel with others) has been licensed to Metabolon Inc. and is under development for clinical use.

2.3 Breast Cancers: Transcriptome Profiling to Guide Use of Adjuvant Chemotherapy

A common clinical question with apparently localized breast cancers is whether adjuvant systemic (toxic) chemotherapy should be used on top of surgery and radiation for the tumor. Mammprint is an FDA-approved 70-gene transcriptomic signature assay for prognosis in breast cancer patients, specifically developed to distinguish women with node-negative, stage I/II breast cancer at age <55 who have metastases within 10 years from those who do not [15]. Those with low risk scores can be advised to forego the chemotherapy. A large outcome study of women with intermediate scores and pathology features is ongoing.

Much more widely used is the test OncotypeDX with 21 genes to produce a risk score for metastases and decision on whether to use chemotherapy or estrogen modulation in ER+, node-negative breast cancer patients. Again, ongoing studies are attempting to refine the predictive value of intermediate risk scores.

2.4 Lung Cancers: Comparing Classifiers for Transcriptome Data

Shedden et al. [11] compared eight newly-developed prognostic models on top of clinical variables in a multi-site, blinded, retrospective study of 442 early-stage lung cancer patients. Only two of the classifiers gave useful results and only one added value to the clinical-pathological information (high score/poor survival).

2.5 Proteomics-Based Cancer Profiling

Cancer protein profiling can be based on blood (plasma) or on tumor specimens. The aims are more specific diagnosis, stratification of patients for treatment or follow-up, probability of response from specific therapies, and probability of recurrence or metastasis. Profiling from plasma has proved quite difficult. Reasons are that cancers are highly heterogeneous, tumor proteins are greatly diluted upon release to the extracellular fluid and blood, the proteins are generally in low abundance especially when tumors are early (small), and plasma is an extraordinarily complex specimen dominated by high abundance-proteins. About 2,000 proteins have now been identified with high confidence in human plasma, as documented in the Human Plasma PeptideAtlas [2].

2.6 Lung Cancers: Aptamer-Based Proteomics Profiling of Plasma

Ostroff et al. [10] developed a large library of aptamer-based protein capture reagents that could identify about 1,100 proteins in plasma, then produced a 44-biomarker and finally a 12-biomarker panel of relevant proteins to distinguish non-small cell lung cancer patients from controls. They reported an area-under-the curve of 0.90 with Bayesian methods. Further clinical validation is ongoing.

2.7 Ovarian Cancers: A Panel of Abundant Plasma Proteins that can Guide Surgical Plans for Patients with Pelvic Mass

Based on empirical results from mass spectrometry-based proteomics, not tied to cancer pathways or mechanisms, Chan and colleagues at Johns Hopkins developed the FDA-approved OVA1 test [18]. Elevated levels of beta-2 microglobulin and CA125 and decreased levels of transthyretin, apolipoprotein A1, and transferrin are combined to predict ovarian

cancer in women who have pelvic masses (80–90 % benign) with sensitivity 0.92 and specificity 0.45; this is not a screening test, but a test to guide whether a general surgeon or a gynecologic surgeon is more likely to be appropriate for definitive surgery.

2.8 A New Class of Biomarker Candidates: Alternative Splice Variant Proteins and Transcripts

One of the great surprises of the Human Genome Project is that 20,000 genes are sufficient to code for all the proteins. Earlier estimates ranged upward from 50,000 or 100,000. These genes account for only 1.5 % of the genome sequence. Complexity has evolved a process of splicing of messenger RNA transcripts, with roughly five or more such variants per gene; what was first thought to be unusual now seems to involve more than 90 % of protein-coding genes. Our group at the University of Michigan has developed an analytical workflow combining Ensembl and ECGene databases, with regular updates, to produce a modified ECGene database with all potential translation products (10M sequences in the mouse; 14M sequences in the human) for matching of peptide sequences detected by mass spectrometry. Menon et al. [4] have published a series of studies identifying splice variants of high biological interest in the KrasG12D activation/Ink4a/Arf double deletion mouse model of human pancreatic cancers and the Her2/neu amplification model of that subtype of breast cancers [5,9]. We reported 420 splice variants (92 novel) in pancreatic cancer, and 608 (68 novel) in the breast cancer. We have mapped these variants to protein–protein interactions with the Michigan Molecular Interactions database (MiMI) [13] and visualized the network with Cytoscape and the MiMI plug-in [4]. Studies of human prostate, breast, and colon cancer datasets are ongoing. Many different splicing mechanisms were identified, and the expected mRNA alternative splice transcripts were demonstrated in tumor specimens with qRT-PCR. For studies that begin with RNA sequencing, it is necessary to perform protein assays to show that the transcripts are, indeed, translated to proteins (Table 1).

A special feature of this line of research is our combining proteomics with computational methods to predict and compare the protein 3-D conformation of pairs of splice variants and deduce the likely functional consequences of the variant. We used I-TASSER and M-ALIGN from the Yang Zhang Lab at the University of Michigan to characterize several pairs of interesting splice variants [6]. For examples, an exon 9/exon 10 swap alters allosteric inhibition and ATP binding of pyruvate kinase M2 and triggers the aerobic glycolysis (Warburg effect) characteristic of rapidly growing cancers; and the difference of six amino acids between variants of annexin-6 exposes a Ser–Pro–Thr sequence to phosphorylation which contributes to greater proliferation of breast cancer cells.

3 Conclusion

There are already several good examples of clinically-useful omics-based tests. There are much larger numbers of promising discovery phase biomarker candidates from genomic, transcriptomic, epigenomic, proteomic, and metabolomic studies of cancers and cancer models. There have been rapid advances in proteomics technologies and associated bioinformatics methods, including orbitrap-based mass spectrometry, triple-quad mass spectrometry for multiple-reaction monitoring of targeted proteins, glycan and N-glycosite

analyses of glycoproteins, better recognition of posttranslational modifications, and detection and characterization of alternative splice variants. A highly integrated, rigorous test development pathway is essential to overcome the inherent overfitting in high-dimensionality assays and move from discovery to confirmation to validation to clinical utility to clinical decisions that help patients.

References

1. Collins FS. Research agenda: opportunities for research and NIH. *Science*. 2010; 327:36–37. [PubMed: 20044560]
2. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, Mallick P, Katz JE, Malmström J, Ossola R, Watts JD, Lin B, Zhang H, Moritz RL, Aebersold R. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics*. 2011; 10:9. doi:10.1074/mcp.M110.006353.
3. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. [PubMed: 11237011]
4. Menon R, Omenn GS. Proteomic characterization of novel alternative splice variant proteins in human epidermal growth factor receptor 2/neu-induced breast cancers. *Cancer Res*. 2010; 70:3440–3449. [PubMed: 20388783]
5. Menon R, Zhang Q, Zhang Y, Fermin D, Bardeesy N, DePinho RA, Lu C, Hanash SM, Omenn GS, States DJ. Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer. *Cancer Res*. 2009; 69:300–309. [PubMed: 19118015]
6. Menon R, Roy A, Mukerjee S, Belkin S, Zhang Y, Omenn GS. Functional implications of structural predictions for alternative splice proteins expressed in Her2/neu-induced breast cancers. *J Proteome Res*. 2011; 10:5503–5511. [PubMed: 22003824]
7. Micheel, C.; Nass, S.; Omenn, GS., editors. *Evolution of translational omics: lessons learned and path forward*. Washington, DC: National Academy Press; 2012.
8. Omenn, GS. Gene-environment interactions: eco-genetics and toxicogenomics. In: Ginsburg, G.; Willard, H., editors. *Genomic and personalized medicine*. 2nd edn. New York: Elsevier; 2012. p. 50-59.
9. Omenn GS, Yocum AK, Menon R. Alternative splice variants, a new class of protein cancer biomarker candidates: findings in pancreatic cancer and breast cancer with systems biology implications. *Dis Mark*. 2010; 28:241–251.
10. Ostroff RM, Bigbee WL, Franklin W, et al. Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS One*. 2010; 5:e15003. doi:10.1371/journal.pone.0015003. [PubMed: 21170350]
11. Shedden K, Taylor JM, Enkemann SA. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med*. 2008; 14:822–827. [PubMed: 18641660]
12. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*. 2009; 457:910–914. [PubMed: 19212411]
13. Tarcea VG, Weymouth T, Ade A, et al. Michigan Molecular Interactions r2: from interacting proteins to pathways. *Nucl Acids Res*. 2009; 37:D642–D646. [PubMed: 18978014]
14. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005; 310:644–648. [PubMed: 16254181]
15. Van de Vijver MJ, He YD, van't Veer LJ. A gene-expression signature as a predictor of survival in breast cancer. *NEJM*. 2002; 347:1999–2009. [PubMed: 12490681]
16. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001; 291:1304–1351. [PubMed: 11181995]

17. Wang XV, Verhaak RG, Purdom E, Spellman PT, Speed TP. Unifying gene expression measures from multiple platforms using factor analysis. *PLoS One*. 2011; 6:e17691. [PubMed: 21436879]
18. Zhang Z, Chan DW. The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiol Biomark Prev*. 2010; 19:2995–2999.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Integrating High-Throughput Measurements with the Phenotype: from Science to Medicine

[GS Omenn & B Athey, NCIBI, 2010]

Fig. 1. Schema for genome to phenome and phenome to genome through multi-dimensional omics analyses and interactions with behavioral and environmental variables [8]

Fig. 2.

An integrated approach to omics-based clinical tests. This schema from the Institute of Medicine shows the key attributes of the discovery phase, the test validation phase, and the evaluation of clinical utility [7]

Table 1

Proteomics-based biomarker discovery strategies

-
1. Start with microarray or next-gen sequencing evidence for carcinogenic pathway mechanisms in tumor and then track corresponding proteins in biofluids (e.g., TMPRSS2/ETS and sarcosine in prostate cancers)
 2. Perform targeted proteomics with SRM/MRM mass spectrometry to quantitate candidate proteins
 3. Detect auto-antibodies in plasma as a biological amplification of tumor protein signals
 4. Identify alternative splice variants of biologically relevant proteins in cancers or in plasma of patients and in genetically-matched mouse models (see below)
-