Published in final edited form as:

Clin Exp Metastasis. 2012 October ; 29(7): 645–652. doi:10.1007/s10585-012-9507-z.

Systematic, Evidence-Based Discovery of Biomarkers at the NCI

John R. Prensner, Ph.D.¹, Arul M. Chinnaiyan, M.D., Ph.D.¹, and Sudhir Srivastava, Ph.D., MPH²

¹Michigan Center for Translational Pathology, University of Michigan Medical School, 1400 E. Medical Center Drive, Ann Arbor, MI 48109

²Cancer Biomarkers Research Group, Division of Cancer Prevention, National Cancer Institute, 6130 Executive Boulevard, Suite 3142, Rockville, MD 20852

Abstract

In the past decade, biomarker discovery has become ubiquitous in cancer research. However, despite this interest in biomarker research, few newly-characterized biomarkers have emerged as clinically-used entities. Here, we review the current state of biomarker research in cancer and identify challenges that stall many biomarker discovery efforts. We outline a model for systematic biomarker discovery, exemplified by recent efforts in prostate cancer, in which bioinformatics plays a central role in identifying promising new candidate biomarkers. Finally, we review the role of the National Cancer Institute's Early Detection Research Network (EDRN) in biomarker studies and the importance of EDRN-led efforts to establish a research standard for more effective biomarker discovery efforts.

Keywords

biomarker; prostate cancer; bioinformatics; early detection

Introduction

The past decade has witnessed a great surge in biomarker research, with thousands of research articles nominating an ever-increasing number of putative cancer biomarkers. Indeed, in 2010 over 20,000 papers investigated biomarkers, including over 8,000 biomarkers for cancer and about 600 that were suggested as useful for early detection. Such a tremendous research output has fueled expectations that effective biomarker-based diagnostics would rapidly take form. Yet, this has not been the case, and only a handful of newly-characterized biomarkers have been approved by the Food and Drug Administration (FDA) [1]. Here, we review general problems with biomarker research that have led to this low success rate and highlight promising biomarkers emerging in the field of prostate cancer.

²Correspondence to: Sudhir Srivastava, Ph.D., MPH, Cancer Biomarkers Research Group, Division of Cancer Prevention, National Cancer Institute, 6130 Executive Boulevard, Suite 3142, Rockville, MD 20852, Phone: 301-435-1594, Fax: 301-402-8990, SS1a@nih.gov.

Conflict of Interest Disclosure

A.M.C. serves as an advisor to Gen-Probe, Inc., who has developed diagnostic tests using PCA3 and TMPRSS2-ERG. The University of Michigan has licensed the development of TMPRSS2-ERG-based prostate cancer diagnostic assays to Gen-Probe and A.M.C. is named as a co-inventor. Gen-Probe was not involved in the writing or approval of this manuscript.

Problems in current biomarker research

The central goal of biomarker research is to elucidate a molecule whose measurement provides information about a patient's disease or risk of disease [2]. The clinical application of biomarkers is, therefore, of paramount importance if a biomarker is to be effectively translated into a clinical setting. However, defining a clinical setting is often neglected for many biomarker candidates, thereby complicating clinical translation and making independent validation of candidate biomarkers difficult [3]. For example, merely discovering a biomarker for cancer versus normal patients is not adequate without defining several key parameters: how is the biomarker measured in human patients (tissue/serum/urine, etc), what patient population receives the test, and how is the test interpreted? These issues plague many previous studies, limiting the impact of biomarker discoveries that are not aligned with the intended clinical use and leading to poorly-designed biomarker studies that commonly fail [4,5].

Another challenge of biomarker research has been the role of serendipity in biomarker discovery. Use of convenience samples in discovery continues to complicate the generalizability of biomarkers for a specific use since these biomarkers suffer from being discovered by chance and are plagued by bias, leading to low rates of validation by independent groups [6]. In part, the continued use of non-human sources, such as cell lines and animal models, further complicates biomarker discovery by introducing biomarkers not reproducible in humans.

Although it is encouraging that, in the last 3 years, the number of approved biomarkers for cancer has increased modestly, progress is proceeding slowly if one looks at the proportionality of approved biomarkers compared to those published each year [7]. Due to perceived underachievement of long-term returns on investment, many private industry research programs avoid biomarker research—a stark contrast to drug discovery efforts. This attitude is in accordance the generally low reimbursement for diagnostics when compared to drugs [7]. Taken together, there is a need for systematic discovery and validation of biomarkers both from public and private settings.

Organizational efforts on biomarkers

The process of biomarker discovery ought to be no different than drug discovery. It requires sustained, long-term efforts starting from basic research all the way to FDA filing. It is estimated that it costs more than one billion dollars to find one effective cancer drug. Biomarkers for early detection may be even harder because they entail finding the disease among asymptomatic populations. The cost of developing robust diagnostic FDA-approved biomarkers is speculative because there is a lack of published historical data.

In light of these challenges and difficulties, NCI established a program called Early Detection Research Network (EDRN) with the goal of developing and evaluating biomarkers for cancer risk, early detection, diagnosis and prognosis [8]. EDRN promotes a vertical approach to conducting biomarker research, whereby biomarkers are developed in Biomarker Developmental Laboratories (BDLs), refined and cross validated by Biomarker Reference Laboratories (BRLs), and validated in collaboration with Clinical Validation Centers (CVCs), all within one organization [1]. The focus is on coordinating multiple resources with a goal of minimizing the barriers to the rapid and efficient transfer of biomarker developments between entities.

The process of biomarker discovery is guided by a statistical five-phase criterion, termed a prospective-specimen-collection, retrospective-blinded-evaluation (PROBE) design (Table 1) [9–11]. This five-phase approach has established both a scientific standard and a roadmap

for successfully translating biomarker research from the laboratory to the clinic. Phase 1 is devoted to discovery and involves exploratory study to identify potentially useful biomarkers. Phase 2 refers to validation where biomarkers are tested to determine their capacity for distinguishing between people with cancer and those without. Phase 3 determines the capacity of a biomarker to detect preclinical disease by testing the marker against tissues collected longitudinally from research cohorts. Phase 4 includes prospective screening studies on biomarker performance in large populations, and determines the false referral rate. Phase 5 refers to the penultimate period in which large-scale population studies evaluate both the role of the biomarker for cancer detection, and its overall screening impact. EDRN has been charged with conducting studies up to Phase 3 (Table 1). Studies conducted beyond Phase 3 require collaborations with the clinical trial community which may have appropriate cohorts and samples for validating biomarkers per our guidelines [9,10].

EDRN uses these guidelines for triaging biomarkers and distinguishing useful biomarkers from those which do not show clinical utility [1]. When judged using the PROBE criteria, it becomes apparent that the vast majority of candidate biomarkers found in the published literature are in Phase 1, lacking the validation or downstream development and refinement necessary for clinical translation of a biomarker discovery effort [6]. While these candidate biomarkers may reflect promising research findings, the issue then arises on how to select individual biomarkers for further evaluation, because biomarker development is a long, labor-intensive process not suitable for most candidates. In light of this, how do we determine which biomarkers are most promising for clinical translation? And at what point in the biomarker development process should this selection occur?

Although there are no definitive answers to these questions, the EDRN has provided the first step by establishing a cohort of clinical reference samples (CRS) to facilitate the initial stage of biomarker validation. The use of CRS for independent validation efforts can be used to determine a “go” or a “no go” decision. In addition, considerations about the method of biomarker evaluation in patients also impact a “go” or “no go” decision, as tissue-based biomarkers often require a biopsy or other invasive procedure. Biomarkers measured in serum or urine therefore offer a less invasive (and often preferable) alternative to tissue-based biomarkers.

Development of a biomarker discovery pipeline

The goal of the EDRN is twofold: (1) to discover novel biomarkers, and (2) to evaluate the efficacy and appropriateness of existing biomarkers. The example of prostate cancer here is instructive. Widespread screening of asymptomatic men for a serum prostate specific antigen (PSA) level has dramatically increased the incidence of prostate cancer in Western populations following its adoption in the early 1990s [12–14]. But PSA is a flawed biomarker, plagued by a low specificity and positive predictive value [15,16]. In fact, typically 66 – 75% of men with an elevated PSA (>4ng/mL) do not have prostate cancer [15]. To help improve prostate cancer screening efforts, EDRN investigators identified about 108 biomarkers that could be potentially tested to serve as an adjunct to PSA (S.S., unpublished data). Out of 108 they found about 58 of them which use reproducible tests which have reproducible data (S.S., unpublished data). After critical reviews of these 58, only five were selected for further validation, indicating that very few biomarkers exhibit promise for clinical translation (S.S., unpublished data).

The EDRN, therefore, emphasizes the need for systematic, rational discovery of biomarkers using a well-design research design and set of clinical samples. Moreover, biomarkers with well-characterized biological roles in tumorigenesis are of particular interest, as these

represent functional components of cancer biology. Candidate biomarkers identified through the use of convenience samples should be approached with caution, as success in biomarker development results from systematic approaches that provide informative data based on sound clinical questions, address clinical questions and plan to implement the clinical needs. A clinical validation study of new biomarkers should be entertained only if likely to provide significant enhancement over the current standard of care; in other words, a poorly designed clinical study is worse than no study at all. Furthermore, identification of novel and relevant biomarkers should be sought by prospectively designing clinical studies with that purpose rather at the forefront rather than piggybacking ongoing studies. Here, we will review the discovery of novel prostate cancer biomarkers of an example of systematic, rational biomarker development.

Bioinformatics-driven biomarker discovery in prostate cancer

Over the past decade, emerging biomarkers in prostate cancer have demonstrated the power of rational and systematic nomination of biomarker candidates. Our research group the University of Michigan has employed bioinformatics-driven analyses of prostate cancer RNA to discover novel components of prostate cancer biology [17]. By integrating large-scale expression profiling datasets of human prostate cancer, we have defined a set of genes with dramatic overexpression in a subset of cases (outliers) that are often overlooked by conventional overexpression or copy-number analyses. Such efforts facilitated the discovery of common chromosomal translocation of the ETS family of transcription factors (ERG, ETV1, ETV4, ETV5), which occur in approximately 50% of patients [18,19]. These translocations result in high expression of ETS factors by generating a gene fusion with the ETS factor under the control of androgen-responsive promoters such as TMPRSS2 [18,20]. The most common of these gene fusions, TMPRSS2-ERG, accounts for ~90% of all ETS fusions found in prostate cancer (Fig. 1A and 1B) [17,20], and *in vitro* studies have demonstrated that induction of AR signaling in prostate cancer cells results in genomic colocalization of the TMPRSS2 and ERG genes, which thereby facilitates the formation of the gene fusion when in the presence of DNA damage (Fig. 1A) [21].

Because the TMPRSS2-ERG gene fusion is both tissue-specific and cancer-specific, it represents a promising biomarker for prostate cancer. Indeed, these chromosomal rearrangements offer distinct advantages over the current biomarkers for prostate cancer, such as PSA, because their detection is indicative of cancer, unlike PSA, and their functional role in promoting prostate cancer progression is well-characterized. Currently, detection of ERG fusions (by fluorescence *in situ* hybridization), ERG protein (by IHC), or TMPRSS2-ERG mRNA in urine can all be used as diagnostic tools for prostate cancer [22–24].

Yet, 50% of prostate cancers do not harbor an ETS gene fusion, meaning that these fusions lack sensitivity as a screening assay. Using similar bioinformatics analyses of ETS-negative prostate cancers led to the identification of the serine protease SPINK1 as an outlier in ~20% of ETS-negative cancers (10% of all prostate cancers) (Fig. 1B) [25]. SPINK1 expression was also observed as an independent predictor of biochemical recurrence after resection [25], and *in vitro* studies demonstrated a role for SPINK1 in tumor cell invasion [25,26]. Similarly rare fusions of the RAF family kinases have also been found in 1–2% of patients (Fig. 1B) [27]. These findings suggest that the systematic nomination of biomarkers can identify complementary biomarkers that also represent aspects of disease biology.

Developing panels of biomarkers for prostate cancer

Because individually outliers such as TMPRSS2-ERG or SPINK1 lack sensitivity, their ideal clinical application is as part of a multiplexed assay that provides information on multiple biomarkers [22]. Detection of outlier RNA in urine sediments is a particularly

promising avenue for biomarker development because acquisition of urine sediments does not require an invasive biopsy procedure.

To this end, multiple groups have investigated multiplexed panels of urine biomarkers for prostate cancer diagnosis in PSA-prescreened cohorts. These studies have cumulatively defined an emerging role for two urine biomarkers, TMPRSS2-ERG and PCA3, whose detection improves the ability to distinguish men with prostate cancer and men without [23,28–30]. PCA3 is a long noncoding RNA (lncRNA) whose 3.7kb RNA product is both spliced and polyadenylated, yet does not encode a protein (Fig. 1C) [31]. lncRNAs have been recently shown to contribute to prostate cancer pathogenesis [32] and are emerging as a poorly understood layer of cancer biology [31]. Although its function in prostate cancer cells is unclear, PCA3 ranks as one of the best prostate cancer biomarkers in tissue-based studies, with >90% of prostate cancers exhibiting PCA3 overexpression (Fig. 1C) [33,34]. Using a first-generation assay in a laboratory setting, retrospective analysis of urine PCA3 values show a sensitivity of 54% and a specificity of 74% [35], and the combination of PCA3 with TMPRSS2-ERG improves the area-under-the-curve (AUC) metric of the receiver-operating-characteristic (ROC) curve over each individually [22].

To translate these findings into clinical practice, commercial CLIA-grade urine assays have been developed for PCA3 and TMPRSS2-ERG (Table 2). Tomlins et al. recently used these assays to conduct a large-scale analysis of PCA3 and TMPRSS2-ERG in PSA-prescreened populations at multiple community clinic settings as well as academic institutions [36]. In both the academic and community cohorts, a combined urine measurement of TMPRSS2-ERG and PCA3 outperformed PSA in distinguishing patients with biopsy-proven cancer (the AUC for TMPRSS2-ERG + PCA3 was between 0.71 – 0.77 and the AUC for PSA was between 0.60 – 0.61) [36]. Of note, TMPRSS2-ERG and PCA3 also provided additional improvement over the Prostate Cancer Prevention Trial (PCPT) risk calculator for prostate cancer, which integrates serum PSA level, age, patient and family history to predict risk for prostate cancer [36]. While the prognostic significance of PCA3 and TMPRSS2-ERG individually has debated in conflicting studies [37], Tomlins et al. found that the combined TMPRSS2-ERG + PCA3 measurement associated with risk of high-grade cancer, and an increased TMPRSS2-ERG score was individually correlated with the number of biopsy cores with cancer and maximum tumor dimension, but not PSA level or ultrasound volume [36]. EDRN investigators are similarly conducting a multi-institutional validation study testing the utility of a PCA3 urine test to assist clinicians in decision making for initial biopsy or repeat biopsy.

While TMPRSS2-ERG and PCA3 are promising biomarkers, they have limitations as well. First, the measurement of TMPRSS2-ERG and PCA3 is determined relative to the urine KLK3 (PSA) mRNA level, and thus a low yield of prostate cells from a urine sediment sample can render the tests uninformative. Second, these assays are more expensive and more complicated to perform, thus requiring specialized personnel to perform the tests.

As a result, the EDRN is also testing biomarkers developed elsewhere that may provide additive value to the panel. New –omic technologies, such as metabolomics and proteomics, have nominated novel biomarkers that have not previously been measured. An example of this is sarcosine, a metabolite produced by cells during the metabolism of the amino acid glycine, which was nominated by systematic analyses of prostate tissue cohorts [38]. Sarcosine overexpression characterizes prostate cancer, particularly metastatic cancers, and it may play a role in cancer aggressiveness in both prostate and lung cancer [38,39]. Development of urine assays for sarcosine is also underway (Table 1) [40], and the EDRN is poised to incorporate promising biomarkers into its programs. We are currently testing a number of biomarkers using the prostate cancer reference sets, which are comprised of

patients samples from patients that are (1) men over 40 years of age, (2) scheduled for prostate biopsy due to an elevated PSA (>2.5 ng/mL) or PSA rising (>0.5 ng/mL/yr), (3) risk factors for prostate cancer (e.g. family history) even in the absence of an elevated PSA, (4) abnormal digital rectal exam (DRE), (5) percent free PSA $<15\%$, (6) no prior history of prostate cancer or prostate biopsy, and (7) prostate biopsy with at least 10 cores taken in a laterally directed fashion. Blood from these patients is collected prior to prostate biopsy.

In addition, the EDRN is also re-evaluating PSA in prostate cancer screening. Despite the long-time use of a 4.0 ng/ml cutoff for a 'normal' PSA value [13], it has been acknowledged that only about 25% of men with such an elevated value will be found to have prostate cancer at prostate biopsy [15] and ~15–20% of men with a PSA <4.0 ng/ml may also have prostate cancer [41]. Because of this, three-quarters of men with an elevated PSA who have a biopsy undergo the procedure unnecessarily. Recent data from large longitudinal screening programs and from the Prostate Cancer Prevention Trial now suggest that the risk of prostate cancer is equally elevated (20–25%) even among men with serum PSA levels from 2.5 ng/ml to 4.0 ng/ml [41]. Additional indications for prostate biopsy include a rising PSA, an abnormal digital rectal examination, or even a lower PSA value for a patient with other risk factors [42]. For example, in accord with the initial demonstration from Hopkins that family history is linked to prostate cancer risk, EDRN group has demonstrated that for a 65 year old man with a first degree relative with prostate cancer, a PSA of 1.8 carries a 25% positive predictive value for prostate cancer (<http://edrn.nci.nih.gov/resources/sample-reference-sets/Prostate%20Ref%20SOP.pdf>). An opportunity exists for a biomarker in this application to reduce the number of unnecessary initial and repeat biopsies in men who are ultimately proven to not have prostate cancer while maintaining a very high level of sensitivity.

EDRN as an accelerator and as a brake

EDRN is an accelerator that drives good biomarkers through the clinic and also as a brake to use good clinical design to eliminate biomarkers without any added value. Two examples of this are [–2]pro-PSA, a precursor molecule of PSA which contains two additional amino acids at the 5' end of the molecule [16,43], and percent free PSA (%fPSA), which represents the relative fraction of serum PSA proteins not complexed with other serum proteins [44,45]. Biochemically, PSA begins as a precursor protein and undergoes several proteolytic cleavage steps to produce the mature PSA protein [43], which can complexes in the blood with alpha-1-antichymotrypsin. In prostate cancer, improper processing and complexing of PSA results in an increase in [–2]pro-PSA [14,46] and a decrease in the %fPSA [44,45]. Changes in [–2]pro-PSA and %fPSA in prostate cancer also associate with more aggressive, higher-grade disease [16,47,48]. Widespread interest in these biomarkers has further spurred development of commercial assays for [–2]pro-PSA and %fPSA (Table 2).

To pursue these biomarkers, the EDRN conducted a multi-center study of [–2]pro-PSA in combination with PSA and fPSA for prostate cancer detection in patients with relatively low PSA levels of 2.0 to 10.0 ng/mL [48]. The objective of the study was to evaluate [–2]pro-PSA, fPSA, and PSA using a mathematical formula (prostate health index [ϕ]) = $([–2]proPSA / fPSA) \times PSA^{1/2}$) to enhance specificity for detecting overall and high-grade prostate cancer. For the 2–10 ng/mL PSA range, at 80–95% sensitivity, the specificity and AUC (0.703) of [–2] ϕ exceeded those of PSA and %fPSA. Increasing ϕ was associated with a 4.7-fold increased risk of prostate cancer and 1.61-fold increased risk of Gleason 7 disease upon biopsy [48]. The AUC for ϕ (0.724) exceeded that of %fPSA (0.670) in discriminating between prostate cancer with Gleason 4+3 vs. lower grade disease or negative biopsies.

Conclusion

The clinical translation of novel biomarkers can transform disease management. However, biomarker discovery and development is fraught with challenges, resulting in the eventual abandonment or failure of the vast majority of candidate biomarkers. The EDNRN has proposed a rational paradigm for biomarker development that focuses on systematic, evidence-based discovery and validation of biomarkers as a prerequisite for further advancement of nominated biomarkers to clinical trials. Biased discovery methods and poorly designed clinical trials compromise many biomarker studies, complicating efforts to independently validate these markers and define their appropriate clinical use. New discoveries of urine RNA biomarkers (TMPRSS2-ERG, PCA3) and novel derivations of serum PSA (fPSA, [-2]pro-PSA) in prostate cancer now promise to change the clinical detection and management of this disease. Together, the combination of multiple biomarkers will offer advantages over each individual biomarker in clinical use. These biomarkers serve as examples demonstrating the potential benefit biomarker development has on clinical oncology.

Acknowledgments

We would like to acknowledge the numerous labs, authors, and publications that we were unable to cite in this review due to space restrictions. This work was supported by the Early Detection Research Network grant U01 CA 11275 (to A.M.C.), the Department of Defense grants PC100171 (to A.M.C.) and PC094290 (to J.R.P.), NIH Prostate Specialized Program of Research Excellence grant P50CA69568 (to A.M.C.). A.M.C. is supported by the Doris Duke Charitable Foundation Clinical Scientist Award, the Prostate Cancer Foundation, the American Cancer Society, and the Howard Hughes Medical Institute. J.R.P. is a Fellow of the University of Michigan Medical Scientist Training Program. A.M.C. is a Taubman Scholar of the University of Michigan.

References

1. Srivastava S. Cancer biomarker discovery and development in gastrointestinal cancers: Early detection research network-a collaborative approach. *Gastrointest Cancer Res.* 2007; 1 Suppl 2(4):S60–S63. [PubMed: 19360150]
2. Sawyers CL. The cancer biomarker problem. *Nature.* 2008; 452(7187):548–552. [PubMed: 18385728]
3. Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: Roles of phases, guidelines, and study design. *J Clin Epidemiol.* 2007; 60(12):1205–1219. [PubMed: 17998073]
4. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat Biotechnol.* 2006; 24(8):971–983. [PubMed: 16900146]
5. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: Design issues. *J Natl Cancer Inst.* 2010; 102(3):152–160. [PubMed: 20075367]
6. Diamandis EP. Cancer biomarkers: Can we turn recent failures into success? *J Natl Cancer Inst.* 2010; 102(19):1462–1467. [PubMed: 20705936]
7. Nass, SJ. Committee on developing biomarker-based tools for cancer screening, diagnosis, and treatment. Washington DC: National Academies Press; 2007.
8. Srivastava S, Rossi SC. Early detection research program at the nci. *Int J Cancer.* 1996; 69(1):35–37. [PubMed: 8600056]
9. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: Standards for study design. *J Natl Cancer Inst.* 2008; 100(20):1432–1438. [PubMed: 18840817]
10. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (remark). *Nat Clin Pract Oncol.* 2005; 2(8): 416–422. [PubMed: 16130938]
11. Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM. Beyond psa: The next generation of prostate cancer biomarkers. *Sci Transl Med.* 2012; 4(127):127rv123.

12. Etzioni R, Cha R, Feuer EJ, Davidov O. Asymptomatic incidence and duration of prostate cancer. *Am J Epidemiol.* 1998; 148(8):775–785. [PubMed: 9786232]
13. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, Petros JA, Andriole GL. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N Engl J Med.* 1991; 324(17):1156–1161. [PubMed: 1707140]
14. Makarov DV, Loeb S, Getzenberg RH, Partin AW. Biomarkers for prostate cancer. *Annu Rev Med.* 2009; 60:139–151. [PubMed: 18947298]
15. Brawer MK. Prostate-specific antigen: Current status. *CA Cancer J Clin.* 1999; 49(5):264–281. [PubMed: 11198954]
16. Stephan C, Jung K, Lein M, Diamandis EP. Psa and other tissue kallikreins for prostate cancer detection. *Eur J Cancer.* 2007; 43(13):1918–1926. [PubMed: 17689069]
17. Prensner JR, Chinnaiyan AM. Oncogenic gene fusions in epithelial carcinomas. *Curr Opin Genet Dev.* 2009; 19(1):82–91. [PubMed: 19233641]
18. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, Yu J, et al. Distinct classes of chromosomal rearrangements create oncogenic ets gene fusions in prostate cancer. *Nature.* 2007; 448(7153):595–599. [PubMed: 17671502]
19. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, et al. Recurrent fusion of *tprss2* and ets transcription factor genes in prostate cancer. *Science.* 2005; 310(5748):644–648. [PubMed: 16254181]
20. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. *Nat Rev Cancer.* 2008; 8(7):497–511. [PubMed: 18563191]
21. Mani RS, Tomlins SA, Callahan K, Ghosh A, Nyati MK, Varambally S, Palanisamy N, Chinnaiyan AM. Induced chromosomal proximity and gene fusions in prostate cancer. *Science.* 2009; 326(5957):1230. [PubMed: 19933109]
22. Laxman B, Morris DS, Yu J, Siddiqui J, Cao J, Mehra R, Lonigro RJ, Tsodikov A, Wei JT, Tomlins SA, Chinnaiyan AM. A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. *Cancer Res.* 2008; 68(3):645–649. [PubMed: 18245462]
23. Hessels D, Smit FP, Verhaegh GW, Witjes JA, Cornel EB, Schalken JA. Detection of *tprss2-erg* fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clin Cancer Res.* 2007; 13(17):5103–5108. [PubMed: 17785564]
24. Park K, Tomlins SA, Mudaliar KM, Chiu YL, Esgueva R, Mehra R, Suleman K, Varambally S, Brenner JC, MacDonald T, Srivastava A, et al. Antibody-based detection of *erg* rearrangement-positive prostate cancer. *Neoplasia.* 2010; 12(7):590–598. [PubMed: 20651988]
25. Tomlins SA, Rhodes DR, Yu J, Varambally S, Mehra R, Perner S, Demichelis F, Helgeson BE, Laxman B, Morris DS, Cao Q, et al. The role of *spink1* in ets rearrangement-negative prostate cancers. *Cancer Cell.* 2008; 13(6):519–528. [PubMed: 18538735]
26. Ateeq B, Tomlins SA, Laxman B, Asangani IA, Cao Q, Cao X, Li Y, Wang X, Feng FY, Pienta KJ, Varambally S, et al. Therapeutic targeting of *spink1*-positive prostate cancer. *Sci Transl Med.* 2011; 3(72):72ra17.
27. Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, Han B, Cao Q, Cao X, Suleman K, Kumar-Sinha C, et al. Rearrangements of the raf kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med.* 2010; 16(7):793–798. [PubMed: 20526349]
28. Haese A, de la Taille A, van Poppel H, Marberger M, Stenzl A, Mulders PF, Huland H, Abbou CC, Remzi M, Tinzl M, Feyerabend S, et al. Clinical utility of the *pca3* urine assay in european men scheduled for repeat biopsy. *Eur Urol.* 2008; 54(5):1081–1088. [PubMed: 18602209]
29. Hessels D, Schalken JA. The use of *pca3* in the diagnosis of prostate cancer. *Nat Rev Urol.* 2009; 6(5):255–261. [PubMed: 19424173]
30. Roobol MJ, Schroder FH, van Leeuwen P, Wolters T, van den Bergh RC, van Leenders GJ, Hessels D. Performance of the prostate cancer antigen 3 (*pca3*) gene and prostate-specific antigen in prescreened men: Exploring the value of *pca3* for a first-line diagnostic test. *Eur Urol.* 2010; 58(4):475–481. [PubMed: 20637539]
31. Prensner JR, Chinnaiyan AM. The emergence of lncrnas in cancer biology. *Cancer Discov.* 2011; 1(5):391–407. [PubMed: 22096659]

32. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, Cao X, et al. Transcriptome sequencing across a prostate cancer cohort identifies pcat-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol*. 2011; 29(8):742–749. [PubMed: 21804560]
33. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB. Dd3: A new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res*. 1999; 59(23):5975–5979. [PubMed: 10606244]
34. de Kok JB, Verhaegh GW, Roelofs RW, Hessels D, Kiemeny LA, Aalders TW, Swinkels DW, Schalken JA. Dd3(pca3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res*. 2002; 62(9):2695–2698. [PubMed: 11980670]
35. Deras IL, Aubin SM, Blase A, Day JR, Koo S, Partin AW, Ellis WJ, Marks LS, Fradet Y, Rittenhouse H, Groskopf J. Pca3: A molecular urine assay for predicting prostate biopsy outcome. *J Urol*. 2008; 179(4):1587–1592. [PubMed: 18295257]
36. Tomlins SA, Aubin SM, Siddiqui J, Lonigro RJ, Sefton-Miller L, Miick S, Williamsen S, Hodge P, Meinke J, Blase A, Penabell Y, et al. Urine tmprss2:Erg fusion transcript stratifies prostate cancer risk in men with elevated serum psa. *Sci Transl Med*. 2011; 3(94):94ra72.
37. Gopalan A, Leversha MA, Satagopan JM, Zhou Q, Al-Ahmadie HA, Fine SW, Eastham JA, Scardino PT, Scher HI, Tickoo SK, Reuter VE, et al. Tmprss2-erg gene fusion is not associated with outcome in patients treated by prostatectomy. *Cancer Res*. 2009; 69(4):1400–1406. [PubMed: 19190343]
38. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*. 2009; 457(7231):910–914. [PubMed: 19212411]
39. Zhang WC, Shyh-Chang N, Yang H, Rai A, Umashankar S, Ma S, Soh BS, Sun LL, Tai BC, Nga ME, Bhakoo KK, et al. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell*. 2012; 148(1–2):259–272. [PubMed: 22225612]
40. Meyer TE, Fox SD, Issaq HJ, Xu X, Chu LW, Veenstra TD, Hsing AW. A reproducible and high-throughput hplc/ms method to separate sarcosine from alpha- and beta-alanine and to quantify sarcosine in human serum and urine. *Anal Chem*. 2011; 83(14):5735–5740. [PubMed: 21635006]
41. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med*. 2004; 350(22):2239–2246. [PubMed: 15163773]
42. Thompson IM, Pauler Ankerst D, Chi C, Goodman PJ, Tangen CM, Lippman SM, Lucia MS, Parnes HL, Coltman CA Jr. Prediction of prostate cancer for patients receiving finasteride: Results from the prostate cancer prevention trial. *J Clin Oncol*. 2007; 25(21):3076–3081. [PubMed: 17634486]
43. Balk SP, Ko YJ, Bubley GJ. Biology of prostate-specific antigen. *J Clin Oncol*. 2003; 21(2):383–391. [PubMed: 12525533]
44. Catalona WJ, Partin AW, Slawin KM, Brawer MK, Flanigan RC, Patel A, Richie JP, deKernion JB, Walsh PC, Scardino PT, Lange PH, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: A prospective multicenter clinical trial. *JAMA*. 1998; 279(19):1542–1547. [PubMed: 9605898]
45. Catalona WJ, Smith DS, Wolfert RL, Wang TJ, Rittenhouse HG, Ratliff TL, Nadler RB. Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. *JAMA*. 1995; 274(15):1214–1220. [PubMed: 7563511]
46. Mikolajczyk SD, Catalona WJ, Evans CL, Linton HJ, Millar LS, Marker KM, Katir D, Amirkhan A, Rittenhouse HG. Proenzyme forms of prostate-specific antigen in serum improve the detection of prostate cancer. *Clin Chem*. 2004; 50(6):1017–1025. [PubMed: 15054080]
47. Sokoll LJ, Chan DW, Mikolajczyk SD, Rittenhouse HG, Evans CL, Linton HJ, Mangold LA, Mohr P, Bartsch G, Klocker H, Horninger W, et al. Proenzyme psa for the early detection of prostate cancer in the 2.5–4.0 ng/ml total psa range: Preliminary analysis. *Urology*. 2003; 61(2):274–276. [PubMed: 12597929]

48. Sokoll LJ, Sanda MG, Feng Z, Kagan J, Mizrahi IA, Broyles DL, Partin AW, Srivastava S, Thompson IM, Wei JT, Zhang Z, et al. A prospective, multicenter, national cancer institute early detection research network study of [-2]proPSA: Improving prostate cancer detection and correlating with cancer aggressiveness. *Cancer Epidemiol Biomarkers Prev.* 2010; 19(5):1193–1200. [PubMed: 20447916]

Figure 1. Outliers in prostate cancer

(A) The TMPRSS2-ERG gene fusion is the product of two genes, TMPRSS2 and ERG, located on chromosome 21. Androgen signaling brings these loci into genomic proximity, which enables fusion formation following a disruptive event to the DNA (e.g. radiation). The gene fusion results in overexpression of ERG protein, which drives oncogenic processes in prostate cancer. (B) A schematic of prostate cancer outliers, demonstrating the presence of a driving gene fusion in approximately 50% of patients. The majority of gene fusions in prostate cancer involve ERG. SPINK1 overexpression marks a subset of gene fusion-negative prostate cancers, and SPINK1 has been shown to drive cancer cell invasion. (C)

PCA3 is a long noncoding RNA (lncRNA) located in the intron of the PRUNE2 gene on chromosome 9. PCA3 overexpression occurs in >90% of prostate cancer patients.

Table 1

Biomarker development paradigm

Phase 1	Biomarker discovery	
	<i>Major considerations</i>	Dynamic range of biomarker expression
		Absolute expression level of the biomarker
Phase 2	Biomarker Validation	
	<i>Major considerations</i>	Reproducibility of biomarker utility in independent cohorts
		Development of a robust, easy assay for measurement of the biomarker
		Cost of the assay
	Phase 3 Retrospective analyses of clinical datasets	
Phase 3	<i>Major considerations</i>	Does the biomarker improve stratification of cancer and normal samples?
		Reproducibility of clinical results in independent cohorts
		Demographic composition of cohorts, reflecting the target population
		Define the clinical endpoints distinguished by the biomarker
	Phase 4 Prospective screening	
Phase 4	<i>Major considerations</i>	Enrollment of the proper target population for the biomarker
		Matching control and experimental cohorts appropriately
		Evaluating the proper clinical endpoints
		Feasibility of biomarker measurement (cost, ease of assay)
		Efficacy in improving stratification of cancer samples
Phase 5	Large-scale population studies	
	<i>Major considerations</i>	Validation of biomarker efficacy in a broad setting
		Biomarker reproducibility across multiple patient cohorts
		Efficacy for the biomarker to improve patient management
		Financial feasibility of biomarker introduction into clinical practice

Table 2

Emerging prostate cancer biomarkers in development

Biomarker	Definition	Change in prostate cancer	Method of detection	Obtained from
% Free PSA	The fraction of PSA not bound to blood serum proteins	Decreased	Immunoassay	Serum
[-2]pro-PSA	Precursor PSA protein with 2 additional 5' amino acids	Increased	Immunoassay	Serum
[-5]pro-PSA	Precursor PSA protein with 5 additional 5' amino acids	Increased	Immunoassay	Serum
[-7]pro-PSA	Precursor PSA protein with 7 additional 5' amino acids	Increased	Immunoassay	Serum
TMPRSS2-ERG	Gene fusion mRNA product between TMPRSS2 and ERG	Increased	Transcription-mediated amplification, PCR	Urine
PCA3	Long noncoding RNA transcript upregulated in cancer	Increased	Transcription-mediated amplification, PCR	Urine
Sarcosine	Small metabolite generated during glycine metabolism	Increased	Mass spectrometry	Urine