Reviews

Integrative proteomics of prostate cancer
Mauro Scaravilli¹, Ebrahim Afyounian²,₃, Matti Nykter²,₃,₄, Tapio Visakorpi²,₃,₅ and Leena Latonen¹

Abstract
Large-scale, genome-wide sequencing efforts have comprehensively characterized genomic and transcriptomic landscape of prostate cancer, revealing recurrently altered genes as well as substantial heterogeneity in cancer-driving mutations. Recently, integrative proteogenomic studies of prostate tumor samples have demonstrated that, in fact, there is a poor correlation between alterations in the genome and transcriptome compared with the proteomic profiles, suggesting that a large fraction of the accrued changes at nucleotide levels are insignificant with respect to protein functions. The proteomic profiling has provided an enhanced view of deregulated pathways during development of prostate cancer and formation of castration resistance. These results underline the importance of proteogenomic analyses to gain better understanding of the cancer phenotypes and to develop more effective therapeutic strategies.

Introduction
Prostate cancer is the second most commonly diagnosed cancer among males worldwide and the most frequently diagnosed malignancy in developed countries [1]. The number of prostate cancer cases has increased by 42% during the last decade [2]. The disease often progresses slowly, and surgical removal of the tumor by radical prostatectomy or radiation therapy are effective treatments for organ-confined prostate cancer. Cancers that have spread outside the prostatic capsule are treated hormonally, either by surgical or chemical castration or by antiandrogens. Although initially effective, hormonal therapy eventually fails in over 80% of the cases, and the cancer becomes hormone-refractory, that is, resistant to therapy (castration-resistant prostate cancer; CRPC) [3]. Currently, there is no effective treatment for CRPC. In addition, there is a lack of reliable markers to identify more aggressive forms of the disease at the time of diagnosis. A recent development in the treatment of advanced disease has sprung from the use of next-generation antiandrogen drugs, such as abiraterone and enzalutamide. However, although effective at slowing down disease progression [4,5], these drugs are responsible for an increase in treatment-induced neuroendocrine (NE) type cases, characterized by a highly aggressive phenotype [6,7].

During the last decade, next-generation sequencing technologies have revolutionized studies of human genetics and diseases, including cancer. Several genome-wide analyses of genetic and transcriptional changes have been performed in prostate cancer with the aim of finding the molecular alterations responsible for the development and progression of this disease. These studies led to the identification of several, recurrent deletions and mutations in tumor-suppressor genes, amplifications of oncogenic loci, aberrations in genes involved in key cellular pathways such as androgen receptor (AR) signaling, novel noncoding RNA species, as well as common chromatin rearrangements [8–14]. Moreover, recent efforts by The Cancer Genome Atlas and the Prostate Cancer Foundation/Stand Up To Cancer (PCF/SU2C) consortia, using significantly larger cohorts of clinical samples, allowed the characterization of additional recurrent mutations at lower frequencies. These efforts revealed a substantial heterogeneity of alterations in prostate cancer, as well as a long-tail distribution in the incidence of mutated oncogenic drivers...
Despite this uncovered complexity of alterations at the genetic, epigenetic, and transcriptomic levels, only recently have we started to discover how these affect the proteomic landscape of prostate cancer (Figure 1).

Highly informative large-scale proteomics analyses (in the order of thousands of quantitated proteins) have been published from primary prostate tumors as well as bone metastases [17–19]. These studies serve as valuable snapshots to the clinical proteome and provide insights into the heterogeneity of the disease at the protein level. Interestingly, recent reports have demonstrated a poor correlation between expression changes at the transcriptomic and proteomic levels in several types of cancer [20–22]. These findings have highlighted the need for studies that investigate the inference from genomic abnormalities to cancer phenotypes by integrative proteogenomic studies, as well as strategies to improve the current methods for the identification and characterization of aberrant protein products [23].

### Integrative proteogenomic studies in prostate cancer

The first integrative study combining genomic and transcriptomic analyses to full proteome from the same clinical prostate tumor samples was published by Lato nen et al., in 2018 [24]. This study compared alterations in primary prostate cancer and CRPC and, thus, revealed alterations occurring during the development of the disease as well as during the formation of treatment resistance. In the study, a cohort of clinical samples including benign prostatic hyperplasia, untreated primary tumors, and locally recurrent CRPC cases were analyzed with mass spectrometry (MS) and SWATH-quantitation-based proteomics. The high-confidence data of 3400 proteins were integrated with previously generated genetic, epigenetic, and transcriptomic profiles of the same samples [13,14]. The results of these analyses showed that, especially in CRPC, gene copy number, DNA methylation, and RNA expression levels do not reliably predict proteomic changes. Differentially expressed proteins and novel regulated pathways that

![Figure 1](https://i.imgur.com/3Q5Q5Q5.png)

**Computational biology lens uncovers deeper insights into prostate cancer by integrating multiple levels of data along the central dogma.**

Samples collected from prostate cancer tissues are subjected to several types of high throughput measurements providing different information and views of molecular alterations. Each level of data (left panel) by itself only explains parts of what is observed in the proteome. Integrating these data using different computational biology methods (right panel), we can determine the extent to which changes at one level contribute to what is observed at the proteome level, which represents the endpoint in the central dogma. Measurements of alterations in the genome (e.g. SNVs, insertions, deletions, and translocations), epigenome (e.g. DNA methylation, histone modifications), transcriptome (RNA expression patterns), and proteome (protein expression patterns, including posttranslational modifications) are represented. Other data modalities, such as protein-DNA interactions as well as chromatin accessibility data, could also be utilized to gain even deeper insights. These insights aid in devising better biomarkers and enhanced future therapeutics for the disease. TAD, topologically associated domain.
are not associated with corresponding changes at the mRNA level were identified in both primary and more advanced stages of the disease. Moreover, several novel miRNA-target gene interactions were found to affect the protein endproduct with no significant changes at mRNA level, providing miRNA targets to further assess for potential clinical applications. Overall, this study provided valuable new insight into proteomic changes throughout the development and progression of the disease and into potential use of proteomic data to develop new therapeutic strategies.

The second integrative large-scale proteomics study from prostate cancer was performed on a clinically homogeneous cohort of localized, treatment-naive primary prostate tumors analyzed with MS shotgun proteomics [25]. Approximately 3400 individual proteins were identified in all samples. Integrative analysis of genome, methylome, and transcriptome showed that genomic and proteomic features are poorly correlated, confirming previous observations in prostate cancer [24] and other cancer types [20–22]. In this study, gene fusions of the erythroblast transformation-specific (ETS) gene family, which are the most frequently occurring somatic aberrations in prostate cancer and not associated with clinical outcome, were found to be significantly associated with changes in protein abundances. The proteins that were most affected by the presence of ETS fusion showed a good correlation with their mRNA level. However, many genes associated with ETS fusion status either at mRNA or protein level only. In this article, five subtypes of primary prostate cancer were identified based on the proteome. Surprisingly, genomic and proteomic subtypes were largely independent, as were proteomic subtypes of mutational burden [genomic rearrangements (GRs) and single nucleotide variants (SNVs)] and AR activity signatures—as expected for treatment-naive cases [25].

**Poor correlation of genomic and epigenomic events with the prostate cancer proteome**

Both of the current integrative proteogenomic studies of prostate cancer show that genomic and proteomic features are poorly correlated [24,25]. Primary prostate cancer is driven by copy number alterations (CNVs) more than SNVs [26], yet, only 2% of proteins have their abundance associated with CNAs [25]. Furthermore, the mutational burden as a whole (genomic rearrangements and SNVs) in primary prostate cancer patients was not significantly associated with tumor protein abundance profile [25]. Latonen et al. (2018) showed that neither the altered gene copy numbers nor the global methylation changes were detectable through the proteome to the same extent as they correlate with the global RNA expression. This was especially prominent in CRPC, where significantly more genomic aberrations exist compared with primary cancers [24]. This indicates that a large proportion of genomic aberrations that accrue to the tumors are untranslated and subsequently left without a functional effect at the protein level.

**The transcriptome is a poor predictor of the proteome in prostate cancer**

Globally, mRNA and protein abundances are only weakly correlated in primary prostate cancers [24,25], and this correlation is further decreased in CRPC [24]. Some genes even show opposite expression patterns at mRNA compared with protein levels, and mRNA abundance explains only around 10% of protein abundance variability in primary prostate cancer [24,25]. E.g. deletion of PTEN affects abundance of around half of the studied genes at the mRNA level, but only 2.7% at the protein level [25]. Similar findings have been reported for breast, ovarian, and colorectal cancer [20,22], indicating that the transcriptome is a poor predictor of the proteome in cancer. Nevertheless, the transcriptomic changes observed are likely to affect cellular functions that are dependent on noncoding RNAs and possible translation-independent functions of mRNA.

Interestingly, 10% of most abundant proteins were better correlated with their respective miRNAs than the 10% least abundant proteins detected in the primary prostate cancer proteomes [25]. This finding likely reflects the differential role and regulation of these protein groups. While the high abundance proteins that are more easily detected by MS approaches often represent structural and core function proteins (e.g. membrane-bound organelles, extracellular proteins, cytoskeletal and attachment proteins, ribosomal proteins of protein synthesis), many highly and transiently regulated proteins such as signal transducers and transcription factors are generally lower in abundance. This is supported by the notion that an overrepresentation of nuclear proteins was found in genes with coding transcripts expressed but no protein detected [25]. Conversely, an overrepresentation of immune-related genes was found for proteins with no mRNA detected [25], suggesting that the source of gene expression lies elsewhere than in the tumor cells.

**Integrative proteomics view of prostate cancer pathways**

In addition to characterization of the overall protein levels, the integrative analyses have also provided views to explain how alterations in protein levels and post-translational regulation affect cellular functions (Figure 2). A valuable study in this direction was presented by Drake et al. (2016) [27], who compared genomic, transcriptomic, and phosphoproteomic (antibody-based extraction of pST- and pY-containing peptides) data sets from metastatic CRPC samples. Differentially expressed master transcriptional regulators, functionally mutated genes, and differentially activated kinases were revealed indicating 6 major
signaling pathways with phosphorylation of several key residues. These included cell cycle, DNA repair, nuclear receptor, PI3K-AKT-mTOR, stemness (including TGFβ, WNT, NOTCH, and MYC pathways), and migration and invasion pathways [27].

When comparing altered pathways based on RNA and protein expression data, Latonen et al. (2018) [24] showed that less than a third of them were common. Interestingly, even though cell metabolism is already well known to undergo alterations in cancer, the integrative proteomics studies revealed novel information to this aspect. In primary cancer, ERG-fusions were found to associate with carboxylic acid metabolism, corroborating the links between ERG and lipid metabolism [25]. In Latonen et al., sequential changes in the citric acid cycle (Krebs cycle) were identified first during cancer development and then during formation of treatment resistance. Metabolic pathway alterations were not identified by phosphoproteomic analyses, likely reflecting the fact that metabolic enzymes, as many other housekeeping proteins, are subject to less posttranslational activity regulation than, for example, signaling proteins. Conversely, DNA repair, migration-related and invasion-related, as well as PI3K-AKT-mTOR pathways were all detected by both proteomic and phosphoproteomic analyses [24,26].

In primary cancer, ETS fusions overall were associated with alterations in pathways governing intracellular and extracellular vesicles and lysosomal genes [25]. Also, translation-related and protein degradation-related pathways as well as GTPase signaling were picked by proteomics better than transcriptomics [17,24]. The notion that proteomic data can reveal pathway alterations not visible through RNA expression was especially true in CRPC, where alterations in DNA metabolism and repair, cell metabolism, and various signaling events were better spotted [18,24]. Further, the poor correlation of protein levels with mRNA expression in CRPC was reflected through the fact that based on mRNA levels, several pathways were identified to be altered which did not appear so based on the protein levels. These divergences in altered, cancer-associated signaling pathways detected by transcriptomics and proteomics [24,25] underline that caution should be taken in interpreting cellular activity patterns and the effects of genomic aberrations by transcriptomic data alone.

Technical considerations and future perspectives

With current technologies, large-scale proteomes are still biased toward high-abundance proteins. This results in, for example, cytoplasmic and structural proteins being overrepresented compared with low abundance nuclear proteins and signaling proteins. Biologically, the latter likely bear a high relevance in cancer development and evolution. For these proteins, the post-transcriptional activity regulation patterns are often also
crucial. Thus, the lack of data on these still hampers our understanding of the prostate cancer proteomes. The National Cancer Institute’s Clinical Proteomic Tumor Analysis Consortium (https://proteomics.cancer.gov/programs/cptac) represents a major effort aiming at better understanding the molecular basis of cancer through the collection of large scale glycoproteomes and phosphoproteomes in addition to full proteomes from several cancer types, including prostate cancer. Further, the development of high-throughput single cell proteomics approaches will be essential to dissect tumor cell heterogeneity at the protein level.

Another major challenge for the future is assessing the functional subproteomes in prostate cancer, such as the interactomes of key cancer driving factors. Several studies performed in cell lines report wildtype and prostate cancer-related T877A mutant AR interactomes based on affinity purification or proximity labeling followed by MS [28–32]. Another relevant subproteome studied in cell lines is the N-Myc-interactome, shedding light on events taking place in NE-type CRPC [33]. While most of these studies correlate their findings with chromatin binding and/or gene expression results, genuine integrative views of activation and mutation status-dependent changes in the relevant transcription factor interactomes are lacking both in cell line models as well as in prostate cancer tumors. Further, the treatment resistance linked to NE development of prostate tumors treated with next generation androgens has so far been explored at the proteomic level only in patient-derived xenografts [34]. The integrative proteome profiling of treatment-induced NE tumors is one key task for future studies (Figure 2), especially for understanding tumor cell plasticity not evidenced at the genetic level, which plays a central role in treatment resistance.

Final words
It has already been shown that prognostic biomarkers based on integrating multiomics significantly outperform those based on a single data type [25]. Beyond biomarkers, discoveries for better treatments in the future require a comprehensive understanding of prostate cancer biology at all molecular levels. After the leap that the current proteogenomic studies have provided us, knowledge and integration of functional subproteomes and metabolomes are warranted for further steps towardcurative prostate cancer therapies also for the advanced disease.

Conflict of interest statement
Nothing declared.

Acknowledgements
The authors’ work is supported by the Academy of Finland (TV 317755, MN 310829, LL 317671), Sigrid Juselius Foundation (TV, LL), Tampere University doctoral school (EA) and the Cancer Society of Finland.

References
Papers of particular interest, published within the period of review, have been highlighted as:
* of special interest
** of outstanding interest


This was the first integrative large-scale proteomics study published of clinical prostate cancer, and it remains the only such study of CRPC to date. 17 primary and 11 CRPC tumors were analysed for genomes, epigenomes, transcriptomes and proteomes in comparison to benign prostatic hyperplasia (BPH) controls. The proteome of 3400 proteins was identified from fresh-frozen tissues by label free mass spectrometry and SWATH analysis. The strength of the study is in comparison of the clinical evolution of the disease, enlightening the protein changes during treatment resistance and identifying previously unknown alterations in e.g. tumor metabolism.
Integrative Proteomics of Prostate Cancer Scaravilli et al. 49

based shotgun proteomics performed on fresh-frozen tissues. Around 3400 individual proteins were identified in all samples and subjected to integrative analysis with data of the genome, methylome and transcriptome. The novel findings in this study where strong association of ETS fusions with changes in protein abundances and demonstration that prognostic biomarkers based on integrating multi-omics significantly outperform those based on a single data type.


In this study, 11 treatment-naive and 16 metastatic CRPC samples obtained at rapid autopsy were analyzed using quantitative label-free mass spectrometry to generate a dataset of phosphorylated proteins. A total of 8348 phosphorylated peptides were identified. Phosphopeptides corresponding to 74 kinases were found, 18 of which were differentially phosphorylated in metastatic tissues. These findings were integrated with transcriptomic and genomic datasets collected from different metastatic CRPC tissues or patients. Overall, this study shows that changes in the phosphoproteome and particularly at kinase level (protein signaling) can explain variations in response to therapy in patients with similar transcriptomes.


