Proteogenomic heterogeneity of localized human prostate cancer progression

Konstantina Charmpi1,2,#, Tiannan Guo3,4,#, Qing Zhong5,#, Ulrich Wagner5, Ulrich Wagner5, Nora C. Toussaint5,6,7, Christine E. Fritz5, Chunhui Yuan4, Hao Chen4, Niels J. Rupp5, Ailsa Christiansen5, Dorothea Rutishauser5, Jan H. Rüschoff5, Christian Fankhauser5, Karim Saba5,8, Cedric Poyet8, Thomas Hermanns8, Kathrin Oehl1, Ariane L. Moore9, Christian Beisel9, Laurence Calzone10, Loredana Martignetti10, Qiushi Zhang4, Yi Zhu3,4, María Rodríguez Martínez11, Matteo Manica11, Michael C. Haffner12, Ruedi Aebersold3,13,* Peter J. Wild5,14,* Andreas Beyer1,2,*

1 CECAD, University of Cologne, Cologne, Germany
2 Center for Molecular Medicine Cologne (CMMC), Medical Faculty, University of Cologne, Cologne, Germany
3 Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland
4 School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China
5 Department of Pathology and Molecular Pathology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
6 NEXUS Personalized Health Technologies, ETH Zurich, Zurich, Switzerland
7 Swiss Institute of Bioinformatics, Zurich, Switzerland
8 Department of Urology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
9 Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland
10 Institut Curie, Paris, France
11 IBM Zurich Research Laboratory, Zurich, Switzerland
12 Fred Hutchinson Cancer Research Center, Seattle, USA
13 Faculty of Science, University of Zurich, Zurich, Switzerland
14 Dr. Senckenberg Institute of Pathology, University Hospital Frankfurt, Goethe-University Frankfurt, Frankfurt, Germany.

*Corresponding authors:
Tiannan Guo: guotiannan@westlake.edu.cn
Ruedi Aebersold: aebersold@imsb.biol.ethz.ch
Peter J. Wild: Peter.Wild@kgu.de
Andreas Beyer: abeyer2@uni-koeln.de

#Equal contribution
Abstract

Tumor-specific genomic aberrations are routinely determined by high throughput genomic measurements. However, it is unclear how complex genome alterations affect molecular networks through changing protein levels, and consequently biochemical states of tumor tissues. Here, we investigated how tumor heterogeneity evolves during prostate cancer progression. In this study, we performed proteogenomic analyses of 105 prostate samples, consisting of both benign prostatic hyperplasia regions and malignant tumors, from 39 prostate cancer (PCa) patients. Exome sequencing, copy number analysis, RNA sequencing and quantitative proteomic data were integrated using a network-based approach and related to clinical and histopathological features. In general, the number and magnitude of alterations (DNA, RNA and protein) correlated with histopathological tumor grades. Although common sets of proteins were affected in high-grade tumors, the extent to which these proteins changed their concentrations varied considerably across tumors. Our multi-layer network integration identified a sub-network consisting of nine genes whose activity positively correlated with increasingly aggressive tumor phenotypes. Importantly, although the effects on individual gene members were barely detectable, together the perturbation of this sub-network was predictive for recurrence-free survival time. The multi-omics profiling of multiple tumor sites from the same patients revealed cases of likely shared clonal origins as well as the occasional co-existence of multiple clonally independent tumors in the same prostate. Overall, this study revealed molecular networks with remarkably convergent alterations across tumor sites and patients, but it also exposed a diversity of network effects: we could not identify a single sub-network that was perturbed in all high-grade tumor regions.
Introduction

Prostate cancer (PCa) represents one of the most common neoplasm among men with almost 1,300,000 new cases and 360,000 deaths in 2018 accounting for 15% of all cancers diagnosed. PCa is the fifth leading cause of cancer death in men and represents 6.6% of total cancer mortality in men. Despite earlier detection and new treatments, the lifetime risk to die of PCa has remained stable at approximately 3% since 1980. (National Cancer Institute SEER data: https://seer.cancer.gov/statfacts/html/prost.html).

In many patients, PCa is indolent and slow growing. The challenge is to identify those patients who are unlikely to experience significant progression while offering radical therapy to those who are at risk. Current risk stratification models are based on clinicopathological variables including histomorphologically defined grade groups, prostate-specific antigen (PSA) levels and clinical stage. Although those variables provide important information for clinical risk assessment and treatment planning, they do not sufficiently predict the course of the disease.

Extensive genomic profiling efforts have provided important insights into the common genomic alterations in primary and metastatic PCa. Interestingly, PCa genomes show a high frequency of recurrent large-scale chromosomal rearrangements such as TMPRSS2-ERG. In addition, extensive copy number alterations (CNAs) are common in PCa, yet point mutations are relatively infrequent in primary PCa compared to other cancers. A major complicating factor is that around 80% of PCAs are multifocal and harbor multiple spatially and often morphologically distinct tumor foci. Several recent studies have suggested that the majority of topographically distinct tumor foci appear to arise independently and show few or no overlap in driver gene alterations. Therefore, a given prostate gland can harbor clonally independent PCAs.

To allow for a more functional assessment of the biochemical state of PCa, it is necessary to go beyond genomic alterations and comprehensively catalogue cancer specific genomic, transcriptomic and proteomic alterations in an integrated manner. Such an approach will provide critical information for basic and translational research and could result into clinically relevant markers. While hundreds of PCa genomes and transcriptomes have been profiled to date, little is known about the PCa proteome. Although recent work has emphasized the need for integrated multi-omics profiling of PCa, we still lack understanding about how genomic changes impact on mRNA and protein levels. Especially the complex relationship between tumor grade, tumor progression and multi-layered molecular network changes remains largely elusive.

For example, previous work has shown that copy number changes may alter transcript levels of many genes, whereas the respective protein levels remain relatively stable. Indeed, there is compelling
evidence across multiple tumor types that many genomic alterations are ‘buffered’ at the protein level and are hence mostly clinically inconsequential²²,²³. To better understand the evolution of PCa and to identify core networks perturbed by genomic alterations and thus central for the tumor phenotype, it is therefore essential to investigate the transmission of CNAs to the transcriptomic and proteomic level.

To this end, it is important to decipher which genomic alterations impact PCa proteomes, which of those proteomic alterations are functionally relevant, and how molecular networks are perturbed at the protein level across tumors.

To address these open questions, we performed a multi-omics profiling of radical prostatectomy (RP) specimens at the level of the genome, transcriptome and proteome from adjacent biopsy-level samples, using state-of-the-art technologies. Unique features of this study are (1) the utilization of PCT (pressure cycling technology)-SWATH (Sequential Window Acquisition of all THeoretical Mass Spectra) mass spectrometry²⁴,²⁵, allowing rapid and reproducible quantification of thousands of proteins from biopsy-level tissue samples collected in clinical cohorts; (2) the simultaneous profiling of all omics layers from the same tissue regions; (3) inclusion and full profiling of benign regions, which provides a matching control for each tumor; and (4) the full multi-omics characterization of multiple tumor regions from the same patients, thus enabling the detailed investigation of tumor heterogeneity. This design resulted in the multi-layered analyses of 105 samples from 39 PCa patients, as well as of the exome of corresponding peripheral blood cells yielding a comprehensive molecular profile for each patient and identified molecular networks that are commonly altered in multiple patients. Importantly, some of the affected genes/proteins exhibited very small individual effect sizes, suggesting that combined network effects of multiple genes may significantly contribute to determining PCa phenotypes.

Results

Proteogenomic analysis of the sample cohort identifies known PCa biomarkers.

In this study, we analyzed 39 PCa patients (Supplementary Fig. 1) belonging to three groups who underwent laparoscopic robotic-assisted RP. The patients were from the PCa Outcomes Cohort (ProCOC) study²⁶,²⁷. Tumor areas were graded using the ISUP (International Society of Urological Pathology) grade groups²⁸, which range from ISUP grade group G1 (least aggressive) to G5 (most aggressive). The more advanced grade groups G4 and G5 are considered jointly (G4/5). The cohort tested included 12 low-grade (G1), 17 intermediate- (G2 and G3), and 10 high-grade (G4/5) patients (Fig. 1a, Supplementary Fig. 1,
Supplementary Table 1). For low-grade PCa patients, we selected two representative regions, one of benign prostatic hyperplasia (BPH) and one of malignant tumor (TA). Since PCa often presents as a multifocal disease with heterogeneous grading within each prostate specimen we analyzed two different tumor regions from the 27 intermediate- and high-grade patients. In those cases three representative regions, including BPH, the most aggressive tumor (TA1) and a secondary, lower-grade tumor (TA2) were analyzed. Thus, TA1 always represented the higher-grade nodule compared to TA2. Note, whereas each patient was assigned a patient-specific overall grade (i.e. ‘low’, ‘intermediate’ or ‘high’), each tumor area was additionally assigned an individual grade group based on its histological appearance. According to current ISUP guidelines, the grading of the entire prostate specimen depends on the size and grade of individual nodules. Thus, it is possible that the patient grading is lower than the grading of the most aggressive nodule, if another lower-grade nodule is larger. Tumor regions contained at least 70% tumor cellularity and the distance between the analyzed areas (TA1 versus TA2) was at least 5 mm. Altogether, we obtained 105 prostate tissue specimens (Supplementary Table 1). Three adjacent tissue biopsies of the dimensions 0.6 x 0.6 x 3.0 mm were punched from each representative region for exome sequencing, CNA (derived from the exome sequencing data), RNA sequencing (RNA-seq), and quantitative proteomic analysis using the PCT-SWATH technology respectively. Proteomic analysis was performed in duplicates for each tissue sample. Peripheral blood samples from each patient were also subjected to exome sequencing and served as the genomic wild-type reference (Fig. 1). All three types of grading (i.e. patient-specific overall grading, TA1 grading and TA2 grading) were predictive of the recurrence-free survival (RFS) in our study.

In agreement with prior reports, we observed relatively few recurrent point mutations across patients (Supplementary Fig. 2, Supplementary Table 2), but substantial CNAs (Supplementary Figs. 3 and 4, Supplementary Table 3). In total, 1,110 genes showed copy number gains in at least five samples or copy number losses in at least five samples (see Supplementary Text for details). Likewise, our data confirmed the differential expression of several transcripts/proteins that had previously been suggested as PCa biomarkers or which are known oncogenes in other tumor types (Supplementary Fig. 5, Supplementary Tables 4 and 5) (see Supplementary Text for details). This consistency with previously published results confirmed the quality of our data and motivated us to go beyond previous work by performing a network-based multi-omics multi-gene analysis.
Molecular perturbations correlate with tumor grade.

As a first step towards a cross-layer analysis, we asked if high-grade PCa would be generally affected by stronger alterations (compared to low-grade PCa) at the genome, transcriptome, and proteome layer. Thus, we devised molecular perturbation scores that quantified the number of affected genes/proteins and the extent to which these genes/proteins were altered in the tumor specimens compared to their benign controls (see Methods for details). Higher-grade tumors (G3 and G4/5) exhibited significantly higher molecular perturbation scores than lower-grade tumors (G1 and G2). Those differences were statistically significant in all but one case ($P$ value $< 0.05$, one-sided Wilcoxon rank sum test, Fig. 2). The CNA perturbation magnitude exhibited the highest correlation with the PCa grading, confirming prior studies documenting the tight association between CNA, histopathological grade and risk of progression. Previous work suggested that copy number changes are to some extent buffered at the protein level. Interestingly, we observed that proteins known to be part of protein complexes were significantly less strongly correlated with the fold changes (FCs) of their coding mRNAs than proteins not known to be part of protein complexes ($P$ value $< 2.6e-11$, one-sided t-test, Supplementary Fig. 6). This result is consistent with the concept that protein complex stoichiometry contributes to the buffering of mRNA changes at the level of proteins. Thus, molecular patterns in high-grade PCa are more strongly perturbed at all layers and the effects of genomic variation are progressively but non-uniformly attenuated along the axis of gene expression.

Inter-patient heterogeneity decreases along protein biosynthesis.

Our analysis of CNA profiles (above and Supplementary Text) already revealed many shared CNAs across patients, suggesting that such common CNAs might represent genomic driver changes. We therefore investigated if such a convergence towards common molecular endpoints could also be observed at the transcript and protein level. To address this question, we first computed a reference molecular signature that is characteristic of the molecular perturbations of tumors in a given grade group. These 'centroid vectors' were obtained by computing the average tumor-to-benign FCs across all samples within a grade group. Consistent with the observation above, we found that the average effect sizes (averaged absolute centroid FCs) were increasing with the grade group for all three layers (CNA, mRNA, and protein; Supplementary Table 6). Next, we compared each individual sample within a group against the matching centroid of the same group. For the quantification of the similarity between a tumor sample and the corresponding centroid we used four similarity/distance measures: Pearson correlation, Mutual Information (MI), Manhattan distance and Euclidean distance. While the first two measures (i.e. Pearson...
correlation and MI) quantify the degree to which the tumor sample and centroid vector co-vary, the other
two measures \textit{(i.e.} Manhattan and Euclidean distance\textit{)} also take into account the magnitude of the FCs in
the two vectors. To illustrate this difference, imagine two patients having perturbations of the same
genes/proteins, but one of them exhibiting overall two-fold greater FCs \textit{(i.e.} all FCs are increased by a
factor of two compared to the other patient\textit{)}. In such a scenario Pearson correlation and MI would yield
identical results for the two patients when compared to the centroid, whereas Manhattan and Euclidean
distance would identify them as different. Using the Pearson correlation and MI, we found that high-grade
PCa (G4/5) were more similar to their respective centroid than low-grade PCa (G1) to their centroid (Fig. 3). This effect was particularly pronounced for protein-level changes. This is consistent with the notion
that protein levels (and not mRNA levels) are subjected to stronger selection. Interestingly, when the
Euclidean distance and the Manhattan distance were used to characterize tumor similarity, we found that
the high-grade tumors were more \textit{dissimilar} to each other than the low-grade tumors (Supplementary
Fig. 7), in sharp contrast to the Pearson correlation and MI. Based on the nature of the different similarity
measures tested, we hypothesized that there is a set of proteins commonly affected in their abundance
by oncogenic alterations in high-grade tumors. This would increase the similarity using the Pearson
correlation or MI. However, although the same proteins are affected, they are affected to a different
extent in different high-grade tumors, \textit{i.e.} the FCs exhibited a high degree of variability (Supplementary
Fig. 7), which would increase the dissimilarity based on the Euclidean distance or Manhattan distance (see
Supplementary Fig. 7 for a schematic explanation).

To further corroborate the notion of common endpoints, we identified the 10 proteins with the
largest average absolute FCs across all tumor specimens (Supplementary Table 6). Among them was PSA
(KLK3), and several other well established PCa-associated proteins like AGR2$^{37}$, MDH2$^{38}$, MFAP4$^{39}$ and
FABP5$^{40}$. We observed that for some of these top 10 proteins, FCs were more extreme in the higher-grade
tumors (G3 and G4/5) compared to lower-grade tumors (G1 and G2), such as MDH2 and SEPHS1 (up-
regulation; Fig. 3c). RABL3 was one of the most strongly down-regulated proteins (Fig. 3c), which is a
surprising finding as RABL3 is known to be up-regulated in other solid tumors$^{41,42}$. Interestingly, in most
cases these proteins were from loci that were not subject to CNAs (Supplementary Fig. 7, Supplementary
Table 6). Taken together, these findings suggest that tumor mechanisms in different patients converged
on common protein endpoints and that the expression levels of these proteins were progressively more
strongly affected during tumor evolution. Thus, we next aimed to identify molecular networks that
underlie these alterations and thus indicate altered biochemical states of the respective tissues.
Gene networks are regulated on CNA, mRNA and protein level.

Although the analyses described above suggested convergent proteomic changes, we did not detect a single mutation that was common to all high-grade tumors, let alone common to all 39 PCa cases tested. It has previously been suggested that mutations affecting different genes could impact common molecular networks if the respective gene products interact at the molecular level. However, previous analyses were mostly restricted to individual molecular layers. For example, it was shown that genes mutated in different patients often cluster together in molecular interaction networks. But, effects of these mutations on transcript and protein levels remained unexplored in this case. Here, we aimed at a multi-layer network analysis, involving the genome, transcriptome and proteome. We applied network analysis at each layer separately and the layers were combined afterwards. To identify sub-networks commonly perturbed, we mapped our data onto the STRING gene interaction network and employed network propagation separately to the CNA, transcriptome and proteome data for each of the tumor samples. We excluded point mutations from this analysis as their frequency was too low in our cohort. By combining published molecular interactome data with a network propagation algorithm, we aimed to 'enrich' network regions with many perturbed genes/proteins. We reasoned that the convergent consequences of genomic variants on common network regions would be indicative of specific biochemical functions that are important for the tumor biology. We therefore identified genes/proteins in network regions that showed a higher score (or a lower score) in high-grade (G4/5) relative to lower-grade (G1) tumor groups at all three levels (Fig. 4a, b; Methods). This analysis identified sub-networks consisting of over- and under-expressed genes (relative to the benign controls). We found 57 amplified genes (Supplementary Table 6) for which transcripts and proteins were often over-expressed in high-grade PCa (Fig. 4a) and 21 genes with copy number loss (Supplementary Table 6) for which transcripts and proteins were often down-regulated compared to lower-grade tumors (Fig. 4b).

Among the up-regulated network nodes, we observed genes modulating the stability of chromatin, such as chromatin-binding protein Chromobox 1 (CBX1), SET Domain Bifurcated 1 (SETDB1), a function linking to H3K27me3 and H3K9me3 in chromatin, and CBX3 (known as HP1-γ). SETDB1 is an oncogene in melanoma and has also been found to be over-expressed in PCa and cell lines. Further, we found genes involved in DNA damage repair, such as SMG7 and ATR, and PRKCZ, which had already been suggested as a biomarker prognostic for survival in PCa. Multiple actin related proteins including ARPC1B, ARPC5, ACTL6A, and CFL1, which are markers for aggressive cancers, were part of the up-regulated network nodes. Moreover, the up-regulated genes contained proteins related to the cell cycle like BANF1 and proteins interacting with the centrosome including LAMTOR1 and RAB7A that
had already been associated with PCa\textsuperscript{60}. Finally, several signaling molecules with known roles in PCa were up-regulated, such as the transcription factor Yin Yang 1 (YY1)\textsuperscript{61}, the TGF-\(\beta\) receptor TGFBR1\textsuperscript{62}, and KPNA4, which promotes metastasis through activation of NF-\(\kappa\)B and Notch signaling\textsuperscript{63}. Thus, up-regulated network nodes are involved in DNA/chromatin integrity and growth control.

Likewise, several of the down-regulated genes had functions associated with PCa. For example, the oxidative stress related gene MGST1, which is recurrently deleted in PCa\textsuperscript{64}. ALDH1A3 is a direct androgen-responsive gene, which encodes NAD-dependent aldehyde dehydrogenase\textsuperscript{65}. DHCR24 is involved in cholesterol biosynthesis and regulated by the androgen receptor\textsuperscript{66}. Polymorphisms in CYP1A1\textsuperscript{67-69} are associated with PCa risk in several meta-analyses among different ethnicities.

Further, our network analysis is suggesting tumor mechanisms converging on genes that are known contributors to PCa tumor biology. For example, the PCa-associated gene SF3B2\textsuperscript{70,71} was only weakly amplified in some of the high-grade tumors (average log\(_2\)FC = 0.016) and mRNA levels showed similarly small changes (average log\(_2\)FC = 0.024). On the other hand, the SF3B2 protein levels were consistently and more strongly up-regulated across tumors (average log\(_2\)FC = 0.31), especially within the high-grade tumors (Supplementary Fig. 8). Another example is UBE2T whose over-expression is known to be associated with PCa\textsuperscript{72}. Unfortunately, we could not quantify the corresponding protein levels. However, we observed a strong and consistent mRNA over-expression across several tumors (average log\(_2\)FC = 0.73), even though at the DNA level the gene was only weakly amplified (average log\(_2\)FC = 0.023; Supplementary Fig. 8). Our findings of more heterogeneous CNAs, but more uniform mRNA and protein alterations point on convergent evolutionary mechanisms, as we move along the axis of gene expression.

Next, we analyzed the largest connected component with genes up-regulated in advanced disease in more detail (see Methods). It consists of the nine nodes EMD, BANF1, ACTL6A, YY1, RUVBL1, KANSL1, MRGBP, VPS72 and ZNHIT1 (Fig. 4a), and is referred to in the following as Network Component 1 (Supplementary Table 6). Seven of these proteins are involved in chromosome organization which may induce genomic alterations and influence the outcome of multiple cancers including PCa\textsuperscript{73}. For example, the actin-related protein ACTL6A is a member of the SWI/SNF (BAF) chromatin remodeling complex\textsuperscript{74}, and a known oncogene and a prognostic biomarker for PCa\textsuperscript{75}. Likewise, KANSL1 is involved in histone post-translation modifications, while VPS72 is a member of histone- and chromatin remodeling complexes\textsuperscript{76}.

Several samples were characterized by a small, but consistent DNA amplification of multiple members of Network Component 1 (Fig. 4c). Out of the 66 tumor samples, there were 30 samples – belonging to all grade groups – with a weak but consistent DNA amplification of Network Component 1 members, while the high-grade samples had stronger amplifications on average (i.e. larger effect sizes).
Importantly, gene members of Network Component 1 were dispersed across eight chromosomes (Supplementary Table 6). The parallel DNA amplification of these genes is therefore the result of multiple independent CNA events, while the signal on any single gene alone was too weak to be significant in isolation. In some but not all cases, the amplifications led to a small, but consistent increase in mRNA expression of the amplified gene loci (Fig. 4d). Unfortunately, only three out of the nine proteins were detected in our proteomics experiments (Fig. 4e). Interestingly, patients where the DNA amplifications led to transcript over-expression were almost always high-grade patients, whereas patients where the amplification affected gene expression to a smaller extent were low- or intermediate-grade patients (Fig. 4c, d). Further, we noticed that TA2 samples graded as G3 from high-grade patients carried amplifications of Network Component 1, whereas tumor areas graded as G3 from intermediate-grade patients did not have amplifications of this network component (Fig. 4c, d). Thus, although the tumor areas were histologically equally classified, tumor areas from high-grade patients carried a CNA signature and expression patterns reminiscent of the high-grade areas from the same patients. Therefore, within the cohort tested the joint DNA amplification of this network component along with RNA up-regulation is a signature of high-grade tumors. Curiously, the higher-grade tumor areas of those high-grades patients (TA1) carried stronger DNA amplifications than the respective lower-grade areas (TA2), which implies that the progressive amplification of Network Component 1 during tumor evolution may contribute to an increasingly aggressive phenotype. In order to further corroborate the clinical relevance of this network perturbation we analyzed published datasets of three additional PCa cohorts (TCGA\textsuperscript{8}, MSKCC\textsuperscript{31}, and Aarhus\textsuperscript{77}), together comprising a total of 513 patients with known clinical outcome. We found, that over-expression of genes from Network Component 1 was a significant predictor of reduced RFS in the TCGA cohort (P value = 8.8e-4, log-rank test), which was the cohort with the largest number of patients. In the other two cohorts we observed the same trend, although the difference in RFS was not statistically significant (P value = 0.11 and 0.093 for MSKCC, and Aarhus; Fig. 4f).

In conclusion, our findings suggest that relatively weak but broad CNAs of entire network components are associated with high-grade tumors and that the presence of some of these perturbations in lower-grade tumors may be predictive of the future development of a more aggressive phenotype.

Analysis of distinct tumor nodules defines intra-patient heterogeneity (TA1 versus TA2 comparison).

The CNA patterns (Supplementary Fig. 4) and the Network Component 1 analysis (Fig. 4c, d) suggest that different tumor areas from the same patient shared several mutations. Such common
signatures are expected if different tumor nodules originate from a common clone. If this was true, we
would expect mutational signatures to be more similar between different nodules from the same patient
than between patients, even though mutated genes may be shared across patients. To compare the intra-
and inter-patient molecular heterogeneity at the levels of CNAs, transcript, and protein FCs, we computed
the Pearson correlation between tumor area 1 (TA1) and its paired tumor area 2 (TA2) for each layer and
all of the 27 patients with two characterized tumor areas (25 for the mRNA, see Methods and
Supplementary Text). As a control, we also computed all pairwise Pearson correlations between the
samples within each of the grade groups (i.e. inter-patient correlation). As expected, paired TA1 and TA2
from the same patient were on average more strongly correlated to each other compared to samples from
different patients within the same grade group. This finding was consistent for all omics layers (Fig. 5a),
and was more pronounced at the CNA and mRNA layers compared to the protein layer.

Next, we tested whether a high correlation at the level of CNA also implies a high correlation at
the level of mRNA and proteins. We tested this idea by ‘correlating the correlations’, i.e. we correlated the
TA1-TA2 correlation of CNA profiles with the correlation between the mRNA and protein profiles of the
same tumor areas (Fig. 5b). Indeed, a higher correlation of two tumor areas at the level of CNA correlated
significantly with a higher correlation at the level of mRNA ($r=0.49$, $P$ value=0.014). In other words,
knowing how similar two tumor areas of a patient are at the CNA level supports a prediction of their
similarity at the mRNA level (and conversely). Although the correlation between protein and CNA was not
statistically significant, it followed the same trend ($r=0.35$, $P$ value=0.076).

Comparing molecular similarity across omics layers allowed us to identify specific types of patients.
The patients H2, H4, M13 had highly correlated tumor areas at all three layers (upper right corner in all
scatterplots of Fig. 5b). Likely, the tumor areas of these patients have a common clonal origin
(Supplementary Fig. 3). In contrast, patients M12 and M14 had weakly correlated tumor areas at all levels
(bottom left corner in all scatterplots of Fig. 5b). These tumor nodules either have independent clonal
origins or they diverged at an earlier stage during tumor evolution (Supplementary Fig. 3). For example,
in the case of patient M12 large parts of the genome were not affected by CNAs in the benign sample as
well as in TA1 and TA2. However, as shown on Supplementary Fig. 3, a large region was amplified in TA1,
whereas the same region was deleted in TA2. This is consistent with a scenario in which TA1 and TA2 show
parallel evolution. A third class of patients is exemplified by the patients M9 and M17, who showed a high
correlation between their tumor areas on the CNA and mRNA levels, but not on the protein level. Yet other
patterns were apparent in patients M6 and H3. They showed very similar protein patterns in the two
tumor areas, but not on the other two layers. We confirmed that protein-level similarity correlated with
similar histological characteristics of the tumor areas. Supplementary Fig. 9 shows formalin-fixed paraffin-embedded (FFPE) tissue microarray images (duplicates) from the analyzed tumor nodules (TA1 and TA2, diameter 0.6 mm), further underlining the hypothesis that ultimately protein-level alterations are responsible for common cellular phenotypes. Although we cannot fully exclude the possibility that some of these results were affected by technical noise in the data, our findings suggest that transcript alterations can frequently be buffered at the level of proteins (patients M9, M17, Supplementary Fig. 6) and that convergent evolutionary processes may lead to the alteration of common proteins (patients M6, H3). We also note that our findings are specific to the two tumor areas available in this study and could be different if other nodules had been sampled for each of the patients. However, our findings on patients with weakly correlated tumor areas at all levels like M12 and M14 suggest that these patients might carry more than one disease.6

Discussion

Despite twenty years of oncological research involving genome-scale (omics) technologies, we know remarkably little about how the discovered genomic alterations affect the biochemical state of a cell and consequently the disease phenotype. In particular, little is known about how genomic alterations propagate along the axis of gene expression17,18. Here, we have exploited recent technological advances in data acquisition that made it possible to characterize small samples of the same tumor specimens at the level of genomes, transcriptomes, and proteomes and advances in computational strategies towards the network-based integration of multi-omics data.

In our study, samples were generated from small, less than 1 mm diameter punches in immediate spatial proximity in the tumor and subsequently profiled at all three ‘omics layers’ (DNA, RNA, proteome). Due to the large spatial heterogeneity of PCa14,25, this design - which is so far uncommon for studies profiling multiple layers from tumor specimens - was instrumental for increasing the comparability of the various omics layers and thus facilitated the analysis of molecular mechanisms. Our key findings are: (1) we confirmed the importance of CNAs for PCa biology and the alteration of many known PCa-associated genes at the transcript- and protein-level; (2) we revealed a generally elevated molecular alteration of high-grade tumors compared to lower-grade tumors; (3) although our study confirmed large within- and between-patient genomic heterogeneity, (4) we detected molecular networks that were commonly altered at the mRNA and protein-level. The fact that many of those target molecules are known drivers of PCa tumorigenesis, supports the notion that these proteins/transcripts are subject to convergent evolutionary mechanisms.
We integrated the three omics layers using a network-based approach as opposed to directly comparing gene perturbations (mutations) to gene products (transcripts and proteins). Using genome data only, it had previously been hypothesized that whereas the identity of specific mutated genes may differ between tumors, those mutations might still affect common molecular networks\textsuperscript{43}. In other words, tumor phenotypes are more dependent on which network regions are perturbed rather than which individual genes are mutated. Our study provides experimental evidence that such network effects are indeed propagated to subsequent molecular layers and that this effect propagation may be clinically relevant.

Our multi-omics network analysis revealed that high-grade PCa tumors distinguished themselves from low-grade tumors in two aspects. The first is a generally higher heterogeneity and loss of controlled gene regulation, which increased the molecular differences among high-grade tumors. It had previously been shown that gene expression in tumors is often less coordinated than in normal samples\textsuperscript{30}. The increasing heterogeneity of protein concentrations suggests that this loss of coordinated expression also affects protein levels. The second aspect is the convergence of molecular alterations towards a consistent set of perturbed specific molecular sub-networks at the genomic, transcriptomic and proteomic layer along the progression from low-grade to high-grade tumors. Thus, although we observed globally a higher degree of variability in gene expression and proteome control among high-grade specimens, a specific subset of the observed alterations appeared crucial for determining the aggressive tumor phenotype.

Tumors are under selective pressure acting on the biochemical function of the cells. It is generally believed that proteins are a closer reflection of the functional state of a cell than the mRNA. Here we could show that the fold changes of proteins like RABL3, MFAP4, and SF3B2 were more pronounced and/or uniform across high-grade tumors than either their coding mRNAs or the underlying CNAs.

Specifically, our analysis led to the identification of Network Component 1, a sub-network involved in chromatin remodeling and consisting of genes that were weakly amplified in intermediate-grade (G3) tumor specimens. Signals of individual gene members of this component were virtually indistinguishable from noise in our cohort. However, their consistent alterations across the network region, across molecular layers and the fact that the same genes showed enhanced signals in high-grade specimens, rendered this component highly interesting. The fact that expression changes of Network Component 1 members were predictive for survival in independent cohorts further supports the potential clinical relevance of this sub-network. Our network-based cross-omics analysis identified nine other network components (Fig. 4) successfully capturing several known and potentially new PCa-associated genes. However, neither Network Component 1 nor any of the other network components was uniformly subject to CNAs across all high-grade patients. Instead, we found different network components modified in
different patients and these sub-networks were involved in cellular processes as diverse as actin remodeling, DNA damage response, and metabolic functions, all of which are known contributors to PCa biology. This further underlines the large patient-to-patient variability of PCa and it demonstrates the diversity of molecular mechanisms leading to histologically similar phenotypes. Future prediction models of PCa including the ISUP grade groups, PSA levels and clinical stage might be improved by exploiting multi-omics network analyses. Detecting more or less aggressive networks in prostate biopsies would help clinicians to advice either active surveillance or active therapy. However, the development of such multidimensional biomarkers would require much larger patient cohorts.

Another distinguishing feature of this study was the simultaneous profiling of two different tumor regions in 27 out of the 39 patients. The profiling of multiple tumor regions from the same prostate helped to further highlight the enormous heterogeneity of PCa within and between patients and provided important insights into PCa evolution. The fact that Network Component 1 was more strongly affected in the paired higher-grade nodules of high-grade patients suggests that at least certain sub-networks are subject to an evolutionary process, that progressively ‘moves’ protein levels towards a more aggressive state. Generally, and at all molecular layers tested, the two paired tumor areas were more similar to each other compared to two samples from the same grade group but different patients, suggesting common evolutionary origins. Although the two tumor areas seemed to mostly originate from the same clone, this was not always the case. In some patients, different nodules exhibited different molecular patterns at all omics layers, suggesting early evolutionary separation. Thus, for the first time, current diagnostic, expert-level consensus guidelines are supported by detailed proteogenomic data. Our findings support earlier claims that clonality itself might be a prognostic marker with implications for future, more tumor-specific treatment when targeted therapies become available also for PCa.

Our study shows that all three molecular layers (genome, transcriptome and proteome) contributed valuable information for understanding the biology of PCa. In particular the DNA layer informed about causal events, clonality, and genomic similarity between tumors. The transcriptome was relevant for understanding the transmission of CNA effects to proteins and served as a surrogate in cases where protein levels remained undetected. The proteome was crucial for revealing protein-level buffering of CNA effects as well as for indicating convergent evolution on functional endpoints. In a routine diagnostic context though, measuring all three layers may not be feasible for the near future due to resource and time limitations. Thus, the identification of improved, routine Usable molecular markers for PCa diagnostics and prognosis remains an open problem.
Materials and methods

Patients and samples

A total of 39 men with localized PCa who were scheduled for RP were selected from a cohort of 1,200 patients within the ProCOC study and processed at the Department of Pathology and Molecular Pathology, University Hospital Zurich, Switzerland. Each of the selected intermediate- and high-grade patients had two different tumor nodules with different ISUP grade groups. H&E (Hematoxylin and Eosin)-stained fresh frozen tissue sections of 105 selected BPH and tumor regions were evaluated by two experienced pathologists (PJW, NJR) to assign malignancy, tumor stage, and Grade Group according to the International Union Against Cancer (UICC) and WHO/ISUP criteria. This study was approved by the Cantonal Ethics Committee of Zurich (KEK-ZH-No. 2008-0040), the associated methods were carried out in accordance with the approved guidelines, and each patient has signed an informed consent form. Patients were followed up on a regular basis (every three months in the first year and at least annually thereafter) or on an individual basis depending on the disease course in the following years. The RFS was calculated with a biochemical recurrence (BCR) defined as a PSA $\geq$0.1 ng/ml. Patients were censored if lost to follow-up or event-free at their most recent clinic visit. Patients with a postoperative PSA persistence or without distinct follow-up data for the endpoint BCR were excluded from the analysis of BCR.

Exome sequencing and somatic variant analysis

The exome sequencing (exome-seq) was performed using the Agilent Sure Select Exome platform for library construction and Illumina HiSeq 2500 for sequencing read generation. We mapped and processed the reads using a pipeline based on bowtie2\textsuperscript{79} (1.1.1) and the Genome Analysis Tools Kit (GATK)\textsuperscript{80} (3.2-2). We detected and reported nonsynonymous variants or variants causing splicing changes using Strelka (1.0.14) and Mutect (1.1.7) combined with post-processing by the CLC Genomics Workbench (8.0.3). In this process, all tissue samples of a patient were compared to the respective blood sample. Trimmomatic\textsuperscript{81} (0.36) was used for adaptor clipping and low-quality subsequence trimming of the FASTQ files. Subsequently, single reads were aligned to the hg19 reference genome with bowtie2 with options “--very-sensitive -k 20”. We applied samtools\textsuperscript{82} (0.1.19) and picard-tools (1.119) to sort the resulting bam files in coordinate order, merge different lanes, filter out all non-primary alignments, and remove PCR duplicates. Quality of the runs was checked using a combination of BEDtools\textsuperscript{83} (2.21), samtools, R (3.1) and FastQC (0.11.2).
Bam files containing the mapped reads were preprocessed in the following way: indel information was used to realign individual reads using the RealignerTargetCreator and IndelRealigner option of the GATK. Mate-pair information between mates was verified and fixed using Picard tools and single bases were recalibrated using GATK’s BaseRecalibrator. After preprocessing, variant calling was carried out by comparing benign or tumor prostate tissue samples with matched blood samples using the programs MuTect and Strelka independently. Somatic variants that were only detected by one of the two programs were filtered out using CLC Genomics Workbench. So were those that had an entry in the dbSNP common database and those that represented synonymous variants without predicted effects on splicing.

CNA analysis of exome-seq data

The Bam files generated during the process of somatic variant calling were processed with the CopywriteR package (v.2.2.0) for the R software. CopywriteR makes use of so-called “off-target” reads, i.e. reads that cover areas outside of the exon amplicons. “Off-target” reads are produced due to inefficient enrichment strategies. In our case on average 28.5% of the total reads were not on target. Briefly, CopywriteR removes low quality and anomalous read pairs, then peaks are called in the respective blood reference, and all reads in this region are discarded. After mapping the reads into bins, those peak regions, in which reads had been removed, were compensated for. Additionally, read counts are corrected based on mappability and GC-content. Finally, a circular binary segmentation is carried out and for each segment the log count ratios between tissue samples and the respective blood sample are reported as copy number gain or loss. The copy number of each gene in each sample was reported based on the log count ratio of the respective segment in which the gene was located. The overall performance of this CNA-calling approach was evaluated by comparing the results of the TA1 (and TA) samples with CNA results obtained by applying the OncoScan Microarray pipeline to FFPE samples from the same tumors (Supplementary Fig. 10).

OncoScan Microarrays

OncoScan copy number assays were carried out and analyzed as described previously. Briefly, DNA was extracted from punches of FFPE cancer tissue blocks. Locus-specific molecular inversion probes were hybridized to complementary DNA and gaps were filled in a nucleotide-specific manner. After amplification and cleavage of the probes, the probes were hybridized to the OncoScan assay arrays. Scanning the fluorescence intensity and subsequent data processing using the Affymetrix® GeneChip®
Command Console and BioDiscovery Nexus express resulted in log intensity ratio data (sample versus Affymetrix reference) and virtual segmentation of the genome into areas with copy number gain, loss or stability.

**RNA Sequencing**

RNA sequencing was performed at the Functional Genomics Center Zurich. RNA-seq libraries were generated using the TruSeq RNA stranded kit with PolyA enrichment (Illumina, San Diego, CA, USA). Libraries were sequenced with 2x126bp paired-end on an Illumina HiSeq 2500 with an average of 105.2 mio reads per sample.

Paired-end reads were mapped to the human reference genome (GRCh37) using the STAR aligner (version 2.4.2a). Quality control of the resulting bam files using QoRTs and mRIN showed strong RNA degradation in a significant fraction of the samples: mRIN classified 31 samples as highly degraded (Supplementary Fig. 11, Supplementary Table 4). In order to correct for this 3’ bias, 3 tag counting was performed as described by Sigurgeirsson et al using a tag length of 1,000. After 3’ bias correction, three samples still showed a clear 3’ bias: the two tumor regions (TA1 and TA2) of the patient M5 and TA2 from patient M8 (Supplementary Fig. 11). These samples were excluded from subsequent analyses. Additionally, the BPH region of the patient M5 was excluded due to the exclusion of both its tumor regions.

FeatureCounts was used to determine read counts for all genes annotated in ENSEMBL v75. Genes for which no read was observed in any of the samples in the original data were excluded from the analysis. Further, after 3 tag counting, all genes with without at least 1 read per million in N of the samples were removed. We chose N to be 10 which corresponds to the size of the smallest grade group (G2). In a last reduction step, all genes with more than one transcript were excluded, yielding a final set of 14,281 genes.

Read count normalization and differential gene expression analysis was performed using the R packages sva and DESeq2. All benign tissues were considered biological replicates and differential gene expression for the individual tumor samples was determined against all benign tissues. Gene expression changes with an adjusted P value < 0.1 were considered significant.

**RNA-seq - 3’ bias correction**

The 3 tag counting approach for 3’ bias correction was used on the RNA-seq dataset. This approach requires changing of the annotation file in two steps: 1) isoform filtering and 2) transcript length restriction. As proposed for each gene we determined the highest expressed isoform within a set of
high quality samples. As high quality samples we used all samples with an mRIN score greater than or equal to 0.02. This set contains 7 benign and 15 tumor samples. Isoform expression was determined using cufflinks\textsuperscript{97}. As transcript length we chose 1,000bp.

**Gene fusions**

FusionCatcher (version 0.99.5a beta) was used to determine gene fusions for all samples. Fusions classified as “probably false positive” are discarded unless they are also classified as “known fusion”.

**PCT assisted sample preparation for SWATH-MS**

We first washed each tissue sample to remove O.C.T., followed by PCT-assisted tissue lysis and protein digestion, and SWATH-MS analysis, as described previously\textsuperscript{24} Briefly, a series of ethanol solutions were used to wash the tissues each tissue, including 70% ethanol / 30% water (30 s), water (30 s), 70% ethanol / 30% water (5 min, twice), 85% ethanol / 15% water (5 min, twice), and 100% ethanol (5 min, twice). Subsequently, the tissue punches were lysed in PCT-MicroTubes with PCT-MicroPestle\textsuperscript{98} with 30 µl lysis buffer containing 8 M urea, 0.1 M ammonium bicarbonate, Complete protease inhibitor cocktail (Roche) and PhosSTOP phosphatase inhibitor cocktail (Roche) using a barocycler (model NEP2320-45k, PressureBioSciences, South Easton, MA). The lysis was performed with 60 cycles of high pressure (45,000 p.s.i., 50 s per cycle) and ambient pressure (14.7 p.s.i., 10 s per cycle). The extracted proteins were then reduced and alkylated prior to lys-C and trypsin-mediated proteolysis under pressure cycling. Lys-C (Wako; enzyme-to-substrate ratio, 1:40) -mediated proteolysis was performed using 45 cycles of pressure alternation (20,000 p.s.i. for 50 s per cycle and 14.7 p.s.i. for 10 s per cycle), followed by trypsin (Promega; enzyme-to-substrate ratio, 1:20)-mediated proteolysis using the same cycling scheme for 90 cycles. The resultant peptides were cleaned using SEP-PAC C18 (Waters Corp., Milford, MA) and analyzed, after spike-in 10% iRT peptides\textsuperscript{51}, using SWATH-MS following the 32-fixed-size-window scheme as described previously\textsuperscript{19, 21} using a 5600 TripleTOF mass spectrometer (Sciex) and a 1D+ Nano LC system (Eksigent, Dublin, CA). The LC gradient was formulated with buffer A (2% acetonitrile and 0.1% formic acid in HPLC water) and buffer B (2% water and 0.1% formic acid in acetonitrile) through an analytical column (75 µm \times 20 cm) and a fused silica PicoTip emitter (New Objective, Woburn, MA, USA) with 3-µm 200 Å Magic C18 AQ resin (Michrom BioResources, Auburn, CA, USA). Peptide samples were separated with a linear gradient of 2% to 35% buffer B over 120 min at a flow rate of 0.3 µl min\textsuperscript{-1}. Ion accumulation time for MS1 and MS2 was set at 100 ms, leading to a total cycle time of 3.3 s.
To build a comprehensive library for SWATH data analysis, we analyzed unfractionated prostate tissue digests prepared by the PCT method using Data Dependent Acquisition (DDA) mode in a tripleTOF mass spectrometer over a gradient of 2 hours as described previously. We spiked iRT peptides into each sample to enable retention time calibration among different samples. We then combined these data with the DDA files from the pan-human library project. All together we analyzed 422 DDA files using X!Tandem and OMSSA against three protein sequence databases downloaded on Oct 21, 2016 from UniProt, including the SwissProt database of curated protein sequences (n=20,160), the splicing variant database (n=21,970), and the trembl database (n=135,369). Using each database, we built target-decoy protein sequence database by reversing the target protein sequences. We allowed maximal two missed cleavages for fully tryptic peptides, and 50 p.p.m. for peptide precursor mass error, and 0.1 Da for peptide fragment mass error. Static modification included carbamidomethyl at cysteine, while variable modification included oxidation at methionine. Search results from X!Tandem and OMSSA were further analyzed through Trans-Proteomic Pipeline (TPP, version 4.6.0) using PeptideProphet and iProphet, followed by SWATH assay library building procedures as detailed previously. Altogether, we identified 167,402 peptide precursors, from which we selected the proteins detected in prostate tissue samples, and built a sample-specific library. SWATH wiff files were converted into mzXML files using ProteoWizard msconvert v.3.0.3316, and then mzML files using OpenMS tool FileConverter. OpenSWATH was performed using the tool OpenSWATHWorkflow with input files including the mzXML file, the TraML library file, and TraML file for iRT peptides.

To obtain consistent quantification of the SWATH files, we obtained the all annotated b and y fragments from the sp, sv and tr libraries. About ten thousand redundant and low-quality assays were removed. Then we extracted the chromatography of these fragments and MS1 signals using OpenSWATHWorkflow, followed by curation using DIA-expert. Briefly, the chromatography of all fragments and MS1 signals were subject to scrutiny by empirically developed expert rules. A reference sample with best q value by pyprophet was picked up to refined fragments. The peptide precursors are further filtered based on the following criteria: i) remove peptide precursors with a q value higher than 1.7783e-06 to achieve a false discovery rate of 0.00977 at peptide level using SWATH2stats; ii) peptides with a FC higher than 2 between the reference sample and its technical replicate were removed; iii) peptides matching to multiple SwissProt protein sequences were removed. The data matrix was first
quantile normalized, \( \log_2 \) transformed, followed by batch correction using the ComBat R package. Finally, for each protein and pair of technical replicates the average value was computed.

**Statistical analysis**

All plots were produced with R. Kaplan-Meier estimators were used for RFS analysis. Differences between survival estimates were evaluated by the log-rank test.

**Computation of molecular perturbation scores**

On the genomic level (mutation and CNA), we kept the tumor samples (66 in total) that contain FCs with respect to the blood. The mutation matrix was further discretized by setting all non-zero events to 1. At the transcriptomics level, the FCs for the 63 tumor samples were computed as described above (see ‘RNA Sequencing’). Finally, on the proteomics level, we computed the FCs for the tumor samples (66 in total) as follows: for each protein, its mean intensity over the normal samples was subtracted from the intensities of the tumor samples. (We chose to compute the FCs for the tumor samples with respect to a global reference (average of all normal samples) and not with respect to their paired benign sample in order to achieve a higher consistency with the transcriptomics level.)

We assigned to each sample two molecular perturbation scores summarizing/quantifying the magnitude of its FCs: DE_count counts the number of mutated/differentially expressed (DE) genes, while the DE_sum score is the sum of absolute FCs of all genes. Thus, while the first score counts the number of events (mutations/DE genes), the second one quantifies their magnitude. A gene is regarded as mutated/DE if its value is 1 in the mutation layer and if its absolute value is above a threshold that has been set to 1 for the mRNA and protein layer. For the CNA layer, the corresponding threshold was set to 0.5 because the range of FCs in the CNA matrix is smaller than the mRNA and protein matrices. Both types of scores were computed for each molecular level, except for the point mutations where only DE_count was computed. Afterwards, the 66 DE_count scores (63 for the mRNA) and the DE_sum scores at each layer were divided into the four grade groups G1, G2, G3 and G4/5 respectively.

**Network propagation/smoothing**

As a network, the STRING gene interaction network (version 10) was used, after removing all edges with combined score smaller or equal to 900 and keeping subsequently the largest connected component. The resulting network consisted of 10,729 nodes and 118,647 (high-confidence) edges. For the network smoothing, the weight matrix was computed as described in Vanunu et al. but for an
unweighted graph and the propagation parameter was set to 0.5. The propagation was iteratively repeated 500 times to ensure convergence of the results. For the mapping from gene symbols to STRING identifiers (Supplementary Table 6) we used the R/Bioconductor package STRINGdb106. The gene symbols with no matching STRING identifier were removed, while for those that mapped to multiple STRING identifiers, the first mapping was kept (default choice in the package). From the multiple gene symbols that mapped to the same STRING identifier, the first mapping was kept. The genes that were not present in the network were removed from the datasets, while those that were present in the network but not in the corresponding dataset were initially filled in with 0’s.

Genes with very small, ‘smoothed’ (absolute) FCs were filtered out as follows: after the network propagation, only network nodes that had protein measurements themselves or at least one direct neighbor (on the filtered STRING network) with protein measurements were considered in the next steps of this analysis. I.e. network nodes without measured FCs at the protein layer that had no direct neighbor with measured protein values were removed from the subsequent analyses.

For significance testing, the one-sided Wilcoxon rank sum test comparing the smoothed FCs between the groups G4/5 and G1 was applied to each network node (after filtering) and layer, once for up-regulation and once for down-regulation. The resulting sub-networks (up-regulated and down-regulated) consisted of those genes that were significant (P value below 0.05) at all three layers and all of the edges connecting them on the filtered STRING network.

**Network Component 1 analysis**

For each tumor sample at the CNA layer, a one-sided, one-sample t-test has been applied testing if its average FC over the genes of the Network Component 1 (and in particular those that have been measured at the CNA) is significantly greater than 0. Due to the presence of outliers in some samples, the non-parametric, one-sided Wilcoxon signed-rank test has been applied as well yielding very similar results (data not shown). A result is considered to be significant if the corresponding P value is below 0.05. The analysis has been repeated for the mRNA and protein layer.

**Independent cohorts validation**

For the validation of Network Component 1, we used published datasets of three PCa cohorts: TCGA, MSKCC, and Aarhus. For TCGA and MSKCC, we downloaded the mRNA data- consisting of 493 and 150 samples respectively- with precomputed z-scores per gene, and corresponding clinical data from
cBioPortal\cite{cBioPortal} (https://www.cbioportal.org/). There were 333 samples with available disease/progression-free survival time in TCGA and 140 samples in MSKCC and only these were used in the subsequent survival analysis.

For the Aarhus study (NCBI GEO dataset GSE46602), we downloaded the mRNA matrix and corresponding clinical information as described in Ycart et al\cite{Ycart}. The resulting mRNA matrix consisted of 20,186 genes and 50 samples- 40 PCa samples with known RFS time and 10 benign samples. Once excluding the benign samples, we computed z-scores per gene in order to have comparable values with the other two studies. These 40 PCa samples were also considered in the subsequent survival analysis.

We reduced all three datasets to the nine genes of Network Component 1. In each of the datasets, we computed for each sample an average z-score across the nine genes of Network Component 1 (combined risk score). Subsequently, we used these combined risk scores to split the samples of each study into two groups: samples with a combined risk score larger or equal to the median combined risk score of the study were considered as ‘altered’ and the rest as ‘unaltered’. Kaplan-Meier curves were generated for the two groups.

**Mutual information**

In order to compute the MI between each individual sample within a group against the matching centroid of the same group, the R package infotheo \cite{infotheo} has been used. For the MI, the data needs to be discrete. For that, we discretize the tumor-to-benign FCs and centroids by setting 1, if the value is above a threshold \( \text{thr} \), -1, if the value is below \(-\text{thr}\) and 0 else. (\( \text{thr} \) is equal to 1 for the mRNA and proteins and 0.5 for the CNAs.)

**Acknowledgments**

This work was supported by the SystemsX.ch project PhosphoNet PPM (to R.A. and P.J.W.), the Swiss National Science Foundation (grant no. 3100A0-688 107679 to R.A.), the European Research Council (grant no. ERC-2008-AdG 233226 and ERC-2014-AdG670821 to R.A.), the Foundation for Scientific Research at the University of Zurich (to P.J.W.), the Westlake Startup Grant (to T.G.), Zhejiang Provincial Natural Science Foundation of China (Grant No. LR19C050001 to T.G.), Hangzhou Agriculture and Society Advancement Program (Grant No. 20190101A04 to T.G.). A.B. and K.C. are supported by the German Federal Ministry of Education and Research Grants Sybacol and PhosphoNetPPM.
**Author contributions**

A.B., T.G., P.J.W. and R.A. designed the project. P.J.W., T.G., Q.Z., C.E.F., N.J.R., A.C, D.R, J.H.R., C.F., K.S., C.P., T.H., A.L.M. and C.B. procured the samples and performed the experiments. K.C., T.G., Q.Z., U.W., R.S., N.C.T, K.O., L.C., L.M., M.R.M, M.M and A.B. designed and performed the statistical analyses with critical inputs from C.Y., H.C., Q.Z., Y.Z., M.H. and other authors. K.C., A.B., T.G. and R.A. interpreted the results. K.C., T.G., P.J.W., A.B. and R.A. wrote the manuscript with inputs from all co-authors. A.B., R.A., P.J.W. and T.G. supported and supervised the project.

**Competing interests**

R.A. holds shares of Biognosys AG, which operates in the field covered by the article. The research groups of R.A. and T.G. are supported by SCIEX, which provides access to prototype instrumentation, and Pressure Biosciences Inc., which provides access to advanced sample preparation instrumentation.
References

1. Ferlay, J., et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**, E359-386 (2015).

2. Epstein, J.I., et al. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol* **69**, 428-435 (2016).

3. Gordetsky, J. & Epstein, J. Grading of prostatic adenocarcinoma: current state and prognostic implications. *Diagn Pathol* **11**, 25 (2016).

4. Berger, M.F., et al. The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-220 (2011).

5. Barbieri, C.E., et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* **44**, 685-689 (2012).

6. Robinson, D., et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228 (2015).

7. Baca, S.C., et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-677 (2013).

8. Cancer Genome Atlas Research, N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011-1025 (2015).

9. Wedge, D.C., et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat Genet* **50**, 682-692 (2018).

10. Tomlins, S.A., et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648 (2005).

11. Grasso, C.S., et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-243 (2012).

12. Boutros, P.C., et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* **47**, 736-745 (2015).

13. Cooper, C.S., et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet* **47**, 367-372 (2015).

14. Lovf, M., et al. Multifocal Primary Prostate Cancer Exhibits High Degree of Genomic Heterogeneity. *Eur Urol* **75**, 498-505 (2019).

15. Lindberg, J., et al. Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins. *Eur Urol* **63**, 347-353 (2013).

16. Espiritu, S.M.G., et al. The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* **173**, 1003-1013 e1015 (2018).

17. Sinha, A., et al. The Proteogenomic Landscape of Curable Prostate Cancer. *Cancer Cell* **35**, 414-427 e416 (2019).

18. Latonen, L., et al. Integrative proteomics in prostate cancer uncovers robustness against genomic and transcriptomic aberrations during disease progression. *Nat Commun* **9**, 1176 (2018).

19. Iglesias-Gato, D., et al. The Proteome of Primary Prostate Cancer. *Eur Urol* **69**, 942-952 (2016).

20. Attard, G., et al. Prostate cancer. *Lancet* **387**, 70-82 (2016).

21. Goncalves, E., et al. Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. *Cell Syst* **5**, 386-398 e384 (2017).

22. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535-550 (2016).

23. Liu, Y., et al. Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. *Nat Commun* **8**, 1212 (2017).
Proteogenomic heterogeneity of prostate cancer progression

24. Guo, T., et al. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat Med* **21**, 407-413 (2015).

25. Guo, T., et al. Multi-region proteome analysis quantifies spatial heterogeneity of prostate tissue biomarkers. *Life Sci Alliance* **1**(2018).

26. Umbehr, M., et al. ProCOC: the prostate cancer outcomes cohort study. *BMC Urol* **8**, 9 (2008).

27. Wettstein, M.S., et al. Prognostic Role of Preoperative Serum Lipid Levels in Patients Undergoing Radical Prostatectomy for Clinically Localized Prostate Cancer. *Prostate* **77**, 549-556 (2017).

28. Epstein, J.I., et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *The American journal of surgical pathology* **40**, 244-252 (2016).

29. Epstein, J.I., Amin, M.B., Reuter, V.E. & Humphrey, P.A. Contemporary Gleason Grading of Prostatic Carcinoma: An Update With Discussion on Practical Issues to Implement the 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *The American journal of surgical pathology* **41**, e1-e7 (2017).

30. Reznik, E. & Sander, C. Extensive decoupling of metabolic genes in cancer. *PLoS Comput Biol* **11**, e1004176 (2015).

31. Taylor, B.S., et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22 (2010).

32. Zhang, B., et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382-387 (2014).

33. Geiger, T., Cox, J. & Mann, M. Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS Genet* **6**, e1001090 (2010).

34. Juschke, C., et al. Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. *Genome Biol* **14**, r133 (2013).

35. Stingele, S., et al. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol Syst Biol* **8**, 608 (2012).

36. Cover, T.M. & Thomas, J.A. *Elements of information theory*, (John Wiley & Sons, 2012).

37. Zhang, J.S., Gong, A., Cheville, J.C., Smith, D.I. & Young, C.Y. AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. *Genes Chromosomes Cancer* **43**, 249-259 (2005).

38. Liu, Q., et al. Malate dehydrogenase 2 confers docetaxel resistance via regulations of JNK signaling and oxidative metabolism. *Prostate* **73**, 1028-1037 (2013).

39. Yang, J., et al. Integrated analysis of microfibrillar-associated proteins reveals MFAP4 as a novel biomarker in human cancers. *Epigenomics* **11**, 1635-1651 (2019).

40. Senga, S., Kobayashi, N., Kawaguchi, K., Ando, A. & Fujii, H. Fatty acid-binding protein 5 (FABP5) promotes lipolysis of lipid droplets, de novo fatty acid (FA) synthesis and activation of nuclear factor-kappa B (NF-kappaB) signaling in cancer cells. *Biochim Biophys Acta Mol Cell Biol Lipids* **1863**, 1057-1067 (2018).

41. Pan, Y., et al. The overexpression of Rabl3 is associated with pathogenesis and clinicopathologic variables in hepatocellular carcinoma. *Tumour Biol* **39**, 1010428317696230 (2017).

42. Zhang, W., Sun, J. & Luo, J. High Expression of Rab-like 3 (Rabl3) is Associated with Poor Survival of Patients with Non-Small Cell Lung Cancer via Repression of MAPK8/9/10-Mediated Autophagy. *Med Sci Monit* **22**, 1582-1588 (2016).

43. Hofree, M., Shen, J.P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nat Methods* **10**, 1108-1115 (2013).

44. Szklarczyk, D., et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-452 (2015).
Proteogenomic heterogeneity of prostate cancer progression

822

821

820

818

815

814

812

811

810

809

808

807

806

805

804

803

802

801

800

799

798

797

796

795

794

793

792

791

790

789

788

787

786

785

784

783

782

781

779

66.

65.

64.

63.

61.

59.

58.

56.

55.

54.

53.

52.

51.

50.

49.

48.

47.

46.

45.

44.

43.

42.

41.

40.

39.

38.

37.

36.

35.

34.

33.

32.

31.

30.

29.

28.

27.

26.

25.

24.

23.

22.

21.

20.

19.

18.

17.

16.

15.

14.

13.

12.

11.

10.

9.

8.

7.

6.

5.

4.

3.

2.

1.

45. Zhou, D., Bousquet, O., Lal, TN, Weston, J., Scholkopf B. Learning with local and global consistency. Advances in Neural Information Processing Systems 16, 312-328 (2004).

46. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. Plos Comput Biol 6, e1000641 (2010).

47. Baude, A., et al. Hepatoma-derived growth factor-related protein 2 promotes DNA repair by homologous recombination. Nucleic Acids Research 44, 2214-2226 (2015).

48. Guler, G.D., et al. Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from Lethal Drug Exposure. Cancer Cell 32, 221 (2017).

49. Sleazak, J., Truong, M., Huang, W. & Jarrard, D. HP1gamma expression is elevated in prostate cancer and is superior to Gleason score as a predictor of biochemical recurrence after radical prostatectomy. BMC Cancer 13, 148 (2013).

50. Ceol, C.J., et al. The SETDB1 histone methyltransferase is recurrently amplified in and accelerates melanoma. Nature 471, 513-517 (2011).

51. Sun, Y., et al. Histone methyltransferase SETDB1 is required for prostate cancer cell proliferation, migration and invasion. Asian J Androl 16, 319-324 (2014).

52. Luo, H., Cowen, L., Yu, G., Jiang, W. & Tang, Y. SMG7 is a critical regulator of p53 stability and function in DNA damage stress response. Cell Discovery 2, 15042 (2016).

53. Karanika, S., et al. Targeting DNA Damage Response in Prostate Cancer by Inhibiting Androgen Receptor-CDC6-ATR-Chk1 Signaling. Cell Reports 18, 1970 (2017).

54. Diouf, B., et al. Somatic deletions of genes regulating MSH2 protein stability cause DNA mismatch repair deficiency and drug resistance in human leukemia cells. Nature Medicine 17, 1298-1303 (2011).

55. Yao, S., et al. PRKCζ Expression Promotes the Aggressive Phenotype of Human Prostate Cancer Cells and Is a Novel Target for Therapeutic Intervention. Genes Cancer 1, 444-464 (2010).

56. Zhou, W., et al. Low-frequency nonsynonymous variants in FKBPL and ARPC1B genes are associated with breast cancer risk in Chinese women. Mol Carcinog 56, 774-780 (2017).

57. De Robertis, M., et al. EphB2 stem-related and EphA2 progression-related miRNA-based networks in progressive stages of CRC evolution: clinical significance and potential miRNA drivers. Mol Cancer 17, 169 (2018).

58. Zeng, Z., Yang, H. & Xiao, S. ACTL6A expression promotes invasion, metastasis and epithelial mesenchymal transition of colon cancer. BMC Cancer 18, 1020 (2018).

59. Castro, M.A., et al. CFL1 expression levels as a prognostic and drug resistance marker in nonsmall cell lung cancer. Cancer 116, 3645-3655 (2010).

60. Overbye, A., et al. Identification of prostate cancer biomarkers in urinary exosomes. Oncotarget 6, 30357-30376 (2015).

61. Deng, Z., et al. Yin Yang 1 regulates the transcriptional activity of androgen receptor. Oncogene 328, 3746-3757 (2009).

62. Fournier, P.G., et al. The TGF-beta Signaling Regulator PMEPA1 Suppresses Prostate Cancer Metastases to Bone. Cancer Cell 27, 809-821 (2015).

63. Yang, J., et al. Inhibition of KPNA4 attenuates prostate cancer metastasis. Oncogene 36, 2868-2878 (2017).

64. Liu, W., et al. Homozygous deletions and recurrent amplifications implicate new genes involved in prostate cancer. Neoplasia 10, 897-907 (2008).

65. Trasino, S.E., Harrison, E.H. & Wang, T.T. Androgen regulation of aldehyde dehydrogenase 1A3 (ALDH1A3) in the androgen-responsive human prostate cancer cell line LNCaP. Exp Biol Med (Maywood) 232, 762-771 (2007).

66. Bonaccorsi, L., et al. Androgen receptor regulation of the seladin-1/DHCR24 gene: altered expression in prostate cancer. Lab Invest 88, 1049-1056 (2008).
Proteogenomic heterogeneity of prostate cancer in high grade and type II endometrial cancer. Noske, A. Kuilman, T. (2001).

Sherry, S.T. sample pairs. cancer samples. Bioinformatics Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. (2009).

Li, Bioinformatics Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. generation DNA sequencing data. McKenna, A. Genome research Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat Genet 45, 357-359 (2012).

Lee, R.S. & Roberts, C.W. Linking the SWI/SNF complex to prostate cancer. Nat Genet 45, 1268-1269 (2013).

Chuu, C.-P., Lin, C.-Y., Huang, S.-H. & Tsai, K.K.-C. MO2-10-6ACTL6A is a novel oncogene and prognostic biomarker for prostate cancer. Annals of Oncology 30 (2019).

Cai, Y., et al. The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. J Biol Chem 280, 13665-13670 (2005).

Mortensen, M.M., et al. Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy. Sci Rep 5, 16018 (2015).

Hussain, M., et al. LBA12_PRPROfound: Phase III study of olaparib versus enzalutamide or abiraterone for metastatic castration-resistant prostate cancer (mCRPC) with homologous recombination repair (HRR) gene alterations. Annals of Oncology 30 (2019).

Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359 (2012).

Mckenna, A., et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20, 1297-1303 (2010).

Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120 (2014).

Li, H., et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).

Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842 (2010).

Cibulskis, K., et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31, 213-219 (2013).

Saunders, C.T., et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811-1817 (2012).

Sherry, S.T., et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29, 308-311 (2001).

Kuillman, T., et al. CopywriteR: DNA copy number detection from off-target sequence data. Genome Biol 16, 49 (2015).

Noske, A., et al. Detection of CCNE1/URI (19q12) amplification by in situ hybridisation is common in high grade and type II endometrial cancer. Oncotarget 8, 14794-14805 (2017).

Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21 (2013).
90. Hartley, S.W. & Mullikin, J.C. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* **16**, 224 (2015).
91. Feng, H., Zhang, X. & Zhang, C. mRIN for direct assessment of genome-wide and gene-specific mRNA integrity from large-scale RNA-sequencing data. *Nat Commun* **6**, 7816 (2015).
92. Shao, W., et al. Comparative analysis of mRNA and protein degradation in prostate tissues indicates high stability of proteins. *Nat Commun* **10**, 2524 (2019).
93. Sigurgeirsson, B., Emanuelsson, O. & Lundeberg, J. Sequencing degraded RNA addressed by 3' tag counting. *PLoS One* **9**, e91851 (2014).
94. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
95. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883 (2012).
96. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
97. Trapnell, C., et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515 (2010).
98. Shao, S., et al. Reproducible Tissue Homogenization and Protein Extraction for Quantitative Proteomics Using MicroPestle-Assisted Pressure-Cycling Technology. *J Proteome Res* **15**, 1821-1829 (2016).
99. Guo, T., et al. Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines. *iScience* **21**, 664-680 (2019).
100. Röst, H.L., et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **32**, 219-223 (2014).
101. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
102. Szklarczyk, D., et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43**, D447-D452 (2015).
103. Franceschini, A., et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**, D808-815 (2013).
104. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating Genes and Protein Complexes with Disease via Network Propagation. *Plos Computational Biology* **6**(2010).
105. Gao, J., et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).
106. Ycart, B., Charmpi, K., Rousseaux, S. & Fourniè, J.-J. Large scale statistical analysis of GEO datasets. *Sci Signal* **4**, 113:111-119 (2014).
Figure legends

**Figure 1.** Proteogenomics analysis of 105 tissue regions from 39 PCa patients. a, Representative immunohistochemistry images of prostate tissues and the selection of BPH and tumorous tissue regions for genome, transcriptome and proteome analysis. b, Kaplan-Meier curves for our cohort when the patients are stratified by the overall grade (left), the TA1 or TA grade group (middle) and the TA2 or TA grade group (right). Point-wise 95% confidence bands are shown for the whole range of time values.

**Figure 2.** Molecular perturbation scores for point mutations, CNAs, transcriptome and proteome data. a, Distributions of the first type of molecular perturbation scores (DE_count’s) for the four grade groups (visualized as violin plots) at the mutation layer (upper left), CNA layer (upper right), mRNA layer (lower left) and protein layer (lower right). Points represent the actual values. The horizontal lines correspond to the median value in each of the four grade groups. b, Distributions of the second type of molecular perturbation scores (DE_sum’s) for the four grade groups (visualized as violin plots) at the CNA layer (upper left), mRNA layer (upper right) and protein layer (lower left). Points represent the actual values. The horizontal lines correspond to the median value in each of the four grade groups. P values (in each of the titles) show the significance of the one-sided Wilcoxon rank sum test where the values of G3 and G4/5 are gathered together and compared to the values of G1 and G2 (also gathered together).

**Figure 3.** Within-group similarity at the different layers quantified by different similarity measures. a, b, Distributions of the similarity scores between the individual tumor samples and the centroid using the Pearson correlation (a) and MI (b) for the four grade groups (visualized as violin plots) at the CNA (upper left), mRNA (upper right) and protein (lower left) layers. Points represent the actual values. The horizontal lines correspond to the median value in each of the four grade groups. P values from the one-sided Wilcoxon rank sum test comparing G4/5 versus G1: 0.0014 for the CNA, 0.89 for the mRNA and 0.053 for the protein layer with the Pearson correlation, and 0.027 for the CNA, 0.0052 for the mRNA and 0.0081 for the protein layer with the MI. c, Density plots of the FCs in the four grade groups for 3 selected proteins among the 10 highest scoring (score: mean of the absolute FCs across all tumor samples) proteins. Vertical lines correspond to the average FC in each of the four grade groups. The selected proteins have more extreme FCs in the high-grade tumors (G3 and G4/5).

**Figure 4.** Consistently dysregulated sub-networks, Network Component 1 heatmaps and validation in three independent cohorts. a, Sub-networks consistently dysregulated in the high-grade tumors compared to the low-grade tumors. There is a significant up-regulation of the FCs of the depicted genes (colored in red) after network smoothing in the group G4/5 compared to the group G1 in all three layers (CNA, mRNA and protein). b, Sub-networks consistently dysregulated in the high-grade tumors compared to the low-grade tumors. There is a significant down-regulation of the FCs of the depicted genes (colored in blue) after network smoothing in the group G4/5 compared to the group G1 in all three layers (CNA, mRNA and protein). Functional annotation of the sub-networks in (a) and (b) with more than one node is given below. c, Heatmap of the CNA matrix reduced to the Network Component 1 genes. The columns/samples are ordered based on the grade group. The bottom colorbar depicts the effect size of each sample, i.e. its average FC across the genes of the Network Component 1. The next colorbar represents the negative logarithm in base 10 of the P value from the t-test. The above colorbar shows whether the result is significant: significant results are colored in black and the rest in white. The top
Supplementary Table 1. Clinicopathological, immunological and other molecular information of the 39 PCa patients. (a) Overall clinicopathological characteristics. (b) Detailed information for each patient. Pat: numeric patient ID; Pat_id: patient ID grouped by the overall grade. L: low grade; M: intermediate grade; H: high grade; Overall_Gleason_GrGp: overall ISUP grade group; pT: tumor stage; pN: nodal status; R: surgical margin status; Age_at_OP: age at operation; PSA_at_Diag: blood PSA level at diagnosis; Time (months): RFS time. A value of 0 corresponds to patients excluded for the reasons explained in Methods (see ‘Patients and samples’); Status: status indicator. 1 means recurrence; DX name: tissue region name; ImageName: name of the scanned images; index_tumor_id: patient ID of TA1 (or TA); TA1_GrGp: grade group for TA1; T_GrGp: grade group for TA2.

Supplementary Table 2. Exome analysis of the peripheral blood cells and 105 prostate tumor punches in 39 patients. (a) Allele frequencies (AF) of somatic single nucleotide variants (SNVs) that were called by our bioinformatics pipeline. Genes with called SNV are indicated by an AF > 0. A value of 0 indicates that no SNV was found in the respective genes. In our data, no gene was found with more than one called somatic SNV. (b) Number of samples per gene with called somatic SNV. (c) Protein domain analysis using DAVID.

Supplementary Table 3. Copy number analysis of 105 PCa samples. (a) Log2 ratios indicating the CNA status are shown for all genes in all samples. Values were determined by overlapping gene locations with CNA segments as calculated by CopywriteR. In case more than one segment overlapped with a gene, number was chosen that had the highest absolute value. (b) Genes are shown with log2 ratios higher than 0.5 or lower than -0.5 in at least one sample.
**Supplementary Table 4. RNA-seq analysis.** (a) Log$_2$FCs (relative to all benign samples) for all genes across the tumor samples. (b) mRIN score per sample generated using mRIN (v1.2.0). (c) ETS family gene fusions observed in tumor samples using FusionCatcher: a value of 1 means that the fusion was observed in the respective sample but not its corresponding benign sample, otherwise the value is 0.

**Supplementary Table 5. Proteomics data of 210 PCa samples with duplicates.** (a) Sample information includes patient ID, clinical diagnosis, sample ID and batch design. (b) Protein matrix of log$_2$ scaled intensity of 2,371 proteins quantified in 210 PCa samples.

**Supplementary Table 6. Integration analysis of 66 tumor samples.** (a) L1-norm of the ‘centroid vectors’ in the three layers (CNA, mRNA and protein) across the four grade groups. (b) Information (i.e. reference linking them to PCa, consistency between observed and reported effect and number of tumor samples with CNAs) for the 10 highest-scoring proteins (those with largest average absolute FCs across all tumor specimens). (c) Consistently up-regulated genes in the high-grade tumors: for each of these genes, there is a significant up-regulation of its FCs after network smoothing in the group G4/5 compared to the group G1 in all three layers (CNA, mRNA and protein). (d) Consistently down-regulated genes in the high-grade tumors: for each of these genes, there is a significant down-regulation of its FCs after network smoothing in the group G4/5 compared to the group G1 in all three layers (CNA, mRNA and protein). (e) Chromosome information for the gene members of Network Component 1. (f) Mapping from gene symbols to STRING identifiers.
Figure 1

(a)

3 punches per region

PCa with 1 BPH and 1 dominant tumor nodules (BPH + TA), N = 12

PCa with 1 BPH and 2 dominant tumor regions (BPH + TA1 + TA2), N = 27

105 regions
30 peripheral blood
Genome: Exome, CNA
Transcriptome: RNA-seq
Proteome: PCT-SWATH

(b)

Recurrence-free survival

Biochemical RFS (months)

low (G1)
intermediate (G2 and G3)
high (G4/5)
p<0.001

Biochemical RFS (months)

p<0.001

Biochemical RFS (months)

p<0.001
Figure 2

(a) Point Mutations (p-value=0.0096)

(b) CNA (p-value=2e-06)

(c) RNA (p-value=0.0493)

(d) CNA (p-value=1e-06)

(e) RNA (p-value=0.0568)

(f) Protein (p-value=0.0131)

(g) Protein (p-value=0.007)
Figure 3

(a) CNA

(b) RNA

(c) Protein

MDH2

SEPHS1

RABL3
Figure 4

**Functional annotation**

**Up-regulation**
1. Chromatin remodelling
2. Mitochondrial import
3. RNA polymerase
4. unknown function in prostate
5. Actin related complex, Arp2/3 complex
6. Protein quality control
7. mRNA splicing
8. DNA damage response

**Down-regulation**
1. Glutathione metabolism
2. Cholesterol biosynthesis

**CNA**

| Log2 FC | Significance | Grade group |
|---------|--------------|-------------|
| -1      | no           | G1          |
| 1       | yes          | G2          |
|         |              | G3          |
|         |              | G4/5        |

-10 log10 p-value (CNA)
Figure 4 (continued)

**d**

Log2 FC

RNA

| Low | Intermediate | High |
|-----|--------------|------|

- \(-\log_{10} \text{p-value} (\text{RNA})\)

**e**

Log2 FC

Protein

| Low | Intermediate | High |
|-----|--------------|------|

- \(\log_{10} \text{p-value} (\text{Protein})\)

**f**

TCGA

Logrank test P-value = 0.00068

MSKCC

Logrank test P-value = 0.11129

Aarhus

Logrank test P-value = 0.00001

Survival probability over time.
Figure 5

a

Protein

RNA

FIGURE 5

b

RNA vs. CNA

Protein vs. CNA

Protein vs. RNA

TA1-TA2 Pears correl CNA

TA1-TA2 Pears correl CNA

TA1-TA2 Pears correl RNA

TA1-TA2 Pears correl RNA

r = 0.486

r = 0.347

r = 0.185

Patient grade

intermediate

high