Convergent network effects along the axis of gene expression during prostate cancer progression

Konstantina Charmpi[1,2,#], Tiannan Guo[3,4,#,*], Qing Zhong[5,#], Ulrich Wagner[5,#], Rui Sun[4], Nora C. Toussaint[5,6,7], Christine E. Fritz[5], Chunhui Yuan[4], Hao Chen[4], Niels J. Rupp[5], Ailsa Christiansen[5], Dorothea Rutishauser[5], Jan H. Rüschoff[5], Christian Fankhauser[5], Karim Saba[5,8], Cedric Poyet[8], Thomas Hermanns[8], Kathrin Oehl[5], Ariane L. Moore[9], Christian Beisel[9], Laurence Calzone[10], Loredana Martignetti[10], Qiushi Zhang[4], Yi Zhu[3,4], María Rodríguez Martínez[11], Matteo Manica[11], Michael C. Haffner[12], Ruedi Aebersold[3,13,*], Peter J. Wild[5,14,*], Andreas Beyer[1,2,*]

[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

Background

Tumor-specific genomic aberrations are routinely determined by high throughput genomic measurements. It remains unclear though, how complex genome alterations affect molecular networks through changing protein levels, and consequently biochemical states of tumor tissues.

Results

Here, we investigated the propagation of genomic effects along the axis of gene expression during prostate cancer progression. For that, we quantified genomic, transcriptomic and proteomic alterations based on 105 prostate samples, consisting of benign prostatic hyperplasia regions and malignant tumors, from 39 prostate cancer patients. Our analysis revealed convergent effects of distinct copy number alterations impacting on common downstream proteins, which are important for establishing the tumor phenotype. We devised a network-based approach that integrates perturbations across different molecular layers, which identified a sub-network consisting of nine genes whose joint activity positively correlated with increasingly aggressive tumor phenotypes and was predictive of recurrence-free survival. Further, our data revealed a wide spectrum of intra-patient network effects, ranging from similar to very distinct alterations on different molecular layers.

Conclusions

This study uncovered 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.

Keywords: molecular aberrations, network effects, prostate cancer, proteogenomic analysis, tumor heterogeneity

61 # Background

62 Prostate cancer (PCa) represents one of the most common neoplasm among men with almost
63 1,300,000 new cases and 360,000 deaths in 2018 [1] accounting for 15% of all cancers diagnosed. PCa is
64 the fifth leading cause of cancer death in men and represents 6.6% of total cancer mortality in men [1].
65 Despite earlier detection and new treatments, the lifetime risk to die of PCa has remained stable at
66 approximately 3% since 1980. (National Cancer Institute SEER data:
67 https://seer.cancer.gov/statfacts/html/prost.html). In many patients, PCa is indolent and slow growing.
68 The challenge is to identify those patients who are unlikely to experience significant progression while
69 offering radical therapy to those who are at risk. Current risk stratification models are based on
70 clinicopathological variables including histomorphologically defined grade groups, prostate-specific
71 antigen (PSA) levels and clinical stage. Although those variables provide important information for
72 clinical risk assessment and treatment planning [2, 3], they do not sufficiently predict the course of the
73 disease.

74 Extensive genomic profiling efforts have provided important insights into the common genomic
75 alterations in primary and metastatic PCa [4-9]. Interestingly, PCa genomes show a high frequency of
76 recurrent large-scale chromosomal rearrangements such as TMPRSS2-ERG [10]. In addition, extensive
77 copy number alterations (CNAs) are common in PCa, yet point mutations are relatively infrequent in
78 primary PCa compared to other cancers [6, 11]. A major complicating factor is that around 80% of PCas
79 are multifocal and harbor multiple spatially and often morphologically distinct tumor foci [12, 13].
80 Several recent studies have suggested that the majority of topographically distinct tumor foci appear to
81 arise independently and show few or no overlap in driver gene alterations [14-16]. Therefore, a given
82 prostate gland can harbor clonally independent PCas.

83 To allow for a more functional assessment of the biochemical state of PCa, it is necessary to go
84 beyond genomic alterations and comprehensively catalogue cancer specific genomic, transcriptomic and
85 proteomic alterations in an integrated manner [17-19]. Such an approach will provide critical
86 information for basic and translational research and could result into clinically relevant markers. While
87 hundreds of PCa genomes and transcriptomes have been profiled to date [20], little is known about the
88 PCa proteome. Although recent work has emphasized the need for integrated multi-omics profiling of
89 PCa, we still lack understanding about how genomic changes impact on mRNA and protein levels [17-19].
90 Especially the complex relationship between tumor grade, tumor progression and multi-layered
91 molecular network changes remains largely elusive.

92 For example, previous work has shown that copy number changes may alter transcript levels of
93 many genes, whereas the respective protein levels remain relatively stable [21]. Indeed, there is
94 compelling evidence across multiple tumor types that many genomic alterations are 'buffered' at the
95 protein level and are hence mostly clinically inconsequential [22]. To better understand the evolution of
96 PCa and to identify core networks perturbed by genomic alterations and thus central for the tumor
97 phenotype, it is therefore essential to investigate the transmission of CNAs to the transcriptomic and
98 proteomic level.

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

102 To address these open questions, we performed a multi-omics profiling of radical prostatectomy
103 (RP) specimens at the level of the genome, transcriptome and proteome from adjacent biopsy-level
104 samples, using state-of-the-art technologies. Unique features of this study are (1) the utilization of PCT
105 (pressure cycling technology)-SWATH (Sequential Window Acquisition of all THeoretical Mass Spectra)
106 mass spectrometry [23, 24], allowing rapid and reproducible quantification of thousands of proteins
107 from biopsy-level tissue samples collected in clinical cohorts; (2) the simultaneous profiling of all omics
108 layers from the same tissue regions; (3) inclusion and full profiling of benign regions, which provides a
109 matching control for each tumor; and (4) the full multi-omics characterization of multiple tumor regions
110 from the same patients, thus enabling the detailed investigation of tumor heterogeneity. This design
111 resulted in the multi-layered analyses of 105 samples from 39 PCa patients, as well as of the exome of
112 corresponding peripheral blood cells yielding a comprehensive molecular profile for each patient and
113 identified molecular networks that are commonly altered in multiple patients. Importantly, some of the
114 affected genes/proteins exhibited very small individual effect sizes, suggesting that combined network
115 effects of multiple genes may significantly contribute to determining PCa phenotypes.

116 Results

117 Proteogenomic analysis of the sample cohort identifies known PCa biomarkers.

118 In this study, we analyzed 39 PCa patients (**Additional file 1: Fig. S1**) belonging to three groups
119 who underwent laparoscopic robotic-assisted RP. The patients were from the PCa Outcomes Cohort
120 (ProCOC) study [25, 26]. Tumor areas were graded using the ISUP (International Society of Urological
121 Pathology) grade groups [27], which range from ISUP grade group G1 (least aggressive) to G5 (most

122 aggressive). The more advanced grade groups G4 and G5 are considered jointly (G4/5). The cohort
123 tested included 12 low-grade (G1), 17 intermediate- (G2 and G3), and 10 high-grade (G4/5) patients (**Fig.**
124 **1a**, **Additional file 1: Fig. S1**, **Additional file 2: Table S1**). For low-grade PCa patients, we selected two
125 representative regions, one of benign prostatic hyperplasia (BPH) and one of malignant tumor (TA). Since
126 PCa often presents as a multifocal disease with heterogeneous grading within each prostate specimen
127 [24] we analyzed two different tumor regions from the 27 intermediate- and high-grade patients. In
128 those cases three representative regions, including BPH, the most aggressive tumor (TA1) and a
129 secondary, lower-grade tumor (TA2) [2] were analyzed. Thus, TA1 always represented the higher-grade
130 nodule compared to TA2. Note, whereas each patient was assigned a patient-specific overall grade (*i.e.*
131 'low', 'intermediate' or 'high'), each tumor area was additionally assigned an individual grade group
132 based on its histological appearance. According to current ISUP guidelines, the grading of the entire
133 prostate specimen depends on the size and grade of individual nodules [28]. Thus, it is possible that the
134 patient grading is lower than the grading of the most aggressive nodule, if another lower-grade nodule is
135 larger. Tumor regions contained at least 70% tumor cellularity and the distance between the analyzed
136 areas (TA1 versus TA2) was at least 5 mm. Altogether, we obtained 105 prostate tissue specimens
137 (**Additional file 2: Table S1**). Three adjacent tissue biopsies of the dimensions 0.6 x 0.6 x 3.0 mm were
138 punched from each representative region for exome sequencing, CNA (derived from the exome
139 sequencing data), RNA sequencing (RNA-seq), and quantitative proteomic analysis using the PCT-SWATH
140 technology [23] respectively. Proteomic analysis was performed in duplicates for each tissue sample.
141 Peripheral blood samples from each patient were also subjected to exome sequencing and served as the
142 genomic wild-type reference (**Fig. 1**). All three types of grading (*i.e.* patient-specific overall grading, TA1
143 grading and TA2 grading) were predictive of the recurrence-free survival (RFS) in our study.

144 In agreement with prior reports, we observed relatively few recurrent point mutations across
145 patients (**Additional file 1: Fig. S2**, **Additional file 3: Table S2**), but substantial CNAs (**Additional file 1:**
146 **Figs. S3** and **S4**, **Additional file 4: Table S3**). In total, 1,110 genes showed copy number gains in at least
147 five samples or copy number losses in at least five samples (see **Additional file 1: Supplementary Text**
148 for details). Likewise, our data confirmed the differential expression of several transcripts/proteins that
149 had previously been suggested as PCa biomarkers or which are known oncogenes in other tumor types
150 (**Additional file 1: Fig. S5**, **Additional file 5: Table S4** and **Additional file 6: Table S5**) (see **Additional file**
151 **1: Supplementary Text** for details). This consistency with previously published results confirmed the
152 quality of our data and motivated us to go beyond previous work by performing a network-based multi-
153 omics multi-gene analysis.

154 ## Molecular perturbations correlate with tumor grade.

155 As a first step towards a cross-layer analysis, we asked if high-grade PCa would be generally
156 affected by stronger alterations (compared to low-grade PCa) at the genome, transcriptome, and
157 proteome layer [29]. Thus, we devised molecular perturbation scores that quantified the number of
158 affected genes/proteins and the extent to which these genes/proteins were altered in the tumor
159 specimens compared to their benign controls (see the '**Methods**' section for details). Higher-grade
160 tumors (G3 and G4/5) exhibited significantly higher molecular perturbation scores than lower-grade
161 tumors (G1 and G2). Those differences were statistically significant in all but one case (P value < 0.05,
162 one-sided Wilcoxon rank sum test, **Fig. 2**). The CNA perturbation magnitude exhibited the highest
163 correlation with the PCa grading, confirming prior studies documenting the tight association between
164 CNA, histopathological grade and risk of progression [4, 5, 30]. Previous work suggested that copy
165 number changes are to some extent buffered at the protein level [31]. Interestingly, we observed that
166 proteins known to be part of protein complexes were significantly less strongly correlated with the fold
167 changes (FCs) of their coding mRNAs than proteins not known to be part of protein complexes (P value <
168 2.6e-11, one-sided t-test, **Additional file 1: Fig. S6**). This result is consistent with the concept that protein
169 complex stoichiometry contributes to the buffering of mRNA changes at the level of proteins [22, 32-34].
170 Thus, molecular patterns in high-grade PCa are more strongly perturbed at all layers and the effects of
171 genomic variation are progressively but non-uniformly attenuated along the axis of gene expression.

172 ## Inter-patient heterogeneity decreases along protein biosynthesis.

173 Our analysis of CNA profiles (above and **Additional file 1: Supplementary Text**) already revealed
174 many shared CNAs across patients, suggesting that such common CNAs might represent genomic driver
175 changes. We therefore investigated if such a convergence towards common molecular endpoints could
176 also be observed at the transcript and protein level. To address this question, we first computed a
177 reference molecular signature that is characteristic of the molecular perturbations of tumors in a given
178 grade group. These 'centroid vectors' were obtained by computing the average tumor-to-benign FCs
179 across all samples within a grade group. Consistent with the observation above, we found that the
180 average effect sizes (averaged absolute centroid FCs) were increasing with the grade group for all three
181 layers (CNA, mRNA, and protein; **Additional file 7: Table S6**). Next, we compared each individual sample
182 within a group against the matching centroid of the same group. For the quantification of the similarity
183 between a tumor sample and the corresponding centroid we used four similarity/distance measures:
184 Pearson correlation, Mutual Information (MI)[35], Manhattan distance and Euclidean distance. While the

185 first two measures (*i.e.* Pearson correlation and MI) quantify the degree to which the tumor sample and
186 centroid vector co-vary, the other two measures (*i.e.* Manhattan and Euclidean distance) also take into
187 account the magnitude of the FCs in the two vectors. To illustrate this difference, imagine two patients
188 having perturbations of the same genes/proteins, but one of them exhibiting overall two-fold greater
189 FCs (*i.e.* all FCs are increased by a factor of two compared to the other patient). In such a scenario
190 Pearson correlation and MI would yield identical results for the two patients when compared to the
191 centroid, whereas Manhattan and Euclidean distance would identify them as different. Using the
192 Pearson correlation and MI, we found that high-grade PCa (G4/5) were more similar to their respective
193 centroid than low-grade PCa (G1) to their centroid (**Fig. 3**). This effect was particularly pronounced for
194 protein-level changes. This is consistent with the notion that protein levels (and not mRNA levels) are
195 subjected to stronger selection. Interestingly, when the Euclidean distance and the Manhattan distance
196 were used to characterize tumor similarity, we found that the high-grade tumors were more *dissimilar* to
197 each other than the low-grade tumors (**Additional file 1: Fig. S7**), in sharp contrast to the Pearson
198 correlation and MI. Based on the nature of the different similarity measures tested, we hypothesized
199 that there is a set of proteins commonly affected in their abundance by oncogenic alterations in high-
200 grade tumors. This would increase the similarity using the Pearson correlation or MI. However, although
201 the same proteins are affected, they are affected to a different extent in different high-grade tumors, *i.e.*
202 the FCs exhibited a high degree of variability (**Additional file 1: Fig. S7**), which would increase the
203 dissimilarity based on the Euclidean distance or Manhattan distance (see the '**Methods**' section and
204 **Additional file 1: Fig. S7** for a schematic explanation).

205 To further corroborate the notion of common endpoints, we focused on the 20 proteins with the
206 largest average absolute FCs across all tumor specimens (**Additional file 1: Fig. S7**, **Additional file 7:**
207 **Table S6**). Among them was PSA (KLK3), and several other well established PCa-associated proteins like
208 AGR2 [36], MDH2 [37], MFAP4 [38] and FABP5 [39]. We observed that for some of these top 20 proteins,
209 FCs were more extreme in the higher-grade tumors (G3 and G4/5) compared to lower-grade tumors (G1
210 and G2), such as MDH2 and SEPHS1 (up-regulation; **Fig. 3c**). RABL3 was one of the most strongly down-
211 regulated proteins (**Fig. 3c**), which is a surprising finding as RABL3 is known to be up-regulated in other
212 solid tumors [40, 41]. Interestingly, in most cases these proteins were from loci that were not subject to
213 CNAs (**Additional file 1: Fig. S7**, **Additional file 7: Table S6**), hinting that independent genomic events
214 would impact on these target proteins *via* network effects.

215 Effects of distinct CNAs converge on common proteins.

216 It has previously been suggested that mutations affecting different genes could impact common
217 molecular networks if the respective gene products interact at the molecular level [42]. However,
218 previous analyses were mostly restricted to individual molecular layers. For example, it was shown that
219 genes mutated in different patients often cluster together in molecular interaction networks [42]. But,
220 effects of these mutations on transcript and protein levels remained unexplored in this case. Here, we
221 aimed at a multi-layer network analysis, involving the genome, transcriptome and proteome using two
222 different network approaches.

223 First, we speculated that distinct genomic events in different patients would commonly impact
224 on at least some of the top 20 target proteins identified above. Among those top targets we selected
225 AGR2, ACPP, POSTN and LGALS3BP, because these proteins/genes had correlated protein- and mRNA
226 FCs; thus, protein level changes were likely caused by cognate mRNA level changes. Importantly, those
227 mRNA/protein level changes could not be explained by CNAs of the coding genes themselves (**Additional**
228 **File 1: Fig. S7**). Thus, these transcriptional changes were likely caused by *trans*-effects. To identify
229 potential regulators for each target gene, we used an independently inferred generic transcriptional
230 regulatory network (Leote *et al. in revision*; preprint available on bioRxiv) and selected putative
231 regulators at most two edges away from the target genes (see the '**Methods**' for details). Using the
232 ElasticNet algorithm we next fitted a linear model regressing the mRNA changes of the target genes
233 against the network neighbors' CNAs. Thereby we identified genes whose CNA changes were associated
234 with mRNA changes of the four target genes AGR2, ACPP (a.k.a ACP3), POSTN and LGALS3BP (**Additional**
235 **file 7: Table S6**). To validate our approach, we used two independent PCa cohorts (TCGA; [8] and MSKCC
236 [30]) and computed the association between the CNAs of each significant regulator and the
237 corresponding mRNA log-FC of the respective target gene. In most cases we observed an agreement in
238 terms of effect directions, *i.e.* the signs of association between CNA changes of the putative regulators
239 and the mRNA log-FCs of the respective target had the same direction as in our cohort (**Additional file 7:**
240 **Table S6**). Here we use ACPP/ACP3 as an illustrative example (**Fig. 3d)**: ACP3 (a.k.a. PAcP) is a prostate-
241 specific acid phosphatase with a critical role PCa etiology and has been suggested as a PCa biomarker
242 long before PSA [43]. ACP3 is known to inhibit cell proliferation and is therefore typically down-regulated
243 in PCa [44], despite elevated ACP3 protein levels in patient blood [43]. In our cohort ACP3 levels were
244 strongly down-regulated in all of the high-grade patients and in the vast majority of low- and
245 intermediate-grade patients, suggesting that ACP3 down-regulation represents an early event during
246 PCa evolution. Despite its established role in PCa, little is known about the oncogenic driver events

247 downregulating ACP3 [43]. Our network modeling identified six putative ACPP/ACP3 regulators
248 (ANKRD22, MS4A3, RHOV, ARL11, DEFB1, and DPYSL2), several of which are already known to be
249 associated with PCa [45-47]. Analyzing the CNA signatures of these six putative regulators reveals at
250 least two larger groups of patients (**Fig. 3d)**: the first one harboring joint deletions of ARL11 and
251 ANKRD22, the second one harboring joint deletions of DEFB1 and DPYSL2. The latter two genes are both
252 encoded on Chromosome 8 and thus, their deletion may be due to single CNA events. ARL11 and
253 ANKRD22 however, are encoded on different chromosomes. Importantly, these events were clonal in
254 most cases, *i.e.* they were mostly common to both tumor samples of a given patient. A remarkable
255 exception was patient M7, who had a joint deletion of ARL11 and ANKRD22 in tumor area 1 (TA1) and a
256 DEFB1/DPYSL2 deletion in tumor area 2 (TA2). Hence, our network analysis hints that distinct deletions
257 in the network vicinity of ACP3 can lead to the repression of this anti-proliferative protein. More
258 examples of distinct CNAs having similar effects on downstream targets were found for the other three
259 focus proteins (**Additional file 1: Fig. S7)**. Taken together, these findings suggest that tumor mechanisms
260 in different patients converged on common protein endpoints and that the expression levels of these
261 proteins were progressively more strongly affected during tumor evolution.

262 Joint network effects of CNAs drive tumor progression.

263 The analysis above identified molecular networks driving tumor alterations and thus indicated
264 altered biochemical states that were common to most tumor specimens. To identify sub-networks that
265 specifically distinguish high-grade from low-grade tumors, we mapped our data onto the STRING gene
266 interaction network [48], and employed network propagation [49, 50] separately to the CNA,
267 transcriptome and proteome data for each of the tumor samples. We excluded point mutations from
268 this analysis as their frequency was too low in our cohort. By combining published molecular
269 interactome data with a network propagation algorithm [42, 49], we aimed to 'enrich' network regions
270 with many perturbed genes/proteins. We reasoned that the convergent consequences of genomic
271 variants on common network regions would be indicative of specific biochemical functions that are
272 important for the tumor biology. We therefore identified genes/proteins in network regions that showed
273 a higher score (or a lower score) in high-grade (G4/5) relative to lower-grade (G1) tumor groups at all
274 three levels (**Fig. 4a**, **b**; '**Methods**' section). This analysis identified sub-networks consisting of over- and
275 under-expressed genes (relative to the benign controls). We found 57 amplified genes (**Additional file 7:**
276 **Table S6**) for which transcripts and proteins were often over-expressed in high-grade PCa (**Fig. 4a**) and

277 21 genes with copy number loss (**Additional file 7: Table S6**) for which transcripts and proteins were
278 often down-regulated compared to lower-grade tumors (**Fig. 4b**).

279 Among the up-regulated network nodes, we observed genes modulating the stability of
280 chromatin, such as chromatin-binding protein Chromobox 1 (CBX1) [51], SET Domain Bifurcated 1
281 (SETDB1) [52], a function linking to H3K27me3 and H3K9me3 in chromatin, and CBX3 (known as HP1-γ)
282 [53]. SETDB1 is an oncogene in melanoma [54] and has also been found to be over-expressed in PCa and
283 cell lines [55]. Further, we found genes involved in DNA damage repair, such as SMG7 [56] and ATR [57],
284 and PRKCZ[58], which had already been suggested as a biomarker prognostic for survival in PCa [59].
285 Multiple actin related proteins including ARPC1B [60], ARPC5 [61], ACTL6A [62], and CFL1 [63], which are
286 markers for aggressive cancers, were part of the up-regulated network nodes. Moreover, the up-
287 regulated genes contained proteins related to the cell cycle like BANF1 and proteins interacting with the
288 centrosome including LAMTOR1 and RAB7A that had already been associated with PCa [64]. Finally,
289 several signaling molecules with known roles in PCa were up-regulated, such as the transcription factor
290 Yin Yang 1 (YY1) [65], the TGF-β receptor TGFBR1 [66], and KPNA4, which promotes metastasis through
291 activation of NF-κB and Notch signaling [67]. Thus, up-regulated network nodes are involved in
292 DNA/chromatin integrity and growth control.

293 Likewise, several of the down-regulated genes had functions associated with PCa. For example,
294 the oxidative stress related gene MGST1, which is recurrently deleted in PCa [68]. ALDH1A3 is a direct
295 androgen-responsive gene, which encodes NAD-dependent aldehyde dehydrogenase [69]. DHCR24 is
296 involved in cholesterol biosynthesis and regulated by the androgen receptor [70]. Polymorphisms in
297 CYP1A1 are associated with PCa risk in several meta-analyses among different ethnicities [71-73].

298 Further, our network analysis is suggesting tumor mechanisms converging on genes that are
299 known contributors to PCa tumor biology. For example, the PCa-associated gene SF3B2 [74, 75] was only
300 weakly amplified in some of the high-grade tumors (average $\log_2$FC = 0.016) and mRNA levels showed
301 similarly small changes (average $\log_2$FC = 0.024). On the other hand, the SF3B2 protein levels were
302 consistently and more strongly up-regulated across tumors (average $\log_2$FC = 0.31), especially within the
303 high-grade tumors (**Additional file 1: Fig. S8**). Another example is UBE2T whose over-expression is
304 known to be associated with PCa [76]. Unfortunately, we could not quantify the corresponding protein
305 levels. However, we observed a strong and consistent mRNA over-expression across several tumors
306 (average $\log_2$FC = 0.73), even though at the DNA level the gene was only weakly amplified (average
307 $\log_2$FC = 0.023; **Additional file 1: Fig. S8**). Our findings of more heterogeneous CNAs, but more uniform

308 mRNA and protein alterations point on convergent evolutionary mechanisms, as we move along the axis
309 of gene expression.

310 Next, we analyzed the largest connected component with genes up-regulated in advanced
311 disease in more detail (see the '**Methods**' section). It consists of the nine nodes EMD, BANF1, ACTL6A,
312 YY1, RUVBL1, KANSL1, MRGBP, VPS72 and ZNHIT1 (**Fig. 4a**), and is referred to in the following as
313 Network Component 1 (**Additional file 7: Table S6**). Seven of these proteins are involved in chromosome
314 organization which may induce genomic alterations and influence the outcome of multiple cancers
315 including PCa [77]. For example, the actin-related protein ACTL6A is a member of the SWI/SNF (BAF)
316 chromatin remodeling complex[78], and a known oncogene and a prognostic biomarker for PCa [79].
317 Further, ACTL6A, RUVBL1 and MRGBP are together part of the NuA4/Tip60-HAT complex, which is
318 another chromatin remodeling complex involved in DNA repair [80]. Likewise, KANSL1 is involved in
319 histone post-translation modifications, while VPS72 is a member of histone- and chromatin remodeling
320 complexes [81]. Thus, Network Component 1 consists of genes involved in chromatin remodeling and
321 DNA repair, many of which are known to be involved in cancers.

322 Several samples were characterized by a small, but consistent DNA amplification of multiple
323 members of Network Component 1 (**Fig. 4c**). Out of the 66 tumor samples, there were 30 samples –
324 belonging to all grade groups – with a weak but consistent DNA amplification of Network Component 1
325 members, while the high-grade samples had stronger amplifications on average (*i.e.* larger effect sizes).
326 Importantly, gene members of Network Component 1 were dispersed across eight chromosomes
327 (**Additional file 7: Table S6**). The parallel DNA amplification of these genes is therefore the result of
328 multiple independent CNA events, while the signal on any single gene alone was too weak to be
329 significant in isolation. In some but not all cases, the amplifications led to a small, but consistent increase
330 in mRNA expression of the amplified gene loci (**Fig. 4d**). Unfortunately, only three out of the nine
331 proteins were detected in our proteomics experiments (**Fig. 4e**). Interestingly, patients where the DNA
332 amplifications led to transcript over-expression were almost always high-grade patients, whereas
333 patients where the amplification affected gene expression to a smaller extent were low- or
334 intermediate-grade patients (**Fig. 4c**, **d**). Further, we noticed that TA2 samples graded as G3 from high-
335 grade patients carried amplifications of Network Component 1, whereas tumor areas graded as G3 from
336 intermediate-grade patients did not have amplifications of this network component (**Fig. 4c**, **d**). Thus,
337 although the tumor areas were histologically equally classified, tumor areas from high-grade patients
338 carried a CNA signature and expression patterns reminiscent of the high-grade areas from the same
339 patients. Therefore, within the cohort tested the joint DNA amplification of this network component

340 along with RNA up-regulation is a signature of high-grade tumors. Curiously, the higher-grade tumor
341 areas of those high-grades patients (TA1) carried stronger DNA amplifications than the respective lower-
342 grade areas (TA2), which implies that the progressive amplification of Network Component 1 during
343 tumor evolution may contribute to an increasingly aggressive phenotype. In order to further corroborate
344 the clinical relevance of this network perturbation we analyzed published datasets of three additional
345 PCa cohorts (TCGA[8], MSKCC [30], and Aarhus [82]), together comprising a total of 713 patients with
346 known clinical outcome. We found that amplification of genes from Network Component 1 was a
347 significant predictor of reduced RFS in the MSKCC cohort (*P* value = 8.8e-3, log-rank test). In the TCGA
348 cohort, we observed the same trend although the difference in RFS was not statistically significant (*P*
349 value = 0.17; **Fig. 4f**). Additionally, we found that over-expression of genes from Network Component 1
350 was a significant predictor of reduced RFS in the TCGA cohort (*P* value = 2.1e-4, log-rank test), which was
351 the cohort with the largest number of patients. In the other two cohorts we observed the same trend,
352 although the difference in RFS was not statistically significant (*P* value = 0.30 and 0.093 for MSKCC, and
353 Aarhus; **Fig. 4f**). Thus, both CNA and RNA changes of Network Component 1 are predictive of the time to
354 relapse in independent cohorts.

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

359 Analysis of distinct tumor nodules defines intra-patient heterogeneity (TA1 versus TA2 360 comparison).

361 The CNA patterns (**Additional file 1: Fig. S4**) and the Network Component 1 analysis (**Fig. 4c**, **d**)
362 suggest that different tumor areas from the same patient shared several mutations. Such common
363 signatures are expected if different tumor nodules originate from a common clone. If this was true, we
364 would expect mutational signatures to be more similar between different nodules from the same patient
365 than between patients, even though mutated genes may be shared across patients. To compare the
366 intra- and inter-patient molecular heterogeneity at the levels of CNAs, transcript, and protein FCs, we
367 computed the Pearson correlation between tumor area 1 (TA1) and its paired tumor area 2 (TA2) for
368 each layer and all of the 27 patients with two characterized tumor areas (25 for the mRNA, see the
369 '**Methods**' section and **Additional file 1: Supplementary Text**). As a control, we also computed all
370 pairwise Pearson correlations between the samples within each of the grade groups (*i.e.* inter-patient

371 correlation). As expected, paired TA1 and TA2 from the same patient were on average more strongly
372 correlated to each other compared to samples from different patients within the same grade group. This
373 finding was consistent for all omics layers (**Fig. 5a**), and was more pronounced at the CNA and mRNA
374 layers compared to the protein layer.

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

383 Comparing molecular similarity across omics layers allowed us to identify specific types of
384 patients. The patients H2, H4, M13 had highly correlated tumor areas at all three layers (upper right
385 corner in all scatterplots of **Fig. 5b**). Likely, the tumor areas of these patients have a common clonal
386 origin (**Additional file 1: Fig. S3**). In contrast, patients M12 and M14 had weakly correlated tumor areas
387 at all levels (bottom left corner in all scatterplots of **Fig. 5b**). These tumor nodules either have
388 independent clonal origins or they diverged at an earlier stage during tumor evolution (**Additional file 1:**
389 **Fig. S3**) [16]. For example, in the case of patient M12 large parts of the genome were not affected by
390 CNAs in the benign sample as well as in TA1 and TA2. However, as shown on **Additional file 1: Fig. S3**, a
391 large region was amplified in TA1, whereas the same region was deleted in TA2. This is consistent with a
392 scenario in which TA1 and TA2 show parallel evolution. A third class of patients is exemplified by the
393 patients M9 and M17, who showed a high correlation between their tumor areas on the CNA and mRNA
394 levels, but not on the protein level. Yet other patterns were apparent in patients M4, M7, and H10. They
395 showed similar mRNA and protein patterns in the two tumor areas, but relatively uncorrelated CNAs. M7
396 was the patient that we identified earlier with two different CNA signatures both reducing the levels of
397 the same protein (ACP3). The results here apply to global proteome patterns and therefore hint that
398 such convergent network effects of CNAs can be frequent. We confirmed that protein-level similarity
399 correlated with similar histological characteristics of the tumor areas. **Additional file 1: Fig. S9** shows
400 formalin-fixed paraffin-embedded (FFPE) tissue microarray images (duplicates) from the analyzed tumor
401 nodules (TA1 and TA2, diameter 0.6 mm), further underlining the hypothesis that ultimately protein-
402 level alterations are responsible for common cellular phenotypes. Although we cannot fully exclude the

403 possibility that some of these results were affected by technical noise in the data, our findings suggest
404 that transcript alterations can frequently be buffered at the level of proteins (patients M9, M17,
405 **Additional file 1: Fig. S6**) and that convergent evolutionary processes may lead to the alteration of
406 common proteins (patients M4, M7, H10). We also note that our findings are specific to the two tumor
407 areas available in this study and could be different if other nodules had been sampled for each of the
408 patients. However, our findings on patients with weakly correlated tumor areas at all levels like M12 and
409 M14 suggest that these patients might carry more than one disease [16].

410 Discussion

411 Despite twenty years of oncological research involving genome-scale (omics) technologies, we
412 know remarkably little about how the discovered genomic alterations affect the biochemical state of a
413 cell and consequently the disease phenotype. In particular, little is known about how genomic
414 alterations propagate along the axis of gene expression [17, 18]. Here, we have exploited recent
415 technological advances in data acquisition that made it possible to characterize small samples of the
416 same tumor specimens at the level of genomes, transcriptomes, and proteomes and advances in
417 computational strategies towards the network-based integration of multi-omics data.

418 In our study, samples were generated from small, less than 1 mm diameter punches in
419 immediate spatial proximity in the tumor and subsequently profiled at all three 'omics layers' (DNA, RNA,
420 proteome). Due to the large spatial heterogeneity of PCa [14, 24], this design - which is so far uncommon
421 for studies profiling multiple layers from tumor specimens - was instrumental for increasing the
422 comparability of the various omics layers and thus facilitated the analysis of molecular mechanisms. Our
423 key findings are: (1) we confirmed the importance of CNAs for PCa biology and the alteration of many
424 known PCa-associated genes at the transcript- and protein-level; (2) we revealed a generally elevated
425 molecular alteration of high-grade tumors compared to lower-grade tumors; (3) although our study
426 confirmed large within- and between-patient genomic heterogeneity, (4) we detected molecular
427 networks that were commonly altered at the mRNA and protein-level. The fact that many of those target
428 molecules are known drivers of PCa tumorigenesis, supports the notion that these proteins/transcripts
429 are subject to convergent evolutionary mechanisms.

430 We integrated the three omics layers using a network-based approach as opposed to directly
431 comparing gene perturbations (mutations) to gene products (transcripts and proteins). Using genome
432 data only, it had previously been hypothesized that whereas the identity of specific mutated genes may
433 differ between tumors, those mutations might still affect common molecular networks [42]. In other

434 words, tumor phenotypes are determined by the perturbation of molecular networks and not by the
435 perturbation of isolated genes. Our study provides experimental evidence that such network effects are
436 indeed propagated to subsequent molecular layers and that this effect propagation may be clinically
437 relevant. A very prominent example is our discovery that the long-known PCa gene ACPP (ACP3) is
438 downregulated through diverse CNA events, while sometimes even within the same patient different
439 CNA events might downregulate this critical tumor suppressor.

440 Our multi-omics network analysis revealed that high-grade PCa tumors distinguished themselves
441 from low-grade tumors in two aspects. The first is a generally higher heterogeneity and loss of controlled
442 gene regulation, which increased the molecular differences among high-grade tumors. It had previously
443 been shown that gene expression in tumors is often less coordinated than in normal samples [29]. The
444 increasing heterogeneity of protein concentrations suggests that this loss of coordinated expression also
445 affects protein levels. The second aspect is the convergence of molecular alterations towards specific
446 molecular sub-networks at the genomic, transcriptomic and proteomic layer along the progression from
447 low-grade to high-grade tumors. Thus, although we observed globally a higher degree of variability in
448 gene expression and proteome control among high-grade specimens, a specific subset of the observed
449 alterations appeared crucial for determining the aggressive tumor phenotype. Tumors are under
450 selective pressure acting on the biochemical function of the cells. It is generally believed that proteins
451 are a closer reflection of the functional state of a cell than the mRNA. Here we could show that the fold
452 changes of proteins like RABL3, MFAP4, and SF3B2 were more pronounced and/or uniform across high-
453 grade tumors than either their coding mRNAs or the underlying CNAs.

454 Specifically, our analysis led to the identification of Network Component 1, a sub-network
455 involved in chromatin remodeling and consisting of genes that were weakly amplified in intermediate-
456 grade (G3) tumor specimens. Signals of individual gene members of this component were virtually
457 indistinguishable from noise in our cohort. However, their consistent alterations across the network
458 region, across molecular layers and the fact that the same genes showed enhanced signals in high-grade
459 specimens, rendered this component highly interesting. The fact that copy number and expression
460 changes of Network Component 1 members were predictive for survival in independent cohorts further
461 supports the potential clinical relevance of this sub-network. Our network-based cross-omics analysis
462 identified nine other network components (**Fig. 4**) successfully capturing several known and potentially
463 new PCa-associated genes. However, neither Network Component 1 nor any of the other network
464 components was uniformly subject to CNAs across all high-grade patients. Instead, we found different
465 network components modified in different patients and these sub-networks were involved in cellular

466 processes as diverse as actin remodeling, DNA damage response, and metabolic functions, all of which
467 are known contributors to PCa biology. This further underlines the large inter-patient variability of PCa
468 and it demonstrates the diversity of molecular mechanisms leading to histologically similar phenotypes.
469 Future prediction models of PCa including the ISUP grade groups, PSA levels and clinical stage might be
470 improved by exploiting multi-omics network analyses. Detecting aggressive networks alterations in
471 prostate biopsies would help clinicians to advice either active surveillance or active therapy. However,
472 the development of such multi-dimensional biomarkers would require much larger patient cohorts.

473 Another distinguishing feature of this study was the simultaneous profiling of two different
474 tumor regions in 27 out of the 39 patients. The profiling of multiple tumor regions from the same
475 prostate helped to further highlight the enormous heterogeneity of PCa within patients and provided
476 important insights into PCa evolution. The fact that Network Component 1 was more strongly affected in
477 the paired higher-grade nodules of high-grade patients suggests that at least certain sub-networks are
478 subject to an evolutionary process, that progressively 'moves' protein levels towards a more aggressive
479 state. Generally, and at all molecular layers tested, the two paired tumor areas were more similar to
480 each other compared to two samples from the same grade group but different patients, suggesting
481 common evolutionary origins. Although the two tumor areas seemed to mostly originate from the same
482 clone, this was not always the case. In some patients, different nodules exhibited different molecular
483 patterns at all omics layers, suggesting early evolutionary separation. Thus, for the first time, current
484 diagnostic, expert-level consensus guidelines [28] are supported by detailed proteogenomic data. Our
485 findings support earlier claims that clonality itself might be a prognostic marker with implications for
486 future, more tumor-specific treatment when targeted therapies become available also for PCa [16, 83].

487 Our study shows that all three molecular layers (genome, transcriptome and proteome)
488 contributed valuable information for understanding the biology of PCa. In particular the DNA layer
489 informed about causal events, clonality, and genomic similarity between tumors. The transcriptome was
490 relevant for understanding the transmission of CNA effects to proteins and served as a surrogate in cases
491 where protein levels remained undetected. The proteome was crucial for revealing protein-level
492 buffering of CNA effects as well as for indicating convergent evolution on functional endpoints. In a
493 routine diagnostic context though, measuring all three layers may not be feasible for the near future due
494 to resource and time limitations. Thus, the identification of improved, routine-usable molecular markers
495 for PCa diagnostics and prognosis remains an open problem [17].

496 # Conclusions

497 This study uncovered molecular networks with remarkably convergent alterations across tumor
498 sites and patients. In particular, we identified a sub-network consisting of nine genes whose joint activity
499 positively correlated with increasingly aggressive tumor phenotypes. The fact that this sub-network was
500 predictive for survival in independent cohorts further supports its potential clinical relevance. At the
501 same time though, our study also exposed a diversity of network effects: we could not identify a single
502 sub-network that was perturbed in all high-grade tumor regions, let alone the observed distinct intra-
503 patient alterations at all omics layers for some patients. Overall, our study has significantly expanded our
504 understanding of PCa biology and serves as a model for future work aiming to explore network effects of
505 mutations with an integrated multi-omics approach.

506 # Methods

507 ## Patients and samples

508 A total of 39 men with localized PCa who were scheduled for RP were selected from a cohort of
509 1,200 patients within the ProCOC study and processed at the Department of Pathology and Molecular
510 Pathology, University Hospital Zurich, Switzerland [25]. Each of the selected intermediate- and high-
511 grade patients had two different tumor nodules with different ISUP grade groups. H&E (Hematoxylin and
512 Eosin)-stained fresh frozen tissue sections of 105 selected BPH and tumor regions were evaluated by two
513 experienced pathologists (PJW, NJR) to assign malignancy, tumor stage, and Grade Group according to
514 the International Union Against Cancer (UICC) and WHO/ISUP criteria. This study was approved by the
515 Cantonal Ethics Committee of Zurich (KEK-ZH-No. 2008-0040), the associated methods were carried out
516 in accordance with the approved guidelines, and each patient has signed an informed consent form.
517 Patients were followed up on a regular basis (every three months in the first year and at least annually
518 thereafter) or on an individual basis depending on the disease course in the following years. The RFS was
519 calculated with a biochemical recurrence (BCR) defined as a PSA ≥0.1 ng/ml. Patients were censored if
520 lost to follow-up or event-free at their most recent clinic visit. Patients with a postoperative PSA
521 persistence or without distinct follow-up data for the endpoint BCR were excluded from the analysis of
522 BCR.

523 ## Exome sequencing and somatic variant analysis

524 The exome sequencing (exome-seq) was performed using the Agilent Sure Select Exome
525 platform for library construction and Illumina HiSeq 2500 for sequencing read generation. We mapped

526 and processed the reads using a pipeline based on bowtie2 [84] (1.1.1) and the Genome Analysis Tools
527 Kit (GATK) [85] (3.2-2). We detected and reported nonsynonymous variants or variants causing splicing
528 changes using Strelka (1.0.14) and Mutect (1.1.7) combined with post-processing by the CLC Genomics
529 Workbench (8.0.3). In this process, all tissue samples of a patient were compared to the respective blood
530 sample.

531 Trimmomatic [86] (0.36) was used for adaptor clipping and low-quality subsequence trimming of
532 the FASTQ files. Subsequently, single reads were aligned to the hg19 reference genome with bowtie2
533 with options "--very-sensitive -k 20". We applied samtools [87] (0.1.19) and picard-tools (1.119) to sort
534 the resulting bam files in coordinate order, merge different lanes, filter out all non-primary alignments,
535 and remove PCR duplicates. Quality of the runs was checked using a combination of BEDtools [88] (2.21),
536 samtools, R (3.1) and FastQC (0.11.2).

537 Bam files containing the mapped reads were preprocessed in the following way: indel
538 information was used to realign individual reads using the RealignerTargetCreator and IndelRealigner
539 option of the GATK. Mate-pair information between mates was verified and fixed using Picard tools and
540 single bases were recalibrated using GATK's BaseRecalibrator. After preprocessing, variant calling was
541 carried out by comparing benign or tumor prostate tissue samples with matched blood samples using
542 the programs MuTect [89] and Strelka [90] independently. Somatic variants that were only detected by
543 one of the two programs were filtered out using CLC Genomics Workbench. So were those that had an
544 entry in the dbSNP [91] common database and those that represented synonymous variants without
545 predicted effects on splicing.

546 CNA analysis of exome-seq data

547 The Bam files generated during the process of somatic variant calling were processed with the
548 CopywriteR package (v.2.2.0) for the R software [92]. CopywriteR makes use of so-called "off-target"
549 reads, *i.e.* reads that cover areas outside of the exon amplicons. "Off-target" reads are produced due to
550 inefficient enrichment strategies. In our case on average 28.5% of the total reads were not on target.
551 Briefly, CopywriteR removes low quality and anomalous read pairs, then peaks are called in the
552 respective blood reference, and all reads in this region are discarded. After mapping the reads into bins,
553 those peak regions, in which reads had been removed, were compensated for. Additionally, read counts
554 are corrected based on mappability and GC-content. Finally, a circular binary segmentation is carried out
555 and for each segment the log count ratios between tissue samples and the respective blood sample are
556 reported as copy number gain or loss. The copy number of each gene in each sample was reported

557 based on the log count ratio of the respective segment in which the gene was located. The overall
558 performance of this CNA-calling approach was evaluated by comparing the results of the TA1 (and TA)
559 samples with CNA results obtained by applying the OncoScan Microarray pipeline to FFPE samples from
560 the same tumors (**Additional file 1: Fig. S10**).

561 ## OncoScan Microarrays

562 OncoScan copy number assays were carried out and analyzed as described previously [93].
563 Briefly, DNA was extracted from punches of FFPE cancer tissue blocks. Locus-specific molecular inversion
564 probes were hybridized to complementary DNA and gaps were filled in a nucleotide-specific manner.
565 After amplification and cleavage of the probes, the probes were hybridized to the OncoScan assay
566 arrays. Scanning the fluorescence intensity and subsequent data processing using the Affymetrix®
567 GeneChip® Command Console and BioDiscovery Nexus express resulted in log intensity ratio data
568 (sample versus Affymetrix reference) and virtual segmentation of the genome into areas with copy
569 number gain, loss or stability.

570 ## RNA Sequencing

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

575 Paired-end reads were mapped to the human reference genome (GRCh37) using the STAR
576 aligner (version 2.4.2a) [94]. Quality control of the resulting bam files using QoRTs [95] and mRIN [96]
577 showed strong RNA degradation[97] in a significant fraction of the samples: mRIN classified 31 samples
578 as highly degraded (**Additional file 1: Fig. S11**, **Additional file 5: Table S4**). In order to correct for this 3'
579 bias, 3 tag counting was performed as described by Sigurgeirsson et al [98] using a tag length of 1,000.
580 After 3' bias correction, three samples still showed a clear 3' bias: the two tumor regions (TA1 and TA2)
581 of the patient M5 and TA2 from patient M8 (**Additional file 1: Fig. S11**). These samples were excluded
582 from subsequent analyses. Additionally, the BPH region of the patient M5 was excluded due to the
583 exclusion of both its tumor regions.

584 FeatureCounts [99] was used to determine read counts for all genes annotated in ENSEMBL v75.
585 Genes for which no read was observed in any of the samples in the original data were excluded from the
586 analysis. Further, after 3 tag counting, all genes with without at least 1 read per million in N of the

587 samples were removed. We chose N to be 10 which corresponds to the size of the smallest grade group
588 (G2). In a last reduction step, all genes with more than one transcript were excluded, yielding a final set
589 of 14,281 genes.

590 Read count normalization and differential gene expression analysis was performed using the R
591 packages sva [100] and DESeq2 [101]. All benign tissues were considered biological replicates and
592 differential gene expression for the individual tumor samples was determined against all benign tissues.
593 Gene expression changes with an adjusted *P* value < 0.1 were considered significant.

594 RNA-seq - 3' bias correction

595 The 3 tag counting approach for 3' bias correction was used on the RNA-seq dataset [98]. This
596 approach requires changing of the annotation file in two steps: 1) isoform filtering and 2) transcript
597 length restriction. As proposed in [98] for each gene we determined the highest expressed isoform
598 within a set of high quality samples. As high quality samples we used all samples with an mRIN score
599 greater than or equal to 0.02. This set contains 7 benign and 15 tumor samples. Isoform expression was
600 determined using cufflinks [102]. As transcript length we chose 1,000bp.

601 Gene fusions

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

605 PCT assisted sample preparation for SWATH-MS

606 We first washed each tissue sample to remove O.C.T., followed by PCT-assisted tissue lysis and
607 protein digestion, and SWATH-MS analysis, as described previously [23]. Briefly, a series of ethanol
608 solutions were used to wash the tissues each tissue, including 70% ethanol / 30% water (30 s), water (30
609 s), 70% ethanol / 30% water (5 min, twice), 85% ethanol / 15% water (5 min, twice), and 100% ethanol (5
610 min, twice). Subsequently, the tissue punches were lysed in PCT-MicroTubes with PCT-MicroPestle [103]
611 with 30 µl lysis buffer containing 8 M urea, 0.1 M ammonium bicarbonate, Complete protease inhibitor
612 cocktail (Roche) and PhosSTOP phosphatase inhibitor cocktail (Roche) using a barocycler (model
613 NEP2320-45k, PressureBioSciences, South Easton, MA). The lysis was performed with 60 cycles of high
614 pressure (45,000 p.s.i., 50 s per cycle) and ambient pressure (14.7 p.s.i., 10 s per cycle). The extracted
615 proteins were then reduced and alkylated prior to lys-C and trypsin-mediated proteolysis under pressure

616 cycling. Lys-C (Wako; enzyme-to-substrate ratio, 1:40) -mediated proteolysis was performed using 45
617 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
618 by trypsin (Promega; enzyme-to-substrate ratio, 1:20)-mediated proteolysis using the same cycling
619 scheme for 90 cycles. The resultant peptides were cleaned using SEP-PAC C18 (Waters Corp., Milford,
620 MA) and analyzed, after spike-in 10% iRT peptides [51], using SWATH-MS following the 32-fixed-size-
621 window scheme as described previously [19, 21] using a 5600 TripleTOF mass spectrometer (Sciex) and a
622 1D+ Nano LC system (Eksigent, Dublin, CA). The LC gradient was formulated with buffer A (2%
623 acetonitrile and 0.1% formic acid in HPLC water) and buffer B (2% water and 0.1% formic acid in
624 acetonitrile) through an analytical column (75 µm × 20 cm) and a fused silica PicoTip emitter (New
625 Objective, Woburn, MA, USA) with 3-µm 200 Å Magic C18 AQ resin (Michrom BioResources, Auburn, CA,
626 USA). Peptide samples were separated with a linear gradient of 2% to 35% buffer B over 120 min at a
627 flow rate of 0.3 µl min^{-1}. Ion accumulation time for MS1 and MS2 was set at 100 ms, leading to a total
628 cycle time of 3.3 s.

629 SWATH assay query library for prostate tissue proteome

630 To build a comprehensive library for SWATH data analysis, we analyzed unfractionated prostate
631 tissue digests prepared by the PCT method using Data Dependent Acquisition (DDA) mode in a tripleTOF
632 mass spectrometer over a gradient of 2 hours as described previously [19]. We spiked iRT peptides [51] into
633 each sample to enable retention time calibration among different samples. We then combined these
634 data with the DDA files from the pan-human library project [104]. All together we analyzed 422 DDA files
635 using X!Tandem [52] and OMSSA [53] against three protein sequence databases downloaded on Oct 21, 2016
636 from UniProt, including the SwissProt database of curated protein sequences (n=20,160), the splicing
637 variant database (n=21,970), and the trembl database (n=135,369). Using each database, we built target-
638 decoy protein sequence database by reversing the target protein sequences. We allowed maximal two
639 missed cleavages for fully tryptic peptides, and 50 p.p.m. for peptide precursor mass error, and 0.1 Da
640 for peptide fragment mass error. Static modification included carbamidomethyl at cysteine, while
641 variable modification included oxidation at methionine. Search results from X!Tandem and OMSSA were
642 further analyzed through Trans-Proteomic Pipeline (TPP, version 4.6.0) [54] using PeptideProphet and
643 iProphet, followed by SWATH assay library building procedures as detailed previously [19, 55]. Altogether,
644 we identified 167,402 peptide precursors, from which we selected the proteins detected in prostate
645 tissue samples, and built a sample-specific library. SWATH wiff files were converted into mzXML files
646 using ProteoWizard [56] msconvert v.3.0.3316, and then mzML files using OpenMS [57] tool FileConverter.

647 OpenSWATH[105] was performed using the tool OpenSWATHWorkflow with input files including the
648 mzXML file, the TraML library file, and TraML file for iRT peptides.

649 Peptide quantification using OpenSWATH

650 To obtain consistent quantification of the SWATH files, we obtained the all annotated *b* and *y*
651 fragments from the sp, sv and tr libraries. About ten thousand redundant and low-quality assays were
652 removed. Then we extracted the chromatography of these fragments and MS1 signals using
653 OpenSWATHWorkflow, followed by curation using DIA-expert[106]. Briefly, the chromatography of all
654 fragments and MS1 signals were subject to scrutiny by empirically developed expert rules. A reference
655 sample with best q value by pyprophet was picked up to refined fragments. The peptide precursors are
656 further filtered based on the following criteria: i) remove peptide precursors with a q value higher than
657 1.7783e-06 to achieve a false discovery rate of 0.00977 at peptide level using SWATH2stats [107]; ii)
658 peptides with a FC higher than 2 between the reference sample and its technical replicate were
659 removed; iii) peptides matching to multiple SwissProt protein sequences were removed. The data matrix
660 was first quantile normalized, $\log_2$ transformed, followed by batch correction using the ComBat R
661 package [108]. Finally, for each protein and pair of technical replicates the average value was computed.

662 Statistical analysis

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

665 Computation of molecular perturbation scores

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

674 We assigned to each sample two molecular perturbation scores summarizing/quantifying the
675 magnitude of its FCs: DE_count counts the number of mutated/differentially expressed (DE) genes, while

676 the DE_sum score is the sum of absolute FCs of all genes. Thus, while the first score counts the number
677 of events (mutations/DE genes), the second one quantifies their magnitude. A gene is regarded as
678 mutated/DE if its value is 1 in the mutation layer and if its absolute value is above a threshold that has
679 been set to 1 for the mRNA and protein layer. For the CNA layer, the corresponding threshold was set to
680 0.5 because the range of FCs in the CNA matrix is smaller than the mRNA and protein matrices. Both
681 types of scores were computed for each molecular level, except for the point mutations where only
682 DE_count was computed. Afterwards, the 66 DE_count scores (63 for the mRNA) and the DE_sum scores
683 at each layer were divided into the four grade groups G1, G2, G3 and G4/5 respectively.

684 Network propagation/smoothing

685 As a network, the STRING gene interaction network (version 10)[48] was used, after removing all
686 edges with combined score smaller or equal to 900 and keeping subsequently the largest connected
687 component. The resulting network consisted of 10,729 nodes and 118,647 (high-confidence) edges. For
688 the network smoothing, the weight matrix was computed as described in Vanunu et al.[49], but for an
689 unweighted graph and the propagation parameter was set to 0.5. The propagation was iteratively
690 repeated 500 times to ensure convergence of the results. For the mapping from gene symbols to STRING
691 identifiers (**Additional file 7: Table S6**) we used the R/Bioconductor package STRINGdb [109]. The gene
692 symbols with no matching STRING identifier were removed, while for those that mapped to multiple
693 STRING identifiers, the first mapping was kept (default choice in the package). From the multiple gene
694 symbols that mapped to the same STRING identifier, the first mapping was kept. The genes that were
695 not present in the network were removed from the datasets, while those that were present in the
696 network but not in the corresponding dataset were initially filled in with 0's.

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

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

707 ## Network Component 1 analysis

708 For each tumor sample at the CNA layer, a one-sided, one-sample t-test has been applied testing
709 if its average FC over the genes of the Network Component 1 (and in particular those that have been
710 measured at the CNA) is significantly greater than 0. Due to the presence of outliers in some samples,
711 the non-parametric, one-sided Wilcoxon signed-rank test has been applied as well yielding very similar
712 results (data not shown). A result is considered to be significant if the corresponding *P* value is below
713 0.05. The analysis has been repeated for the mRNA and protein layer.

714 ## Independent cohorts validation

715 For the validation of Network Component 1, we used published datasets of three PCa cohorts:
716 TCGA, MSKCC, and Aarhus. For TCGA and MSKCC, we downloaded the CNA, mRNA with precomputed z-
717 scores per gene, and corresponding clinical data from cBioPortal[110] (https://www.cbioportal.org/).
718 There were 489 samples with $\log_2$CNA data and 493 samples with mRNA profiles in TCGA. In MSKCC,
719 there were 157 primary tumors with CNA data and 131 primary tumors with mRNA data. The clinical
720 endpoint used in TCGA was the progression-free survival time and the disease-free survival in MSKCC. All
721 previous samples had known survival time.

722 For the Aarhus study (NCBI GEO dataset GSE46602), we downloaded the mRNA matrix and
723 corresponding clinical information as described in Ycart et al [111]. The resulting mRNA matrix consisted
724 of 20,186 genes and 50 samples- 40 PCa samples with known RFS time and 10 benign samples. Once
725 excluding the benign samples, we computed z-scores per gene in order to have comparable values with
726 the other two studies. These 40 PCa samples were also considered in the subsequent survival analysis.
727 CNA data was not available for the Aarhus study.

728 We reduced all datasets to the nine genes of Network Component 1. In each of the datasets, we
729 computed for each sample an average copy number change (CNA) or an average z-score (mRNA) across
730 the nine genes of Network Component 1 (combined risk score). Subsequently, we used these combined
731 risk scores to split the samples of each dataset into two groups: samples with a combined risk score
732 larger or equal to the median combined risk score of the study were considered as 'altered' and the rest
733 as 'unaltered'. Kaplan-Meier curves were generated for the two groups. Due to the high level of
734 discretized values in MSKCC at the CNA layer, a sample is considered to be 'altered' in that dataset if its
735 combined risk score is above zero.

736 ## Analysis of regulators and target genes

737 For this analysis, we used an independently inferred (partially directed) generic transcriptional
738 regulatory network (Leote *et al. in revision*; preprint available on bioRxiv). For each target gene the
739 following procedure was applied: firstly we identified its neighborhood of order two in the
740 transcriptional network by considering incoming edges only, *i.e.* all nodes from which the target gene
741 was reachable in at most two steps (equivalently the node itself, its parents and the parents of parents).
742 These are potential regulators of the target under consideration. Of these nodes, genes with no copy
743 number measurements and genes altered (*i.e.* with $\log_2$ copy number ratio greater than 0.5 in absolute)
744 in fewer than two tumor samples across the 66 tumor samples were filtered out. Subsequently, we fitted
745 an elastic net model with alpha=0.5. We used as output variable the mRNA FC of the target gene and as
746 input variables the CNAs of the regulators after the filtering. The value for the regularization parameter
747 lambda was chosen through 10-fold cross validation (default in the R package glmnet (https://cran.r-
748 project.org/web/packages/glmnet/)). Predictors/regulators with a non-zero beta coefficient were
749 deemed significant. We have used the elastic net model with alpha=0.5 because it is a method giving
750 sparse solutions and can deal with correlated predictors at the same time.

751 For the validation of our approach, we used the two independent PCa cohorts described above
752 (TCGA and MSKCC) and reduced the samples to those having both CNA and mRNA profile. This resulted
753 in 488 samples for TCGA and 109 samples for MSKCC. Next, for each of the significant
754 regulators/predictors we computed the Spearman correlation between its CNAs and the corresponding
755 mRNA z-scores of the target gene in each of the two independent studies. Finally, for each target gene
756 and each study we counted how many times the sign of the Spearman correlation matched the sign of
757 the Spearman correlation computed for our cohort, *i.e.* there was an agreement regarding the direction
758 of the association (**Additional file 7: Table S6**).

759 ## Differences between similarity/distance measures

760 A mathematical explanation for the observed differences between the similarity/distance
761 measures is provided. Let $\boldsymbol{y_i^G}$ denote the vector of $\log_2$ fold changes ($\log_2$FCs) corresponding to the *i*-th
762 tumor sample in the grade group G ($G \in \{G1, G2, G3, G4/5\}$), *i.e.* the fold changes of protein/mRNA
763 concentration in the respective tumor sample versus the global benign control. Then we can write:

$$\boldsymbol{y_i^G} = \begin{pmatrix} y_{i1}^G \\ y_{i2}^G \\ \vdots \\ y_{iN}^G \end{pmatrix} = \begin{pmatrix} \mu_1^G \\ \mu_2^G \\ \vdots \\ \mu_N^G \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1}^G \\ \varepsilon_{i2}^G \\ \vdots \\ \varepsilon_{iN}^G \end{pmatrix}, \qquad (1)$$

where y_{ij}^G corresponds to the $\log_2$FC of protein/mRNA *j* in the respective tumor sample (*i.e. i*-th tumor sample in the grade group *G*), μ_j^G to its (population) mean $\log_2$FC in the grade group *G* and ε_{ij}^G to the deviation of the respective individual $\log_2$FC from the population mean (*j*=1,…*N*), *N* being the total number of proteins/genes. The population mean μ_j^G is estimated by computing the average $\log_2$FC across all tumor samples in that grade group for the protein/mRNA *j*. Equivalently, (1) can be re-written:

$$\begin{pmatrix} y_{i1}^G \\ y_{i2}^G \\ \vdots \\ y_{iN}^G \end{pmatrix} - \begin{pmatrix} \mu_1^G \\ \mu_2^G \\ \vdots \\ \mu_N^G \end{pmatrix} = \begin{pmatrix} \varepsilon_{i1}^G \\ \varepsilon_{i2}^G \\ \vdots \\ \varepsilon_{iN}^G \end{pmatrix}.$$

On the one hand, the Euclidean and Manhattan distance between one tumor sample and the corresponding mean/centroid vector are functions of these error terms ε_{ij}^G's alone and thus only dependent on the variances of the proteins/genes in the corresponding grade group – these are shown to be higher in the high-grade tumors. For the Pearson correlation *r* on the other hand, one can use the equivalent simple linear regression modeling in (1) with the beta coefficient being equal to 1. By standardizing the vector $\boldsymbol{y_i^G}$ and the centroid vector, we have that $r = 1\sqrt{var(centroid\ vector)/var(\boldsymbol{y_i^G})} = \sqrt{var(centroid\ vector)/(var(centroid\ vector) + var(residuals))} = \sqrt{1/(1 + var(centroid\ vector)/var(residuals))}$ where *var* denotes the variance. The Pearson correlation will thus depend both on the variance of the residuals and the variance of the predictor, *i.e.* mean/centroid vector in the corresponding grade group. Both quantities are expected to be higher in the high-grade samples, while our data implies that the variance of the centroids (predictors) dominates over the variance of the residuals.

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 (http://cran.r-project.org/web/packages/infotheo/index.html) has been used. For the MI, the data needs to be

789 discrete. For that, we discretize the tumor-to-benign FCs and centroids by setting 1, if the value is above
790 a threshold *thr*, -1, if the value is below -*thr* and 0 else. (*thr* is equal to 1 for the mRNA and proteins and
791 0.5 for the CNAs.)

792 Availability of data and materials

793 Exome and RNA sequencing data were submitted to the Sequence Read Archive (SRA) at NCBI under
794 accession numbers PRJNA577801 (exome-seq) and PRJNA579899 (RNA-seq), respectively. The SWATH
795 proteomics data were deposited in PRIDE. Project accession code is PXD004589. The published datasets
796 of the two PCa cohorts (TCGA and MSKCC) analyzed during the current study can be downloaded from
797 cBioPortal[110] while the third (Aarhus) is available at the NCBI GEO repository under the accession
798 number GSE46602 .

799 Acknowledgements

800 Not applicable.

801 Funding

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

809 Authors' contributions

810 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.,
811 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.,
812 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
813 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
814 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.,
815 P.J.W. and T.G. supported and supervised the project.

Ethics declarations

Ethics approval and consent to participate

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.

Consent for publication

Not applicable.

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.

Supplementary information

Additional file 1: Supplementary text and supplementary figures.

Additional file 2: Table S1. 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 the '**Methods**' section (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.

Additional file 3: Table S2. 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.

844 **Additional file 4: Table S3. Copy number analysis of 105 PCa samples.** (a) Log_2 ratios indicating the CNA
845 status are shown for all genes in all samples. Values were determined by overlapping gene locations with
846 CNA segments as calculated by CopywriteR. In case more than one segment overlapped with a gene,
847 number was chosen that had the highest absolute value. (b) Genes are shown with log_2 ratios higher
848 than 0.5 or lower than -0.5 in at least one sample.

849 **Additional file 5: Table S4. RNA-seq analysis.** (a) Log_2FCs (relative to all benign samples) for all genes
850 across the tumor samples. (b) mRIN score per sample generated using mRIN (v1.2.0). (c) ETS family gene
851 fusions observed in tumor samples using FusionCatcher: a value of 1 means that the fusion was observed
852 in the respective sample but not its corresponding benign sample, otherwise the value is 0.

853 **Additional file 6: Table S5. Proteomics data of 210 PCa samples with duplicates.** (a) Sample information
854 includes patient ID, clinical diagnosis, sample ID and batch design. (b) Protein matrix of log_2 scaled
855 intensity of 2,371 proteins quantified in 210 PCa samples.

856 **Additional file 7: Table S6. Integration analysis of 66 tumor samples.** (a) L1-norm of the 'centroid
857 vectors' in the three layers (CNA, mRNA and protein) across the four grade groups. (b) Information (i.e.
858 reference linking them to PCa, consistency between observed and reported effect and number of tumor
859 samples with CNAs) for the first 10 highest-scoring proteins (those with largest average absolute FCs
860 across all tumor specimens). (c) For each target gene the regulators with a non-zero beta coefficient
861 from the elastic net model are given. Further, the proportion of the elastic net predictors with the same
862 sign of Spearman correlation in our cohort and the independent study are noted for the two
863 independent PCa cohorts. (d) Consistently up-regulated genes in the high-grade tumors: for each of
864 these genes, there is a significant up-regulation of its FCs after network smoothing in the group G4/5
865 compared to the group G1 in all three layers (CNA, mRNA and protein). (e) Consistently down-regulated
866 genes in the high-grade tumors: for each of these genes, there is a significant down-regulation of its FCs
867 after network smoothing in the group G4/5 compared to the group G1 in all three layers (CNA, mRNA
868 and protein). (f) Chromosome information for the gene members of Network Component 1. (g) Mapping
869 from gene symbols to STRING identifiers.

870

871 # References

872

873 1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F:
874 **Cancer incidence and mortality worldwide: sources, methods and major patterns in**
875 **GLOBOCAN 2012.** *Int J Cancer* 2015, **136:**E359-386.
876 2. Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, Vickers AJ, Parwani AV,
877 Reuter VE, Fine SW, et al: **A Contemporary Prostate Cancer Grading System: A Validated**
878 **Alternative to the Gleason Score.** *Eur Urol* 2016, **69:**428-435.
879 3. Gordetsky J, Epstein J: **Grading of prostatic adenocarcinoma: current state and prognostic**
880 **implications.** *Diagn Pathol* 2016, **11:**25.
881 4. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R,
882 Pflueger D, Sougnez C, et al: **The genomic complexity of primary human prostate cancer.**
883 *Nature* 2011, **470:**214-220.
884 5. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P,
885 Van Allen E, Stransky N, et al: **Exome sequencing identifies recurrent SPOP, FOXA1 and MED12**
886 **mutations in prostate cancer.** *Nat Genet* 2012, **44:**685-689.
887 6. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin
888 ME, Pritchard CC, Attard G, et al: **Integrative clinical genomics of advanced prostate cancer.** *Cell*
889 2015, **161:**1215-1228.
890 7. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N,
891 MacDonald TY, Ghandi M, et al: **Punctuated evolution of prostate cancer genomes.** *Cell* 2013,
892 **153:**666-677.
893 8. Cancer Genome Atlas Research N: **The Molecular Taxonomy of Primary Prostate Cancer.** *Cell*
894 2015, **163:**1011-1025.
895 9. Wedge DC, Gundem G, Mitchell T, Woodcock DJ, Martincorena I, Ghori M, Zamora J, Butler A,
896 Whitaker H, Kote-Jarai Z, et al: **Sequencing of prostate cancers identifies new cancer genes,**
897 **routes of progression and drug targets.** *Nat Genet* 2018, **50:**682-692.
898 10. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X,
899 Tchinda J, Kuefer R, et al: **Recurrent fusion of TMPRSS2 and ETS transcription factor genes in**
900 **prostate cancer.** *Science* 2005, **310:**644-648.
901 11. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ,
902 Brenner JC, et al: **The mutational landscape of lethal castration-resistant prostate cancer.**
903 *Nature* 2012, **487:**239-243.
904 12. Boutros PC, Fraser M, Harding NJ, de Borja R, Trudel D, Lalonde E, Meng A, Hennings-Yeomans
905 PH, McPherson A, Sabelnykova VY, et al: **Spatial genomic heterogeneity within localized,**
906 **multifocal prostate cancer.** *Nat Genet* 2015, **47:**736-745.
907 13. Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, Kremeyer B, Butler A,
908 Lynch AG, Camacho N, et al: **Analysis of the genetic phylogeny of multifocal prostate cancer**
909 **identifies multiple independent clonal expansions in neoplastic and morphologically normal**
910 **prostate tissue.** *Nat Genet* 2015, **47:**367-372.
911 14. Lovf M, Zhao S, Axcrona U, Johannessen B, Bakken AC, Carm KT, Hoff AM, Myklebost O, Meza-
912 Zepeda LA, Lie AK, et al: **Multifocal Primary Prostate Cancer Exhibits High Degree of Genomic**
913 **Heterogeneity.** *Eur Urol* 2019, **75:**498-505.
914 15. Lindberg J, Klevebring D, Liu W, Neiman M, Xu J, Wiklund P, Wiklund F, Mills IG, Egevad L,
915 Gronberg H: **Exome sequencing of prostate cancer supports the hypothesis of independent**
916 **tumour origins.** *Eur Urol* 2013, **63:**347-353.

917 16. Espiritu SMG, Liu LY, Rubanova Y, Bhandari V, Holgersen EM, Szyca LM, Fox NS, Chua MLK,
918 Yamaguchi TN, Heisler LE, et al: **The Evolutionary Landscape of Localized Prostate Cancers**
919 **Drives Clinical Aggression.** *Cell* 2018, **173:**1003-1013 e1015.
920 17. Sinha A, Huang V, Livingstone J, Wang J, Fox NS, Kurganovs N, Ignatchenko V, Fritsch K, Donmez
921 N, Heisler LE, et al: **The Proteogenomic Landscape of Curable Prostate Cancer.** *Cancer Cell* 2019,
922 **35:**414-427 e416.
923 18. Latonen L, Afyounian E, Jylha A, Nattinen J, Aapola U, Annala M, Kivinummi KK, Tammela TTL,
924 Beuerman RW, Uusitalo H, et al: **Integrative proteomics in prostate cancer uncovers robustness**
925 **against genomic and transcriptomic aberrations during disease progression.** *Nat Commun*
926 2018, **9:**1176.
927 19. Iglesias-Gato D, Wikstrom P, Tyanova S, Lavallee C, Thysell E, Carlsson J, Hagglof C, Cox J, Andren
928 O, Stattin P, et al: **The Proteome of Primary Prostate Cancer.** *Eur Urol* 2016, **69:**942-952.
929 20. Attard G, Parker C, Eeles RA, Schroder F, Tomlins SA, Tannock I, Drake CG, de Bono JS: **Prostate**
930 **cancer.** *Lancet* 2016, **387:**70-82.
931 21. Goncalves E, Fragoulis A, Garcia-Alonso L, Cramer T, Saez-Rodriguez J, Beltrao P: **Widespread**
932 **Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer.** *Cell Syst* 2017,
933 **5:**386-398 e384.
934 22. Liu Y, Beyer A, Aebersold R: **On the Dependency of Cellular Protein Levels on mRNA**
935 **Abundance.** *Cell* 2016, **165:**535-550.
936 23. Guo T, Kouvonen P, Koh CC, Gillet LC, Wolski WE, Rost HL, Rosenberger G, Collins BC, Blum LC,
937 Gillessen S, et al: **Rapid mass spectrometric conversion of tissue biopsy samples into**
938 **permanent quantitative digital proteome maps.** *Nat Med* 2015, **21:**407-413.
939 24. Guo T, Li L, Zhong Q, Rupp NJ, Charmpi K, Wong CE, Wagner U, Rueschoff JH, Jochum W,
940 Fankhauser CD, et al: **Multi-region proteome analysis quantifies spatial heterogeneity of**
941 **prostate tissue biomarkers.** *Life Sci Alliance* 2018, **1**.
942 25. Umbehr M, Kessler TM, Sulser T, Kristiansen G, Probst N, Steurer J, Bachmann LM: **ProCOC: the**
943 **prostate cancer outcomes cohort study.** *BMC Urol* 2008, **8:**9.
944 26. Wettstein MS, Saba K, Umbehr MH, Murtola TJ, Fankhauser CD, Adank JP, Hofmann M, Sulser T,
945 Hermanns T, Moch H, et al: **Prognostic Role of Preoperative Serum Lipid Levels in Patients**
946 **Undergoing Radical Prostatectomy for Clinically Localized Prostate Cancer.** *Prostate* 2017,
947 **77:**549-556.
948 27. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, Grading C: **The 2014**
949 **International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading**
950 **of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System.**
951 *Am J Surg Pathol* 2016, **40:**244-252.
952 28. Epstein JI, Amin MB, Reuter VE, Humphrey PA: **Contemporary Gleason Grading of Prostatic**
953 **Carcinoma: An Update With Discussion on Practical Issues to Implement the 2014**
954 **International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading**
955 **of Prostatic Carcinoma.** *Am J Surg Pathol* 2017, **41:**e1-e7.
956 29. Reznik E, Sander C: **Extensive decoupling of metabolic genes in cancer.** *PLoS Comput Biol* 2015,
957 **11:**e1004176.
958 30. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E,
959 Reva B, et al: **Integrative genomic profiling of human prostate cancer.** *Cancer Cell* 2010, **18:**11-
960 22.
961 31. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S,
962 et al: **Proteogenomic characterization of human colon and rectal cancer.** *Nature* 2014, **513:**382-
963 387.

964 32. Geiger T, Cox J, Mann M: **Proteomic changes resulting from gene copy number variations in**
965 **cancer cells.** *PLoS Genet* 2010, **6:**e1001090.
966 33. Juschke C, Dohnal I, Pichler P, Harzer H, Swart R, Ammerer G, Mechtler K, Knoblich JA:
967 **Transcriptome and proteome quantification of a tumor model provides novel insights into**
968 **post-transcriptional gene regulation.** *Genome Biol* 2013, **14:**r133.
969 34. Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z: **Global analysis of genome,**
970 **transcriptome and proteome reveals the response to aneuploidy in human cells.** *Mol Syst Biol*
971 2012, **8:**608.
972 35. Cover TM, Thomas JA: *Elements of information theory.* John Wiley & Sons; 2012.
973 36. Zhang JS, Gong A, Cheville JC, Smith DI, Young CY: **AGR2, an androgen-inducible secretory**
974 **protein overexpressed in prostate cancer.** *Genes Chromosomes Cancer* 2005, **43:**249-259.
975 37. Liu Q, Harvey CT, Geng H, Xue C, Chen V, Beer TM, Qian DZ: **Malate dehydrogenase 2 confers**
976 **docetaxel resistance via regulations of JNK signaling and oxidative metabolism.** *Prostate* 2013,
977 **73:**1028-1037.
978 38. Yang J, Song H, Chen L, Cao K, Zhang Y, Li Y, Hao X: **Integrated analysis of microfibrillar-**
979 **associated proteins reveals MFAP4 as a novel biomarker in human cancers.** *Epigenomics* 2019,
980 **11:**1635-1651.
981 39. Senga S, Kobayashi N, Kawaguchi K, Ando A, Fujii H: **Fatty acid-binding protein 5 (FABP5)**
982 **promotes lipolysis of lipid droplets, de novo fatty acid (FA) synthesis and activation of nuclear**
983 **factor-kappa B (NF-kappaB) signaling in cancer cells.** *Biochim Biophys Acta Mol Cell Biol Lipids*
984 2018, **1863:**1057-1067.
985 40. Pan Y, Liu Z, Feng Z, Hui D, Huang X, Tong D, Jin Y: **The overexpression of Rabl3 is associated**
986 **with pathogenesis and clinicopathologic variables in hepatocellular carcinoma.** *Tumour Biol*
987 2017, **39:**1010428317696230.
988 41. Zhang W, Sun J, Luo J: **High Expression of Rab-like 3 (Rabl3) is Associated with Poor Survival of**
989 **Patients with Non-Small Cell Lung Cancer via Repression of MAPK8/9/10-Mediated**
990 **Autophagy.** *Med Sci Monit* 2016, **22:**1582-1588.
991 42. Hofree M, Shen JP, Carter H, Gross A, Ideker T: **Network-based stratification of tumor**
992 **mutations.** *Nat Methods* 2013, **10:**1108-1115.
993 43. Muniyan S, Chaturvedi NK, Dwyer JG, Lagrange CA, Chaney WG, Lin MF: **Human prostatic acid**
994 **phosphatase: structure, function and regulation.** *Int J Mol Sci* 2013, **14:**10438-10464.
995 44. Araujo CL, Quintero IB, Ovaska K, Herrala AM, Hautaniemi S, Vihko PT: **Transmembrane**
996 **prostatic acid phosphatase (TMPAP) delays cells in G1 phase of the cell cycle.** *Prostate* 2016,
997 **76:**151-162.
998 45. Qiu Y, Yang S, Pan T, Yu L, Liu J, Zhu Y, Wang H: **ANKRD22 is involved in the progression of**
999 **prostate cancer.** *Oncol Lett* 2019, **18:**4106-4113.
1000 46. Donald CD, Sun CQ, Lim SD, Macoska J, Cohen C, Amin MB, Young AN, Ganz TA, Marshall FF,
1001 Petros JA: **Cancer-specific loss of beta-defensin 1 in renal and prostatic carcinomas.** *Lab Invest*
1002 2003, **83:**501-505.
1003 47. Siltanen S, Fischer D, Rantapero T, Laitinen V, Mpindi JP, Kallioniemi O, Wahlfors T, Schleutker J:
1004 **ARLTS1 and prostate cancer risk--analysis of expression and regulation.** *PLoS One* 2013,
1005 **8:**e72040.
1006 48. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A,
1007 Santos A, Tsafou KP, et al: **STRING v10: protein-protein interaction networks, integrated over**
1008 **the tree of life.** *Nucleic Acids Research* 2015, **43:**D447-D452.
1009 49. Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R: **Associating Genes and Protein Complexes**
1010 **with Disease via Network Propagation.** *Plos Computational Biology* 2010, **6**.

1011 50. Zhou D, Bousquet O, Lal TN, Weston J, Schölkopf B: **Learning with Local and Global Consistency.**
1012 *Advances in Neural Information Processing Systems* 2004, **16:**312-328.
1013 51. Baude A, Aaes TL, Zhai B, Al-Nakouzi N, Oo HZ, Daugaard M, Rohde M, Jaattela M: **Hepatoma-**
1014 **derived growth factor-related protein 2 promotes DNA repair by homologous recombination.**
1015 *Nucleic Acids Res* 2016, **44:**2214-2226.
1016 52. Guler GD, Tindell CA, Pitti R, Wilson C, Nichols K, Kaiwai CT, Kim HJ, Wongchenko M, Yan Y, Haley
1017 B: **Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from**
1018 **Lethal Drug Exposure.** *Cancer Cell* 2017, **32:**221.
1019 53. Slezak J, Truong M, Huang W, Jarrard D: **HP1gamma expression is elevated in prostate cancer**
1020 **and is superior to Gleason score as a predictor of biochemical recurrence after radical**
1021 **prostatectomy.** *BMC Cancer* 2013, **13:**148.
1022 54. Ceol CJ, Yariv H, Judit JV, Steve B, Orlando DA, Valentine B, Lauriane F, Lin WM, Hollmann TJ,
1023 Fabrizio F: **The SETDB1 histone methyltransferase is recurrently amplified in and accelerates**
1024 **melanoma.** *Nature* 2011, **471:**513-517.
1025 55. Sun Y, Wei M, Ren SC, Chen R, Xu WD, Wang FB, Lu J, Shen J, Yu YW, Hou JG, et al: **Histone**
1026 **methyltransferase SETDB1 is required for prostate cancer cell proliferation, migration and**
1027 **invasion.** *Asian J Androl* 2014, **16:**319-324.
1028 56. Luo H, Cowen L, Yu G, Jiang W, Tang Y: **SMG7 is a critical regulator of p53 stability and function**
1029 **in DNA damage stress response.** *Cell Discovery* 2016, **2:**15042.
1030 57. Karanika S, Karantanos T, Li L, Wang J, Park S, Yang G, Zuo X, Song JH, Maity SN, Manyam GC:
1031 **Targeting DNA Damage Response in Prostate Cancer by Inhibiting Androgen Receptor-CDC6-**
1032 **ATR-Chk1 Signaling.** *Cell Reports* 2017, **18:**1970.
1033 58. Diouf B, Cheng Q, Krynetskaia NF, Yang W, Cheok M, Pei D, Fan Y, Cheng C, Krynetskiy EY, Geng
1034 H: **Somatic deletions of genes regulating MSH2 protein stability cause DNA mismatch repair**
1035 **deficiency and drug resistance in human leukemia cells.** *Nature Medicine* 2011, **17:**1298-1303.
1036 59. Yao S, Bee A, Brewer D, Dodson A, Beesley C, Ke Y, Ambroisine L, Fisher G, Møller H, Dickinson T:
1037 **PRKC-ζ Expression Promotes the Aggressive Phenotype of Human Prostate Cancer Cells and Is**
1038 **a Novel Target for Therapeutic Intervention.** *Genes Cancer* 2010, **1:**444-464.
1039 60. Zhou W, Jiang Y, Zhu M, Hang D, Chen J, Zhou J, Dai J, Ma H, Hu Z, Jin G, et al: **Low-frequency**
1040 **nonsynonymous variants in FKBPL and ARPC1B genes are associated with breast cancer risk in**
1041 **Chinese women.** *Mol Carcinog* 2017, **56:**774-780.
1042 61. De Robertis M, Mazza T, Fusilli C, Loiacono L, Poeta ML, Sanchez M, Massi E, Lamorte G, Diodoro
1043 MG, Pescarmona E, et al: **EphB2 stem-related and EphA2 progression-related miRNA-based**
1044 **networks in progressive stages of CRC evolution: clinical significance and potential miRNA**
1045 **drivers.** *Mol Cancer* 2018, **17:**169.
1046 62. Zeng Z, Yang H, Xiao S: **ACTL6A expression promotes invasion, metastasis and epithelial**
1047 **mesenchymal transition of colon cancer.** *BMC Cancer* 2018, **18:**1020.
1048 63. Castro MA, Dal-Pizzol F, Zdanov S, Soares M, Muller CB, Lopes FM, Zanotto-Filho A, da Cruz
1049 Fernandes M, Moreira JC, Shacter E, Klamt F: **CFL1 expression levels as a prognostic and drug**
1050 **resistance marker in nonsmall cell lung cancer.** *Cancer* 2010, **116:**3645-3655.
1051 64. Overbye A, Skotland T, Koehler CJ, Thiede B, Seierstad T, Berge V, Sandvig K, Llorente A:
1052 **Identification of prostate cancer biomarkers in urinary exosomes.** *Oncotarget* 2015, **6:**30357-
1053 30376.
1054 65. Deng Z, Wan M, Cao P, Rao A, Cramer SD, Sui G: **Yin Yang 1 regulates the transcriptional activity**
1055 **of androgen receptor.** *Oncogene* 2009, **28:**3746-3757.
1056 66. Fournier PG, Juarez P, Jiang G, Clines GA, Niewolna M, Kim HS, Walton HW, Peng XH, Liu Y,
1057 Mohammad KS, et al: **The TGF-beta Signaling Regulator PMEPA1 Suppresses Prostate Cancer**
1058 **Metastases to Bone.** *Cancer Cell* 2015, **27:**809-821.

1059 67. Yang J, Lu C, Wei J, Guo Y, Liu W, Luo L, Fisch G, Li X: **Inhibition of KPNA4 attenuates prostate**
1060 **cancer metastasis.** *Oncogene* 2017, **36:**2868-2878.
1061 68. Liu W, Xie CC, Zhu Y, Li T, Sun J, Cheng Y, Ewing CM, Dalrymple S, Turner AR, Sun J, et al:
1062 **Homozygous deletions and recurrent amplifications implicate new genes involved in prostate**
1063 **cancer.** *Neoplasia* 2008, **10:**897-907.
1064 69. Trasino SE, Harrison EH, Wang TT: **Androgen regulation of aldehyde dehydrogenase 1A3**
1065 **(ALDH1A3) in the androgen-responsive human prostate cancer cell line LNCaP.** *Exp Biol Med*
1066 *(Maywood)* 2007, **232:**762-771.
1067 70. Bonaccorsi L, Luciani P, Nesi G, Mannucci E, Deledda C, Dichiara F, Paglierani M, Rosati F, Masieri
1068 L, Serni S, et al: **Androgen receptor regulation of the seladin-1/DHCR24 gene: altered**
1069 **expression in prostate cancer.** *Lab Invest* 2008, **88:**1049-1056.
1070 71. Li H, Xiao D, Hu L, He T: **Association of CYP1A1 polymorphisms with prostate cancer risk: an**
1071 **updated meta-analysis.** *Mol Biol Rep* 2012, **39:**10273-10284.
1072 72. Aktas D, Hascicek M, Sozen S, Ozen H, Tuncbilek E: **CYP1A1 and GSTM1 polymorphic genotypes**
1073 **in patients with prostate cancer in a Turkish population.** *Cancer Genet Cytogenet* 2004, **154:**81-
1074 85.
1075 73. Acevedo C, Opazo JL, Huidobro C, Cabezas J, Iturrieta J, Quinones Sepulveda L: **Positive**
1076 **correlation between single or combined genotypes of CYP1A1 and GSTM1 in relation to**
1077 **prostate cancer in Chilean people.** *Prostate* 2003, **57:**111-117.
1078 74. Cao HM, Wan Z, Wu Y, Wang HY, Guan C: **Development and internal validation of a novel**
1079 **model and markers to identify the candidates for lymph node metastasis in patients with**
1080 **prostate cancer.** *Medicine (Baltimore)* 2019, **98:**e16534.
1081 75. Kawamura N, Nimura K, Saga K, Ishibashi A, Kitamura K, Nagano H, Yoshikawa Y, Ishida K,
1082 Nonomura N, Arisawa M, et al: **SF3B2-Mediated RNA Splicing Drives Human Prostate Cancer**
1083 **Progression.** *Cancer Res* 2019, **79:**5204-5217.
1084 76. Wen M, Kwon Y, Wang Y, Mao JH, Wei G: **Elevated expression of UBE2T exhibits oncogenic**
1085 **properties in human prostate cancer.** *Oncotarget* 2015, **6:**25226-25239.
1086 77. Kleppe A, Albregtsen F, Vlatkovic L, Pradhan M, Nielsen B, Hveem TS, Askautrud HA, Kristensen
1087 GB, Nesbakken A, Trovik J, et al: **Chromatin organisation and cancer prognosis: a pan-cancer**
1088 **study.** *Lancet Oncol* 2018, **19:**356-369.
1089 78. Lee RS, Roberts CW: **Linking the SWI/SNF complex to prostate cancer.** *Nat Genet* 2013,
1090 **45:**1268-1269.
1091 79. Chuu C-P, Lin C-Y, Huang S-H, Tsai KK-C: **MO2-10-6ACTL6A is a novel oncogene and prognostic**
1092 **biomarker for prostate cancer.** *Annals of Oncology* 2019, **30**.
1093 80. Doyon Y, Selleck W, Lane WS, Tan S, Cote J: **Structural and functional conservation of the NuA4**
1094 **histone acetyltransferase complex from yeast to humans.** *Mol Cell Biol* 2004, **24:**1884-1896.
1095 81. Cai Y, Jin J, Florens L, Swanson SK, Kusch T, Li B, Workman JL, Washburn MP, Conaway RC,
1096 Conaway JW: **The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone**
1097 **acetyltransferase and SRCAP complexes.** *J Biol Chem* 2005, **280:**13665-13670.
1098 82. Mortensen MM, Hoyer S, Lynnerup AS, Orntoft TF, Sorensen KD, Borre M, Dyrskjot L: **Expression**
1099 **profiling of prostate cancer tissue delineates genes associated with recurrence after**
1100 **prostatectomy.** *Sci Rep* 2015, **5:**16018.
1101 83. Hussain M, Mateo J, Fizazi K, Saad F, Shore ND, Sandhu S, Chi KN, Sartor O, Agarwal N, Olmos D,
1102 et al: **LBA12_PRPROfound: Phase III study of olaparib versus enzalutamide or abiraterone for**
1103 **metastatic castration-resistant prostate cancer (mCRPC) with homologous recombination**
1104 **repair (HRR) gene alterations.** *Annals of Oncology* 2019, **30**.
1105 84. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9:**357-
1106 359.

1107 85. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,
1108 Gabriel S, Daly M, DePristo MA: **The Genome Analysis Toolkit: a MapReduce framework for**
1109 **analyzing next-generation DNA sequencing data.** *Genome Res* 2010, **20:**1297-1303.
1110 86. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.**
1111 *Bioinformatics* 2014, **30:**2114-2120.
1112 87. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
1113 Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.**
1114 *Bioinformatics* 2009, **25:**2078-2079.
1115 88. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.**
1116 *Bioinformatics* 2010, **26:**841-842.
1117 89. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M,
1118 Lander ES, Getz G: **Sensitive detection of somatic point mutations in impure and**
1119 **heterogeneous cancer samples.** *Nat Biotechnol* 2013, **31:**213-219.
1120 90. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK: **Strelka: accurate somatic**
1121 **small-variant calling from sequenced tumor-normal sample pairs.** *Bioinformatics* 2012,
1122 **28:**1811-1817.
1123 91. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K: **dbSNP: the NCBI**
1124 **database of genetic variation.** *Nucleic Acids Res* 2001, **29:**308-311.
1125 92. Kuilman T, Velds A, Kemper K, Ranzani M, Bombardelli L, Hoogstraat M, Nevedomskaya E, Xu G,
1126 de Ruiter J, Lolkema MP, et al: **CopywriteR: DNA copy number detection from off-target**
1127 **sequence data.** *Genome Biol* 2015, **16:**49.
1128 93. Noske A, Brandt S, Valtcheva N, Wagner U, Zhong Q, Bellini E, Fink D, Obermann EC, Moch H,
1129 Wild PJ: **Detection of CCNE1/URI (19q12) amplification by in situ hybridisation is common in**
1130 **high grade and type II endometrial cancer.** *Oncotarget* 2017, **8:**14794-14805.
1131 94. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR:
1132 **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29:**15-21.
1133 95. Hartley SW, Mullikin JC: **QoRTs: a comprehensive toolset for quality control and data**
1134 **processing of RNA-Seq experiments.** *BMC Bioinformatics* 2015, **16:**224.
1135 96. Feng H, Zhang X, Zhang C: **mRIN for direct assessment of genome-wide and gene-specific mRNA**
1136 **integrity from large-scale RNA-sequencing data.** *Nat Commun* 2015, **6:**7816.
1137 97. Shao W, Guo T, Toussaint NC, Xue P, Wagner U, Li L, Charmpi K, Zhu Y, Wu J, Buljan M, et al:
1138 **Comparative analysis of mRNA and protein degradation in prostate tissues indicates high**
1139 **stability of proteins.** *Nat Commun* 2019, **10:**2524.
1140 98. Sigurgeirsson B, Emanuelsson O, Lundeberg J: **Sequencing degraded RNA addressed by 3' tag**
1141 **counting.** *PLoS One* 2014, **9:**e91851.
1142 99. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning**
1143 **sequence reads to genomic features.** *Bioinformatics* 2014, **30:**923-930.
1144 100. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects**
1145 **and other unwanted variation in high-throughput experiments.** *Bioinformatics* 2012, **28:**882-
1146 883.
1147 101. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq**
1148 **data with DESeq2.** *Genome Biol* 2014, **15:**550.
1149 102. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ,
1150 Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts**
1151 **and isoform switching during cell differentiation.** *Nat Biotechnol* 2010, **28:**511-515.
1152 103. Shao S, Guo T, Gross V, Lazarev A, Koh CC, Gillessen S, Joerger M, Jochum W, Aebersold R:
1153 **Reproducible Tissue Homogenization and Protein Extraction for Quantitative Proteomics Using**
1154 **MicroPestle-Assisted Pressure-Cycling Technology.** *J Proteome Res* 2016, **15:**1821-1829.

1155 104. Rosenberger G, Koh CC, Guo T, Rost HL, Kouvonen P, Collins BC, Heusel M, Liu Y, Caron E,
1156 Vichalkovski A, et al: **A repository of assays to quantify 10,000 human proteins by SWATH-MS.**
1157 *Sci Data* 2014, **1:**140031.
1158 105. Röst HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, Wolski W, Collins BC,
1159 Malmstrom J, Malmstrom L, Aebersold R: **OpenSWATH enables automated, targeted analysis of**
1160 **data-independent acquisition MS data.** *Nat Biotechnol* 2014, **32:**219-223.
1161 106. Guo T, Luna A, Rajapakse VN, Koh CC, Wu Z, Liu W, Sun Y, Gao H, Menden MP, Xu C, et al:
1162 **Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines.** *iScience* 2019, **21:**664-680.
1163 107. Blattmann P, Heusel M, Aebersold R: **SWATH2stats: An R/Bioconductor Package to Process and**
1164 **Convert Quantitative SWATH-MS Proteomics Data for Downstream Analysis Tools.** *PLoS One*
1165 2016, **11:**e0153160.
1166 108. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using**
1167 **empirical Bayes methods.** *Biostatistics* 2007, **8:**118-127.
1168 109. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P,
1169 von Mering C, Jensen LJ: **STRING v9.1: protein-protein interaction networks, with increased**
1170 **coverage and integration.** *Nucleic Acids Res* 2013, **41:**D808-815.
1171 110. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson
1172 E, et al: **Integrative analysis of complex cancer genomics and clinical profiles using the**
1173 **cBioPortal.** *Sci Signal* 2013, **6:**pl1.
1174 111. Ycart B, Charmpi K, Rousseaux S, Fournié J-J: **Large scale statistical analysis of GEO datasets.**
1175 2014, **4:**113:111-119.

1176

1177

1178 # Figure legends

1179

1180 **Figure 1. Proteogenomics analysis of 105 tissue regions from 39 PCa patients. a** Representative
1181 immunohistochemistry images of prostate tissues and the selection of BPH and tumorous tissue regions
1182 for genome, transcriptome and proteome analysis. **b** Kaplan-Meier curves for our cohort when the
1183 patients are stratified by the overall grade (left), the TA1 or TA grade group (middle) and the TA2 or TA
1184 grade group (right). Point-wise 95% confidence bands are shown for the whole range of time values.

1185

1186 **Figure 2. Molecular perturbation scores for point mutations, CNAs, transcriptome and proteome data.**
1187 **a** Distributions of the first type of molecular perturbation scores (DE_count's) for the four grade groups
1188 (visualized as violin plots) at the mutation layer (upper left), CNA layer (upper right), mRNA layer (lower
1189 left) and protein layer (lower right). Points represent the actual values. The horizontal lines correspond
1190 to the median value in each of the four grade groups. **b** Distributions of the second type of molecular
1191 perturbation scores (DE_sum's) for the four grade groups (visualized as violin plots) at the CNA layer
1192 (upper left), mRNA layer (upper right) and protein layer (lower left). Points represent the actual values.
1193 The horizontal lines correspond to the median value in each of the four grade groups. *P* values (in each
1194 of the titles) show the significance of the one-sided Wilcoxon rank sum test where the values of G3 and
1195 G4/5 are gathered together and compared to the values of G1 and G2 (also gathered together).

1196

1197 **Figure 3**. **Within-group similarity at the different layers quantified by different similarity measures. a,**
1198 **b** Distributions of the similarity scores between the individual tumor samples and the centroid using the
1199 Pearson correlation (**a**) and MI (**b**) for the four grade groups (visualized as violin plots) at the CNA (upper
1200 left), mRNA (upper right) and protein (lower left) layers. Points represent the actual values. The
1201 horizontal lines correspond to the median value in each of the four grade groups. *P* values from the one-
1202 sided Wilcoxon rank sum test comparing G4/5 versus G1: 0.0014 for the CNA, 0.89 for the mRNA and
1203 0.053 for the protein layer with the Pearson correlation, and 0.027 for the CNA, 0.0052 for the mRNA
1204 and 0.0081 for the protein layer with the MI. **c** Density plots of the FCs in the four grade groups for three
1205 selected proteins among the 20 highest scoring (score: mean of the absolute FCs across all tumor
1206 samples) proteins. Vertical lines correspond to the average FC in each of the four grade groups. The
1207 selected proteins have more extreme FCs in the high-grade tumors (G3 and G4/5). **d** Heatmap of the
1208 CNA matrix reduced to the significant regulators of the target gene ACPP output by the fitted elastic net
1209 model (*i.e.* those with a non-zero beta coefficient). The columns are ordered based on the grade group
1210 while there is a hierarchical clustering of the rows. The added colorbar depicts the mRNA FCs of the
1211 target gene ACPP.

1212

1213 **Figure 4**. **Consistently dysregulated sub-networks, Network Component 1 heatmaps and validation in**
1214 **three independent cohorts. a** Sub-networks consistently dysregulated in high-grade compared to low-
1215 grade tumors. There is a significant up-regulation of the FCs of the depicted genes (colored in red) after
1216 network smoothing in the group G4/5 compared to the group G1 in all three layers (CNA, mRNA and
1217 protein). **b** Same as in (**a**) but here there is a significant down-regulation of the FCs of the depicted genes
1218 (colored in blue) after network smoothing in the group G4/5 compared to G1 in all layers. Functional
1219 annotation of the sub-networks in (**a**) and (**b**) with more than one node is given. **c** Heatmap of the CNA
1220 matrix reduced to the Network Component 1 genes. The columns are ordered based on the grade group.

1221 The bottom colorbar depicts the effect size of each sample, *i.e.* its average FC across the genes of the
1222 Network Component 1. The next colorbar represents the negative logarithm in base 10 of the *P* value
1223 from the t-test. The above colorbar shows whether the result is significant: significant results are colored
1224 in black and the rest in white. The top colorbar depicts the grade group of each sample. A black box has
1225 marks the 'interesting' area with amplification of Network Component 1. Gray rectangles at the bottom
1226 show the grade group of the patients (low, intermediate, high) where the samples have (mainly) come
1227 from. **d, e** Same as in (**c**) but for the mRNA and the proteins. The samples removed due to degradation
1228 (mRNA) are shown in gray. **f** Kaplan-Meier curves for 'altered' and 'unaltered' samples in the three
1229 independent studies, TCGA (left), MSKCC (middle) and Aarhus (right) using the corresponding CNA data
1230 when available (first row) and mRNA data (second row).

1231

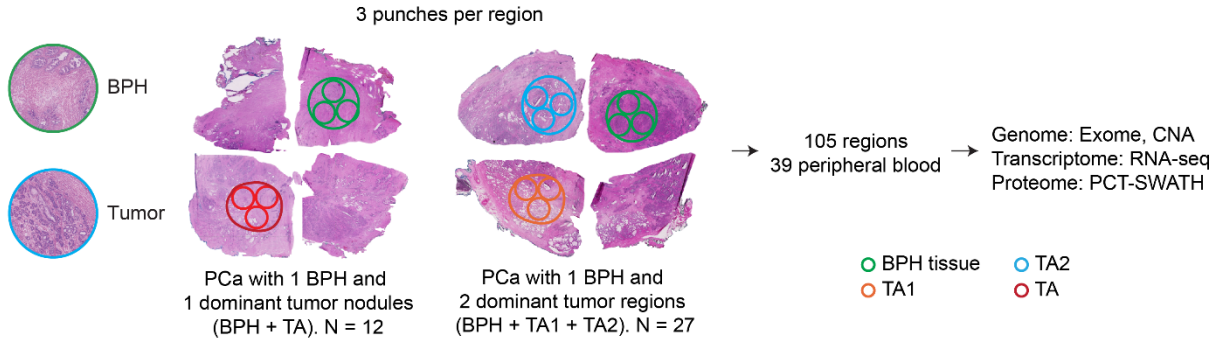
1232 **Figure 5. Within-patient similarity at the different layers. a** Distributions of the within-group similarities
1233 for the four grade groups (visualized as violin plots) based on the Pearson correlation at the CNA layer
1234 (upper left), mRNA layer (upper right) and protein layer (lower left). A 'violin' with the correlations
1235 between TA1 and paired TA2 for the different patients has been added to all three plots and colored in
1236 purple. Points represent the actual values. The horizontal lines correspond to the median value in each
1237 of the groups. This analysis is similar to **Fig. 3a** – but it is not identical. *P* values from the one-sided
1238 Wilcoxon rank sum test comparing the within-patient to the within-group similarities (where all values
1239 from the four groups are gathered together): 8.97e-09 for the CNA, 4.42e-08 for the mRNA and 6.27e-04
1240 for the protein layer. **b** The correlations between TA1 and paired TA2 for the different patients at one
1241 layer are plotted against the corresponding correlations at another layer for each pair of layers: mRNA
1242 versus CNA (upper left), protein versus CNA (upper right) and protein versus mRNA (bottom left). The
1243 points are labeled and colored based on the overall grade in all plots; r: Pearson correlation.

1244

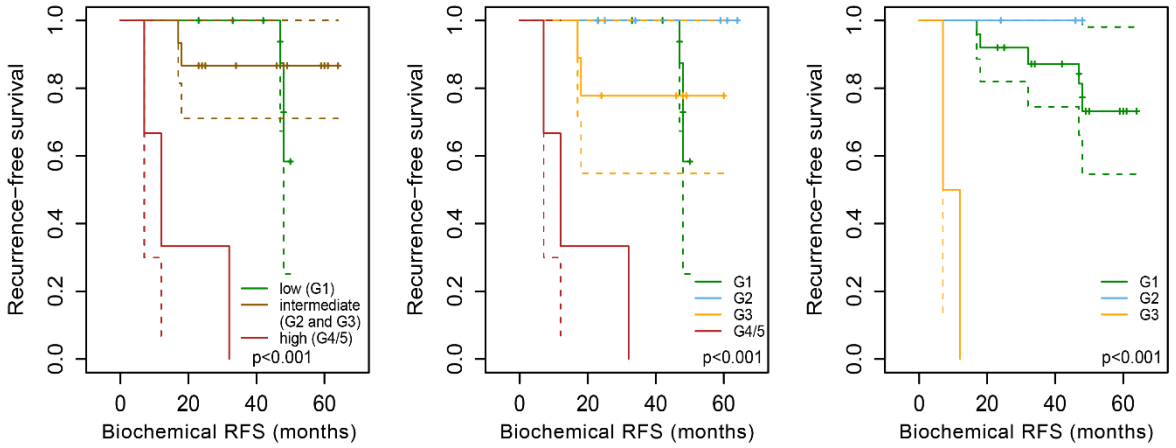
1245

1246 **Figure 1**

a



b

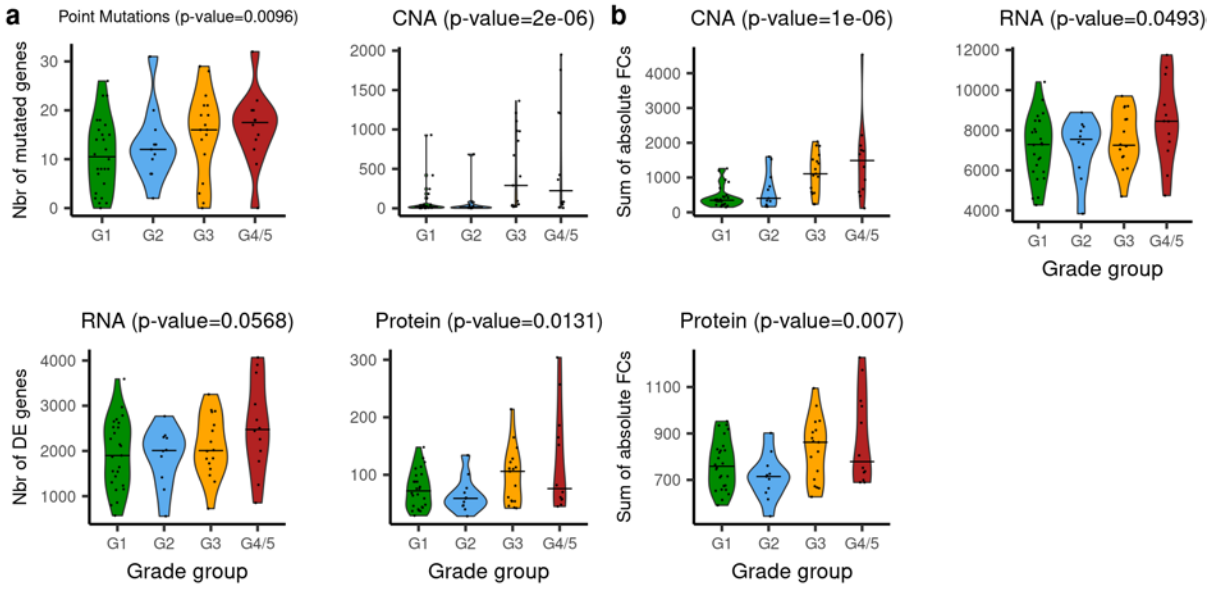


1247

1248

1249 **Figure 2**

1250



1251

1252

Figure 3

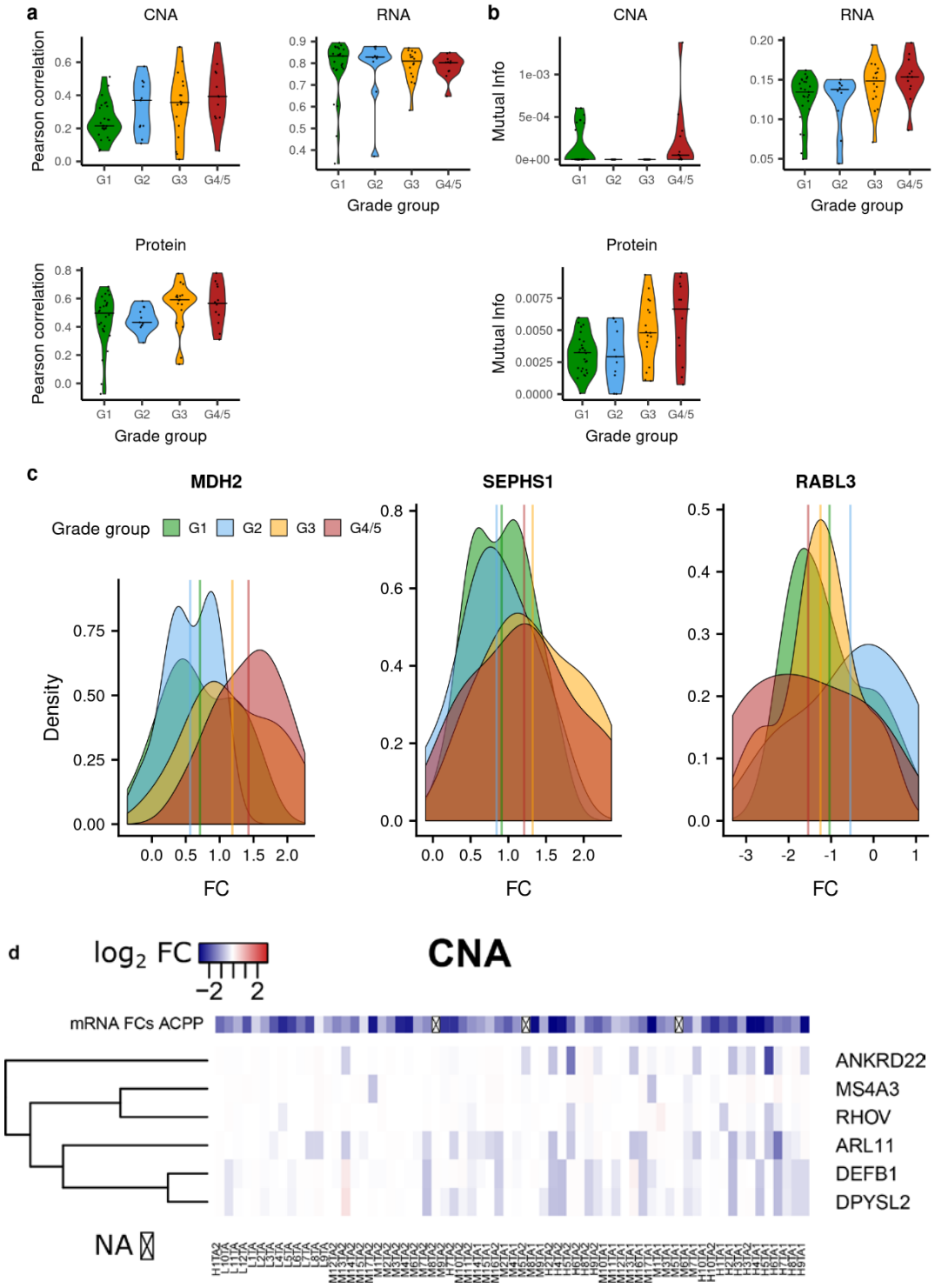
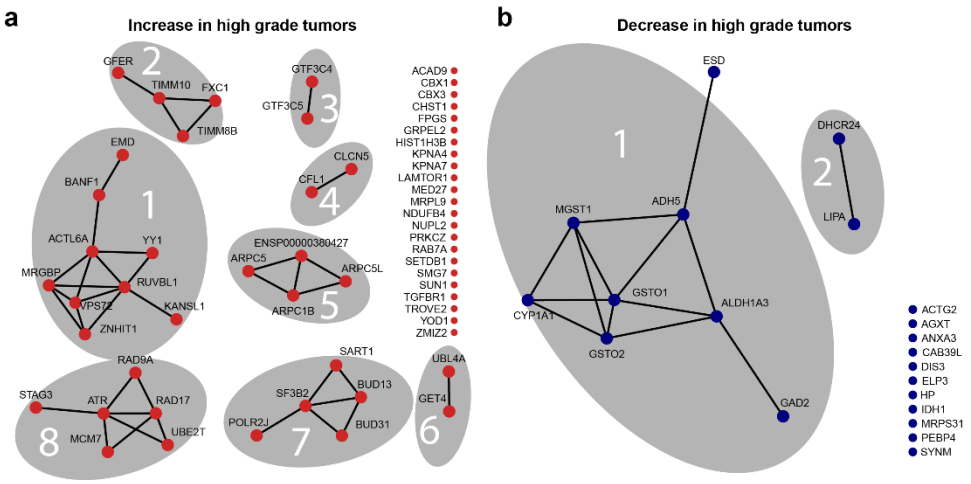

a
CNA
RNA
Pearson correlation
Grade group
G1
G2
G3
G4/5
b
Mutual Info
Protein
c
MDH2
SEPHS1
RABL3
Grade group
Density
FC
d
log2 FC
−2
2
CNA
mRNA FCs ACPP
ANKRD22
MS4A3
RHOV
ARL11
DEFB1
DPYSL2
NA

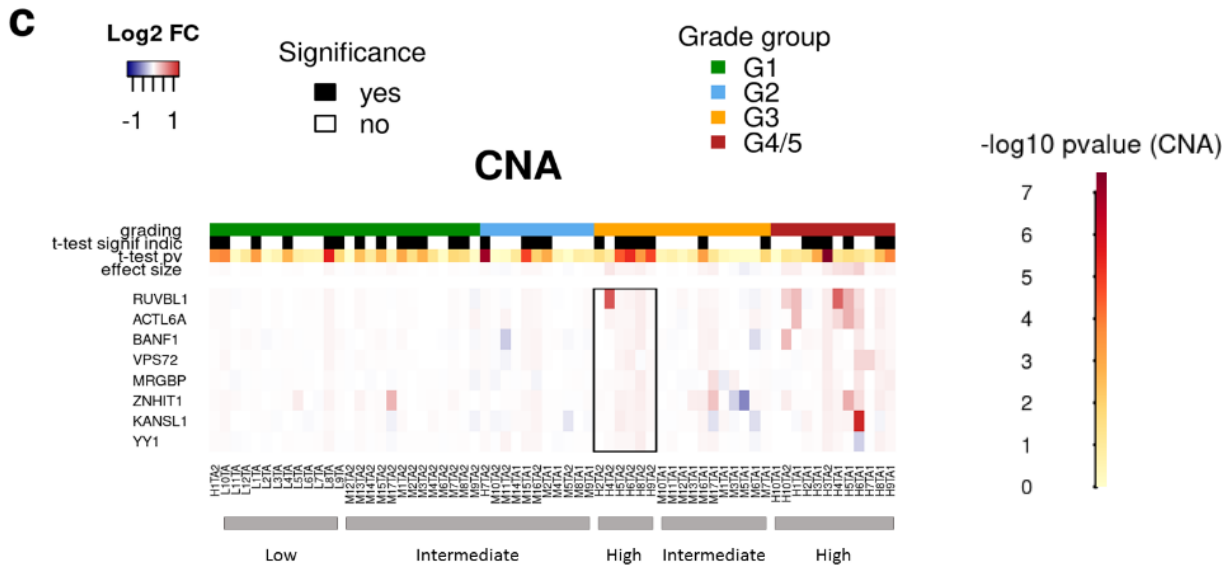
Figure 4



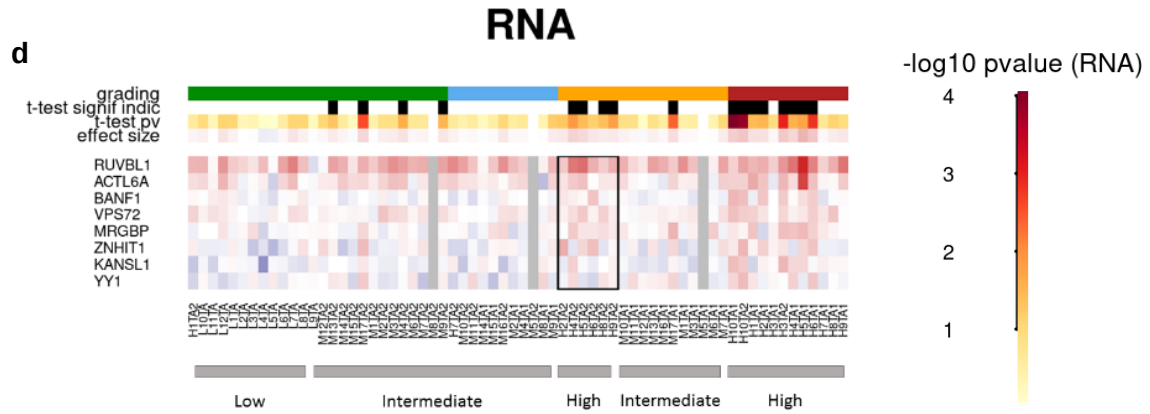
Functional annotation

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

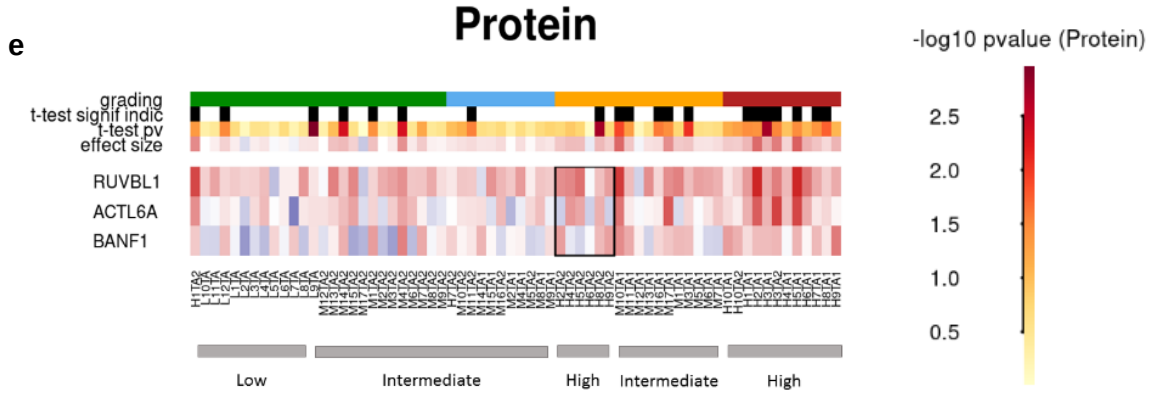
Down-regulation
1. Glutathione metabolism
2. Cholesterol biosynthesis



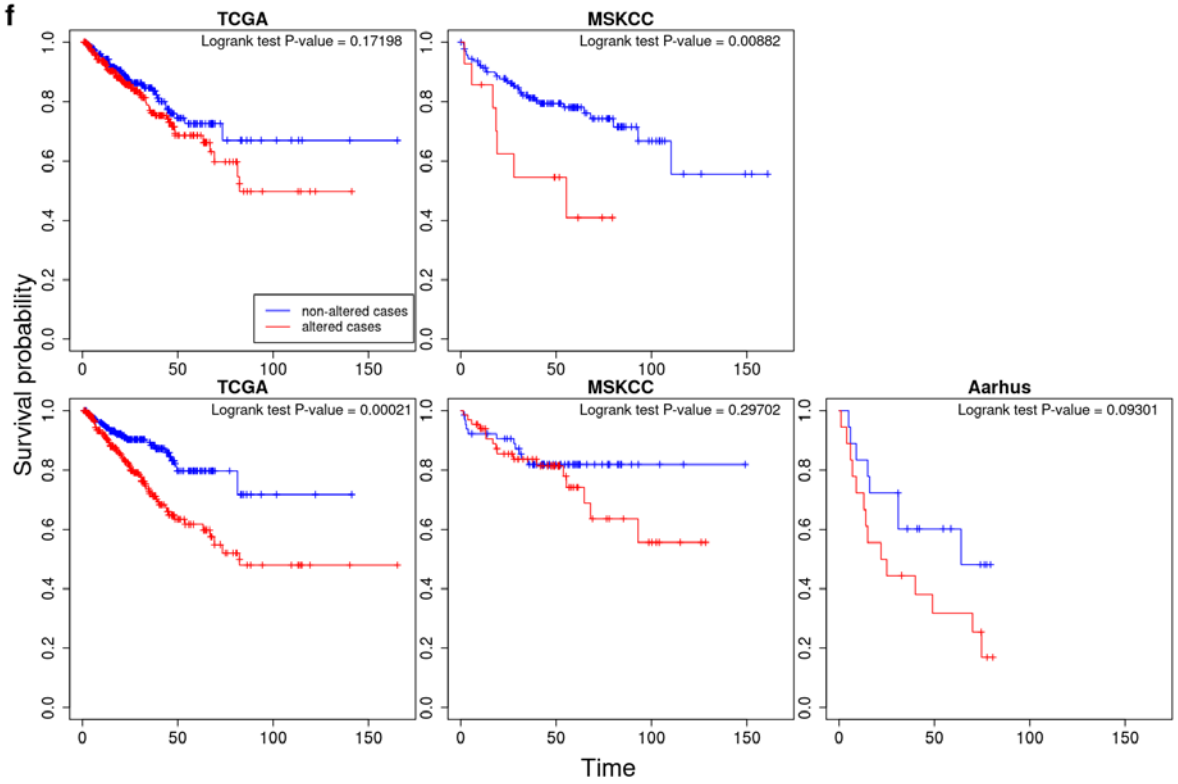
1264 **Figure 4 (*continued*)**



1265



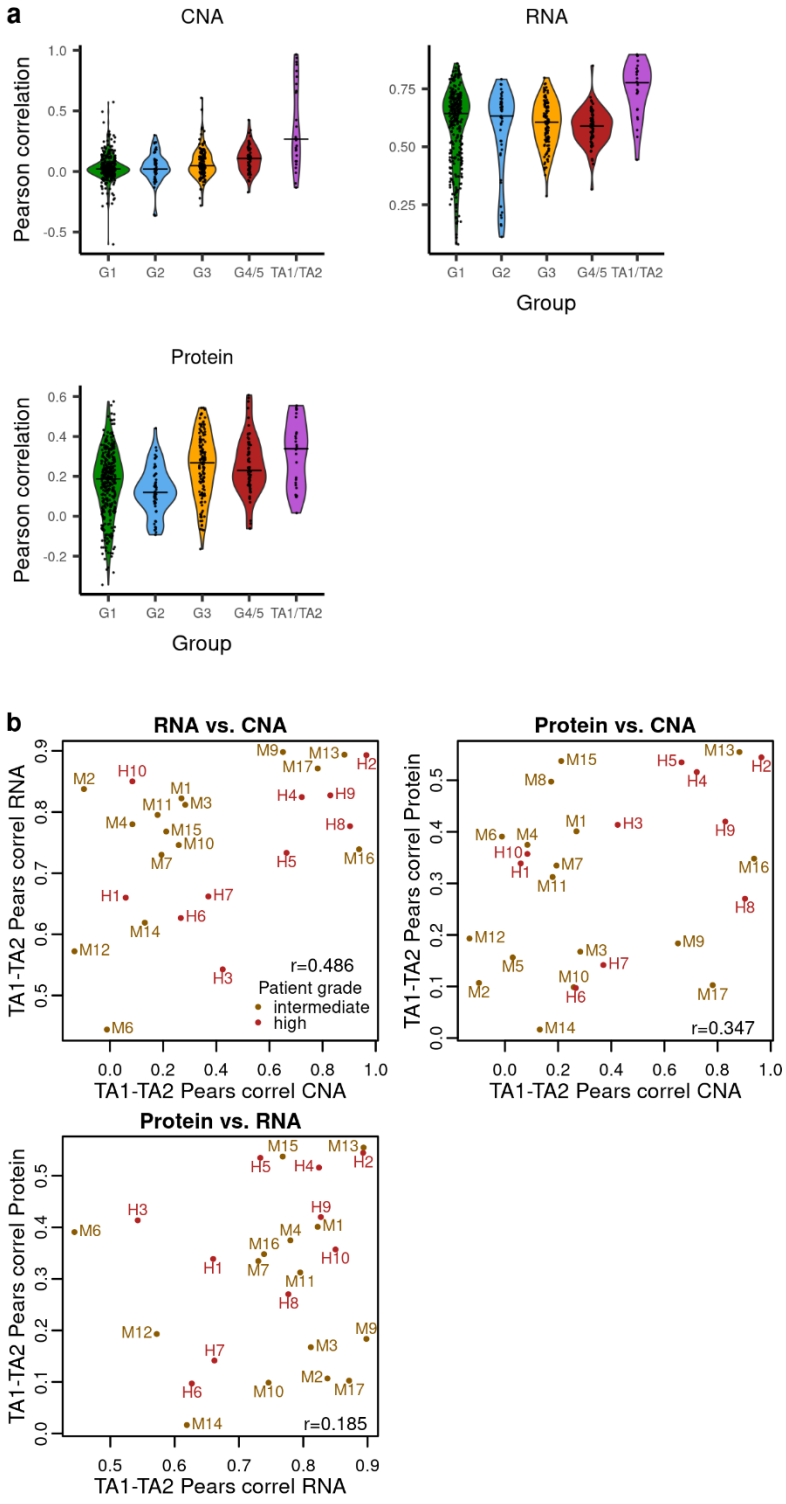
1266



1267

1268 **Figure 5**

1269



1270
1271

1272