Proteogenomic heterogeneity of localized human 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

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

58 ## Introduction

59 Prostate cancer (PCa) represents one of the most common neoplasm among men with almost
60 1,300,000 new cases and 360,000 deaths in 2018 [1] accounting for 15% of all cancers diagnosed. PCa is the
61 fifth leading cause of cancer death in men and represents 6.6% of total cancer mortality in men[1]. Despite
62 earlier detection and new treatments, the lifetime risk to die of PCa has remained stable at approximately
63 3% since 1980. (National Cancer Institute SEER data: https://seer.cancer.gov/statfacts/html/prost.html).
64 In many patients, PCa is indolent and slow growing. The challenge is to identify those patients who are
65 unlikely to experience significant progression while offering radical therapy to those who are at risk.
66 Current risk stratification models are based on clinicopathological variables including histomorphologically
67 defined grade groups, prostate-specific antigen (PSA) levels and clinical stage. Although those variables
68 provide important information for clinical risk assessment and treatment planning[2,3], they do not
69 sufficiently predict the course of the disease.

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

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

88 For example, previous work has shown that copy number changes may alter transcript levels of
89 many genes, whereas the respective protein levels remain relatively stable [21]. Indeed, there is compelling

90 evidence across multiple tumor types that many genomic alterations are 'buffered' at the protein level
91 and are hence mostly clinically inconsequential [22,23]. To better understand the evolution of PCa and to
92 identify core networks perturbed by genomic alterations and thus central for the tumor phenotype, it is
93 therefore essential to investigate the transmission of CNAs to the transcriptomic and proteomic level.
94 To this end, it is important to decipher which genomic alterations impact PCa proteomes, which
95 of those proteomic alterations are functionally relevant, and how molecular networks are perturbed at
96 the protein level across tumors.
97 To address these open questions, we performed a multi-omics profiling of radical prostatectomy
98 (RP) specimens at the level of the genome, transcriptome and proteome from adjacent biopsy-level
99 samples, using state-of-the-art technologies. Unique features of this study are (1) the utilization of PCT
100 (pressure cycling technology)-SWATH (Sequential Window Acquisition of all THeoretical Mass Spectra)
101 mass spectrometry [24,25], allowing rapid and reproducible quantification of thousands of proteins from
102 biopsy-level tissue samples collected in clinical cohorts; (2) the simultaneous profiling of all omics layers
103 from the same tissue regions; (3) inclusion and full profiling of benign regions, which provides a matching
104 control for each tumor; and (4) the full multi-omics characterization of multiple tumor regions from the
105 same patients, thus enabling the detailed investigation of tumor heterogeneity. This design resulted in the
106 multi-layered analyses of 105 samples from 39 PCa patients, as well as of the exome of corresponding
107 peripheral blood cells yielding a comprehensive molecular profile for each patient and identified molecular
108 networks that are commonly altered in multiple patients. Importantly, some of the affected
109 genes/proteins exhibited very small individual effect sizes, suggesting that combined network effects of
110 multiple genes may significantly contribute to determining PCa phenotypes.
111

112 Results

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

114 In this study, we analyzed 39 PCa patients (**Supplementary Fig. 1**) belonging to three groups who
115 underwent laparoscopic robotic-assisted RP. The patients were from the PCa Outcomes Cohort (ProCOC)
116 study [26 27]. Tumor areas were graded using the ISUP (International Society of Urological Pathology) grade
117 groups[28], which range from ISUP grade group G1 (least aggressive) to G5 (most aggressive). The more
118 advanced grade groups G4 and G5 are considered jointly (G4/5). The cohort tested included 12 low-grade
119 (G1), 17 intermediate- (G2 and G3), and 10 high-grade (G4/5) patients (**Fig. 1a**, **Supplementary Fig. 1**,

120 **Supplementary Table 1**). For low-grade PCa patients, we selected two representative regions, one of
121 benign prostatic hyperplasia (BPH) and one of malignant tumor (TA). Since PCa often presents as a
122 multifocal disease with heterogeneous grading within each prostate specimen[25] we analyzed two different
123 tumor regions from the 27 intermediate- and high-grade patients. In those cases three representative
124 regions, including BPH, the most aggressive tumor (TA1) and a secondary, lower-grade tumor (TA2)[2] were
125 analyzed. Thus, TA1 always represented the higher-grade nodule compared to TA2. Note, whereas each
126 patient was assigned a patient-specific overall grade (*i.e.* 'low', 'intermediate' or 'high'), each tumor area
127 was additionally assigned an individual grade group based on its histological appearance. According to
128 current ISUP guidelines, the grading of the entire prostate specimen depends on the size and grade of
129 individual nodules[29]. Thus, it is possible that the patient grading is lower than the grading of the most
130 aggressive nodule, if another lower-grade nodule is larger. Tumor regions contained at least 70% tumor
131 cellularity and the distance between the analyzed areas (TA1 versus TA2) was at least 5 mm. Altogether,
132 we obtained 105 prostate tissue specimens (**Supplementary Table 1**). Three adjacent tissue biopsies of
133 the dimensions 0.6 x 0.6 x 3.0 mm were punched from each representative region for exome sequencing,
134 CNA (derived from the exome sequencing data), RNA sequencing (RNA-seq), and quantitative proteomic
135 analysis using the PCT-SWATH technology [24] respectively. Proteomic analysis was performed in duplicates
136 for each tissue sample. Peripheral blood samples from each patient were also subjected to exome
137 sequencing and served as the genomic wild-type reference (**Fig. 1**). All three types of grading (*i.e.* patient-
138 specific overall grading, TA1 grading and TA2 grading) were predictive of the recurrence-free survival (RFS)
139 in our study.

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

149 ## Molecular perturbations correlate with tumor grade.

150 As a first step towards a cross-layer analysis, we asked if high-grade PCa would be generally
151 affected by stronger alterations (compared to low-grade PCa) at the genome, transcriptome, and
152 proteome layer[30]. Thus, we devised molecular perturbation scores that quantified the number of affected
153 genes/proteins and the extent to which these genes/proteins were altered in the tumor specimens
154 compared to their benign controls (see **Methods** for details). Higher-grade tumors (G3 and G4/5) exhibited
155 significantly higher molecular perturbation scores than lower-grade tumors (G1 and G2). Those differences
156 were statistically significant in all but one case (P value < 0.05, one-sided Wilcoxon rank sum test, **Fig. 2**).
157 The CNA perturbation magnitude exhibited the highest correlation with the PCa grading, confirming prior
158 studies documenting the tight association between CNA, histopathological grade and risk of
159 progression[4,5,31]. Previous work suggested that copy number changes are to some extent buffered at the
160 protein level [32]. Interestingly, we observed that proteins known to be part of protein complexes were
161 significantly less strongly correlated with the fold changes (FCs) of their coding mRNAs than proteins not
162 known to be part of protein complexes (P value < 2.6e-11, one-sided t-test, **Supplementary Fig. 6**). This
163 result is consistent with the concept that protein complex stoichiometry contributes to the buffering of
164 mRNA changes at the level of proteins[22,33-35]. Thus, molecular patterns in high-grade PCa are more strongly
165 perturbed at all layers and the effects of genomic variation are progressively but non-uniformly attenuated
166 along the axis of gene expression.

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

168 Our analysis of CNA profiles (above and **Supplementary Text**) already revealed many shared CNAs
169 across patients, suggesting that such common CNAs might represent genomic driver changes. We
170 therefore investigated if such a convergence towards common molecular endpoints could also be
171 observed at the transcript and protein level. To address this question, we first computed a reference
172 molecular signature that is characteristic of the molecular perturbations of tumors in a given grade group.
173 These 'centroid vectors' were obtained by computing the average tumor-to-benign FCs across all samples
174 within a grade group. Consistent with the observation above, we found that the average effect sizes
175 (averaged absolute centroid FCs) were increasing with the grade group for all three layers (CNA, mRNA,
176 and protein; **Supplementary Table 6**). Next, we compared each individual sample within a group against
177 the matching centroid of the same group. For the quantification of the similarity between a tumor sample
178 and the corresponding centroid we used four similarity/distance measures: Pearson correlation, Mutual
179 Information (MI)[36], Manhattan distance and Euclidean distance. While the first two measures (*i.e.* Pearson

180 correlation and MI) quantify the degree to which the tumor sample and centroid vector co-vary, the other
181 two measures (*i.e.* Manhattan and Euclidean distance) also take into account the magnitude of the FCs in
182 the two vectors. To illustrate this difference, imagine two patients having perturbations of the same
183 genes/proteins, but one of them exhibiting overall two-fold greater FCs (*i.e.* all FCs are increased by a
184 factor of two compared to the other patient). In such a scenario Pearson correlation and MI would yield
185 identical results for the two patients when compared to the centroid, whereas Manhattan and Euclidean
186 distance would identify them as different. Using the Pearson correlation and MI, we found that high-grade
187 PCa (G4/5) were more similar to their respective centroid than low-grade PCa (G1) to their centroid (**Fig.**
188 **3**). This effect was particularly pronounced for protein-level changes. This is consistent with the notion
189 that protein levels (and not mRNA levels) are subjected to stronger selection. Interestingly, when the
190 Euclidean distance and the Manhattan distance were used to characterize tumor similarity, we found that
191 the high-grade tumors were more *dissimilar* to each other than the low-grade tumors (**Supplementary**
192 **Fig. 7**), in sharp contrast to the Pearson correlation and MI. Based on the nature of the different similarity
193 measures tested, we hypothesized that there is a set of proteins commonly affected in their abundance
194 by oncogenic alterations in high-grade tumors. This would increase the similarity using the Pearson
195 correlation or MI. However, although the same proteins are affected, they are affected to a different
196 extent in different high-grade tumors, *i.e.* the FCs exhibited a high degree of variability (**Supplementary**
197 **Fig. 7**), which would increase the dissimilarity based on the Euclidean distance or Manhattan distance (see
198 **Supplementary Fig. 7** for a schematic explanation).

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

211 Gene networks are regulated on CNA, mRNA and protein level.

212 Although the analyses described above suggested convergent proteomic changes, we did not
213 detect a single mutation that was common to all high-grade tumors, let alone common to all 39 PCa cases
214 tested. It has previously been suggested that mutations affecting different genes could impact common
215 molecular networks if the respective gene products interact at the molecular level[43]. However, previous
216 analyses were mostly restricted to individual molecular layers. For example, it was shown that genes
217 mutated in different patients often cluster together in molecular interaction networks [43]. But, effects of
218 these mutations on transcript and protein levels remained unexplored in this case. Here, we aimed at a
219 multi-layer network analysis, involving the genome, transcriptome and proteome. We applied network
220 analysis at each layer separately and the layers were combined afterwards. To identify sub-networks
221 commonly perturbed, we mapped our data onto the STRING gene interaction network [44], and employed
222 network propagation [45,46] separately to the CNA, transcriptome and proteome data for each of the tumor
223 samples. We excluded point mutations from this analysis as their frequency was too low in our cohort. By
224 combining published molecular interactome data with a network propagation algorithm[43,46], we aimed to
225 'enrich' network regions with many perturbed genes/proteins. We reasoned that the convergent
226 consequences of genomic variants on common network regions would be indicative of specific
227 biochemical functions that are important for the tumor biology. We therefore identified genes/proteins
228 in network regions that showed a higher score (or a lower score) in high-grade (G4/5) relative to lower-
229 grade (G1) tumor groups at all three levels (**Fig. 4a**, **b**; **Methods**). This analysis identified sub-networks
230 consisting of over- and under-expressed genes (relative to the benign controls). We found 57 amplified
231 genes (**Supplementary Table 6**) for which transcripts and proteins were often over-expressed in high-
232 grade PCa (**Fig. 4a**) and 21 genes with copy number loss (**Supplementary Table 6**) for which transcripts
233 and proteins were often down-regulated compared to lower-grade tumors (**Fig. 4b**).

234 Among the up-regulated network nodes, we observed genes modulating the stability of
235 chromatin, such as chromatin-binding protein Chromobox 1 (CBX1) [47], SET Domain Bifurcated 1
236 (SETDB1)[48], a function linking to H3K27me3 and H3K9me3 in chromatin, and CBX3 (known as HP1-γ)[49].
237 SETDB1 is an oncogene in melanoma[50] and has also been found to be over-expressed in PCa and cell lines
238 [51]. Further, we found genes involved in DNA damage repair, such as SMG7[52] and ATR[53], and PRKCZ[54], which
239 had already been suggested as a biomarker prognostic for survival in PCa [55]. Multiple actin related proteins
240 including ARPC1B[56], ARPC5[57], ACTL6A[58], and CFL1[59], which are markers for aggressive cancers, were part
241 of the up-regulated network nodes. Moreover, the up-regulated genes contained proteins related to the
242 cell cycle like BANF1 and proteins interacting with the centrosome including LAMTOR1 and RAB7A that

243 had already been associated with PCa[60]. Finally, several signaling molecules with known roles in PCa were
244 up-regulated, such as the transcription factor Yin Yang 1 (YY1) [61], the TGF-β receptor TGFBR1 [62], and KPNA4,
245 which promotes metastasis through activation of NF-κB and Notch signaling [63]. Thus, up-regulated
246 network nodes are involved in DNA/chromatin integrity and growth control.

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

252 Further, our network analysis is suggesting tumor mechanisms converging on genes that are
253 known contributors to PCa tumor biology. For example, the PCa-associated gene SF3B2[70,71] was only
254 weakly amplified in some of the high-grade tumors (average $\log_2$FC = 0.016) and mRNA levels showed
255 similarly small changes (average $\log_2$FC = 0.024). On the other hand, the SF3B2 protein levels were
256 consistently and more strongly up-regulated across tumors (average $\log_2$FC = 0.31), especially within the
257 high-grade tumors (**Supplementary Fig. 8**). Another example is UBE2T whose over-expression is known to
258 be associated with PCa[72]. Unfortunately, we could not quantify the corresponding protein levels. However,
259 we observed a strong and consistent mRNA over-expression across several tumors (average $\log_2$FC = 0.73),
260 even though at the DNA level the gene was only weakly amplified (average $\log_2$FC = 0.023; **Supplementary**
261 **Fig. 8**). Our findings of more heterogeneous CNAs, but more uniform mRNA and protein alterations point
262 on convergent evolutionary mechanisms, as we move along the axis of gene expression.

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

271 Several samples were characterized by a small, but consistent DNA amplification of multiple
272 members of Network Component 1 (**Fig. 4c**). Out of the 66 tumor samples, there were 30 samples –
273 belonging to all grade groups – with a weak but consistent DNA amplification of Network Component 1
274 members, while the high-grade samples had stronger amplifications on average (*i.e.* larger effect sizes).

275 Importantly, gene members of Network Component 1 were dispersed across eight chromosomes
276 (**Supplementary Table 6**). The parallel DNA amplification of these genes is therefore the result of multiple
277 independent CNA events, while the signal on any single gene alone was too weak to be significant in
278 isolation. In some but not all cases, the amplifications led to a small, but consistent increase in mRNA
279 expression of the amplified gene loci (**Fig. 4d**). Unfortunately, only three out of the nine proteins were
280 detected in our proteomics experiments (**Fig. 4e**). Interestingly, patients where the DNA amplifications led
281 to transcript over-expression were almost always high-grade patients, whereas patients where the
282 amplification affected gene expression to a smaller extent were low- or intermediate-grade patients (**Fig.
283 4c**, **d**). Further, we noticed that TA2 samples graded as G3 from high-grade patients carried amplifications
284 of Network Component 1, whereas tumor areas graded as G3 from intermediate-grade patients did not
285 have amplifications of this network component (**Fig. 4c**, **d**). Thus, although the tumor areas were
286 histologically equally classified, tumor areas from high-grade patients carried a CNA signature and
287 expression patterns reminiscent of the high-grade areas from the same patients. Therefore, within the
288 cohort tested the joint DNA amplification of this network component along with RNA up-regulation is a
289 signature of high-grade tumors. Curiously, the higher-grade tumor areas of those high-grades patients
290 (TA1) carried stronger DNA amplifications than the respective lower-grade areas (TA2), which implies that
291 the progressive amplification of Network Component 1 during tumor evolution may contribute to an
292 increasingly aggressive phenotype. In order to further corroborate the clinical relevance of this network
293 perturbation we analyzed published datasets of three additional PCa cohorts (TCGA[8], MSKCC[31], and
294 Aarhus[77]), together comprising a total of 513 patients with known clinical outcome. We found, that over-
295 expression of genes from Network Component 1 was a significant predictor of reduced RFS in the TCGA
296 cohort (*P* value = 8.8e-4, log-rank test), which was the cohort with the largest number of patients. In the
297 other two cohorts we observed the same trend, although the difference in RFS was not statistically
298 significant (*P* value = 0.11 and 0.093 for MSKCC, and Aarhus; **Fig. 4f**).

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

302 Analysis of distinct tumor nodules defines intra-patient heterogeneity (TA1 versus TA2 303 comparison).

304 The CNA patterns (**Supplementary Fig. 4**) and the Network Component 1 analysis (**Fig. 4c**, **d**)
305 suggest that different tumor areas from the same patient shared several mutations. Such common

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

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

325 Comparing molecular similarity across omics layers allowed us to identify specific types of patients.
326 The patients H2, H4, M13 had highly correlated tumor areas at all three layers (upper right corner in all
327 scatterplots of **Fig. 5b**). Likely, the tumor areas of these patients have a common clonal origin
328 (**Supplementary Fig. 3**). In contrast, patients M12 and M14 had weakly correlated tumor areas at all levels
329 (bottom left corner in all scatterplots of **Fig. 5b**). These tumor nodules either have independent clonal
330 origins or they diverged at an earlier stage during tumor evolution (**Supplementary Fig. 3**)[16]. For example,
331 in the case of patient M12 large parts of the genome were not affected by CNAs in the benign sample as
332 well as in TA1 and TA2. However, as shown on **Supplementary Fig. 3**, a large region was amplified in TA1,
333 whereas the same region was deleted in TA2. This is consistent with a scenario in which TA1 and TA2 show
334 parallel evolution. A third class of patients is exemplified by the patients M9 and M17, who showed a high
335 correlation between their tumor areas on the CNA and mRNA levels, but not on the protein level. Yet other
336 patterns were apparent in patients M6 and H3. They showed very similar protein patterns in the two
337 tumor areas, but not on the other two layers. We confirmed that protein-level similarity correlated with

338 similar histological characteristics of the tumor areas. **Supplementary Fig. 9** shows formalin-fixed paraffin-
339 embedded (FFPE) tissue microarray images (duplicates) from the analyzed tumor nodules (TA1 and TA2,
340 diameter 0.6 mm), further underlining the hypothesis that ultimately protein-level alterations are
341 responsible for common cellular phenotypes. Although we cannot fully exclude the possibility that some
342 of these results were affected by technical noise in the data, our findings suggest that transcript alterations
343 can frequently be buffered at the level of proteins (patients M9, M17, **Supplementary Fig. 6**) and that
344 convergent evolutionary processes may lead to the alteration of common proteins (patients M6, H3). We
345 also note that our findings are specific to the two tumor areas available in this study and could be different
346 if other nodules had been sampled for each of the patients. However, our findings on patients with weakly
347 correlated tumor areas at all levels like M12 and M14 suggest that these patients might carry more than
348 one disease[16].

349 Discussion

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

357 In our study, samples were generated from small, less than 1 mm diameter punches in immediate
358 spatial proximity in the tumor and subsequently profiled at all three 'omics layers' (DNA, RNA, proteome).
359 Due to the large spatial heterogeneity of PCa[14,25], this design - which is so far uncommon for studies
360 profiling multiple layers from tumor specimens - was instrumental for increasing the comparability of the
361 various omics layers and thus facilitated the analysis of molecular mechanisms. Our key findings are: (1)
362 we confirmed the importance of CNAs for PCa biology and the alteration of many known PCa-associated
363 genes at the transcript- and protein-level; (2) we revealed a generally elevated molecular alteration of
364 high-grade tumors compared to lower-grade tumors; (3) although our study confirmed large within- and
365 between-patient genomic heterogeneity, (4) we detected molecular networks that were commonly
366 altered at the mRNA and protein-level. The fact that many of those target molecules are known drivers of
367 PCa tumorigenesis, supports the notion that these proteins/transcripts are subject to convergent
368 evolutionary mechanisms.

369 We integrated the three omics layers using a network-based approach as opposed to directly
370 comparing gene perturbations (mutations) to gene products (transcripts and proteins). Using genome data
371 only, it had previously been hypothesized that whereas the identity of specific mutated genes may differ
372 between tumors, those mutations might still affect common molecular networks[43]. In other words, tumor
373 phenotypes are more dependent on which network regions are perturbed rather than which individual
374 genes are mutated. Our study provides experimental evidence that such network effects are indeed
375 propagated to subsequent molecular layers and that this effect propagation may be clinically relevant.
376 Our multi-omics network analysis revealed that high-grade PCa tumors distinguished themselves
377 from low-grade tumors in two aspects. The first is a generally higher heterogeneity and loss of controlled
378 gene regulation, which increased the molecular differences among high-grade tumors. It had previously
379 been shown that gene expression in tumors is often less coordinated than in normal samples[30]. The
380 increasing heterogeneity of protein concentrations suggests that this loss of coordinated expression also
381 affects protein levels. The second aspect is the convergence of molecular alterations towards a consistent
382 set of perturbed specific molecular sub-networks at the genomic, transcriptomic and proteomic layer
383 along the progression from low-grade to high-grade tumors. Thus, although we observed globally a higher
384 degree of variability in gene expression and proteome control among high-grade specimens, a specific
385 subset of the observed alterations appeared crucial for determining the aggressive tumor phenotype.
386 Tumors are under selective pressure acting on the biochemical function of the cells. It is generally believed
387 that proteins are a closer reflection of the functional state of a cell than the mRNA. Here we could show
388 that the fold changes of proteins like RABL3, MFAP4, and SF3B2 were more pronounced and/or uniform
389 across high-grade tumors than either their coding mRNAs or the underlying CNAs.
390 Specifically, our analysis led to the identification of Network Component 1, a sub-network involved
391 in chromatin remodeling and consisting of genes that were weakly amplified in intermediate-grade (G3)
392 tumor specimens. Signals of individual gene members of this component were virtually indistinguishable
393 from noise in our cohort. However, their consistent alterations across the network region, across
394 molecular layers and the fact that the same genes showed enhanced signals in high-grade specimens,
395 rendered this component highly interesting. The fact that expression changes of Network Component 1
396 members were predictive for survival in independent cohorts further supports the potential clinical
397 relevance of this sub-network. Our network-based cross-omics analysis identified nine other network
398 components (**Fig. 4**) successfully capturing several known and potentially new PCa-associated genes.
399 However, neither Network Component 1 nor any of the other network components was uniformly subject
400 to CNAs across all high-grade patients. Instead, we found different network components modified in

401 different patients and these sub-networks were involved in cellular processes as diverse as actin
402 remodeling, DNA damage response, and metabolic functions, all of which are known contributors to PCa
403 biology. This further underlines the large patient-to-patient variability of PCa and it demonstrates the
404 diversity of molecular mechanisms leading to histologically similar phenotypes. Future prediction models
405 of PCa including the ISUP grade groups, PSA levels and clinical stage might be improved by exploiting multi-
406 omics network analyses. Detecting more or less aggressive networks in prostate biopsies would help
407 clinicians to advice either active surveillance or active therapy. However, the development of such multi-
408 dimensional biomarkers would require much larger patient cohorts.

409 Another distinguishing feature of this study was the simultaneous profiling of two different tumor
410 regions in 27 out of the 39 patients. The profiling of multiple tumor regions from the same prostate helped
411 to further highlight the enormous heterogeneity of PCa within and between patients and provided
412 important insights into PCa evolution. The fact that Network Component 1 was more strongly affected in
413 the paired higher-grade nodules of high-grade patients suggests that at least certain sub-networks are
414 subject to an evolutionary process, that progressively 'moves' protein levels towards a more aggressive
415 state. Generally, and at all molecular layers tested, the two paired tumor areas were more similar to each
416 other compared to two samples from the same grade group but different patients, suggesting common
417 evolutionary origins. Although the two tumor areas seemed to mostly originate from the same clone, this
418 was not always the case. In some patients, different nodules exhibited different molecular patterns at all
419 omics layers, suggesting early evolutionary separation. Thus, for the first time, current diagnostic, expert-
420 level consensus guidelines[29] are supported by detailed proteogenomic data. Our findings support earlier
421 claims that clonality itself might be a prognostic marker with implications for future, more tumor-specific
422 treatment when targeted therapies become available also for PCa[16,78].

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

432

433 # Materials and methods

434 ## Patients and samples

435 A total of 39 men with localized PCa who were scheduled for RP were selected from a cohort of
436 1,200 patients within the ProCOC study and processed at the Department of Pathology and Molecular
437 Pathology, University Hospital Zurich, Switzerland [26]. Each of the selected intermediate- and high-grade
438 patients had two different tumor nodules with different ISUP grade groups. H&E (Hematoxylin and Eosin)-
439 stained fresh frozen tissue sections of 105 selected BPH and tumor regions were evaluated by two
440 experienced pathologists (PJW, NJR) to assign malignancy, tumor stage, and Grade Group according to the
441 International Union Against Cancer (UICC) and WHO/ISUP criteria. This study was approved by the
442 Cantonal Ethics Committee of Zurich (KEK-ZH-No. 2008-0040), the associated methods were carried out
443 in accordance with the approved guidelines, and each patient has signed an informed consent form.
444 Patients were followed up on a regular basis (every three months in the first year and at least annually
445 thereafter) or on an individual basis depending on the disease course in the following years. The RFS was
446 calculated with a biochemical recurrence (BCR) defined as a PSA ≥0.1 ng/ml. Patients were censored if lost
447 to follow-up or event-free at their most recent clinic visit. Patients with a postoperative PSA persistence
448 or without distinct follow-up data for the endpoint BCR were excluded from the analysis of BCR.

449 ## Exome sequencing and somatic variant analysis

450 The exome sequencing (exome-seq) was performed using the Agilent Sure Select Exome platform
451 for library construction and Illumina HiSeq 2500 for sequencing read generation. We mapped and
452 processed the reads using a pipeline based on bowtie2 [79] (1.1.1) and the Genome Analysis Tools Kit (GATK)
453 [80] (3.2-2). We detected and reported nonsynonymous variants or variants causing splicing changes using
454 Strelka (1.0.14) and Mutect (1.1.7) combined with post-processing by the CLC Genomics Workbench
455 (8.0.3). In this process, all tissue samples of a patient were compared to the respective blood sample.

456 Trimmomatic [81] (0.36) was used for adaptor clipping and low-quality subsequence trimming of the
457 FASTQ files. Subsequently, single reads were aligned to the hg19 reference genome with bowtie2 with
458 options "--very-sensitive -k 20". We applied samtools [82] (0.1.19) and picard-tools (1.119) to sort the
459 resulting bam files in coordinate order, merge different lanes, filter out all non-primary alignments, and
460 remove PCR duplicates. Quality of the runs was checked using a combination of BEDtools [83] (2.21),
461 samtools, R (3.1) and FastQC (0.11.2).

462 Bam files containing the mapped reads were preprocessed in the following way: indel information
463 was used to realign individual reads using the RealignerTargetCreator and IndelRealigner option of the
464 GATK. Mate-pair information between mates was verified and fixed using Picard tools and single bases
465 were recalibrated using GATK's BaseRecalibrator. After preprocessing, variant calling was carried out by
466 comparing benign or tumor prostate tissue samples with matched blood samples using the programs
467 MuTect [84] and Strelka[85] independently. Somatic variants that were only detected by one of the two
468 programs were filtered out using CLC Genomics Workbench. So were those that had an entry in the dbSNP
469 [86] common database and those that represented synonymous variants without predicted effects on
470 splicing.

471 CNA analysis of exome-seq data

472 The Bam files generated during the process of somatic variant calling were processed with the
473 CopywriteR package (v.2.2.0) for the R software [87]. CopywriteR makes use of so-called "off-target" reads,
474 *i.e.* reads that cover areas outside of the exon amplicons. "Off-target" reads are produced due to
475 inefficient enrichment strategies. In our case on average 28.5% of the total reads were not on target.
476 Briefly, CopywriteR removes low quality and anomalous read pairs, then peaks are called in the respective
477 blood reference, and all reads in this region are discarded. After mapping the reads into bins, those peak
478 regions, in which reads had been removed, were compensated for. Additionally, read counts are corrected
479 based on mappability and GC-content. Finally, a circular binary segmentation is carried out and for each
480 segment the log count ratios between tissue samples and the respective blood sample are reported as
481 copy number gain or loss. The copy number of each gene in each sample was reported based on the log
482 count ratio of the respective segment in which the gene was located. The overall performance of this
483 CNA-calling approach was evaluated by comparing the results of the TA1 (and TA) samples with CNA
484 results obtained by applying the OncoScan Microarray pipeline to FFPE samples from the same tumors
485 (**Supplementary Fig. 10**).

486 OncoScan Microarrays

487 OncoScan copy number assays were carried out and analyzed as described previously [88]. Briefly,
488 DNA was extracted from punches of FFPE cancer tissue blocks. Locus-specific molecular inversion probes
489 were hybridized to complementary DNA and gaps were filled in a nucleotide-specific manner. After
490 amplification and cleavage of the probes, the probes were hybridized to the OncoScan assay arrays.
491 Scanning the fluorescence intensity and subsequent data processing using the Affymetrix® GeneChip®

492 Command Console and BioDiscovery Nexus express resulted in log intensity ratio data (sample versus
493 Affymetrix reference) and virtual segmentation of the genome into areas with copy number gain, loss or
494 stability.

495 RNA Sequencing

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

500 Paired-end reads were mapped to the human reference genome (GRCh37) using the STAR aligner
501 (version 2.4.2a) [89]. Quality control of the resulting bam files using QoRTs [90] and mRIN [91] showed strong
502 RNA degradation[92] in a significant fraction of the samples: mRIN classified 31 samples as highly degraded
503 (**Supplementary Fig. 11**, **Supplementary Table 4**). In order to correct for this 3' bias, 3 tag counting was
504 performed as described by Sigurgeirsson et al [93] using a tag length of 1,000. After 3' bias correction, three
505 samples still showed a clear 3' bias: the two tumor regions (TA1 and TA2) of the patient M5 and TA2 from
506 patient M8 (**Supplementary Fig. 11**). These samples were excluded from subsequent analyses.
507 Additionally, the BPH region of the patient M5 was excluded due to the exclusion of both its tumor regions.

508 FeatureCounts [94] was used to determine read counts for all genes annotated in ENSEMBL v75.
509 Genes for which no read was observed in any of the samples in the original data were excluded from the
510 analysis. Further, after 3 tag counting, all genes with without at least 1 read per million in N of the samples
511 were removed. We chose N to be 10 which corresponds to the size of the smallest grade group (G2). In a
512 last reduction step, all genes with more than one transcript were excluded, yielding a final set of 14,281
513 genes.

514 Read count normalization and differential gene expression analysis was performed using the R
515 packages sva [95] and DESeq2 [96]. All benign tissues were considered biological replicates and differential
516 gene expression for the individual tumor samples was determined against all benign tissues. Gene
517 expression changes with an adjusted *P* value < 0.1 were considered significant.

518 RNA-seq - 3' bias correction

519 The 3 tag counting approach for 3' bias correction was used on the RNA-seq dataset [93]. This
520 approach requires changing of the annotation file in two steps: 1) isoform filtering and 2) transcript length
521 restriction. As proposed in [93] for each gene we determined the highest expressed isoform within a set of

522 high quality samples. As high quality samples we used all samples with an mRIN score greater than or
523 equal to 0.02. This set contains 7 benign and 15 tumor samples. Isoform expression was determined using
524 cufflinks [97]. As transcript length we chose 1,000bp.

525 Gene fusions

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

528 PCT assisted sample preparation for SWATH-MS

529 We first washed each tissue sample to remove O.C.T., followed by PCT-assisted tissue lysis and
530 protein digestion, and SWATH-MS analysis, as described previously[24] Briefly, a series of ethanol solutions
531 were used to wash the tissues each tissue, including 70% ethanol / 30% water (30 s), water (30 s), 70%
532 ethanol / 30% water (5 min, twice), 85% ethanol / 15% water (5 min, twice), and 100% ethanol (5 min,
533 twice). Subsequently, the tissue punches were lysed in PCT-MicroTubes with PCT-MicroPestle [98] with 30
534 µl lysis buffer containing 8 M urea, 0.1 M ammonium bicarbonate, Complete protease inhibitor cocktail
535 (Roche) and PhosSTOP phosphatase inhibitor cocktail (Roche) using a barocycler (model NEP2320-45k,
536 PressureBioSciences, South Easton, MA). The lysis was performed with 60 cycles of high pressure (45,000
537 p.s.i., 50 s per cycle) and ambient pressure (14.7 p.s.i., 10 s per cycle). The extracted proteins were then
538 reduced and alkylated prior to lys-C and trypsin-mediated proteolysis under pressure cycling. Lys-C (Wako;
539 enzyme-to-substrate ratio, 1:40) -mediated proteolysis was performed using 45 cycles of pressure
540 alternation (20,000 p.s.i. for 50 s per cycle and 14.7 p.s.i. for 10 s per cycle), followed by trypsin (Promega;
541 enzyme-to-substrate ratio, 1:20)-mediated proteolysis using the same cycling scheme for 90 cycles. The
542 resultant peptides were cleaned using SEP-PAC C18 (Waters Corp., Milford, MA) and analyzed, after spike-
543 in 10% iRT peptides [51], using SWATH-MS following the 32-fixed-size-window scheme as described
544 previously [19, 21] using a 5600 TripleTOF mass spectrometer (Sciex) and a 1D+ Nano LC system (Eksigent,
545 Dublin, CA). The LC gradient was formulated with buffer A (2% acetonitrile and 0.1% formic acid in HPLC
546 water) and buffer B (2% water and 0.1% formic acid in acetonitrile) through an analytical column (75 µm
547 × 20 cm) and a fused silica PicoTip emitter (New Objective, Woburn, MA, USA) with 3-µm 200 Å Magic C18
548 AQ resin (Michrom BioResources, Auburn, CA, USA). Peptide samples were separated with a linear
549 gradient of 2% to 35% buffer B over 120 min at a flow rate of 0.3 µl min^{-1}. Ion accumulation time for MS1
550 and MS2 was set at 100 ms, leading to a total cycle time of 3.3 s.

551 ## SWATH assay query library for prostate tissue proteome

552 To build a comprehensive library for SWATH data analysis, we analyzed unfractionated prostate
553 tissue digests prepared by the PCT method using Data Dependent Acquisition (DDA) mode in a tripleTOF
554 mass spectrometer over a gradient of 2 hours as described previously [19]. We spiked iRT peptides [51] into
555 each sample to enable retention time calibration among different samples. We then combined these data
556 with the DDA files from the pan-human library project [99]. All together we analyzed 422 DDA files using
557 X!Tandem [52] and OMSSA [53] against three protein sequence databases downloaded on Oct 21, 2016 from
558 UniProt, including the SwissProt database of curated protein sequences (n=20,160), the splicing variant
559 database (n=21,970), and the trembl database (n=135,369). Using each database, we built target-decoy
560 protein sequence database by reversing the target protein sequences. We allowed maximal two missed
561 cleavages for fully tryptic peptides, and 50 p.p.m. for peptide precursor mass error, and 0.1 Da for peptide
562 fragment mass error. Static modification included carbamidomethyl at cysteine, while variable
563 modification included oxidation at methionine. Search results from X!Tandem and OMSSA were further
564 analyzed through Trans-Proteomic Pipeline (TPP, version 4.6.0) [54] using PeptideProphet and iProphet,
565 followed by SWATH assay library building procedures as detailed previously [19, 55]. Altogether, we identified
566 167,402 peptide precursors, from which we selected the proteins detected in prostate tissue samples, and
567 built a sample-specific library. SWATH wiff files were converted into mzXML files using ProteoWizard [56]
568 msconvert v.3.0.3316, and then mzML files using OpenMS [57] tool FileConverter. OpenSWATH[100] was
569 performed using the tool OpenSWATHWorkflow with input files including the mzXML file, the TraML
570 library file, and TraML file for iRT peptides.

571 ## Peptide quantification using OpenSWATH

572 To obtain consistent quantification of the SWATH files, we obtained the all annotated *b* and *y*
573 fragments from the sp, sv and tr libraries. About ten thousand redundant and low-quality assays were
574 removed. Then we extracted the chromatography of these fragments and MS1 signals using
575 OpenSWATHWorkflow, followed by curation using DIA-expert[101]. Briefly, the chromatography of all
576 fragments and MS1 signals were subject to scrutiny by empirically developed expert rules. A reference
577 sample with best q value by pyprophet was picked up to refined fragments. The peptide precursors are
578 further filtered based on the following criteria: i) remove peptide precursors with a q value higher than
579 1.7783e-06 to achieve a false discovery rate of 0.00977 at peptide level using SWATH2stats [102]; ii) peptides
580 with a FC higher than 2 between the reference sample and its technical replicate were removed; iii)
581 peptides matching to multiple SwissProt protein sequences were removed. The data matrix was first

582 quantile normalized, $\log_2$ transformed, followed by batch correction using the ComBat R package [103].
583 Finally, for each protein and pair of technical replicates the average value was computed.

584 Statistical analysis

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

587 Computation of molecular perturbation scores

588 On the genomic level (mutation and CNA), we kept the tumor samples (66 in total) that contain
589 FCs with respect to the blood. The mutation matrix was further discretized by setting all non-zero events
590 to 1. At the transcriptomics level, the FCs for the 63 tumor samples were computed as described above
591 (see 'RNA Sequencing'). Finally, on the proteomics level, we computed the FCs for the tumor samples (66
592 in total) as follows: for each protein, its mean intensity over the normal samples was subtracted from the
593 intensities of the tumor samples. (We chose to compute the FCs for the tumor samples with respect to a
594 global reference (average of all normal samples) and not with respect to their paired benign sample in
595 order to achieve a higher consistency with the transcriptomics level.)
596 We assigned to each sample two molecular perturbation scores summarizing/quantifying the
597 magnitude of its FCs: DE_count counts the number of mutated/differentially expressed (DE) genes, while
598 the DE_sum score is the sum of absolute FCs of all genes. Thus, while the first score counts the number of
599 events (mutations/DE genes), the second one quantifies their magnitude. A gene is regarded as
600 mutated/DE if its value is 1 in the mutation layer and if its absolute value is above a threshold that has
601 been set to 1 for the mRNA and protein layer. For the CNA layer, the corresponding threshold was set to
602 0.5 because the range of FCs in the CNA matrix is smaller than the mRNA and protein matrices. Both types
603 of scores were computed for each molecular level, except for the point mutations where only DE_count
604 was computed. Afterwards, the 66 DE_count scores (63 for the mRNA) and the DE_sum scores at each
605 layer were divided into the four grade groups G1, G2, G3 and G4/5 respectively.

606 Network propagation/smoothing

607 As a network, the STRING gene interaction network (version 10)[104] was used, after removing all
608 edges with combined score smaller or equal to 900 and keeping subsequently the largest connected
609 component. The resulting network consisted of 10,729 nodes and 118,647 (high-confidence) edges. For
610 the network smoothing, the weight matrix was computed as described in Vanunu et al.[105] but for an

611 unweighted graph and the propagation parameter was set to 0.5. The propagation was iteratively
612 repeated 500 times to ensure convergence of the results. For the mapping from gene symbols to STRING
613 identifiers (**Supplementary Table 6**) we used the R/Bioconductor package STRINGdb [106]. The gene symbols
614 with no matching STRING identifier were removed, while for those that mapped to multiple STRING
615 identifiers, the first mapping was kept (default choice in the package). From the multiple gene symbols
616 that mapped to the same STRING identifier, the first mapping was kept. The genes that were not present
617 in the network were removed from the datasets, while those that were present in the network but not in
618 the corresponding dataset were initially filled in with 0's.

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

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

629 Network Component 1 analysis

630 For each tumor sample at the CNA layer, a one-sided, one-sample t-test has been applied testing
631 if its average FC over the genes of the Network Component 1 (and in particular those that have been
632 measured at the CNA) is significantly greater than 0. Due to the presence of outliers in some samples, the
633 non-parametric, one-sided Wilcoxon signed-rank test has been applied as well yielding very similar results
634 (data not shown). A result is considered to be significant if the corresponding *P* value is below 0.05. The
635 analysis has been repeated for the mRNA and protein layer.
636

637 Independent cohorts validation

638 For the validation of Network Component 1, we used published datasets of three PCa cohorts:
639 TCGA, MSKCC, and Aarhus. For TCGA and MSKCC, we downloaded the mRNA data- consisting of 493 and
640 150 samples respectively- with precomputed z-scores per gene, and corresponding clinical data from

641 cBioPortal[107] (https://www.cbioportal.org/). There were 333 samples with available disease/progression-
642 free survival time in TCGA and 140 samples in MSKCC and only these were used in the subsequent survival
643 analysis.

644 For the Aarhus study (NCBI GEO dataset GSE46602), we downloaded the mRNA matrix and
645 corresponding clinical information as described in Ycart et al[108]. The resulting mRNA matrix consisted of
646 20,186 genes and 50 samples- 40 PCa samples with known RFS time and 10 benign samples. Once
647 excluding the benign samples, we computed z-scores per gene in order to have comparable values with
648 the other two studies. These 40 PCa samples were also considered in the subsequent survival analysis.

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

655 ## Mutual information

656 In order to compute the MI between each individual sample within a group against the matching
657 centroid of the same group, the R package infotheo (http://cran.r-
658 project.org/web/packages/infotheo/index.html) has been used. For the MI, the data needs to be discrete.
659 For that, we discretize the tumor-to-benign FCs and centroids by setting 1, if the value is above a threshold
660 *thr*, -1, if the value is below -*thr* and 0 else. (*thr* is equal to 1 for the mRNA and proteins and 0.5 for the
661 CNAs.)

662 # Acknowledgments

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

670 ## Author contributions

671 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.,
672 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.,
673 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
674 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
675 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.,
676 P.J.W. and T.G. supported and supervised the project.

677 ## Competing interests

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

681

682 # References

683

684 1. Ferlay, J.*, et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns
685 in GLOBOCAN 2012. *Int J Cancer* **136**, E359-386 (2015).
686 2. Epstein, J.I.*, et al.* A Contemporary Prostate Cancer Grading System: A Validated Alternative to the
687 Gleason Score. *Eur Urol* **69**, 428-435 (2016).
688 3. Gordetsky, J. & Epstein, J. Grading of prostatic adenocarcinoma: current state and prognostic
689 implications. *Diagn Pathol* **11**, 25 (2016).
690 4. Berger, M.F.*, et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-
691 220 (2011).
692 5. Barbieri, C.E.*, et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in
693 prostate cancer. *Nat Genet* **44**, 685-689 (2012).
694 6. Robinson, D.*, et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228
695 (2015).
696 7. Baca, S.C.*, et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-677 (2013).
697 8. Cancer Genome Atlas Research, N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**,
698 1011-1025 (2015).
699 9. Wedge, D.C.*, et al.* Sequencing of prostate cancers identifies new cancer genes, routes of
700 progression and drug targets. *Nat Genet* **50**, 682-692 (2018).
701 10. Tomlins, S.A.*, et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate
702 cancer. *Science* **310**, 644-648 (2005).
703 11. Grasso, C.S.*, et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature*
704 **487**, 239-243 (2012).
705 12. Boutros, P.C.*, et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer.
706 *Nat Genet* **47**, 736-745 (2015).
707 13. Cooper, C.S.*, et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies
708 multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue.
709 *Nat Genet* **47**, 367-372 (2015).
710 14. Lovf, M.*, et al.* Multifocal Primary Prostate Cancer Exhibits High Degree of Genomic Heterogeneity.
711 *Eur Urol* **75**, 498-505 (2019).
712 15. Lindberg, J.*, et al.* Exome sequencing of prostate cancer supports the hypothesis of independent
713 tumour origins. *Eur Urol* **63**, 347-353 (2013).
714 16. Espiritu, S.M.G.*, et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical
715 Aggression. *Cell* **173**, 1003-1013 e1015 (2018).
716 17. Sinha, A.*, et al.* The Proteogenomic Landscape of Curable Prostate Cancer. *Cancer Cell* **35**, 414-427
717 e416 (2019).
718 18. Latonen, L.*, et al.* Integrative proteomics in prostate cancer uncovers robustness against genomic
719 and transcriptomic aberrations during disease progression. *Nat Commun* **9**, 1176 (2018).
720 19. Iglesias-Gato, D.*, et al.* The Proteome of Primary Prostate Cancer. *Eur Urol* **69**, 942-952 (2016).
721 20. Attard, G.*, et al.* Prostate cancer. *Lancet* **387**, 70-82 (2016).
722 21. Goncalves, E.*, et al.* Widespread Post-transcriptional Attenuation of Genomic Copy-Number
723 Variation in Cancer. *Cell Syst* **5**, 386-398 e384 (2017).
724 22. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA
725 Abundance. *Cell* **165**, 535-550 (2016).
726 23. Liu, Y.*, et al.* Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells.
727 *Nat Commun* **8**, 1212 (2017).

728 24. Guo, T.*, et al.* Rapid mass spectrometric conversion of tissue biopsy samples into permanent
729 quantitative digital proteome maps. *Nat Med* **21**, 407-413 (2015).
730 25. Guo, T.*, et al.* Multi-region proteome analysis quantifies spatial heterogeneity of prostate tissue
731 biomarkers. *Life Sci Alliance* **1**(2018).
732 26. Umbehr, M.*, et al.* ProCOC: the prostate cancer outcomes cohort study. *BMC Urol* **8**, 9 (2008).
733 27. Wettstein, M.S.*, et al.* Prognostic Role of Preoperative Serum Lipid Levels in Patients Undergoing
734 Radical Prostatectomy for Clinically Localized Prostate Cancer. *Prostate* **77**, 549-556 (2017).
735 28. Epstein, J.I.*, et al.* The 2014 International Society of Urological Pathology (ISUP) Consensus
736 Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and
737 Proposal for a New Grading System. *The American journal of surgical pathology* **40**, 244-252
738 (2016).
739 29. Epstein, J.I., Amin, M.B., Reuter, V.E. & Humphrey, P.A. Contemporary Gleason Grading of
740 Prostatic Carcinoma: An Update With Discussion on Practical Issues to Implement the 2014
741 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of
742 Prostatic Carcinoma. *The American journal of surgical pathology* **41**, e1-e7 (2017).
743 30. Reznik, E. & Sander, C. Extensive decoupling of metabolic genes in cancer. *PLoS Comput Biol* **11**,
744 e1004176 (2015).
745 31. Taylor, B.S.*, et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22
746 (2010).
747 32. Zhang, B.*, et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**,
748 382-387 (2014).
749 33. Geiger, T., Cox, J. & Mann, M. Proteomic changes resulting from gene copy number variations in
750 cancer cells. *PLoS Genet* **6**, e1001090 (2010).
751 34. Juschke, C.*, et al.* Transcriptome and proteome quantification of a tumor model provides novel
752 insights into post-transcriptional gene regulation. *Genome Biol* **14**, r133 (2013).
753 35. Stingele, S.*, et al.* Global analysis of genome, transcriptome and proteome reveals the response
754 to aneuploidy in human cells. *Mol Syst Biol* **8**, 608 (2012).
755 36. Cover, T.M. & Thomas, J.A. *Elements of information theory*, (John Wiley & Sons, 2012).
756 37. Zhang, J.S., Gong, A., Cheville, J.C., Smith, D.I. & Young, C.Y. AGR2, an androgen-inducible secretory
757 protein overexpressed in prostate cancer. *Genes Chromosomes Cancer* **43**, 249-259 (2005).
758 38. Liu, Q.*, et al.* Malate dehydrogenase 2 confers docetaxel resistance via regulations of JNK signaling
759 and oxidative metabolism. *Prostate* **73**, 1028-1037 (2013).
760 39. Yang, J.*, et al.* Integrated analysis of microfibrillar-associated proteins reveals MFAP4 as a novel
761 biomarker in human cancers. *Epigenomics* **11**, 1635-1651 (2019).
762 40. Senga, S., Kobayashi, N., Kawaguchi, K., Ando, A. & Fujii, H. Fatty acid-binding protein 5 (FABP5)
763 promotes lipolysis of lipid droplets, de novo fatty acid (FA) synthesis and activation of nuclear
764 factor-kappa B (NF-kappaB) signaling in cancer cells. *Biochim Biophys Acta Mol Cell Biol Lipids*
765 **1863**, 1057-1067 (2018).
766 41. Pan, Y.*, et al.* The overexpression of Rabl3 is associated with pathogenesis and clinicopathologic
767 variables in hepatocellular carcinoma. *Tumour Biol* **39**, 1010428317696230 (2017).
768 42. Zhang, W., Sun, J. & Luo, J. High Expression of Rab-like 3 (Rabl3) is Associated with Poor Survival
769 of Patients with Non-Small Cell Lung Cancer via Repression of MAPK8/9/10-Mediated Autophagy.
770 *Med Sci Monit* **22**, 1582-1588 (2016).
771 43. Hofree, M., Shen, J.P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor
772 mutations. *Nat Methods* **10**, 1108-1115 (2013).
773 44. Szklarczyk, D.*, et al.* STRING v10: protein-protein interaction networks, integrated over the tree of
774 life. *Nucleic Acids Res* **43**, D447-452 (2015).

775 45. Zhou, D., Bousquet, O., Lal, TN, Weston, J., Scholkopf B. Learning with local and global consistency.
776 *Advances in Neural Information Processing Systems* **16**, 312-328 (2004).
777 46. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating genes and protein
778 complexes with disease via network propagation. *PLoS Comput Biol* **6**, e1000641 (2010).
779 47. Baude, A.*, et al.* Hepatoma-derived growth factor-related protein 2 promotes DNA repair by
780 homologous recombination. *Nucleic Acids Research* **44**, 2214-2226 (2015).
781 48. Guler, G.D.*, et al.* Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell
782 Subpopulations from Lethal Drug Exposure. *Cancer Cell* **32**, 221 (2017).
783 49. Slezak, J., Truong, M., Huang, W. & Jarrard, D. HP1gamma expression is elevated in prostate cancer
784 and is superior to Gleason score as a predictor of biochemical recurrence after radical
785 prostatectomy. *BMC Cancer* **13**, 148 (2013).
786 50. Ceol, C.J.*, et al.* The SETDB1 histone methyltransferase is recurrently amplified in and accelerates
787 melanoma. *Nature* **471**, 513-517 (2011).
788 51. Sun, Y.*, et al.* Histone methyltransferase SETDB1 is required for prostate cancer cell proliferation,
789 migration and invasion. *Asian J Androl* **16**, 319-324 (2014).
790 52. Luo, H., Cowen, L., Yu, G., Jiang, W. & Tang, Y. SMG7 is a critical regulator of p53 stability and
791 function in DNA damage stress response. *Cell Discovery* **2**, 15042 (2016).
792 53. Karanika, S.*, et al.* Targeting DNA Damage Response in Prostate Cancer by Inhibiting Androgen
793 Receptor-CDC6-ATR-Chk1 Signaling. *Cell Reports* **18**, 1970 (2017).
794 54. Diouf, B.*, et al.* Somatic deletions of genes regulating MSH2 protein stability cause DNA mismatch
795 repair deficiency and drug resistance in human leukemia cells. *Nature Medicine* **17**, 1298-1303
796 (2011).
797 55. Yao, S.*, et al.* PRKC-ζ Expression Promotes the Aggressive Phenotype of Human Prostate Cancer
798 Cells and Is a Novel Target for Therapeutic Intervention. *Genes Cancer* **1**, 444-464 (2010).
799 56. Zhou, W.*, et al.* Low-frequency nonsynonymous variants in FKBPL and ARPC1B genes are
800 associated with breast cancer risk in Chinese women. *Mol Carcinog* **56**, 774-780 (2017).
801 57. De Robertis, M.*, et al.* EphB2 stem-related and EphA2 progression-related miRNA-based networks
802 in progressive stages of CRC evolution: clinical significance and potential miRNA drivers. *Mol*
803 *Cancer* **17**, 169 (2018).
804 58. Zeng, Z., Yang, H. & Xiao, S. ACTL6A expression promotes invasion, metastasis and epithelial
805 mesenchymal transition of colon cancer. *BMC Cancer* **18**, 1020 (2018).
806 59. Castro, M.A.*, et al.* CFL1 expression levels as a prognostic and drug resistance marker in nonsmall
807 cell lung cancer. *Cancer* **116**, 3645-3655 (2010).
808 60. Overbye, A.*, et al.* Identification of prostate cancer biomarkers in urinary exosomes. *Oncotarget*
809 **6**, 30357-30376 (2015).
810 61. Deng, Z.*, et al.* Yin Yang 1 regulates the transcriptional activity of androgen receptor. *Oncogene*
811 **28**, 3746-3757 (2009).
812 62. Fournier, P.G.*, et al.* The TGF-beta Signaling Regulator PMEPA1 Suppresses Prostate Cancer
813 Metastases to Bone. *Cancer Cell* **27**, 809-821 (2015).
814 63. Yang, J.*, et al.* Inhibition of KPNA4 attenuates prostate cancer metastasis. *Oncogene* **36**, 2868-
815 2878 (2017).
816 64. Liu, W.*, et al.* Homozygous deletions and recurrent amplifications implicate new genes involved in
817 prostate cancer. *Neoplasia* **10**, 897-907 (2008).
818 65. Trasino, S.E., Harrison, E.H. & Wang, T.T. Androgen regulation of aldehyde dehydrogenase 1A3
819 (ALDH1A3) in the androgen-responsive human prostate cancer cell line LNCaP. *Exp Biol Med*
820 *(Maywood)* **232**, 762-771 (2007).
821 66. Bonaccorsi, L.*, et al.* Androgen receptor regulation of the seladin-1/DHCR24 gene: altered
822 expression in prostate cancer. *Lab Invest* **88**, 1049-1056 (2008).

823 67. Li, H., Xiao, D., Hu, L. & He, T. Association of CYP1A1 polymorphisms with prostate cancer risk: an
824 updated meta-analysis. *Mol Biol Rep* **39**, 10273-10284 (2012).
825 68. Aktas, D., Hascicek, M., Sozen, S., Ozen, H. & Tuncbilek, E. CYP1A1 and GSTM1 polymorphic
826 genotypes in patients with prostate cancer in a Turkish population. *Cancer Genet Cytogenet* **154**,
827 81-85 (2004).
828 69. Acevedo, C*., et al.* Positive correlation between single or combined genotypes of CYP1A1 and
829 GSTM1 in relation to prostate cancer in Chilean people. *Prostate* **57**, 111-117 (2003).
830 70. Cao, H.M., Wan, Z., Wu, Y., Wang, H.Y. & Guan, C. Development and internal validation of a novel
831 model and markers to identify the candidates for lymph node metastasis in patients with prostate
832 cancer. *Medicine (Baltimore)* **98**, e16534 (2019).
833 71. Kawamura, N*., et al.* SF3B2-Mediated RNA Splicing Drives Human Prostate Cancer Progression.
834 *Cancer Res* **79**, 5204-5217 (2019).
835 72. Wen, M., Kwon, Y., Wang, Y., Mao, J.H. & Wei, G. Elevated expression of UBE2T exhibits oncogenic
836 properties in human prostate cancer. *Oncotarget* **6**, 25226-25239 (2015).
837 73. Kleppe, A*., et al.* Chromatin organisation and cancer prognosis: a pan-cancer study. *Lancet Oncol*
838 **19**, 356-369 (2018).
839 74. Lee, R.S. & Roberts, C.W. Linking the SWI/SNF complex to prostate cancer. *Nat Genet* **45**, 1268-
840 1269 (2013).
841 75. Chuu, C.-P., Lin, C.-Y., Huang, S.-H. & Tsai, K.K.-C. MO2-10-6ACTL6A is a novel oncogene and
842 prognostic biomarker for prostate cancer. *Annals of Oncology* **30**(2019).
843 76. Cai, Y*., et al.* The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone
844 acetyltransferase and SRCAP complexes. *J Biol Chem* **280**, 13665-13670 (2005).
845 77. Mortensen, M.M*., et al.* Expression profiling of prostate cancer tissue delineates genes associated
846 with recurrence after prostatectomy. *Sci Rep* **5**, 16018 (2015).
847 78. Hussain, M*., et al.* LBA12_PRPROfound: Phase III study of olaparib versus enzalutamide or
848 abiraterone for metastatic castration-resistant prostate cancer (mCRPC) with homologous
849 recombination repair (HRR) gene alterations. *Annals of Oncology* **30**(2019).
850 79. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359
851 (2012).
852 80. McKenna, A*., et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
853 generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
854 81. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
855 *Bioinformatics* **30**, 2114-2120 (2014).
856 82. Li, H*., et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079
857 (2009).
858 83. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features.
859 *Bioinformatics* **26**, 841-842 (2010).
860 84. Cibulskis, K*., et al.* Sensitive detection of somatic point mutations in impure and heterogeneous
861 cancer samples. *Nat Biotechnol* **31**, 213-219 (2013).
862 85. Saunders, C.T*., et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal
863 sample pairs. *Bioinformatics* **28**, 1811-1817 (2012).
864 86. Sherry, S.T*., et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311
865 (2001).
866 87. Kuilman, T*., et al.* CopywriteR: DNA copy number detection from off-target sequence data.
867 *Genome Biol* **16**, 49 (2015).
868 88. Noske, A*., et al.* Detection of CCNE1/URI (19q12) amplification by in situ hybridisation is common
869 in high grade and type II endometrial cancer. *Oncotarget* **8**, 14794-14805 (2017).
870 89. Dobin, A*., et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

871 90. Hartley, S.W. & Mullikin, J.C. QoRTs: a comprehensive toolset for quality control and data
872 processing of RNA-Seq experiments. *BMC Bioinformatics* **16**, 224 (2015).
873 91. Feng, H., Zhang, X. & Zhang, C. mRIN for direct assessment of genome-wide and gene-specific
874 mRNA integrity from large-scale RNA-sequencing data. *Nat Commun* **6**, 7816 (2015).
875 92. Shao, W*., et al.* Comparative analysis of mRNA and protein degradation in prostate tissues
876 indicates high stability of proteins. *Nat Commun* **10**, 2524 (2019).
877 93. Sigurgeirsson, B., Emanuelsson, O. & Lundeberg, J. Sequencing degraded RNA addressed by 3' tag
878 counting. *PLoS One* **9**, e91851 (2014).
879 94. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning
880 sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
881 95. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch
882 effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883
883 (2012).
884 96. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
885 seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
886 97. Trapnell, C*., et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated
887 transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515 (2010).
888 98. Shao, S*., et al.* Reproducible Tissue Homogenization and Protein Extraction for Quantitative
889 Proteomics Using MicroPestle-Assisted Pressure-Cycling Technology. *J Proteome Res* **15**, 1821-
890 1829 (2016).
891 99. Rosenberger, G*., et al.* A repository of assays to quantify 10,000 human proteins by SWATH-MS.
892 *Sci Data* **1**, 140031 (2014).
893 100. Röst, H.L*., et al.* OpenSWATH enables automated, targeted analysis of data-independent
894 acquisition MS data. *Nat Biotechnol* **32**, 219-223 (2014).
895 101. Guo, T*., et al.* Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines. *iScience* **21**, 664-
896 680 (2019).
897 102. Blattmann, P., Heusel, M. & Aebersold, R. SWATH2stats: An R/Bioconductor Package to Process
898 and Convert Quantitative SWATH-MS Proteomics Data for Downstream Analysis Tools. *PloS one*
899 **11**, e0153160 (2016).
900 103. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using
901 empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
902 104. Szklarczyk, D*., et al.* STRING v10: protein-protein interaction networks, integrated over the tree of
903 life. *Nucleic Acids Research* **43**, D447-D452 (2015).
904 105. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating Genes and Protein
905 Complexes with Disease via Network Propagation. *Plos Computational Biology* **6**(2010).
906 106. Franceschini, A*., et al.* STRING v9.1: protein-protein interaction networks, with increased coverage
907 and integration. *Nucleic Acids Res* **41**, D808-815 (2013).
908 107. Gao, J*., et al.* Integrative analysis of complex cancer genomics and clinical profiles using the
909 cBioPortal. *Sci Signal* **6**, pl1 (2013).
910 108. Ycart, B., Charmpi, K., Rousseaux, S. & Fournié, J.-J. Large scale statistical analysis of GEO datasets.
911 **4**, 113:111-119 (2014).

912

913

914 # Figure legends

915

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

921

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

932

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

944

945 **Figure 4**. **Consistently dysregulated sub-networks, Network Component 1 heatmaps and validation in**
946 **three independent cohorts. a,** Sub-networks consistently dysregulated in the high-grade tumors
947 compared to the low-grade tumors. There is a significant up-regulation of the FCs of the depicted genes
948 (colored in red) after network smoothing in the group G4/5 compared to the group G1 in all three layers
949 (CNA, mRNA and protein). **b,** Sub-networks consistently dysregulated in the high-grade tumors compared
950 to the low-grade tumors. There is a significant down-regulation of the FCs of the depicted genes (colored
951 in blue) after network smoothing in the group G4/5 compared to the group G1 in all three layers (CNA,
952 mRNA and protein). Functional annotation of the sub-networks in (**a**) and (**b**) with more than one node is
953 given below. **c,** Heatmap of the CNA matrix reduced to the Network Component 1 genes. The
954 columns/samples are ordered based on the grade group. The bottom colorbar depicts the effect size of
955 each sample, *i.e.* its average FC across the genes of the Network Component 1. The next colorbar
956 represents the negative logarithm in base 10 of the *P* value from the t-test. The above colorbar shows
957 whether the result is significant: significant results are colored in black and the rest in white. The top

958 colorbar depicts the grade group of each sample. A black box has been drawn around the 'interesting' area
959 with amplification of Network Component 1. Gray rectangles at the bottom show the grade group of the
960 patients (low, intermediate, high) where the samples have (mainly) come from. **d,** Same as in (**c**) but for
961 the mRNA layer. The samples/columns that have been removed due to degradation are shown in gray. **e,**
962 Same as in (**c**) but for the protein layer. **f,** Kaplan-Meier curves for 'altered' and 'unaltered' samples in the
963 three independent studies, TCGA (left), MSKCC (middle) and Aarhus (right). A sample is considered to be
964 'altered' if its average z-score across the nine genes of Network Component 1 is not below the median
965 value (*i.e.* median average z-score) in the respective study.

966

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

979

980 **Supplementary Table 1. Clinicopathological, immunological and other molecular information of the 39**
981 **PCa patients.** (a) Overall clinicopathological characteristics. (b) Detailed information for each patient. Pat:
982 numeric patient ID; Pat_id: patient ID grouped by the overall grade. L: low grade; M: intermediate grade;
983 H: high grade; Overall_Gleason_GrGp: overall ISUP grade group; pT: tumor stage; pN: nodal status; R:
984 surgical margin status; Age_at_OP: age at operation; PSA_at_Diag: blood PSA level at diagnosis; Time
985 (months): RFS time. A value of 0 corresponds to patients excluded for the reasons explained in **Methods**
986 (see 'Patients and samples'); Status: status indicator. 1 means recurrence; DX name: tissue region name;
987 ImageName: name of the scanned images; index_tumor_id: patient ID of TA1 (or TA); TA1_GrGp: grade
988 group for TA1; T_GrGp: grade group for TA2.

989

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

996

997 **Supplementary Table 3. Copy number analysis of 105 PCa samples.** (a) Log_2 ratios indicating the CNA
998 status are shown for all genes in all samples. Values were determined by overlapping gene locations with
999 CNA segments as calculated by CopywriteR. In case more than one segment overlapped with a gene,
1000 number was chosen that had the highest absolute value. (b) Genes are shown with log_2 ratios higher than
1001 0.5 or lower than -0.5 in at least one sample.

1002

1003 **Supplementary Table 4. RNA-seq analysis.** (a) Log_2FCs (relative to all benign samples) for all genes across
1004 the tumor samples. (b) mRIN score per sample generated using mRIN (v1.2.0). (c) ETS family gene fusions
1005 observed in tumor samples using FusionCatcher: a value of 1 means that the fusion was observed in the
1006 respective sample but not its corresponding benign sample, otherwise the value is 0.

1007

1008 **Supplementary Table 5. Proteomics data of 210 PCa samples with duplicates.** (a) Sample information
1009 includes patient ID, clinical diagnosis, sample ID and batch design. (b) Protein matrix of log_2 scaled intensity
1010 of 2,371 proteins quantified in 210 PCa samples.

1011

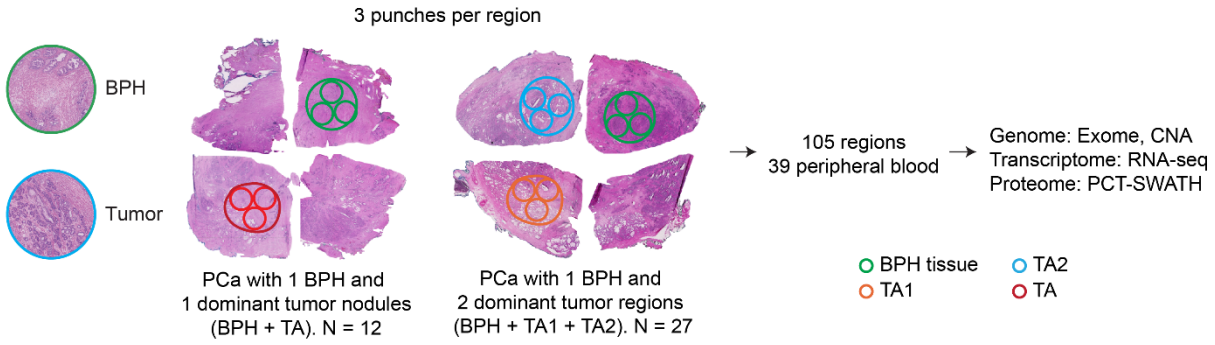
1012 **Supplementary Table 6. Integration analysis of 66 tumor samples.** (a) L1-norm of the 'centroid vectors'
1013 in the three layers (CNA, mRNA and protein) across the four grade groups. (b) Information (*i.e.* reference
1014 linking them to PCa, consistency between observed and reported effect and number of tumor samples
1015 with CNAs) for the 10 highest-scoring proteins (those with largest average absolute FCs across all tumor
1016 specimens). (c) Consistently up-regulated genes in the high-grade tumors: for each of these genes, there
1017 is a significant up-regulation of its FCs after network smoothing in the group G4/5 compared to the group
1018 G1 in all three layers (CNA, mRNA and protein). (d) Consistently down-regulated genes in the high-grade
1019 tumors: for each of these genes, there is a significant down-regulation of its FCs after network smoothing
1020 in the group G4/5 compared to the group G1 in all three layers (CNA, mRNA and protein). (e) Chromosome
1021 information for the gene members of Network Component 1. (f) Mapping from gene symbols to STRING
1022 identifiers.

1023

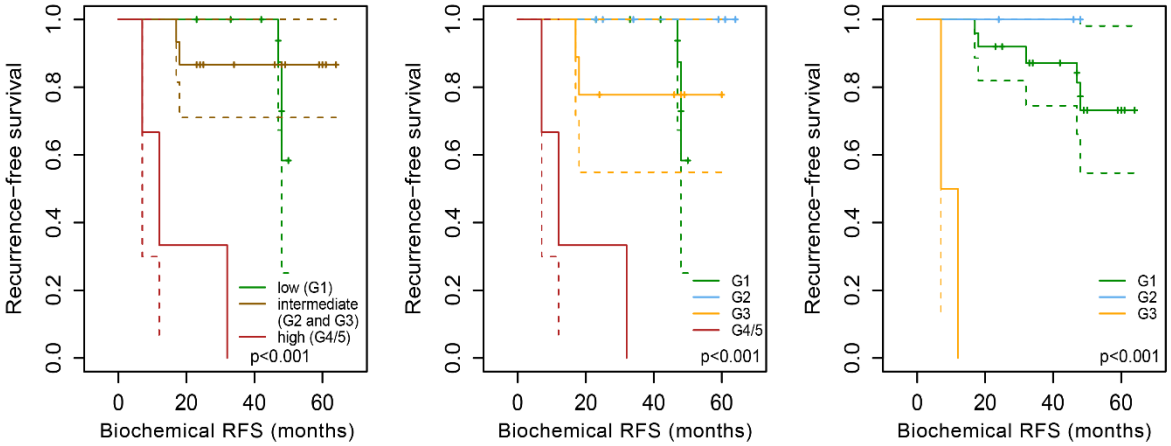
1024

1025 **Figure 1**

a



b

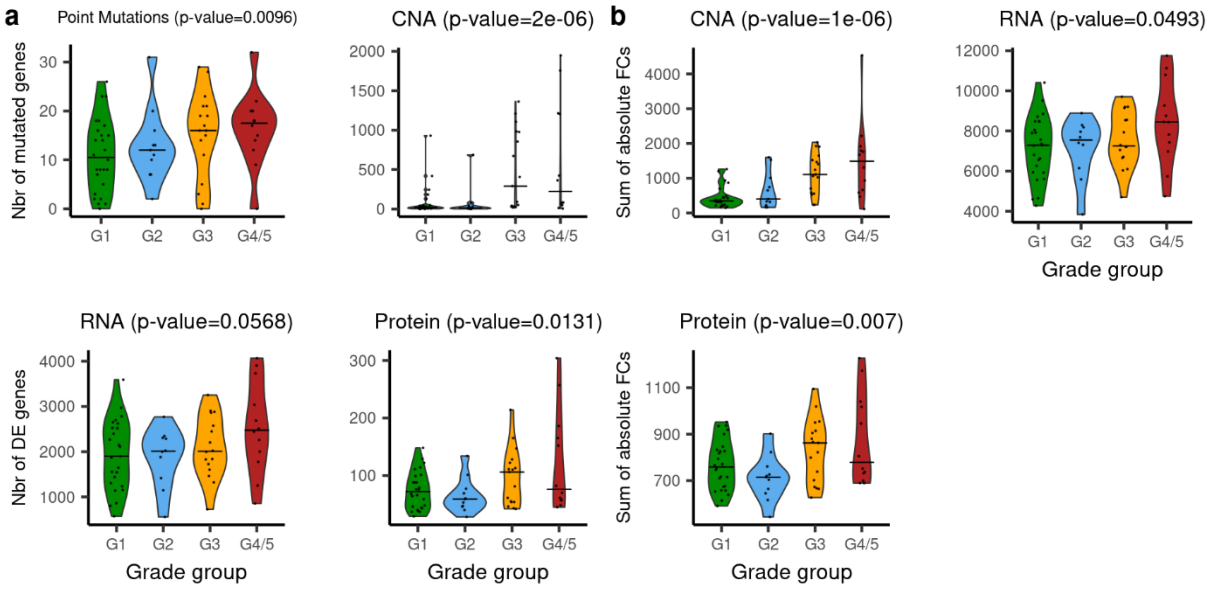


1026

1027

1028 **Figure 2**

1029



1030

1031

1032

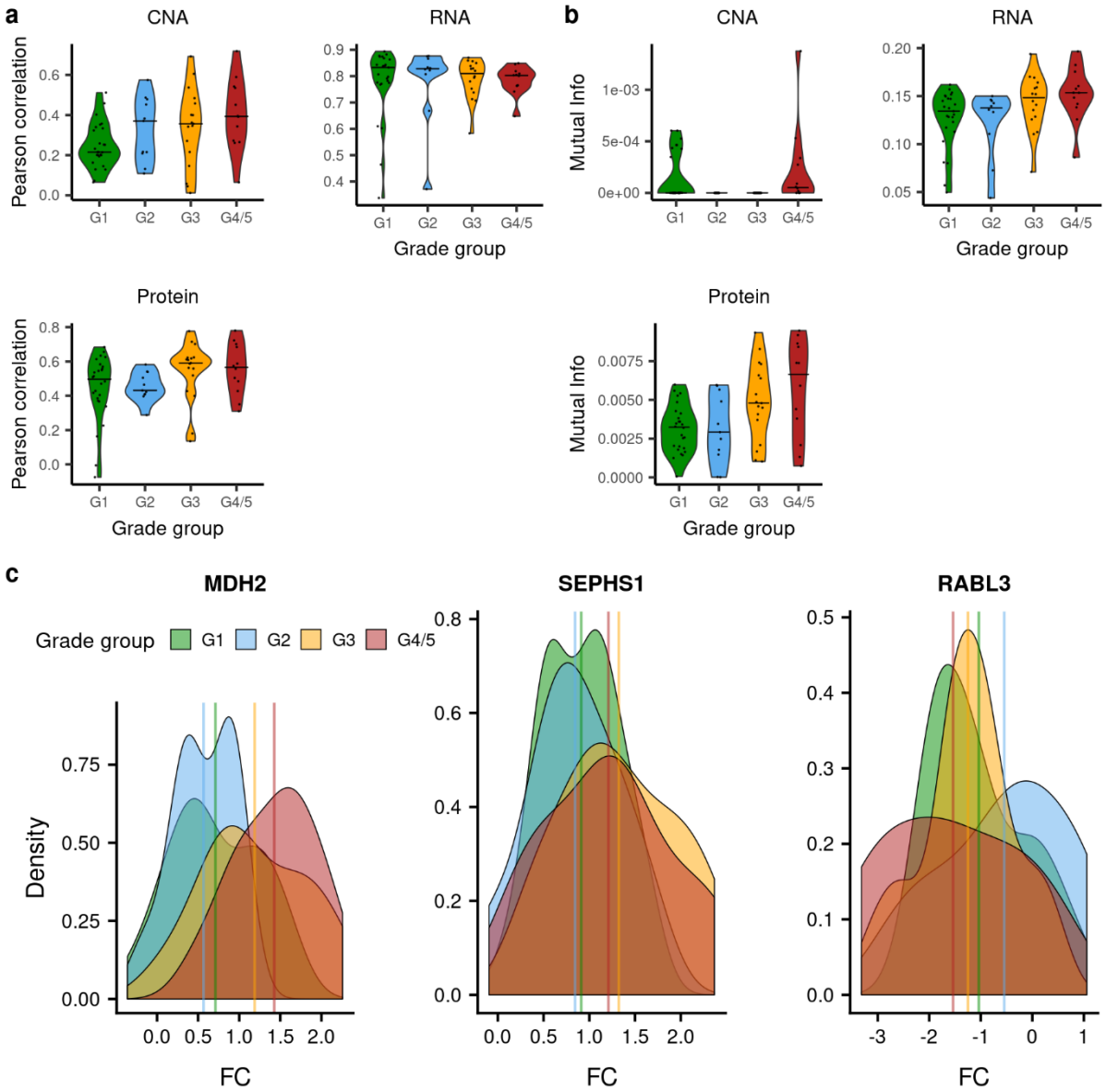
1033

1034

1035

1036 **Figure 3**

1037



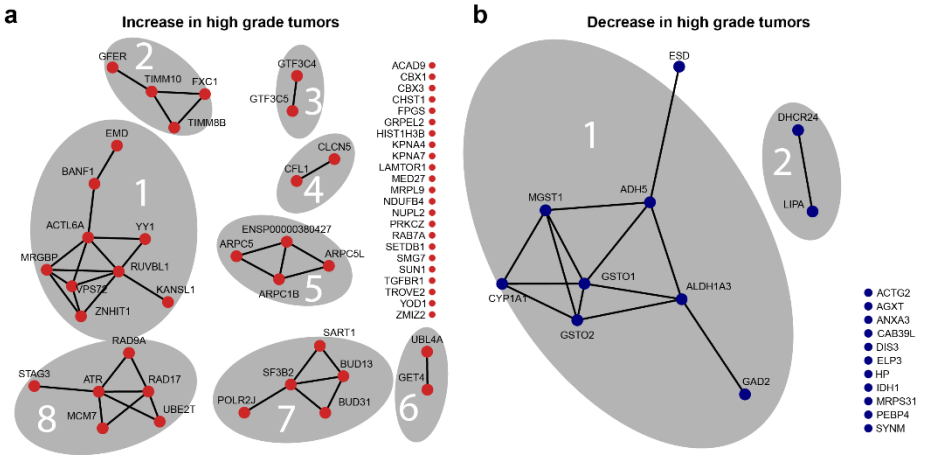
1038

1039

1040

1041 **Figure 4**

1042



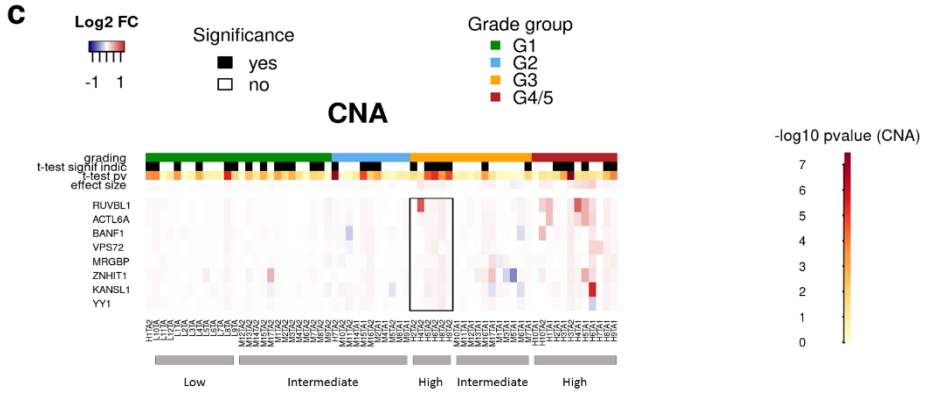
1043

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

1044



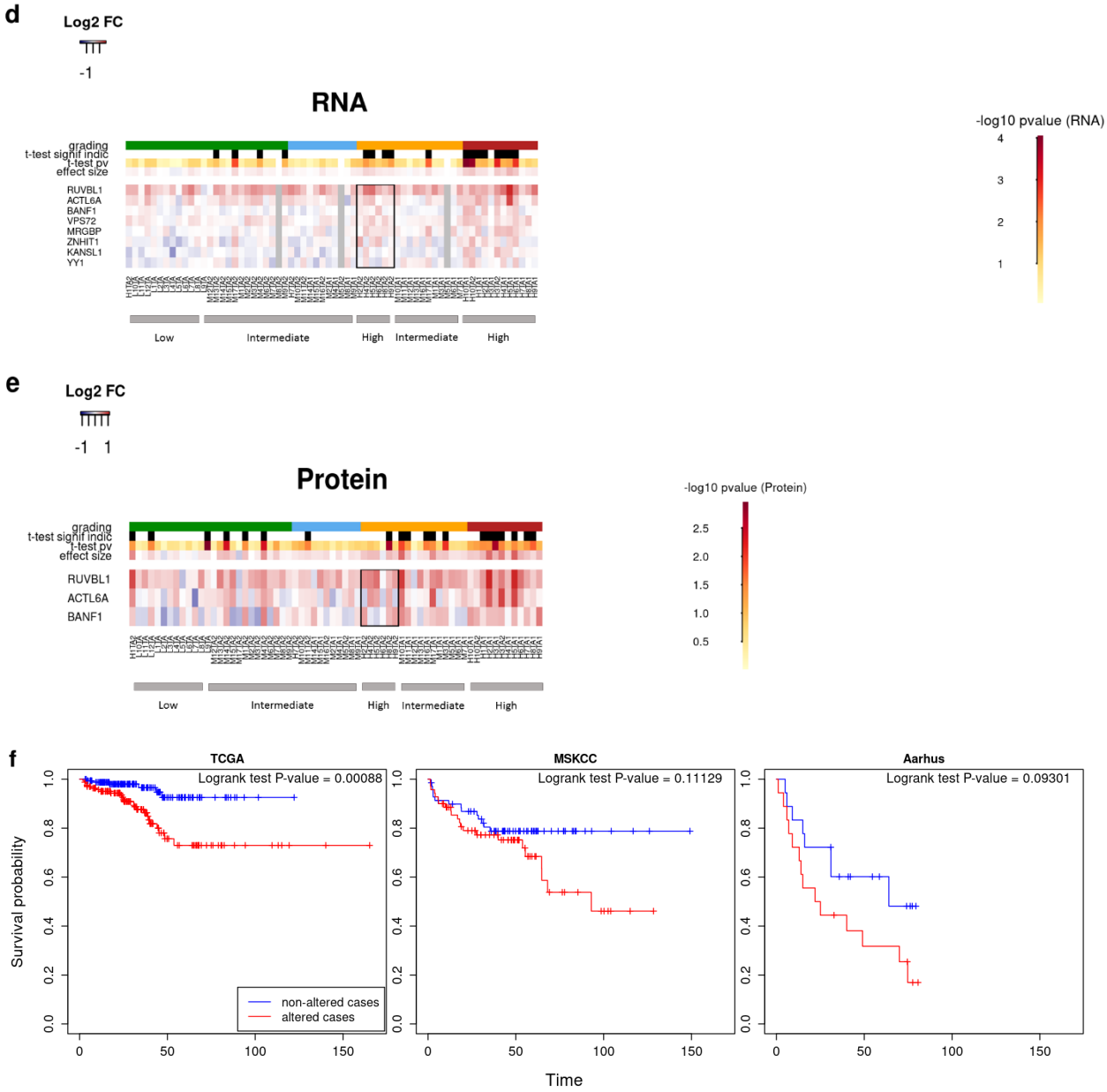
1045

1046

1047

1048 **Figure 4 (*continued*)**

1049



1050

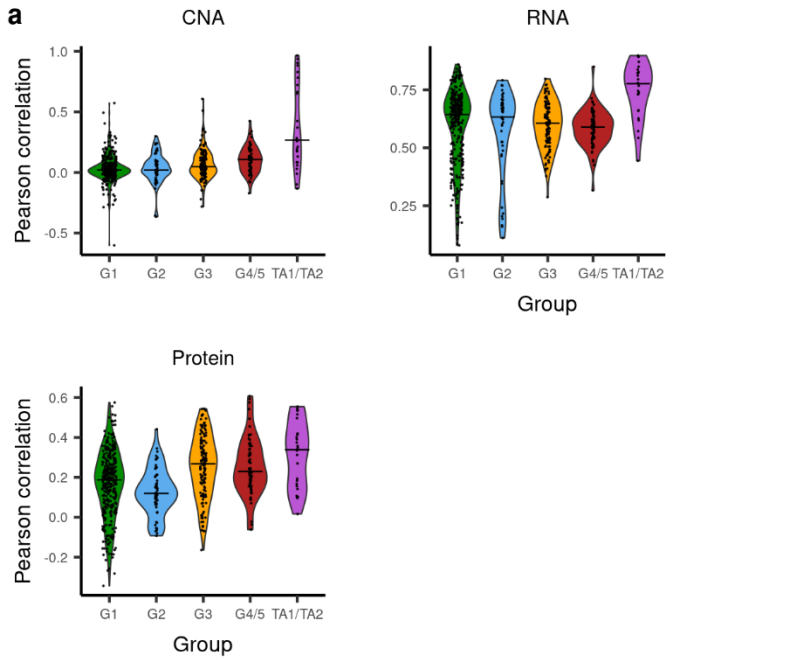
1051

1052

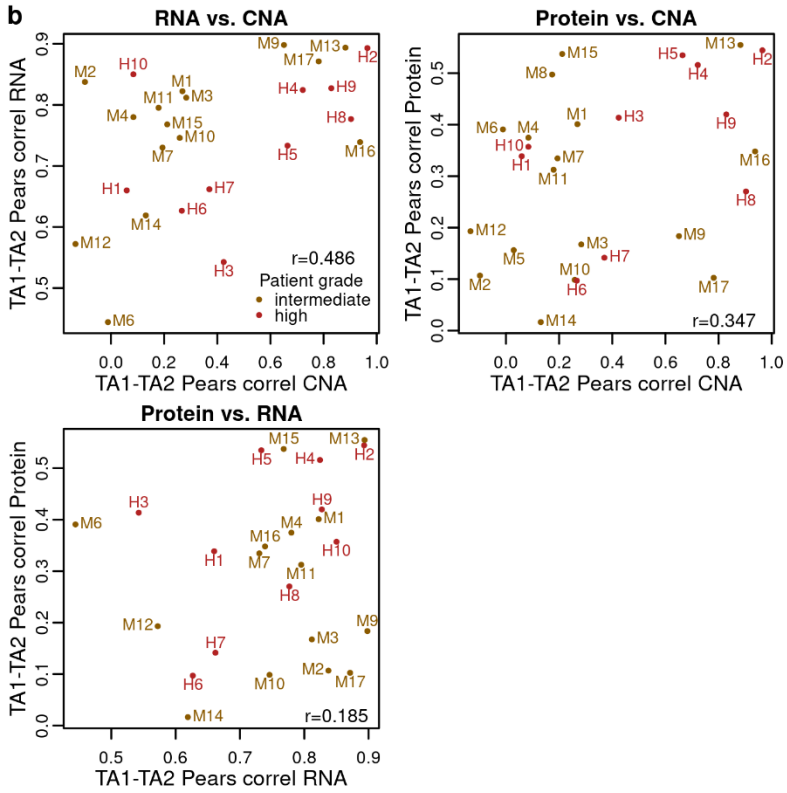
1053

1054 **Figure 5**

1055



1056
1057



1058
1059