

1 **Network integration of multi-tumour omics data suggests novel targeting**
2 **strategies**

3

4 Ítalo Faria do Valle^{1,2*}, Giulia Menichetti^{3*}, Giorgia Simonetti^{4*}, Samantha Bruno⁴, Isabella
5 Zironi¹, Danielle Fernandes Durso^{4,5}, José C M Mombach⁶, Giovanni Martinelli⁴,
6 Gastone Castellani¹, Daniel Remondini^{1§}

7

8 ¹ Department of Physics and Astronomy, University of Bologna – Bologna, Italy

9 ² CAPES Foundation, Ministry of Education of Brazil – Brasília (DF), Brazil

10 ³ Department of Physics, Center for Complex Network Research, Northeastern
11 University, Boston, Massachusetts 02115, USA

12 ⁴ Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna
13 – Bologna, Italy

14 ⁵ National Counsel of Technological and Scientific Development (CNPq), Ministry of
15 Science Technology and Innovation (MCTI), Brasilia, Brazil

16 ⁶ Department of Physics, Universidade Federal de Santa Maria, Santa Maria, RS, Brazi

17 * These authors equally contributed to the paper

18 § Corresponding author: daniel.remondini@unibo.it

19

20

21

22

23

24

25 **Abstract**

26

27 We characterize different tumour types in the search for multi-tumour drug targets, in
28 particular aiming for drug repurposing or novel drug combinations. Starting from 11
29 tumour types from The Cancer Genome Atlas, we obtain three clusters based on
30 transcriptomic correlation profiles. A network-based analysis, integrating gene
31 expression profiles and protein interactions of cancer-related genes, allowed us to
32 define three cluster-specific signatures, with genes belonging to NF- κ B signaling,
33 chromosomal instability, ubiquitin-proteasome system, DNA metabolism, and apoptosis
34 biological processes. These signatures have been characterized by different
35 approaches based on mutational, pharmacological and clinical evidences,
36 demonstrating the validity of our selection. Moreover, we defined new pharmacological
37 strategies validated by *in vitro* experiments that showed inhibition of cell growth in two
38 tumour cell lines, with significant synergistic effect. Our study thus provides a list of
39 genes and pathways with the potential to be used, singularly or in combination, for the
40 design of novel treatment strategies.

41 **Introduction**

42

43 High-throughput molecular profiling has changed the approach to study cancer. For
44 decades, anatomical localization and histological features have guided the identification
45 of cancer subtypes, but the genomic profiling of tumour samples has revealed
46 differences and similarities that go beyond the histopathological classification. The
47 diversity in genomic alteration patterns often stratifies tumours from the same organ or
48 tissue, while tumours in different tissues may present similar patterns¹⁻³. For example,
49 mutational profiling of transcription factors/regulators show tissue specificity, while
50 histone modifiers can be mutated similarly across several cancer types⁴. Hoadley et. al.²
51 suggests that lung squamous, head and neck, and a subset of bladder cancers form a
52 unique cancer category typified by specific alterations, while copy number, protein
53 expression, somatic mutations and activated pathways divide bladder cancer into
54 different subtypes. The analysis of cancer transcriptomes revealed that the same
55 tumour may originate from several cell types, and different biological processes may
56 lead to malignant transformation⁴. Moreover, similar pathways may be activated in
57 different cancers, like ovarian, endometrial and basal-like breast carcinomas^{6,7}.

58 Notwithstanding the enormous increase of knowledge on tumour processes,
59 actually, a practical application of this knowledge to new treatment strategies has not
60 advanced with the same pace. For example, common genetic alterations can predict
61 similar responses to pharmacological therapies across multiple cancer cell lines⁸⁻¹⁰, thus
62 such common molecular and functional profiles could enable the repurposing of
63 therapies from one cancer to another.

64 The huge amount of heterogeneous types of data for a large number of tumours
65 requires novel approaches capable to integrate such information into a unified
66 framework: for this aim, we propose a study of gene networks based on expression
67 profiling and mutational data, in combination with cancer-specific functional annotation.
68 Starting from whole-genome transcriptional profiling extracted from The Cancer
69 Genome Atlas (TCGA) data portal (<https://gdc-portal.nci.nih.gov/>), we selected a
70 curated subset of cancer-related genes and pathways described in the Ontocancro
71 database (<http://ontocancro.inf.ufsm.br/>), and mapped these data onto the BioPlex
72 protein-protein interaction network¹¹. A structural analysis of the obtained networks,
73 based on node centrality, allowed us to rank their relevance and to obtain specific
74 signatures, that may provide multi-tumour drug targets, prognostic markers, and a
75 molecular taxonomy for effective cancer categorization.

76 The validation of our signatures through literature interrogation, clinical
77 information and by *in vitro* testing, makes us confident that this study can help both
78 clinical and research communities, providing novel targets for multi-drug approaches
79 and for repurposing of existing drugs.

80

81 **Results**

82

83 We analyzed transcriptomic data of 2378 samples from 11 tumour types (Table 1)
84 considering 760 cancer-related genes with protein-protein interaction annotation
85 (Bioplex-Ontocancro network, see Methods). The tumour datasets were clustered in
86 three groups based on their gene-gene correlation matrices (see Methods) containing ,
87 respectively, 2, 6 and 3 cancer types: 1) Colon adenocarcinoma (COAD) and Rectum
88 Adenocarcinoma (READ); 2) Lung Adenocarcinoma (LUAD), Lung Squamous Cell
89 Carcinoma (LUSC), Glioblastoma Multiforme (GBM), Ovarian Serous
90 Cystadenocarcinoma (OV), Breast Invasive Carcinoma (BRCA), and Uterine Corpus
91 Endometrial Carcinoma (UCEC); and 3) Brain Lower Grade Glioma (LGG), Kidney
92 Renal Clear Cell Carcinoma (KIRC), and Kidney Renal Papillary Cell Carcinoma (KIRP)
93 (Figure 1). By superimposing the correlation matrices (specific to each cluster) onto the
94 BioPlex-Ontocancro network (common to all tumours), we obtained three weighted
95 networks with approximately 80% of the original nodes and 60% of the original edges
96 (Table 2, see Supplementary Figures 1-4).

97 We hypothesized that the most central genes in each network should play a
98 fundamental role in the tumours represented in the cluster. To find the most central
99 genes we measured the Spectral Centrality (SC)¹², related to the changes in network
100 global diffusivity by node perturbation through a Laplacian formalism, and considered
101 the nodes with SC above the 90th percentile (25, 27 and 24 genes for clusters 1, 2, 3
102 respectively, Table 3). We remark that the chosen signatures have only a small overlap
103 with the most central nodes on the original “full” Bioplex-Ontocancro network not filtered
104 by the cluster-specific correlation matrices (3/25, 13/27 and 4/24 common genes for
105 clusters 1, 2, 3, respectively) showing how the information on gene expression profile is
106 highly specific for the considered tumour clusters. The top-ranking nodes also differ
107 significantly from those obtained from other centrality measures such as degree and
108 betweenness centrality (see Supplementary table 5). Moreover, even if some signature
109 genes overlap between clusters, their links are different (Figures 2, 3, 4, and
110 Supplementary Figure 5) evidencing a specific interaction pattern.

111 We observed that all signatures contain genes related to three biological
112 categories: NF- κ B signaling pathway, chromosomal instability and ubiquitin-proteasome
113 system (Table 4). The chromosomal instability category relates to genes involved in
114 kinetochore formation, microtubule dynamics and chromosome segregation functions.
115 All signatures have at least one substrate recognition component of E3 ubiquitin ligase
116 complexes: *BTRC* in clusters 1 and 2; and *FBXW11* in cluster 3. Cluster 1 has genes
117 involved in spindle checkpoint (*BUB1*, *CDC20*). The cluster 2 signature has many genes
118 related to DNA repair (*CETN2*, *FANCB*, *H2AFX*, *ERCC1*, *ERCC4*, *PARP1*, *XPA*) and
119 DNA replication (*RPA2*, *MCM10*). Moreover, it has three important genes in the
120 signaling path that activates the *STAT3* transcription factor: *SRC*, *NFKB1* and *IL6R*.
121 Indeed, the *STAT3* gene expression levels are significantly higher in cluster 2 (ANOVA
122 p-value: 5.58×10^{-15}) both in comparison with cluster 1 (T-Test p-value: 1.08×10^{-9}) and
123 cluster 3 (T-Test p-value: 1.14×10^{-8}) patients (see Supplementary Figure 6). The cluster
124 3 signature contains genes involved in three different apoptotic mechanisms: induced
125 by TNF- α (*TNFRSF1A* and *BAG4*), induced by Endoplasmatic Reticulum stress
126 (*CAPN1* and *CAPN2*) and caspase-independent apoptosis (*ENDO G*).

127 Then, we searched for possible relationships between the gene signatures and
128 genes commonly mutated in the studied tumours. We observed that some signature
129 genes also present somatic mutations (*REL* and *RAD21* in cluster 1, *ERCC4* and *XPA* in
130 cluster 2, and *AKT2* in cluster 3) or that mutated genes are direct neighbors of the
131 signature genes in the network (see Figures 2, 3, 4). A permutation test over the
132 signature labels (see Methods) reveals a significant proximity of signature genes to
133 mutated genes for cluster 1 and cluster 2 (p-value= 8.76×10^{-4} and p-value= 6.9×10^{-3}
134 respectively, Supplementary Figure 9). For the particular case of cluster 3, only one
135 mutated gene is present in the network and it is successfully selected as a signature
136 gene. These outcomes highlight the strict relationship between signature genes and key
137 processes in tumour development (in analogy with the network-based approach of
138 Novarino et. al.¹³).

139 Since the signature genes are the most central nodes in each cluster, we
140 hypothesized that they might be suitable drug targets. For this purpose we collected,
141 from the DrugBank database, the drugs that target genes in the signatures
142 (Supplementary Table 1) and we evaluated in the ClinicalTrials repository if these drugs
143 are under ongoing clinical trials for cancer treatment. We observed that 11 genes from
144 the cluster signatures are being tested: 4 and 3 genes, from cluster 1 and 2,
145 respectively; 3 genes from both cluster 1 and 2; and 1 gene from both cluster 1 and 3
146 (Table 5, Supplementary Table 2).

147 We then asked whether the expression level of the signature genes could predict
148 the patients survival in each cluster, independently of the tumour type. For cluster 1 and
149 3, survival information were available only for 17 and 32 patients, respectively, which
150 resulted in non-significantly different survival curves, possibly due to the low power of
151 the test (see Supplementary Figures 7 and 8). For cluster 2, we retrieved the clinical
152 information for 448 patients: the survival curves showed that the gene signature
153 significantly separated the patients in two groups according to good or bad survival
154 outcome (Log-rank test p-value = 4.54×10^{-3} , see Methods and Figure 5).

155 We tried to translate our results into novel therapeutic strategies by applying, for
156 a subset of tumours in cluster 2 (which contained the largest and most heterogeneous
157 set of tumours), a set of drugs on targets taken from the signature and from related
158 biological functions. We selected three drugs: Bortezomib, for targeting the proteasome
159 and the NF- κ B pathway; BI6727, for targeting the cluster 2 signature gene *PLK1*; and
160 the PF-00477736 drug, to target the *CHK1/2* genes, which are not in the signature, but
161 also plays a role in the DNA damage response. We tested these drugs, alone or in
162 combination, on the glioblastoma cell line T98G and the breast adenocarcinoma model
163 MCF-7. Both cell lines were highly sensitive to Bortezomib, with an IC50 of 200 nM for
164 MCF-7 and 0.6 nM for T98G (Figure 6). BI6727 treatment reduced viability in a
165 concentration-dependent manner in both models, with the glioblastoma model showing
166 increased responsiveness (IC50 of 69.2 nM *versus* 1.8 μ M for MCF-7, Figure 6).
167 Moreover, both cell lines showed low response to *CHK1/2* inhibition, with IC50 of 26.9
168 μ M for MCF-7 and 15.1 μ M for T98G (Figure 6). We then asked whether these drugs
169 might synergize in the selected models. Although the combinations of PF-00477736
170 with either BI6727 or Bortezomib did not show any additive or synergistic effect in both
171 cell lines (data not shown), we observed a cooperation effect between inhibition of PLK1
172 and proteasome activity (Figure 7A-B). Indeed, we observed in the treatment with the
173 drug combination that the cell viability was significantly lower compared with single
174 agent treatments in MCF-7 cells (Figure 7A, $p < 0.05$), showing a general additive effect
175 (Supplementary Table 4). We observed low Combination Index values (< 1) for both cell
176 lines, indicating synergistic effect for all concentrations tested in the breast cancer
177 model, and for selected concentrations in the glioblastoma model (Figure 7,
178 Supplementary Tables 3 and 4).

179

180

181 **Discussion and conclusion**

182

183 We studied the expression profiles of 11 tumours by considering a selected set of
184 genes from the Ontocancro database and the BioPlex protein-protein interaction
185 network. This knowledge-based selection reduced the dimensionality of the data to a
186 highly curated list of cancer-related genes, involved in pathways that are hallmarks of
187 cancer as cell cycle, inflammation, and apoptosis¹⁴. This approach also ensured that all
188 studied genes had protein-protein interaction annotations, which are crucial to the
189 understanding of how the signaling transduction propagates in the cell¹⁵. We clustered
190 tumours by their gene-gene relationships defined by the Pearson's correlation matrices,
191 to evaluate the functional relationships between genes and their impact on
192 transcriptome organization^{16,17}. tumours from the same organ tended to group together,
193 in agreement with previous studies showing that tissue-of-origin features provide the
194 dominant signals in the identification of cancer subtypes^{2,18}. However, the clustering also
195 grouped tumours originating from different tissues, according to similarities in genomic
196 alterations, as in the case of BRCA, OV, LUSC, and UCEC, which share common
197 characteristics as presence of *TP53* mutations and multiple recurrent chromosomal
198 gains and losses³. In particular, BRCA and UCEC grouped into a well defined sub-
199 cluster, which may reflect their better prognosis when compared to other 10 tumour
200 types².

201 We integrated different types of biological information by a network approach,
202 that allowed us to identify functional modules and to rank genes as network
203 elements^{19,20}. We created a network for each cluster (starting from a common template
204 of protein interactions and superimposing cluster-specific correlation profiles) and
205 obtained specific gene signatures based on node ranking by centrality measures. These
206 signatures presented genes mainly involved in three biological processes: NF- κ B
207 signaling, chromosomal instability and the ubiquitin-proteasome system (Table 4). The
208 NF- κ B signaling pathway regulates genes that participate in cell proliferation, innate and
209 adaptive immune responses, inflammation, cell migration, and apoptosis regulation
210 processes. The aberrant activity of NF- κ B may act as survival factor for transformed
211 cells which would otherwise become senescent or apoptotic²¹. The genes classified into
212 the chromosomal instability category involve kinetochore formation, microtubule
213 dynamics and chromosome segregation functions. The dysfunction in these genes may
214 cause cell inability to faithfully segregate chromosomes, generating genomic alterations
215 as DNA mutation, chromosomal translocation, and gene amplification. The mutant
216 genotypes may confer beneficial phenotypic traits to cancer cells, such as sustained
217 proliferative signaling and resistance to cell death¹⁴. Two genes classified into this
218 category have already been related to clinical practice: the prognostic marker *KIF2C*^{22,23};
219 and the *BUB1* gene, which expression correlates with poor clinical diagnosis^{24,25}. The
220 ubiquitin-proteasome system is the major degradation machinery that controls the
221 abundance of critical regulatory proteins. Perturbation of the regulatory proteins
222 turnover disturbs the intricate balance of signaling pathways and the cellular
223 homeostasis, contributing to the multi-step process of malignant transformation²⁶.
224 Proteasome inhibitors have become valuable tools in the treatment of certain types of
225 cancer, mainly because malignant cells show greater sensitivity to the cytotoxic effects
226 of proteasome inhibition than non-cancer cells²⁷.

227 In addition to common features, cluster 2 signature has several genes related to
228 DNA repair (*CETN2*, *FANCB*, *H2AFX*, *ERCC1*, *ERCC4*, *PARP1*, *XPA*) and DNA
229 replication (*RPA2*, *MCM10*). Interestingly, the tumours in this cluster usually present
230 high rates (50% to 90%) of samples with mutated *TP53*, which is an important sensor
231 for the cell DNA damage response^{2,4,28}. The cluster 2 signature also presents the genes
232 *SRC*, *NFKB*, and *IL6R*, which participates in the activation of *STAT3*, a transcription
233 factor that is necessary for cell transformation²⁹. We observed that *STAT3* gene
234 expression is higher in the tumours of cluster 2 when compared with the tumours of
235 clusters 1 and 3 (Anova p-value: 5.58×10^{-15}). The cluster 3 signature has genes
236 involved in three apoptotic mechanisms, which are induced by TNF- α (*TNFRSF1A* and
237 *BAG4*), or Endoplasmatic Reticulum stress (*CAPN1* and *CAPN2*) and caspase-
238 independent apoptosis (*ENDO*). As the regulation of cell death serves as a natural
239 barrier to cancer development, these processes may reflect different strategies that
240 these tumours use in response to various cellular stresses.

241 Since the transcriptional disturbances observed in cancer can sometimes be
242 explained by underlying somatic mutations^{30,31} we retrieved TCGA mutational data, and
243 focused on cancer related mutations reported in the Catalogue of Somatic Mutations in
244 Cancer (COSMIC) database. Many signature genes resulted also somatically mutated,
245 or first neighbours to mutated genes (Figures 2, 3, 4), showing their strict relationship
246 and the functional relevance of the biologically processes they are involved in.

247 In addition, several genes in the signatures or in their direct network
248 neighborhood are already under clinical investigation in a variety of tumour conditions
249 (as annotated in Clinicaltrials.org database). For example, the AKT pathway has been
250 described as a potential drug intervention in clear cell renal carcinoma³²: *AKT2* gene
251 belongs to the signature of cluster 3 (comprising LGG, KIRC, and KIRP), it is
252 somatically mutated in the tumours of cluster 3 and it has been annotated as drug-target
253 according to the Drug Bank database.

254 We also asked whether the gene signatures could predict survival outcomes in
255 each cluster, thus independently on the tumour type. Our results show that in cluster 2
256 (the only one with enough available samples) the gene signature defined two groups of
257 patients with significantly different Kaplan-Meier survival curves (log-rank test p-value:
258 4.54×10^{-3}).

259 Finally, we tested 3 existing drugs (targeting 2 genes belonging to cluster 2
260 signature, and 1 involved in a related biological process, but not directly belonging to
261 the signature) on 2 tumour types of the cluster, T98G and MCF-7 models. PF-00477736
262 drug (a *CHK1/2* inhibitor, not in the signature)³³ had poor effect on both cell lines, but
263 they resulted highly sensitive to BI6727 (an inhibitor of the signature gene *PLK1*³⁴) and
264 to Bortezomib (proteasome activity inhibitor^{35,36}), with a significant synergic action at
265 several dosages, suggesting a novel therapeutic strategy to be further explored in
266 preclinical models of cluster 2 tumours.

267 These observations indicate that our study succeeded in: 1) clustering tumours
268 highlighting common functional mechanisms related to their transcriptional profile, and
269 2) selecting genes with a relevant functional role in the studied tumours, thus amenable
270 of drug targeting. The combination of these results may thus provide the rationale for
271 choosing novel drug targets and drug combinations, or for repurposing existing drugs
272 towards tumours of the same cluster. As a possible future direction, once obtained an
273 enlarged list of novel and repurposed drugs, the specific transcriptional and mutational
274 profile of single patients, prioritized onto our signatures, might suggest specific
275 combinations of drugs for a more targeted and personalized therapeutic approach.

276

277 **Methods**

278

279 **Gene expression datasets**

280 The gene expression datasets used in this study were retrieved from The Cancer
281 Genome Atlas (TCGA) Data Portal, and included Agilent expression arrays of 2378
282 samples from 11 tumour types, with a different number of samples each (from 16 to 595,
283 see Table 1). We selected for our analysis the genes from the BioPlex protein-protein
284 interaction network¹¹ (n=10961) that were also present in the Ontocancro database
285 (n=1104), resulting in a list of 760 cancer-associated genes related to specific biological
286 functions (such as cell cycle control, DNA damage response, and inflammation).

287

288 **tumour clustering**

289

290 For each tumour dataset, we calculated a correlation matrix containing pairwise
291 Pearson r_{ij} coefficients between genes across all samples available for the tumour. In
292 order to eliminate false correlations and indirect influences, the absolute correlation
293 values ($|r_{ij}|$) were adjusted with the Context Likelihood of Relatedness (CLR)
294 algorithm^{37,38} implemented in the R/Bioconductor package ‘minet’³⁹. The matrices were
295 clustered using hierarchical clustering analysis (with Ward linkage) based on the
296 element-wise Euclidean distance between each pair of tumour matrices A and B,
297 calculated as follows:

298

299
$$d(A, B) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (a_{ij} - b_{ij})^2}$$

300

301 where a_{ij} is the correlation between the genes i and j in the tumour A and b_{ij} is the
302 correlation between the genes i and j in the tumour B.

303

304 **Multi-tumour gene signatures**

305

306 A network approach was applied to find gene signatures that characterized the clusters
307 of tumours. First, we created a backbone network (BioPlex-Ontocancro) by selecting the
308 genes present in the BioPlex protein-protein interaction network that were also
309 annotated in the Ontocancro database. Then, for each cluster the gene-gene correlation
310 coefficients r_{ij} were computed, and their absolute values $|r_{ij}|$ were adjusted with the CLR
311 algorithm, producing the z_{ij} scores^{37,38}. Each score matrix was superimposed to the
312 BioPlex-Ontocancro network, producing three weighted networks (one for each cluster)
313 in which genes were linked only if having correlated expression profiles (with weights
314 given by positive z_{ij} scores, specific to each cluster) and a physical interaction at protein
315 level (given by Bioplex-Ontocancro network, common to all clusters). We remark that
316 the three cluster-related networks result different because of different weight values, or
317 missing links (due to negative z scores set to zero). The networks were analyzed and
318 visualized by Networkx Python package, Matlab and Cytoscape⁴⁰.

319 For the networks of clusters 1 and 3, we selected the giant components (245 and
320 244 nodes, respectively), and for the cluster 2 we selected the two biggest components
321 (149 and 118 nodes). After this selection, we retrieved a gene signature for each cluster
322 composed by the most central genes (nodes), which were defined as those having the
323 Spectral Centrality¹² topological measure SC above the 90th percentile. The SC
324 calculates the effect of node removal on the network diffusivity based on the spectral
325 properties of the Laplacian graph, and it has already been applied successfully to
326 biological data such as the Immune System mediator network. Different results were
327 obtained by considering Betweenness Centrality or weighted degree (Strength W) as
328 centrality measures, as shown in Supplementary Table 5.

329

330 **Validation of the multi-tumour gene signatures**

331

332 We evaluated the relevance of the genes in the signatures by several approaches.
333 First, we verified the proximity with the somatic mutational data extracted from the
334 TCGA data portal for the considered tumours. To avoid cancer unrelated mutations, we
335 considered only mutations that were reported also in the Catalogue of Somatic
336 Mutations in Cancer (COSMIC) database⁴¹. We checked whether the signature genes
337 had been reported as somatic mutated or if they occurred in the neighborhood of
338 mutated genes in the networks. To quantify the proximity of gene signatures to mutated
339 genes we located the nearest mutation (in terms of shortest paths on the network) for
340 each signature gene, resulting in a collection of minimum distance values for each
341 cluster. The average minimum distance from the mutated genes $\langle d_{min} \rangle^{real}$ was then
342 calculated for each cluster and tested with a permutation test. We performed 10^6
343 permutations of the signature labels and recalculated the average minimum distance of
344 each new signature from the mutated genes. The p-values were calculated as

$$345 \quad p = \frac{\sum_{i=1}^{10^6} \langle d_{min} \rangle^i < \langle d_{min} \rangle^{real}}{10^6}$$

346 The results of the proximity analysis are reported in Supplementary Material Figure 9:
347 the signatures of cluster 1 and cluster 2 are significantly closer to mutated genes than
348 expected (p-values 9×10^{-4} and 6.9×10^{-3} , respectively). The permutation test for cluster
349 3 is not completely meaningful because there is only one mutated gene, that anyway
350 results to be one of the signature genes.

351 Secondly, we retrieved from the DrugBank⁴² (<http://www.drugbank.ca/>) and Drug
352 Gene Interaction (DGIdb)⁴³ databases which genes in the signatures were also mapped
353 as drug targets. Third, we checked in the Aggregate Analysis of ClinicalTrials.gov
354 (AACT) database (<https://www.ctti-clinicaltrials.org/aact-database>) for the existence of
355 ongoing clinical trials evaluating the inhibition of genes in the signatures. Fourth, the
356 prognostic potential of each gene signature was evaluated by considering the clinical
357 data (days to death) available in the TCGA data portal. The patients having clinical
358 information were clustered according to the expression levels of the gene signatures by
359 using the k-means algorithm (Python package 'scikit'), considering two patient groups:
360 good versus bad survival outcome. Survival curves were calculated for both groups:
361 we applied the Kaplan-Meier method and evaluated their significance with the log-rank
362 test (Python package "lifelines"). Fifth, we tested the effect of drugs inhibiting genes in
363 our signatures or strictly related to them. The glioblastoma T98G and the breast
364 adenocarcinoma MCF-7 cell lines were obtained from ATCC and DSMZ, respectively.
365 Cells were cultured at a density of 10^5 cells/ml in RPMI medium plus 10% FBS (plus 5%
366 Sodium orthovanadate for T98G) for 72h with increasing concentrations of the following
367 drugs: Bortezomib, BI6727, PF-00477736 (Selleckchem), alone or in combination. One
368 hour and 30 minutes before the end of treatment, WST-1 reagent was added to the cell
369 medium and cell viability was measured according to manufacturer's instruction
370 (Roche). The dose-effect response and the IC50 of each drug were calculated using
371 GraphPad Prism 6 (GraphPad Software). To determine synergy, combination indexes
372 were obtained with the CompuSyn software (ComboSyn Inc.): combination index values
373 <1 , $=1$, and >1 indicate synergism, additive effect and antagonism, respectively.

374

375

376

377

378 **Tables**

379

380 Table 1 – The datasets. List of tumours and their respective number of gene expression
381 arrays

Abbreviation	Cancer	Number of patients
BRCA	Breast invasive carcinoma	593
COAD	Colon adenocarcinoma	172
GBM	Glioblastoma multiforme	595
KIRC	Kidney renal clear cell carcinoma	72
KIRP	Kidney renal papillary carcinoma	16
LGG	Brain lower grade glioma	27
LUAD	Lung adenocarcinoma	32
LUSC	Lung squamous cell carcinoma	155
OV	Ovarian serous cystadenocarcinoma	590
READ	Rectum adenocarcinoma	72
UCEC	Uterine corpus endometrial carcinoma	54
	Total	2378

382

383

384 Table 2 – **Network Properties.** The table shows the main topological features of the
385 cluster networks. Cluster 1: COAD and READ; Cluster 2: LUAD, LUSC, GBM, OV,
386 BRCA, and UCEC; and Cluster 3: LGG, KIRC and KIRP.

	BioPlex- Ontocancro	Cluster 1	Cluster 2	Cluster 3
Clustering Coefficient	0.25	0.21	0.19	0.18
Connected Components	24	41	42	41
Network Diameter	16	18	19	18
Avg Path Length	6.52	7.41	6.88	7.31
Avg Degree	3.84	3.2	3.14	2.98
Number of Nodes	511	406	408	410
Number of Edges	981	650	642	612

387 Table 3 – **Gene Signatures**. List of signature genes for the three tumour clusters.

	BioPlex-Ontocancro	Cluster 1	Cluster 2	Cluster 3
Spectral Centrality > 90 th percentile	ALOX5	ALOX5	BTRC	AKT2
	APP	BTRC	CENPC1	ALOX5
	C17orf70	BUB1	CETN2	BAG4
	CCDC99	CDC20	DSN1	CAPN1
	CETN2	CENPC1	ERCC1	CAPN2
	CSNK2A1	CHUK	ERCC4	CDC16
	CSNK2A2	CUL1	FANCB	CDC27
	EME1	MIS12	FYN	CDT1
	ERCC1	MLF1IP	H2AFX	ENDOG
	ERCC4	NDC80	IL6R	FBXW11
	ERCC6L	NFKB1	MCM10	FNTA
	FANCB	NFKB2	MIS12	GMNN
	GAB1	NFKBIA	MLF1IP	KIF2B
	GRB2	PMF1	NEDD1	KIF2C
	H2AFX	PPP2CB	NFKB1	LSP1
	IL6R	PPP2R5D	NFKBIA	NEDD1
	MAP4K5	PSMB9	NUP43	PRKACG
	MCM10	PSMC2	PARP1	PSMC3
	MLF1IP	PSMF1	PLK1	PSMD9
	MUS81	RAD21	PSMB3	SKP2
	NFKBIA	REL	PSMC3	TNFRSF1A
	NRP1	RELB	RPA2	TUBGCP5
	PIK3CA	RPS27	SRC	UBB
	PIK3CB	SRC	TNFRSF10B	VIM
	PIK3CD	STAG1	TUBGCP5	
	PIK3R2		TUBGCP6	
	PIK3R3		XPA	
	PLK1			
	POLA1			
	POLA2			
	PRIM1			
	PRIM2			
	PSMB3			
	PSMC3			
	RAC1			
	SEC13			
	TNF			

TNFRSF1A
 TRAF6
 UBB
 UBE2T
 XPA
 XPC

388

389

390 Table 4 – **Common biological categories present in the gene signatures.** All cluster
 391 signatures have genes that can be grouped in the following categories: NF- κ B signaling,
 392 chromosomal instability and ubiquitin-proteasome system.

393

	NF-κB Signaling	Chromosomal Instability	Ubiquitin-Proteasome System
Cluster 1	BTRC, CUL1, SRC, NFKBIA, NFKB1, NFKB2, REL, RELB, CHUK	CDC20, BUB1, MLFPIP, CENPC1, MIS12, PMF1, NDC80, RAD21, STAG1	BTRC, CUL1, PSMB9, PSMC2, PSMF1
Cluster 2	BTRC, SRC, NFKBIA, TNFRS10B, IL6R	MIS12, DSN1, MLFPIP, CENPC1, PLK1, NEDD1, TUBGCP5, TUBGCP6	BTRC, PSMB3, PSMC3
Cluster 3	FBXW11, AKT2, TNFR1A	CDC16, CDC27, NEDD1, TUBGCP5, KIF2B, KIF2C	FBXW11, PSMC3, PSMD9

394

395

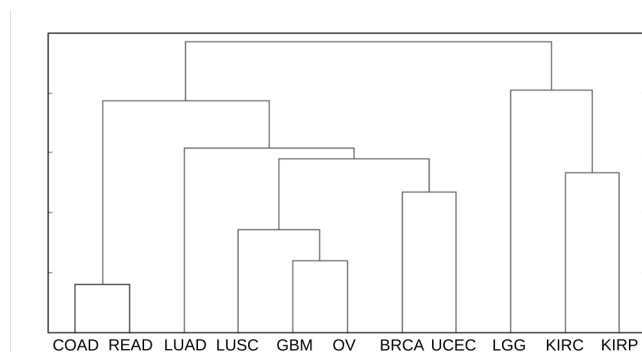
396 Table 5 – List of genes from the signatures that are also being tested in ongoing clinical

397 trials studies (according ClinicalTrials.gov).

Inhibition target	Number of clinical trials	Cluster signature
ALOX5	18	1, 3
CHUK	9	1
FYN	97	2
IL6R	2	2
NFKB1	40	1, 2
NKFB2	8	1
NKFBIA	8	1, 2
PARP1	106	2
PPP2CB	5	1
PSMB9	25	1
SRC	135	1, 2

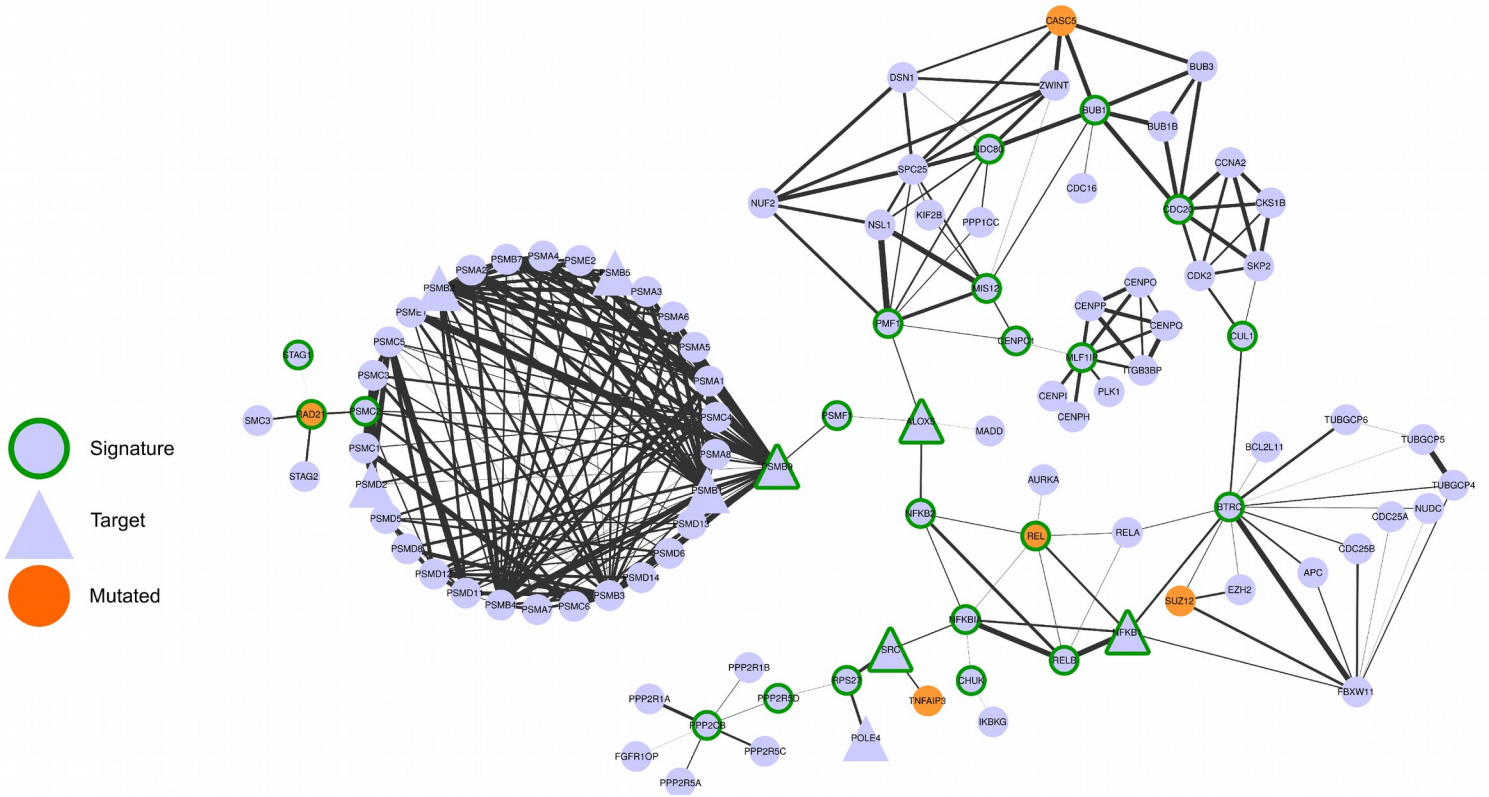
398 **Figures**

399



401 **Figure 1 – tumour clustering.** For each tumour, we produced a matrix from the
402 correlation (Pearson) of the expression profiles among 760 genes. The correlations
403 values were adjusted by the CLR algorithm. Then, we clustered the resulting matrices
404 by euclidean metrics.

405

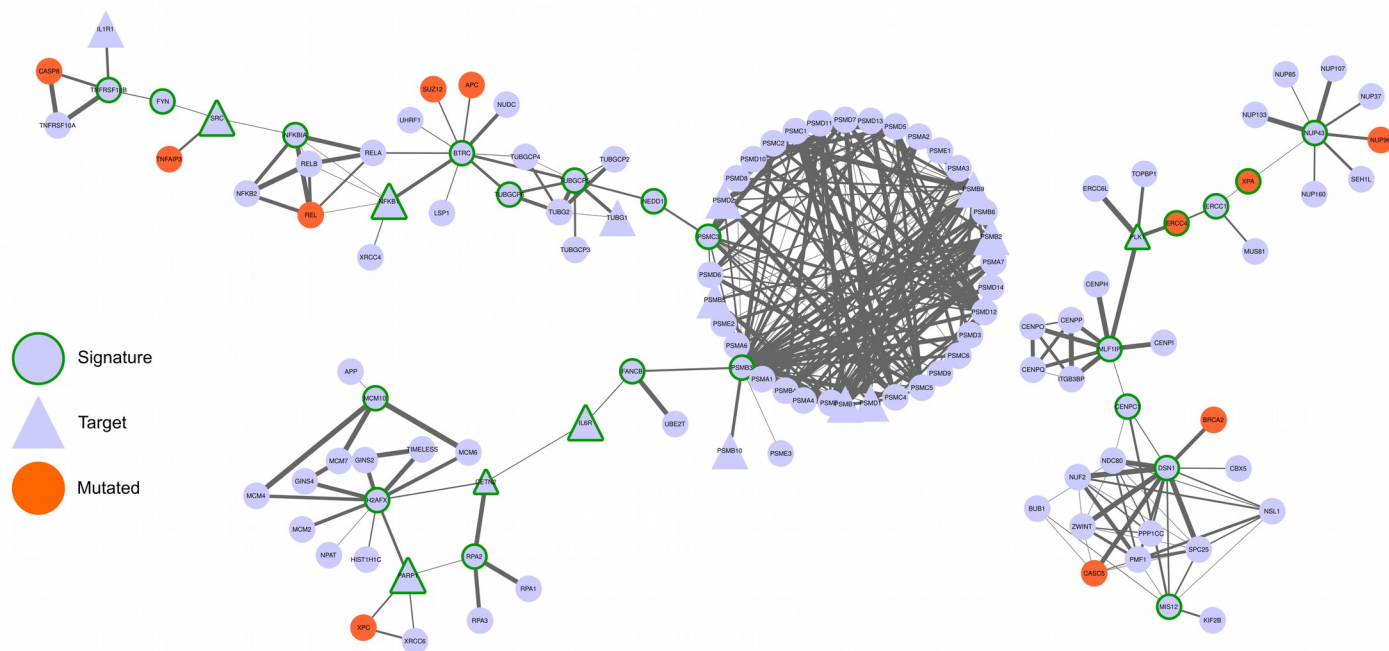


407 Figure 2 – Network composed by the first neighbors of the cluster 1 signature

408 genes

409

410

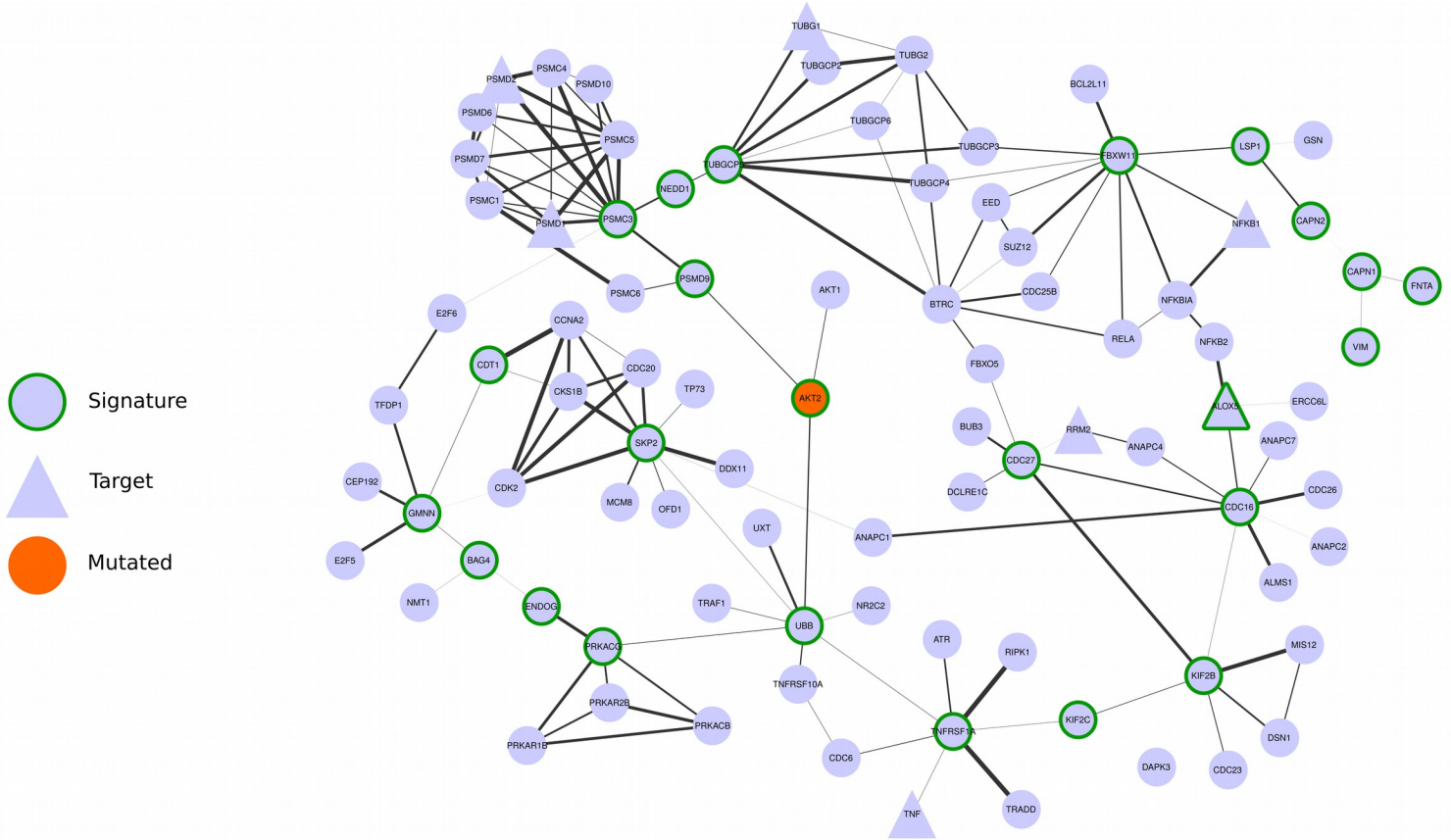


411 Figure 3 – Network composed by the first neighbors of the cluster 2 signature

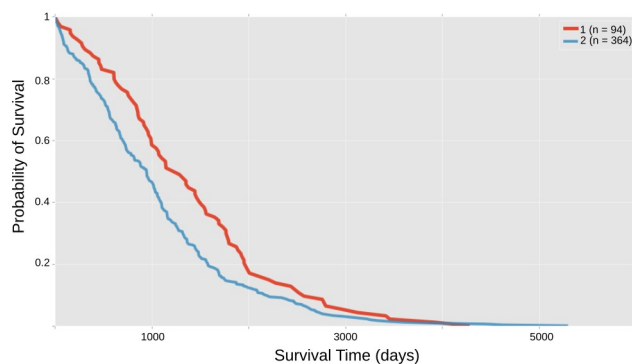
412 genes.

413

414

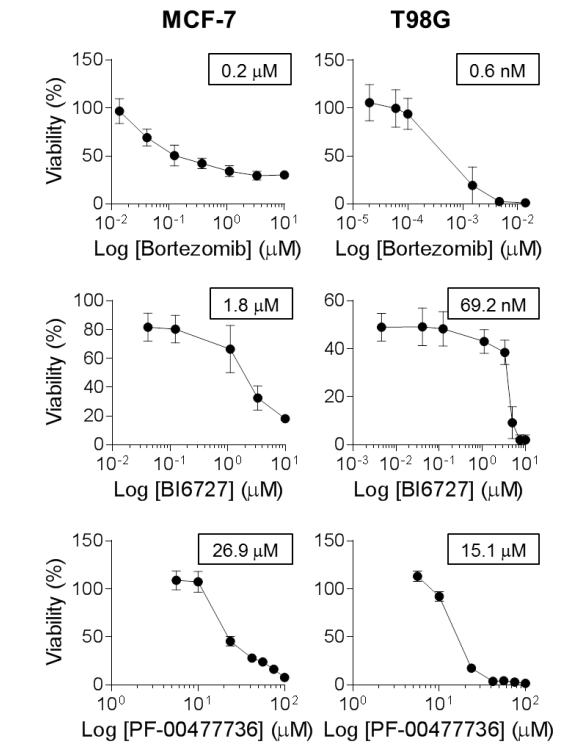


415 Figure 4 – Network composed by the first first neighbors of the cluster 3 signature
416 genes
417



419 Figure 5 – Gene signature and survival outcome. The cluster 2 signature defined two
420 groups of patients with significantly different Kaplan-Meier survival curves.

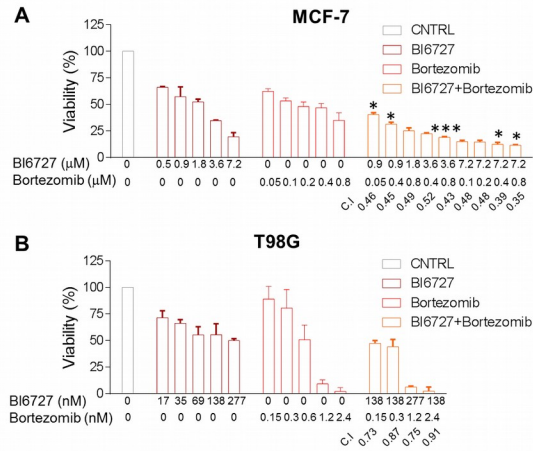
421



423 Figure 6 – *In vitro* response of cancer cell lines from signature 2 to treatment with
424 Bortezomib, BI6727 and PF-00477736 as single agents. MCF-7 and T98G cells were
425 treated with increasing doses of Bortezomib (0.01 to 10 μ M for MCF-7, 0.02 to 10 nM
426 for T98G), BI6727 (0.04 to 10 μ M for MCF-7, 0.004 to 10 μ M for T98G), PF-00477736
427 (5.6 to 100 μ M) and cell viability was measured 72h after drug administration by WST-1
428 assay (three independent experiments). Cell viability is represented as (mean \pm SEM).
429 IC50 values are reported in the boxes (GraphPad Prism 6).

430

431



432 Figure 7 – **Sensitivity of MCF-7 and T98G cells to combined inhibition of PLK1 and**
 433 **proteasome activity.** MCF-7 and T98G cells were treated with increasing doses of
 434 Bortezomib (0.05 to 0.8 μM for MCF-7, 0.15 to 2.4 nM for T98G) and BI6727 (0.5 to 7.2
 435 μM for MCF-7, 17 to 277 nM for T98G), alone or in combination and cell viability was
 436 measured 72h after drug administration by WST-1 assay (three independent
 437 experiments). Statistical significance was determined by Student’s t test (*, P < 0.05; ***,
 438 P < 0.001). Combination index (C.I.) was calculated by CompuSyn software. (A) MCF-7
 439 cells: combinations with a C.I. lower than 0.5 are shown. (B) T98G cells: combinations
 440 showing synergistic effect are shown.

441

442 **Supplementary Information**

443

444 Supplementary Figure 1 – Bioplex-Ontocancro Network. Network built from the 760

445 genes found both in BioPlex protein-protein interaction network and Ontocancro

446 database

447

448 Supplementary Figure 2 – Overview of the cluster 1 network. Diamonds with red
449 borders represent the genes in the cluster 1 signature and orange circles represent the
450 mutated genes

451

452 Supplementary Figure 3 – Overview of the cluster 2 network. Diamonds with red
453 borders represent the genes in the cluster 2 signature and orange circles represent the
454 mutated genes

455

456 Supplementary Figure 4 – Overview of the cluster 3 network. Diamonds with red
457 borders represent the genes in the cluster 3 signature and orange circles represent the
458 mutated genes

459

460 Supplementary Figure 5 – Network of signature genes common to cluster 1 and 2. Even
461 though the signatures can present common genes, they have different set of interactors
462 in each cluster network.

463

464 Supplementary Figure 6 – Boxplot of STAT3 levels for clusters 1, 2, 3. Cluster 2 patients
465 presented higher STAT3 gene expression in comparison with cluster 1 (T-Test p-value:
466 1.08×10^{-9}) and cluster 3 (T-Test p-value: 1.14×10^{-8}).

467

468 Supplementary Figure 7 – Kaplan-Meier curves for the two groups of cluster 1 patients
469 defined by K-means clustering approach. The clustering was applied only to the genes
470 in cluster 1 signature. Only 17 patients had the survival information in the TCGA data
471 portal (logrank-test pvalue = 0.9118)

472

473 Supplementary Figure 8 – Kaplan-Meier curves for the two groups of cluster 3 patients
474 defined by K-means clustering approach. The clustering was applied considering only
475 the genes in cluster 3 signature. Only 32 patients had the survival information in the
476 TCGA data portal (logrank-test pvalue = 0.9056)

477

478 Supplementary Figure 9 - Plot of the distribution of the 10^6 permutations for the 3
479 clusters (from left to right). The inboxes show the minimum average distances for the
480 signatures (represented in the plots as red vertical lines), and the p-values with respect
481 to the permutations.

482

483 Supplementary Table 1 – List of drug-gene interactions for the genes in the signatures,
484 extracted from the Drug Gene Interaction database (DGIdb).

485

486 Supplementary Table 2 – List of ongoing clinical trials (according to ClinicalTrials.gov)
487 that evaluate the inhibition of the genes in the signatures.

488

489 Supplementary Table 3 – Combination Indexes for BI6727 and Bortezomib treatment at
490 different concentrations in the T98G cell line.

491

492 Supplementary Table 4 – Combination Indexes for BI6727 and Bortezomib treatment at
493 different concentrations in the MCF-7 cell line.

494

495 Supplementary Table 5 - Spearman's rank correlation values for the centrality measures
496 (Spectral Centrality SC, Betweenness Centrality BC, strength W) on the nodes for the 3
497 clusters. The results refer to the whole node list ("All") or only to the signatures,
498 obtained as the top 10% of the ranked measures ("90th"). We remark the drop in
499 correlation when considering only the gene signatures obtained by the different
500 centrality measures.

501

502

503 **List of Abbreviations**

504

505 TCGA: The Cancer Genome Atlas; SC: Spectral Centrality; COAD: Colon
506 Adenocarcinoma; READ: Rectum Adenocarcinoma; LUAD: Lung Adenocarcinoma;
507 LUSC: Lung Squamous Cell Carcinoma; GBM: Glioblastoma Multiforme; OV: Ovarian
508 Serous Cystadenocarcinoma; BRCA: Breast Invasive Carcinoma; UCEC: Uterine
509 Corpus Endometrial Carcinoma; LGG: Brain Lower Grade Glioma; KIRC: Kidney Renal
510 Clear Cell Carcinoma; KIRP: Kidney Renal Papillary Cell Carcinoma.

511

512 **Competing interests**

513 The others authors declare that they have no competing interests.

514

515 **Authors' Contributions**

516 IFV, GM and GS contributed equally to the paper. GS, SB, IZ and DFD performed the
517 experiments and interpreted the results. IFV, GM and JCM collected and analyzed the
518 data. GM, GC and DR designed the research, interpreted the results, wrote the paper.

519

520 **Authors' information**

521

522 ¹ Department of Physics and Astronomy, University of Bologna – Bologna, Italy

523 ² CAPES Foundation, Ministry of Education of Brazil – Brasília (DF), Brazil

524 ³ Department of Physics, Center for Complex Network Research, Northeastern

525 University, Boston, Massachusetts 02115, USA

526 ⁴ Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna

527 – Bologna, Italy

528 ⁵ National Counsel of Technological and Scientific Development (CNPq), Ministry of

529 Science Technology and Innovation (MCTI), Brasilia, Brazil

530 ⁶ Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

531 * These authors equally contributed to the paper

532 § Corresponding author: daniel.remondini@unibo.it

533

534 **Acknowledgements**

535 We thank the Science Without Borders project of CAPES foundation (Ministry of

536 Education of Brazil - Brasília -DF, Brazil) for the doctoral scholarship (grant number:

537 10186-13-1) for IFV. This study was supported by the Interomics CNR Flagship

538 Initiative; Mimomics EU FP7 Project n. 305280; NGS PTL EU FP7 Project n. 306242;

539 the CNPq Project n. 402547/2012-8; the Associazione Italiana per la Ricerca sul Cancro

540 AIRC (Investigator Grant to GM, n. 19226); the Programma di ricerca Regione-

541 Università 2010-2012 (L. Bolondi); and the Innovative Medicines Initiative (IMI) 2 project

542 "HARMONY", n 116026.

543

544

545 References

546

- 547 1. Chang, K. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat.*
548 *Genet.* **45**, 1113–1120 (2013).
- 549 2. Hoadley, K. a. *et al.* Multiplatform Analysis of 12 Cancer Types Reveals
550 Molecular Classification within and across tissues of origin. *Cell* **158**, 929–944
551 (2013).
- 552 3. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human
553 cancers. *Nat Genet.* **45**, 1127–1133 (2013).
- 554 4. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer
555 types. *Nature* **502**, 333–9 (2013).
- 556 5. Verhaak, R. G. W. *et al.* Integrated genomic analysis identifies clinically relevant
557 subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1,
558 EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
- 559 6. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast
560 tumours. *Nature* **490**, 61–70 (2012).
- 561 7. Network, C. G. A. R. *et al.* Integrated genomic characterization of endometrial
562 carcinoma. *Nature* **497**, 67–73 (2013).
- 563 8. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive
564 modelling of anticancer drug sensitivity. *Nature* **483**, 603–7 (2012).
- 565 9. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug
566 sensitivity in cancer cells. *Nature* **483**, 570–5 (2012).
- 567 10. Heiser, L. M. *et al.* Subtype and pathway specific responses to anticancer
568 compounds in breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2724–9 (2012).
- 569 11. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human
570 Interactome. *Cell* **162**, 425–440 (2015).
- 571 12. Pauls, S. D. & Remondini, D. Measures of centrality based on the spectrum of
572 the Laplacian. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **85**, 066127
573 (2012).
- 574 13. Novarino, G. *et al.* Exome sequencing links corticospinal motor neuron disease
575 to common neurodegenerative disorders. *Science* **343**, 506–511 (2014)
- 576 14. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell*
577 **144**, 646–74 (2011).
- 578 15. Vinayagam, A *et al.* A Directed Protein Interaction Network for Investigating
579 Intracellular Signal Transduction. *Sci. Signal.* **4**, rs8–rs8 (2011).
- 580 16. Lee, H. K., Hsu, A. K., Sajdak, J., Qin, J. & Pavlidis, P. Coexpression analysis of
581 human genes across many microarray data sets. *Genome Res.* **14**, 1085–94
582 (2004).

- 583 17. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and
584 display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.*
585 **95**, 14863–8 (1998).
- 586 18. Martínez, E. *et al.* Comparison of gene expression patterns across 12 tumour
587 types identifies a cancer supercluster characterized by TP53 mutations and
588 cell cycle defects. *Oncogene* **34**, 2732–2740 (2015).
- 589 19. Ideker, T. & Sharan, R. Protein networks in diseases. *Genome Res.* **18**, 644–
590 652 (2008).
- 591 20. Castellani, G. C. *et al.* Systems medicine of inflammaging. *Brief. Bioinform.*
592 1–14 (2015). doi:10.1093/bib/bbv062
- 593 21. Hoesel, B. & Schmid, J. A. The complexity of NF- κ B signaling in inflammation
594 and cancer. *Mol. Cancer* **12**, 86 (2013).
- 595 22. Bie, L., Zhao, G., Wang, Y. & Zhang, B. Kinesin family member 2C
596 (KIF2C/MCAK) is a novel marker for prognosis in human gliomas. *Clin.*
597 *Neurol. Neurosurg.* **114**, 356–60 (2012).
- 598 23. Sanhaji, M., Friel, C. T., Wordeman, L., Louwen, F. & Yuan, J. Mitotic
599 centromere-associated kinesin (MCAK): a potential cancer drug target.
600 *Oncotarget* **2**, 935–947 (2011).
- 601 24. Bie, L. *et al.* The accuracy of survival time prediction for patients with glioma is
602 improved by measuring mitotic spindle checkpoint gene expression. *PLoS*
603 *One* **6**, e25631 (2011).
- 604 25. Finetti, P. *et al.* Sixteen-kinase gene expression identifies luminal breast
605 cancers with poor prognosis. *Cancer Res.* **68**, 767–76 (2008).
- 606 26. Sahasrabudhe, A. A. & Elinitoba-Johnson, K. S. J. Role of the ubiquitin
607 proteasome system in hematologic malignancies. *Immunol. Rev.* **263**,
608 224–239 (2015).
- 609 27. Saez, I. & Vilchez, D. The Mechanistic Links Between Proteasome Activity,
610 Aging and Age-related Diseases. *Curr. Genomics* **15**, 38–51 (2014).
- 611 28. Collisson, E. a. *et al.* Comprehensive molecular profiling of lung
612 adenocarcinoma. *Nature* **511**, 543–50 (2014).
- 613 29. Iliopoulos, D., Hirsch, H. A. & Struhl, K. An Epigenetic Switch Involving NF-
614 κ B, Lin28, Let-7 MicroRNA, and IL6 Links Inflammation to Cell
615 Transformation. *Cell* **139**, 693–706 (2009).
- 616 30. Shlien, A. *et al.* Direct Transcriptional Consequences of Somatic Mutation in
617 Breast Cancer. *Cell Rep.* **16**, 2032–2046 (2016).
- 618 31. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish
619 tumour subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.*
620 **98**, 10869–74 (2001).
- 621 32. Creighton, C. J. *et al.* Comprehensive molecular characterization of clear cell
622 renal cell carcinoma. *Nature* **499**, 43–9 (2013).

- 623 33. Lee, C. *et al.* Polo-like kinase 1 inhibition kills glioblastoma multiforme brain
624 tumour cells in part through loss of SOX2 and delays tumour progression
625 in mice. *Stem Cells* **30**, 1064–75 (2012).
626 34. Bholra, N. E. *et al.* Kinome-wide functional screen identifies role of PLK1 in
627 hormone-independent, ER-positive breast cancer. *Cancer Res.* **75**, 405–
628 14 (2015).
629 35. Thaler, S. *et al.* The proteasome inhibitor Bortezomib (Velcade) as potential
630 inhibitor of estrogen receptor-positive breast cancer. *Int. J. Cancer* **137**, 686–
631 97 (2015).
632 36. Yin, D. *et al.* Proteasome inhibitor PS-341 causes cell growth arrest and
633 apoptosis in human glioblastoma multiforme (GBM). *Oncogene* **24**, 344–54
634 (2005).
635 37. [Saccanti E](#), *et al.* Entropy-Based Network Representation of the Individual
636 Metabolic Phenotype. *J. Proteome Res.* **15** (9), 3298-307 (2016).
637 38. Faith, J. J. *et al.* Large-scale mapping and validation of *Escherichia coli*
638 transcriptional regulation from a compendium of expression profiles. *PLoS*
639 *Biol.* **5**, 0054–0066 (2007).
640 39. Meyer, P. E., Lafitte, F. & Bontempi, G. minet: A R/Bioconductor package for
641 inferring large transcriptional networks using mutual information. *BMC*
642 *Bioinformatics* **9**, 461 (2008).
643 40. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of
644 biomolecular interaction networks. *Genome Res.* **13**, 2498–504 (2003).
645 41. Forbes, S. A. *et al.* The Catalogue of Somatic Mutations in Cancer (COSMIC).
646 *Curr. Protoc. Hum. Genet.* **Chapter 10**, Unit 10.11 (2008).
647 42. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug
648 discovery and exploration. *Nucleic Acids Res.* **34**, D668–72 (2006).
649 43. Wagner, A. H. *et al.* DGIdb 2.0: mining clinically relevant drug-gene interactions.
650 *Nucleic Acids Res.* **44**, D1036–44 (2016).
651

652

653

654