

ARTICLE

DOI: 10.1038/s41467-018-06992-7

OPEN

# Network integration of multi-tumour omics data suggests novel targeting strategies

Ítalo Faria do Valle<sup>1,2</sup>, Giulia Menichetti<sup>3</sup>, Giorgia Simonetti<sup>4</sup>, Samantha Bruno<sup>4</sup>, Isabella Zironi<sup>1</sup>, Danielle Fernandes Durso<sup>4,5</sup>, José C.M. Mombach<sup>6</sup>, Giovanni Martinelli<sup>4,7</sup>, Gastone Castellani<sup>1</sup> & Daniel Remondini<sup>1</sup>

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

<sup>1</sup>Department of Physics and Astronomy, University of Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy. <sup>2</sup>CAPES Foundation, Ministry of Education of Brazil – Setor Bancário Norte (SBN), Quadra 2, Bloco L, Lote 06, Edifício CAPES, 70.040-031 Brasília, DF, Brazil. <sup>3</sup>Department of Physics, Center for Complex Network Research, Northeastern University, 177 Huntington Avenue, 02115 Boston, MA, USA. <sup>4</sup>Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna and Institute of Hematology “L. and A. Seràgnoli”, Via Massarenti 9, 40138 Bologna, Italy. <sup>5</sup>National Council of Technological and Scientific Development (CNPq), Ministry of Science Technology and Innovation (MCTI), Brasília 70000-000 DF, Brazil. <sup>6</sup>Department of Physics, Universidade Federal de Santa Maria, Av. Roraima 1000, 97105-900 Santa Maria, RS, Brazil. <sup>7</sup>Present address: Istituto Scientifico Romagnolo per lo Studio e Cura dei Tumori (IRST) IRCCS, Via Piero Maroncelli 40, 47014 Meldola, Italy. These authors contributed equally: Ítalo Faria do Valle, Giulia Menichetti, Giorgia Simonetti. Correspondence and requests for materials should be addressed to D.R. (email: [daniel.remondini@unibo.it](mailto:daniel.remondini@unibo.it))

High-throughput molecular profiling has changed the approach to study cancer. For decades, anatomical localization and histological features have guided the identification of cancer subtypes, but the genomic profiling of tumour samples has revealed differences and similarities that go beyond the histopathological classification. The diversity in genomic alteration patterns often stratifies tumours from the same organ or tissue, while tumours in different tissues may present similar patterns<sup>1–3</sup>. For example, mutational profiling of regulatory proteins shows tissue specificity, while histone modifiers can be mutated similarly across several cancer types<sup>4</sup>. Hoadley et al.<sup>2</sup> suggests that lung squamous, head and neck, and a subset of bladder cancers form a unique cancer category typified by specific alterations, while copy number, protein expression, somatic mutations and activated pathways divide bladder cancer into different subtypes. The analysis of cancer transcriptomes revealed that the same tumour may originate from several cell types, and different biological processes may lead to malignant transformation<sup>4</sup>. Moreover, similar pathways may be activated in different cancers, like ovarian, endometrial and basal-like breast carcinomas<sup>5,6</sup>. Notwithstanding the enormous increase of knowledge on tumour processes, a practical application of this knowledge to new treatment strategies has not advanced with the same pace. For example, common genetic alterations can predict similar responses to pharmacological therapies across multiple cancer cell lines<sup>7–9</sup>, thus common molecular and functional profiles could enable the repurposing of therapies from one cancer to another.

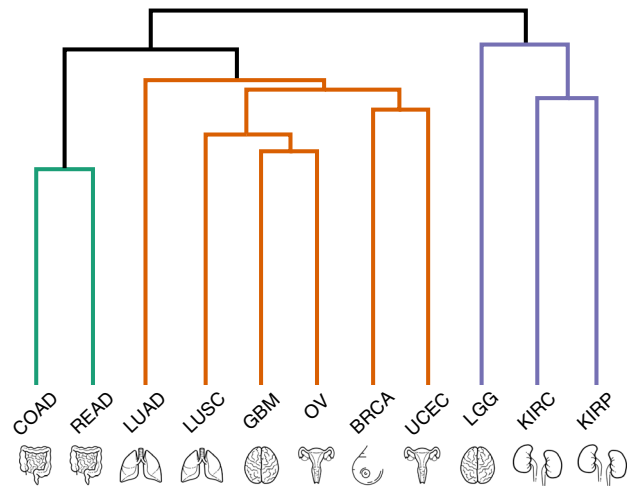
Several methods have been proposed and applied for the analysis of omics data in cancer<sup>10</sup>. Generally, they refer to: (a) reconstruction of regulatory networks from expression data,<sup>11</sup> (b) identification of network modules by clustering or network diffusion techniques (usually starting from an a priori selection of seed genes as somatic mutations and differentially expressed genes);<sup>12–15</sup> and (c) evaluation of cancer alterations at the pathway-level comparing many samples<sup>16,17</sup>. However, the search for new drug targeting and repurposing strategies requires different network approaches able to evaluate a broad list of genes and identify their individual impact in the underlying regulatory networks of several tumour types at the same time.

For this aim, we propose a study of gene networks based on expression profiling and interactome topology, in combination with cancer-specific functional annotation.

Starting from whole-genome transcriptional profiling extracted from The Cancer Genome Atlas (TCGA) data portal, we selected a curated subset of 760 cancer-related genes described both in the Ontocancro database<sup>18</sup>, and in the BioPlex protein–protein interaction—PPI-network<sup>19,20</sup>. We defined three tumour clusters starting from the gene–gene correlation matrices of each tumour (see Methods). Then we performed a topological analysis of the corresponding networks based on node centrality, obtaining specific signatures for multi-tumour drug targeting and survival prognosis. The validation of our signatures through literature interrogation, clinical information and by in vitro testing, makes us confident that this study can help both clinical and research communities, providing novel targets for multi-drug approaches and for repurposing of existing drugs.

## Results

**Identification of multi-tumour gene signatures.** We analyzed transcriptomics data of 2378 samples from 11 tumour types (Supplementary Table 1) considering 760 cancer-related genes with both oncogenic and PPI annotation (BioPlex-Ontocancro network, see Methods). The tumour datasets were clustered in three groups based on their gene–gene correlation matrices (see Methods) containing,



**Fig. 1** Tumour clustering. Tumour dendrogram according to the hierarchical clustering of 760-gene z-scores filtered by CLR algorithm. Brain, lung, intestine, uterus, and kidneys icons made by Kirill Kazachek (<https://www.flaticon.com/authors/kirill-kazachek>) and breast icon made by Freepik (<https://www.flaticon.com/authors/freepik>)

respectively, 2, 6 and 3 cancer types: (1) Colon adenocarcinoma (COAD) and Rectum Adenocarcinoma (READ); (2) Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LUSC), Glioblastoma Multiforme (GBM), Ovarian Serous Cystadenocarcinoma (OV), Breast Invasive Carcinoma (BRCA), and Uterine Corpus Endometrial Carcinoma (UCEC); and (3) Brain Lower Grade Glioma (LGG), Kidney Renal Clear Cell Carcinoma (KIRC), and Kidney Renal Papillary Cell Carcinoma (KIRP) (Fig. 1). We superimposed the gene–gene correlation matrices (calculated with the samples of all tumours inside each cluster) onto the BioPlex-Ontocancro network: a network of genes resulting from the intersection of the BioPlex network (containing physical protein–protein interactions measured via mass spectrometry) with the Ontocancro database (with annotations for cancer-related genes and pathways, see Methods). We obtained three weighted networks, each with approximately 80% nodes and 60% edges of the original BioPlex-Ontocancro network (Supplementary Note 1, Supplementary Table 2, Supplementary Figures 1–4).

We hypothesized that the most central genes in each network should play a fundamental role in the tumours represented in the cluster, in accordance to similar approaches proposed in previous works<sup>21,22</sup>.

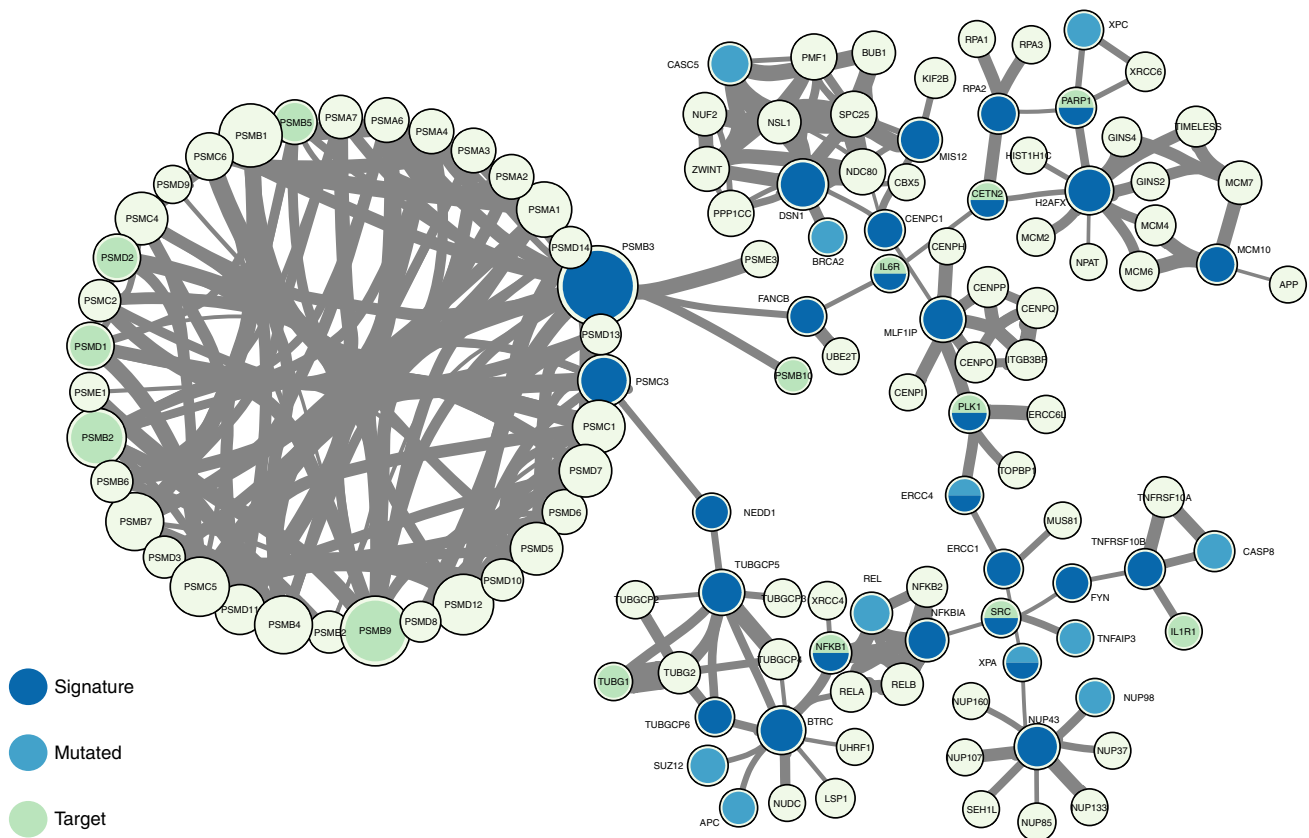
We used Spectral Centrality (SC)<sup>23</sup> as centrality measure, which is related to the changes in network global diffusivity by node perturbation through a Laplacian formalism. We applied SC in the largest connected components and considered as cluster-specific signatures the nodes with SC above the 90th percentile (25, 27 and 24 genes for clusters 1, 2, 3 respectively, Table 1 and Supplementary Data 6). As the BioPlex-Ontocancro network was differentially filtered by each cluster-specific correlation matrix, we compared the resulting signatures with the most central genes in the network prior to the filtering by any correlation matrix, and observed small overlaps between them (3/25, 13/27 and 4/24 common genes for clusters 1, 2, 3, respectively) showing how the added information of gene expression profiling correlation contributes in the creation of networks that are highly specific for the tumours in each cluster (Supplementary Table 12). The top-ranking nodes also differ significantly from those obtained from other centrality measures such as degree and betweenness centrality (see Supplementary Table 3) and even if some signature genes overlap between clusters, their links are different (Fig. 2,

**Table 1 Signature genes. List of signature genes for the three tumour clusters**

	Cluster 1	Cluster 2	Cluster 3
Spectral centrality > 90th percentile	<i>ALOX5</i>	<i>BTRC</i>	<i>AKT2</i>
	<i>BTRC</i>	<i>CENPC1</i>	<i>ALOX5</i>
	<i>BUB1</i>	<i>CETN2</i>	<i>BAG4</i>
	<i>CDC20</i>	<i>DSN1</i>	<i>CAPN1</i>
	<i>CENPC1</i>	<i>ERCC1</i>	<i>CAPN2</i>
	<i>CHUK</i>	<i>ERCC4</i>	<i>CDC16</i>
	<i>CUL1</i>	<i>FANCB</i>	<i>CDC27</i>
	<i>MIS12</i>	<i>FYN</i>	<i>CDT1</i>
	<i>MLF1IP</i>	<i>H2AFX</i>	<i>ENDOG</i>
	<i>NDC80</i>	<i>IL6R</i>	<i>FBXW11</i>
	<i>NFKB1</i>	<i>MCM10</i>	<i>FNTA</i>
	<i>NFKB2</i>	<i>MIS12</i>	<i>GMNN</i>
	<i>NFKBIA</i>	<i>MLF1IP</i>	<i>KIF2B</i>
	<i>PMF1</i>	<i>NEDD1</i>	<i>KIF2C</i>
	<i>PPP2CB</i>	<i>NFKB1</i>	<i>LSP1</i>
	<i>PPP2R5D</i>	<i>NFKBIA</i>	<i>NEDD1</i>
	<i>PSMB9</i>	<i>NUP43</i>	<i>PRKACG</i>
	<i>PSMC2</i>	<i>PARP1</i>	<i>PSMC3</i>
	<i>PSMF1</i>	<i>PLK1</i>	<i>PSMD9</i>
	<i>RAD21</i>	<i>PSMB3</i>	<i>SKP2</i>
	<i>REL</i>	<i>PSMC3</i>	<i>TNFRSF1A</i>
	<i>RELB</i>	<i>RPA2</i>	<i>TUBGCP5</i>
	<i>RPS27</i>	<i>SRC</i>	<i>UBB</i>
	<i>SRC</i>	<i>TNFRSF10B</i>	<i>VIM</i>
	<i>STAG1</i>	<i>TUBGCP6</i>	
		<i>XPA</i>	

Supplementary Figures 5–6, Supplementary Note 3, Supplementary Figure 9), evidencing specific interaction patterns. We observed an overlap (>50%) of our signatures with signatures obtained with the same procedure when protein–protein interactions from different experiments<sup>24,25</sup> (i.e., yeast-two hybrid) were added to the BioPlex-Ontocancro network, supporting the robustness and biological relevance of the observed signatures (Supplementary Note 4, Supplementary Tables 4–5).

The signatures contain genes related to three biological categories: NF-κB signaling, ubiquitin-proteasome system and chromosomal instability, the last category referring to genes involved in kinetochore formation, microtubule dynamics and chromosome segregation functions (Table 2 and Supplementary Data 2 for the list of Gene Ontology enriched terms). The signatures have at least one substrate recognition component of E3 ubiquitin ligase complexes: *BTRC* in clusters 1 and 2; and *FBXW11* in cluster 3. Cluster 1 has genes involved in spindle checkpoint (*BUB1*, *CDC20*). The cluster 2 signature has many genes related to DNA repair (*CETN2*, *FANCB*, *H2AFX*, *ERCC1*, *ERCC4*, *PARP1*, *XPA*) and DNA replication (*RPA2*, *MCM10*). Moreover, it has three important genes in the signaling path that activates the *STAT3* transcription factor: *SRC*, *NFKB1* and *IL6R*. Indeed, the *STAT3* gene expression levels are significantly higher in cluster 2 (one-way ANOVA *p*-value:  $5.58 \times 10^{-15}$ ) both in comparison with cluster 1 and cluster 3 (one-way ANOVA Student's *t* test *p*-values  $1.08 \times 10^{-9}$  and  $1.14 \times 10^{-8}$ , Supplementary Figure 10). The cluster 3 signature contains genes involved in three different apoptotic mechanisms: induced by TNF-α (*TNFRSF1A* and *BAG4*), induced by endoplasmatic reticulum stress (*CAPN1* and *CAPN2*) and caspase-independent apoptosis (*ENDOG*).



**Fig. 2** Network context of cluster 2 signature genes. Network composed by the first neighbours of cluster 2 signature genes in the BioPlex-Ontocancro network. Node sizes are proportional to their degree in the network and edge thickness is proportional to the normalized (CLR) co-expression between genes

**Table 2 Common biological categories in gene signatures**

	<b>NF-κB signaling</b>	<b>Chromosomal instability</b>	<b>Ubiquitin-proteasome system</b>
Cluster 1	<i>BTRC, CUL1, SRC, NFKBIA, NFKB1, NFKB2, REL, RELB, CHUK</i>	<i>CDC20, BUB1, MLFPIP, CENPC1, MIS12, PMF1, NDC80, RAD21, STAG1</i>	<i>BTRC, CUL1, PSMB9, PSMC2, PSMF1</i>
Cluster 2	<i>BTRC, SRC, NFKBIA, TNFRSF10B, IL6R</i>	<i>MIS12, DSN1, MLFPIP, CENPC1, PLK1, NEDD1, TUBGCP5, TUBGCP6</i>	<i>BTRC, PSMB3, PSMC3</i>
Cluster 3	<i>FBXW11, AKT2, TNFR1A</i>	<i>CDC16, CDC27, NEDD1, TUBGCP5, KIF2B, KIF2C</i>	<i>FBXW11, PSMC3, PSMD9</i>

All signatures have genes that can be grouped into the following categories: NF-κB signaling, chromosomal instability and ubiquitin-proteasome system

Then, we searched for possible relationships between the signature genes and genes commonly mutated across patients (Supplementary Note 4, Supplementary Tables 4–5). We observed that some signature genes also present somatic mutations (*REL* and *RAD21* in cluster 1, *ERCC4* and *XPA* in cluster 2, and *AKT2* in cluster 3) or that mutated genes are direct neighbours of the signature genes in the network (Fig. 2, Supplementary Figures 2–6). A permutation test over the signature labels reveals a significant proximity of signature genes to mutated genes for cluster 1 and cluster 2 (random permutation test  $p$ -value =  $8.76 \times 10^{-4}$  and  $p$ -value =  $6.9 \times 10^{-3}$  respectively, Supplementary Figure 11). For the particular case of cluster 3, only one mutated gene is present in the signature neighbourhood and it is successfully selected as a signature gene (Fig. 2). These outcomes highlight the strict relationship between signature genes and key processes in tumour development (in analogy with the network-based approach of Novarino et al.<sup>26</sup>).

Since the signature genes are the most central nodes in the three tumour cluster networks, we hypothesized that they might be suitable drug targets. For this purpose, we collected the drugs that target genes in the signatures (Supplementary Data 4) and we evaluated in the ClinicalTrials repository if these drugs were under ongoing clinical trials for cancer treatment. We observed that 11 genes from the cluster signatures are being tested: four and three genes, from cluster 1 and 2, respectively; three genes from both cluster 1 and 2; and one gene from both cluster 1 and 3 (Table 3, Supplementary Data 5). Interestingly, by considering all drugs that target genes in the BioPlex-Ontocancro network, we observed a significant overrepresentation of drugs that target genes in cluster 2 gene signature (Supplementary Note 8, Supplementary Table 9).

**Gene signatures differentiate survival outcome.** We then asked whether the expression levels of the signature genes could predict patient survival in each cluster, independently of the tumour type. For cluster 1 and 3, survival information was available only for 17 and 32 patients, respectively, which resulted in non-significantly different survival curves, possibly due to the low power of the test (Supplementary Note 9, Supplementary Figures 14–15). For cluster 2, we retrieved the clinical information for 1884 patients: the survival curves showed that the 27-gene signature significantly separated the patients in two groups according to good or bad survival outcome (Log-rank test  $p$ -value =  $7.26 \times 10^{-18}$ , Fig. 3a). We tested the significance of this separation against randomly generated signatures of the same size. Since one of the main factors affecting signature performance was the imbalance of samples from the same tumour in the two groups, we introduced a measure of sample imbalance to be considered together with the  $p$ -value (see Supplementary Figure 16). Our results show that our gene signature is quite balanced, and that it outperforms random signatures with similar extent of sample imbalance (Supplementary Figure 16). Although confounding factors may affect the performance, a multivariate analysis that considered patient age and tumour type as covariates, confirmed that the

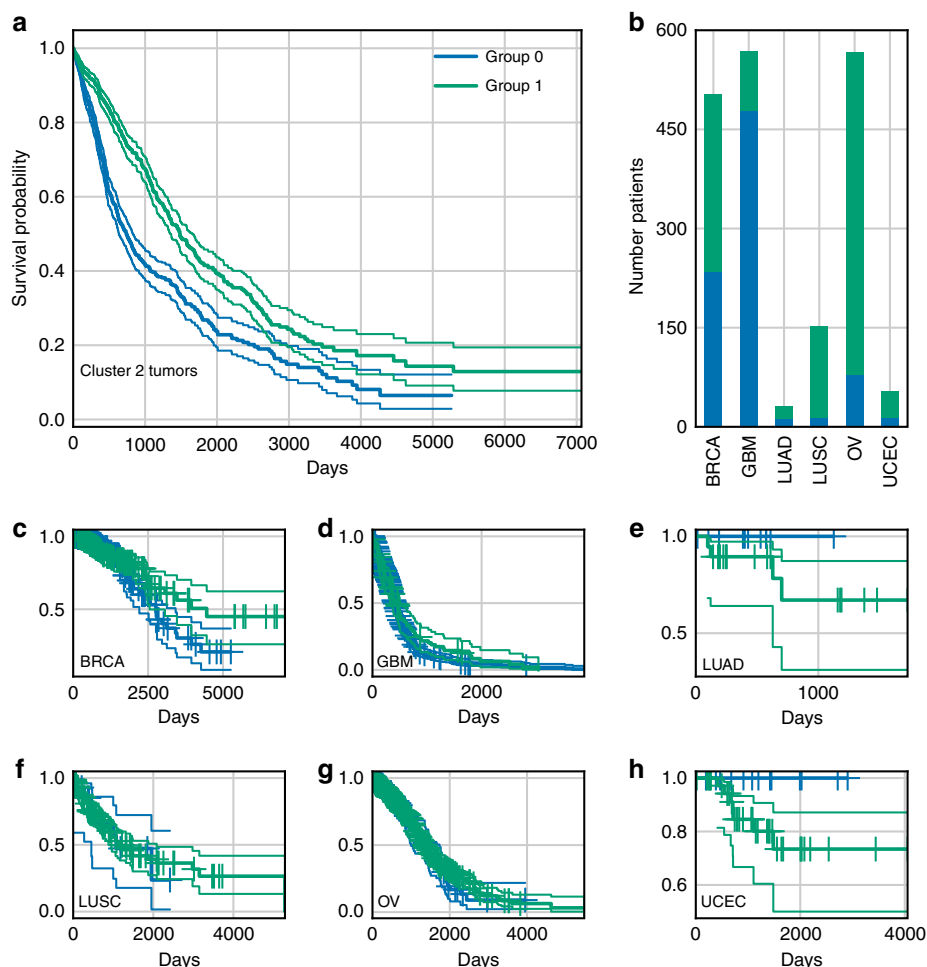
**Table 3 Clinical trials**

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

List of signature genes that are also being tested in ongoing clinical trials (according to AACT database)

separation of samples due to the gene signature played the major role in the observed difference of survival outcomes (Supplementary Data 1). When we further stratified the analysis at a single tumour level (considering its stage or grade when available) the separation resulted significant only for a set of tumours, with differences possibly due also to the different number of samples in each stage or grade considered (see Supplementary Data 1).

**Experimental validation of cluster 2 signature.** We tried to translate our results into novel therapeutic strategies by applying, for a subset of tumours in cluster 2 (which contained the largest and most heterogeneous set of tumours), a set of drugs on targets taken from the signature or from related biological categories. We selected three drugs: BI6727, for targeting the cluster 2 signature gene *PLK1*; Bortezomib, for targeting proteasome and NF-κB pathway; and the PF-00477736 drug, to target the *CHK1/2* genes, which play a role in DNA damage response but are not in the signature. The choice was based on signature gene list (considered as equally relevant, since many factors could affect exact SC node ranking), drug availability and required experimental settings, in order to avoid potential culture biases (e.g. the need of adding specific compounds to the medium or co-culturing with other cells to test some drugs). We tested these drugs, alone or in combination, on T98G glioblastoma cell line (modelling a clinically very aggressive tumour) and MCF-7 breast adenocarcinoma cell line (modelling a very common tumour type, see Methods for details on the experiments). Both cell lines resulted highly sensitive to Bortezomib, with an IC50 of 200 nM for MCF-7 and 0.6 nM for T98G (Fig. 4). BI6727 treatment reduced viability in a concentration-dependent manner in both models, with the glioblastoma model showing increased responsiveness (IC50 of 69.2 nM versus 1.8 μM for MCF-7). Moreover, both cell lines showed low response to PF-00477736, with IC50 of 26.9 μM for MCF-7 and 15.1 μM for T98G. We then asked whether these

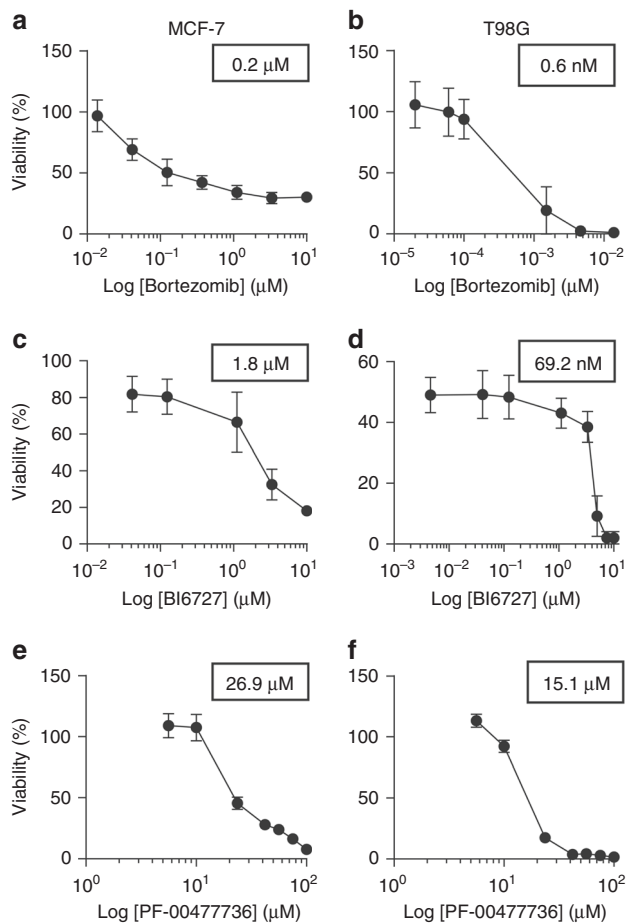


**Fig. 3** Gene signature and survival outcome. Kaplan–Meier survival curves for two groups of patients obtained via k-mean clustering of expression levels of cluster 2 signature genes. **a** Plot of the survival curves for worse (group 0) and better (group 1) outcome patient groups of all cluster 2 tumours, with a 95% confidence interval (group 0: 790 observations, 362 censored; group 1: 1094 observations, 648 censored; log-rank test  $p$ -value:  $7.26 \times 10^{-18}$ ). **b** Distribution of group 0 and group 1 patients inside cluster 2 tumours, showing a fairly homogeneous distribution of samples (see Supplementary Table 10 for further discussion). **c–h** Kaplan–Meier survival curves for groups 0 and 1 considering each tumour separately

drugs might synergize in the selected models. Although the combinations of PF-00477736 with either BI6727 or Bortezomib did not show any additive or synergistic effect in both cell lines, we observed a cooperation effect between BI6727 and Bortezomib (Fig. 5a, b). Indeed, we observed that cell viability was significantly lower compared with single agent treatments in MCF-7 cells (Fig. 5a, two-sided Student's  $t$  test  $p$ -value  $< 0.05$ ), showing a general additive effect (Supplementary Data 7). We observed low Combination Index values ( $< 1$ ) for both cell lines, indicating synergistic effect for all concentrations tested in MCF-7, and for selected concentrations in TG98 (Fig. 5, Supplementary Data 7).

Moreover, we integrated our experimental validation results with a wider set of experimental outcomes available from the Genomics of Drug Sensitivity in Cancer (GDSC) project<sup>27</sup>, which reports the drug screening experiments for 224,510 drug-cell line pairs (265 drugs, 1074 cell lines) (Supplementary Note 10). We observed that the IC<sub>50</sub> values obtained in our experimental validation are lower than the majority of those reported in the GDSC project, being in the 1st, 6th, 10th, and 24th percentiles of the IC<sub>50</sub> distribution for all drug-cell lines pairs (Supplementary Figure 17) and in the 1st, 9th, 7th, and 22nd percentiles of the IC<sub>50</sub> distribution of all drugs specifically tested on MCF-7 and T98G cell lines (Supplementary Figure 18).

**Cell line sensitivity to drugs targeting signature genes.** We also asked whether signature genes in each cluster could predict a better response of cell lines to the related drugs when compared with all drugs tested in GDSC. We identified 56, 218, and 49 cell lines associated to the tumour types grouped in cluster 1, 2, and 3, respectively (according to TCGA classification, Supplementary Data 8). We identified 7, 15, and 7 drugs targeting genes belonging to signature 1, 2, and 3, respectively. We observed that most of the cell lines in each cluster (51/53, 103/218, and 47/49, respectively) were more sensitive to drugs targeting signature genes as compared with all other drugs (according to IC<sub>50</sub> values, Supplementary Figures 19–21). Specifically, we found that 4, 18, and 10 cell lines in clusters 1, 2, and 3, respectively, showed significant differences (two-sided Student's  $t$  test  $p$ -value  $< 0.05$ , Fig. 6), all of them displaying lower IC<sub>50</sub> value when treated with drugs targeting signature genes. Remarkably, no significant differences with the opposite trend (higher IC<sub>50</sub> values) were found. The trend remained unaltered when the IC<sub>50</sub> values were compared between drugs targeting signature genes versus drugs targeting non-signature genes that belonged to the BioPlex-Ontocancro network: 50/53, 63/218, and 40/49 cell lines presented lower IC<sub>50</sub> values and for 2, 6, and 4 cell lines in clusters 1, 2, and 3, respectively, the difference



**Fig. 4** Cluster 2 cell lines response to Bortezomib, BI6727 and PF-00477736. MCF-7 and T98G cells were treated with increasing doses of Bortezomib (0.01–10 μM for MCF-7 (a), 0.02–10 nM for T98G (b)), BI6727 (0.04–10 μM for MCF-7 (c), 0.004–10 μM for T98G (d)), PF-00477736 (5.6–100 μM (e, f)) and cell viability was measured 72 h after drug administration by WST-1 assay (three independent experiments). Cell viability is represented as mean (error bars: s.e.m.). IC50 values are shown in the boxes

was statistically significant (two-sided Student's *t* test *p*-value < 0.05, Supplementary Figure 22).

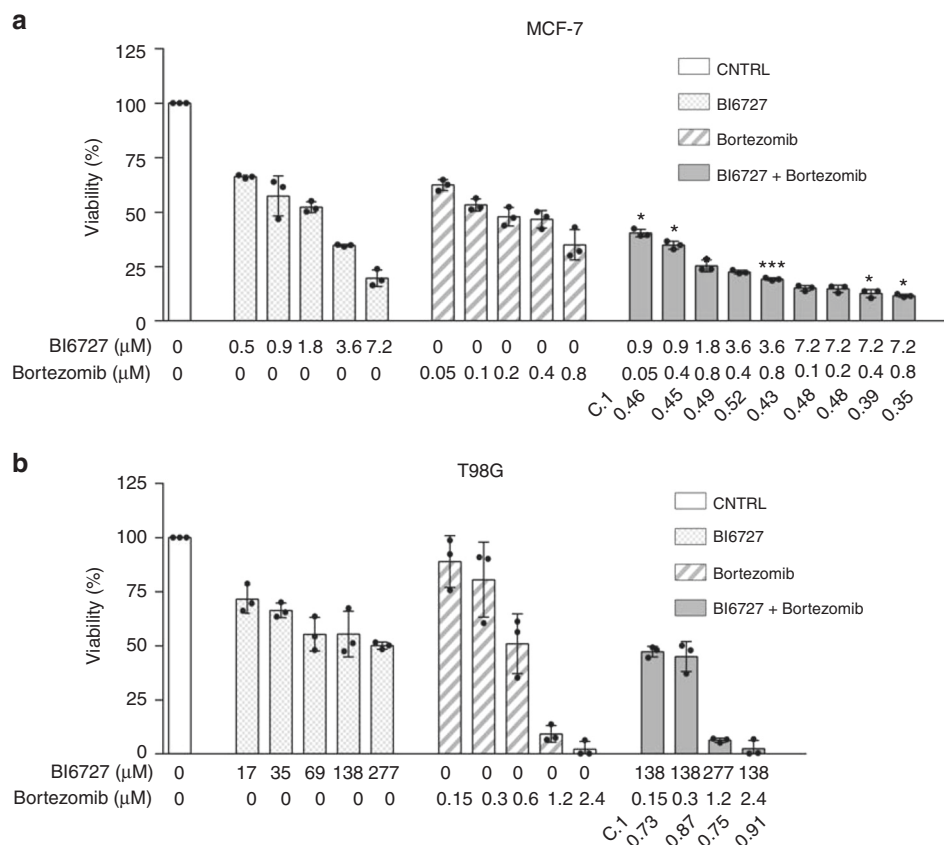
## Discussion

We studied the expression profiles of 11 tumours by considering a selected set of genes annotated in the Ontocancro database and BioPlex PPI network. This knowledge-based selection reduced the dimensionality of the data to a highly curated list of cancer-related genes, involved in pathways that are hallmarks of cancer as cell cycle, inflammation, and apoptosis<sup>28</sup>. This approach also ensured that all studied genes had protein–protein interaction annotations, which are crucial to the understanding of how the signaling transduction propagates in the cell<sup>29</sup>. We clustered tumours by their gene–gene correlation matrices, to evaluate the functional relationships between genes and their impact on transcriptome organization. Many tumours from the same organ grouped together, in agreement with previous studies showing that tissue-of-origin features provide the dominant signals in the identification of cancer subtypes<sup>2,30</sup>. However, the clustering also grouped tumours originating from different tissues, according to similarities in genomic alterations, as in the case of BRCA, OV, LUSC, and UCEC, which share common characteristics as presence of *TP53* mutations and multiple recurrent chromosomal gains and

losses<sup>3</sup>. In particular, BRCA and UCEC grouped into a well-defined subcluster, which may reflect their better prognosis when compared to other ten tumour types<sup>2</sup>. Interestingly, the tumours LGG, KIRC and KIRP clustered in the same group, suggesting that these tumours might activate the same pathways and oncogenic processes. Although Hoadley et al.<sup>2</sup> did not evaluate the tumours LGG and KIRP, the authors observed that glioblastoma GBM and kidney cancer KIRC clustered in the same group when using gene expression and copy number alteration profiles.

We integrated different types of biological information by a network approach, that allowed us to identify functional modules and to rank genes as network elements<sup>21,22,31,32</sup>. We created a network for each cluster (starting from a common template of protein interactions and superimposing cluster-specific expression correlation profiles) and obtained specific gene signatures based on node ranking by centrality measures. Gene Ontology enrichment analysis resulted in pathways mainly associated to NF-κB signaling, chromosomal instability and ubiquitin-proteasome system. NF-κB signaling regulates genes that participate in cell proliferation, innate and adaptive immune responses, inflammation, cell migration, and apoptosis regulation processes. The aberrant activity of NF-κB may act as survival factor for transformed cells which would otherwise become senescent or apoptotic<sup>33</sup>. Chromosomal instability category involves kinetochore formation, microtubule dynamics and chromosome segregation functions. The dysfunction in these genes may cause cell inability to faithfully segregate chromosomes, generating genomic alterations as DNA mutation, chromosomal translocation, and gene amplification. The mutant genotypes may confer beneficial phenotypic traits to cancer cells, such as sustained proliferative signaling and resistance to cell death<sup>28</sup>. Two genes classified into this category have already been related to clinical practice: the prognostic marker *KIF2C*,<sup>34,35</sup> and the *BUB1* gene, whose expression correlates with poor clinical diagnosis<sup>36,37</sup>. The ubiquitin-proteasome system is the major degradation machinery that controls the abundance of critical regulatory proteins. Perturbation of the regulatory proteins turnover disturbs the intricate balance of signaling pathways and the cellular homeostasis, contributing to the multi-step process of malignant transformation<sup>38</sup>. Proteasome inhibitors have become valuable tools in the treatment of certain types of cancer, mainly because malignant cells show greater sensitivity to the cytotoxic effects of proteasome inhibition than non-cancer cells<sup>39</sup>.

In addition to common features, cluster 2 signature has several genes related to DNA repair (*CETN2*, *FANCB*, *H2AFX*, *ERCC1*, *ERCC4*, *PARP1*, *XPA*) and DNA replication (*RPA2*, *MCM10*). Interestingly, the tumours in this cluster usually present high rates (50–90%) of samples with mutated *TP53*, which is an important sensor for the cell DNA damage response<sup>2,4,40</sup>. The cluster 2 signature also presents the *SRC*, *NFKB*, and *IL6R* genes, which participate in the activation of *STAT3*, a transcription factor that is necessary for cell transformation<sup>41</sup>. We observed that *STAT3* gene expression is higher in the tumours of cluster 2 when compared with the tumours of clusters 1 and 3 (one-way ANOVA *p*-value:  $5.58 \times 10^{-15}$ , see Supplementary Figure 10). The cluster 3 signature has genes involved in three apoptotic mechanisms, which are induced by TNF-α (*TNFRSF1A* and *BAG4*), or endoplasmic reticulum stress (*CAPN1* and *CAPN2*) and caspase-independent apoptosis (*ENDO1*). As the regulation of cell death serves as a natural barrier to cancer development, these processes may reflect different strategies that these tumours use in response to various cellular stresses. We remark that different experimental approaches for protein–protein interaction (e.g. yeast two-hybrid and affinity-purification followed by mass spectrometry) usually present low overlap, mainly due to assay complementarity, and differences in sensitivity and search



**Fig. 5** Combined inhibition of PLK1 and proteasome in MCF-7 and T98G cells. MCF-7 (**a**) and T98G (**b**) cells were treated with increasing doses of Bortezomib (0.05–0.8 μM for MCF-7, 0.15–2.4 nM for T98G) and BI6727 (0.5–7.2 μM for MCF-7, 17–277 nM for T98G), alone or in combination, and cell viability was measured 72 h after drug administration by WST-1 assay (three independent experiments). Cell viability is represented as mean (error bars: s. d.). Statistical significance was determined by two-sided Student’s *t* test (\**P* < 0.05; \*\*\**P* < 0.001). C.I. Combination Index (CompuSyn software). **a** MCF-7 cells: combinations with a C.I. lower than 0.5 are shown. **b** T98G cells: combinations showing synergistic effect are shown

space<sup>42</sup>. Consequently, the found gene signatures might be dependent on the protein network used in this study, despite the fact that we observed a >50% overlap in all tumour clusters when protein–protein interactions obtained from different experimental approaches were included in the BioPlex–Ontocancro network to generate different signatures (Supplementary Table 4).

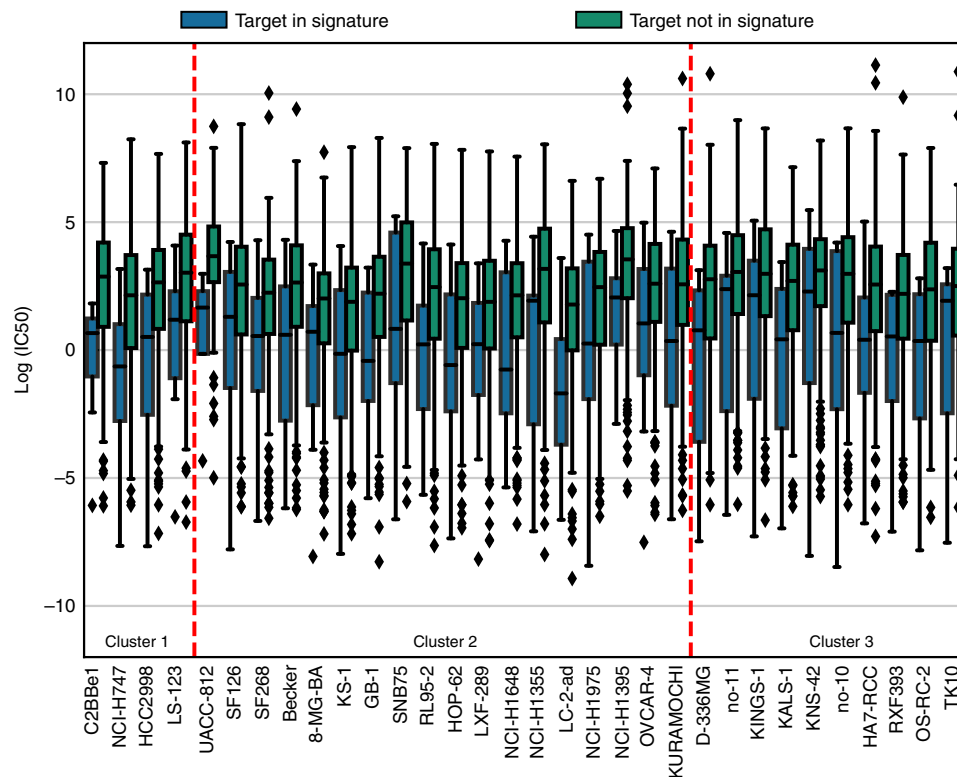
Since the transcriptional disturbances observed in cancer can sometimes be explained by underlying somatic mutations<sup>43,44</sup> we retrieved TCGA mutational data, and focused on cancer-related mutations reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. Many signature genes resulted also somatically mutated, or first neighbours to mutated genes (Fig. 2; Supplementary Figures 5–6, Supplementary Figures 11–13, and Supplementary Table 8), showing their strict relationship and the functional relevance of the biological processes they are involved in.

In addition, several genes in the signatures or in their direct network neighbourhood are already under clinical investigation in a variety of tumour conditions (as annotated in Clinicaltrials.org database). For example, the *AKT* pathway has been described as a potential drug intervention in clear cell renal carcinoma:<sup>45</sup> *AKT2* gene belongs to the signature of cluster 3 (comprising LGG, KIRC, and KIRP), it is somatically mutated in the tumours of cluster 3 and it has been annotated as drug-target according to the DrugBank database.

We asked whether gene signatures could predict survival outcomes in each cluster, thus independently on the single tumour type. Our results show that in cluster 2 (the only one with enough available samples) the 27-gene signature defined two groups of patients with significantly different Kaplan–Meier survival curves,

also in comparison with randomly generated signatures (log-rank test *p*-value:  $7.26 \times 10^{-18}$ , see Supplementary Figure 16). However, we remark that the clinical meaning of this separation requires further investigation, considering the high correlation of gene expression profiles in tumours<sup>46,47</sup> and since differences exist at the level of single tumour due to their stage or grade (see Fig. 3, Supplementary Table 11, Supplementary Data 1). We tested three existing drugs (two targeting elements of cluster 2 signature, and one involved in a related biological process, but not directly belonging to the signature) on two tumour types of the cluster, a very aggressive one (GBM, T98G cell line) and a common one (breast cancer, MCF-7 model). PF-0047736 drug (a *CHK1/2* inhibitor, not in the signature)<sup>48</sup> had poor effect on both cell lines. On the contrary, they resulted highly sensitive to the combination of BI6727 (an inhibitor of the signature gene *PLK1*<sup>49</sup>) and Bortezomib (proteasome activity inhibitor<sup>50</sup>). A number of tumour models is sensitive to proteasome and *PLK1* inhibitors. Indeed, genes belonging to the ubiquitin–proteasome system can be found in all clusters and *PLK1* expression is higher in cluster 2 tumours compared to the others. However, many cell lines have been reported throughout literature to be insensitive to these drugs<sup>51–56</sup>. Moreover, we observed a significant synergic action of Bortezomib and *PLK1* inhibitors at several dosages on both cell lines, not found when BI6727 was considered in combination with the other two signature-related drugs.

Finally, we performed *in silico* experiments on a broader range of cell lines and drugs by exploiting the drug screening data of the GDSC project. We observed that our signature genes predict high sensitivity to the related targeted agents on most cell lines in



**Fig. 6** Sensitivity to drugs targeting signature genes. We retrieved drug screening data from the GDSC project and considered the cell lines associated to the tumours in our clusters. The significance was evaluated with Student's *t* test (two-sided *p*-value < 0.05). The continuous horizontal line is the median, the lower and upper boundary represents the 25th and 75th percentiles, respectively. Whiskers extend to data points that lie within 1.5 interquartile ranges of the lower and upper quartiles; and observations that fall outside this range are displayed independently. Blue boxplots: IC50 values for signature genes; green boxplots: IC50 values for all other genes in GDSC targeting the same cell line

GDSC database that could be associated to our tumour clusters. Taken together, these findings suggest novel potential therapeutic strategies to be further explored in preclinical models.

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

## Methods

**Gene expression datasets.** The gene expression datasets used in this study were retrieved from The Cancer Genome Atlas (TCGA, <https://gdc-portal.nci.nih.gov>) Data Portal, and included Agilent expression arrays of 2378 samples from 11 tumour types, with a different number of samples each (from 16 to 595, Supplementary Table 1). We selected for our analysis the genes from the BioPlex protein–protein interaction network<sup>19,20</sup> ( $n = 10,961$ ) that were also present in the Ontocancro database <http://ontocancro.inf.ufsm.br/>, specific for tumour-related biological processes ( $n = 1104$ ), resulting in a list of 760 cancer-associated genes related to specific biological functions (such as cell cycle control, DNA damage response, and inflammation).

**Tumour clustering.** For each tumour dataset, we calculated a correlation matrix containing pairwise Pearson  $r_{ij}$  coefficients between the previously selected 760 genes across all samples available for the tumour. In order to eliminate non-significant correlations and indirect influences, the absolute correlation values ( $|r_{ij}|$ ) were adjusted with the Context Likelihood of Relatedness (CLR) algorithm<sup>57,58</sup>

(implemented in the R/Bioconductor package 'minet'<sup>59</sup>) obtaining  $z_{ij}$  positive and negative scores. We considered only positive values ( $z_{ij} > 0$ ), but the obtained clusters resulted fairly robust when different threshold values were applied (Supplementary Note 2, Supplementary Figures 7–8). The filtered  $z$ -score matrices were clustered using hierarchical clustering analysis (with Ward linkage) based on the element-wise Euclidean distance between each pair of tumour matrices A and B, calculated as follows:

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

where  $a_{ij}$  is the  $z$ -score between the genes  $i$  and  $j$  in Tumour A and  $b_{ij}$  the  $z$ -score between the same genes in Tumour B. The clusters were defined using a dynamic branch cutting algorithm implemented in the R package `dynamicTreeCut`<sup>60</sup> (parameters: `minClusterSize = 2`, `cutHeight = 590` and `method = 'Hybrid'`).

**Multi-tumour gene signatures.** A network approach was applied to find gene signatures that characterized the clusters of tumours. First, we created a template network (BioPlex-Ontocancro PPI) by selecting the genes present in the BioPlex protein–protein interaction network that were also annotated in the Ontocancro database. In the initial list of 760 genes used for tumour clustering, only 591 were connected to each other (169 isolated nodes) thus considered for network analysis. Then, for each cluster the gene–gene correlation coefficients  $r_{ij}$  were computed, considering all the samples of all tumours in the cluster, and their absolute values  $|r_{ij}|$  were adjusted with the CLR algorithm, producing the  $z_{ij}$  scores<sup>57,58</sup>. Each score matrix was superimposed to the BioPlex-Ontocancro network, producing three weighted networks (one for each cluster) in which genes were linked only if having correlated expression profiles ( $z_{ij} > 0$ , specific to each cluster) and a physical interaction at protein level (given by Bioplex-Ontocancro network, common to all clusters). We remark that the three cluster-related networks can differ for their weight values or for missing links (due to negative  $z$ -scores set to zero). The networks were analyzed and visualized by Networkx Python package, Matlab and Cytoscape.

For the networks of clusters 1 and 3, we selected the giant components (245 and 244 nodes, respectively), and for cluster 2 we selected the two largest components, since they had similar size much larger than the other components (149 and 118 nodes) and since together resulted in a number of nodes comparable to the giant



component of cluster 1 and 3 (Supplementary Data 3). After this selection, we retrieved a gene signature for each cluster composed by the most central genes (network nodes), which were defined as those having the Spectral Centrality<sup>23</sup> topological measure (SC) above the 90th percentile (see Table 1, Supplementary Figures 2–6, Supplementary Data 6 with genes sorted by SC). SC calculates the effect of node removal on the network diffusivity based on the spectral properties of the Laplacian graph, and it has already been applied successfully to biological data such as the Immune System mediator network. Different results were obtained by considering Betweenness Centrality or weighted degree (Strength *W*) as centrality measures, as shown in Supplementary Table 3. The robustness of our signatures was tested by considering a random subsampling (50% of the initial number of patients in each cluster) for 100 times, and then counting the number of occurrences of original signature genes into the signatures obtained by applying the analysis pipeline onto the patient subsampling (see Supplementary Table 6).

**Validation of the multi-tumour gene signatures.** We evaluated the relevance of the genes in the signatures by several approaches.

First, we verified the proximity with the somatic mutational data extracted from the TCGA data portal for the considered tumours. To avoid cancer unrelated mutations, we considered only mutations that were reported also in the COSMIC database<sup>61</sup>. We checked whether the signature genes had been reported as somatically mutated or if they occurred in the neighbourhood of mutated genes in the networks. To quantify the proximity of gene signatures to mutated genes we located the nearest mutation (in terms of shortest paths on the network) for each signature gene, resulting in a collection of minimum distance values for each cluster, and we compared these distances with random gene signatures of the same size (Supplementary Figure 11–13, Supplementary Table 8).

Secondly, we retrieved from the DrugBank <http://www.drugbank.ca/> and Drug Gene Interaction (DGIdb)<sup>62</sup> databases which genes in the signatures were also mapped as drug targets. Third, we checked in the Aggregate Analysis of Clinical Trials (AACT) database (<https://www.ctti-clinicaltrials.org/aact-database>) for the existence of ongoing clinical trials evaluating the inhibition of signature genes. Fourth, the prognostic potential of each gene signature was evaluated by considering the clinical data (days to death) available in the TCGA data portal. The patients having clinical information were clustered according to the expression levels of the gene signatures by a k-means algorithm (Python package ‘scikit’) considering two patient groups: good versus bad survival outcome. Survival curves were calculated for both groups: we applied the Kaplan–Meier method with censored data and evaluated their significance with the log-rank test (Python package ‘lifelines’). We observed that the effect of the gene signature on dividing patients remained prevailing even when considering confounding factors such age and tumour type, and some differences emerged only when considering tumour stage or grade (see Supplementary Table 11, Supplementary Data 1). Moreover, we compared our signature with randomly generated signatures, by using two indices: *P*-value and *S<sub>T</sub>*, a parameter indicating the imbalance of tumour samples inside both patient groups (details and results in Supplementary Figure 16). Fifth, we tested the effect of drugs inhibiting genes in our signatures or strictly related to them. The glioblastoma T98G and the breast adenocarcinoma MCF-7 cell lines were obtained from ATCC® ([lglstandards-atcc.org/](http://lglstandards-atcc.org/)) and Leibniz Institut DSMZ ([dsmz.de](http://dsmz.de)), respectively. Cells were cultured at a density of 10<sup>5</sup> cells/ml in RPMI medium plus 10% FBS (plus 5% Sodium orthovanadate for T98G) for 72 h with increasing concentrations of the following drugs: Bortezomib, BI6727, PF-00477736 (Selleckchem), alone or in combination. The dose range was chosen based on insights from the literature on the same cell lines or related cellular models<sup>49,50,63–70</sup>. One hour and thirty minutes before the end of treatment, WST-1 reagent was added to the cell medium and cell viability was measured according to manufacturer’s instruction (Roche). The cell response to different drug dosages was tested in three independent experiments by duplicates of each experimental condition. The dose-effect response and the IC50 of each drug were calculated using GraphPad Prism 6 (GraphPad Software) with a 95% confidence interval. To determine synergy, combination indexes were obtained with the CompuSyn software (ComboSyn Inc.): combination index values <1, =1, and >1 indicate synergism, additive effect and antagonism, respectively (Supplementary Data 7).

Finally, we also validated the signature genes as potential drug targets by using drug screening data from the Genomics of Drug Sensitivity in Cancer (GDSC) project (<http://www.cancerrxgene.org/>). The drug–gene interactions were retrieved both from GDSC project and DGIdb, considering only drugs with known mechanistic actions and approved by the FDA. We associated cell lines in the GDSC project to our cluster according to TCGA tumour classification (Supplementary Figure 17–22, Supplementary Data 8).

**Code availability.** The code for the calculation of the Spectral Centrality (SC) measure is available in <https://zenodo.org/record/1420602#.W6Bju5NKbW> [<https://doi.org/10.5281/zenodo.1420602>]. Codes used for the other analyses in this study are available from the corresponding author upon reasonable request.

## Data availability

The gene expression datasets used in this study were retrieved from The Cancer Genome Atlas (TCGA, <https://gdc-portal.nci.nih.gov>) Data Portal. Other data

supporting the findings of this study are available within the paper and its supplementary information files.

Received: 13 February 2017 Accepted: 4 October 2018

Published online: 30 October 2018

## References

1. Cancer Genome Atlas Research Network. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
2. Hoadley, K.A. et al. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell* **158**, 929–944 (2013).
3. Ciriello, G. et al. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
4. Kandath, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
5. Koboldt, D. C. et al. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
6. Network, C. G. A. R. et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
7. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
8. Garnett, M. J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
9. Heiser, L. M. et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc. Natl Acad. Sci. USA* **109**, 2724–2729 (2012).
10. Creixell, P. et al. Pathway and network analysis of cancer genomes. *Nat. Methods* **12**, 615–621 (2015).
11. Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinforma.* **7** (Suppl 1), S7 (2006).
12. Lan, A. et al. ResponseNet: revealing signaling and regulatory networks linking genetic and transcriptomic screening data. *Nucleic Acids Res.* **39**, 424–429 (2011).
13. Vandin, F., Clay, P., Upfal, E. & Raphael, B. J. Discovery of mutated subnetworks associated with clinical data in cancer. *Pac. Symp. Biocomput.* 55–66 (2012).
14. Leiserson, M. D. M. et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114 (2015).
15. Bersanelli, M. et al. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinforma.* **17**, 15 (2016).
16. Drier, Y., Sheffer, M. & Domany, E. Pathway-based personalized analysis of cancer. *Proc. Natl Acad. Sci. USA* **110**, 6388–6393 (2013).
17. Greenblum, S. I. et al. The PathOlogist: an automated tool for pathway-centric analysis. *BMC Bioinforma.* **12**, 133 (2011).
18. Librelotto, G. R. et al. in *Advances in Bioinformatics and Computational Biology: 4th Brazilian Symposium on Bioinformatics, BSB 2009, Porto Alegre, Brazil, July 29–31, 2009. Proceedings* (eds Guimarães, K. S., Panchenko, A. & Przytycka, T. M.) 164–167 (Springer Berlin Heidelberg, 2009). [https://doi.org/10.1007/978-3-642-03223-3\\_18](https://doi.org/10.1007/978-3-642-03223-3_18)
19. Huttlin, E. L. et al. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425–440 (2015).
20. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
21. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333–408 (2013).
22. Guney, E., Menche, J., Vidal, M. & Barabási, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 10331 (2016).
23. Pauls, S. D. & Remondini, D. Measures of centrality based on the spectrum of the Laplacian. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **85**, 066127 (2012).
24. Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
25. Menche, J. et al. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
26. Novarino, G. et al. Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* **343**, 506–511 (2014).
27. Yang, W. et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2013).
28. Hanahan, D. & Weinberg, R. a. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
29. Vinayagam, A. et al. A directed protein interaction network for investigating intracellular signal transduction. *Sci. Signal.* **4**, rs8–rs8 (2011).

30. Martínez, E. et al. Comparison of gene expression patterns across 12 tumor types identifies a cancer supercluster characterized by TP53 mutations and cell cycle defects. *Oncogene* **34**, 2732–2740 (2015).
31. Ideker, T. & Sharan, R. Protein networks in diseases. *Genome Res.* **18**, 644–652 (2008).
32. Castellani, G. C. et al. Systems medicine of inflammaging. *Brief. Bioinform.* **17**, 527–540 (2016).
33. Hoessel, B. & Schmid, J. A. The complexity of NF- $\kappa$ B signaling in inflammation and cancer. *Mol. Cancer* **12**, 86 (2013).
34. Bie, L., Zhao, G., Wang, Y. & Zhang, B. Kinesin family member 2C (KIF2C/MCAK) is a novel marker for prognosis in human gliomas. *Clin. Neurol. Neurosurg.* **114**, 356–360 (2012).
35. Sanhaji, M., Friel, C. T., Wordeman, L., Louwen, F. & Yuan, J. Mitotic centromere-associated kinesin (MCAK): a potential cancer drug target. *Oncotarget* **2**, 935–947 (2011).
36. Bie, L. et al. The accuracy of survival time prediction for patients with glioma is improved by measuring mitotic spindle checkpoint gene expression. *PLoS ONE* **6**, e25631 (2011).
37. Finetti, P. et al. Sixteen-kinase gene expression identifies luminal breast cancers with poor prognosis. *Cancer Res.* **68**, 767–776 (2008).
38. Sahasrabudhe, A. A. & Elinitoba-Johnson, K. S. J. Role of the ubiquitin proteasome system in hematologic malignancies. *Immunol. Rev.* **263**, 224–239 (2015).
39. Saez, I. & Vilchez, D. The mechanistic links between proteasome activity, aging and age-related diseases. *Curr. Genom.* **15**, 38–51 (2014).
40. Collisson, E. A. et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
41. Iliopoulos, D., Hirsch, H. A. & Struhl, K. An epigenetic switch involving NF- $\kappa$ B, Lin28, Let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell* **139**, 693–706 (2009).
42. Luck, K., Sheynkman, G. M., Zhang, I. & Vidal, M. Proteome-scale human interactomics. *Trends Biochem. Sci.* **42**, 342–354 (2017).
43. Shlien, A. et al. Direct transcriptional consequences of somatic mutation in breast cancer. *Cell Rep.* **16**, 2032–2046 (2016).
44. Sørlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
45. Creighton, C. J. et al. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
46. Venet, D., Dumont, J. E. & Detours, V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput. Biol.* **7**, e1002240 (2011).
47. Cleary, B., Cong, L., Cheung, A., Lander, E. S. & Regev, A. Efficient generation of transcriptomic profiles by random composite measurements. *Cell* **171**, 1424–1436.e18 (2017).
48. Lee, C. et al. Polo-like kinase 1 inhibition kills glioblastoma multiforme brain tumor cells in part through loss of SOX2 and delays tumor progression in mice. *Stem Cells* **30**, 1064–1075 (2012).
49. Bhola, N. E. et al. Kinome-wide functional screen identifies role of PLK1 in hormone-independent, ER-positive breast cancer. *Cancer Res.* **75**, 405–414 (2015).
50. Thaler, S. et al. The proteasome inhibitor Bortezomib (Velcade) as potential inhibitor of estrogen receptor-positive breast cancer. *Int. J. Cancer* **137**, 686–697 (2015).
51. Rückrich, T. et al. Characterization of the ubiquitin-proteasome system in bortezomib-adapted cells. *Leukemia* **23**, 1098–1105 (2009).
52. Ri, M. et al. Bortezomib-resistant myeloma cell lines: a role for mutated PSMB5 in preventing the accumulation of unfolded proteins and fatal ER stress. *Leukemia* **24**, 1506–1512 (2010).
53. Franke, N. E. et al. Impaired bortezomib binding to mutant  $\beta$ 5 subunit of the proteasome is the underlying basis for bortezomib resistance in leukemia cells. *Leukemia* **26**, 757–768 (2012).
54. de Wilt, L. H. A. M. et al. Proteasome-based mechanisms of intrinsic and acquired bortezomib resistance in non-small cell lung cancer. *Biochem. Pharmacol.* **83**, 207–217 (2012).
55. Wu, Y.-X., Yang, J.-H. & Saitou, H. Bortezomib-resistance is associated with increased levels of proteasome subunits and apoptosis-avoidance. *Oncotarget* **7**, 77622–77634 (2016).
56. Nonomiya, Y. et al. Effect of AKT3 expression on MYC- and caspase-8-dependent apoptosis caused by polo-like kinase inhibitors in HCT 116 cells. *Cancer Sci.* **107**, 1877–1887 (2016).
57. Saccenti, E., Suarez-Diez, M., Luchinat, C., Santucci, C. & Tenori, L. Probabilistic networks of blood metabolites in healthy subjects as indicators of latent cardiovascular risk. *J. Proteome Res.* **14**, 1101–1111 (2015).
58. Faith, J. J. et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, 0054–0066 (2007).
59. Meyer, P. E., Lafitte, F. & Bontempi, G. minet: a R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinforma.* **9**, 461 (2008).
60. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
61. Forbes, S. A. et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
62. Wagner, A. H. et al. DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res.* **44**, D1036–D1044 (2016).
63. Vlachostergios, P. J. & Papandreou, C. N. Efficacy of low dose temozolomide in combination with bortezomib in U87 glioma cells: a flow cytometric analysis. *Arch. Med. Sci.* **11**, 307–310 (2015).
64. Pezuk, J. A. et al. Inhibition of polo-like kinase 1 induces cell cycle arrest and sensitizes glioblastoma cells to ionizing radiation. *Cancer Biother. Radiopharm.* **28**, 516–522 (2013).
65. Hideshima, H. et al. IKK $\beta$  inhibitor in combination with bortezomib induces cytotoxicity in breast cancer cells. *Int. J. Oncol.* **44**, 1171–1176 (2014).
66. Han, J., Ma, L., Hendzel, M. J. & Allalunis-Turner, J. The cytotoxicity of gamma-secretase inhibitor I to breast cancer cells is mediated by proteasome inhibition, not by gamma-secretase inhibition. *Breast Cancer Res.* **11**, R57 (2009).
67. Bryant, C., Rawlinson, R. & Massey, A. J. Chk1 inhibition as a novel therapeutic strategy for treating triple-negative breast and ovarian cancers. *Bmc. Cancer* **14**, 570 (2014).
68. Ma, Z. et al. The Chk1 inhibitor AZD7762 sensitizes p53 mutant breast cancer cells to radiation in vitro and in vivo. *Mol. Med. Rep.* **6**, 897–903 (2012).
69. Tang, Y., Dai, Y., Grant, S. & Dent, P. Enhancing CHK1 inhibitor lethality in glioblastoma. *Cancer Biol. Ther.* **13**, 379–388 (2012).
70. Yao, F. et al. An autophagy inhibitor enhances the inhibition of cell proliferation induced by a proteasome inhibitor in MCF-7 cells. *Mol. Med. Rep.* **5**, 84–88 (2012).

## Acknowledgements

This study was supported by the Interomics CNR Flagship Initiative; Mimomics EU FP7 Project no. 305280; NGS PTL EU FP7 Project no. 306242; the CNPq Project no. 402547/2012–8; the Associazione Italiana per la Ricerca sul Cancro AIRC (Investigator Grant to G. M., no. 19226); the Programma di ricerca Regione-Università 2010–2012 (L. Bolondi); and the Innovative Medicines Initiative (IMI) 2 project “HARMONY”, no. 116026. We thank the Science Without Borders project of CAPES foundation (Ministry of Education of Brazil - Brasília -DF, Brazil) for the doctoral scholarship (grant number: 10186–13–1) for IFV.

## Author contributions

I.F.V., G. Menichetti and G.S. contributed equally to the paper. G.S., S.B., I.Z. and D.F.D. performed the experiments and interpreted the results. I.F.V., G. Menichetti and J.C.M. collected and analyzed the data. D.R. supervised data analysis. G.C., G. Martinelli and D. R. designed the research, interpreted the results, wrote the paper.


## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-06992-7>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018