

A PanorOmic view of personal cancer genomes

Lidia Mateo^{1,†}, Oriol Guitart-Pla^{1,†}, Carles Pons¹, Miquel Duran-Frigola¹, Roberto Mosca¹ and Patrick Aloy^{1,2,*}

¹Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology, Barcelona 08028, Catalonia, Spain and ²Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Catalonia, Spain

Received January 30, 2017; Revised March 24, 2017; Editorial Decision April 05, 2017; Accepted April 12, 2017

ABSTRACT

The massive molecular profiling of thousands of cancer patients has led to the identification of many tumor type specific driver genes. However, only a few (or none) of them are present in each individual tumor and, to enable precision oncology, we need to interpret the alterations found in a single patient. Cancer PanorOmics (<http://panoromics.irbbarcelona.org>) is a web-based resource to contextualize genomic variations detected in a personal cancer genome within the body of clinical and scientific evidence available for 26 tumor types, offering complementary cohort- and patient-centric views. Additionally, it explores the cellular environment of mutations by mapping them on the human interactome and providing quasi-atomic structural details, whenever available. This ‘PanorOmic’ molecular view of individual tumors, together with the appropriate genetic counselling and medical advice, should contribute to the identification of actionable alterations ultimately guiding the clinical decision-making process.

INTRODUCTION

Large-scale cancer genomics studies (1,2) have shown that every personal cancer genome harbors thousands of genomic alterations—including somatic mutations, copy number alterations, gene expression changes and epigenetic modifications—that are not present in the patient’s germline. Of these, a very small subset might be ‘driver’ mutations, which confer a selective growth advantage through the enhancement of some biological capabilities like proliferation, angiogenesis, invasion or metastasis or through the evasion of homeostatic mechanisms such as growth suppression, cell cycle arrest, cell death or immune surveillance (3). However, the vast majority of genomic changes detected in cancer cells are just ‘passenger’ and do not confer oncogenic properties (4).

Distinguishing between driver and passenger mutations is a challenging task that has now become a matter of major interest, especially if we envision that the analysis of personal cancer genomes will become a common clinical practice. Current methods are based on the identification of signatures of positive selection at gene level that have been observed in large cohorts (i.e. identification of genes showing high somatic mutation rate, mutation functional impact bias and/or mutational clustering patterns that are more often observed in cancer than expected by chance) (5). These statistical approaches have permitted the identification of, for instance, over 200 driver genes for cutaneous melanoma or breast cancer (6). However, when focusing on individual tumors, it is difficult to find >2–4 driver genes mutated and, in some patients, none of the usual suspects seems to be altered (6). Moreover, the effect of distinct mutations in the same gene might be radically different (7–9), being it necessary to consider their exact location and molecular environment.

Cancer PanorOmics integrates and displays high-throughput cancer sequencing analyses, together with data obtained from individual patients. The user can upload a list of somatic mutations, copy number alterations and/or gene expression changes detected in a personal cancer genome, and choose a reference cohort among the 26 tumor types available. Then, these genomic alterations can be explored in the light of what is known about the reference cohort. Additionally, the server maps mutations on the high-resolution 3D structure of proteins and protein–protein interactions to provide a molecular context to the genomic alterations analyzed. Cancer PanorOmics is available at <http://panoromics.irbbarcelona.org>.

DATA

The current version of Cancer PanorOmics catalogs 2 335 564 mutations found in 17 613 gene products from 20 683 cancer patients, representing 26 tumor types. These data have been compiled from IntOGen (10) and COSMIC (11) databases and complemented with interactions and struc-

*To whom correspondence should be addressed. Tel: +34 93 403 96 90; Fax: +34 93 403 71 14; Email: patrick.aloy@irbbarcelona.org

†These authors contributed equally to this work as first authors.

Table 1. Cancer PanorOmics database content

Tumors—patients		Molecular Data	
Non small cell lung carcinoma	1515	Mutations	
Head and neck squamous cell carcinoma	785	Mutations (in total)	2 335 564
Stomach adenocarcinoma	705	Mutations in reference proteome	2 277 702
Esophageal carcinoma	1113	Mutations with structural annotation	731 531
Glioblastoma multiforme	884	- Surface mutations	598 640
Hepatocarcinoma	1891	- Buried mutations	132 916
Lower grade glioma	86	- Interface mutations	81 422
Renal clear cell carcinoma	850	Proteins	
Uterine corpus endometrioid carcinoma	628	Proteins in the human proteome	20 298
Medulloblastoma	459	Proteins affected by somatic mutations	17 613
Prostate adenocarcinoma	1263	Mutated proteins with structural data	10 934
Colorectal adenocarcinoma	1264	- Mutated proteins with complete structures	1669
Acute lymphoblastic leukemia	254	- Mutated proteins with complete models	2900
Thyroid carcinoma	666	- Mutated proteins with partial struct/models	6365
Difuse large B cell lymphoma	274	Mutated proteins w/o structural data	6679
Serous ovarian adenocarcinoma	637	Mutated proteins in the interactome	11 868
Cutaneous melanoma	901	Mut. proteins with struct. data in the interactome	8407
Chronic lymphocytic leukemia	807	Proteins affected by interaction mutations	5020
Breast carcinoma	2006	Interactions	
Lung adenocarcinoma	867	Interactions in the human binary interactome	66 452
Neuroblastoma	720	Interactions affected by somatic mutations	9257
Acute myeloid leukemia	773	Interactions with structural data	9875
Pancreatic adenocarcinoma	1386	- Interactions with exp. structures	5235
Bladder carcinoma	650	- Interactions with global models	3254
Lung squamous cell carcinoma	641	- Interactions with domain-domain models	1386
Any tumor type	20 683	Interactions w/o structural data	56 577

tural information from Interactome3D (12) and dSysMap (7).

We downloaded the COSMIC v79 mutation data file containing 3 301 941 unique coding point mutations identified in genome-wide screens. We excluded polymorphic SNPs and mutations identified in cell-lines, organoids or xenografts, which yielded a total of 2 758 163 unique coding point mutations. We transferred somatic mutations annotated by COSMIC from HGNC gene symbol to the coordinates in the corresponding protein Uniprot Accession Code(s) (Uniprot AC). When needed, we performed an alignment between the protein sequence deposited in COSMIC and the corresponding Uniprot AC sequence in order to determine the precise coordinates of the mutation in the protein Uniprot AC. We could successfully remap 2 277 702 unique mutations, out of which 731 531 could be structurally annotated and classified into ‘Surface’, ‘Buried’ or ‘Interface’ using Interactome3D and dSysMap databases. The chromosomal coordinates of the proteins were obtained from UCSC. Regarding the information about clinical actionability, we considered 1480 compounds from ChEMBL database (13) with annotated protein target(s) and clinical indications, according to Medical Subject Headings (MeSH), Anatomical Therapeutic Chemical classification (ATC) or Experimental Factor Ontology (EFO). We defined a sublist of cancer related drugs by collecting compounds that contained the keywords ‘cancer’, ‘carcinoma’ or ‘neoplasm’ in the MeSH headings or in the EFO term, or drugs classified as antineoplastic agents (‘L01’ ATC code and descendants). The list of druggable genes was obtained from the Supplementary material of (14).

Cancer PanorOmics will be updated every 6 months. Please, see Table 1, or the *Stats* page at the server, for more detailed information.

WORKFLOW

User input

The user can upload a list of somatic mutations, copy number and/or gene expression alterations detected in a personal cancer genome. It is important to note that the main identifier used by Cancer PanorOmics is Uniprot Accession code (Uniprot AC). The user can also provide the HGNC gene symbol or the Uniprot gene name and Cancer PanorOmics will be able to structurally annotate the somatic mutations after ensuring that the reference aminoacid of the mutation corresponds to the aminoacid encoded at the same position in the Uniprot AC sequence. Before submitting the query, the user should also choose one of the 26 tumor types available as a reference cohort. The server provides several examples to show the accepted formats of the input data.

Clinical context

Cancer PanorOmics offers a contextualized view of the genomic alterations uploaded by the user within the available knowledge for the selected tumor type. More specifically, the user can easily see which altered genes in a given patient are known drivers, and how often each protein is mutated or copy number altered in the reference cohort. The results are displayed in two complementary Patient and Cohort Centric Views, as shown in Figure 1.

Patient centric view. In the ‘Patient Centric View’, known cancer driver genes are mapped on top of all the genes that are mutated, amplified or deleted in the patient. Chromosomes are depicted as grey boxes in the outermost ring and each circular section represents an altered gene, sorted by chromosomal coordinates. Each gene section contains



Figure 1. Clinical Context of the alterations detected in a ductal breast carcinoma patient (TCGA-A2-A4RX-01). The outermost rings represent the chromosomal location of genes and the next three data tracks show information about gene expression, copy number variation or somatic mutation in each gene. In the Patient Centric View (left panel), all somatic mutations and/or copy number alterations detected in this patient are displayed. Additionally, known drivers for the selected tumor type (filled circles), together with other cancer drivers (empty circles), are mapped onto patient alterations. In the Cohort Centric View (right panel), all known tumor type specific drivers are displayed and the frequency of each driver alteration is represented using a color gradient (green represents somatic mutation frequency, whereas red/blue represent copy number gain/loss or gene over/underexpression frequency). The alterations detected in this patient (filled circles) are mapped onto the known drivers. PIK3R1 gene is highlighted in yellow.

three cells that correspond to gene expression, copy number and somatic mutations, which are the three concentric data tracks displayed. Known cancer driver genes are mapped on the patient's alterations using filled circles for tumor type specific drivers or empty circles for drivers in any other tumor type.

Cohort centric view. In this complementary view, patient alterations are mapped on top of known cancer driver genes. Chromosomes are again depicted as grey boxes but here each circular section represents a known cancer driver gene in the reference cohort. Each gene section is colored according to the frequency of each type of genomic alteration in the cohort. Gene expression and copy number alteration frequencies are represented in a red-white-blue scale in the first two data tracks, whereas somatic mutation frequency is represented in green in the inner data track. Patient alterations affecting known drivers are represented by black circles.

Molecular context

Cancer PanorOmics maps mutations on the high-resolution 3D structure of proteins and interactions. When the user selects a protein in the Patient Centric View or the Cohort Centric View, a new page is loaded to display the molecular context of the corresponding genomic alterations.

Protein-protein interaction network. The current human interactome contains 66 452 interactions (12). We zoom in

to the protein of interest and its neighbors, and represent on each of them the recurrence of genomic alterations in the reference cohort. Small black circles indicate the presence of a genomic alteration either on a protein or at the interaction interface between two proteins. Every node contains three concentric circles displaying the information available for three molecular data types: gene expression variations, copy number changes and somatic mutations. Additional details about each of the proteins describing their known or potential role in cancer, as well as information about their clinical actionability, are shown in the Molecular Context page, and the corresponding downloadable table.

Structural details. Whenever available, structural 3D information is shown in an interactive panel. The mutations selected by the user are highlighted with a sphere representation and classified as 'buried', 'surface' or 'interface', based on their location in the 3D structure. We provide 3D structural details for 731 531 point mutations mapped onto 10 934 proteins, representing >30% of the cancer point mutations currently cataloged. Of these, 81 422 mutations lay at protein-protein interaction interfaces (i.e. edgetic) and are thus likely to affect the connectivity of the human interactome (7,9). Below the structural representation, we report information about the genomic alterations detected in the selected proteins, such as the mutation type, somatic status and FATHMM (15) predicted functional impact for those mutations contained in COSMIC, as well as a brief description of the protein function or the method used to identify

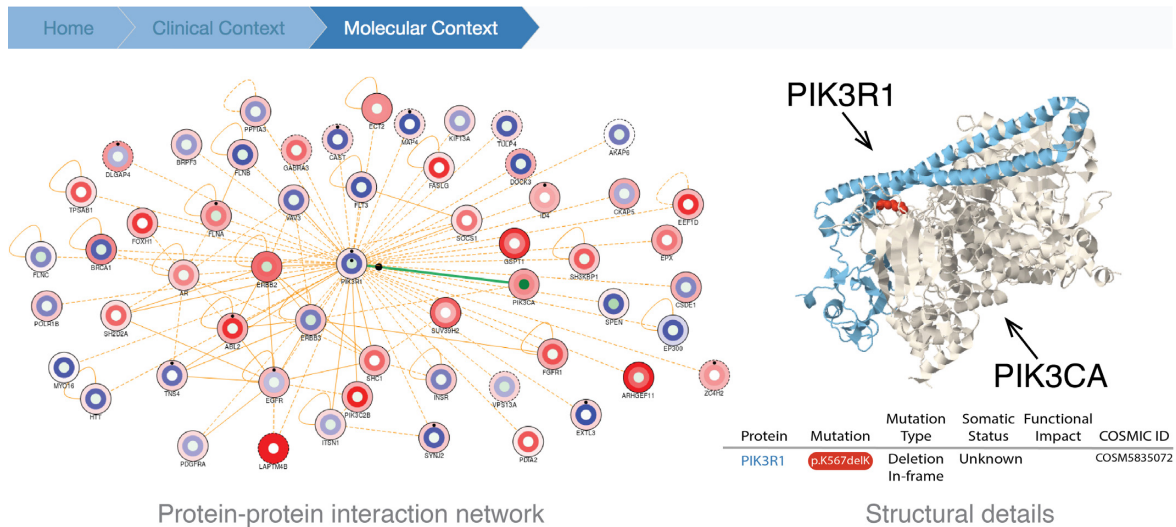


Figure 2. Molecular Context of $PIK3R1^{K567delK}$ and $PIK3R1^{overexpr.}$ alterations detected in a ductal breast carcinoma patient (TCGA-A2-A4RX-01). The protein–protein interaction network shows the 50 most frequently altered interactors of $PIK3R1$. Each node describes a protein with three concentric circles informing about the frequency of each type of alteration: the inner green circle represents somatic mutation frequency, the middle red/blue circle represents copy number gain/loss frequency and the outer red/blue circle represents over/underexpression frequency in the reference cohort. The edges represent physical interactions between proteins. Genomic alterations detected in this patient are represented by small black dots. Genomic alterations can affect one or more molecular layers of a protein. Somatic mutations can also affect the protein–protein interaction interface between two proteins. Nodes or edges represented with a solid line in the network have structural information available. The structural details show the two proteins involved in the selected interaction: $PIK3R1$ (blue) and $PIK3CA$ (gray). The precise location of the amino acid affected by the somatic mutation $PIK3R1^{K567delK}$ detected in this patient is highlighted in red.

the selected interaction. We also list the mutations detected in the reference cohort that map onto the same protein or protein–protein interaction interface, so that the user can identify highly mutated protein domains or regions.

USE CASE

To illustrate the applicability of Cancer PanorOmics, we studied the genomic alterations detected in a 67 years old woman diagnosed with ductal breast carcinoma (TCGA-A2-A4RX-01). The molecular profile of this patient consisted of 21 somatic mutations, 5 copy number alterations and 1120 gene expression changes.

The Patient Centric View (left panel in Figure 1) showed that six genomic alterations affect either known breast cancer driver genes ($PIK3R1^{K567delK}$, $STAG1^{E449Q}$, $KMT2C^{CN_{loss}}$, $PIK3R1^{overexpr.}$) or a driver gene that is not annotated to any specific tumor type ($LMNA^{G567V}$, $LMNA^{overexpr.}$). The Cohort Centric View (right panel in Figure 1) indicates that $STAG1$ and $PIK3R1$ are only rarely mutated in this reference cohort (mutated in <2% of patients). It also shows that $KMT2C$ is more frequently altered (deleted in 12% of patients). Many other gene expression alterations appear in this view, being the upregulation of $NDRG1$ the most recurrent one (>12% of patients).

We next explored the molecular context of the $PIK3R1^{K567delK}$ mutation. The network view (left panel in Figure 2) shows that 11 of the direct interactors of $PIK3R1$ exhibit gene expression changes (e.g. $EGFR$ and $ABL2$ upregulation, which occur in >5% of breast cancer patients). Moreover, $PIK3R1^{K567delK}$ lays at the interaction interface with $PIK3CA$ (right panel in Figure 2), one of

the most frequently mutated breast cancer drivers, likely triggering an edgetic perturbation. However, experimental validation is needed to know whether this mutation is really disrupting the $PIK3R1$ – $PIK3CA$ interaction, and whether this might have an impact on the oncogenic potential. Additionally, we observed that patients with putative $PIK3R1$ edgetic mutations have an overall survival curve that is more similar to that of patients with edgetic mutations in the $PIK3CA$ side of the interaction than patients with non-edgetic $PIK3R1$ mutations (Figure 3). Although we did not account for any other covariates that might influence the overall survival (i.e. age, tumor type or stage at diagnosis), this serves as an example of how a survival analysis could be used to identify edgetic mutations that might have a prognostic value and would deserve further investigation.

DISCUSSION

In the era of personalized medicine, we can now generate comprehensive and accurate portraits of individual tumors by putting together multiple layers of molecular profiles, such as somatic mutations, DNA copy-number alterations or mRNA expression. Computational solutions to large-scale data integration and visualization are needed to facilitate the interpretation of such complex data. Cancer PanorOmics has a significant added value with respect to other resources separately, since it switches between patient- and cohort-centric perspectives, contextualizing the alterations of interest, and offering a seamless systemic (network) view annotated with 3D structures. In addition, no programming skills are needed to submit an integrated

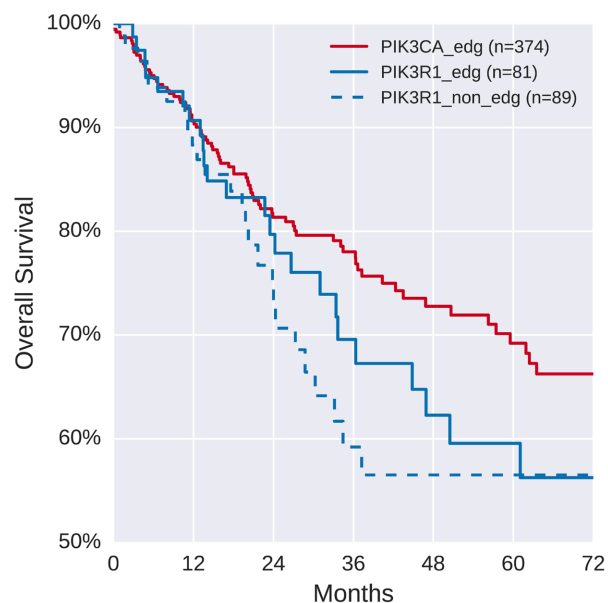


Figure 3. Kaplan–Meyer estimate of probability of overall survival (OS) in patients with non-edgetic PIK3R1 mutations and patients with putative PIK3R1–PIK3CA edgetic mutations at both sides of the interaction interface. The estimated 3-year survival probability of patients with putative PIK3R1–PIK3CA edgetic mutations is 69.58% (PIK3R1 interface, $n = 81$) and 78.02% (PIK3CA interface, $n = 374$), which are considerably higher than the 59.21% ($n = 89$) observed in patients with non-edgetic PIK3R1 mutations. The survival analysis has been generated with somatic mutations and survival data of 4795 cancer patients obtained from cBioPortal (18,19).

query and visualize the results with an enhanced and interactive display. We believe that servers like this one will help bridging the gap between cancer genomics and clinical oncologists, and play a central role in future personalized cancer management.

SERVER INFORMATION

Cancer PanorOmics runs on Apache HTTP Server, and it is written in PHP. After the automated validation by the server, the submitted mutations are stored into a PostgreSQL database and a user id is generated to identify access to the data. The server frontend is designed using the Bootstrap css framework. The interactive display of the mutation data relies on BioCircos.js (16) and Cytoscape.js (17) javascript libraries, which were modified to fulfill the visualization needs. D3.js (<https://d3js.org/>) javascript library is also used to display legend information, and Jsmol (<http://www.jmol.org/>) to provide interactive visualization of protein 3D structures.

ACKNOWLEDGEMENTS

We would like to thank the people involved in cancer genomics projects, as well as those annotating and cataloging in databases all the relevant findings. We also acknowledge the help of the Structural Bioinformatics & Network Biology group for extensive testing of the server.

FUNDING

Spanish Ministerio de Economía y Competitividad [BIO2016-77038-R]; European Commission [SyStemAge: 306240] and European Research Council [SysPharmAD: 614944] to P.A.; FPI and Juan de la Cierva fellowships, respectively to L.M. and C.P. Funding for open access charge: Spanish Ministerio de Economía y Competitividad. *Conflict of interest statement.* None declared.

REFERENCES

1. Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
2. International Cancer Genome, C., Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabe, R.R., Bhan, M.K., Calvo, F., Eerola, I. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
3. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
4. Vogelstein, B. and Kinzler, K.W. (2015) The path to cancer—three strikes and you're out. *N. Engl. J. Med.*, **373**, 1895–1898.
5. Tokheim, C.J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B. and Karchin, R. (2016) Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 14330–14335.
6. Rubio-Perez, C., Tamborero, D., Schroeder, M.P., Antolin, A.A., Deu-Pons, J., Perez-Llamas, C., Mestres, J., Gonzalez-Perez, A. and Lopez-Bigas, N. (2015) In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell*, **27**, 382–396.
7. Mosca, R., Tenorio-Laranga, J., Olivella, R., Alcalde, V., Ceol, A., Soler-Lopez, M. and Aloy, P. (2015) dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods*, **12**, 167–168.
8. Porta-Pardo, E., Hrabe, T. and Godzik, A. (2015) Cancer3D: understanding cancer mutations through protein structures. *Nucleic Acids Res.*, **43**, D968–D973.
9. Zhong, Q., Simonis, N., Li, Q.R., Charlotiaux, B., Heuze, F., Klitgord, N., Tam, S., Yu, H., Venkatesan, K., Mou, D. *et al.* (2009) Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.*, **5**, 321.
10. Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M.P., Jene-Sanz, A., Santos, A. and Lopez-Bigas, N. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods*, **10**, 1081–1082.
11. Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L. *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.*, **45**, D777–D783.
12. Mosca, R., Ceol, A. and Aloy, P. (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.
13. Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M., Kruger, F.A., Light, Y., Mak, L., McGlinchey, S. *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–D1090.
14. Finan, C., Gaulton, A., Kruger, F., Lumbers, T., Shah, T., Engmann, J., Galver, L., Kelly, R., Karlsson, A., Santos, R. *et al.* (2017) The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.*, **29**, doi:10.1126/scitranslmed.aag1166.
15. Shihab, H.A., Rogers, M.F., Gough, J., Mort, M., Cooper, D.N., Day, I.N., Gaunt, T.R. and Campbell, C. (2015) An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, **31**, 1536–1543.
16. Cui, Y., Chen, X., Luo, H., Fan, Z., Luo, J., He, S., Yue, H., Zhang, P. and Chen, R. (2016) BioCircos.js: an interactive Circos JavaScript library for biological data visualization on web applications. *Bioinformatics*, **32**, 1740–1742.
17. Franz, M., Lopes, C.T., Huck, G., Dong, Y., Sumer, O. and Bader, G.D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.

18. Cerami,E., Gao,J., Dogrusoz,U., Gross,B.E., Sumer,S.O., Aksoy,B.A., Jacobsen,A., Byrne,C.J., Heuer,M.L., Larsson,E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.
19. Gao,J., Aksoy,B.A., Dogrusoz,U., Dresdner,G., Gross,B., Sumer,S.O., Sun,Y., Jacobsen,A., Sinha,R., Larsson,E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, **6**, p11.