

# Oncogenomic portals for the visualization and analysis of genome-wide cancer data

Katarzyna Klonowska<sup>1</sup>, Karol Czubak<sup>1</sup>, Marzena Wojciechowska<sup>1</sup>, Luiza Handschuh<sup>1,2</sup>, Agnieszka Zmienko<sup>1,3</sup>, Marek Figlerowicz<sup>1,3</sup>, Hanna Dams-Kozłowska<sup>4,5</sup> and Piotr Kozłowski<sup>1</sup>

<sup>1</sup> European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

<sup>2</sup> Department of Hematology and Bone Marrow Transplantation, Poznan University of Medical Sciences, Poznan, Poland

<sup>3</sup> Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland

<sup>4</sup> Department of Diagnostics and Cancer Immunology, Greater Poland Cancer Centre, Poznan, Poland

<sup>5</sup> Chair of Medical Biotechnology, Poznan University of Medical Sciences, Poznan, Poland

**Correspondence to:** Piotr Kozłowski, *email:* kozlowp@yahoo.com

**Keywords:** cBioPortal, COSMIC, IntOGen, PPISURV/MIRUMIR, Tumorscape

**Received:** June 18, 2015

**Accepted:** September 28, 2015

**Published:** October 15, 2015

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Somatically acquired genomic alterations that drive oncogenic cellular processes are of great scientific and clinical interest. Since the initiation of large-scale cancer genomic projects (e.g., the Cancer Genome Project, The Cancer Genome Atlas, and the International Cancer Genome Consortium cancer genome projects), a number of web-based portals have been created to facilitate access to multidimensional oncogenomic data and assist with the interpretation of the data. The portals provide the visualization of small-size mutations, copy number variations, methylation, and gene/protein expression data that can be correlated with the available clinical, epidemiological, and molecular features. Additionally, the portals enable to analyze the gathered data with the use of various user-friendly statistical tools. Herein, we present a highly illustrated review of seven portals, i.e., Tumorscape, UCSC Cancer Genomics Browser, ICGC Data Portal, COSMIC, cBioPortal, IntOGen, and BioProfiling.de. All of the selected portals are user-friendly and can be exploited by scientists from different cancer-associated fields, including those without bioinformatics background. It is expected that the use of the portals will contribute to a better understanding of cancer molecular etiology and will ultimately accelerate the translation of genomic knowledge into clinical practice.**

## INTRODUCTION

Cancer encompasses a broad spectrum of diseases (>100) that arise from somatically acquired genetic, epigenetic, transcriptomic, and proteomic alterations that have accumulated in the genomes of cancer cells [1]. These alterations are implicated in hallmark oncogenic cellular processes that are characterized by, e.g., sustained proliferative signaling, resistance to apoptosis, induction of invasion and metastasis, and neoangiogenesis [2]. The somatic loss-of-function or gain-of-function alterations are overrepresented in specific genomic regions, which

could indicate their potential suppressive or oncogenic roles, respectively. However, it must be noted that somatic mutations occur on different genetic backgrounds and can sometimes interact with germline mutations, which could modify predisposition to cancer when such mutations occur in cancer-associated genes.

Recent advances in technologies for high-throughput genome analysis, such as microarray-based methods and next-generation sequencing (NGS), have enhanced progress in the field of oncogenomics [3]. These tools were fundamental for the initiation and development of multi-centered cancer genomic projects, such as (i) the

Wellcome Trust Sanger Institute's Cancer Genome Project (CGP) [4, 5], (ii) The Cancer Genome Atlas (TCGA) [6-8], and (iii) the International Cancer Genome Consortium (ICGC) cancer genome projects [9, 10]. These projects have been launched for genome-wide analyses of genetic, epigenetic, transcriptomic, and proteomic alterations in hundreds or even thousands of cancer samples. Their general aim is to provide publicly available oncogenomic datasets for the better understanding of the molecular mechanisms that underlie cancer and for the assessment of the influence of specific alterations on clinical phenotypes. Application of the appropriate pipeline for computational interpretation and thought-provoking visualization of the results of oncogenomic projects is crucial to exploring the multidimensional character of genome-wide cancer data [11]. In response to this need, a number of oncogenomic portals were created to assist with accessing the abundant cancer datasets. These portals gather and facilitate the analysis of data with regard to small-size mutation, copy number variation (CNV), methylation, and gene/protein expression. Moreover, they offer a wide range of analysis tools that include the testing of correlations of specific genomic alterations with available clinical information.

Herein, we provide a highly illustrated guide through several web-based oncogenomic portals that were generated to facilitate scientists from different cancer-associated fields, including molecular and clinical oncologists, epidemiologists, and bioinformaticians, with the extraction of meaningful information from expanding oncogenomic sources. Browsing through the portals, prospective users will find a variety of data regarding cancer types and subtypes, oncogenic molecular pathways and cancer-associated genes of interest. All of the portals described below are user-friendly and provide intuitive integration as well as interactive oncogenomic dataset visualizations, and thus, bioinformatics skills and knowledge are not essential to exploring and using these tools. The individual paragraphs listed below present the characteristics and possible utilization of selected web portals. Descriptions and figures that present specific portals were prepared according to their versions from the first half of 2015, and they are summarized in Table 1.

## Tumorscape

Tumorscape [12, 13] was developed at The Broad Institute of MIT and Harvard in Cambridge, MA USA. This website was one of the first oncogenomic portals to provide information about cancer copy number changes in a format that was easily accessible to non-bioinformaticians. With this portal, the copy number profiles of over 3,700 cancers (both primary cancers and cell lines) are mapped to the human genome reference sequence and are visualized as heatmap tracks, with the use of the Integrative Genomics Viewer (The Broad Institute). Genomic regions with increased ( $>2$ ) and

decreased ( $<2$ ) copy number are marked, respectively, in red and blue colors, the intensity of which indicates the amplitude of the copy number changes (Figure 1). The tracks that represent all of the analyzed samples are shown next to one another, forming a panel that allows direct comparison and visualization of all of the analyzed samples. In addition, Tumorscape provides tools that allow "cancer-centric" and "gene-centric" data analyses. The "cancer-centric" analysis (Figure 1A) provides a list of genomic regions that are either significantly amplified or deleted in a specific cancer along with information about the genes that are located in the altered regions. The "gene-centric" (Figure 1B) analysis provides summary statistics of the copy number alterations that affect a gene of interest in a specific cancer type and/or across all cancer types. This summary enables the interpretation of the role of an analyzed gene as a potential oncogene or tumor suppressor.

## UCSC cancer genomics browser

The University of California at Santa Cruz (UCSC) Cancer Genomics Browser [14-19] integrates oncogenomic CNV, small-size mutations, methylation, transcriptomic, and proteomic datasets that were obtained in a variety of experiments that were conducted with the use of samples from different cancer types and subtypes. With this portal, all of the oncogenomic information is mapped to the human genome reference sequence and presented as color-coded heatmap tracks. As in Tumorscape, the data from specific experiments are visualized as panels of heatmap tracks in which each track represents an individual sample. Using this portal, the required data can be browsed from the perspective of the whole genome, the exome, a specific chromosome, or a gene. Additionally, there is also the possibility of viewing PARADIGM datasets to gather a sample-specific "gene activity level." This parameter (obtained using the PARADIGM method) [20] provides the incorporation of pathway interactions (which are deposited in the NCI Pathway Interaction Database) [21] and the integration of data with regard to different types of oncogenomic alterations, e.g., changes in the expression or copy number of a given gene [16]. In the UCSC Cancer Genomics Browser, multiple panels can be simultaneously displayed to visualize different categories of oncogenomic information for a specific cancer type and/or the same category of oncogenomic data for different cancer types (Figure 2A-2D). With this browser, analyses can be concurrently conducted for thousands of samples (oncogenomic datasets) that are sorted by different clinical, epidemiological, and molecular features (Figure 2). These features include survival, histological type, tumor nuclei percent, followup treatment success, new tumor event after initial treatment, neoplasm histologic grade, and tumor necrosis percent, as well as gender, age

**Table 1: Main characteristics of the selected oncogenomic portals.**

database	data source	sites of analysed cancer <sup>1</sup>	organisation of data <sup>2</sup>	oncogenomic data/analyses	link/literature
Tumorscape	Broad Institute	Bd; Bld; Br; Bra; Clr; Eso; GIST; HN; Htp; Kd; Lng; Lvr; Lymph; Msh; Ov; Pnc; Prst; Sk; ST; Stc; Swn; Thr; Utr; also in: cancer cell lines	level i-iii	copy number alterations	<a href="http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf">http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf</a> ; [12]
UCSC Cancer Genomics Browser	TCGA, SU2C Breast Cell Line, Cancer Cell Line Encyclopedia, The Connectivity Map, TARGET, cancer data from literature	Bd; Bld; Br; Bra; Chl; Col; Clr; EG; Eso; HN; Kd; Lng; Lvr; Lymph; Msh; Ov; Pan; Pnc; Prc / Prn; Prst; Rc; Sk; ST; Stc; Thm; Thr; Utr; also: cancer cell lines; cancer data from mouse models	level i-iii	DNA copy number, miRNA/exon/gene/protein expression, DNA methylation, gene-level mutations, PARADIGM pathway activity; clinical, epidemiological, and molecular information	<a href="https://genome-cancer.ucsc.edu">https://genome-cancer.ucsc.edu</a> ; [14-18]
ICGC Data Portal	ICGC, TCGA, TARGET	Bd; Bld; Bo; Br; Bra; Clr; Col; Eso; HN; Kd; Lng; Lvr; Lymph; Nb; Ov; Pnc; Prst; Rc; Sk; ST; Stc; Thr; Utr;	level i-iv	simple somatic mutations, copy number somatic alterations, structural somatic mutations, simple germline variants, DNA methylation, gene/protein expression, miRNA expression, exon junction; epidemiological and clinical data	<a href="https://dcc.icgc.org">https://dcc.icgc.org</a> ; [32]
COSMIC	TCGA, ICGC, cancer data from literature	Bo; Br; EA; Eso; GIST; Htp; Kd; Lvr; Lng; Ov; Pnc; Prst; Sk; Stc; Tst; Thm; Thr; Utr	level iii-iv	somatic mutations, copy number alterations, gene expression	<a href="http://www.sanger.ac.uk/genetics/CGP/cosmic">http://www.sanger.ac.uk/genetics/CGP/cosmic</a> ; [39-43]
cBioPortal	AMC, BCCRC, BGI, British Columbia, Broad, Broad/Cornell, CCLE, CLCGP, Genentech, ICGC, JHU, Michigan, MKSCC, MKSCC/Broad, NCCS, NUS, PCGP, Pfizer UHK, Riken, Sanger, Singapore, TCGA, TSP, UTokyo, Yale	ACC; Bd; Bld; Br; Bra; Chl; Clr; Eso; HN; Kd; Lng; Lvr; Lymph; MM; Npx; Ov; Pnc; Prst; Sk; ST; Stc; Thr; Utr; also: cancer cell lines	level iii-iv	mutations, putative copy number alterations; mRNA expression, protein/phosphoprotein level; survival analyses	<a href="http://www.cbioportal.org">http://www.cbioportal.org</a> ; [57, 58]
IntOGen (2014.12)	TCGA, ICGC, cancer data from literature	Bd; Bld; Br; Bra; Clr; Eso; HN; Kd; Lng; Lvr; Lymph; Ov; Pnc; Prst; Sk; Stc; Thr; Utr	level iii-iv	results of the analyses indicating driver alterations and genes; therapies tailored to the mutation profiles of the analyzed patients	<a href="http://www.intogen.org/">http://www.intogen.org/</a> ; [67-70]
BioProfiling.de					
PPISURV	for gene expression: Gene Expression Omnibus; for interactome: IntAct, HPRD, Reactome, HumanCyc, NCI_NATURE, PhosphoSitePlus	Bd; Bld; Br; Bra; Col; Htp; Lng; Lvr; Lymph; Ov; Prst; ST; Utr	level iv	survival analyses	<a href="http://bioprofiling.de/GEO/PPISURV/ppisurv.html">http://bioprofiling.de/GEO/PPISURV/ppisurv.html</a> ; [81]
MIRUMIR	Gene Expression Omnibus	Br; Eso; Lvr; Lng; Npx; Ov; Prst; Sk	level iv	survival analyses	<a href="http://www.bioprofiling.de/GEO/MIRUMIR/mirumir.html">http://www.bioprofiling.de/GEO/MIRUMIR/mirumir.html</a> ; [83]
DRUGSURV	for gene expression: Gene Expression Omnibus; for drugs modulating a gene of interest: DrugBank, Pubchem Bioassay	Bld; Br; Bd; Col; Bra; Lng; Lvr; Lymph; Prst; ST; Utr	level iv	list of drugs targeting specific genes/cancer types; survival analyses	<a href="http://www.bioprofiling.de/GEO/DRUGSURV/index.html">http://www.bioprofiling.de/GEO/DRUGSURV/index.html</a> ; [85]

<sup>1</sup>List of abbreviations of cancer sites. In the brackets there are exemplary cancer subtypes included in the portals.

ACC – adenoid cystic carcinoma; Bd – bladder; Bld – blood; Bo – bone; Br – breast; Bra – brain; Chl – cholangiocarcinoma; Clr – colorectal; Col – colon; EA – eye and adnexa; EG - endocrine glands; Eso – esophagus; GIST – gastrointestinal; HN – head and neck; Htp – hematopoietic; Kd – kidney; Lng – lung; Lvr – liver and biliary tract; Lymph – Lymphoma; Msh – mesothelioma; Mth – mouth; Nb – neuroblastoma; Npx – nasopharynx; Ov – ovary; Pan – pancancer; Pnc – pancreas; Pnx – pharynx; Prc/Prn - pheochromocytoma and paraganglioma; Prst – prostate; Rc – rectum; Sk – skin; ST – soft tissues; Stc – stomach; Swn – schwannoma; Thm – thymus; Thr – thyroid; Tst – testis; Utr – uterine (cervix and corpus).

<sup>2</sup>In oncogenomic portals cancer resources are arranged in different levels of organisation, including: (i) raw, (ii) computationally processed/normalized, (iii) interpreted and (iv) summarized data [3].

### A Analysis by Cancer Type

Summary Amplifications Deletions

Cancer Type: Lung NSC Available Data: 62 peaks Page: 1 2 3 4

Peak Region	#Genes in Peak	Residual Q-value	Frequency of Amplification			Genes in Selected Peak
			Overall	Focal	High-level	
<a href="#">chr14:35755817-35835900</a>	0	5.34E-78	0.438	0.243	0.1	EGFR
<a href="#">chr12:67484379-68036772</a>	4	1.39E-26	0.323	0.153	0.053	
<a href="#">chr8:129271608-129288419</a>	0	3.86E-25	0.557	0.196	0.087	
<a href="#">chr7:54995340-55363609</a>	1	7.19E-25	0.45	0.101	0.067	
<a href="#">chr12:56419523-56444825</a>	6	1.03E-22	0.322	0.153	0.041	

(truncated for presentation purpose)

### B Analysis by Gene

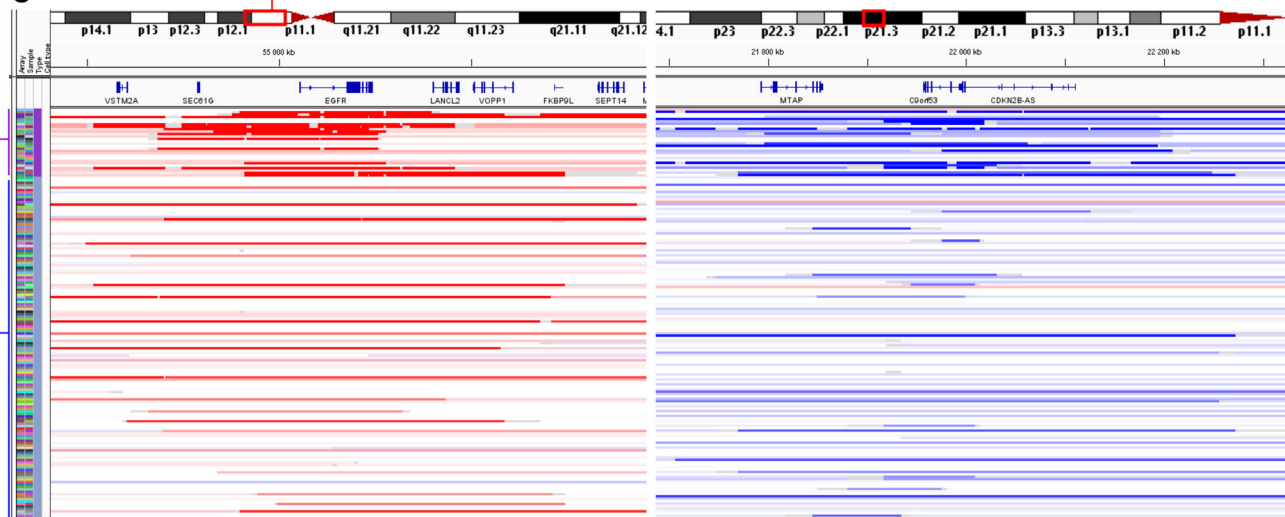
Summary Amplifications Deletions

EGFR ([chr7:55054218-55206232](#))

Cancer Subset	In Peak?	Nearest Peak	#Genes in Peak	Q-value	Frequency of Amplification		
					Overall	Focal	High-level
<a href="#">all_cancers</a>	Yes	<a href="#">chr7:54911575-55275039</a>	1	1.57E-46	0.3159	0.0562	0.0438
<a href="#">all_epithelial</a>	Yes	<a href="#">chr7:54911575-55363609</a>	1	5.62E-37	0.4161	0.0725	0.0575
<a href="#">Lung NSC</a>	Yes	<a href="#">chr7:54995340-55363609</a>	1	9.29E-25	0.4529	0.0996	0.0668
<a href="#">all_lung</a>	Yes	<a href="#">chr7:54946418-55363609</a>	1	4.39E-23	0.4496	0.0956	0.0633
<a href="#">Esophageal squamous</a>	Yes	<a href="#">chr7:54762764-55366816</a>	2	7.55E-8	0.7273	0.3409	0.2273
<a href="#">Breast</a>	Yes	<a href="#">chr7:54662395-57367942</a>	14	0.0318	0.3498	0.0617	0.0494
<a href="#">Glioma</a>	No	<a href="#">chr7:55196509-55258847</a>	0	3.4E-36	0.8293	0.4634	0.439
<a href="#">all_neural</a>	No	<a href="#">chr7:55196509-55258847</a>	0	1.79E-23	0.3917	0.1014	0.0922
<a href="#">Hepatocellular</a>	No	No peak on chromosome	0	0.895	0.3636	0.0248	0.0413

(truncated for presentation purpose)

### C



(truncated for presentation purpose)

**Figure 1: Examples of Tumorscape data analysis and visualization.** **A.** An example of the results that were obtained with the “cancer-centric” analysis. The table shows a list of genomic regions that were most frequently amplified in lung adenocarcinoma. The q-value represents the likelihood of a random occurrence of the specific amplification/deletion that is calculated based on the background copy number variation. The fourth most frequently amplified region that spans *EGFR* is highlighted. **B.** Results obtained with “gene-centric” analysis; the table depicts a list of cancers in which the representative gene (*EGFR*) is located in or near the frequently amplified region (orange and yellow rows, respectively). **C.** Visualization of chromosomal regions that span the exemplary *EGFR* and *CDKN2A* genes, which are undergoing frequent amplifications and deletions, respectively. The heatmaps show copy number variations of glioma and lung adenocarcinoma samples. Each row represents an individual sample, and red and blue indicate amplification and deletion, respectively.



at initial pathologic diagnosis, tobacco smoking history, cytogenetic abnormalities, and expression subtypes. Apart from the heatmap tracks (Figure 2B), the data presented in specific panels can be summarized and plotted as box-and-whiskers or proportions (Figure 2C).

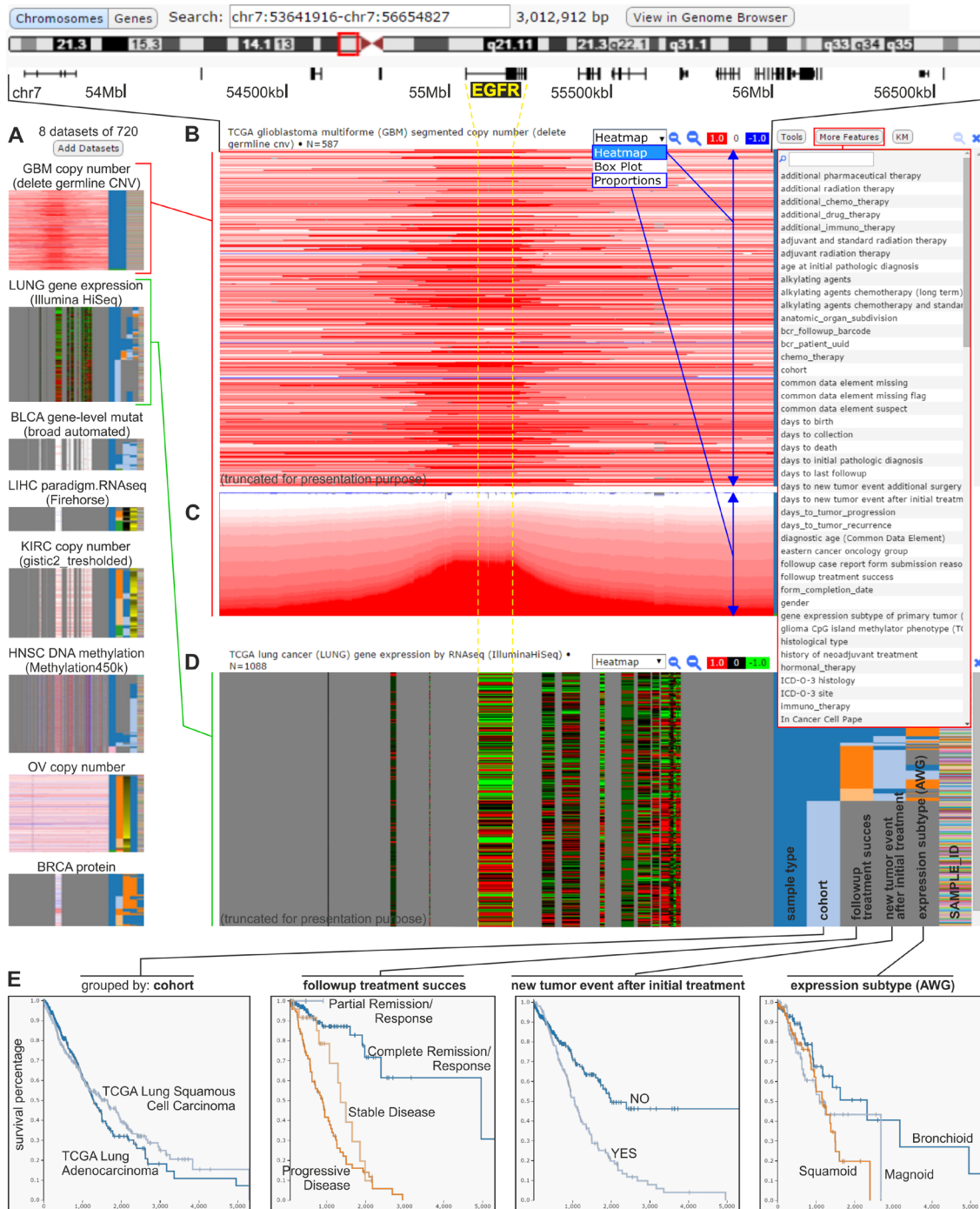
The datasets can also be statistically processed and depicted with the use of a number of tools, such as the hgSignature, which enables the simultaneous analysis of the expression of several genes, to incorporate an algebraic expression signature as a clinical feature. The inclusion of such a feature to the statistical analysis of cancer data could allow the correlation of the molecular and clinical phenotypes or the subdivision of the clinical phenotypes based on the molecular data [15]. Additionally, a correlation of the available clinical, epidemiological, and molecular features with a patient's survival can be depicted in a Kaplan-Meier plot (Figure 2E). Subgroups of samples (distinguished based on the associated features or genomic signatures) can be compared in terms of the obtained oncogenomic data with the use of various statistical tests [i.e., differences in mean, Wilcoxon, Fisher's exact, Fisher's linear discriminant, Jarque Bera normality, Levene homogeneity of variances (HOV), Brown - Forsythe HOV, and Student's T-tests], which can be adjusted for multiple hypotheses p-values through the Bonferonni and Benjamini-Hochberg false discovery rate (FDR) corrections. Importantly, all of the genomic information that is stored in the UCSC database can be easily downloaded for external analyses.

Successful applications of the UCSC Cancer Genomics Browser in cancer-associated research are described in many papers [22-30]. For example, Wu and colleagues [22] used the statistical tool for the generation of a Kaplan-Meier plot to support the significance of their experimental data. Their study revealed that the up-regulated expression level of *HNFI1A-AS1* in lung adenocarcinoma is significantly correlated with the TNM stage, tumor size, and lymph node metastasis. These results are in line with the Kaplan-Meier plot, which indicates that patients with high *HNFI1A-AS1* expression overall experienced worse survival compared to patients with low *HNFI1A-AS1* expression. The UCSC Cancer Genomics Browser is also broadly used for downloading genomic and clinical data for external analyses [24-26, 30].

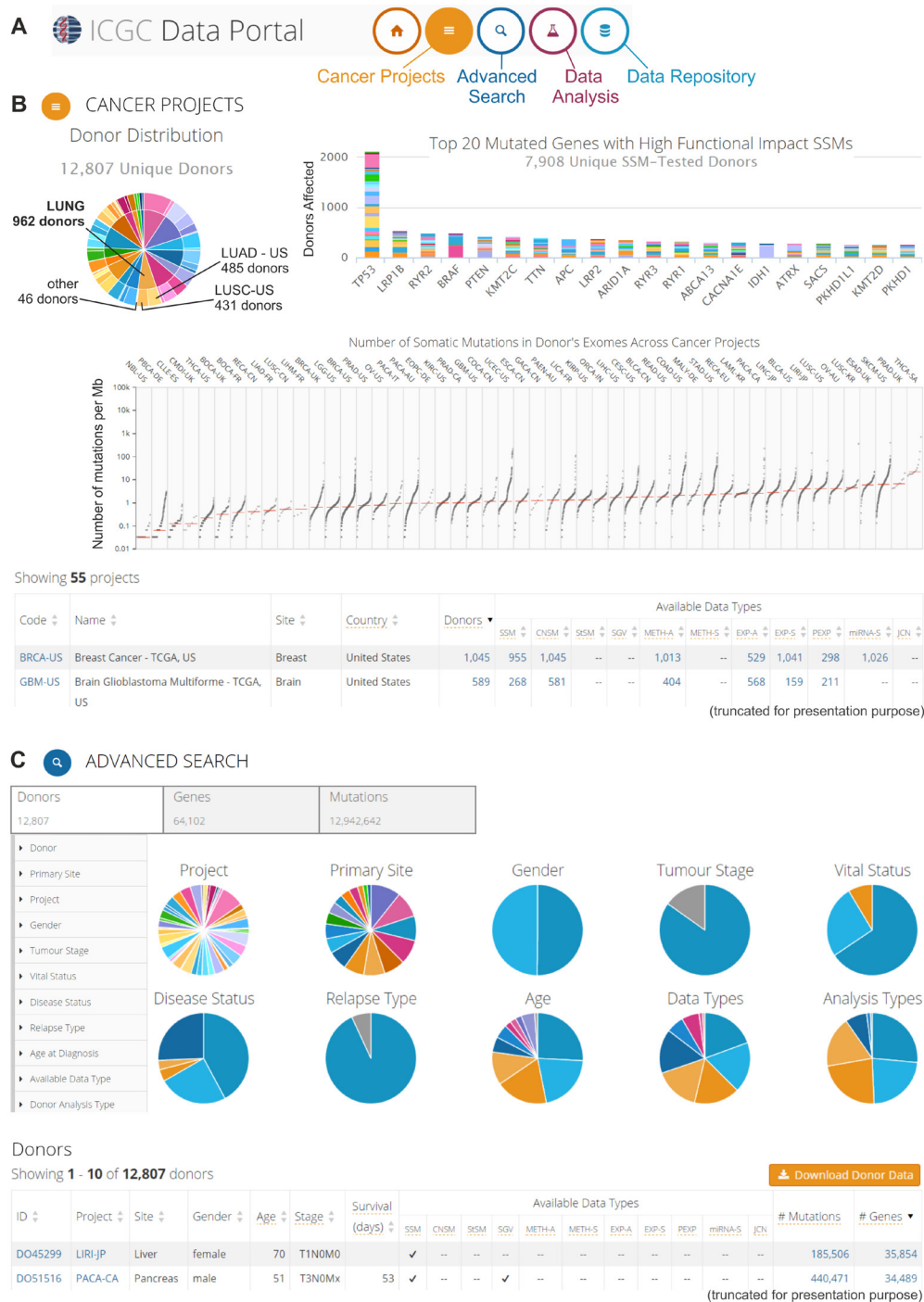
It is also noteworthy that the authors of the UCSC Cancer Genomics Browser are currently developing a new oncogenomic platform called UCSC Xena [31], which allows users to upload, visualize, and analyze a custom genomic dataset in the context of the large projects data stored in the web browser. Although the UCSC Cancer Genomics Browser and the UCSC Xena currently coexist, it is anticipated that after adding some vital functionalities, UCSC Xena will replace the UCSC Cancer Genomics Browser [18].

## ICGC data portal

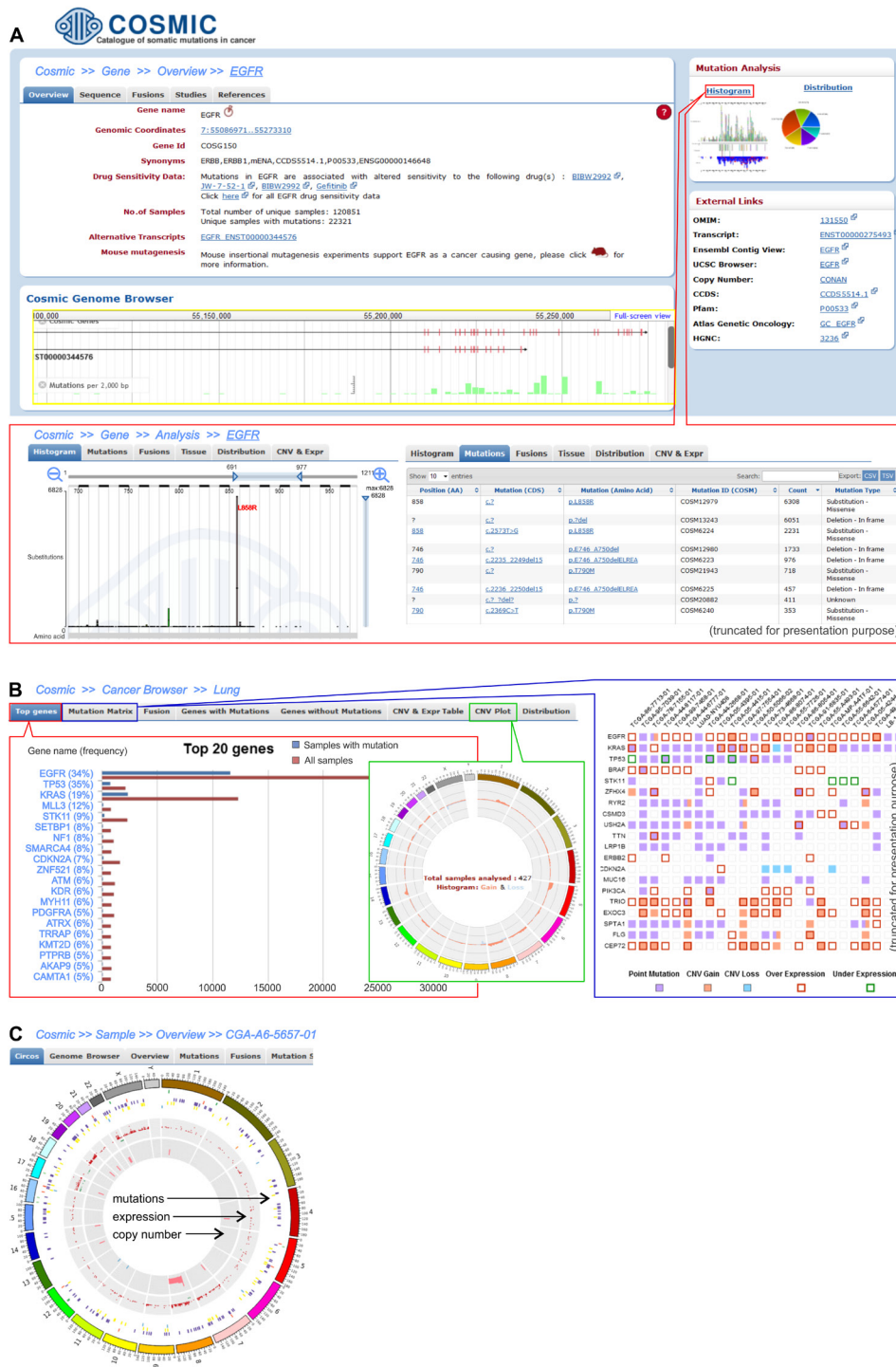
The ICGC Data Portal [32, 33] provides integration and visualization of the results of 55 cancer projects. This portal was created for the analysis of genomic sequence alterations in relation to clinical patient characteristics, such as ethnicity and epidemiological information. With this portal, the oncogenomic data can be analyzed using four interactive entry points: "Cancer Projects," "Advanced Search," "Data Analysis" and "Data Repository" (Figure 3A). The "Cancer Projects" (Figure 3B) enables data browsing from distinct projects that focus on the oncogenomic analysis of specific cancer types and subtypes. For each dataset, the provided summary includes a list of available oncogenomic data types, most affected donors, genes most frequently affected by cancer alterations, and most common mutations. It is also possible to use the "keyword search" tool to browse all of the gathered oncogenomic data in terms of a specific gene, mutation, donor, or molecular pathway that is of interest. The integration of external databases, such as the Ensembl [34], OMIM [35], Reactome [36], and COSMIC [37], enables the user to look more broadly at a specific gene, molecular pathway, or mutation in terms of its role in carcinogenesis. The "Advanced Search" (Figure 3C) allows extending the analysis and correlating data with additional clinical (e.g., tumor stage, relapse type, disease status), epidemiological (e.g., gender, age at diagnosis, vital status), molecular (e.g., type of the mutation and its consequence), and technical (e.g., type of sequencing platform used for the analysis) information. The "Data Analysis" entry point allows launching three types of analyses: "Enrichment Analysis," "Phenotype Comparison," and "Set Operations." The "Enrichment Analysis" permits the user to identify groups of gene sets from the selected "universe," i.e., Reactome Pathways, Gene Ontology (GO) Molecular Function, GO Biological Process or GO Cellular Component, which appear to be statistically significantly over-represented when compared with a custom gene set that is uploaded by the user. The uploaded custom gene set can consist of up to 10,000 genes. The "Enrichment Analysis" is based on a hypergeometric test and Benjamini-Hochberg adjustment for multiple test corrections with the FDR value threshold selected by the user. The "Phenotype Comparison" analysis allows the user to compare some clinical and epidemiological characteristics across patients with various cancer types, whereas the "Set Operations" can be used to distinguish the shared fraction of the analyzed sets, which are depicted in a Venn diagram (e.g., mutations that are causative across several cancer types). "Data Repository" allows all of the ICGC Cancer Project data to be downloaded and analyzed with the use of external programs and tools of interest. An example of ICGC Data Portal utilization for downloading oncogenomic data has already been published [38].



**Figure 2: The UCSC Cancer Genomics Browser.** An example of analysis focused on the *EGFR* genomic region that is conducted concurrently on various oncogenomic data across different cancer types and subtypes. **A.** Small-scale images (icons) of selected datasets that are simultaneously visualized in the browser. Datasets represented by icons are displayed in a column, similar to the datasets from panels B-D. **B.** A heatmap panel that presents the results of the TCGA genome-wide copy number analysis of glioblastoma multiforme (GBM) samples. A screenshot of the GBM dataset was used for presentation, based on the presence of considerable amplification of the genomic region that spans the representative *EGFR*. Each horizontal line (track) represents a specific sample. The red or blue colors indicate, respectively, a gain or loss in the copy number. On the right side of panel B, there is a drop-down list with epidemiological, clinical, and molecular attributes that can be used to sort the presented data (as shown in panel D). **C.** The TCGA copy number data identified in patients with GBM visualized as a proportions plot. **D.** A heatmap panel showing the results of TCGA analysis of gene expression in lung cancer samples in the genes that are indicated above (e.g., *EGFR*). Red and green colors indicate, respectively, upregulation and downregulation of the relative gene expression. The samples are sorted by epidemiological, clinical, and molecular attributes (selected from a drop-down list of attributes), as in panels B and C, shown on the right side of the expression panel. The copy number and expression data presented in panels B-D correspond to the same genomic region indicated above panel B. **E.** Kaplan-Meier plots generated using the attributes of lung cancer samples (shown in the right side of panel D).

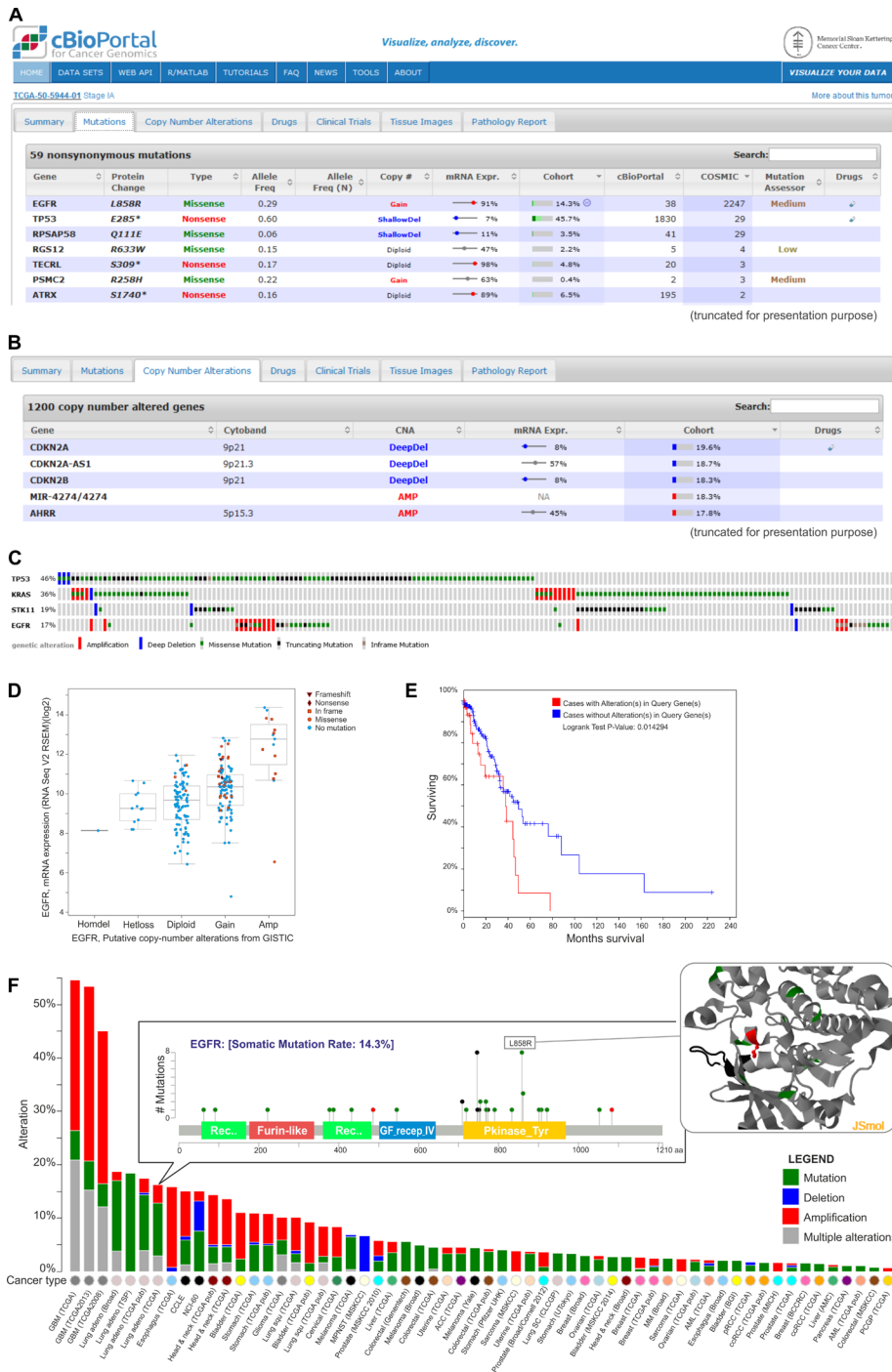


**Figure 3: The ICGC Data Portal.** An example of possible data analyses and visualizations. **A.** Three interactive entry points to the ICGC Data Portal. **B.** The “Cancer Projects” entry point. Screenshot of summary results from all 55 cancer projects. The upper left-hand panel: pie chart that depicts the distribution of cancer types (internal circle) and cancer subtypes/projects (external circle) among the donors, e.g., different lung cancer types and subtypes/projects (indicated in the pie chart). The upper right-hand panel: bar plot that represents the top 20 most frequently mutated genes. Different colors indicate different projects. The middle panel: scatter plot that depicts the distribution of the number of somatic mutations in the donors’ exomes across cancer projects. Each dot represents the number of somatic mutations (per 1 Mb) that are identified in the analyzed sample. Vertical lines indicate the median number of mutations. The bottom part of panel B shows a summary of each project. More information about the specific project (types of experimental analyses, available genomic data, most commonly mutated genes, most common mutations, and most affected donors) can be found by clicking at specific project code. **C.** The “Advanced Search” entry point, which enables extended analysis of the oncogenomic data. This screenshot shows the browsing of donor features. The upper left-hand panel depicts features that can be used for filtering the donor data. The middle panel (pie charts) provides a summary of the clinical, epidemiological, and molecular attributes of the donors. The bottom panel represents summary data about specific donors. More information (clinical and genetic) can be found by clicking at the donor ID.



**Figure 4: Three levels of data analysis in the COSMIC browser.** **A.** Screenshot shows exemplary *EGFR* gene data. The upper left-hand panel demonstrates basic information about the gene, whereas the right-hand panel of “Mutation analysis” provides links to the detailed data of mutations that were detected in the *EGFR*. Within the panel, there is a “Histogram” link that allows detailed analysis of the gene alterations, whose features are shown in the framed panel. One of the histograms shows the distribution of *EGFR* tyrosine kinase domain mutations, with the most frequently occurring mutation being L858R. The distribution can also be visualized as a table (on the right). **B.** The screenshots present the results for the representative lung adenocarcinoma cancer type. The left framed panel shows a list of the 20 most frequently mutated genes, whereas the middle and right framed panels display a CNV plot and the Mutation Matrix, respectively. The CNV circular plot shows a summary of the copy number variations across the whole genome of the lung adenocarcinoma. The height of the corresponding bars shows the total number of samples with CNV in a specific region. The Mutation Matrix presents alterations in the most frequently mutated genes ( $y$ -axis) in the adenocarcinoma samples that have the highest number of alterations ( $x$ -axis). **C.** Circular plot of all of the alterations (coding mutations, gene expression and CNV) that are detected in an individual exemplary sample (TCGA-A6-5657-01) of adenocarcinoma.





## COSMIC

The Catalogue of Somatic Mutations in Cancer (COSMIC) was developed at the Wellcome Trust Sanger Institute in Hinxton, UK [37, 39-43]. It is the most comprehensive database of somatic mutations in cancer. The portal provides information about the CNV and the expression level of cancer-associated genes that is obtained via the analysis of all of the samples that were tested for specific mutations (both positive and negative results are reported). This tool enables the calculation of the objectivized frequency of mutations in different types of tumors. The records included in COSMIC are derived from two sources: (i) a literature review of over 21,000 research papers and (ii) two projects: TCGA and ICGC. Together, these sources provide information that is obtained from more than a million samples. For almost 20,000 samples, whole-genome sequencing was conducted, which provided complete information about alterations in their genomes. In addition to the above, literature curation allowed the generation of the Cancer Gene Census, which is available under the COSMIC external links; thus far, it is the most reliable list of cancer-associated genes.

The data integrated in COSMIC can be searched by sample name, by gene name, and via cancer browser (Figure 4). Searching by the sample name allows the user to obtain a genome-wide overview of all of the cancer-associated events (e.g., mutations, gene fusions, and CNV) associated with a sample of interest. The second approach enables the user to overview all of the data that is related to a specific gene, such as its sequence, mutations, fusions, copy number variations, and expression. The data that refer to a specific cancer type (mutations, fusion and copy number and expression alterations of genes) can be retrieved via the cancer browser.

Due to its comprehensiveness, COSMIC is widely used and has been cited in hundreds of publications (e.g., [44-56]). For example, Chen et al. [46] used this database to confirm the presence of specific mutations in the *KRAS*, *NRAS*, and *BRAF* genes in myeloma cell lines. In another study, Ostrow and colleagues [48] took advantage of the Cancer Gene Census to select well-known cancer-associated genes for further analyses of the dynamics of the evolutionary process within tumors, with a focus on breast cancer.

## cBioPortal

The cBioPortal [57-59] was developed at the Memorial Sloan-Kettering Cancer Center in New York City, NY USA. This portal contains genomic data, including copy number alterations, mRNA and microRNA expression, DNA methylation and protein and phosphoprotein abundance, which were obtained for

multiple types of cancer. Currently, the portal collects records that were derived from 91 individual cancer studies, in which 31 types of cancer were analyzed with the use of over 21,000 samples. Because the tools that were integrated in the portal perform different types of analyses, different statistical tests can be used to assess the significance in specific analysis (for example, Fisher's exact test can be used to calculate the significance of mutual exclusivity of two genes or the log-rank test can be used to calculate survival analysis significance). All of the portal data can be retrieved in a format that is compatible with the R framework for statistical computing and graphics.

Cancer-associated alterations deposited in the cBioPortal can be browsed as (i) the overview of all of the genomic events that were detected in an individual cancer sample (Figure 5A, 5B), (ii) alterations in a specific gene across all of the samples that were included in one study (Figure 5C-5E), and (iii) a comparison of the frequency of the alterations in a given gene across all 91 studies (Figure 5F). For each study, it is also possible to inquire which genes are most frequently altered in the analyzed set of samples. In the cBioPortal, the genomic data are integrated with clinical outcomes, which allows determining whether a specific gene plays a potentially oncogenic role in a given cancer type. Apart from the on-line analysis of data deposited in the portal, there is also the possibility to download the results that were obtained for a specific study. Additionally, the browser enables the visualization of data that is uploaded by the user.

A wide range of tools that are available makes the portal useful in various types of analyses, which has resulted in its popularity and applicability (e.g., [51, 60-65]). For example, the authors of this paper used this portal to determine the correlation between copy number changes and expression level of two miRNA biogenesis genes (*DROSHA* and *DICER1*) that were found to be frequently amplified in lung cancer [63]. Other authors used the cBioPortal for the analysis of the *PARK2* deletion in low-grade glioma and glioblastoma and for the analysis of the correlation between *PARK2* mRNA expression and prognosis in patients [60]. Lu and colleagues used the portal to retrieve copy number data for the design of the model that predicts genetic interactions in human cancer [61].

## IntOGen

The Integrative Oncogenomics Cancer Browser (IntOGen) [66] was developed by the Biomedical Genomics Group integrated in the Research Unit on Biomedical Informatics of the University Pompeu Fabra, Biomedical Research Park in Barcelona. The browser contains the results of computational secondary analyses of oncogenomic data from several large genome-wide projects. The analyses were focused on the selection

**A**

**Summary of survival analyses of EGFR interactome**

Survival profile of EGFR itself (click here to view !!!)

The table below reports summary for each type of EGFR interactions with respect to survival across available datasets.

Interactome source	Number of partners	Significant (p-val<0.05) hits	Best p-value*	Link to full report
IntAct	206	14	8.95e-05	<a href="#">view details</a>
HPRD	167	5	0.0015	<a href="#">view details</a>
Reactome	60	2	0.022	<a href="#">view details</a>
HumanCyc	1041	1	0.0025	<a href="#">view details</a>
NCI_NATURE	52	1	0.011	<a href="#">view details</a>
PhosphoSitePlus	29	0	0.31	<a href="#">view details</a>
<b>all</b>	<b>1385</b>	<b>14</b>	<b>0.00015</b>	<a href="#">view details</a>

\* the p-values are adjusted for multiple testing by FDR

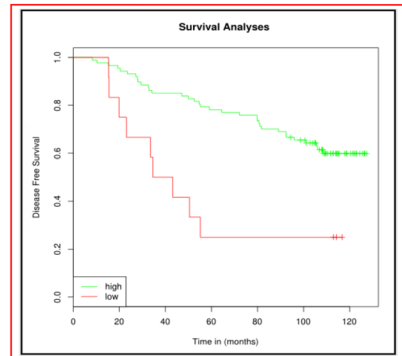
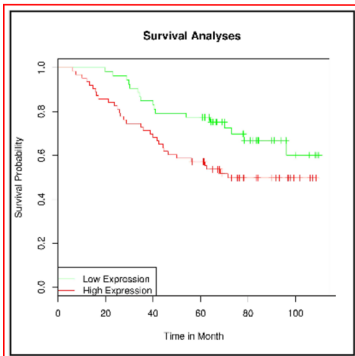
\*\* Effect sign ("positive", "negative", "no") indicates weather significant proportion of genes from the list is associated either positively or negatively with survival.

GEO Dataset	Cancer Type	P-value, FDR adjusted (unadjusted)	Odds ratio	k	K	m	M	Effect sign**	View Details
Gene expression data for pathological stage I-II lung adenocarcinomas	Lung cancer	8.95e-05 (1.98e-06)	1.54	58	191	3407	20387	negative	<a href="#">view details</a>
Expression data from untreated CLL patients	chronic lymphocytic leukemia	0.00065 (2.89e-05)	1.76	40	191	2200	20386	negative	<a href="#">view details</a>
An expression signature for p53 in breast cancer predicts mutation status, transcriptional effects, and patient survival	Breast cancer	0.000808 (5.39e-05)	2.05	30	187	967	12940	negative	<a href="#">view details</a>
Prediction of survival in diffuse large B cell lymphoma treated with chemotherapy plus Rituximab	diffuse large B cell lymphoma	0.0018 (0.00016)	1.08	74	191	5432	20387	positive	<a href="#">view details</a>
Whole-transcript expression data for liposarcoma	liposarcoma	0.0031 (0.00035)	1.47	44	187	1827	12940	negative	<a href="#">view details</a>

(truncated for presentation purpose)

GEO dataset	Cancer Type	GENE (Probe ID)	P-value	Effect Sign	View Details
human meningioma fixed tumour tissue	meningioma	352	2.71e-10	Negative	<a href="#">view details</a>
discovery cohort for genomic predictor of response and survival following neoadjuvant taxane-anthracycline chemotherapy in breast cancer	breast cancer	201984_S_AT	6.28e-05	Negative	<a href="#">view details</a>
whole-transcript expression data for liposarcoma	liposarcoma	211607_X_AT	0.000485	Positive	<a href="#">view details</a>
validation cohort for genomic predictor of response and survival following neoadjuvant taxane-anthracycline chemotherapy in breast cancer	breast cancer	201983_S_AT	0.0017	Negative	<a href="#">view details</a>
an eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia	chronic lymphocytic leukemia	201983_S_AT	0.00832	Negative	<a href="#">view details</a>
metastasis gene expression profile predicts recurrence and death in colon cancer patients (moffitt samples)	colon cancer	210984_X_AT	0.0113	Positive	<a href="#">view details</a>
molecular characterization of breast cancer subtypes derived from joint analysis of high throughput miRNA and mRNA data	breast cancer	A_23_P215790	0.0126	Negative	<a href="#">view details</a>
relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis	lung cancer	A_23_P215790	0.0213	Negative	<a href="#">view details</a>

(truncated for presentation purpose)



**B**

#	MicroRNA	probe ID	PVAL*	Effect sign**	GEO ID	TYPE	DATASET TITLE	Details
1	HSA-MIR-21	hsa-miR-21	0.000322 (0.0338)	positive	SSE19783	Disease Free Survival	"Molecular Characterization of Breast Cancer Subtypes from Joint Analysis of High Throughput mRNA and mRNA Data (III)"	<a href="#">view details</a>
2	HSA-MIR-21	27165	0.00445 (0.1635)	positive	SSE59334	Overall Survival	"microRNA expression profiling and its integration with mRNA in metastatic melanoma reveal associations with BRAF mutation status and patient prognosis"	<a href="#">view details</a>
3	HSA-MIR-21	5740	0.00667 (0.1635)	positive	SSE17405	Overall Survival	"Global microRNA expression profiling of high-risk ER+ breast cancers from patients receiving adjuvant Tamoxifen mono-therapy: a DECG study"	<a href="#">view details</a>
4	HSA-MIR-21	62462	0.00729 (0.1635)	positive	SSE59334	Overall Survival	"microRNA expression profiling and its integration with mRNA in metastatic melanoma reveal associations with BRAF mutation status and patient prognosis"	<a href="#">view details</a>
5	HSA-MIR-21	35988	0.00972 (0.1635)	positive	SSE59334	Overall Survival	"microRNA expression profiling and its integration with mRNA in metastatic melanoma reveal associations with BRAF mutation status and patient prognosis"	<a href="#">view details</a>
6	HSA-MIR-21	9706	0.00997 (0.1635)	positive	SSE59334	Overall Survival	"microRNA expression profiling and its integration with mRNA in metastatic melanoma reveal associations with BRAF mutation status and patient prognosis"	<a href="#">view details</a>

(truncated for presentation purpose)

**Figure 6: Exemplary results generated in the PPISURV and MIRUMIR databases. A.** Results generated with the PPISURV. Survival analysis shown for representative EGFR and its interactome. From the top: the first table depicts the summary of EGFR interactions that are annotated according to different interactomes across the available datasets. The last column of the table provides a link for more detailed characteristics of a selected interactome (shown in the second table). It includes the results of the analysis of the influence of the particular interactome on survival determined for all of the available datasets. The third table presents datasets on the direct correlation between EGFR expression and survival. The last column of the table is a link for the visualization of the data in the Kaplan-Meier graph. The exemplary graph shows the influence of EGFR expression on survival in lung adenocarcinoma patients. **B.** Results generated with MIRUMIR. The table shows a summary analysis for a representative microRNA-21 on the influence of its expression on survival in a specific cancer type. The inset represents the Kaplan-Meier graph of the effect of the microRNA expression on disease-free survival in breast cancer.

of cancer-associated genes that are known as drivers. IntOGen is one of the most dynamically developing and updating oncogenomic browsers.

In the initial release of the browser, catalogued cancer data were provided in a set of three integrated web-based sub-portals, namely, the IntOGen Arrays [67], IntOGen TCGA [68], and IntOGen Mutations [69], which allowed the browsing of visualized cancer data from different perspectives. The first sub-portal, i.e., the IntOGen Arrays, exploited cancer data on genome-wide expression and copy number for analyses aimed at selecting genes and molecular pathways that are associated with specific cancer types and subtypes [67]. Analyses provided by the other two IntOGen sub-portals were performed on a partially different set of oncogenomic data but with the use of a similar rationale. In the IntOGen TCGA, the set of somatic sequence alterations identified by exome sequencing of over 3,000 tumors from 12 cancer types (TCGA pan cancer data) was used for analyses focused on the identification of cancer-associated genes, i.e., drivers [68]. The IntOGen Mutations was focused on the evaluation of the role of somatic sequence variants in carcinogenesis and the identification of cancer drivers. In addition to the TCGA data, this sub-portal took advantage of the results from other large projects, e.g., the ICGC. The portal provided results obtained via the analyses of over 4,500 cancer exomes/genomes from 13 cancer types [69]. The results previously gathered in the interactive web-based platforms are currently available in the form of downloadable databases at the IntOGen site [66].

The introduction of a new release of IntOGen (release 2014.12) was aimed at building a bridge between molecular oncogenomics and clinical practice (the personalization of medicine) [70]. Nuria Lopez-Bigas and other co-authors of the browser proposed a strategy of “*in silico* prescription” of tailored anticancer therapy. In the first stage of the strategy, a secondary computational analysis of oncogenomic data from 6,792 patients of 28 different cancer types was performed. The analysis was focused on the evaluation of the role of somatic sequence alterations (including simple somatic variants, copy number alterations and fusion events) in carcinogenesis and the identification of cancer drivers. The drivers were selected when focusing on the following factors: mutation frequency in comparison to background (MutSigCV tool [47]), the presence of highly functional mutations (Oncodrive FM tool [71]), and regional clustering of mutations (Oncodrive CLUST tool [72]) [68, 70]. Although, all of the above tools take advantage of various algorithms and statistical methods, they all are based on similar principles and utilize similar oncogenic gene features. It is important to note that all of the implemented algorithms are supported by appropriate statistical tests. Information about the 459 identified driver genes, including their “mode of action” [loss-of-function (LoF), gain-of-function (GoF) or switch-of-function (SoF)] as

assessed with the use of the OncodriveROLE tool [73], is deposited in the Cancer Drivers Database. It can be either interactively visualized in the IntOGen web site [66] or downloaded for external analysis. In further stages of the strategy, Rubio-Perez and colleagues created the Cancer Drivers Actionability Database, which catalogues the already available and candidate therapies (under preliminary research or clinical trials) that are tailored to the cancer genomes of patients who were analyzed in the first stage. The Cancer Drivers Actionability Database can also be downloaded from the IntOGen website [66]. Additionally, the IntOGen portal can be exploited for the analysis of external data in the context of a single tumor or a cohort of tumors.

IntOGen is increasingly used by scientists from various cancer-associated fields for confirmation or identification of a potential driver role of genes of interest (e.g., selected based on experimental results) [74-78]. For example, Kovac and colleagues used IntOGen and MutSigCV programs for computational validation of 20 candidate papillary renal cell carcinoma (pRCC)-specific driver genes, which were selected based on the sequencing analysis of 31 exomes or genomes of pRCCs. The computational analysis of TCGA pRCC data for somatic single nucleotide variants (SNVs) in the candidate genes revealed significantly mutated genes and confirmed *SETD2*, *BAP1*, *NFE2L2* and *CUL3* as drivers, with a more modest degree of support for some other genes from a set of experimentally predefined candidates [74].

### BioProfiling.de portal

The BioProfiling.de portal [79, 80] contains three distinct databases: PPISURV [81, 82], MIRUMIR [83, 84], and DRUGSURV [85, 86]. The main purpose of PPISURV [81, 82] is the identification of important cancer-associated genes that do not have direct impact on the cancer survival outcome but nevertheless affect cancer by various interactions with other genes. Such a map of connections is called a “gene interactome”; it is created based on several external databases, which deposit information about the following: direct protein interactions (deposited in the IntAct Molecular Interaction Database [87]), regulatory and signaling pathways (Reactome, NCI Pathway Interaction Database, and HumanCyc databases) [21, 36, 88], and protein post-translational modifications (PhosphoSitePlus database) [89]. PPISURV allows users to analyze the influence of the gene interactome as well as a gene of interest on survival (Figure 6A). These analyses are performed with the use of over 40 whole transcriptome expression studies that were performed with the use of approximately 8,000 samples that represent 17 types of cancer.

The MIRUMIR provides a similar type of analysis as the PPISURV; however, it is focused on the impact of specific microRNA gene expression on survival in specific



cancer types. Either MIRUMIR or PPISURV enable the visualization of survival data via Kaplan-Meier graphs, showing the influence of the expression of a gene of interest on survival in a specific cancer type (Figure 6B).

The third database that is incorporated in the BioProfiling.de portal is DRUGSURV [85]. DRUGSURV provides the opportunity to explore the survival effect of expression alterations of genes that are known to be modulated by a selected drug. This database includes information about approximately 1,700 drugs that were approved by the Food and Drug Administration (FDA), along with approximately 5,000 experimental drugs. A specific drug, cancer type or gene can be queried and investigated in terms of its anticancer potential.

The advantage of the tools that are available in the BioProfiling.de portal is that all of them provide results that are supported by appropriate statistical analysis (the R statistical package), which is not always available for the tools in the other oncogenomic portals. A false discovery rate control procedure is implemented to adjust the p-values when there is multiple testing.

The usefulness of the above-mentioned databases has been confirmed in a number of publications (e.g., [63, 90-98]). For example, Schittek et al., [90] used the PPISURV to perform survival analysis on patients who were stratified based on the expression of *CK1* gene isoforms (*CSNK1A1*, *CSNK1D*, and *CSNK1E*) in different cancers. In another study [99], MIRUMIR was used to evaluate the potential of miR-200c and miR-141 to serve as biomarkers in breast cancer.

## CONCLUSIONS

Since the initiation of large-scale oncogenomic projects, a variety of databases and web-based portals have been created to enable the interactive visualization and interpretation of the abundant genome-wide cancer data. The range of available web-based portals is not limited to those described in our review. Among other noteworthy portals that provide sets of visualization tools that are helpful for oncogenomic data analysis are Oasis [100, 101], Oncomine [102, 103], Cancer Genetics Web [104], and CaSNP [105, 106]. In short, Oasis is a recently launched open-access web portal for explanatory analysis of cancer data. This portal was developed based on a custom version of the BioMart framework that was designed for oncogenomics data analysis, and it provides a unique set of visualization tools. Oncomine is another portal that provides useful visualization and analytical tools, which can browse and analyze over 715 expression and sequence alteration datasets. The Cancer Genetics Web is a web-based tool that can be used to gather literature that is related to a specific cancer type/predisposing syndrome or a gene of interest that is potentially associated with cancer. This tool provides a short summary about a disease and gene of interest,

as well as a list of the latest publications and useful external links. Another interesting feature of the Cancer Genetics Web portal is a colorful panel of summarizing keywords that are available for each gene. The fourth portal is CaSNP, which gathers the results of genome-wide CNA profiling that was performed with the use of SNP arrays across 34 different cancer types. In most of the portals, the datasets and methods that are applied in their analyses and graphical presentations are continually updated. As a result, the portals deliver complex pictures of cancer genome alterations and their potential impact on cancer molecular pathogenesis. Importantly, the portals are very intuitive and address a wide community of researchers, who are not necessarily familiar with advanced computational methods. The users can take advantage of oncogenomic portals to further explore the cancer molecular basis and select new candidate cancer-associated genes for experimental validation. Regardless of current interest in exploring data that is gathered in the portals, the usefulness of the tools that are available in the oncogenomic portals will be verified in time by the users. Ultimately, it is expected that the utilization of the portals for the analysis of expanding oncogenomic data will make a substantial contribution to our understanding of cancer molecular etiology and the translation of extended cancer genomic knowledge into clinical practice.

## ACKNOWLEDGMENTS

This work was supported by the National Science Centre grant 2011/01/B/NZ5/02773.

## CONFLICTS OF INTEREST

All authors declare no conflict of interest.

## REFERENCES

1. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458: 719-724.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144: 646-674.
3. Chin L, Hahn WC, Getz G, Meyerson M. Making sense of cancer genomic data. *Genes Dev*. 2011; 25: 534-555.
4. Dickson D. Wellcome funds cancer database. *Nature*. 1999; 401: 729.
5. Cancer Genome Project. <https://www.sanger.ac.uk/research/projects/cancergenome/>.
6. Mc Lendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogiannakis GM, Olson JJ, Mikkelsen T, Lehman N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455: 1061-1068.
7. Collins FS, Barker AD. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart

- a new course across the complex landscape of human malignancies. *Sci Am.* 2007; 296: 50-57.
8. The Cancer Genome Atlas. <http://cancergenome.nih.gov/>.
  9. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, et al. International network of cancer genome projects. *Nature.* 2010; 464: 993-998.
  10. The International Cancer Genome Consortium. <https://icgc.org/>.
  11. Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N. Visualizing multidimensional cancer genomics data. *Genome Med.* 2013; 5: 9.
  12. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, et al. The landscape of somatic copy-number alteration across human cancers. *Nature.* 2010; 463: 899-905.
  13. Tumorscape. <http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf>.
  14. Zhu J, Sanborn JZ, Benz S, Szeto C, Hsu F, Kuhn RM, Karolchik D, Archie J, Lenburg ME, Esserman LJ, Kent WJ, Haussler D, Wang T. The UCSC Cancer Genomics Browser. *Nat Methods.* 2009; 6: 239-240.
  15. Sanborn JZ, Benz SC, Craft B, Szeto C, Kober KM, Meyer L, Vaske CJ, Goldman M, Smith KE, Kuhn RM, Karolchik D, Kent WJ, Stuart JM, et al. The UCSC Cancer Genomics Browser: update 2011. *Nucleic Acids Res.* 2011; 39: D951-959.
  16. Goldman M, Craft B, Swatloski T, Ellrott K, Cline M, Diekhans M, Ma S, Wilks C, Stuart J, Haussler D, Zhu J. The UCSC Cancer Genomics Browser: update 2013. *Nucleic Acids Res.* 2013; 41: D949-954.
  17. Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D, Zhu J. Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. *Sci Rep.* 2013; 3: 2652.
  18. Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, Haussler D, Zhu J. The UCSC Cancer Genomics Browser: update 2015. *Nucleic Acids Res.* 2015; 43: D812-817.
  19. UCSC Cancer Genomics Browser. <https://genome-cancer.ucsc.edu>.
  20. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics.* 2010; 26: i237-245.
  21. NCI Pathway Interaction Database. <http://pid.nci.nih.gov/>.
  22. Wu Y, Liu H, Shi X, Yao Y, Yang W, Song Y. The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. *Oncotarget.* 2015; 6: 9160-9172. Doi: 10.18632/oncotarget.3247.
  23. Mahauad-Fernandez WD, Borchering NC, Zhang W, Okeoma CM. Bone marrow stromal antigen 2 (BST-2) DNA is demethylated in breast tumors and breast cancer cells. *PLoS One.* 2015; 10: e0123931.
  24. Sokhi UK, Bacolod MD, Emdad L, Das SK, Dumur CI, Miles MF, Sarkar D, Fisher PB. Analysis of global changes in gene expression induced by human polynucleotide phosphorylase (hPNPase(old-35)). *J Cell Physiol.* 2014; 229: 1952-1962.
  25. Brown DV, Daniel PM, D'Abaco GM, Gogos A, Ng W, Morokoff AP, Mantamadiotis T. Coexpression analysis of CD133 and CD44 identifies proneural and mesenchymal subtypes of glioblastoma multiforme. *Oncotarget.* 2015; 6: 6267-6280. Doi: 10.18632/oncotarget.3365.
  26. Yin S, Yang J, Lin B, Deng W, Zhang Y, Yi X, Shi Y, Tao Y, Cai J, Wu CI, Zhao G, Hurst LD, Zhang J, et al. Exome sequencing identifies frequent mutation of MLL2 in non-small cell lung carcinoma from Chinese patients. *Sci Rep.* 2014; 4: 6036.
  27. Chen X, Meng J, Yue W, Yu J, Yang J, Yao Z, Zhang L. Fibulin-3 suppresses Wnt/beta-catenin signaling and lung cancer invasion. *Carcinogenesis.* 2014; 35: 1707-1716.
  28. Zhao C, Qiao Y, Jonsson P, Wang J, Xu L, Rouhi P, Sinha I, Cao Y, Williams C, Dahlman-Wright K. Genome-wide profiling of AP-1-regulated transcription provides insights into the invasiveness of triple-negative breast cancer. *Cancer Res.* 2014; 74: 3983-3994.
  29. Poaty H, Coullin P, Peko JF, Dessen P, Diatta AL, Valent A, Leguern E, Prevot S, Gombé-Mbalawa C, Candelier JJ, Picard JY, Bernheim A. Genome-wide high-resolution aCGH analysis of gestational choriocarcinomas. *PLoS One.* 2012; 7: e29426.
  30. Veeraraghavan J, Tan Y, Cao XX, Kim JA, Wang X, Chamness GC, Maiti SN, Cooper LJ, Edwards DP, Contreras A, Hilsenbeck SG, Chang EC, Schiff R, et al. Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. *Nat Commun.* 2014; 5: 4577.
  31. Xena. <http://xena.ucsc.edu/>.
  32. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B, Wong-Erasmus M, Yao L, Kasprzyk A. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database (Oxford).* 2011; 2011: bar026.
  33. ICGC Data Portal. <https://dcc.icgc.org>.
  34. Ensembl. <http://www.ensembl.org/index.html>.
  35. OMIM. <http://www.omim.org/>.
  36. Reactome. <http://www.reactome.org/>.
  37. COSMIC. <http://www.sanger.ac.uk/genetics/CGP/cosmic>.
  38. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature.* 2013; 494: 366-370.
  39. Bamford S, Dawson E, Forbes S, Clements J, Pettett R,

- Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, Wooster R. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*. 2004; 91: 355-358.
40. Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, Flanagan A, Teague J, Wooster R, Futreal PA, Stratton MR. Cosmic 2005. *Br J Cancer*. 2006; 94: 318-322.
  41. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet*. 2008; Chapter 10: Unit 10.11.
  42. Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, Futreal PA. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res*. 2010; 38: D652-657.
  43. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*. 2011; 39: D945-950.
  44. Gaston D, Hansford S, Oliveira C, Nightingale M, Pinheiro H, Macgillivray C, Kaurah P, Rideout AL, Steele P, Soares G, Huang WY, Whitehouse S, Blowers S, et al. Germline mutations in MAP3K6 are associated with familial gastric cancer. *PLoS Genet*. 2014; 10: e1004669.
  45. Kelley RK, Magbanua MJ, Butler TM, Collisson EA, Hwang J, Sidiropoulos N, Evason K, McWhirter RM, Hameed B, Wayne EM, Yao FY, Venook AP, Park JW. Circulating tumor cells in hepatocellular carcinoma: a pilot study of detection, enumeration, and next-generation sequencing in cases and controls. *BMC Cancer*. 2015; 15: 206.
  46. Chen Y, Huang R, Ding J, Ji D, Song B, Yuan L, Chang H, Chen G. Multiple myeloma acquires resistance to EGFR inhibitor via induction of pentose phosphate pathway. *Sci Rep*. 2015; 5: 9925.
  47. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013; 499: 214-218.
  48. Ostrow SL, Barshir R, DeGregori J, Yeger-Lotem E, Hershberg R. Cancer evolution is associated with pervasive positive selection on globally expressed genes. *PLoS Genet*. 2014; 10: e1004239.
  49. Li Y, Liang C, Wong KC, Jin K, Zhang Z. Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. *Nucleic Acids Res*. 2014; 42: e76.
  50. Gulukota K, Helseth DL, Jr., Khandekar JD. Direct observation of genomic heterogeneity through local haplotyping analysis. *BMC Genomics*. 2014; 15: 418.
  51. Arcila ME, Drilon A, Sylvester BE, Lovly CM, Borsu L, Reva B, Kris MG, Solit DB, Ladanyi M. MAP2K1 (MEK1) Mutations Define a Distinct Subset of Lung Adenocarcinoma Associated with Smoking. *Clin Cancer Res*. 2015; 21: 1935-1943.
  52. Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, Shin A, Jee SH, Kim DH, Cai Q, Long J, Shi J, Wen W, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet*. 2014; 46: 533-542.
  53. Hu B, Ying X, Wang J, Piriyaopongsa J, Jordan IK, Sheng J, Yu F, Zhao P, Li Y, Wang H, Ng WL, Hu S, Wang X, et al. Identification of a tumor-suppressive human-specific microRNA within the FHIT tumor-suppressor gene. *Cancer Res*. 2014; 74: 2283-2294.
  54. Shtivelman E, Davies MQ, Hwu P, Yang J, Lotem M, Oren M, Flaherty KT, Fisher DE. Pathways and therapeutic targets in melanoma. *Oncotarget*. 2014; 5: 1701-1752. Doi: 10.18632/oncotarget.1892.
  55. Riester M, Werner L, Bellmunt J, Selvarajah S, Guancial EA, Weir BA, Stack EC, Park RS, O'Brien R, Schutz FA, Choueiri TK, Signoretti S, Lloreta J, et al. Integrative analysis of 1q23.3 copy-number gain in metastatic urothelial carcinoma. *Clin Cancer Res*. 2014; 20: 1873-1883.
  56. Li K, Liu Y, Zhou Y, Zhang R, Zhao N, Yan Z, Zhang Q, Zhang S, Qiu F, Xu Y. An integrated approach to reveal miRNAs' impacts on the functional consequence of copy number alterations in cancer. *Sci Rep*. 2015; 5: 11567.
  57. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2: 401-404.
  58. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013; 6: p11.
  59. cBioPortal. <http://www.cbioportal.org>.
  60. Lin DC, Xu L, Chen Y, Yan H, Hazawa M, Doan N, Said JW, Ding LW, Liu LZ, Yang H, Yu S, Kahn M, Yin D, et al. Genomic and Functional Analysis of the E3 Ligase PARK2 in Glioma. *Cancer Res*. 2015; 75: 1815-1827.
  61. Lu X, Megchelenbrink W, Notebaart RA, Huynen MA. Predicting human genetic interactions from cancer genome evolution. *PLoS One*. 2015; 10: e0125795.
  62. Medico E, Russo M, Picco G, Cancelliere C, Valtorta E, Corti G, Buscarino M, Isella C, Lamba S, Martinoglio B, Veronese S, Siena S, Sartore-Bianchi A, et al. The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. *Nat Commun*. 2015; 6:

7002.

63. Czubak K, Lewandowska MA, Klonowska K, Roszkowski K, Kowalewski J, Figlerowicz M, Kozłowski P. High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer. *Oncotarget*. 2015; 6: 23399-416. Doi: 10.18632/oncotarget.4351.
64. Powers S. Cooperation between MYC and companion 8q genes in hepatocarcinogenesis. *Hepatology*. 2015; 61: 757-758.
65. Chen H, Song R, Wang G, Ding Z, Yang C, Zhang J, Zeng Z, Rubio V, Wang L, Zu N, Weiskoff AM, Minze LJ, Jeyabal PV, et al. OLA1 regulates protein synthesis and integrated stress response by inhibiting eIF2 ternary complex formation. *Sci Rep*. 2015; 5: 13241.
66. IntOGen. <http://www.intogen.org/>.
67. Gundem G, Perez-Llamas C, Jene-Sanz A, Kedzierska A, Islam A, Deu-Pons J, Furney SJ, Lopez-Bigas N. IntOGen: integration and data mining of multidimensional oncogenomic data. *Nat Methods*. 2010; 7: 92-93.
68. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandath C, Reimand J, Lawrence MS, Getz G, Bader GD, Ding L, Lopez-Bigas N. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep*. 2013; 3: 2650.
69. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N. IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods*. 2013; 10: 1081-1082.
70. Rubio-Perez C, Tamborero D, Schroeder MP, Antolin AA, Deu-Pons J, Perez-Llamas C, Mestres J, Gonzalez-Perez A, Lopez-Bigas N. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell*. 2015; 27: 382-396.
71. Gonzalez-Perez A, Lopez-Bigas N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res*. 2012; 40: e169.
72. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*. 2013; 29: 2238-2244.
73. Schroeder MP, Rubio-Perez C, Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. *Bioinformatics*. 2014; 30: i549-555.
74. Kovac M, Navas C, Horswell S, Salm M, Bardella C, Rowan A, Stares M, Castro-Giner F, Fisher R, de Bruin EC, Kovacova M, Gorman M, Makino S, et al. Recurrent chromosomal gains and heterogeneous driver mutations characterise papillary renal cancer evolution. *Nat Commun*. 2015; 6: 6336.
75. Pickering CR, Zhou JH, Lee JJ, Drummond JA, Peng SA, Saade RE, Tsai KY, Curry JL, Tetzlaff MT, Lai SY, Yu J, Muzny DM, Doddapaneni H, et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin Cancer Res*. 2014; 20: 6582-6592.
76. Moncunill V, Gonzalez S, Bea S, Andrieux LO, Salaverria I, Royo C, Martinez L, Puiggros M, Segura-Wang M, Stutz AM, Navarro A, Royo R, Gelpi JL, et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat Biotechnol*. 2014; 32: 1106-1112.
77. Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, Pooley KA, Pritchard AL, Tiffen JC, Petljak M, Palmer JM, Symmons J, Johansson P, et al. POT1 loss-of-function variants predispose to familial melanoma. *Nat Genet*. 2014; 46: 478-481.
78. Yang CY, Lu RH, Lin CH, Jen CH, Tung CY, Yang SH, Lin JK, Jiang JK, Lin CH. Single nucleotide polymorphisms associated with colorectal cancer susceptibility and loss of heterozygosity in a Taiwanese population. *PLoS One*. 2014; 9: e100060.
79. Antonov AV. BioProfiling.de: analytical web portal for high-throughput cell biology. *Nucleic Acids Res*. 2011; 39: W323-327.
80. BioProfiling.de. <http://bioprofiling.de/>.
81. Antonov AV, Krestyaninova M, Knight RA, Rodchenkov I, Melino G, Barlev NA. PPISURV: a novel bioinformatics tool for uncovering the hidden role of specific genes in cancer survival outcome. *Oncogene*. 2014; 33: 1621-1628.
82. PPISURV. <http://bioprofiling.de/GEO/PPISURV/ppisurv.html>.
83. Antonov AV, Knight RA, Melino G, Barlev NA, Tsvetkov PO. MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. *Cell Death Differ*. 2013; 20: 367.
84. MIRUMIR. <http://www.bioprofiling.de/GEO/MIRUMIR/mirumir.html>.
85. Amelio I, Gostev M, Knight RA, Willis AE, Melino G, Antonov AV. DRUGSURV: a resource for repositioning of approved and experimental drugs in oncology based on patient survival information. *Cell Death Dis*. 2014; 5: e1051.
86. DRUGSURV. <http://www.bioprofiling.de/GEO/DRUGSURV/index.html>.
87. IntAct Molecular Interaction Database. <http://www.ebi.ac.uk/intact/>.
88. HumanCyc. <http://humancyc.org/>.
89. PhosphoSitePlus. <http://www.phosphosite.org>.
90. Schitteck B, Sinnberg T. Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis. *Mol Cancer*. 2014; 13: 231.
91. Althubiti M, Lezina L, Carrera S, Jukes-Jones R, Giblett SM, Antonov A, Barlev N, Saldanha GS, Pritchard CA, Cain K, Macip S. Characterization of novel markers of senescence and their prognostic potential in cancer. *Cell Death Dis*. 2014; 5: e1528.
92. Antonov A, Agostini M, Morello M, Minieri M, Melino G,



- Amelio I. Bioinformatics analysis of the serine and glycine pathway in cancer cells. *Oncotarget*. 2014; 5: 11004-11013. Doi: 10.18632/oncotarget.2668.
93. Salah Z, Arafeh R, Maximov V, Galasso M, Khawaled S, Abou-Sharieha S, Volinia S, Jones KB, Croce CM, Aqeilan RI. miR-27a and miR-27a\* contribute to metastatic properties of osteosarcoma cells. *Oncotarget*. 2015; 6: 4920-4935. Doi: 10.18632/oncotarget.3025.
  94. Suh SS, Yoo JY, Cui R, Kaur B, Huebner K, Lee TK, Aqeilan RI, Croce CM. FHIT suppresses epithelial-mesenchymal transition (EMT) and metastasis in lung cancer through modulation of microRNAs. *PLoS Genet*. 2014; 10: e1004652.
  95. Braun FK, Mathur R, Sehgal L, Wilkie-Grantham R, Chandra J, Berkova Z, Samaniego F. Inhibition of methyltransferases accelerates degradation of cFLIP and sensitizes B-cell lymphoma cells to TRAIL-induced apoptosis. *PLoS One*. 2015; 10: e0117994.
  96. Li XJ, Ren ZJ, Tang JH. MicroRNA-34a: a potential therapeutic target in human cancer. *Cell Death Dis*. 2014; 5: e1327.
  97. Tsimokha AS, Kulichkova VA, Karpova EV, Zaykova JJ, Aksenov ND, Vasilishina AA, Kropotov AV, Antonov A, Barlev NA. DNA damage modulates interactions between microRNAs and the 26S proteasome. *Oncotarget*. 2014; 5: 3555-3567. Doi: 10.18632/oncotarget.1957.
  98. Bongiorno-Borbone L, Giacobbe A, Compagnone M, Eramo A, De Maria R, Peschiaroli A, Melino G. Antitumoral effect of desmethylclomipramine in lung cancer stem cells. *Oncotarget*. 2015; 6: 16926-16938. Doi: 10.18632/oncotarget.4700.
  99. Antolin S, Calvo L, Blanco-Calvo M, Santiago MP, Lorenzo-Patino MJ, Haz-Conde M, Santamarina I, Figueroa A, Anton-Aparicio LM, Valladares-Ayerbes M. Circulating miR-200c and miR-141 and outcomes in patients with breast cancer. *BMC Cancer*. 2015; 15: 297.
  100. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, Bardou P, Beck T, Blake A, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res*. 2015; 43: W589-598.
  101. OASIS. <http://www.oasis-genomics.org/>.
  102. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004; 6: 1-6.
  103. Oncomine. <https://www.oncomine.org>.
  104. Cancer Genetics Web. <http://www.cancerindex.org/geneweb/>.
  105. Cao Q, Zhou M, Wang X, Meyer CA, Zhang Y, Chen Z, Li C, Liu XS. CaSNP: a database for interrogating copy number alterations of cancer genome from SNP array data. *Nucleic Acids Res*. 2011; 39: D968-974.
  106. CaSNP. <http://cistrome.dfci.harvard.edu/CaSNP/>.