Published online 8 November 2019 *Nucleic Acids Research, 2020, Vol. 48, Database issue* ***D863–D870***
doi: 10.1093/nar/gkz964

DriverDBv3: a multi-omics database for cancer driver gene research

Shu-Hsuan Liu[1], Pei-Chun Shen[1], Chen-Yang Chen[2], An-Ni Hsu[1], Yi-Chun Cho[1], Yo-Liang Lai[1,3], Fang-Hsin Chen[4,5,6], Chia-Yang Li[7], Shu-Chi Wang[8], Ming Chen[9], I-Fang Chung[10] and Wei-Chung Cheng[1,11,*]

[1]Graduate Institute of Biomedical Science, China Medical University, Taichung 40403, Taiwan, [2]Cytoaurora Biotechnologies, Inc. Hsinchu Science Park, Hsinchu 30261, Taiwan, [3]Department of Radiation Oncology, China Medical University Hospital, Taichung 40403, Taiwan, [4]Department of Medical Imaging and Radiological Sciences, Chang Gung University, Taoyuan 33302, Taiwan, [5]Department of Radiation Oncology, Chang Gung Memorial Hospital at Linkou, Taoyuan 33302, Taiwan, [6]Institute for Radiological Research, Chang Gung University and Chang Gung Memorial Hospital, Taoyuan 33302, Taiwan, [7]Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung 80708, Taiwan, [8]Department of Medical Laboratory Science and Biotechnology, College of Health Sciences, Kaohsiung Medical University, Kaohsiung 80708, Taiwan, [9]Center for Medical Genetics, Changhua Christian Hospital, Changhua 50006, Taiwan, [10]Institute of BioMedical Informatics, National Yang-Ming University, Taipei 11221, Taiwan and [11]Research Center for Tumor Medical Science, China Medical University, Taichung 40403, Taiwan

Received September 14, 2019; Revised October 09, 2019; Editorial Decision October 10, 2019; Accepted November 06, 2019

ABSTRACT

An integrative multi-omics database is needed urgently, because focusing only on analysis of one-dimensional data falls far short of providing an understanding of cancer. Previously, we presented DriverDB, a cancer driver gene database that applies published bioinformatics algorithms to identify driver genes/mutations. The updated DriverDBv3 database (http://ngs.ym.edu.tw/driverdb) is designed to interpret cancer omics' sophisticated information with concise data visualization. To offer diverse insights into molecular dysregulation/dysfunction events, we incorporated computational tools to define CNV and methylation drivers. Further, four new features, CNV, Methylation, Survival, and miRNA, allow users to explore the relations from two perspectives in the 'Cancer' and 'Gene' sections. The 'Survival' panel offers not only significant survival genes, but gene pairs synergistic effects determine. A fresh function, 'Survival Analysis' in 'Customized-analysis,' allows users to investigate the co-occurring events in user-defined gene(s) by mutation status or by expression in a specific patient group. Moreover, we redesigned the web interface and provided interactive figures to interpret cancer omics' sophisticated information, and also constructed a Summary panel in the 'Cancer' and 'Gene' sections to visualize the features on multi-omics levels concisely. DriverDBv3 seeks to improve the study of integrative cancer omics data by identifying driver genes and contributes to cancer biology.

INTRODUCTION

With the advanced development of next generation sequencing (NGS) technology and progressively decreasing economic cost and time, the amount of sequencing data has led to the era of 'big data science' (1). Several large sequencing projects have been accomplished in recent years, such as the 1000 Genome Project, The Cancer Genome Atlas (TCGA), and the International Cancer Genome Consortium (ICGC) (2). TCGA is one of the most comprehensive and largest sequencing projects, which the National Institutes of Health (NIH) launched in 2006. TCGA contains multi-omics data at different molecular levels, and provides >10 000 patients with approximately 33 types of cancer (3). Further, the 21 European members have just signed the 1 million genome project that is intended to be completed in 2022 (4). These projects offer an excellent opportunity to disclose diverse molecular signatures in various disease types. The application of precision medicine to various cancer patients has become critical because of the disease's heterogeneous functions and morphology (5). Integrating multi-omics data is the key to link cancer genetics, clinical, and epidemiological information to ensure that pa-

*To whom correspondence should be addressed. Tel: +886 4 2205 2121 (Ext. 7820); Fax: +886 4 2233 3496; Email: cwc0702@gmail.com

tients obtain effective and proper diagnosis and therapeutics in precision medicine (6), and genomics data's utility has the potential to transition from research alone to wide use in healthcare (4). Apparently, the pressing challenge associated with the implementation of precision medicine is the way to integrate different multi-omics data to build a comprehensive picture of genomic biology.

Thus, an integrative multi-omics database is needed urgently because focusing only on analysis of one-dimensional data falls far short of providing an understanding of the entire cancer environment. Few databases have offered certain investigative approaches and visualizations to link different omics data, including LinkedOmics (7), TCOA (8), MEXPRESS (9) and cbioportal (10,11). However, many current studies still are addressing the need for varied interpretation and integration because of the disease's complexity. Hence, contributions in the field of interpreting multi-omics data stand in need of different perspectives.

Survival analysis is the foundational indicator of the clinical outcome in investigating the importance of specific molecules, and the extensive TCGA data also have the potential to establish the associations between molecular events and clinical use (3). Various databases, including PROGgene (12), PrognoScan (13) and SurvExpress (14), have contributed to single gene survival analysis based on RNA expression profiles, yet few databases have provided clinical information for user-defined subgroups of specific cancer patients. Further, analyses of a single target gene frequently are inadequate in cancer research, because a large number of studies has shown co-occurring events of two or more targets in aberrant expression that lead ultimately to poor survival outcomes (15–19). Amelio *et al.* also developed a web tool, SynTarget, to examine the synergistic effect between two genes based on survival outcomes in several cancer datasets (20). However, there is no current data portal to explore the complicated relations between two/more gene targets and clinical outcome in a sub-group of patients.

DriverDB is a cancer driver gene database featured previously in 2014 and 2016, which applies published bioinformatics algorithms to dedicated driver gene/mutation identification. In this updated version, our goal is to interpret cancer omics' sophisticated information through concise data visualization. There are four major improvements in this version: First, we collected ∼11,000 copy number variation (CNV), ∼12 000 methylation, and ∼11 000 smRNA-seq datasets from the public domain. Further, 3000 RNA-seq and 2000 exome-seq datasets have been incorporated newly into DriverDBv3. Second, to offer diverse insights into molecular dysregulation/dysfunction events, we incorporated four computational tools that define CNV and methylation drivers, as well as multiple mutation tools into our analysis pipeline. Third, four new features, 'CNV,' 'methylation,' 'survival,' and 'miRNA,' in the 'Cancer' and 'Gene' sections allow users to obtain a more comprehensive picture of the relations from two perspectives. 'CNV' and 'methylation' display the tool-defined drivers in various cancers. 'Survival' offers not only significant survival genes, but gene pairs that have been determined to have synergistic effects. In 'miRNA,' cancer-related miRNAs are gathered to depict their interactions with driver genes Fourth, a new function, 'Survival Analysis' in 'Customized-analysis,' allows users to investigate user-defined gene(s)' survival significance by mutation status or by expression in a specific group of patients users can define according to dozens of clinical criteria. This new function allows users to establish the connection between molecular events and clinical practice. Moreover, we redesigned the DriverDB web interface and provided interactive figures that allow users to explore the information when the mouse moves to specific regions of an interactive plot. Users can investigate the data from different perspectives to produce views that are informative and easy to interpret. Further, our database incorporates the cancer-related genes that are defined in CGC (21) and NCG6.0 (22) to provide better illustrations of driver gene identification and increase our interpretation's importance.

MATERIALS AND METHODS

Data collection

TCGA's updated RNA sequencing and exome sequencing data were collected from the GDC data portal (https://portal.gdc.cancer.gov/), for which the data pre-processing approach is described in previous publications (23,24). Level 3 CNV data were downloaded by applying the TCGA2BED tool (25). Level 3 Methylation data were collected from firehose (https://gdac.broadinstitute.org/). TCGA clinical data were downloaded using an R package, 'TCGAbiolinks' (26). In addition, cancer-related genes are defined according to CGC, which was downloaded from COSMIC (https://cancer.sanger.ac.uk/census) and the NCG6.0 database.

Synergistic survival analysis

A synergistic effect of co-expression genes is defined as that in which two high-level genes co-expression results in very poor outcomes in survival, which has been observed in much research (15–19). Thus, we incorporate synergistic effect to improve the understanding of the relationships between genes and clinical outcome. To determine the genes with significant synergistic effects in expression level, we established an analytic model to evaluate the synergistic effects between co-expression driver genes and identified them according to the following two steps. Firstly, to filter the proper candidates for further calculation, three criteria were applied: (i) genes with a basal expression level were included, meaning that gene expression actually affects survival; (ii) the coefficient of gene expression variation (CV, defined as the standard deviation normalised to expression mean) >1, which is used to represent a larger dispersion in the sample variability and is used frequently to filter candidates, such as in Shukla *et al.*'s study (27) and (iii) single gene survival is significant with a log-rank P-value <0.05. By following these criteria, genes with basal expression, and significant variation and survival are filtered out. Secondly, to define the synergistic survival effect further, we considered the definition from previous studies and incorporated the difference when the hazard ratio (HR) between two genes combined is >1.5-fold of each gene. Because HR is a common indicator used to denote increased survival probability,

we considered that the difference in the HR is able to interpret the level of relative risk (28). Hence, by filtering highly distinguishable shifts in HR, the synergistic pairs can be defined.

CNV, MET and miRNA—define dysregulation events

The previous DriverDB focused on driver mutation identification, and was concerned with gene 'dysfunction' events in cancer studies. However, now we have included 'dysregulation' events by integrating multi-omics data, including CNV, methylation, and miRNA datasets. To characterise the 'dysregulation' event, aberrant molecular mechanisms, such as abnormal DNA methylation, plays the primary role in altering the transcription level (29,30). To define CNV and methylation dysregulation events, only tools that take both features, aberrant shifts and changes in gene expression, into consideration are applied. Therefore, iGC (31) and DIGGIT (32) are used to identify CNV dysregulation events, while methylmix (33,34) and ELMER (35) are used to define methylation dysregulation events. Our database integrates computational algorithms of CNV and methylation and presents them in different interpretations. Moreover, abnormal miRNA regulations' role has been studied well in cancer research for years, and thus, we incorporated the data from our previous database, YM500, to DriverDBv3 to establish the relations with negative correlation coefficients between driver genes and miRNA (36–38).

WEB INTERFACE

To investigate cancer's different molecular features, three main functions— 'Cancer,' 'Gene' and 'Customized-analysis'—are characterised in our web interface. These functions are provided to help users study multi-omics data from different perspectives. In addition, cancer-related genes from CGC and NCG6.0 database are provided in every network to allow users to inspect the relations between those cancer genes.

Cancer

The 'Cancer' function summarizes the results of driver genes' different molecular features calculated by published bioinformatics algorithms/tools for a specific cancer type/dataset. As shown in Figure 1, we offer multiple new sections of omics features and a summary in the first panel in the web. We present an outlined network that gathers mutation, CNV, methylation, and miRNA drivers defined in each feature and marked by different colors in the node (Figure 1A). Different panels by features, including mutation, CNV, methylation, survival, and miRNA, offer further the details and visualizations analyzed for a specific cancer type. Supplementary Figure S1 shows the top 30 crucial mutation drivers that multiple computational algorithms identify, including the following new tools: CoMET (39), Mutex (40) and DriverML (41). Novel features characterised, CNV, and Methylation panels demonstrated in similar fashion, are shown in Figure 1B, C, and D. We also highlight the top 30 drivers by using heatmap (Figure 1B) and percentage barcharts (Figure 1C), and perform locus enrichment to understand those regions that contain CNV/differentially methylated events, as shown in Figure 1D. Unlike determining cancer drivers, as previously mentioned, the 'Survival' panel in the 'Cancer' section provides distinct networks constructed by applying the gene pairs with synergistic effects to illustrate the synergistic relations between genes (Figure 1E). The survival genes with HR >1 and HR <1 are denoted as orange and green nodes, respectively. The two genes with synergistic effect defined in 'Synergistic Survival Analysis' are connected with a grey line. Kaplan Meier plots of each synergistic survival event also are provided, and are shown from two approaches: all high vs. others and four groups based on expression (all high, high/low, low/high, and all low), as shown in Figure 1F. The 'miRNA' panel is designed to illustrate the relation between differentially expressed genes and miRNAs (Supplementary Figure S2A).

Gene

The 'Gene' section is designed to illustrate different features of a user-selected gene at multi-omics levels. Compared to the previous DriverDB, we also provide four additional panels: 'CNV,' 'Methylation,' 'Survival' and 'miRNA.' Figure 2 incorporates visualizations added recently in the 'Gene' section. Figure 2A illustrates the summary features in the various cancer types for a single target gene selected. The 'CNV' and 'Methylation' panels exhibit bioinformatic algorithms' combined analytical results, which are shown in Figure 2B and Supplementary Figure S3A, respectively. As shown in Figure 2C, the scatter plot illustrates the correlation between gene expression (y-axis) and CNV value (x-axis). The left boxplot indicates the expression levels, as well as the bottom boxplot displays CNV values in each CNV type. The 'Methylation' panel also displays the relation between gene expression and beta value in a similar fashion (Supplementary Figure S3B). The 'Survival' panel manages single gene survival and survival of two genes with synergistic effects. The network of synergistic effects for a single target gene also is provided, as Figure 2D shows. The width of the lines indicates the number of cancers in which two genes have synergistic effects. The two directions of the HR (>1 or <1) are illustrated in grey and pink, respectively. The 'miRNA' panels displays the miRNAs regulating the selected target gene (Supplementary Figure S2B).

Customized analysis

'Customized Analysis' is a unique function provided in our database to investigate a specific group of patients by 'Survival analysis' and 'Driver gene identification' (Figure 3A). This allows researchers to select well-defined cancer samples based on one or multiple clinical criteria, and the selection panel for user-defined samples allows genes, datasets, and clinical criteria to be filtered (Supplementary Figure S4A). In contrast with the function of 'Driver gene identification', which was the 'Meta-analysis' function in the previous DriverDBv2, 'Survival analysis' helps researchers investigate the co-occurring events that affect patients' survival. Here, we provide different insights into survival analysis through two approaches, which are to stratify patients 'By expression' level or 'By mutation' status of the user-selected gene(s) in user-defined patients (Figure 3A). The

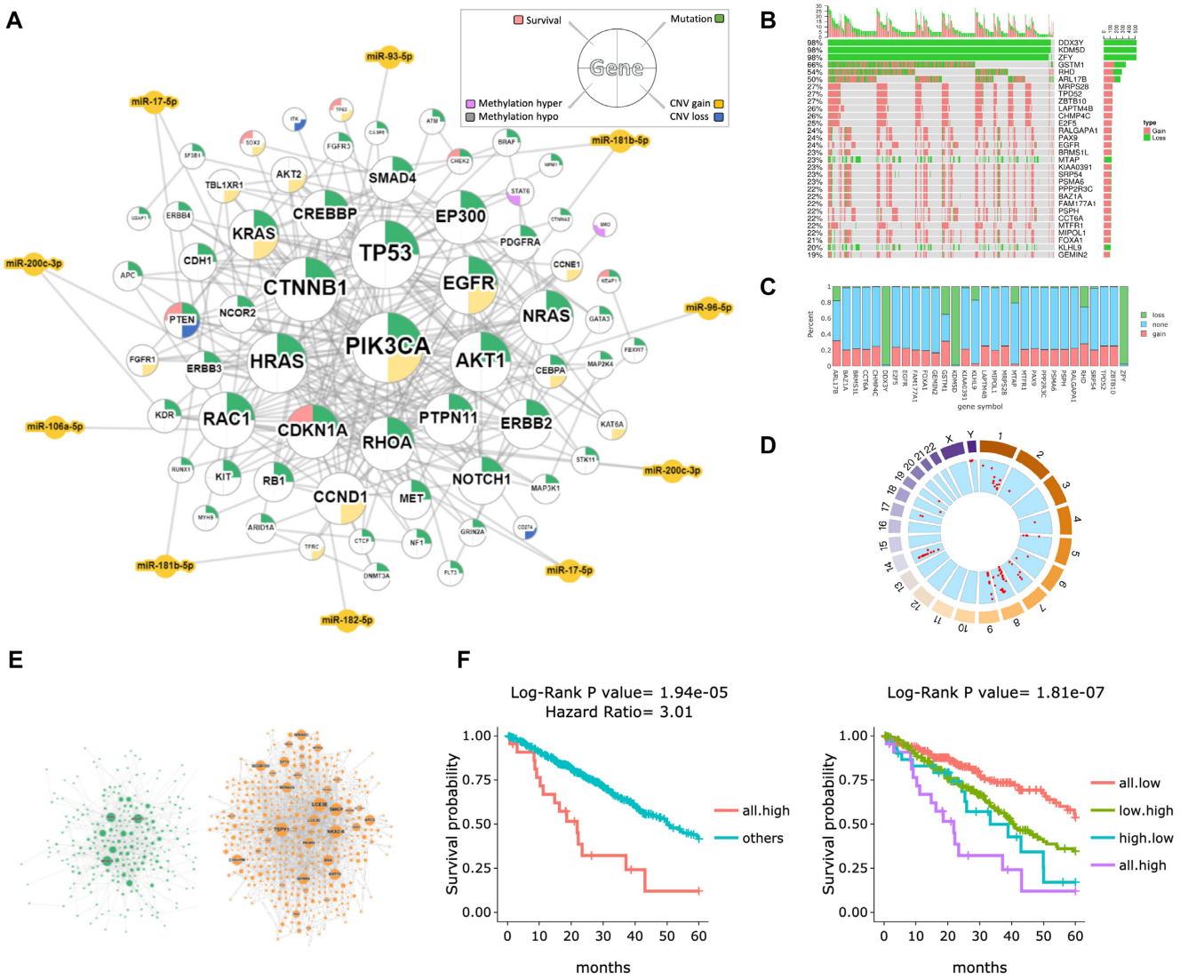

Figure 1. Omics features provided in the 'Cancer' section. (**A**) A network in 'Summary' panel illustrates mutation, CNV, methylation and miRNA drivers which are presented by different color grids in the node. The interactions between nodes are protein-protein interactions (PPI) in STRING database and synergistic effect gene pairs. (**B**) A heatmap in the 'CNV' panel is the display of the top 30 CNV drivers. (**C**) A percentage barchart in the 'CNV' panel represents the top 30 CNV drivers' sample proportions. (**D**) A circle graph in the 'CNV' panel marks the drivers' loci on each chromosome using a red dot based on the result of locus enrichment analysis. (**E**) A network illustrates the synergistic effect gene pairs defined in the 'Survival' panel with two directions (HR > 1 and HR < 1). The orange nodes indicate the synergistic effect genes in HR >1; the green nodes represent HR<1 genes. The larger nodes in the networks represent more synergistic effects defined. (**F**) Kaplan–Meier plots of each synergistic survival event are shown by two approaches: all high versus others (left) and four groups based on the expression (right).

analysis of 'By expression' allows researchers to investigate the co-expression events that affect patient survival by entering more than one target and defining a subgroup of specific patients. If more than one gene is selected for survival analysis by expression, we provide three stratification methods (all high versus others, high versus low, num. of high), as shown in Supplementary Figure S4B. In addition, we also provide four categories of survival time in this function (Supplementary Figure S4C), including overall survival (OS), progression-free interval (PFI), disease-free interval (DFI), and disease-specific survival (DSS), previously defined in the Liu *et al.* study (3). Alternatively, we offer a 'By mutation' function, which helps users investigate in a similar sense. After submitting the final request, the user receives a notification email with a Result ID that allows them to explore the results of 'Customized-analysis' in the 'Result and Download' when the calculation is finished (Supplementary Figure S5).

DISCUSSION

Although cancer hallmarks with different molecular features have been discovered at multi-omics levels, including genome, epigenome, and transcriptome, integrative omics research's importance on the basis of system biology is still the crucial issue at present. Vasaikar *et al.* presented Linke-

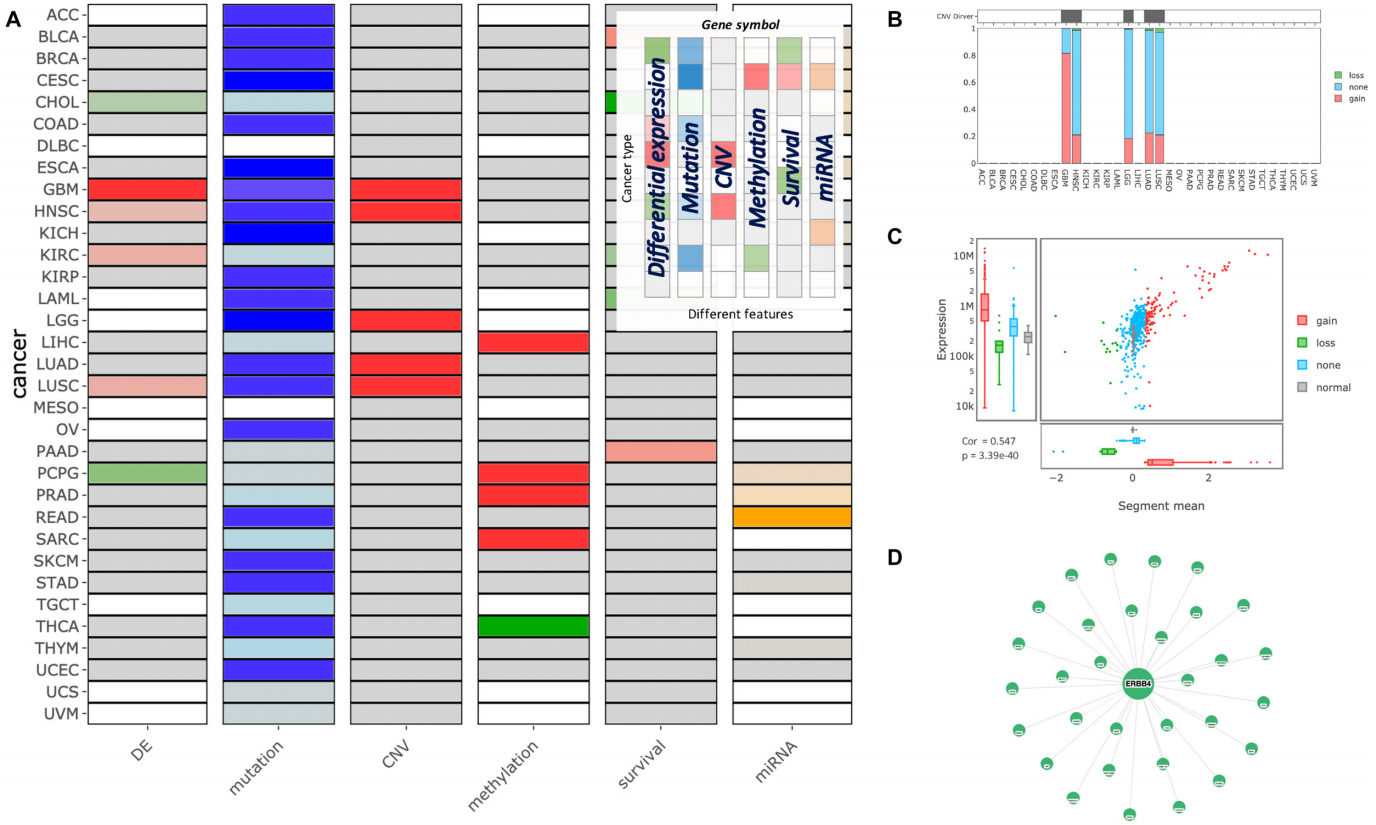

Figure 2. Omics features provided in the 'Gene' section. (**A**) A summary graph represents multi-omics features in the various cancer types for a single target gene in the 'Summary' panel. (**B**) The graph illustrates the bioinformatic algorithm combined analytical results. The upper panel indicates the driver defined by the number of tools, and the percentage barchart shows the sample proportion of each CNV types. (**C**) The details of a CNV driver are depicted, showing by boxplot and correlations between two omics data. The boxplot showing on the left side represents RNA expressions based on the CNV types, while the boxplot on the bottom shows segment mean values of each CNV types. (**D**) The network of synergistic effects for a single target gene is displayed which the width of the lines indicates the number of cancers. The two directions of the HR (>1 or <1) are illustrated in gray and pink, respectively.

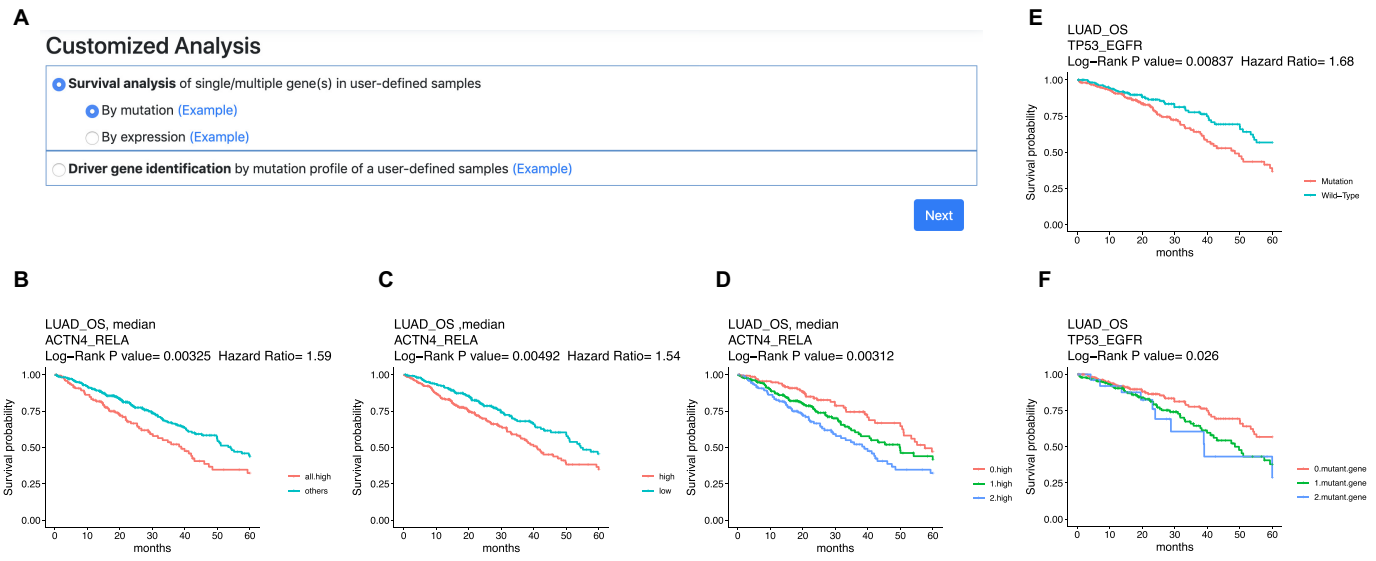

Figure 3. Novel functions in the 'Customized-analysis' section. (**A**) A illustration of 'Customized analysis' function is shown. In the 'By expression' function of 'Survival analysis,' three stratification methods, all high versus other (**B**), high versus low (**C**), and num. of high (**D**), used to explore co-occurring events. Two stratification methods for 'By mutation' function are mutation vs wild-type (**E**) and num. of mutant genes (**F**).

dOmics, which explores the associations and interpretations between different omics data for single gene-based studies (7). TCOA (8) and cbioportal (11) offer the ability to query omics data and establish the relations among them, while MEXPRESS (9) provides advanced visualization to combine omics data. In this updated version of DriverDB, we integrated multi-omics information and used published bioinformatic algorithms/tools to address the cancer driver events at distinct molecular levels (Figures 1 and 2). To interpret cancer omics' sophisticated information, we also designed 'Summary' panels in the 'Cancer' (Figure 1A) and 'Gene' (Figure 2A) sections to visualize the features at multi-omics levels concisely. To the best of our knowledge, no other database provides similar characteristics to define cancer driver genes and visualize cancer omics data concisely.

Various survival relevant databases indicate the importance of using survival analysis in cancer research. We have endeavored to enhance survival analysis' feasibility and scope of application. Firstly, not only is single gene survival provided in the 'Gene' section, but synergistic effects of two significant survival genes are illustrated in both the 'Cancer' and 'Gene' sections. Then, to cope with different clinical circumstances, we provide four clinical endpoints (OS, PFI, DFI and DSS) for a single gene survival analysis in the 'Gene' section.

Further, 'Survival Analysis' in the 'Customized-Analysis' has achieved a new step that allows researchers to address a variety of requirements. 'By expression' function offers the possibility to reveal other co-occurring events that are not included in the 'Gene' and 'Cancer' sections because of the strict criteria in 'Synergistic Survival Analysis.' Some research requires several targets to be simultaneously examined, and this function allows users to enter single or multiple genes. More importantly, a subgroup of patients, such as those with triple negative breast cancer, can be identified easily by dozens of clinical characteristics (Supplementary Figure S4A). In addition, we provide complex analytical reports in customized survival analysis to address multiple possibilities. The analytical report consists of four categories of clinical endpoints (OS, PFI, DFI, and DSS), three stratification methods (All high versus others, high vs. low, and num. of high) and two intervals (5 years and all), as shown in Supplementary Figure S4B. Supplementary Figure S4C displays four survival endpoints, which can be applied in different clinical trials depending upon the research purpose. Another remarkable contribution of this function is that it provides three different stratification methods, which have not been incorporated in other databases. All high vs. others are used most frequently in co-expression survival studies (Figure 3B) (15). High versus low, which stratifies patients by applying the median/mean of the z-score, the normalized value of variance, helps researchers evaluate the survival risk based on two equal groups (Figure 3C) (42). Num. of high usually is used to investigate the trends in gene modules' power (Figure 3D) (16). Stated simply, if researchers can discover the true targets, more gene combinations lead to a worse likelihood of survival. As shown in Figure 3B–D, not only the three stratification methods' outcomes show clearly that the co-occurrence of a high level of ACTN4 and RELA leads to a poor prognosis (18), but we have indicated that a significant survival difference is observed by the method of 'the number of high.' This demonstrates that different numbers of high level expressions of ACTN4 and RELA can contribute to different survival outcome, which data strengthen the verification of the study further, as well as provide different perspectives.

The importance of co-occurring mutation events in cancer biology also have been addressed in numerous studies (43–45). Those genomic alteration events have become increasingly important because many studies have demonstrated that co-mutations of crucial oncogenes, such as KRAS, affect the cancer microenvironment severely and influence therapeutic responses (45). To address this issue, the 'By mutation' function in 'Survival Analysis' in the 'Customized-analysis' allows co-mutation genes to be explored in a subset of patients according to specific clinical criteria (Figure 3A). We also provide two stratification methods (mutation vs. wild type, and num. of mutant genes) in the 'By mutation' function (Supplementary Figure S4B). Figure 3E indicates that there is a significant difference between the mutation and wild type groups, and Figure 3F shows clearly that the trend in the number of mutant genes results in a worse survival outcome. As shown in Figure 3E and F, we have confirmed co-mutation events between TP53 and EGFR in the early stage of lung cancer, reported in previous studies (45).

Recent advances in high-throughput technologies have brought a paradigm shift from single omics studies to large scale multi-omics research. DriverDBv3 seeks to improve the study of integrative cancer omics data by identifying driver genes' different molecular features, and providing a summary interpretation and informative visualization. We hope this updated version of DriverDB will make a contribution to integrative omics-based research.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to the National Center for High performance Computing for computer time and facilities and thank the TCGA research network for the availability of data.

FUNDING

Ministry of Science and Technology [MOST 106-2221-E-039-011-MY3; MOST 108-2314-B-039-060; MOST 108-2622-E-039-005-CC2]; China Medical University [CMU108-MF-93]. Funding for open access charge: China Medical University.
Conflict of interest statement. None declared.

REFERENCES

1. Bender,E. (2015) Big data in biomedicine. *Nature*, **527**, S1.
2. Mashl,R.J., Scott,A.D., Huang,K.L., Wyczalkowski,M.A., Yoon,C.J., Niu,B., DeNardo,E., Yellapantula,V.D., Handsaker,R.E., Chen,K. *et al.* (2017) GenomeVIP: a cloud platform for genomic variant discovery and interpretation. *Genome Res.*, **27**, 1450–1459.

3. Liu,J., Lichtenberg,T., Hoadley,K.A., Poisson,L.M., Lazar,A.J., Cherniack,A.D., Kovatich,A.J., Benz,C.C., Levine,D.A., Lee,A.V. *et al.* (2018) An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*, **173**, 400–416.
4. Saunders,G., Baudis,M., Becker,R., Beltran,S., Beroud,C., Birney,E., Brooksbank,C., Brunak,S., Van den Bulcke,M., Drysdale,R. *et al.* (2019) Leveraging European infrastructures to access 1 million human genomes by 2022. *Nat. Rev. Genet.*, **20**, 693–701.
5. Zukotynski,K.A. and Gerbaudo,V.H. (2017) Molecular imaging and precision medicine in lung cancer. *PET Clin.*, **12**, 53–62.
6. Robles,A.I. and Harris,C.C. (2017) Integration of multiple "OMIC" biomarkers: a precision medicine strategy for lung cancer. *Lung Cancer*, **107**, 50–58.
7. Vasaikar,S.V., Straub,P., Wang,J. and Zhang,B. (2018) LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.*, **46**, D956–D963.
8. Sun,Q., Li,M. and Wang,X. (2018) The Cancer Omics Atlas: an integrative resource for cancer omics annotations. *BMC Med Genomics*, **11**, 63.
9. Koch,A., Jeschke,J., Van Criekinge,W., van Engeland,M. and De Meyer,T. (2019) MEXPRESS update 2019. *Nucleic Acids Res.*, **47**, W561–W565.
10. 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.
11. 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**, l1.
12. Goswami,C.P. and Nakshatri,H. (2014) PROGgeneV2: enhancements on the existing database. *BMC Cancer*, **14**, 970.
13. Mizuno,H., Kitada,K., Nakai,K. and Sarai,A. (2009) PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med. Genomics*, **2**, 18.
14. Aguirre-Gamboa,R., Gomez-Rueda,H., Martinez-Ledesma,E., Martinez-Torteya,A., Chacolla-Huaringa,R., Rodriguez-Barrientos,A., Tamez-Pena,J.G. and Trevino,V. (2013) SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. *PLoS One*, **8**, e74250.
15. Reedijk,M., Odorcic,S., Chang,L., Zhang,H., Miller,N., McCready,D.R., Lockwood,G. and Egan,S.E. (2005) High-level coexpression of JAG1 and NOTCH1 is observed in human breast cancer and is associated with poor overall survival. *Cancer Res.*, **65**, 8530–8537.
16. Wiseman,S.M., Makretsov,N., Nielsen,T.O., Gilks,B., Yorida,E., Cheang,M., Turbin,D., Gelmon,K. and Huntsman,D.G. (2005) Coexpression of the type 1 growth factor receptor family members HER-1, HER-2, and HER-3 has a synergistic negative prognostic effect on breast carcinoma survival. *Cancer*, **103**, 1770–1777.
17. Aytes,A., Mitrofanova,A., Lefebvre,C., Alvarez,M.J., Castillo-Martin,M., Zheng,T., Eastham,J.A., Gopalan,A., Pienta,K.J., Shen,M.M. *et al.* (2014) Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell*, **25**, 638–651.
18. Lomert,E., Turoverova,L., Kriger,D., Aksenov,N.D., Nikotina,A.D., Petukhov,A., Mittenberg,A.G., Panyushev,N.V., Khotin,M., Volkov,K. *et al.* (2018) Co-expression of RelA/p65 and ACTN4 induces apoptosis in non-small lung carcinoma cells. *Cell Cycle*, **17**, 616–626.
19. Yang,X.H., Liu,L., Hu,Y.J., Zhang,P. and Hu,Q.G. (2019) Co-expression of XIAP and CIAP1 play synergistic effect on patient's prognosis in head and neck cancer. *Pathol. Oncol. Res.*, **25**, 1111–1116.
20. Amelio,I., Tsvetkov,P.O., Knight,R.A., Lisitsa,A., Melino,G. and Antonov,A.V. (2016) SynTarget: an online tool to test the synergetic effect of genes on survival outcome in cancer. *Cell Death Differ.*, **23**, 912.
21. Sondka,Z., Bamford,S., Cole,C.G., Ward,S.A., Dunham,I. and Forbes,S.A. (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer*, **18**, 696–705.
22. Repana,D., Nulsen,J., Dressler,L., Bortolomeazzi,M., Venkata,S.K., Tourna,A., Yakovleva,A., Palmieri,T. and Ciccarelli,F.D. (2019) The Network of Cancer Genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biol.*, **20**, 1.
23. Cheng,W.C., Chung,I.F., Chen,C.Y., Sun,H.J., Fen,J.J., Tang,W.C., Chang,T.Y., Wong,T.T. and Wang,H.W. (2014) DriverDB: an exome sequencing database for cancer driver gene identification. *Nucleic Acids Res.*, **42**, D1048–D1054.
24. Chung,I.F., Chen,C.Y., Su,S.C., Li,C.Y., Wu,K.J., Wang,H.W. and Cheng,W.C. (2016) DriverDBv2: a database for human cancer driver gene research. *Nucleic Acids Res.*, **44**, D975–D979.
25. Cumbo,F., Fiscon,G., Ceri,S., Masseroli,M. and Weitschek,E. (2017) TCGA2BED: extracting, extending, integrating, and querying The Cancer Genome Atlas. *BMC Bioinformatics*, **18**, 6.
26. Colaprico,A., Silva,T.C., Olsen,C., Garofano,L., Cava,C., Garolini,D., Sabedot,T.S., Malta,T.M., Pagnotta,S.M., Castiglioni,I. *et al.* (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.*, **44**, e71.
27. Shukla,S., Evans,J.R., Malik,R., Feng,F.Y., Dhanasekaran,S.M., Cao,X., Chen,G., Beer,D.G., Jiang,H. and Chinnaiyan,A.M. (2017) Development of a RNA-Seq based prognostic signature in lung adenocarcinoma. *J. Natl. Cancer Inst.*, **109**, djw200.
28. Sashegyi,A. and Ferry,D. (2017) On the interpretation of the hazard ratio and communication of survival benefit. *Oncologist*, **22**, 484–486.
29. Gevaert,O., Villalobos,V., Sikic,B.I. and Plevritis,S.K. (2013) Identification of ovarian cancer driver genes by using module network integration of multi-omics data. *Interface Focus*, **3**, 20130013.
30. Gevaert,O., Tibshirani,R. and Plevritis,S.K. (2015) Pancancer analysis of DNA methylation-driven genes using MethylMix. *Genome Biol.*, **16**, 17.
31. Lai,Y.P., Wang,L.B., Wang,W.A., Lai,L.C., Tsai,M.H., Lu,T.P. and Chuang,E.Y. (2017) iGC-an integrated analysis package of gene expression and copy number alteration. *BMC Bioinformatics*, **18**, 35.
32. Alvarez,M.J., Chen,J.C. and Califano,A. (2015) DIGGIT: a Bioconductor package to infer genetic variants driving cellular phenotypes. *Bioinformatics*, **31**, 4032–4034.
33. Gevaert,O. (2015) MethylMix: an R package for identifying DNA methylation-driven genes. *Bioinformatics*, **31**, 1839–1841.
34. Cedoz,P.L., Prunello,M., Brennan,K. and Gevaert,O. (2018) MethylMix 2.0: an R package for identifying DNA methylation genes. *Bioinformatics*, **34**, 3044–3046.
35. Silva,T.C., Coetzee,S.G., Gull,N., Yao,L., Hazelett,D.J., Noushmehr,H., Lin,D.C. and Berman,B.P. (2019) ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics*, **35**, 1974–1977.
36. Chung,I.F., Chang,S.J., Chen,C.Y., Liu,S.H., Li,C.Y., Chan,C.H., Shih,C.C. and Cheng,W.C. (2017) YM500v3: a database for small RNA sequencing in human cancer research. *Nucleic Acids Res.*, **45**, D925–D931.
37. Cheng,W.C., Chung,I.F., Huang,T.S., Chang,S.T., Sun,H.J., Tsai,C.F., Liang,M.L., Wong,T.T. and Wang,H.W. (2013) YM500: a small RNA sequencing (smRNA-seq) database for microRNA research. *Nucleic Acids Res.*, **41**, D285–D294.
38. Cheng,W.C., Chung,I.F., Tsai,C.F., Huang,T.S., Chen,C.Y., Wang,S.C., Chang,T.Y., Sun,H.J., Chao,J.Y., Cheng,C.C. *et al.* (2015) YM500v2: a small RNA sequencing (smRNA-seq) database for human cancer miRNome research. *Nucleic Acids Res.*, **43**, D862–D867.
39. Leiserson,M.D., Wu,H.T., Vandin,F. and Raphael,B.J. (2015) CoMEt: a statistical approach to identify combinations of mutually exclusive alterations in cancer. *Genome Biol.*, **16**, 160.
40. Babur,O., Gonen,M., Aksoy,B.A., Schultz,N., Ciriello,G., Sander,C. and Demir,E. (2015) Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. *Genome Biol.*, **16**, 45.
41. Han,Y., Yang,J., Qian,X., Cheng,W.C., Liu,S.H., Hua,X., Zhou,L., Yang,Y., Wu,Q., Liu,P. *et al.* (2019) DriverML: a machine learning algorithm for identifying driver genes in cancer sequencing studies. *Nucleic Acids Res.*, **47**, e45.

42. Kao,S.H., Cheng,W.C., Wang,Y.T., Wu,H.T., Yeh,H.Y., Chen,Y.J., Tsai,M.H. and Wu,K.J. (2019) Regulation of miRNA biogenesis and histone modification by K63-polyubiquitinated DDX17 controls cancer stem-like features. *Cancer Res.*, **79**, 2549–2563.
43. Mina,M., Raynaud,F., Tavernari,D., Battistello,E., Sungalee,S., Saghafinia,S., Laessle,T., Sanchez-Vega,F., Schultz,N., Oricchio,E. *et al.* (2017) Conditional selection of genomic alterations dictates cancer evolution and oncogenic dependencies. *Cancer Cell*, **32**, 155–168.
44. Sanchez-Vega,F., Mina,M., Armenia,J., Chatila,W.K., Luna,A., La,K.C., Dimitriadoy,S., Liu,D.L., Kantheti,H.S., Saghafinia,S. *et al.* (2018) Oncogenic signaling pathways in the cancer genome atlas. *Cell*, **173**, 321–337.
45. Skoulidis,F. and Heymach,J.V. (2019) Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat. Rev. Cancer*, **19**, 495–509.