

RESEARCH

Open Access

# CoMutPlotter: a web tool for visual summary of mutations in cancer cohorts



Po-Jung Huang<sup>1,2,3</sup>, Hou-Hsien Lin<sup>4</sup>, Chi-Ching Lee<sup>5</sup>, Ling-Ya Chiu<sup>4</sup>, Shao-Min Wu<sup>2</sup>, Yuan-Ming Yeh<sup>3</sup>, Petrus Tang<sup>2</sup>, Cheng-Hsun Chiu<sup>3</sup>, Ping-Chiang Lyu<sup>4\*</sup> and Pei-Chien Tsai<sup>1,2,3\*</sup>

From The 8th Annual Translational Bioinformatics Conference  
Seoul, South Korea. 31 October - 2 November 2018

## Abstract

**Background:** CoMut plot is widely used in cancer research publications as a visual summary of mutational landscapes in cancer cohorts. This summary plot can inspect gene mutation rate and sample mutation burden with their relevant clinical details, which is a common first step for analyzing the recurrence and co-occurrence of gene mutations across samples. The cBioPortal and iCoMut are two web-based tools that allow users to create intricate visualizations from pre-loaded TCGA and ICGC data. For custom data analysis, only limited command-line packages are available now, making the production of CoMut plots difficult to achieve, especially for researchers without advanced bioinformatics skills. To address the needs for custom data and TCGA/ICGC data comparison, we have created CoMutPlotter, a web-based tool for the production of publication-quality graphs in an easy-of-use and automatic manner.

**Results:** We introduce a web-based tool named CoMutPlotter to lower the barriers between complex cancer genomic data and researchers, providing intuitive access to mutational profiles from TCGA/ICGC projects as well as custom cohort studies. A wide variety of file formats are supported by CoMutPlotter to translate cancer mutation profiles into biological insights and clinical applications, which include Mutation Annotation Format (MAF), Tab-separated values (TSV) and Variant Call Format (VCF) files.

**Conclusions:** In summary, CoMutPlotter is the first tool of its kind that supports VCF file, the most widely used file format, as its input material. CoMutPlotter also provides the most-wanted function for comparing mutation patterns between custom cohort and TCGA/ICGC project. Contributions of COSMIC mutational signatures in individual samples are also included in the summary plot, which is a unique feature of our tool.

CoMutPlotter is freely available at <http://tardis.cgu.edu.tw/comutplotter>.

**Keywords:** Cancer mutational profile mutational signature TCGA

## Background

With the rapid evolution of next-generation technologies (NGS) combined with dropping costs, whole-exome sequencing (WES) has become a widely-accepted application for clinical research and diagnostic purposes. In the past few years, over 10,000 exomes across 40 distinct types of human cancer were generated by The Cancer Genome

Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). The Broad institute has released the GATK Best Practice workflow tailored to somatic variant discovery. Researchers can follow this standardize analysis protocol, making their results comparable to TCGA/ICGC projects. Variant annotation is a relatively mature and feasible work because of the state-of-the-art packages like ANNOVAR [1], VEP [2], SnpEff [3], and Oncotator [4]. However, an intuitive and convenient way for visualizing and interpreting genomic data from high-throughput technologies continues to be challenging. Inconsistent file formats used in handling mutation profiles may introduce

\* Correspondence: [pclyu@mx.nthu.edu.tw](mailto:pclyu@mx.nthu.edu.tw); [pct sai@gap.cgu.edu.tw](mailto:pct sai@gap.cgu.edu.tw)

<sup>4</sup>Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan

<sup>1</sup>Department of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan

Full list of author information is available at the end of the article



additional problems in subsequent data integration, visualization and comparison.

CoMut plot [5–7] is widely used in cancer research publications as a visual summary of mutational landscapes in cancer cohorts. This summary plot can inspect gene mutation rate and sample mutation burden with their relevant clinical details, which is a common first step for analyzing the recurrence and co-occurrence of gene mutations across samples. There are two web-based applications, the cBioPortal [8] and iCoMut (<http://firebrowse.org/iCoMut/>), which allow users to create intricate visualizations from pre-loaded TCGA data. For custom data analysis, only certain file formats such as MAF and TSV format are supported at this stage, which are based on command-line packages [6, 7], making the production of customizable plots difficult to achieve, especially for non-bioinformatics researchers.

To address the needs for custom data and TCGA/ICGC data comparison, we have created CoMutPlotter, a web-based tool, for the production of publication quality graphs and to translate cancer mutation profiles into biological insights and clinical applications. A wide variety of file formats are supported by CoMutPlotter, which include Mutation Annotation Format (MAF), Tab-separated values (TSV) and Variant Call Format (VCF) files. It is worth noting that CoMutPlotter is the first tool of its kind that directly supports VCFs, a dominant output format of all variant discovery pipelines like the GATK Toolkit [9], VarScan [10], and SAMtools [11]. Deciphering signatures of the mutational processes in human cancer is a new trend in cancer research community [12–14] because these signatures are footprints of molecular aberrations occurring in tumors. Alexandrov et al. identified a list of 30 reference signatures and about half of these signatures can be attributed to endogenous processes such as enzymatic activity of DNA cytidine deaminases (AID/APOBEC), the deficiency of DNA mismatch repair, or mutations in *POLE* and to exogenous mutagens like tobacco, ultraviolet light and toxic chemicals [15].

Our specific aim to construct CoMutPlotter is to lower the barriers between complex cancer genomic data and researchers. In addition to specifying the mutation burden and types of individual samples, we also allow the user to plot clinical features with their respective samples, providing intuitive access to mutational profiles from TCGA/ICGC as well as custom cohort studies alongside their clinical attributes. CoMutPlotter also provides the most-wanted function for comparing mutational landscapes between custom cohort and TCGA/ICGC project. To gain insight into the mutational processes that have altered the cancer genome, contributions of COSMIC signatures are quantified at sample resolution and integrated in the summary plot as dot

matrix, which is a unique feature of CoMutPlotter. CoMutPlotter is freely available at <http://tardis.cgu.edu.tw/comutplotter>.

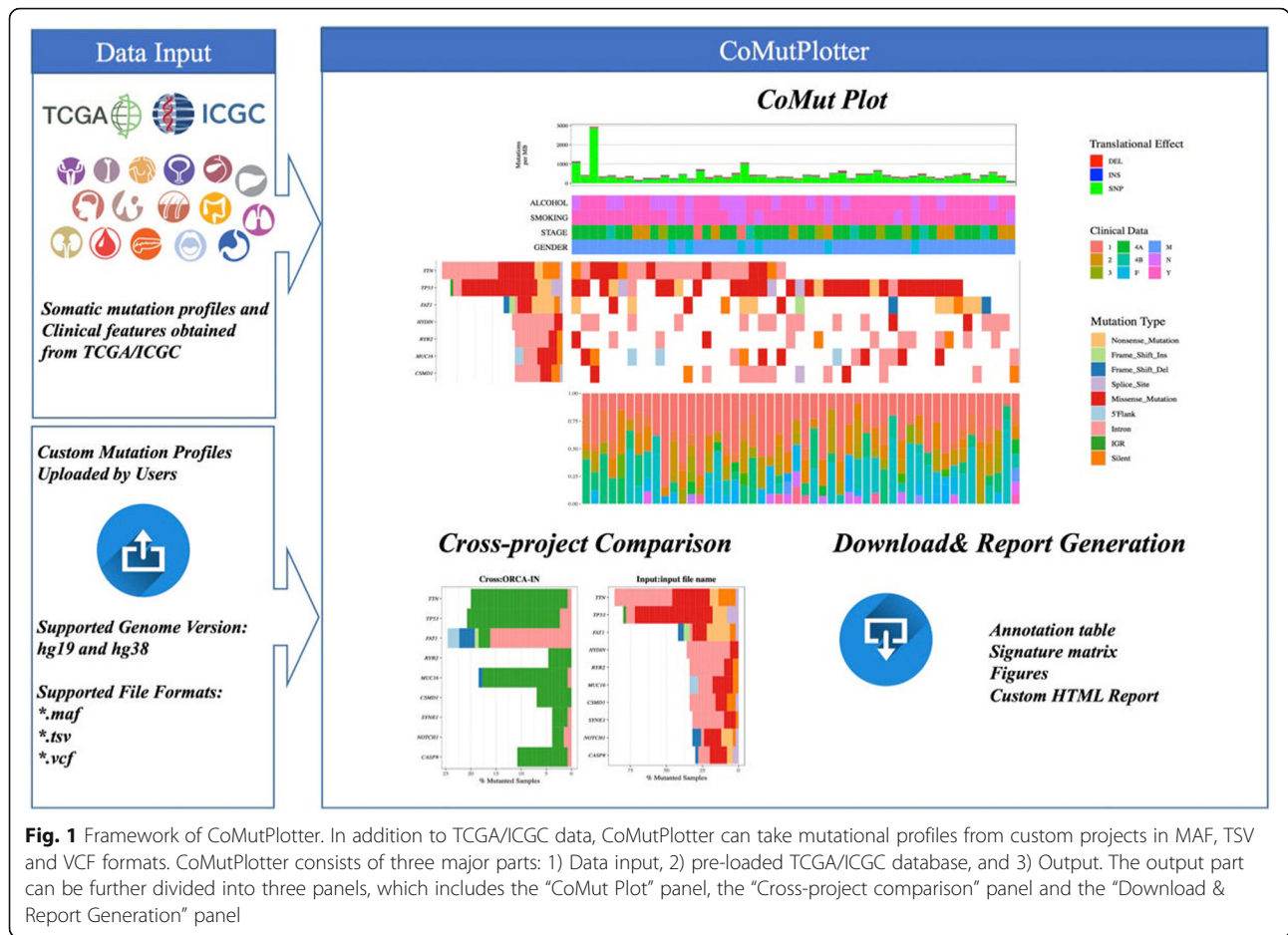
## Implementation

### CoMutPlotter framework

CoMutPlotter provides an intuitive web interface to receive mutation profiles obtained from cancer sequencing projects. Mutation Annotation Format (MAF) is widely used in TCGA cancer studies for storing mutation profiles, which is also the basis for many downstream analyses such as variant annotation, driver gene detection, mutual exclusivity analysis, and mutational signature identification. In addition to MAF file, CoMutPlotter also includes function to convert ICGC tab-separated values (TSV) file and standard Variant Call Format (VCF) file to MAF file, making this tool more accessible to wider researchers. CoMutPlotter not only provides complete functions for performing analyses mentioned above but also creates an interactively framework to present and summarize the important characteristics of the multidimensional analysis results from a custom cancer cohort. For the convenience of comparative analysis between custom data and TCGA/ICGC data, 73 mutation profiles were downloaded from TCGA and ICGC Data Portal and compiled as pre-loaded database. The PHP and R script are used to summarize all the generated results into an integrative plot to grasp the global characteristics of a mutation profile and to reveal the co-occurrence of mutations and samples. Download links are also provided to download publication-quality figures, significantly mutated gene list and detailed annotation table (Fig. 1).

### Data input

CoMutPlotter accepts three dominant formats of mutation profiles, including MAF, TSV, and VCF formats. To make data management and analysis more efficient, mutation profiles in diverse formats are converted to MAF format before entering subsequent analyses. A custom script for file format conversion is available for download ([http://tardis.cgu.edu.tw/comutplotter/comutplotter\\_tutorial/implementation.html#for-custom-study-with-large-number-of-vcf-files](http://tardis.cgu.edu.tw/comutplotter/comutplotter_tutorial/implementation.html#for-custom-study-with-large-number-of-vcf-files)) when users try to deal with a study cohort with large number of VCF files. To perform in-depth comparisons between clinical features or study designs within a cancer cohort, the demographic profile can also be uploaded along with the mutation profiles. Detailed instructions on the usage of the custom script and the acceptable format of the demographic file can be found on the tutorial page ([http://tardis.cgu.edu.tw/comutplotter/Tutorial/comutTutorial.html#2\\_data\\_input](http://tardis.cgu.edu.tw/comutplotter/Tutorial/comutTutorial.html#2_data_input)).



**Functional consequence annotation**

Functional annotation of variants is a key step [16] in the analysis of cancer sequencing data, and annotation results may have a substantial influence on the final conclusions of cohort studies. Despite using the same transcript sets (e.g., REFSEQ or ENSEMBL) as the basis for annotation, there still have about 20% disagreement between annotation results generated from well-recognized methodologies such as ANNOVAR, SnpEff, and Variant Effect Predictor. To capture the expected variant annotations in concordance with TCGA published cancer studies, GENCODE release 19 was used to construct cancer-relevant transcripts as instructed by previous study ([https://www.broadinstitute.org/~lichtens/onco-beta/tx\\_exact\\_uniprot\\_matches.AKT1\\_CRLF2\\_FGFR1.txt](https://www.broadinstitute.org/~lichtens/onco-beta/tx_exact_uniprot_matches.AKT1_CRLF2_FGFR1.txt)) [4]. The local installed version of Broad’s Onco-tator [4] was used to perform the annotation tasks, making functional annotation of variants become a reproducible step and ensuring the annotation results are comparable between custom cohort and TCGA/ICGC studies. The mutation rates of synonymous and non-synonymous variants can be

calculated in individual samples, which are subsequently rendered into a stacked bar chart for monitoring selective pressure acting on protein-coding genes. Gene mutations can be further classified into missense, nonsense, stop-gain, insertion and deletion, frameshift and splice site mutations, depending on where they occur and whether they alter the composition of proteins.

**Cancer driver gene identification**

International cancer projects are underway through The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) aim to establish a comprehensive catalogue of cancer associated genes across all cancer types. However, most of the existing analytical methods fail to account for mutational heterogeneity that affects the background mutation rate and may led to the identification of many specious genes. Lawrence et al. has developed a new method, named MutSigCV [17], to address the issue of mutational heterogeneity, which is correlated with transcriptional activity, DNA replication timing, and mutation frequency variability across patients. To facilitate the

identification of genes truly associated with cancer and to make driver gene detection more accessible to users, CoMutPlotter has incorporated MutSigCV as a critical analysis module. The mutation profiles uploaded by users are converted to MAF format as mentioned above and then subjected to MutSigCV to determine significantly mutated genes with false discovery rates (q-value) less than or equal to 0.1. Since the mutation profiles of 73 cancer projects have been downloaded from TCGA/ICGC Data Portal, we also applied the MutSigCV method to identify driver genes in individual cancer projects. Based on the pre-calculated results, users can easily compare the resulting gene lists between custom study cohort and published cancer projects.

### Mutational signature recognition

Mutational signatures are patterns of somatic mutations hidden in cancer genomes, which can be represented as different combinations of 96 available trinucleotide mutation contexts. Each mutational signature may be associated with specific kinds of mutational processes resulting from exogenous and endogenous mutagens such as ultraviolet radiations, tobacco-related exposures and abnormal activity of enzymes. Up to date, 30 distinct mutational signatures have been identified and categorized in COSMIC database using the WTSI Mutational Signature Analysis Framework [12]. However, large cohorts and sufficient computing resources are required by existing analysis framework of WTSI. Moreover, quantifying known signatures in individual samples is not possible under the current WTSI framework when sample sizes are small. For known signatures identification and quantification, the R *deconstructSigs* package [18] was used to determine the composition of mutational signatures in individual tumor samples. A dot matrix plot is used to show the percentage contribution of the identified signatures in each sample. The proposed etiology of each signature can be downloaded as a summary table, which may be beneficial to explore different combinations of mutational signatures that are representative in distinct groups of patients, to depict potential therapeutic targets and to reveal new connections between mutational processes and clinical features.

### Report generation

With the improved completeness of software packages over the past few years, data analysis in cancer research has gradually become a feasible task. Many state-of-the-art analysis packages such as GATK [9], Oncotator [4] and MutSigCV [17] have been released by the Broad Institute and users can apply these packages to analyze their own data when computing power is not a concern. However, most of the existing packages lack a mechanism to create a visual summary for effectively communicating personal findings

to research community, which can be the most important and challenging step of scientific research. As shown in Fig. 2, CoMutPlotter has summarized all the analysis results mentioned in above sections into a single integrative plot. The mutation profile of custom cohort is displayed as a heatmap in the main body of the plot, using different glyphs and colors to reveal diverse type of gene mutations in different patients. The significantly mutated genes identified by MutSigCV are displayed as bar chart along the right axis of the plot, ordered according to negative q-values in log transformations. Along the top axis is the density of mutations for each patient while the mutation frequency for each gene is rendered along the left axis. The clinical features can be retrieved from the uploaded demographic profile aligned according to the respective patients and rendered as a second heatmap on the top panel of the plot. Contributions of COSMIC mutational signatures in each patient are shown as dot matrix, rendered at the bottom panel of the plot. The dynamic framework of CoMutPlotter provides both sorting and filtering functions on the left panel. Users can sort the list of genes according to mutation frequencies or the FDR q-values. Filters are provided based on items such as custom gene list and mutation types, facilitating users to focus on their target of interest. A “report generation” button is provided to create a publication-quality figure, often seen in cancer research publications as a visual summary of genetic aberrations in cancer cohorts along with table with detailed annotation information.

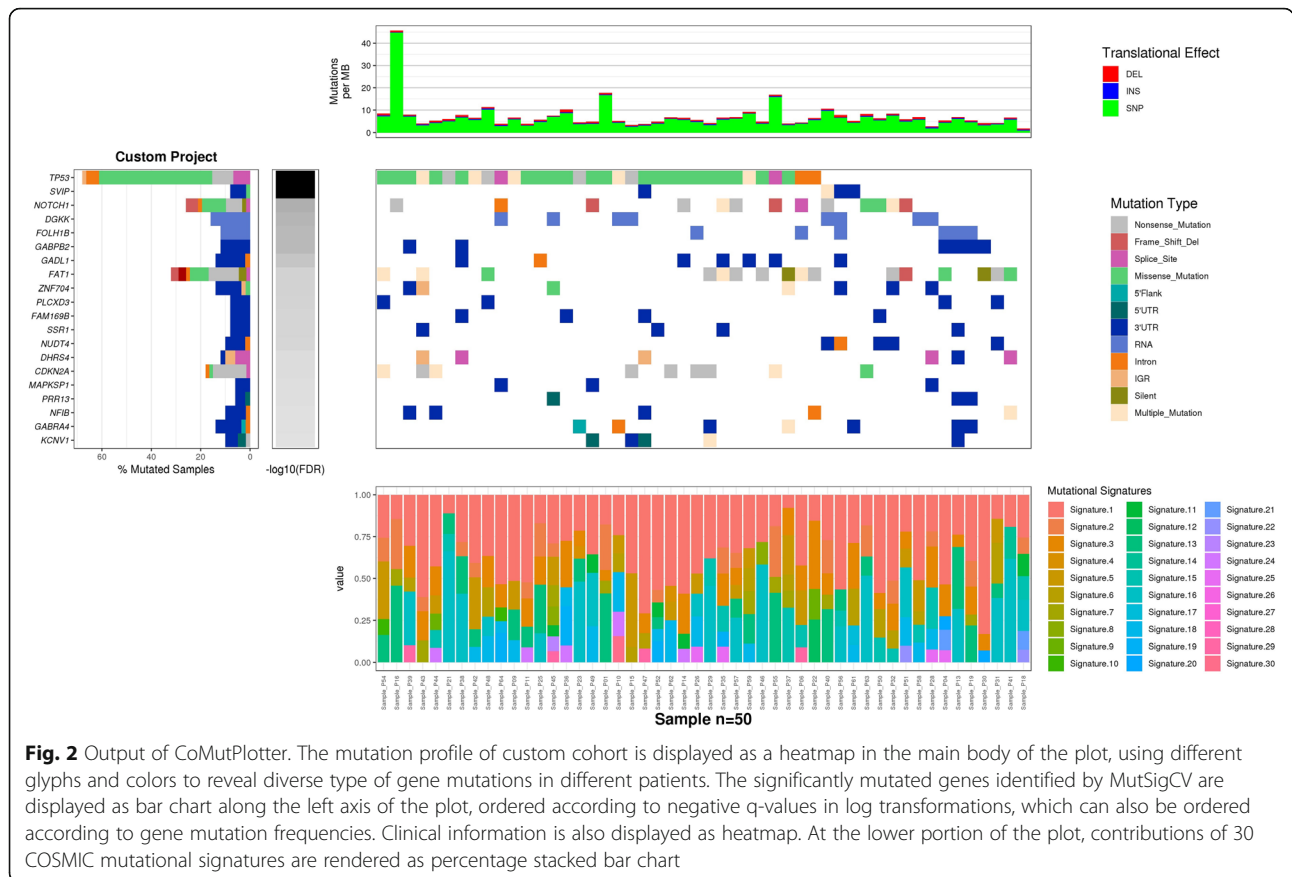
## Results and discussion

### Example of use

As a proof-of-concept experiment, we applied CoMutPlotter to analyze our published datasets [17, 18], which contain 50 sets of whole-exome sequencing data from oral cancer patients in Taiwan. In this study cohort, 24,051 mutation events that correspond to 23,495 unique somatic mutations were identified and recorded in the MAF file, which can be downloaded through the following link: ([http://tardis.cgu.edu.tw/comutplotter/oscc\\_50.maf.zip](http://tardis.cgu.edu.tw/comutplotter/oscc_50.maf.zip)). For cross-project comparison, the same analysis workflow was also applied to analyze 178 sets of whole-exome sequencing data from oral cancer patients in India [19], which can be downloaded from the ICGC Data Portal through the following link: ([https://dcc.icgc.org/api/v1/download?fn=/release\\_27/Projects/ORCA-IN/simple\\_somatic\\_mutation.open.ORCA-IN.tsv.gz](https://dcc.icgc.org/api/v1/download?fn=/release_27/Projects/ORCA-IN/simple_somatic_mutation.open.ORCA-IN.tsv.gz)).

Detailed exemplary outputs for 50 oral tumors can be found on the CoMutPlotter demonstration page at (<http://tardis.cgu.edu.tw/comutplotter/Demo/>). Detailed instructions can refer to the following link [http://tardis.cgu.edu.tw/comutplotter/comutplotter\\_tutorial/example-of-use.html](http://tardis.cgu.edu.tw/comutplotter/comutplotter_tutorial/example-of-use.html).





**Output summary**

After the successful submission of a job, a dynamic progress bar will be displayed, indicating processing statuses such as job queuing, format conversion, variant annotation, significantly mutated gene identification, mutational signature decomposition, and CoMut plot generation. The standard output can be separated into three webpage panels, including CoMut plot, Cross-project comparison, and Download & Report generation.

In the “CoMut plot” panel, stacked bar graphs are used to represent the mutation burden of individual samples, the compositions of translational effects and the most frequently affected genes in a study cohort, rendered at the top and left-hand side of the resulting CoMut plot. In the main body of the plot, heat map is used to visualize multiple genomic alteration events in individual samples and to render diverse mutation types by different color schemes. Percentage stacked bar is used to represent the identified COSMIC signatures in each sample, which can be switched to dot matrix to better convey the contributions of respective mutational processes. The resulting plot can be ordered not only by gene mutation frequency but also significant FDR values calculated from MutSigCV algorithm. Moreover, users can create custom plot according to

the function for mutation types selection and custom gene list.

In the “Cross-project comparison” panel, users can easily compare their study cohort to pre-loaded cancer projects from TCGA/ICGC. Despite that users can retrieve or create CoMut plot for each TCGA/ICGC project using on-line resources or command-line tools, only CoMutPlotter provides the function to render the comparison result in the same plot and in the same gene order, making cross-project comparison become an easy task. As shown in Fig. 2, users can easily depict the convergent and divergent gene mutation frequencies between Taiwan and India populations of the same cancer type.

In the “Download & Report generation” panel, detailed information about the significantly mutated genes, contributions of mutational signatures in individual samples and the resulting CoMut plot can be downloaded from our server as separated tables, figures or integrated HTML file.

**Comparison of the features across similar tools**

Over the past few years, many packages have been developed to meet the needs for visual summary of mutations in cancer cohorts. These packages can be further

**Table 1** Comparison of the features of similar tools for CoMut-like plot generation

Software	CoMutPlotter	iCoMut <sup>a</sup>	cBioPortal <sup>b</sup>	GenVisR <sup>c</sup>	maftools <sup>d</sup>
Usage	Web	Web	Web	Command	Command
Supported file format	MAF TSV VCF	TCGA data	TCGA/ICGC data	MAF TSV	MAF
Significantly mutated gene identification	✓	✓	✗	✗	✗
Cross-project comparison	✓	✗	✗	✗	✗
Mutational signature analysis	✓	✗	✗	✗	✓

<sup>a</sup> <http://firebrowse.org/iCoMut/><sup>b</sup> <http://www.cbioportal.org><sup>c</sup> <https://bioconductor.org/packages/release/bioc/html/GenVisR.html><sup>d</sup> <https://bioconductor.org/packages/release/bioc/html/maftools.html>

classified into two groups. One group is web-based tools and the other group is command line tools. The cBioPortal and iCoMut are two representative packages of web-based tools and the benefit part is easy-to-use while the shortcoming is limited to the cancer projects from TCGA or ICGC. The command-line tools have their inherited problem, only support MAF format as their input format and likely limited to specific users with bioinformatics background. Furthermore, the issue of cross-project comparison has never been covered by existing packages as well as the functionality of mutational signature analysis. CoMutPlotter aims to provide the most comprehensive set of features to address all of these issues. A more detailed comparisons of similar existing software is summarized in Table 1.

### Future development

The planned future development of new features will be focused on incorporating the copy number variation and gene expression data into the resulting CoMut plot.

### Conclusions

CoMutPlotter is the first tool of its kind that supports VCF file, the most widely used file format, as its input material. CoMutPlotter provides the most complete solution starting from file format conversion all the way to variant annotation, driver gene identification, mutational signature recognition and CoMut plot generation. Moreover, CoMutPlotter also provides the most-wanted function for comparing mutation patterns between custom cohort and TCGA/ICGC project. Contributions of COSMIC mutational signatures in individual samples are also included in the summary plot, which is a unique feature of our tool.

CoMutPlotter is freely available at <http://tardis.cgu.edu.tw/comutplotter>.

### Availability and requirements

Project name: CoMutPlotter.

Project home page: <http://tardis.cgu.edu.tw/comutplotter>

Operating system(s): Platform independent.

Programming language(s): R, PHP, Shell Script and JavaScript.

Other requirements: Supported browsers Safari, Google Chrome, Firefox, Internet Explorer 11 and Microsoft Edge.

License: GNU GPL version 3.

Any restrictions to use by non-academics: none.

### Abbreviations

COSMIC: Catalogue of Somatic Mutations in Cancer; GATK: Genome Analysis Toolkit; ICGC: International Cancer Genome Consortium; MAF: Mutation Annotation Format; TCGA: The Cancer Genome Atlas; TSV: Tab-separated Values; VCF: Variant Call Format; WES: Whole-exome sequencing

### Funding

This work was supported by the Chang Gung Memorial Hospital, Linkou, Taiwan (CMRPD110131/2/3). Publication costs are funded by grants from the Ministry of Science and Technology, Taiwan (MOST 107-2314-B-182-062).

### Availability of data and materials

The demonstration datasets used in this study are available in the ICGC Data Portal ([https://dcc.icgc.org/api/v1/download?fn=/release\\_27/Projects/ORCA-IN/simple\\_somatic\\_mutation.open.ORCA-IN.tsv.gz](https://dcc.icgc.org/api/v1/download?fn=/release_27/Projects/ORCA-IN/simple_somatic_mutation.open.ORCA-IN.tsv.gz)) and Chang Gung University ([http://tardis.cgu.edu.tw/comutplotter/oscc\\_50.maf.zip](http://tardis.cgu.edu.tw/comutplotter/oscc_50.maf.zip)).

### About this supplement

This article has been published as part of *BMC Medical Genomics Volume 12 Supplement 5, 2019: Selected articles from the 8th Translational Bioinformatics Conference: Medical Genomics*. The full contents of the supplement are available online at <https://bmcmgenomics.biomedcentral.com/articles/supplements/volume-12-supplement-5>.

### Authors' contributions

PJH designed the overall framework of this application and prepared the manuscript. HHL established the system and prepared the demonstration data sets. CCL created the uploading and job-queuing system. LYC, WSM, YMY and KYH contributed to the visual interactive interface. HHL, CHC and PT tested the software and prepared the tutorial page. PCL and PCT were the main supervisors of the project. All authors of this paper have read and approved the final manuscript.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Department of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan. <sup>2</sup>Graduate Institute of Biomedical Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan. <sup>3</sup>Genomic Medicine Core Laboratory, Chang Gung Memorial Hospital, Linkou, Taiwan. <sup>4</sup>Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan. <sup>5</sup>Department and Graduate Institute of Computer Science and Information Engineering, Chang Gung University, Taoyuan, Taiwan.

Published: 11 July 2019

### References

1. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucl Acids Res.* 2010;38:e164.
2. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl variant effect predictor. *Genome Biol BioMed Central.* 2016;17:122.
3. De Baets G, Van Durme J, Reumers J, Maurer-Stroh S, Vanhee P, Dopazo J, et al. SnpEff 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucl. Acids Res.* 2012;40:D935–9.
4. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: Cancer variant annotation tool. *Hum Mutat.* 2015;36:E2423–9.
5. Bi WL, Greenwald NF, Ramkissoon SH, Abedalthagafi M, Coy SM, Ligon KL, et al. Clinical identification of oncogenic drivers and copy-number alterations in pituitary tumors. *Endocrinology.* 2017;158:2284–91.
6. Yu L, Kim HT, Kasar S, Benien P, Du W, Hoang K, et al. Survival of Del17p CLL depends on genomic complexity and somatic mutation. *Clin Cancer Res.* 2017;23:735–45.
7. Huang P-J, Lee C-C, Chiu L-Y, Huang K-Y, Yeh Y-M, Yang C-Y, et al. VARReporter: variant reporter for cancer research of massive parallel sequencing. *BMC Genomics.* 2018;19:86.
8. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6:p11.
9. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
10. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;22:568–76.
11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
12. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* 2013;3:246–59.
13. Huang P-J, Chiu L-Y, Lee C-C, Yeh Y-M, Huang K-Y, Chiu C-H, et al. mSignatureDB: a database for deciphering mutational signatures in human cancers. *Nucl Acids Res.* 2017;46:D964–70.
14. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature.* 2016;534:47–54.
15. Petljak M, Alexandrov LB. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis.* 2016;37:531–40.
16. McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier J-B, et al. Choice of transcripts and software has a large effect on variant annotation. *Genome Med.* 2014;6:26.
17. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499:214–8.
18. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* 2016;17:31.
19. India Project Team of the International Cancer Genome Consortium. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nat Commun.* 2013;4:2873.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

