SOFTWARE TOOL ARTICLE

MAFDash: An easy-to-use dashboard builder for mutation data [version 1; peer review: 2 not approved]

Ashish Jain1,2, Mayank Tandon1,2

1CCR Collaborative Bioinformatics Resource (CCBR), Center for Cancer Research, National Cancer Institute, Bethesda, MD, 20814, USA
2Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research, Frederick, MD, 21701, USA

Abstract
Characterizing the somatic mutation landscape of a cohort of patients has become a routine task in cancer research in recent years. Such studies are often highly interdisciplinary, requiring iterative analysis that must be evaluated at each step by many researchers. Therefore, there is a growing need for reporting tools that can easily generate interactive reports for sharing data and results with collaborators. Here we present an R package, MAFDash, that tries to simplify summarization and visualization of mutation data from Mutation Annotation Format (MAF) files. The output HTML dashboard is a self-contained report that can be used for downstream analysis and sharing results. MAFDash is freely available on Github (https://github.com/CCBR/MAFDash).

Keywords
MAF, Mutation, Single Nucleotide Variants, Visualization, Dashboard, WES, WGS

Open Peer Review

Approval Status

1
2

version 1
06 Jul 2022

1. Sigve Nakken1, Oslo University Hospital, Oslo, Norway
   University of Oslo, Oslo, Norway
   University of Oslo, Oslo, Norway
2. Heiko Brennenstuhl1, University Hospital Heidelberg, Heidelberg, Germany
   University Heidelberg, Heidelberg, Germany

Any reports and responses or comments on the article can be found at the end of the article.
Corresponding authors: Ashish Jain (jain.ashishjain1@gmail.com), Mayank Tandon (mtandon09@gmail.com)

Author roles: Jain A: Conceptualization, Data Curation, Methodology, Software, Visualization, Writing – Review & Editing; Tandon M: Conceptualization, Data Curation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN26120150003I. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Jain A and Tandon M. 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 work is properly cited.

How to cite this article: Jain A and Tandon M. MAFDash: An easy-to-use dashboard builder for mutation data [version 1; peer review: 2 not approved] F1000Research 2022, 11:748 https://doi.org/10.12688/f1000research.118761.1

First published: 06 Jul 2022, 11:748 https://doi.org/10.12688/f1000research.118761.1
Introduction
In the last decade, the cost of next-generation sequencing (NGS) has gone down exponentially as both throughput and novel methods continue to advance.1 For human clinical research, this has been reflected in an ever-growing number of datasets describing genomic variation among both normal and disease cohorts, including the 1000 Genomes Project Consortium,2 and the more recent gnomAD project,3 which still serve as important benchmarks of normal genomic variation in humans. Similar efforts for characterizing somatic mutations in cancer research have been completed for 33 tumor types by The Cancer Genome Atlas (TCGA) consortium,4 and over 1,700 cancer cell lines in the Cancer Cell Line Encyclopedia (CCLE) project from the Broad Institute.5 Although both TCGA and CCLE provide mult-omics data, single nucleotide polymorphisms (SNPs) and small insertion/deletions (Indels) from NGS data are often used as starting points for downstream analyses diving deep into the biological pathways and identifying drug target genes in these cancers.6

Both TCGA and CCLE provide somatic mutations freely as Mutation Annotation Format (MAF) files. This format is used to report high quality somatic variants for cohorts of cancer patients as it is more readable and portable than the traditional variant call format (VCF), and is therefore a common starting point for downstream analysis of somatic SNP/Indel data. R packages like ‘maftools’7 are frequently used by bioinformaticians to read, summarize, and perform statistical tests on data from MAF files, and they provide excellent functions for basic visualization, and flexible manipulation of the underlying data.

Since MAF files can contain a large number of annotations (e.g. the vcf2maf tool from MSKCC8 produces 136 columns of annotations with a default installation of Variant Effect Predictor9), selecting useful information and preparing it for discussion with researchers requires expertise. To simplify this task, we have developed MAFDash, an R package that helps to quickly create HTML dashboards for summarizing and visualizing data from MAF files. The resulting HTML file serves as a self-contained report that can be used to explore and share the results. MAFDash provides preset functions for extracting and organizing somatic variant data into interactive tables and figures. The goal of this package is to provide a simplified interface to filter and present data from MAF files suitable both for highly customized reports, as well as routine output from variant calling pipelines. The package also provides functions to generate individual plots as a ggplot210 or ComplexHeatmap11 object giving users more flexibility.

Methods
Implementation
MAFDash is a package intended for use with the R programming language.12 The report is generated with a parameterized R Markdown script to arrange all the information. If a MAF object is provided, an interactive table is generated to provide client-side, dynamic filtering of the variant data. In addition to the dashboard generation, it also consists of a variety of functions to generate high quality figures to visualize mutation data. We also provided detailed documentation and a test dataset to demonstrate usage of these functions. Static plots are generated using the R packages ‘maftools’,7 ComplexHeatmap,11 ‘circlize’,13 and ‘ggplot2’.10 Interactive visualizations are implemented using ‘canvasXpress’14 and ‘plotly’.15

Operation
MAFDash was developed and tested on 2019 Macbook Pros with 2.4GHz 8-core Intel Core i9 processors and 16 Gb of memory, running Mac OS X 10.15.7 (Catalina). The source code and documentation is hosted on Github (https://github.com/CCBR/MAFDash).

Functions for TCGA data
The function getMAFdataTCGA(…) retrieves TCGA mutation data in MAF format. This function takes the cancer code(s) as input and outputs the TCGA mutation data called from Mutect2,16 or other callers as available. This function internally uses the ‘TCGAbiolinks’ R package17 to download the data and then uses internal processing to output the mutation data in a clean format. For annotation information, the getTCGAClinicalAnnotations(…) function extracts and processes common clinical features provided with the TCGA data including pathological state, tissue site, age, gender, race, and vital status, and generates reasonable preset colors suitable for use with ‘ComplexHeatmap’. The processed mutation data along with the clinical annotations can be further analyzed by utilizing the various visualization functions in MAFDash.

Filtering of mutations
The filterMAF(…) function in MAFDash automatically detects the presence of relevant columns and re-casts them appropriately for numeric or text-based filtering. These include tumor read frequency and depth, frequency in population databases (gnomAD11 and ExAC15), and consensus mutation calls from multiple variant callers. Such criteria are frequently used for determining tumor mutational burden (TMB) from whole-exome sequencing data.20
This function also can also remove a preset list of commonly mutated genes,\textsuperscript{21} or a custom set of genes. Finally, data is processed in definable chunks of lines (default of 10,000 lines), which is intended to help filter large MAF files without getting “out of memory” issues.

**Visualizations of summarized mutation data**

MAFDash consists of various functions for visualizing summarized mutation data across a cohort of samples. Below are the different functions that are provided.

- \texttt{generateBurdenPlot(...)}: It generates a dotplot and a barplot to show the comparison of the total number of mutations across the samples. The mutations are also grouped based on its type.
- \texttt{generateMutationTypePlot(...)}: It generates a barplot showing the distribution of the silent and non-silent mutations across the input samples.
- \texttt{generateOncoPlot(...)}: It generates a heatmap that summarized the top mutated genes across the input samples.
- \texttt{generateOverlapPlot(...)}: It generates a circular plot to show the common mutations across the input samples.
- \texttt{generateRibbonPlot(...)}: It generates a heatmap to show the cosine similarity between the mutated genes using the result from maftools' \texttt{somaticInteractions(...)} function.
- \texttt{generateTiTvPlot(...)}: It plots the frequency of transitions and transversions of the gene mutations in the input datasets.
- \texttt{generateTCGAComparePlot(...)}: It computes and plots the mutation load of the input MAF against all 33 of the TCGA cohorts derived from MC3 project. It also calculates the significant mutational load differences between the cancers.

**Mutational signatures and etiologies**

Mutational signature matrix for single-base substitutions (SBS) were retrieved from COSMIC v3.2.\textsuperscript{22} Text in the “Acceptance criteria” section of each signature page was retrieved from the COSMIC website using R scripts. This free text was lightly filtered and manually curated yielding 25 broad categories for 78 total signatures and is provided with the package repository in tabular format (Table 1).

| Signature | Etiology (Scraped) | Etiology (Broad) |
|-----------|--------------------|------------------|
| SBS45     | 8-oxo-guanine introduced during sequencing | 8-oxo-guanine introduced during sequencing |
| SBS24     | Aflatoxin exposure | Aflatoxin exposure |
| SBS5      | Aging/Tobacco smoking/NER deficiency | Aging/Tobacco smoking/NER deficiency |
| SBS84     | AID activity | AID activity |
| SBS85     | AID activity | AID activity |
| SBS2      | APOBEC activity | APOBEC activity |
| SBS13     | APOBEC activity | APOBEC activity |
| SBS22     | Aristolochic acid exposure | Aristolochic acid exposure |
| SBS32     | Azathioprine exposure | Azathioprine exposure |
| SBS30     | BER deficiency | BER deficiency |
| SBS36     | BER deficiency | BER deficiency |
| SBS88     | Colibactin exposure | Colibactin exposure |
| SBS17a    | Damage by ROS | Damage by ROS |
| SBS18     | Damage by ROS | Damage by ROS |
| Signature | Etiology (Scraped)                  | Etiology (Broad)                  |
|-----------|-----------------------------------|----------------------------------|
| SBS17b    | Damage by ROS/5FU chemotherapy     | Damage by ROS                    |
| SBS10c    | Defective POLD1 proofreading       | Defective POLD1 proofreading     |
| SBS10d    | Defective POLD1 proofreading       | Defective POLD1 proofreading     |
| SBS90     | Duocarmycin exposure               | Duocarmycin exposure             |
| SBS54     | Germline variants contamination    | Germline variants contamination  |
| SBS42     | Haloalkanes exposure               | Haloalkanes exposure             |
| SBS3      | HR deficiency                      | HR deficiency                    |
| SBS8      | HR deficiency/NER deficiency       | HR deficiency                    |
| SBS6      | MMR deficiency                     | MMR deficiency                   |
| SBS15     | MMR deficiency                     | MMR deficiency                   |
| SBS21     | MMR deficiency                     | MMR deficiency                   |
| SBS26     | MMR deficiency                     | MMR deficiency                   |
| SBS44     | MMR deficiency                     | MMR deficiency                   |
| SBS20     | MMR deficiency + POLD1 mutation    | MMR deficiency                   |
| SBS14     | MMR deficiency + POLE mutation     | MMR deficiency                   |
| SBS31     | Platinum chemotherapy              | Platinum chemotherapy            |
| SBS35     | Platinum chemotherapy              | Platinum chemotherapy            |
| SBS10a    | POLE exonuclease domain mutation   | POLE exonuclease domain mutation |
| SBS10b    | POLE exonuclease domain mutation   | POLE exonuclease domain mutation |
| SBS28     | POLE exonuclease domain mutation   | POLE exonuclease domain mutation |
| SBS9      | Polymerase eta somatic hypermutation|                                 |
| SBS43     | Possible sequencing artifact       | Sequencing artifact              |
| SBS51     | Possible sequencing artifact       | Sequencing artifact              |
| SBS55     | Possible sequencing artifact       | Sequencing artifact              |
| SBS56     | Possible sequencing artifact       | Sequencing artifact              |
| SBS57     | Possible sequencing artifact       | Sequencing artifact              |
| SBS58     | Possible sequencing artifact       | Sequencing artifact              |
| SBS59     | Possible sequencing artifact       | Sequencing artifact              |
| SBS27     | Sequencing artifact                | Sequencing artifact              |
| SBS60     | Sequencing artifact                | Sequencing artifact              |
| SBS47     | Sequencing artifact (blacklisted cancer samples for poor quality) | Sequencing artifact |
| SBS48     | Sequencing artifact (blacklisted cancer samples for poor quality) | Sequencing artifact |
| SBS49     | Sequencing artifact (blacklisted cancer samples for poor quality) | Sequencing artifact |
| SBS50     | Sequencing artifact (blacklisted cancer samples for poor quality) | Sequencing artifact |
| SBS52     | Sequencing artifact (blacklisted cancer samples for poor quality) | Sequencing artifact |
| SBS53     | Sequencing artifact (blacklisted cancer samples for poor quality) | Sequencing artifact |
| SBS46     | Sequencing artifact (early releases of TCGA) | Sequencing artifact |
| SBS1      | Spontaneous deamination of 5-methylcytosine | Spontaneous deamination of 5-methylcytosine |
Etiology annotations for mutational signature analysis

To aid interpretation of mutational signature analysis, we have curated COSMIC signatures etiologies from COSMIC v3.2.22 Specifically, we scraped the COSMIC website to retrieve the proposed etiology for all 78 COSMIC single-base substitution (SBS) signatures, yielding 36 unique etiologies, which we further manually curated into 25 broad categories.

The generateCOSMICMutSigSimHeatmap(…) function shows these categorized proposed etiologies as colored row annotations, aimed at quickly identifying distinct or common etiologies across a cohort. Figure 1 shows the SBS signature in each sample in columns, COSMIC mutation signatures in rows, and each cell is colored to indicate the level of similarity between the two.

Use cases
Interactive HTML reports for MAF data

MAFDash has a function (getMAFDashboard(…)) that generates an HTML dashboard for visualization and analysis of mutation data in MAF format. The dashboard consists of arbitrarily defined or preset interactive plots describing the data. By default, if MAF data is provided, the dashboard visualizes the mutations data in five different tabs.

- Summary plots: Static multi-part figure describing cohort summaries of variant classification, variant type, number of variants per samples and nucleotide change (from ‘maftools’).
- Burden plots: Interactive plots showing the number of variants per samples in the form of a dotplot and barplot, with hover text containing sample and mutation information.

| Signature | Etiology (Scraped)                                | Etiology (Broad) |
|-----------|--------------------------------------------------|------------------|
| SBS11     | Temozolomide chemotherapy/MMR deficiency +       | MMR deficiency   |
|           | temozolomide                                     |                  |
| SBS87     | Thiopurine chemotherapy                          | Thiopurine       |
|           | chemotherapy                                     | chemotherapy     |
| SBS29     | Tobacco chewing                                  | Tobacco          |
| SBS4      | Tobacco smoking                                  | Tobacco          |
| SBS92     | Tobacco smoking                                  | Tobacco          |
| SBS12     | Unknown                                          | Unknown          |
| SBS16     | Unknown                                          | Unknown          |
| SBS19     | Unknown                                          | Unknown          |
| SBS23     | Unknown                                          | Unknown          |
| SBS33     | Unknown                                          | Unknown          |
| SBS34     | Unknown                                          | Unknown          |
| SBS37     | Unknown                                          | Unknown          |
| SBS39     | Unknown                                          | Unknown          |
| SBS40     | Unknown                                          | Unknown          |
| SBS41     | Unknown                                          | Unknown          |
| SBS89     | Unknown                                          | Unknown          |
| SBS91     | Unknown                                          | Unknown          |
| SBS93     | Unknown                                          | Unknown          |
| SBS94     | Unknown                                          | Unknown          |
| SBS25     | Unknown chemotherapy                             | Unknown          |
| SBS86     | Unknown chemotherapy                             | Unknown          |
| SBS7a     | UV light exposure                                 | UV light exposure |
| SBS7b     | UV light exposure                                 | UV light exposure |
| SBS7c     | UV light exposure                                 | UV light exposure |
| SBS7d     | UV light exposure                                 | UV light exposure |
| SBS38     | UV light exposure (indirect effect)              | UV light exposure |
Figure 1. Heatmap showing the cosine similarity between the mutational signatures from The Cancer Genome Atlas’s Adrenocortical carcinoma cohort with Catalogue Of Somatic Mutations In Cancer (COSMIC) signatures.

Figure 2. Snapshot of the oncoplot tab of the HTML dashboard created by the getMAFDashboard(...) function using the TCGA Adrenocortical Carcinoma (ACC) dataset.
Oncoplot: Plot summarizing the top mutated genes across the samples.

Co-occurrence of mutated genes: A circular ribbon plot showing co-occurrence of the mutations, inspired by the `somaticInteractions(...)` function in `maftools`.

Interactive heatmap: An interactive version of the oncoplot with hover text showing the number of mutations in a gene for a particular sample.

In addition to these plots, an interactive table is generated using the `DT` and `crosstalk` R packages to provide client-side, dynamic filtering of the variant data. The generated dashboard is self-contained for sharing with collaborators. MAFDash will automatically account for missing data and also provides reasonable defaults for filtering mutation data. Figure 2 shows the dashboard output for Adrenocortical carcinoma (ACC) downloaded from TCGA.

HTML reports for arbitrary plots
Even without MAF data, MAFDash can be used to generate an HTML report with user generated plot objects. Users can pass any `ggplot2`, `ComplexHeatmap`, or `plotly` objects, or the location of an image file to include it in the dashboard as a list, and have it rendered as a dashboard with each element as a tab in the report. Figure 3 shows an example dashboard using the `iris` dataset provided with R.

Conclusions
We developed MAFDash to simplify the process of generating interactive reports for somatic mutation analysis. The ‘maftools’ R package already provides a comprehensive toolkit for organizing and analyzing MAF data, but it exclusively uses base R graphics for plotting, which is not amenable to further modification or interactivity. For example,
the tcaCompare(…) function is an excellent visual comparison of mutation burden with all cancer types in TCGA. To allow interactivity in MAFDash, we implemented the same visualization using ‘ggplot2’, which can trivially be converted to an interactive HTML widget using ‘plotly’. Finally, the self-contained nature of the HTML report, as well as a range of choices for interactive plots, is aimed at easily sharing data and interpretations. Overall, we hope that MAFDash will allow for quick iterations of analysis during collaborations between bioinformaticians and bench scientists.

Data availability
All data underlying the results are available as part of the article and no additional source data are required.

Software availability
- Source code available at: https://github.com/CCBR/MAFDash
- Archived source code at time of publication: https://doi.org/10.5281/zenodo.6421833
- License: MIT License

Acknowledgements
We would like to thank CCR Collaborative Bioinformatics Resource (CCBR) members for their feedback.

References
1. Kris A: Wetterstrand. The cost of sequencing a human genome. 2022. Reference Source
2. Consortium Genomes ProjectAuton A, Brooks LD, et al.: A global reference for human genetic variation. Nature. 2015; 526(7577): 68–74. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). PubMed Abstract | Publisher Full Text
3. Gudmundsson S, Singer-Berk M, Watts NA, et al.: Variant interpretation using population databases: Lessons from gnomad. Hum. Mutat. 2021. ISSN 1098-1004 (Electronic) 1059-7794 (Linking). PubMed Full Text | Reference Source
4. Cancer Genome Atlas Research NetworkWeinstein JN, Collisson EA, et al.: The cancer genome atlas pan-cancer analysis project. Nat. Genet. 2013; 45(10): 1133–1150. ISSN 1546-1718 (Electronic) 1061-4036 (Linking). PubMed Abstract | Publisher Full Text
5. Barretina J, Caponigro G, Stransky N, et al.: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012; 483(7391): 603–607. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). PubMed Abstract | Publisher Full Text
6. Zheng S, Cherniack AD, Dewal N, et al.: Comprehensive pan-genomic characterization of adenocarcinoma carcinoma. Cancer Cell. 2016; 29(3): 723–736. ISSN 1535-6108. PubMed Abstract | Publisher Full Text | Free Full Text | Reference Source
7. Mayakonda A, Lin DC, Assenov Y, Plass C, et al.: Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018; 28(11): 1747–1756. ISSN 1549-5469 (Electronic) 1088-9051 (Linking). PubMed Abstract | Publisher Full Text
8. Memorial Sloan Kettering Cancer Center: vcf2maf. 2013. Reference Source
9. McLaren W, Gil L, Hunt SE, et al.: The ensemble variant effect predictor. Genome Biol. 2016; 17(1):122. ISSN 1473-7384 (Electronic) 1474-7596 (Linking). PubMed Abstract | Publisher Full Text
10. Wickham H: ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016. ISBN 978-3-319-24277-4. Reference Source
11. Gu Z, Eils R, Schlesner M: Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016; 32(18): 2847–2849. ISSN 1367-4811 (Electronic) 1367-4853 (Linking). PubMed Abstract | Publisher Full Text
12. R Core Team: R: A language and environment for statistical computing. 2020.
13. Gu Z, Gu L, Eils R, et al.: Circize implements and enhances circular visualization in r. Bioinformatics. 2014; 30(19): 2811–2. ISSN 1367-4811 (Electronic) 1367-4803 (Linking). PubMed Abstract | Publisher Full Text
14. Neuhaus I, Brett C: canvashxpress: Visualization Package for CanvasXpress in R. 2022. R package version 1.3.7.4. Reference Source
15. Sievert C: Interactive web-based data visualization with r, plotly, and shiny. 2020. Reference Source
16. Benjamin D, Sato T, Cibuśki K, et al.: Calling somatic SNVs and indels with mutect2. bioRxiv. Publisher Full Text | Reference Source
17. Colaprico A, Silva TC, Olsen C, et al.: Tcgabiolinks: an r/biocpackage for integrative analysis of tcga data. Nucleic Acids Res. 2016; 44(8): e71. ISSN 1362-4922 (Electronic) 0305-1048 (Linking). PubMed Abstract | Publisher Full Text
18. Karczewski KJ, Franciolli LC, Tiao G, et al.: The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020; 581(7809): 434–443. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). PubMed Abstract | Publisher Full Text
19. Karczewski KJ, Weisburd B, Thomas B, et al.: The exac browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res. 2017; 45(D1): D840–D845. ISSN 1362-4962 (Electronic) 0305-1048 (Linking). PubMed Abstract | Publisher Full Text
20. Vilimas T: Measuring tumor mutational burden using whole-exome sequencing. Methods Mol. Biol. 2020. 2055: 63–91. ISSN 1940-6029 (Electronic) 1064-3745 (Linking). PubMed Abstract | Publisher Full Text
21. Shyr C, Tarailo-Graovac M, Gottlieb M, et al.: Flags, frequently mutated genes in public exomes. BMC Med. Genet. 2014; 15: 64. ISSN 1755-8794 (Electronic) 1755-8794 (Linking). PubMed Abstract | Publisher Full Text
22. Alexandrov LB, Kim J, Haradhvala NJ, et al.: The repertoire of mutational signatures in human cancer. Nature. 2020; 578(7793): 94–101. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). PubMed Abstract | Publisher Full Text
23. Xie Y, Cheng J, Tan X: DT: A Wrapper of the JavaScript Library ‘DataTables’. 2022. R package version 0.21. Reference Source
24. Cheng J, Sievert C: *crosstalk: Inter-Widget Interactivity for HTML Widgets*. 2021. R package version 1.2.0.

Reference Source

25. Jain A, Tandon M: *MAFDash: An easy-to-use dashboard builder for mutation data* (0.2.2). Zenodo. 2022.

Publisher Full Text
Open Peer Review

Current Peer Review Status:  ×  ×

Version 1

Reviewer Report 02 September 2022

https://doi.org/10.5256/f1000research.130605.r147677

© 2022 Brennenstuhl H. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Heiko Brennenstuhl

1 Division of Child Neurology and Metabolic Medicine, Centre for Child and Adolescent Medicine, University Hospital Heidelberg, Heidelberg, Germany
2 Institute for Human Genetics, University Heidelberg, Heidelberg, Germany

In their manuscript "MAFDash: An easy-to-use dashboard builder for mutation data" Jain and Tandon report an R package which allows processing and visualization of mutation data from standardized and widely available MAF files and displays the results in an HTML dashboard. Various forms of presentation are generated automatically (tables, burden plots, OncoPlots, interactive heat maps, etc.), which should enable an analysis of the data set through sophisticated visualizations.

The manuscript is well written in an understandable manner and all considerations are clearly and convincingly presented.

Nevertheless, I have some small points of criticism:

1. MAF files are extremely impractical to handle and can be a great challenge, especially for researchers with limited experience in bioinformatics. In my experience, VCF files are much more common and should therefore be included in the workflow of this tool.

2. Some of the links within the documentation on github are broken (e.g. https://mtandon09.github.io/MAFDashRPackage/examples/LAML.mafdash.html and https://mtandon09.github.io/MAFDashRPackage/examples/articles/Quick_Start.html)

3. Local installation does not work (at least in my application): Some dependencies are not available for MAFDash according to my R version (including 'TCGAbiolinks', 'maftools', 'ComplexHeatmap', 'BSgenome-Hsapiens.UCSC.hg38'), the installation is aborted with the remark 'MAFDash_0.2.2.tar.gz' had non-zero exit status'

Is the rationale for developing the new software tool clearly explained?
Yes
Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: human genetics, inborn errors of metabolism

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 21 July 2022

https://doi.org/10.5256/f1000research.130605.r143350

© 2022 Nakken S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sigve Nakken

1 Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
2 Centre for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway
3 Centre for Cancer Cell Reprogramming, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

Jain and Tandon have created a tool for visualization and analysis of cancer mutation data represented through the MAF format. The tool generates a dashboard as its output, hence the name, MAFdash. The tool offers multiple features to analyse mutation data from a cohort of cases/samples, exemplified through oncoplots (highly mutated genes), burden plots, mutational signature analysis, etc.

Currently, the tool suffers from some key limitations that I encourage the authors to handle in a
revised version.

**Major points**

- The tool is intended for analysis of MAF data, and the authors exemplify this through data from TCGA. However, most users will primarily be interested in using MAFdash on their own data rather than TCGA/CCLE. Importantly, users who have done cancer genome sequencing typically end up with VCF files after calling (I am not aware of any somatic variant callers that produce MAF?). If the tool does not support any transformation of VCF towards MAF, I am afraid the tool will not be used according to the intentions outlined by the authors, that is to support users with analysis/visualization of mutation data. In other words, the whole premise of the tool is to have a MAF file at hand, but obtaining this from a large-scale sequencing project is not covered by the tool, nor does the tool provide pointers/workflows to how this can be accomplished. Showcasing a complete workflow from variant calling towards MAF towards MAFdash would be helpful for the users, and strengthen the tool significantly. Based on the reasoning above, I think the slogan "Once you call the variants, it's a MAFDash to the finish line" is in my opinion somewhat misleading.

- Technically, the tool suffers from multiple issues:

  1. [https://github.com/CCBR/MAFDash](https://github.com/CCBR/MAFDash) contains a number of pointers to [https://github.com/ashishjain1988/MAFDash](https://github.com/ashishjain1988/MAFDash), this needs to be cleaned up. Similarly, the documentation site [https://ashishjain1988.github.io/MAFDash/](https://ashishjain1988.github.io/MAFDash/) contains links to [https://mtandon09.github.io/MAFDashRPackage/](https://mtandon09.github.io/MAFDashRPackage/). Please clean up the GitHub page and the accompanying documentation site. Most importantly, links to the example reports are non-functioning, which are critically important to showcase the output of the tool. Currently, I am unable to explore any output examples from the tool.

  2. The installation procedure is not working properly, the DESCRIPTION file needs cleaning:
     - Addition of *biocViews*: (for Bioconductor packages)
     - Move the (large) BSgenome package to *Suggests*
     - Clean out the *Depends* stuff, just keep R there. Rest go into *Suggests/Imports*.
     - Remove most of the version pinning in the *Imports*
     - Make sure the installation works both on Mac OSX and Linux

     - Looking at the function reference (https://ashishjain1988.github.io/MAFDash/reference/index.html), there are a number of misleading elements, i.e.

       ```r
generateOncoPlot() - Function to generate a dashboard from a MAF file?
filterMAF() and filterMAF2()?  
compute_exome_coverage() - is this relevant function for MAFdash?
```

**Minor points**

The example case for arbitrary plots using the *iris* dataset is misleading, please provide a relevant dataset.

**Is the rationale for developing the new software tool clearly explained?**
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
No

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, translational bioinformatics, precision cancer medicine

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com