
Application Notes

BodyMapR: an R package and Shiny application designed to generate anatomical visualizations of cancer lesions

David M. Miller ^{1,2} and Sophia Z. Shalhout ¹

¹Division of Hematology/Oncology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, and
²Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts, USA

Corresponding Author: David M. Miller, MD, PhD, Division of Hematology/Oncology, Department of Medicine, Massachusetts General Hospital, 15 Parkman Street, Room 132, Boston MA 02114, USA; dmiller4@mgh.harvard.edu

Received 11 December 2021; Revised 7 February 2022; Editorial Decision 10 February 2022; Accepted 22 February 2022

ABSTRACT

Objectives: Structured real-world data (RWD), such as those found in cancer registries, provide a rich source of information regarding the natural history of cancer. Interactive data visualizations of cancer lesions can provide insights into certain clinical tumor characteristics (CTC). Software that can be integrated into an oncological data collection effort and generate anatomical data visualizations of CTC are limited.

Materials and Methods: We created BodyMapR: an R package and Shiny application that generates anatomical visualizations of cancer lesions from structured data.

Results: BodyMapR is a Shiny application that transposes structured data from REDCap[®] onto an anatomical map to yield an interactive data visualization.

Conclusions: BodyMapR is freely available under the MIT license and can be obtained from GitHub. BodyMapR is executed in R and deployed as a Shiny application. It can be integrated into an existing cancer research platform and produces an interactive data visualization of CTC.

Key words: data visualization, cancer, Shiny app, REDCap, Merkel cell carcinoma

LAY SUMMARY

Large-scale data collection efforts in rare cancers are challenging and uncommon. Consequently, we lack a comprehensive understanding of clinical tumor characteristics (CTC), such as patterns of metastatic spread and biomarkers predictive of treatment response, for most rare tumors. Data collection efforts that incorporate data captured during real-world practice (a.k.a real-world data) can improve our understanding of CTC. Depicting RWD, for example, from a cancer registry, onto graphical representations of anatomical structures can provide a user-friendly technique to process information regarding CTC. However, displaying large amounts of RWD onto anatomical data visualizations is labor-intensive and time consuming. Currently, there is a dearth of software that can facilitate this process. Here, we present BodyMapR, a novel software that generates an interactive visualization of CTC from RWD. The package is freely available and modifiable by end users.

INTRODUCTION

Large-scale data collection efforts in rare cancers, such as Merkel cell carcinoma (MCC), are challenging and uncommon.¹ The dearth of available data sets have limited our understanding of the natural history of rare cancers, such as MCC.²⁻⁴ Consequently, we lack a comprehensive understanding of clinical tumor characteristics (CTC), such as patterns of metastatic spread and biomarkers predictive of treatment response, for most rare tumors. Data collection efforts that incorporate structured data captured during real-world practice (a.k.a Real-World Data or RWD) can improve our understanding of CTC.

Depicting RWD, for example, from a cancer registry, onto graphical representations of anatomical structures can provide a user-friendly technique to process information regarding CTC. An improved understanding of the geographical lesion profile (GLP) of a cancer type may provide important insights. For example, although cancer types have been historically grouped based on the tissue of origin (eg, “lung cancer” or “pancreatic cancer”), neoplasms originating from the same tissue can have clinically relevant distinctions in pathogenesis. For example, in MCC, at least 2 distinct transforming mechanisms (eg, the Merkel cell polyoma virus and ultraviolet radiation), with distinct underlying mutational landscapes, have been described.² Data demonstrating geographical differences in virus-positive versus virus-negative MCC has emerged.⁵ Thus, further definition of the relationship between topography and mutational landscape may lead to insights in pathogenesis across cancer types.

However, displaying large amounts of RWD onto anatomical data visualizations is labor-intensive and time consuming. While informatic packages that generate modular visualizations of anatomi-

grams and tissues are available,⁶ software that fully integrates data collection instruments for real-time anatomical data visualizations of cancer registry data are lacking.

We previously published an overview of a methodology and design of a Research Electronic Data Capture (REDCap®)⁷-based system to facilitate capture of RWD.^{1,8} That platform incorporates a form entitled the Lesion Information instrument, which provides a structured format for the collection of CTC.⁹ This instrument is freely available and can be incorporated into any existing REDCap® project. It is currently being used by the Project Data Sphere led MCC Patient Registry.¹

Here, we present BodyMapR, an R package with a Shiny application front-end, which generates an interactive data visualization of CTC. Its software wrangles and transforms structured data from a REDCap® project and provides graphing functions (Figure 1). BodyMapR is executed in R but is deployed as a Shiny application to enhance the user interface for users with limited programming capabilities. In this article, we provide (1) instructions on how to obtain and execute BodyMapR, (2) the R code for the server-side functions to allow for project-specific adaptations, (3) a Biorender-generated png file in which data is overlaid and displayed, and (4) a sample dataset for demonstration purposes.

MATERIALS AND METHODS

Software dependencies

BodyMapR is written in R (version 4.0.0), organized using roxygen2,¹⁰ and utilizes the following packages dplyr,¹¹ tidyr,¹² readr,¹³ stringr,¹⁴ purrr,¹⁵ magrittr,¹⁶ plotly,¹⁷ shinydashboard,¹⁸ and Shiny.¹⁹ For full details, instructions and examples refer to the video

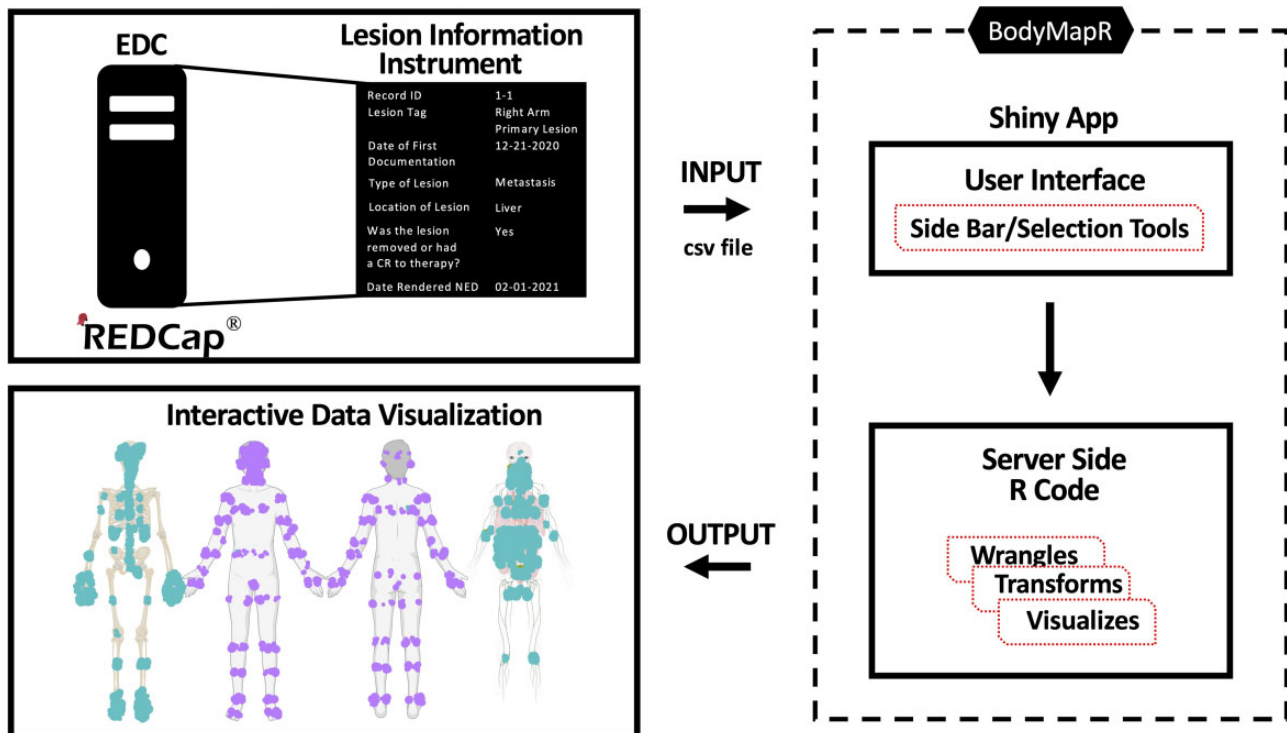


Figure 1. Schema of BodyMapR. BodyMapR takes data from a REDCap® project that incorporates the Lesion Information and Genomics instruments. This csv file is loaded into the Shiny application and end users engage BodyMapR via a browser-based interface. Server-side R code executes the functions of BodyMapR to generate an interactive Plotly visualization of clinical tumor characteristic data displayed onto an anatomical body map. Anatomical images created with BioRender.com.

demonstration (https://github.com/TheMillerLab/BodyMapR/blob/main/Video_Demo.md), or README file (<https://github.com/TheMillerLab/BodyMapR/blob/main/README.md>), both of which can be viewed on the package GitHub page.

Clinical informatics dependencies

BodyMapR facilitates data visualizations from structured data contained in the Lesion Information instrument stored within REDCap® project. The data dictionary for this form has been previously published.²⁰ BodyMapR also integrates clinico-genomic data from the Genomics data capture instrument, which has been previously described^{8,21} and is freely available on GitHub (https://github.com/TheMillerLab/genetex/blob/main/data-raw/genomics_data_dictionary.csv).

RESULTS

BodyMapR inputs

As depicted in Figure 1, BodyMapR takes data from a REDCap® project that has incorporated the Lesion Information and Genomics instruments as the input. The BodyMapR Shiny application is launched via the function `launch_BodyMapR()`. This function takes one argument, “Data,” a raw csv file exported from REDCap®. `launch_BodyMapR()` is the only function an end user needs to call to execute and utilize BodyMapR. Once launched, clinical researchers interface with BodyMapR in a web browser. The application’s browser-based user interface (UI) facilitates its use by investigators with limited programming skills. `launch_BodyMapR()` has a built-in default data set “BodyMapR_mock_dataset.” If the argument “Data” is not specified by an end user, the default data set will be incorporated into the application for demonstration purposes.

“BodyMapR_mock_dataset” is a synthetic data set and contains no protected health information.

BodyMapR UI

The centerpiece of the BodyMapR UI is an anatomical landscape, henceforth referred to as the “Body Map” (Figure 2). The Body Map includes a skeleton, the anterior and posterior likeness of an androgynous adult, and representations of visceral and lymphatic structures. We designed the Body Map using images from BioRender.com.²² Users control what information is displayed onto the Body Map via the application’s UI sidebar. Given that an improved understanding of the GLP of a cancer type may provide insight into patterns of spread, the default settings of BodyMapR display the GLP of the entire cohort, color-coded by tumor morphology (eg, primary vs metastasis vs recurrence). In contrast, a personalized Body Map at the single-subject level can be obtained by selecting a Record ID from the `selectizeInput()` selector “Filter on Record ID” in the application’s sidebar (Supplementary Figure S1).

As stated above, further definition of the relationship between topography and mutational landscape may lead to insights in pathogenesis across cancer types. Thus, BodyMapR incorporates the Genomics Instruments in order to display over 900 genes found in common clinico-genomics platforms.⁸ Users can select which genes to be visualized on the Body Map using the “Filter on Gene Mutation” selector (Figure 2).

BodyMapR server-side functions

The server side of the Shiny application contains the executable code of BodyMapR. The package contains a set of functions that wrangles, transforms, and graphs CTC data from a REDCap® project (Figure 1). Table 1 summarizes the package’s functions and their respective action.

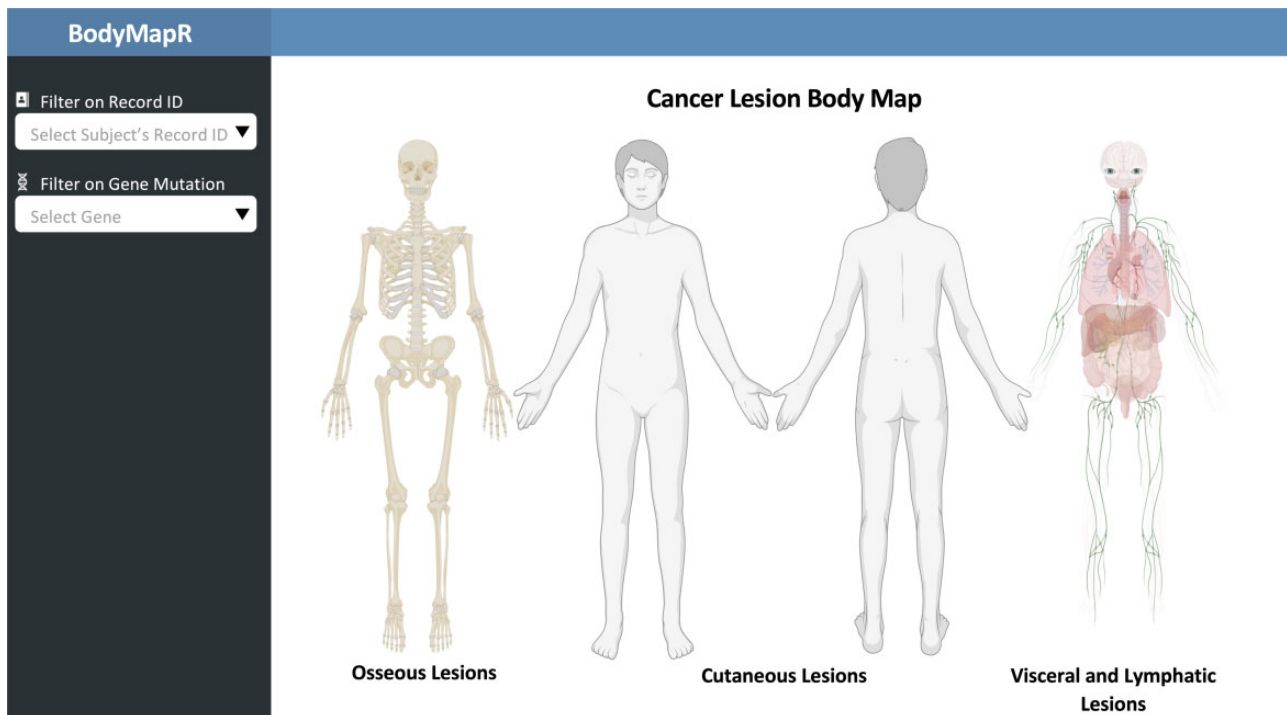


Figure 2. Browser-based user interface. Users control what input is displayed onto the BodyMapR anatomical graphic using the sidebar selectors. Anatomical images created with BioRender.com.

Table 1. BodyMapR functions

Functions	Function
bodymapr_df()	Creates a data frame of clinical tumor characteristics from a structured electronic data collection instrument (eg, the Lesion Information instrument) and maps them to lookup tables that contain a coordinate system for the Body Map
bodymapr_plot()	Creates an interactive data visualization of clinical tumor characteristics from a structured EDC lesion instrument that has been processed by bodymapr_df()
genomics.df.unite()	Wrangles and processes genomics data from a REDCap project that has incorporated the Genomics Instrument. Genetic alterations are listed in wide format with a concatenation of genomic alterations in 1 cell. This function is called within BodyMapR's bodymapr_df() function
genomics.df.long()	Wrangles and processes genomics data from a REDCap project that has incorporated the Genomics Instrument. This allows for expedited analysis of patient-level data from REDCap. Genetic alterations are listed in long format. This function is called within BodyMapR's genomics.df.unite() function
shiny_server()	Contains the server side of the Shiny application. It incorporates both bodymapr_df() and bodymapr_plot(). Therefore, the data are wrangled, processed, and graphed with this function
shiny_ui()	Creates the user interface of BodyMapR
launch_BodyMapR()	Launches the BodyMapR Shiny application

Note: Key functions unique to BodyMapR with a brief description of their action are shown.

BodyMapR outputs

Topographical mapping

As described previously,²⁰ the Lesion Information instrument was designed to capture 651 distinct anatomical structures, including skeletal, cutaneous, lymphatic, mucosal, and visceral locations (Supplementary Table S1). To construct a data visualization of these anatomical structures, individual structures were mapped to the Body Map using an X-Y grid-based coordinate system. To assist in topographical mapping, an interactive coordinate system was generated via plotly¹⁷ (code below).

```
body_map.png <- BodyMapR::BodyMapR_biorender.png
grid.2 <- data.frame(x = rep((seq(from = 1 to = 100, by = 0.5)),
                        each = 200),
                    y = rep((seq(from = 1, to = 100, by = 1)),
                        len = 100))
```

```
df.grid <- data.frame()
grid.plotly <- ggplot(df.grid) +
  xlim(0, 100) +
  ylim(0, 100) +
  theme_bw() +
  theme(panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_blank()) +
  theme(axis.text.x = element_blank(),
```

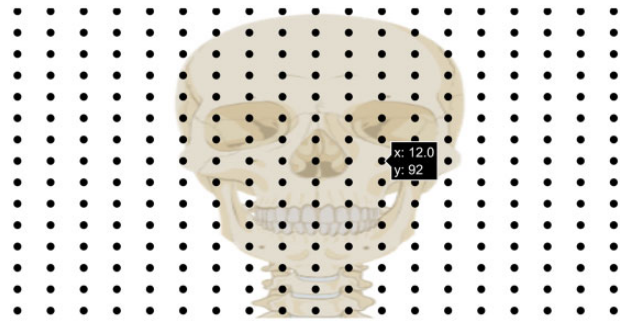


Figure 3. X-Y coordinate system for mapping topographical elements. Using plotly, individual X-Y coordinates are easily visualized and mapped to X-Y coordinates on the Body Map topographical image. Anatomical images created with BioRender.com.

```
axis.text.y = element_blank(),
axis.ticks = element_blank(),
axis.title = element_blank()) +
annotation_raster(body_map.png,
                  ymin = 0,
                  xmin = 0,
                  xmax = 100,
                  ymax = 100) +
geom_point(data = grid.2,
           aes(x = x,
              y = y))
ggplotly(grid.plotly)
```

As seen in Figure 3, individual X-Y coordinates are visualized by utilizing the hover text functionality of plotly. For example, using this method the left zygomatic bone is mapped to $x=12.0$ and $y=92.0$ (Figure 3). This approach was used to map every anatomical structure in the Lesion Information instrument.

The derived X-Y coordinates were then incorporated into lookup tables for each topographical region. For example, Supplementary Table S2 displays the osseous lookup table, in which each of the 282 skeletal structures in the Lesion Information instrument are mapped to a unique X and Y combination. Lookup tables for each portion of the topographical map are embedded in the BodyMapR function body_map.df().

The data-processing function body_map.df() takes one argument (“data”). It returns a data frame that is then used for the BodyMapR graphing functions embedded in the server side of the Shiny application. The “data” argument taken by bodymapr_df() is a raw csv file exported from a REDCap[®] project for which the Lesion Information and Genomics instruments have been incorporated. bodymapr_df() wrangles the raw csv file and uses the dplyr function left_join() to combine the various anatomical lookup tables to the REDCap[®] project data to map each lesion to its corresponding topographical location (Supplementary Figure S2). An example of the output of bodymapr_df() is seen in Supplementary Table S3.

Interactive topography

The output of bodymapr_df() is then used as the argument for the BodyMapR graphing function bodymapr_plot() (Supplementary Figure S3). In addition to providing a static GLP of the cohort, BodyMapR was designed to display an interactive graphic to increase the amount of information presented by the package without compromising the user experience with data overload. Utilizing hover text functionality via plotly, bodymapr_df() provides detailed

and clinically relevant CTC information (Supplementary Figure S1). Given that therapeutic strategies in oncology are increasingly determined by genomic alterations in tumor specimens, the default parameters of `bodymapr_df()` display individual genomic alterations and tumor mutational burden in the hover text. In addition, since tumor size is an important CTC, often with staging implications, individual lesions are sized proportional to the longest axis of their clinical measurement.

BodyMapR customizations

The BodyMapR source code has been made freely available to provide end users with the ability to customize the user interface and server-side functions to best suit their clinical research needs. As an example, BodyMapR is currently being used by the Project Data Sphere-supported MCC patient registry.³ Therefore, modifications have been made to incorporate specific CTC salient to MCC. For example, the MCC BodyMapR application has incorporated sidebar selectors with disease-specific information such as Merkel Cell polyoma virus-related data (Supplementary Figure S4). Analogous adaptations can be made on a project-specific basis by end users to meet their clinical research needs.

Furthermore, because BodyMapR can function as a stand-alone R package (ie, outside of a Shiny application), users are able to customize which data are mapped onto the Body Map inside an integrated development environment (IDE), such as RStudio[®].²³ For example, if an investigator intends to visualize a subset of the cohort that contained liver lesions, the following code could be used:

```
# Create a data frame containing only those observations with liver
# or hepatic metastases
liver_lesions<- BodyMapR::BodyMapR_mock_dataset %>%
  filter(str_detect(string = BodyMapR::BodyMapR_mock_dataset$
  lesion_tag,
    pattern = regex("liver|hepatic", ignore_case = TRUE)))
# Filter the mock dataset with liver lesions to map only those sub-
# jects that have liver lesions
BodyMapR::BodyMapR_mock_dataset %>%
  filter(record_id %in% liver_lesions$record_id) %>%
  BodyMapR::bodymapr_df() %>%
  BodyMapR::bodymapr_plot()
```

The above code will generate an interactive Body Map data visualization within an IDE (Supplementary Figure S5). Of note, as seen above, the functions of BodyMapR are compatible with tidyverse syntax and use of the magrittr pipe operator.¹⁴

LIMITATIONS AND SOLUTIONS

BodyMapR is to be used in conjunction with the Lesion Information and Genomics Information instrument; therefore, its functions optimally when those forms are installed into a REDCap[®] project. Thus, we have made the data dictionaries freely available so that others may incorporate them into their individual projects. However, if an investigator has incorporated a lesion field that uses other common anatomical coding systems, such as the International Classification of Diseases for Oncology, 3rd edition (ICD-O-3) or the Systematized Nomenclature of Medicine (SNOMED), we have incorporated a lookup table that converts these fields into coordinates on the BodyMapR Body Map (Supplementary Table S4).

In addition, although BodyMapR is designed to be used with a specific anatomical image, which is incorporated into the package, the bespoke X-Y coordinate system we described above can be used to create lookup tables for any graphic. Therefore, an end user can utilize alternate images when appropriate.

CONCLUSIONS

Data visualizations of CTC can improve our understanding of the natural history of malignancies. Tools that streamline topographical profiling of cancer lesions from structured data collection systems are time consuming to design and require knowledge of software engineering and application development. These resources may be limited on clinical research teams and thus, customizable and freely available, web-based tools are an unmet need. BodyMapR is an open-sourced R package that is deployed as a Shiny application that can integrate into an existing cancer research platform and produces an interactive data visualization of CTC.

FUNDING

The Harvard Cancer Center Merkel Cell Carcinoma patient registry is supported by grants from Project Data Sphere, ECOG-Acrin, and the American skin association. SZS was also funded by the MGH-ECOR Fund for Medical Discovery Clinical Research Grant.

AUTHOR CONTRIBUTIONS

DMM created and developed the BodyMapR package, authored the article, and granted final approval of the article. SZS contributed to the development of the BodyMapR package, including code writing, participated in the authorship of the article, and granted final approval of the article.

SUPPLEMENTARY MATERIAL

Supplementary material is available at JAMIA Open online.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data/code for this application can be found in our GITHUB repository: (<https://github.com/TheMillerLab/BodyMapR>).

REFERENCES

1. Miller DM, Shalhout SZ, Saqlain F, *et al*. *The Merkel Cell Carcinoma Patient Registry-From Promise to Prototype to Patient*. 2021. <https://themillerlab.io>. Accessed January 10, 2022.
2. Harms PW, Harms KL, Moore PS, *et al*.; International Workshop on Merkel Cell Carcinoma Research (IWMCC) Working Group. The biology and treatment of Merkel cell carcinoma: Current understanding and research priorities. *Nat Rev Clin Oncol* 2018; 15 (12): 763–76.
3. Shalhout SZ, Emerick KS, Kaufman HL, Miller DM. Immunotherapy for non-melanoma skin cancer. *Curr Oncol Rep* 2021; 23 (11): 125.
4. Shalhout SZ, Kaufman HL, Emerick KS, Miller DM. Immunotherapy for nonmelanoma skin cancer: facts and hopes [published online ahead of

- print February 4, 2022]. *Clin Cancer Res* 2022; doi:10.1158/1078-0432.CCR-21-2971.
5. Schrama D, Peitsch WK, Zaparka M, et al. Merkel cell polyomavirus status is not associated with clinical course of Merkel cell carcinoma. *J Invest Dermatol* 2011; 131 (8): 1631–8.
 6. Maag JLV. Gganatogram: An r package for modular visualisation of anagrams and tissues based on ggplot2. *F1000Res* 2018; 7: 1576.
 7. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42 (2): 377–81.
 8. Shalhout SZ, Miller DM. *Optimizing Real-World Data Collection: Clinical Genomics*. 2020. https://www.themillerlab.io/post/optimizing_rwd_collection-clinical_genomics/. Accessed October 16, 2021.
 9. Miller DM, Shalhout SZ. *Optimizing Real-World Data Collection: Genomics Electronic Data Capture Instrument*. 2021. https://www.themillerlab.io/post/optimizing_rwd_collection-lesion_information_instrument/. Accessed December 10, 2021.
 10. Wickham H, Danenberg P, Eugster M. *Roxygen2: In-Source Documentation for R (version 6.0.1)*. R package Version 7.1.1. 2013. <https://cran.r-project.org/package=roxygen2>.
 11. Wickham H, Francois R, Henry L, Muller K. *Dplyr: A Grammar of Data Manipulation*. R Package Version 1.0.5. 2021.
 12. Wickham H. *Tidyr: Tidy Messy Data*. R package Version 1.1.3. 2013.
 13. Wickham H, Hester J, Francois R. *Readr: Read Rectangular Text Data*. R package Version 1.4.0. 2020.
 14. Wickham H. *Strings: Simple, Consistent Wrappers for Common String*. R package Version 1.4.0. 2019.
 15. Henry L, Wickham H. *Purrr: Functional Programming Tools*. R package Version 0.3.4. 2020.
 16. Bache SM, Wickham H, Henry L. *Magrittr: A Forward-Pipe Operator for r*. R package Version 2.0.1. 2020.
 17. Sievert C. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Boca Raton, FL: Chapman and Hall; 2020.
 18. Chang W, Ribeiro BB; RStudio, Almasaeed Studio and Adobe Systems Incorporated. *Shinydashboard: Create Dashboards With ‘Shiny’*. R Package Version 1.0.5. 2021.
 19. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. *Shiny: Web Application Framework for r*. R package Version 1.6.0. 2018.
 20. Miller DM, Saqlain F, Shalhout SZ. *Optimizing Real-World Data Collection: The Lesion Information Electronic Data Capture Instrument*. 2021. <https://themillerlab.io>. Accessed January 10, 2022.
 21. Miller DM, Shalhout SZ. GENETEX-a GENomics report TEXT mining R package and Shiny application designed to capture real-world clinico-genomic data. *JAMIA Open* 2021; 4 (3): ooab082.
 22. BioRender. 2021. <https://biorender.com>. Accessed October 16, 2021.
 23. RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC; 2020.