The Journal of Open Source Software

gimap: An R Package for Genetic Interaction Mapping in Dual-Target CRISPR Screens

Candace Savonen [1,2], **Phoebe C. R. Parrish** [1], **Kathryn J. Isaac** [1], **Daniel Groso**[1], **Marissa Fujimoto**[1], **Siobhan O'Brien** [1], **and Alice H. Berger** [1]

1 Fred Hutchinson Cancer Center, United States **2** Synthesize Bio, United States

DOI: 10.21105/joss.09891

Software

- Review
- Repository
- Archive

Editor: Frederick Boehm

Reviewers:

- @kalyanidhusia
- @btmonier
- @tomsing1

Submitted: 04 August 2025
Published: 06 August 2026

Summary

The gimap (Genetic Interaction MAPping) R package addresses a fundamental challenge in genomic research: the difficulty of understanding combinatorial interactions among genes. Gene redundancy makes traditional single-gene knockout methods ineffective for identifying therapeutic targets, as backup genes can mask the effects when a single gene is disabled. gimap offers a solution by providing a comprehensive framework for analyzing dual-target CRISPR screening data, where two genes are simultaneously disabled to reveal their backup relationships. This software implements the methods used by Parrish et al. (2021). The package processes raw count data through a multi-step pipeline that includes normalization, calculation of expected and observed CRISPR scores, computation of genetic interaction scores, and statistical analysis to identify significant interactions. Unlike general tools, gimap is specifically tailored for paired guide CRISPR data with built-in quality control reporting and visualization tools. The package makes best practices the default options and is available on GitHub with comprehensive documentation to support the research community in extracting meaningful insights from complex genetic screening experiments.

Statement of Need

When multiple genes have the same function, a common result of evolutionary processes, it becomes challenging to isolate their true functions. This redundancy means that many possible therapeutic targets are missed by traditional methods that disable just one gene at a time (De Kegel & Ryan, 2019; Parrish et al., 2021). A more complementary approach involves disabling two genes simultaneously to reveal these backup relationships (Thompson et al., 2021).

Recent advances in CRISPR technology now allow researchers to knock out gene pairs at once, offering a powerful solution to this problem (Gonatopoulos-Pournatzis et al., 2020). Although software solutions exist for single knockout CRISPR, such as MAGeCK, there is no standardized software solution for paired gene CRISPR studies (Li et al., 2014).

The R package, called `gimap` (Genetic Interaction MAPping), was developed specifically for analyzing these dual-target CRISPR experiments. It helps researchers identify important relationships between genes, such as when two genes work together or when disabling both creates a dramatic effect that wouldn't occur by disabling either one alone.

`gimap` is specifically tailored to handle the unique characteristics of paired guide CRISPR data, including the distinction between single-targeting and double-targeting constructs and the need to account for differential double-strand break effects. The package seamlessly integrates with data generated using a specialized pgPEN library but can be adapted for most paired guide CRISPR screening approaches (Parrish et al., 2021).

Implementation

gimap addresses this need by providing a comprehensive analytical framework for dual-target CRISPR screening data. The package performs several critical functions: (1) normalization of read count data to account for variable sequencing depth and technical biases, (2) calculation of CRISPR scores that reflect the effect of gene knockouts on cell proliferation, (3) determination of expected CRISPR scores for gene pairs based on single-gene effects, (4) computation of genetic interaction scores that quantify deviations from expected effects, and (5) statistical analysis to identify significant interactions.

Overall design philosophy

In order to ensure usability for the research community, we built gimap using the following design philosophy:

1. Making best practices the default options and including warning messages for when alternative options are chosen (e.g. if filtering has not been applied).
2. Using elements from familiar packages such as fastqc reports (our `run_qc()` function creates such a report) (*Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data*, n.d.).
3. Trying to document and inform users of the statistics and decisions that have been made by the software clearly.

gimap data handling

gimap implements a multi-step analysis pipeline:

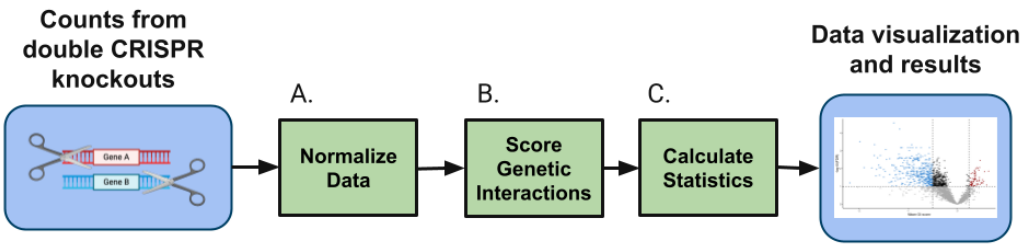

Figure 1: gimap workflow completes 3 main steps. Part A, B, and C of the figure show the major steps of the workflow which are to normalize the data through a multi step process, score genetic interactions based on the expected versus observed CRISPR scores, and finally to calculate statistics to identify statistically significant genetic interactions.

Normalize data

Raw count data is transformed into log2 counts per million (CPM) and adjusted by subtracting pre-treatment values to obtain log2 fold changes. These are further normalized based on the distribution of negative controls (e.g. safe-targeting or non-targeting) and positive controls (pgRNAs targeting known essential genes), analogous to the normalization methods employed by the Cancer Dependency Map (depmap.org) (Arafeh et al., 2025; DepMap, Broad, 2025). Full equations for this normalization procedure are provided in the package documentation.

Score genetic interactions

The goal of this step is to quantify deviations from expected additive effects when two genes are simultaneously targeted, which allows us to identify true genetic interactions beyond what would be predicted from single-gene effects alone. We model a control distribution for

non-interacting genes by assuming that, in the absence of genetic interactions, the effect of simultaneously targeting two genes should be additive (i.e., the sum of their individual effects). We further assume that the observed genetic interaction (GI) score is a linear transformation of the expected GI score plus a consistent error term sampled from an approximately normal, homoscedastic distribution, allowing us to use linear modeling approaches to account for systematic biases.

For double-targeting constructs, expected CRISPR scores are calculated as the sum of the corresponding single-targeting scores. For single-targeting constructs, the expected score combines the single-target effect with the mean effect of control constructs. Interaction scores represent the difference between observed and expected CRISPR scores, adjusted using a linear model to account for systematic biases:

$$I_{i,j} = S_{i,j}^{\text{obs}} - (\beta_0 + \beta_1 \cdot S_{i,j}^{\text{exp}})$$

where $S_{i,j}^{\text{obs}}$ and $S_{i,j}^{\text{exp}}$ are the observed and expected CRISPR scores for genes i and j, and β_0 and β_1 are the intercept and slope from a linear regression of observed vs. expected scores. Full derivations for expected-score calculation and bias adjustment are available in the package documentation.

Calculate statistics

T-tests compare the distribution of double-targeting genetic interaction scores for one pair against the background distribution of single-targeting scores, with false discovery rate correction using the Benjamini-Hochberg procedure for multiple hypothesis testing (Benjamini & Hochberg, 1995). Full statistical formulas are provided in the package documentation.

The package also provides comprehensive visualization tools including volcano plots to highlight significant genetic interactions and detailed result tables for further analysis.

Use Cases

`gimap` has been successfully used to identify synthetic lethal interactions among paralog genes in cancer cell lines, revealing potential therapeutic targets that single-gene approaches have missed. The package accommodates various experimental designs, including time-course studies and treatment comparisons, offering flexibility for diverse research questions.

Example applications include:

- Identification of genes that provide functional redundancy in critical cellular pathways
- Discovery of context-dependent genetic interactions that emerge under specific conditions or treatments
- Systematic mapping of gene networks based on functional interactions rather than physical associations

Conclusion

`gimap` provides a robust, accessible framework for analyzing paired guide CRISPR screening data and identifying genetic interactions with potential biological and therapeutic significance. By streamlining the computational workflow from raw counts to statistically rigorous interaction scores, `gimap` enables researchers to efficiently extract meaningful insights from complex genetic screening experiments. The package is available on GitHub (https://github.com/FredHutch/gimap) with comprehensive documentation and tutorials to facilitate adoption by the research community.

Acknowledgements

This work was funded in part by the Translational Data Science Integrated Research Center (TDS-IRC) of Fred Hutchinson Cancer Center, NIH/NCI R01CA262556 to A.H.B., the American Cancer Society to A.H.B., NIH/NHGRI R25HG012337 to M.F., and a Washington Research Foundation postdoctoral fellowship to S.O.

Related Publications

This software implements methods originally described in Parrish et al. (2021). The current submission focuses on the software implementation and its accessibility to the research community.

Conflict of Interest

The authors declare no conflicts of interest.

AI Usage Disclosure

No generative AI tools were used in the development of this software. AI was used for minor polishing in the preparation of this manuscript.

References

Arafeh, R., Shibue, T., Dempster, J. M., & others. (2025). The present and future of the cancer dependency map. *Nature Reviews Cancer*, *25*, 59–73. https://doi.org/10.1038/s41568-024-00763-x

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. (n.d.). Retrieved April 4, 2025, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

De Kegel, B., & Ryan, C. J. (2019). Paralog buffering contributes to the variable essentiality of genes in cancer cell lines. *PLoS Genetics*, *15*(10), e1008466. https://doi.org/10.1371/journal.pgen.1008466

DepMap, Broad. (2025). *DepMap public 25Q2*. Dataset. https://doi.org/10.1146/annurev-cancerbio-062722-024239

Gonatopoulos-Pournatzis, T., Aregger, M., Brown, K. R., Farhangmehr, S., Braunschweig, U., Ward, H. N., Ha, K. C. H., Weiss, A., Billmann, M., Durbic, T., Myers, C. L., Blencowe, B. J., & Moffat, J. (2020). Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9-Cas12a platform. *Nature Biotechnology*, *38*(5), 638–648. https://doi.org/10.1038/s41587-020-0437-z

Li, W., Xu, H., Xiao, T., Cong, L., Love, M. I., Zhang, F., Irizarry, R. A., Liu, J. S., Brown, M., & Liu, X. S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology*, *15*(12), 554. https://doi.org/10.1186/s13059-014-0554-4

Parrish, P. C. R., Thomas, J. D., Gabel, A. M., Kamlapurkar, S., Bradley, R. K., & Berger, A. H. (2021). Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Reports*, *36*(9), 109597. https://doi.org/10.1016/j.celrep.2021.109597

Thompson, N. A., Ranzani, M., Weyden, L. van der, Iyer, V., Offord, V., Droop, A., Behan, F., Gonçalves, E., Speak, A., Iorio, F., Hewinson, J., Harle, V., Robertson, H., Anderson, E., Fu, B., Yang, F., Zagnoli-Vieira, G., Chapman, P., Del Castillo Velasco-Herrera, M., … Adams, D. J. (2021). Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nature Communications*, *12*(1), 1302. https://doi.org/10.1038/s41467-021-21478-9