The Journal of Open Source Software

drugfindR: Transcriptomic signature analysis and drug repurposing using iLINCS in R

Ali Sajid Imami [1*¶], **Smita Sahay** [1*], **Justin Fortune Creeden** [3], and **Robert Erne McCullumsmith** [1,2]

1 Department of Neurosciences and Psychiatry, University of Toledo, Toledo, OH, USA **2** Promedica Neurosciences Institute, Toledo, OH, USA **3** Department of Psychiatry, University of California San Diego, La Jolla, Diego, CA, USA ¶ Corresponding author * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 02 December 2025
Published: unpublished

Summary

drugfindR is an R package that facilitates mining and analyzing an extensive collection of transcriptomic datasets, precomputed signatures, and their connections, available through the integrated web-based platform iLINCS (Pilarczyk et al., 2022). drugfindR enables users to compare differential gene expression (DGE), or transcriptomic, signatures of interest against standardized chemical perturbation, gene knockdown, or gene overexpression microarray data contained within the Library of Integrated Network-Based Cellular Signatures (LINCS) database (Keenan et al., 2018). The package supports precomputed and user-defined signatures, streamlining hypothesis generation in functional genomics and pharmacological research. The workflows implemented in drugfindR are based on methods developed and validated in our prior publications O'Donovan et al. (2021), which demonstrate the effectiveness of transcriptomic signature analysis for drug repurposing and functional genomics research. In summary, the output data generated by drugfindR allows researchers to understand how the overexpression or knockdown of a gene may affect the expression of genes within the same cellular system, identify downstream molecular consequences of gene perturbation within a system, and investigate drugs that may be repurposed for other physiological reasons.

Statement of Need

Traditional drug discovery is a resource-intensive process that often requires over a decade of research and billions of dollars to develop a single novel compound, followed by rigorous preclinical studies and multiple phases of clinical trials (Sertkaya et al., 2024). Despite these investments, many drugs fail during development due to unforeseen toxicity, lack of efficacy, or poor pharmacokinetics (Sun et al., 2022). In contrast, drug repurposing—the strategy of identifying new therapeutic uses for existing United States Food and Drug Administration (FDA)-approved drugs—offers a more time- and cost-efficient alternative. By leveraging prior safety and pharmacology data, repurposing can accelerate the path to clinical use while reducing the risk of failure (Pinzi et al., 2024). This approach has proven successful in several cases, such as using Allopurinol, originally developed for kidney stones, to treat gout. Gemcitabine, initially an antiviral compound, is now widely used in cancer treatment (Park, 2019).

The National Institute of Health (NIH) LINCS project (Keenan et al., 2018) offers a large-scale resource of transcriptomic profiles in response to various chemical and biological perturbations; however, leveraging LINCS programmatically remains difficult due to limitations in existing tools. The integrated web-based platform, iLINCS (Pilarczyk et al., 2022), provides access to the LINCS data but lacks support for batch analyzes and scriptable workflows, limiting reproducibility and throughput (Pilarczyk et al., 2022). For example, if an end-user of iLINCS

were to investigate all the signatures associated with the gene knockdown signature of the gene DISC1, they would gain access to all the signatures. Still, they would need to process each signature individually to perform any meaningful statistical analyzes.

`drugfindR` fills this gap by offering an R-based, user-friendly interface that supports parallel processing, filtering, and analysis of transcriptomic signatures. `drugfindR` enables researchers to systematically mine concordant or discordant gene expression profiles across drug, knockdown, and overexpression conditions—unlocking new and efficient opportunities for downstream analysis pipelines, drug repurposing, systems-level discovery, and hypothesis generation.

Package Design

`drugfindR` is built on top of R's S4 object-oriented system. It has five core, modular functions that encapsulate key stages of signature processing and matching, and two higher-level wrapper functions that allow streamlined analysis using sensible defaults. These functions are outlined and described in Table 1.

Function	Type of Function	Description
`getSignature()`	Core Modular function	Retrieves L1000 gene expression signatures based on entered iLINCS ID.
`prepareSignature()`	Core Modular function	Processes user-supplied DGE / transcriptomic signature and returns the corresponding L1000 signature.
`filterSignature()`	Core Modular function	Filters a given L1000 signature based on user-defined parameter thresholds.
`getConcordants()`	Core Modular function	Retrieves the concordant LINCS signatures via the iLINCS application programming interface (API) for a given L1000 signature.
`consensusConcordants()`	Core Modular function	Takes concordant data results to return a ranked list of top candidate (gene or drug) matches to the input signature.
`investigateTarget()`	Higher-level wrapper function	Uses all signatures for a given drug or gene in the LINCS database and identifies concordant LINCS signatures.
`investigateSignature()`	Higher-level wrapper function	Processes user-submitted DGE data and queries LINCS to identify relevant matches (i.e., concordant signatures).

The Journal of Open Source Software

Typical Workflow

A typical analysis begins with a user-defined or LINCS-supplied transcriptomic signature using the getSignature() function. After formatting and processing the signature with the prepareSignature() function to ensure that the data shape and column names are correct, the data is filtered either by a log2-fold change threshold or a percentile threshold (e.g., the top and bottom 5% of the signature genes), based on user preference of these parameters, using the filterSignature() function. This curated signature is then passed to the iLINCS API using the getConcordants() function along with the relevant metadata to query the LINCS database and calculate the concordance values of the curated signature with the relevant ones available in the database. Finally, the consensusConcordants() function aggregates results based on the concordance values to generate a ranked list of the highest likelihood candidate targets (i.e., genes or drugs) that exhibit similar or opposing profiles relative to the input signature.

When querying gene perturbation signatures (e.g., knockdown or overexpression), concordant gene signatures reveal functionally similar genes or shared pathway components. In contrast, discordant gene perturbations may point to compensatory mechanisms or potential therapeutic antagonists. When the output consists of chemical perturbagens (drugs), discordant signatures are of particular interest, as they may reverse disease-associated gene expression patterns, suggesting therapeutic potential and candidates for drug repurposing. Conversely, concordant drug signatures may mimic the input condition and can serve as negative controls or drugs to avoid in a disease context.

Future Directions

Planned enhancements for `drugfindR` include support for additional omics data types (e.g., proteomics), improved integration with enrichment analysis tools, and expanded compatibility with signature databases beyond LINCS as they continue to be developed. User feedback will guide iterative development, with a focus on scalability, reproducibility, and accessibility for non-programmers through future Shiny-based graphic user interface (GUI) support.

Acknowledgements

We want to acknowledge the support of Dr. Sinead O'Donovan, who was present at the conception of this project and has been a champion for `drugfindR` throughout all stages of its development. ASI would like to acknowledge his family, Saadia Farooq, Romana Imami, and Samara Imami, for their support in the development of this software and for their patience through every debugging session.

Availability

`drugfindR` is an open-source R package with the source code available on GitHub under the GNU General Public License v3 (GNU GPLv3). It can be installed directly from GitHub and from the CogDisResLab repository at r-universe.

Funding

This project was supported by National Institutes of Health grants

- NIMH R01; MH107487
- NIMH R01; MH121102

References

Creeden, J. F., Imami, A. S., Eby, H. M., Gillman, C., Becker, K. N., Reigle, J., Andari, E., Pan, Z. K., O'Donovan, S. M., McCullumsmith, R. E., & McCullumsmith, C. B. (2021). Fluoxetine as an anti-inflammatory therapy in SARS-CoV-2 infection [Journal Article]. *Biomed Pharmacother*, *138*, 111437. https://doi.org/10.1016/j.biopha.2021.111437

Imami, A. S., O'Donovan, S. M., Creeden, J. F., Wu, X., Eby, H., McCullumsmith, C. B., Uvnäs-Moberg, K., McCullumsmith, R. E., & Andari, E. (2020). Oxytocin's anti-inflammatory and proimmune functions in COVID-19: A transcriptomic signature-based approach [Journal Article]. *Physiol Genomics*, *52*(9), 401–407. https://doi.org/10.1152/physiolgenomics.00095.2020

Keenan, A. B., Jenkins, S. L., Jagodnik, K. M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A. B., Silverstein, M. C., Lachmann, A., Kuleshov, M. V., Ma'ayan, A., Stathias, V., Terryn, R., Cooper, D., Forlin, M., Koleti, A., Vidovic, D., Chung, C., … others. (2018). The library of integrated network-based cellular signatures NIH program: System-level cataloging of human cells response to perturbations [Journal Article]. *Cell Syst*, *6*(1), 13–24. https://doi.org/10.1016/j.cels.2017.11.001

O'Donovan, S. M., Imami, A., Eby, H., Henkel, N. D., Creeden, J. F., Asah, S., Zhang, X., Wu, X., Alnafisah, R., Taylor, R. T., Reigle, J., Thorman, A., Shamsaei, B., Meller, J., & McCullumsmith, R. E. (2021). Identification of candidate repurposable drugs to combat COVID-19 using a signature-based approach [Journal Article]. *Sci Rep*, *11*(1), 4495. https://doi.org/10.1038/s41598-021-84044-9

Park, K. (2019). A review of computational drug repurposing [Journal Article]. *Translational and Clinical Pharmacology*, *27*(2), 59. https://doi.org/10.12793/tcp.2019.27.2.59

Pilarczyk, M., Fazel-Najafabadi, M., Kouril, M., Shamsaei, B., Vasiliauskas, J., Niu, W., Mahi, N., Zhang, L., Clark, N. A., Ren, Y., White, S., Karim, R., Xu, H., Biesiada, J., Bennett, M. F., Davidson, S. E., Reichard, J. F., Roberts, K., Stathias, V., … Medvedovic, M. (2022). Connecting omics signatures and revealing biological mechanisms with iLINCS [Journal Article]. *Nat Commun*, *13*(1), 4678. https://doi.org/10.1038/s41467-022-32205-3

Pinzi, L., Bisi, N., & Rastelli, G. (2024). How drug repurposing can advance drug discovery: Challenges and opportunities [Journal Article]. *Frontiers in Drug Discovery, Volume 4 - 2024*. https://doi.org/10.3389/fddsv.2024.1460100

Sertkaya, A., Beleche, T., Jessup, A., & Sommers, B. D. (2024). Costs of drug development and research and development intensity in the US, 2000-2018 [Journal Article]. *JAMA Network Open*, *7*(6), e2415445–e2415445. https://doi.org/10.1001/jamanetworkopen.2024.15445

Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90 [Journal Article]. *Acta Pharm Sin B*, *12*(7), 3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002