NF Research Tools Central: A disease-specific knowledgebase of experimental tools

Robert J. Allaway PhD[1], Ashley Clayton MS[1], Mialy DeFelice PhD[1], Brynn Zalmanek MLIS[1], Jay Hodgson[1], Caroline Morin MS[2], Stockard Simon[1], James A. Eddy PhD[1], Milen Nikolov PhD[1], Christina Conrad PhD[1], Kenneth Chan PhD[3], Felicia Sabatino[3], Dzmitry Fedarovich[3], Adam Lafontaine PhD[3], Jineta Banerjee PhD[1], Kalyan Vinnakota PhD[4], Marco Marasca[1], Kevin J. Boske[1], Bruce Hoff PhD[1], Ljubomir Bradic[1], James Goss PhD[1], YooRi Kim MS[1], Julie A. Bletz PhD[1]

[1]Sage Bionetworks , Seattle, WA, USA   [2]No affiliation
[3]Rancho Biosciences, Santa Fe, CA, US   [4]Gilbert Family Foundation, Detroit, MI, USA
Correspondence: nf-osi@sagebionetworks.org

Highlights

- We have developed a centralized resource database and companion web-app (**Fig 1**, https://tools.nf.synapse.org) for neurofibromatosis (NF) related research tools, including animal models, cell lines, antibodies, genetic reagents, and biobank.
- This database provides information about tools as well as links to data that have been generated using these tools.
- By mid-2023, the database will leverage AI-assisted data curation and contain information on over 1000 NF1 and RAS-relevant research tools
- We invite the research community to contribute tools as well as feedback, usage notes, and other observations to the database.

Figure 1. The NF Research Tools Database homepage allows users to search for specific tools, or to browse by the type of tool.

Background & Introduction

- Research tools, such as model organisms, cell lines, and antibodies, are essential to designing and executing successful biological experiments.
- Our experience in the NF field has shown that researchers struggle to identify the research tools available to them, determine where tools can be acquired, and understand what tools are most well-suited for which experiments.
- A variety of databases exist to help researchers find useful research tools, these databases often:
    - are specific to one type of research tool while being disease-agnostic
    - provide only high-level information
    - do not contain information about in-development models, and
    - do not contain observational data for the research tools.
- To address this, we created the NF Research Tools Database, a user-friendly, open-access database and companion portal designed to help the neurofibromatosis type 1 (NF1) research community easily find, obtain, and use NF1-relevant research tools.
- This prototype database catalogs a wide variety of NF1-relevant research tools using databases such as Cellosaurus,[1] AntibodyRegistry,[2] RRID Portal,[3] among others, as well as information provided in literature and from the NF community.

Approach

Evaluating External Data Sources & Identifying User Stories

- We conducted a survey of NF-related tools by collaborating with domain experts as well as evaluating generalist tool databases (**Fig 2**); using this information, we identified what categories of metadata were important or feasible to collect
- Relying on our experience working with the experimental research community, we developed a series of "user stories" to help ensure that the resulting product fulfills the needs of the community.
- For example, bench scientists, clinicians, and bioinformaticians might want to "view which driver mutations a disease model has, so that they know if an animal model listed in the database has a specific genetic mutation they are interested in studying"
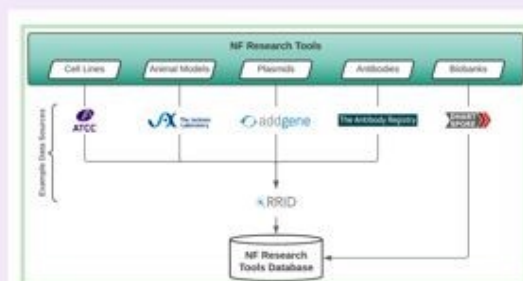

Figure 2. Flow of information on cell lines, animal models, plasmids, antibodies and biobanks into the NF Research Tools Database.

Designing the Database

- Using the external data sources as a model, we enumerated important attribute fields for each research resource and developed a relational data model (**Fig 3**) to store this information
- A machine-readable data model[4] was developed using the "schematic" python library,[5] which in turn was used to generate and populate a MySQL database using information collected from the multiple resources (**Fig 2**)
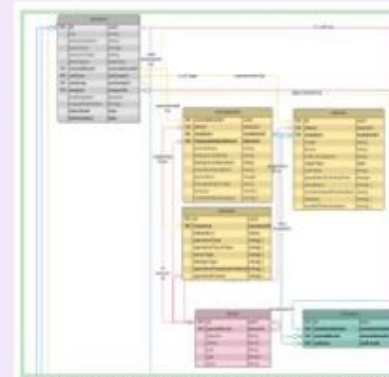

Figure 3. Example modules from the NF Research Tools Database data model.

Web-app Prototyping & Evaluation

- Using the use-cases and data model, we iteratively designed and solicited feedback on a companion web app to allow users to explore the database.
- We then hosted the database tables on the Synapse.org collaborative science platform,[6] and built the prototype application using the Synapse React Client[7] (**Fig 4**) as a new component of the NF Data Portal[8]
- Finally, we conducted an evaluative design research study on the prototype to determine if the design and features of the database website meet the goals and needs of its users, informing improvements and long-term development goals for the database.

Figure 4. Interactively exploring the NF Research Tools Database. (A) Users can filter tools using standardized metadata and free-text search terms. (B) "Tool Detail" pages allow users to perform a deep-dive into information about specific research tools such as, the information described in panels C-E. (C) A "canonical" publication, when available, is listed for each tool. (D) When available, the status of selected genes - for example, genes that have been modified with the Cre-lox system - is provided. (E) When tools are commercially available, the database provides a link to an external vendor. (F) This page also presents "Observations," which are user-submitted details that describe pathology, usage notes, issues, or other information relevant to a particular experimental tool.

New for 2023: Expanding database, easier submission, AI-powered tool descriptions

- Expanding database:
    - 🌱 By mid-2022, the database will be updated to contain over 1,000 tools and resources – a 600% increase from 2022!
- Easier Submission:
    - ✍️ We've introduced a streamlined submission process with fewer steps – Share your research tools in minutes and foster collaboration!
- AI-Powered Tool Descriptions:
    - 🎂 We are harnessing the power of AI large language models to generate more insightful and comprehensive tool descriptions - these will be rolling out soon!

Acknowledgements

References

1. Bairoch,A. (2018) The Cellosaurus, a Cell-Line Knowledge Resource. J. Biomol. Tech., 29, 25–38.
2. Bandrowski,A. et al. (2011) An antibody registry for biological sciences. In, Front. Neuroinform. Conference Abstract: 4th INCF Congress of Neuroinformatics. doi: 10.3389/conf. fninf.
3. Bandrowski,A. et al. (2015) The Resource Identification Initiative: A cultural shift in publishing. F1000Research, 4, 134.
4. www.github.com/nf-osi/nf-research-tools-schema
5. www.github.com/Sage-Bionetworks/schematic
6. Omberg, L., et al. Enabling transparent and collaborative computational analysis of 12 tumor types within The Cancer Genome Atlas. Nat Genet 45, 1121–1126 (2013). https://doi.org/10.1038/ng.2761
7. www.github.com/Sage-Bionetworks/Synapse-React-Client
8. Allaway,R.J. et al. (2019) Engaging a community to enable disease-centric data sharing with the NF Data Portal. Sci Data, 6, 319.