



St. Jude BioHackathon

Title

Simplify patient sample (meta)data tracking and querying

Category

Data Management

Challenge

Overcome the challenge of coordinating multiple different identifiers and metadata from SRM orders, Comp Bio ID requests, and identifiers assigned during Hartwell processing, and experimental metadata. Currently this information is compiled from the different sources manually and combined with necessary metadata, but this time-consuming process can result in incomplete or inaccurate records.

As such, a way to easily track sample metadata, associated identifiers, and data locations in a unified, yet flexible way to account for specific group or lab needs is highly needed. Designing a system that's compatible with existing data repositories and systems at St. Jude (e.g. TBANK, SRM) would also be a huge advantage.

Benefit

This would reduce errors and incorrect analyses due to sample swaps or sample confusion, reduce data redundancy by preventing re-sequencing of the same sample repeatedly, decrease data loss, and save an immense amount of time and money by reducing the time highly paid employees spend hunting down these connections. A proof of concept unified data storage and tracking schema applied on a lab or group-scale would provide a foundation for addressing this on a larger scale.

This would be particularly beneficial (and easy to implement) for new initiatives like PTNI that don't have enormous numbers of samples yet collected.

Helpful Tools, Packages, or Software

SQL, interactive front-ends for display/editing of tables (dash, Shiny, etc).

Test Data

Dummy data of typical identifiers and metadata (SRM #, SJID, cohort info, file locations, etc).