The more detail that can be provided, the better.

Succinct Summary: Given a database linking compounds, genes, and pathways, build a tool that will: 1) allow users to input a list of compounds and retrieve a list of pathways which are enriched in the targets (genes) of those compounds. 2) Allow user to visually explore these pathways—targets hit by compounds in the list should be color coded by activity (ie: red, increased activity; blue, decreased activity). 3) Targets which have a compound not in the list (weren’t screened or were not hits), should be annotated.

Descriptive Summary: The compound discovery team is in need of a tool that enables querying compounds using a user-provided list of compounds to identify enriched pathways and activity in target genes within those given pathways. In addition, the tool should enable querying of pathway-specific targets to enable users to identify a list of compounds that target all start, intermediate, and end-points within a given pathway and their activities on those targets. In example, querying for TLR pathway targets yields a list of compounds that target genes within that pathway + a visual overlay indicates those targets are activated or inactivated by the compound. Bonus points if you can incorporate ranking pathways/compounds based on increased or decreased activity in the pathway!

The tool should provide a visual overlay of the genes targeted by a compound after querying for a pathway and vice versa: 1) querying compound list yields list of pathways + provides an enrichment analysis of the targeted genes to determine which compounds offer the most coverage, activation, or inactivation in a given pathway. In the results of the query, enriched pathways should indicate which targets in that pathway are activated/repressed by a given compound, 2) querying a pathway results in a list of compounds + a coverage/enrichment score for each compound based on the number of genes in that pathway that are targeted and their activities, and 3) after querying, the tool identifies related compounds that target the same pathway and provides a score/readout that allows users to quickly identify if there are any additional compounds to the current query that offer additional coverage for a given pathway.

\* There are existing databases linking compounds and gene targets, and we don’t want to replicate these, as keeping them up-to-date is a full time job better left for others. Likewise, we have good tools that quantify pathway enrichment using a gene list. What we are seeking here is an exploratory data analysis tool that lets users upload their list of compounds to check for pathway enrichment, and to visually explore their hits in those pathways.

3.How would solutions to your proposed challenge benefit your lab or group and the greater St. Jude/scientific community?

The solutions to this challenge would greatly benefit researchers attempting to identify compounds that target various biomarkers/markers/pathways in any given experiment and vice versa.

4.List any specific tools, packages, or software that may be helpful in addressing this problem.

R, Python, SQL, Spotfire, etc.

5.List any datasets that may be used for testing solutions to this problem. GEO, SRA, or other public datasets are preferred.

MSIGDB pathway databases, Novartis set of compounds, St. Jude-provided set of compounds currently available in compound discovery database