Sig2Lead

User Manual

Table of Contents

Installation/Configuration:

A. Installation of R and RStudio

To download R:

http://www.r-project.org/

To download R Studio:

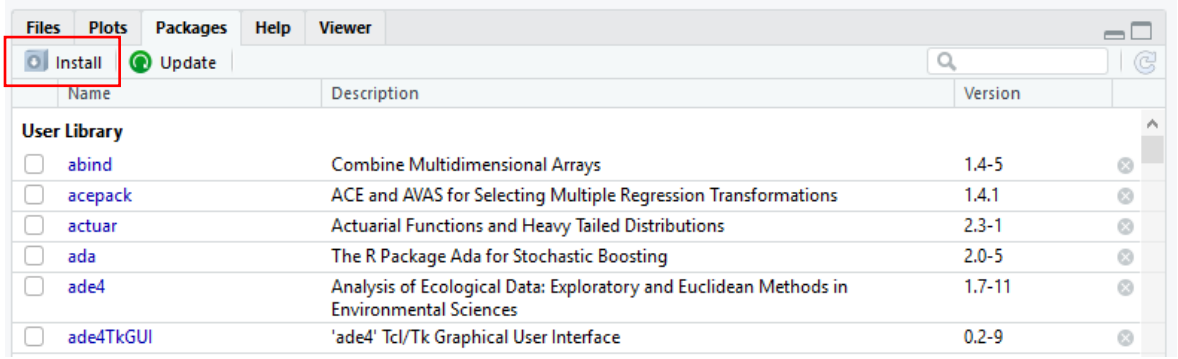
https://www.rstudio.com/products/rstudio/download/

B. Download of Sig2Lead from Github

https://github.com/sig2lead/sig2lead_dev

C. Installation of Dependencies/Libraries

Upon download, all dependencies must also be installed in order to run the program. These dependency libraries include: ChemmineR, gplots, cluster, ggplot2, shiny, httr, jsonlite, DT, ChemmineOB, plyr, dendextend, colorspace, ggforce, rlist, scatterpie, ggrepel, visNetwork, bazar, XML, RCurl, bitops igraph and plotly



Installation of Packages in R Studio. To install packages required for running Sig2Lead, click on the "Packages" tab on the right side of the interface. Under this tab, click "Install" and then type the package for installation. For installation of Bioconductor packages (ChemmineR and ChemmineOB), the biocLite command is required in the console.

D. Instructions for Installation of ChemmineOB for Windows PC

To install the Bioconductor packages (ChemmineR and ChemmineOB), type into the console:

Source("https://bioconductor.org/biocLite.R")

biocLite("ChemmineR")

biocLite("ChemmineOB")

E. Special Instructions for Installation of ChemmineOB for MAC (Brew-Based Install)

In order for Sig2Lead to run on a MAC, a brew-based install of the Open Babel library needs to be performed since there is no binary for R 3.5.0.

To perform a brew-based install of the Open Babel library, complete the following instructions after installing Sig2Lead:

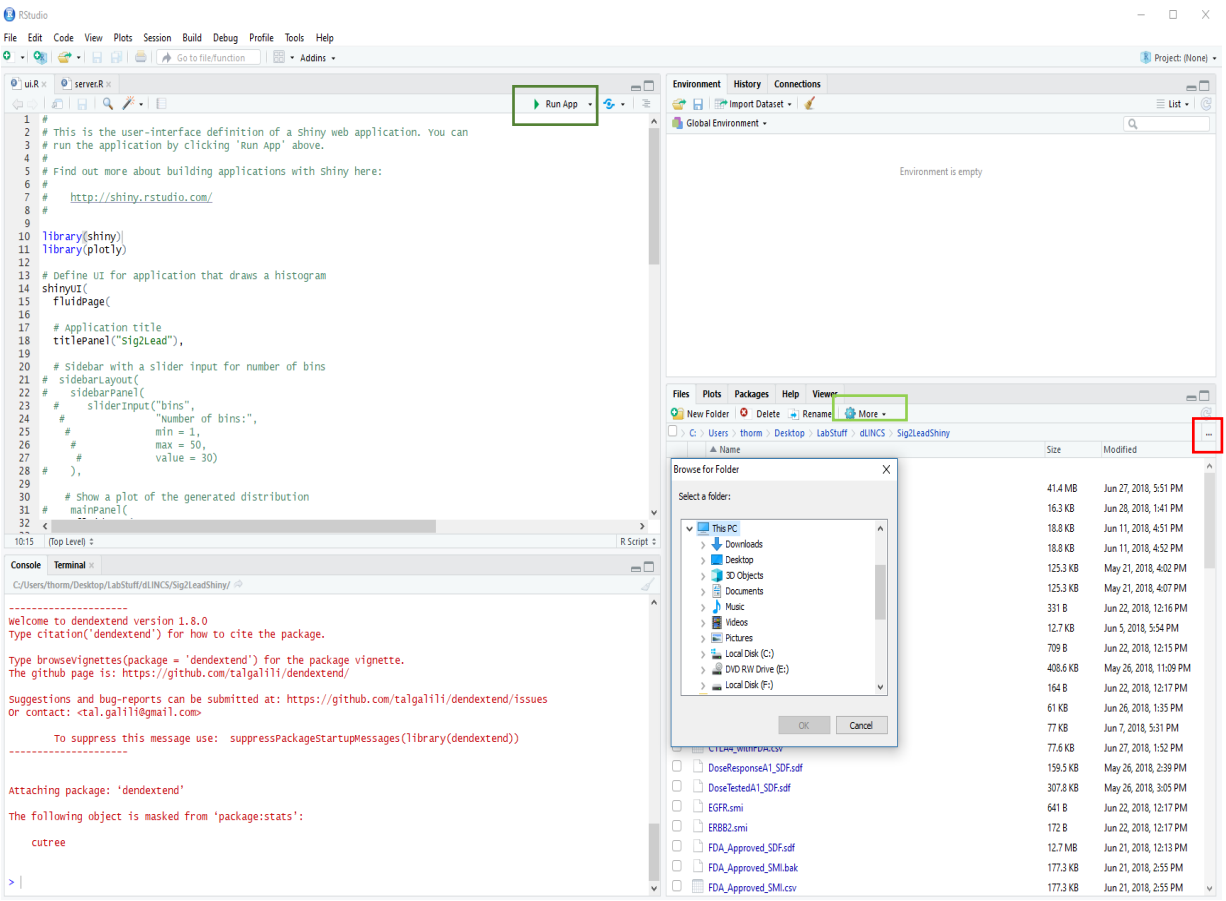
1. Install Brew from https://brew.sh
(Open a terminal window and run /usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)")

2. Install open-babel
brew open-babel

3. Reinstall chemmineOB
remove.packages("chemmineOB")
install.packages("chemmineOB"

F. Change to Working Directory

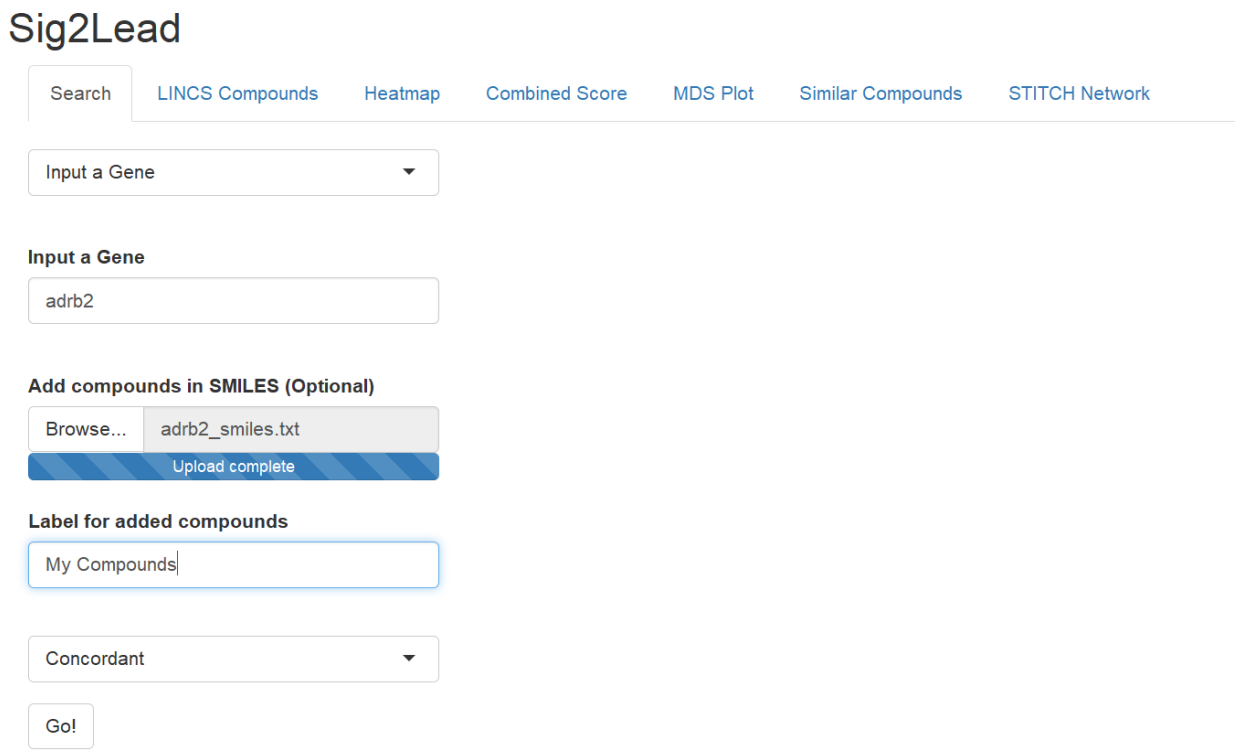
Sig2Lead is downloaded as a directory which includes two levels: Sig2LeadShiny and lib. Under Sig2LeadShiny, there should be three files: ui, server and LINCSCompounds, along with the lib and similarity search subdirectories. The lib folder includes all scripts necessary for running within the ui and server of the application itself. When downloading, make sure this structure is preserved, as this is the structure the application internally references. Once installed, set your working directory in RStudio to the Sig2LeadShiny directory.



Setting Sig2LeadShiny as the working directory in RStudio. To set a working directory, under the "Files" tab, click the "…" button highlighted in red and navigate to the Sig2LeadShiny folder. Then, click the "More" dropdown and select "Set As Working Directory." Once the working directory is set, select "Run App" highlighted in orange to open the Sig2Lead application.

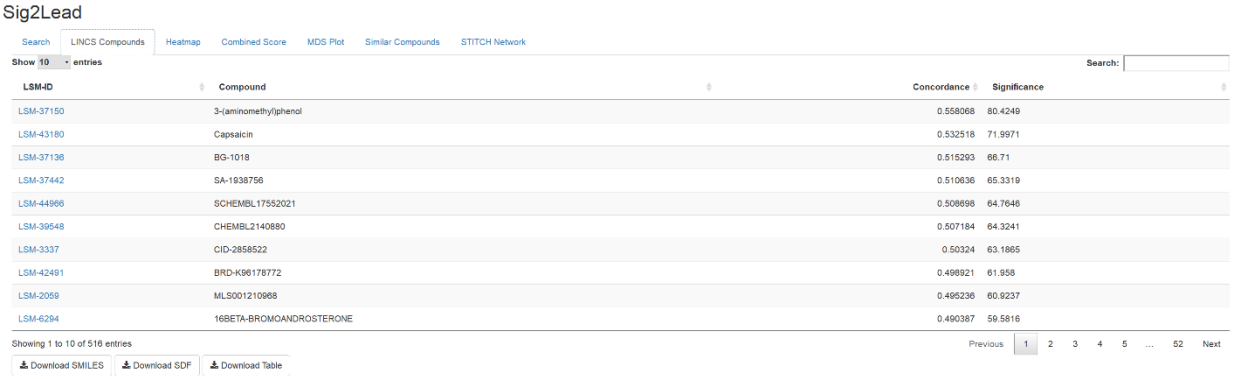
Operational Workflow:

A. Search Tab



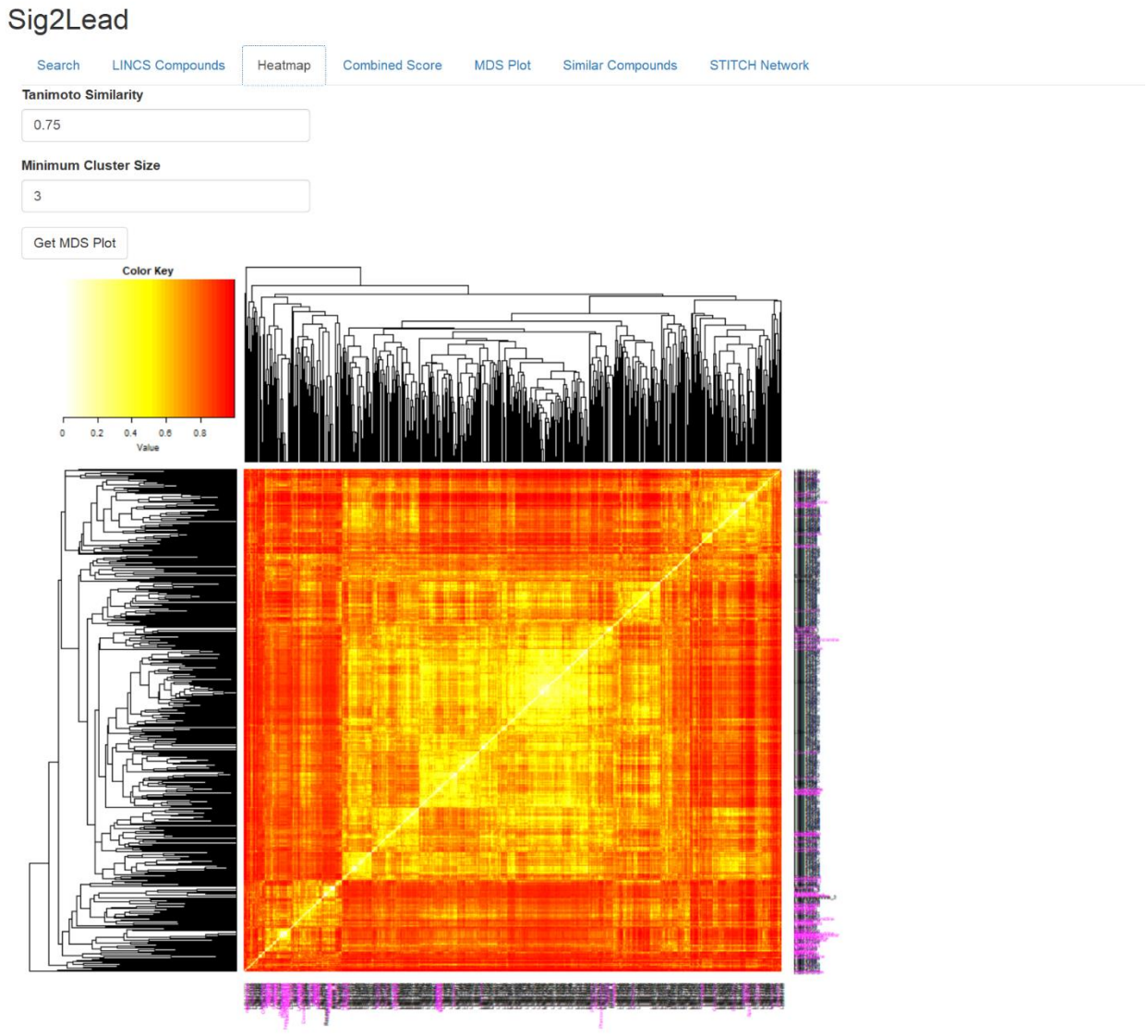
Sig2Lead landing/search page. On this page, the user inputs target genes of interest and optionally compounds in SMILES format for inclusion in clustering steps of Sig2Lead analysis. Once inputting a gene target and optional SMILES compounds, click the "Go!" button and navigate to the "LINCS Compounds" tab to await results. Alternatively, users can upload a user defined profile that has overlap with the L1000 genes by changing the "Input a Gene" dropdown to "Upload a Signature." Additionally, the LINCS data can be queried for concordant (inhibitor) or discordant (activator) profiles as compared to the user defined gene knockdown or user defined signature.

B. LINCS Compounds Tab



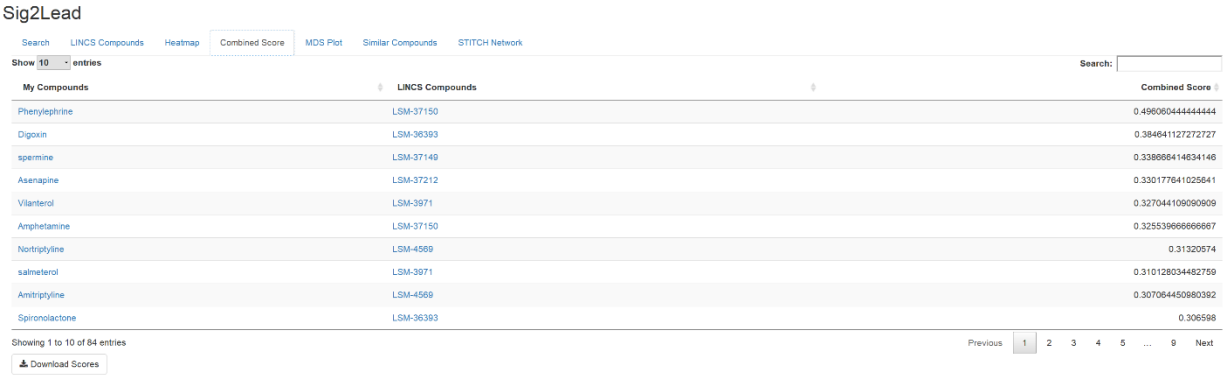
LINCS Compounds Tab for display of all LINCS compounds with concordant expression signatures to user-defined target gene. This tab displays the output of an iLINCS query, along with the SMILES code for all LINCS compounds with concordant signatures to the target gene. These compounds can be downloaded in either SMILES or SDF format, allowing the user to have structural information about all relevant compounds. Data for LSM compounds are available by clicking the hyperlinked LSM IDs, redirecting the user to the LINCS data portal entry for the compound of interest

C. Heatmap Tab



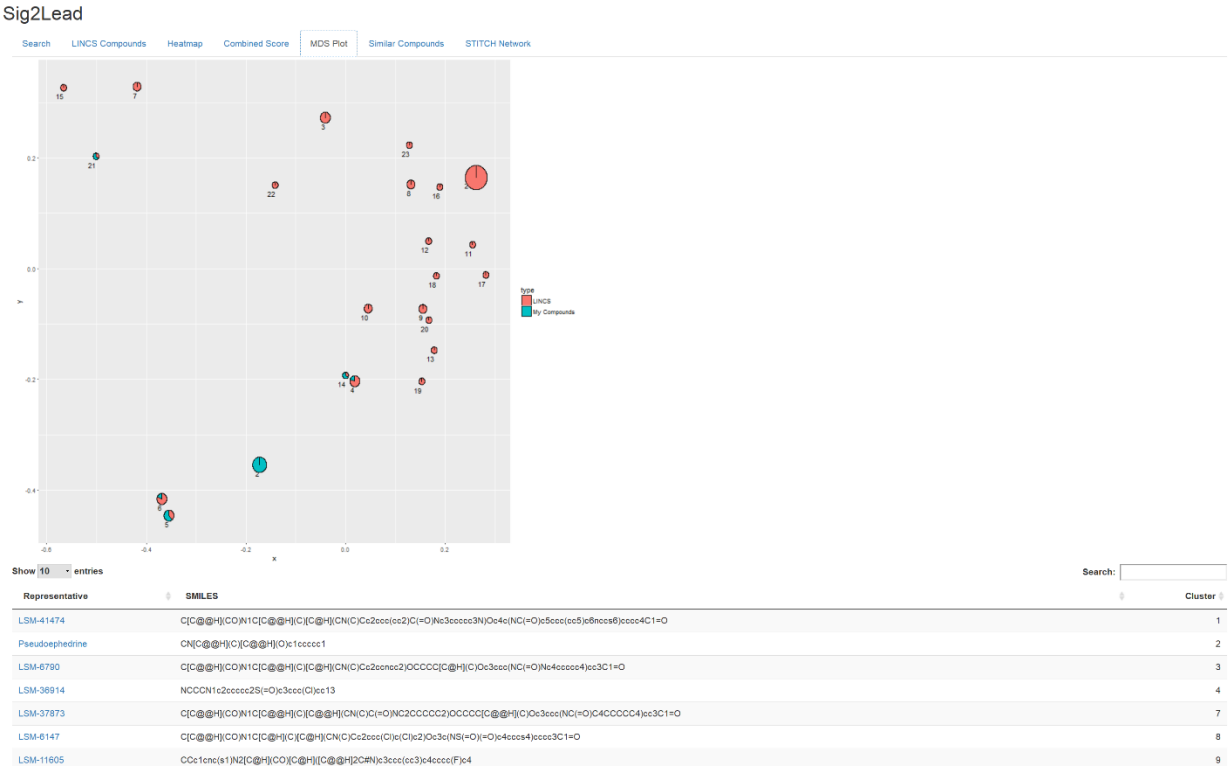
Heatmap tab of Sig2Lead for viewing chemical similarity of LINCS small molecules with concordance to a knockdown of a user-defined target gene along with any user defined small molecules (magenta). This tab allows users to visualize a heatmap of hierarchically clustered small molecules derived from LINCS and user defined compounds if applicable. From this page, representatives and MDS plot can be generated by providing a Tanimoto similarity cutoff and minimum cluster size and clicking “Get MDS Plot.”

D. Combined Score Tab



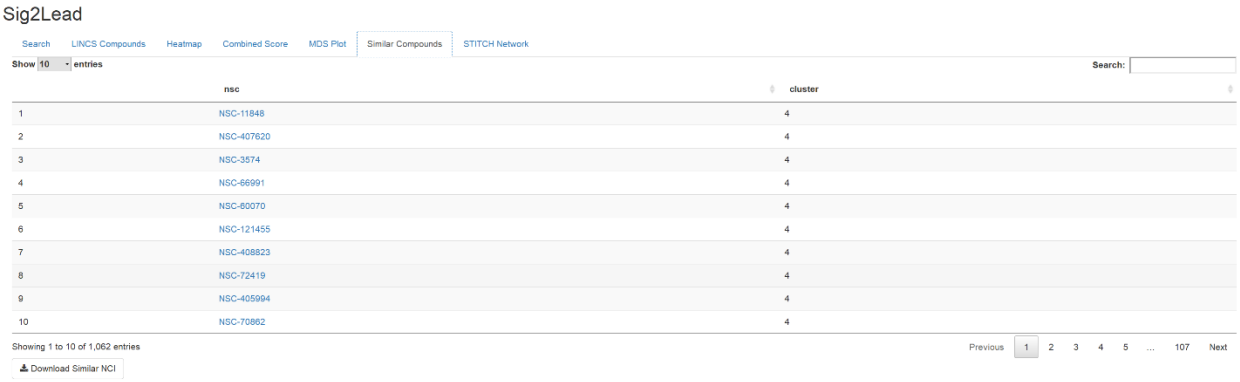
Combined Score Tab for display of the combined score of concordant LINCS compounds and the user-added compounds ("My Compounds"). The compound score is calculated by multiplying the concordance value of the each LINCS compound concordant to the gene of interest by the Tanimoto similarity of the user-added compound to the LINCS compound. This effectively takes into account both the Tanimoto similarity and concordance values. The generated table can be downloaded by clicking on the "Download Scores" button.

E. MDS Plot Tab



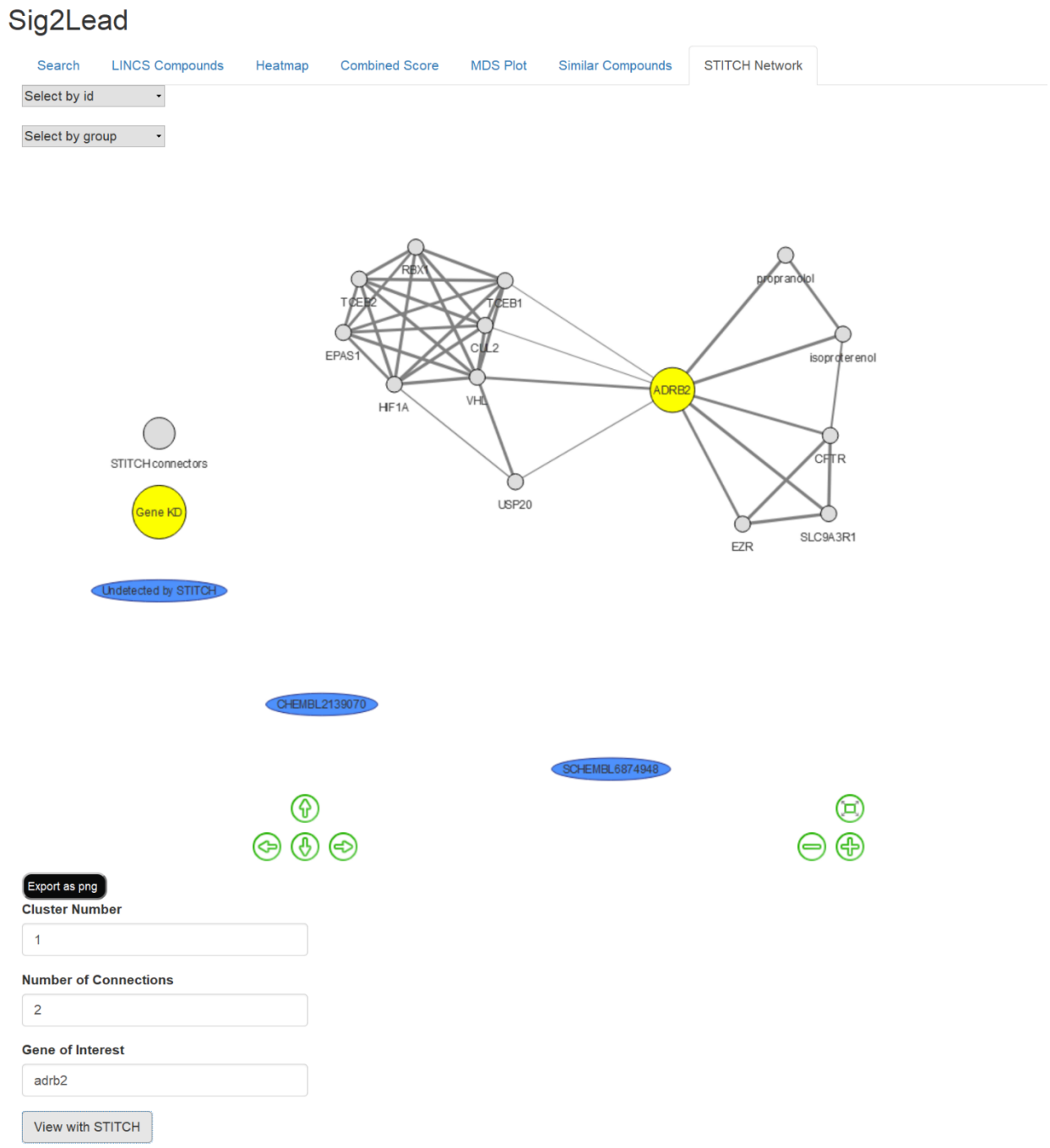
MDS Plot tab of Sig2Lead for an alternative view of hierarchically clustered small molecules derived from concordant signatures to easily view added compounds. MDS displays relative Tanimoto distance of each representative compound to one another with pie chart radius and cluster order correlated to cluster size. Red in pie charts above correspond to compounds with concordant signatures to knockdowns of the target gene in LINCS. Blue in the pie charts above are compounds added by the user, showing structural similarity to others within the same cluster. These representatives can be downloaded with the "Download Representatives" button and the whole clusters can be downloaded with the "Download Clusters" button at the bottom of the page. Related NCI compounds can be identified on the "Get Related NCI Compounds" button, which will identify compounds structurally related to each cluster present within the NCI compound library.

F. Similar Compounds Tab



The Similar Compounds tab shows NSC compounds with similar chemical structures to the centroids of each cluster. Using this feature, users are able to identify compounds within the NCI library that can be ordered and tested in vitro. This library can be ordered for only the shipping cost, allowing researchers to generate preliminary data for projects without the high costs in ordering compounds

G. STITCH Tab



STITCH Network Interface shows known interactions reported in the literature.
Using the "STITCH Network" tab, users can search their gene of interest, along with compound clusters identified via Sig2Lead to see any known interactions reported in the literature. The yellow circle signifies the query gene under the "Gene of Interest" field, and red ovals represent added compounds which are associated with selected gene or others within the same pathway.

Blue ovals represent other compounds within the specified cluster that were not known to associate with the network, but are structurally related to others that may be in the network. Gray circles are other genes identified by STITCH that fall in a pathway with the gene of interest. The thickness of the edges of the network shows confidence of a given association with thicker lines representing more confident associations. This map is interactive and users can change the cluster or gene of interest for updating with the same clusters identified within the previous analysis.