

GUITars: A GUI Tool for Analysis of High-Throughput RNA Interference Screening Data

Asli N. Goktug, Su Sien Ong, Taosheng Chen*

Department of Chemical Biology and Therapeutics, St. Jude Children's Research Hospital, Memphis, Tennessee, United States of America

Abstract

Background: High-throughput RNA interference (RNAi) screening has become a widely used approach to elucidating gene functions. However, analysis and annotation of large data sets generated from these screens has been a challenge for researchers without a programming background. Over the years, numerous data analysis methods were produced for plate quality control and hit selection and implemented by a few open-access software packages. Recently, strictly standardized mean difference (SSMD) has become a widely used method for RNAi screening analysis mainly due to its better control of false negative and false positive rates and its ability to quantify RNAi effects with a statistical basis. We have developed GUITars to enable researchers without a programming background to use SSMD as both a plate quality and a hit selection metric to analyze large data sets.

Results: The software is accompanied by an intuitive graphical user interface for easy and rapid analysis workflow. SSMD analysis methods have been provided to the users along with traditionally-used z-score, normalized percent activity, and *t*-test methods for hit selection. GUITars is capable of analyzing large-scale data sets from screens with or without replicates. The software is designed to automatically generate and save numerous graphical outputs known to be among the most informative high-throughput data visualization tools capturing plate-wise and screen-wise performances. Graphical outputs are also written in HTML format for easy access, and a comprehensive summary of screening results is written into tab-delimited output files.

Conclusion: With GUITars, we demonstrated robust SSMD-based analysis workflow on a 3840-gene small interfering RNA (siRNA) library and identified 200 siRNAs that increased and 150 siRNAs that decreased the assay activities with moderate to stronger effects. GUITars enables rapid analysis and illustration of data from large- or small-scale RNAi screens using SSMD and other traditional analysis methods. The software is freely available at <http://sourceforge.net/projects/guitars/>.

Citation: Goktug AN, Ong SS, Chen T (2012) GUITars: A GUI Tool for Analysis of High-Throughput RNA Interference Screening Data. PLoS ONE 7(11): e49386. doi:10.1371/journal.pone.0049386

Editor: Gajendra P. S. Raghava, CSIR-Institute of Microbial Technology, India

Received: June 5, 2012; **Accepted:** October 10, 2012; **Published:** November 20, 2012

Copyright: © 2012 Goktug et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the American Lebanese Syrian Associated Charities (ALSAC), St. Jude Children's Research Hospital, and National Cancer Institute grant P30CA027165. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: taosheng.chen@stjude.org.

Background

High-throughput RNA interference (RNAi) screening has gained popularity in recent years as an efficient approach to elucidating gene functions. The availability of small interfering RNA (siRNA) and short hairpin RNA (shRNA) libraries that target the entire genome, the relative ease of use, and the efficiency of gene knockdown allows this reverse genetic approach to be amenable in a high-throughput manner [1]. When coupled with the use of small molecules, RNAi screening allows for the development of powerful chemical genetics-based synthetic lethality screens. However, the analysis and subsequent interpretation of large data sets obtained from genome-wide RNAi screens remains a tedious and slow process for researchers. Therefore, researchers who do not have the required resources to develop an in-house data analysis pipeline that meets their specific needs or to obtain commercial data analysis software heavily depend on open-access software packages available to them. One potential issue with these open-access packages is that some require substantial programming experience, which prevents researchers with limited

programming skills to benefit from these resources. Additionally, currently available open-access high-throughput data analysis software uses the most common statistical methods that have been developed for the analysis of small molecule and RNAi screens, including mean difference, percent activity, z-score, and *t*-test statistics for normalization and hit selection purposes [2–9]. In recent years, a new statistical parameter, strictly standardized mean difference (SSMD), proposed by Zhang [10,11] has become a widely used criterion for both screen quality control (QC) and hit selection, because the previously mentioned methods are associated with certain statistical drawbacks for the analysis of high-throughput screening data. Unlike these measures, SSMD addresses the magnitude of the RNAi effect and is more robust to sample size, which leads to comparable values across screens [12,13]. By eliminating the effect of sample size and improving the control of false hit rates, SSMD was proved to be a more reliable parameter to be used in the analysis of high-throughput RNAi screens than the previously mentioned parameters [14–16].

To our knowledge, there has been no user-friendly open-access software package available to researchers that implements the

SSMD algorithm for hit selection in high-throughput RNAi screening analysis. As such, one major advantage of GUItars over other tools (*e.g.* cellHTS2 [17,18], RNAiR [19]) is that it provides an SSMD-based high-throughput analysis tool for researchers working on RNAi screens. The traditional parameters for hit selection, including percent activity, z-score, and *t*-test are also available in our software package for the users who wish to compare their SSMD-based results with the traditional methods. With its automatic workflow, GUItars aims to facilitate the data analysis and visualization process for high-throughput RNAi screens. Notably, its user-friendly design enables researchers with little or no programming knowledge to set up the analysis via its graphical user interface (GUI) and carry out the entire calculation and visualization process by a single button click.

Implementation

GUItars was programmed in MATLAB and made publicly available at (<http://sourceforge.net/projects/guitars/>) for researchers to analyze data from primary and confirmatory RNAi screens, performed either with or without replicates. It is a standalone executable, and no licensure is required. A signal intensity data file directory, a plate ID list (optional), and an RNAi annotation file (optional) need to be provided to run the software. The GUI for GUItars is designed for the end-user to easily enter the file destinations via popup dialog boxes. Alternatively, users can simply enter character strings containing the file paths in the corresponding fields. The GUI window enables the user to enter additional information such as screening method details (*e.g.*, single vs. replicate, 96 vs. 384-well), hit criterion, and plate configuration to carry out a smooth analysis. The user-modifiable fields on the GUI window are grouped into 3 major sections: Input Parameters, Output Parameters, and Plate Configuration (Figure 1).

Input Parameters

The “data file directory” contains the readouts from each assay plate, with each plate data point saved in a separate file (accepted file formats are tsv, tab-delimited txt for all operating systems; and additionally, csv, xls and xlsx for Windows users) (Figure 2). Instead of entering the well coordinates and the intensity values as a list, data must be provided in a matrix form in 16-by-24 or 8-by-12 well format for 384- or 96-well plate screens, respectively. The well coordinates are automatically captured, and the well IDs are assigned accordingly. Number of header lines, which should be identical for each data file, must be specified in the “# of header lines in each data file” field on the GUI window. The processed data content is summarized in the “Raw_data_compiled.tsv” file output. Since GUItars is capable of running analysis for screens with or without replicates, the user must define the screening method of interest by selecting either the “single copy” or “replicates” radio button. If there are replicate assay plates with the same RNAi content, the “replicates” option should be selected, and the number of replicates must be specified. An exception to this rule is if the user prefers to evaluate each assay plate independently, keeping in mind that a different SSMD calculation will be performed. For screens with replicates (interplate), the number of assay plate replicates corresponding to each RNAi source plate has to be equal; otherwise, each condition should be analyzed separately.

The “hit selection method” drop-down menu contains a list of available analysis methods with GUItars. For user convenience, the list contents and the associated parameter options are interactively updated and become visible for further analysis specifications. For screens without replicates, the list includes

percent activity, z-score, robust z-score, SSMD, and robust SSMD options. For screens with replicates (3 or more), the user can choose between *t*-test and SSMD. The “hit selection cut-off” field is provided for user-defined threshold selection, and the value to be entered should be a reasonable cut-off value for the analysis method that will be used. To guide the user through the SSMD cutoff selection, predefined cutoff values are provided in a drop-down menu based on criteria developed by Zhang [11]. With methods other than percent activity and *t*-test, the user-defined “hit selection cut-off” is applied to automatically identify hits that either increase or decrease assay signals, corresponding with assay output readings above the (+) cutoff and below the (–) cutoff, respectively. Another drop-down list is also provided with options to perform the analysis either on raw or log-transformed (\log_2 or \log_{10}) data.

“Hit mapping to the RNAi annotations” is an optional feature that allows the users to choose whether or not matching the assay plate (hit) wells to the RNAi annotations is desired. The following fields will be enabled or disabled based on the user’s selection: “plate ID file” and “RNAi annotation file”. The “plate ID file” contains the assay and RNAi source plate ID information in one of the accepted file formats as mentioned above (Figure 2). For accurate mapping purposes, the RNAi source plate IDs must match the plate IDs in the first column of the “RNAi annotation file”. If the hit mapping option is selected, the “plate ID file” will be used as the master guide for data import. The file names within the data file directory must contain the unique assay plate IDs defined in the plate ID file. If the plate ID field is left empty for either the assay or RNAi source plate, it will be replaced with “no plateID” notation, and the rest of the analysis and mapping will be affected. In screens with replicates, hit mapping is required, and the assay plate IDs for the replicate plates must be entered consecutively in the “plate ID file”.

The “RNAi annotation file” lists the RNAi source plate contents/annotations containing a header line in the first line (Figure 2). It is mandatory that the first and second columns of the file contain RNAi source plate IDs and the corresponding assay well IDs, respectively. The rest of the columns can be as many as desired comprising any relevant gene information (*e.g.*, gene ID, accession number).

Output Parameters

In addition to an Excel file generated with multiple tabs containing comprehensive analysis results (Windows only), output data are also written into individual tab-delimited files for user convenience (all operating systems). All the graphical outputs are stored as JPEG images and written in HTML format, which can be viewed on any web browser. The images are also saved in MATLAB figure format (.fig) which can be printed or saved in higher resolution in various formats such as tif, png, eps, and pdf, as needed. A separate GUI tool to open the MATLAB figures is included in the software package available at the project home page (<http://sourceforge.net/projects/guitars/>) for users without licensed MATLAB software.

A desired directory and a folder name (maximum 25 characters) to save all the output files can be specified in the “output directory” and “output folder name” fields, respectively. The output folder name will be further customized automatically with date and time information to avoid overwriting existing files.

Plate Configuration

For ease of use, well configurations are defined on an interactive plate map provided in the GUI window. In this section, the plate format and the well coordinates should be assigned using the

GUITars - A GUI Tool for High-throughput Analysis of RNAi Screens

INPUT PARAMETERS

Data file directory:

of header lines in each data file:

Single copy Replicates:

Hit Selection Method:

Hit Selection Cut-off:

Hit mapping to the RNAi annotation desired?

Plate ID file:

RNAi annotation file:

OUTPUT PARAMETERS

Output directory:

Output folder name:

PLATE CONFIGURATION

96 - well 384 - well

Plate Quality Metrics Calculations

Automatically Knockout Outlier Control Points Max % of points to knockout: Outlier threshold:

Positive Control #1 Positive Control #2 Positive Control #3 Negative Control Negative Reference Blank (Ignore)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
B	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
C	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
E	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
F	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
G	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
H	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
I	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
J	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
K	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
L	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
M	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
N	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
O	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
P	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Status: Ready

Figure 1. GUITars user interface. Graphical user interface of the standalone executable consists of three major sections: Input Parameters, Output Parameters, and Plate Configuration. Push buttons, pop-up menus, and checkboxes provide a user-friendly tool to easily enter the parameters and automatically carry out the entire analysis workflow with a single button click.
doi:10.1371/journal.pone.0049386.g001

corresponding radio buttons and tabs above the plate configuration panel. The screening plate format can be either in 96-well or 384-well layout, which must match the data format in the “data file”. Based on the user’s selection, the plate configuration panel is dynamically updated to the selected plate format. If the “plate quality metrics calculations” option is checked, negative controls and at least one of the positive control well positions must be specified in the plate map.

The user can specify the well positions on the plate map by checking the boxes corresponding to each well and can navigate through different well types using the colored tabs. GUITars is capable of handling three sets of positive controls in plate QC calculations. To avoid conflicts arising from selecting the same well in more than one tab, the selected wells in the active tab will automatically be disabled in the inactive tabs, and the user will be prompted with a warning message when attempting to check a disabled well position, except in certain circumstances (*e.g.*, a negative control well can also be defined as a negative reference

well). For ease of use, the “Check All” and “Uncheck All” buttons can be used to perform these actions simultaneously on all wells within the active tab.

As a side note, for a screen in which a high hit rate is not necessarily expected (*i.e.*, a screen not using a focused library), all wells containing the RNAi samples should be designated as negative references. On the other hand, negative controls can be used as negative reference if a focused library or a confirmatory screen is being used as described by Zhang [13]. In GUITars, negative reference wells are used as a primary data source in scoring calculations.

Wells that do not contain any controls or RNAi samples should be designated as blank, so they will then be excluded from the calculations for all plates. If only a subset of the plates contain blank wells, or if certain outliers are desired to be manually excluded from the analysis at plate level, then the data points corresponding to those wells should be replaced with the “NaN” notation in the individual data files. GUITars also includes an

7918880	289040	483480	316920	554920	405120	360640	3858520	339080	505520	411680	298320	357280	283240	515520	397480	530520	390360	393080	356720	364560	334640	8172600	123240		
6562640	277960	376280	263880	489680	467680	492960	1004240	792720	688960	1031280	478920	432120	387240	545520	489840	484640	472960	436280	530000	569040	7245360	249760			
7735040	369280	852440	652200	751840	489280	537560	510520	416400	715680	404440	716480	421960	454720	482840	1048120	616200	483400	524560	344360	596360	559960	5168760	131160		
7484600	227200	763080	415920	523680	538280	758960																			
9049400	8655840	795040	365880	539120	716120	585760	siRNA Plate ID	Assay Well ID	Pool Catalog ID	Sequence	Gene ID	Duplex Catalog ID	Gene Accession	GINUMBER	Gene Symbol							511760	669840	7334240	287920
8241840	8850640	420880	489200	292440	399320	535440	SR384MG01P01	A03	M-053401-00	UGUAUAAAG	207157	D-053401-01-D-0534	NM_177579:NM_31341435:31	0610005104								474360	534080	9008560	9051040
8799600	8200920	619480	500480	542800	527560	388040	SR384MG01P01	A04	M-060577-00	GGAAGCUAUL	71664	D-060577-01-D-0605	NM_027853:NM_13382631:31	0610006F0R1K								397600	895000	1.2E+07	1.2E+07
8924280	9639400	722200	836800	632680	692360	467240	SR384MG01P01	A05	M-061790-00	GAGAAGACGL	71670	D-061790-01-D-0617	NM_027857:NM_13382631:31	0610006H0R1K								650760	434200	1.1E+07	9806600
5598360	523920	541080	488840	777040	455320	574320	SR384MG01P01	A06	M-050384-00	GAGGUGAAGL	66836	D-050384-01-D-0503	NM_025791:NM_13385261:31	0610006I0R1K								468000	833880	1.1E+07	1.2E+07
6416480	615520	614400	819240	403120	496040	491640	SR384MG01P01	A07	M-045577-00	CAGAGAAGGL	68323	D-045577-01-D-0455	NM_026675:NM_21311968:21	0610006K0R1K								615880	571480	7386160	426280
3710480	947280	783960	520480	454960	423720	529600	SR384MG01P01	A08	M-051255-00	CUACUCUGU	76252	D-051255-01-D-0512	NM_133764:NM_19526977:19	0610006L0R1K								787400	706200	5469480	447640
6482240	460640	417320	430520	490680	494240	451640	SR384MG01P01	A09	M-045538-00	GAACUCUUAU	68212	D-045538-01-D-0455	NM_026617:NM_21311864:21	0610007H0R1K								707920	503400	4771240	375760
8106840	8356360	483560	483320	455640	545320	490000	SR384MG01P01	A10	M-049622-00	GCAUCUCGGU	67669	D-049622-01-D-0496	NM_026304:NM_13399317:13	0610007P0R1K								576960	476040	2749040	416840
7085440	8024200	698840	506440	292680	678680	936800	SR384MG01P01	A11	M-049172-00	ACAUAUGACL	0	D-049172-01-D-0491	NM_023507:NM_31340591:31	0610007P0R1K								570600	392600	9214160	9786920
1E+07	8757520	692080	359080	441320	472520	629240	SR384MG01P01	A12	M-048216-00	UUAAGAAGCU	58520	D-048216-01-D-0482	NM_021446:NM_10946821:10	0610007P14R1K								755600	638640	1.2E+07	1.3E+07
8786080	8327080	482120	456600	309240	648560	479640	SR384MG01P01	A13	M-059761-00	GGAAGAAGGL	68497	D-059761-01-D-0597	NM_178065:NM_40538837:40	1110018G07R1K								574480	353200	1.1E+07	7305720
							SR384MG01P01	A14	M-059681-00	GAAAGAGGA	73825	D-059681-01-D-0596	NM_026617:NM_21311968:21	0610007H0R1K								488080	1.3E+07	1.1E+07	
							SR384MG01P01	A15	M-050919-00	GGAUAUGGU	66129	D-050919-01-D-0509	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	A16	M-050162-00	GCAAGGACAL	67006	D-050162-01-D-0501	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	A17	M-055083-00	GCACGGACUU	224250	D-055083-01-D-0550	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	A18	M-046748-00	GCAGAGAULC	68530	D-046748-01-D-0467	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	A19	M-046907-00	CAAGUUCGGA	68523	D-046907-01-D-0469	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	A20	M-050447-00	CCAGAGAGCC	70397	D-050447-01-D-0504	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	A21	M-053420-00	GUUGAUGAGL	71795	D-053420-01-D-0534	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	A22	M-050575-00	GGAAGAACCC	68625	D-050575-01-D-0505	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	B03	M-045731-00	GGAAGGGAL	68327	D-045731-01-D-0457	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	B04	M-059602-00	AUUCAUAGL	68318	D-059602-01-D-0596	NM_026617:NM_21311968:21	0610007H0R1K											
							SR384MG01P01	B05	M-041048-00	GAAUAUAGL	68116	D-041048-01-D-0410	NM_026617:NM_21311968:21	0610007H0R1K											

Figure 2. User input files required by GUItars. Three separate input files are required by GUItars: A “data file directory” containing individual files for each plate, an “annotation file” with first two columns containing RNAi source plate ID and assay plate well ID with a single header line, and a “plate ID file” with a single header line. An “annotation file” and a “plate ID” file are mandatory only if the “hit mapping to the RNAi annotation file” option is checked.
doi:10.1371/journal.pone.0049386.g002

algorithm to automatically knock out the outlying data points from the control wells based on user demand. For that, the user should check the “automatically knockout outlier control points” option and fill out the corresponding enabled fields.

The current analysis session can be saved and retrieved via the “Save Protocol” and “Open Existing Protocol” buttons, allowing the users to reuse or share their analysis setup with others. Once all the required fields are completed in the GUI window, analysis can be started by clicking the “Start Analysis” button, and the status of the process can be monitored via the status bar. When the analysis is started, preprocessing of the user input files is performed, and the data sets are checked for completeness. If any unexpected data formatting or analysis issues are encountered, the user is informed by a pop-up warning message, and the program is aborted. GUItars is developed with a robust error capturing mechanism against operator errors with 20 various warning messages to pinpoint the problem and provide an easy-to-use tool.

Results and Discussion

Data analysis is performed and visualized according to the workflow presented in Figure 3. Data sets are processed in three major steps, as follows: Plate QC, scoring and hit selection, and hit annotation. We demonstrated the analysis process and graphical outputs using a 12-plate siRNA (3840-gene pooled mouse siRNA library) screen without replicates in a 384-well format with a luminescence-based assay readout with an emphasis in the importance of distinct visualization approaches, which are chosen to be implemented in GUItars as default. In the demonstrated data set, 320 sample wells were used as negative reference in each plate. For plate QC calculations, 16 different wells were defined as negative control, positive #1 and #2, and 8 wells were defined as positive control #3.

Plate QC

As a common practice in RNAi screens, logarithmic transformation of the raw data is often performed to achieve a symmetrically distributed data set by normalizing highly skewed distributions [16,20]. If the user chooses to log-transform the raw

data, data distribution histograms of each plate are generated for raw data as well as the transformed data, as seen in Figure 4A. With the aid of histograms, one can decide whether the applied transformation method has met the data distribution expectations and whether the hits identified from any of the plates should be evaluated with skepticism.

Plate heat maps are provided to help with the inspection of systematic errors, which may commonly include patterns introduced by liquid handling instruments or edge effects associated with incubation quality (Figure 4B). However, in assays with a high signal window, one can misleadingly conclude an inexistence of

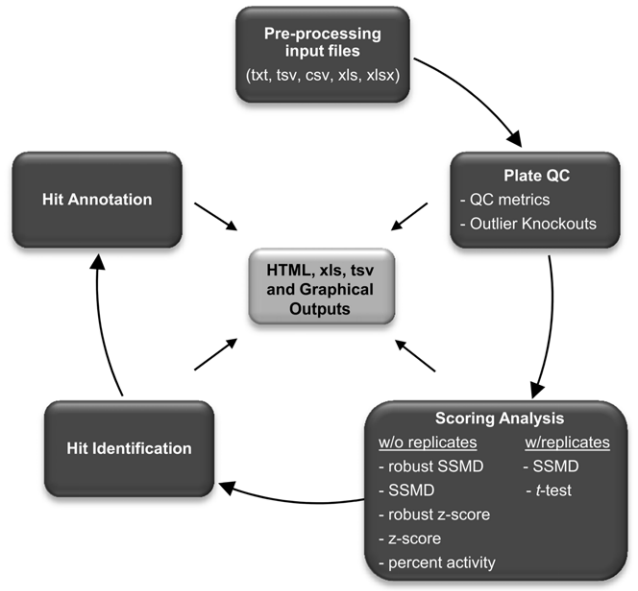


Figure 3. General workflow of high-throughput data analysis with GUItars.
doi:10.1371/journal.pone.0049386.g003

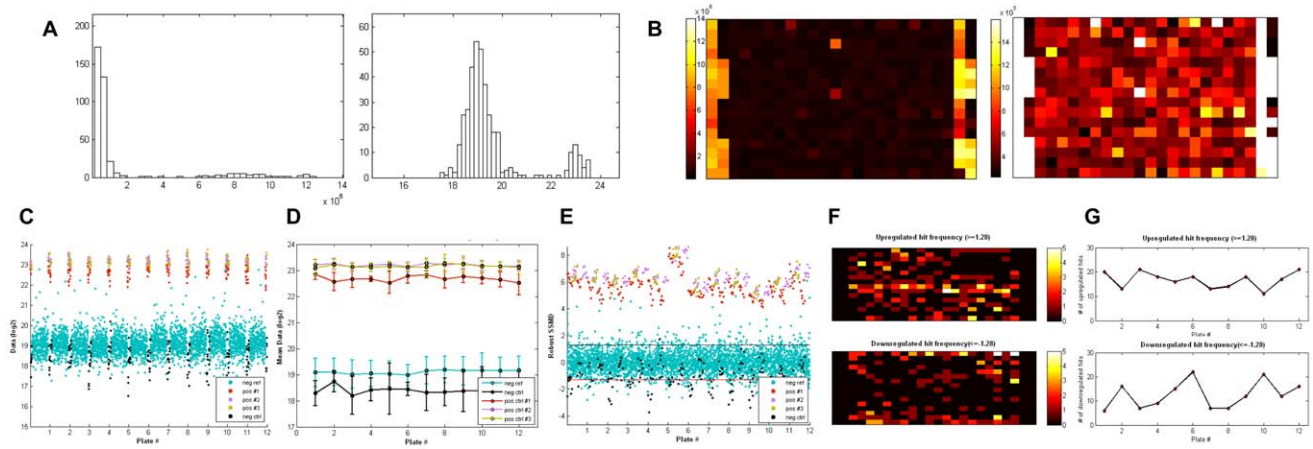


Figure 4. Graphical outputs demonstrated on a 12-plate siRNA screen analyzed with the robust SSMD method with GUItars. (A) Raw data (left) and \log_2 -transformed data (right) histograms of each plate showing the original data distribution and effect of data transformation (one representative plate is shown). (B) Original scale (left) and rescaled (right) heat maps of each plate helping to capture systematic errors (one representative plate is shown). (C) Column-wise plate-series plot. (D) Screen-wise line plot for average control readings showing a clear separation between negative control and positive controls that is consistent throughout the screen. (E) Screen-wise SSMD score scatter plots with cutoff lines at 1.28 and -1.28 for signal-increasing and signal-decreasing hits, respectively. (F) Hit distribution heat maps for signal-increasing (top) and signal-decreasing (bottom) hits. (G) Screen-wise hit counts for signal-increasing (top) and signal-decreasing (bottom) hits. doi:10.1371/journal.pone.0049386.g004

systematic errors because the color scaling is biased. Therefore, a second set of heat maps is provided with a color scaling ranging between $\text{minimum}(\text{mean of negative reference}) - 2 * \text{minimum}(\text{standard deviation of negative reference})$ and $\text{maximum}(\text{mean of negative reference}) + 2 * \text{maximum}(\text{SD of negative reference})$. As seen in Figure 4B, it is easier to observe systematic patterns in the heat map on the right than in the one on the left, although there are no evident errors observed in the screen shown.

While a heat map is a good tool for visually identifying systematic errors at the individual plate level, a plate-series plot is also an expedient way to review the overall performance of the screen at a glance. With GUItars, plate-series plots are generated in row-wise and column-wise formats as so that the row and column effects can be captured easily. A representative column-wise plot is shown in Figure 4C, and it is confirmed that there are no major systematic errors.

With GUItars, calculations of the comprehensive plate QC parameters are performed based on the user’s positive control, negative control, and negative reference well designations (after data transformation, if selected) when the “plate quality metrics” option is checked. Although RNAi screens share certain similarities with small molecule screens, there are many aspects in which they differ. Notably, the Z' -factor [21], a measure of the quality of a screen, is considered acceptable for small molecule screens when it is greater than 0.5. However, for RNAi screens, Z' is usually less than 0.5. As such, using the Z' -factor, RNAi screens are usually less robust than small molecule screens. As proposed by Zhang [10], SSMD can be a better alternative plate QC metric than the Z' -factor. Zhang has shown that, as a quality measurement in high-throughput assays, SSMD is based on a firm statistical theory and clear probabilistic meaning, as opposed to the Z' -factor, which lacks a statistical basis and entails a relatively arbitrarily chosen pass/fail cutoff. Moreover, Zhang has illustrated that quality assessment by the Z' -factor may be misleading for screens with moderate or strong positive controls, while the SSMD-based QC metric can quantify plate quality for screens with positive controls with various activities [22]. Although the Z' -factor is applicable to assays with single readout, another version of the formula which can incorporate multiple readouts via linear

projection of the data is developed by Kümmel et al. [23] for applications such as high content screening. However, it was not integrated into GUItars’ plate quality calculations since multivariate applications are beyond the scope of this study. In GUItars, SSMD (method-of-moment estimate), as a primary plate QC metric, is calculated in addition to the Z' -factor and signal window for each positive control independently in a plate-wise manner. A verbose quality metrics summary is written to an output file, which also includes mean, SD, and coefficient of variation values of positive control, negative control, and negative reference wells (Figure 5). SSMD, Z' -factor, and mean values of controls (Figure 4D) are also graphed on separate line plots for easy plate-to-plate comparison.

The outlier knockout is then performed if it is selected by the user in the analysis setup. The knockout algorithm is designed to disregard the wells whose values are greater or less than the mean $\pm x$ SD of all the wells in the particular control set, starting from the most outlier well, where x is the user-defined outlier threshold. The knockout process ends once the maximum number of points to knock out is reached or when there are no points left satisfying the above criteria, whichever occurs first. In the screen shown, we chose to knock out 30% of the points with values more extreme than ± 2 SD. All the plate quality metrics and visualizations are regenerated after the control outliers are knocked out so that the user can see the comparable results of the process.

Scoring and Hit Selection

GUItars is primarily designed to enhance and facilitate the application of the SSMD-based scoring and hit selection method on high-throughput RNAi screens. In addition to SSMD and robust SSMD, GUItars can apply other commonly used methods, including normalized percent activity, z-score, and robust z-score for screens without replicates and the paired t -test for screens with replicates, as summarized by Zhang [24]. Hence, the user has the flexibility to choose among various scoring methods from the user interface. The other methods mentioned above, which are incorporated into the currently available open-access software packages, however, are associated with certain statistical draw-

w/Positive Control #1																
Assay Plate	Z-prime w/Neg Ctrl	Signal Window w/Neg Ctrl	Plate SSMD w/Neg Ctrl	Z-prime w/Neg Ref	Signal Window w/Neg Ref	Plate SSMD w/Neg Ref	Mean Pos Ctrl	Mean Neg Ref	Mean Pos Ctrl	Stdev Pos Ctrl	Stdev Neg Ref	CV Pos Ctrl	CV Neg Ctrl	CV Neg Ref		
777000288164	0.235	1.320	5.153	0.282	1.891	5.738	22.747	18.446	19.082	0.330	0.767	0.547	1.450	4.157	2.868	
777000288157	0.092	0.421	4.275	0.250	1.658	5.540	22.566	18.499	19.099	0.344	0.887	0.523	1.525	4.194	2.737	
777000288133	0.238	1.430	5.318	0.224	1.481	5.390	22.607	18.204	19.016	0.386	0.733	0.543	1.707	4.024	2.855	
777000288126	0.338	2.229	6.001								0.672	0.493	1.343	3.684	2.587	
777000288119	-0.037	-0.152	3.805								0.982	0.495	1.884	5.320	2.604	
777000288102	0.421	3.242	6.926								0.590	0.405	1.262	3.231	2.134	
777000288287	0.273	1.551	5.316	777000288164	SR384MG01P01	A01	6.5729				0.798	0.519	1.323	4.385	2.711	
777000288270	0.256	1.494	5.327	777000288164	SR384MG01P01	B01	6.3446				0.744	0.509	1.467	4.064	2.654	
777000288263	0.134	0.705	4.665	777000288164	SR384MG01P01	C01	6.3316				0.814	0.550	1.839	4.425	2.871	
777000288225	0.035	0.139	3.804	777000288164	SR384MG01P01	D01	5.9574				1.048	0.672	1.228	5.658	3.506	
777000288249	0.425	3.343	7.013	777000288164	SR384MG01P01	E01	6.4350							3.108	2.626	
777000288232	0.103	0.566	4.622	777000288164	SR384MG01P01	F01	6.4350							4.039	2.788	
w/Positive Control #1 After Knockout																
Assay Plate	Z-prime w/Neg Ctrl	Signal Window w/Neg Ctrl	Plate SSMD w/Neg Ctrl	Assay Plate	siRNA Plate	Assay Well	Robust SSMD	Group								
777000288164	0.554	5.251	8.773	777000288164	SR384MG01P01	D04	1.9743	upregulated (>= 1.28)								
777000288157	0.408	3.815	7.144	777000288164	SR384MG01P01	O04	-1.7479	downregulated (<= -1.28)								
777000288133	0.285	1.733	5.548	777000288164	SR384MG01P01	M05	1.3332	upregulated (>= 1.28)								
777000288126	0.618	5.967	9.412	777000288164	SR384MG01P01	L07	-1.9560	downregulated (<= -1.28)								
				777000288164	SR384MG01P01	A08	3.7129	upregulated (>= 1.28)								
				777000288164	SR384MG01P01	B08	1.4950	upregulated (>= 1.28)								
				777000288164	SR384MG01P01	J09	1.8180	upregulated (>= 1.28)								
Threshold: 1.28																
siRNA Plate ID	Assay Well ID	Pool Catalog Number	Sequence	Gene ID	Duplex Catalog Number	Gene Accession	GINUMBER	Gene Symbol	Robust SSMD Value	Assay Plate ID	Category					
SR384MG01P01	A08	M-051255-00	CUACUCUGU	76252	D-051255-01:D-051255-02	NM_133764:NM_119526977:195	0610006014RIK		3.7129	777000288164	upregulated (>= 1.28)					
SR384MG01P01	A10	M-049622-00	GCAUCUCGG	67669	D-049622-01:D-049622-02	NM_026304:NM_C_133999317:133	0610007P06RIK		-1.4325	777000288164	downregulated (<= -1.28)					
SR384MG01P01	A15	M-050919-00	GGAUUAAGG	66129	D-050919-01:D-050919-02	NM_025370:NM_C_13384743:133	1110018J18RIK		-1.9584	777000288164	downregulated (<= -1.28)					
SR384MG01P01	A16	M-050162-00	GCAAGAGCAL	67606	D-050162-01:D-050162-02	NM_026271:NM_C_13385775:133	1110018M03RIK		1.3205	777000288164	upregulated (>= 1.28)					
SR384MG01P01	A17	M-055083-00	GCAACGACUL	224250	D-055083-01:D-055083-02	NM_171826:NM_C_25188196:251	1110019C08RIK		2.1353	777000288164	upregulated (>= 1.28)					
SR384MG01P01	A19	M-046907-00	CAAGUUCGG	68523	D-046907-01:D-046907-02	NM_026753:NM_C_21312045:213	1110019N10RIK		2.8464	777000288164	upregulated (>= 1.28)					
SR384MG01P01	B08	M-045084-00	AGAAGAAGG	70337	D-045084-01:D-045084-02	NM_027391:NM_C_21312561:213	0610009A07RIK		1.4950	777000288164	upregulated (>= 1.28)					
SR384MG01P01	B14	M-047495-00	GAAAGGAA	73738	D-047495-01:D-047495-02	NM_028633:NM_C_47458821:474	1110020L19RIK		-1.3309	777000288164	downregulated (<= -1.28)					
SR384MG01P01	C12	M-063422-00	GCAAGGAUCU	66847	D-063422-01:D-063422-02	NM_025798:NM_C_31981180:319	0610010I17RIK		5.9546	777000288164	upregulated (>= 1.28)					
SR384MG01P01	D04	M-042211-00	GCUAUACUG	78885	D-042211-01:D-042211-02	NM_030205:NM_C_31560027:315	0610011B16RIK		1.9743	777000288164	upregulated (>= 1.28)					
SR384MG01P01	H12	M-056087-00	GAACUACGG	67446	D-056087-01:D-056087-02	NM_175118:NM_C_31341168:313	0710001B24RIK		4.1712	777000288164	upregulated (>= 1.28)					
SR384MG01P01	I18	M-059352-00	GAACACGAA	71750	D-059352-01:D-059352-02	NM_027900:NM_C_31541816:315	1300003K24RIK		1.3009	777000288164	upregulated (>= 1.28)					
SR384MG01P01	J09	M-040446-00	GCAACUCGG	66117	D-040446-01:D-040446-02	NM_025363:NM_C_27228964:272	1110001J03RIK		1.8180	777000288164	upregulated (>= 1.28)					
SR384MG01P01	J18	M-062744-00	GACACAGCAU	74766	D-062744-01:D-062744-02	NM_138303:NM_C_119923858:199	1300010K09RIK		2.0040	777000288164	upregulated (>= 1.28)					
SR384MG01P01	J19	M-046796-00	GGAUUUGGC	67998	D-046796-01:D-046796-02	NM_026501:NM_C_21312665:213	1300010M03RIK		1.4950	777000288164	upregulated (>= 1.28)					
SR384MG01P01	J21	M-045944-00	GCAACCGCA	67480	D-045944-01:D-045944-02	NM_026186:NM_C_27229044:272	1300013D05RIK		2.1734	777000288164	upregulated (>= 1.28)					

Figure 5. Excel readable output file. Individual tab-delimited output files as well as a comprehensive Excel file are generated with the following information: Plate QC calculations before and after control outlier knockout, scores for all wells classified by well type, scores for hit wells classified by hit type (i.e., signal-increasing or signal-decreasing), and annotated hit list (optional) with corresponding scores. doi:10.1371/journal.pone.0049386.g005

backs for the analysis of high-throughput screening data. In a primary screen without replicates, the mean difference and percent activity measures fail to incorporate data variability in the hit selection [2]. On the other hand, the regular z-score and SSMD methods assume that the variability of a sample RNAi is same as the variability in a negative reference group [25]. It is known that the robust versions of the z-score and SSMD are less sensitive to outliers due to the replacement of mean and SD with median and median absolute deviation (MAD) [9,26,27]. Therefore, the use of the robust versions of these methods is recommended for high-throughput screen analysis. Although z-score and SSMD are linearly related parameters for screens without replicates, the primary difference arises when the SSMD is applied for screens with replicates, in which case z-score is not a valid criterion to score RNAi samples with various variability [13]. On the other hand, in confirmatory screens with replicates, p-values associated with a z-score or t-test determines whether the null hypothesis, which is the mean of an RNAi sample being equal to the mean of the negative reference group, is accepted or not. Also, sample size has major influence on the resulting p-values due to the formulation of these tests. In contrast, SSMD across replicates provides a direct means of measuring the RNAi effect compared with the negative reference wells without the effect of sample size. In GUItars, we use the uniformly minimal variance unbiased estimate (UMVUE) of SSMD in the sample scoring [11].

In the siRNA screen shown, we applied robust SSMD hit selection at a cutoff of 1.28 on log₂-transformed data. As a result of the analysis, associated scores for all wells in every plate were written into an output file with the corresponding assay plate ID, RNAi plate ID, and assay plate well ID, which is categorized based on the well types (i.e., positive control, negative control, and

negative reference) (Figure 5). To visualize the scores and the hit distributions in each plate, individual scatter plots are generated, in which horizontal lines are drawn corresponding to the cutoff used to select hits that either increase or decrease the assay activities. A series plot showing the well scores of all the screened plates is also necessary to explore the plate-to-plate variance within the screen. As seen in a representative graph in Figure 4E, the SSMD values and hit distributions are consistent among all the plates, except in plate 6, which has smaller plate median and MAD values, yielding higher SSMD scores.

In GUItars, the positive control and the negative reference wells are used as the high and low signals in the normalized percent activity calculations whereas only the negative reference wells are used in all the other methods. When the percent activity method is applied, the user has to define which control corresponds to the high signal and whether the hits are selected above or below the selected cutoff. If the t-test method is chosen, the “hit selection cutoff” is treated as the maximum p-value below which the hits are considered significant. Unlike SSMD, by the nature of the t-test statistics, the program outputs cannot classify whether an RNAi hit increases or decreases the assay signal.

The scoring calculations are followed by hit selection based on the cutoff defined by the user excluding the user-defined control and blank wells. The hit well IDs and the hit categories (i.e., increased or decreased signal) are determined above and below the positive and negative cutoff values (in percent activity and t-test, only the positive cutoff is considered), and written into an output file (Figure 5). A hit distribution heat map is utilized as a means to detect the effects of systematic errors in the final hit distribution. It is expected that the hits are distributed randomly among the wells in all plates; therefore, if a particular well is a “hotspot” for hits in

most of the plates, that may be an indication of systematic errors. To provide a visual summary of the hit distribution, the frequency of a well being identified as a hit in all plates within a screen is calculated for each well and presented in separate heat maps for signal-increasing and signal-decreasing hits (Figure 4F). Additionally, the number of hits identified from each plate should also be consistent within a screen unless certain plates are specifically designed to lead to higher hit rates. Therefore, a screen-wise line plot of the number of signal-increasing and signal-decreasing hits is selected as another graphical output mode, allowing the user to inspect for the overall hit selection performance (Figure 4G). Thus, a hit distribution map along with a hit counts line plot provides crucial information for determining the reliability of the screen results.

Hit Annotations

If the “hit mapping to the RNAi annotation” option is selected, a hit mapping algorithm is used to match the hit well IDs of the assay plates with the corresponding RNAi plate annotations according to the user input “plate ID” and “RNAi annotation” files. For screens with replicates, one should use extra care to input the same RNAi source plate ID for all corresponding copies of the assay plates. An output file containing the assay plate ID, score, and the category of the hit wells is generated with the matched RNAi source annotations (Figure 5).

Comparisons

The two leading open access high-throughput analysis software packages for RNAi screens are cellHTS2 [17,18] and RNAither [19] in the BioConductor package, and both are developed in the R programming language. Compared with both cellHTS2 and RNAither, the most important advantage of GUItars is the implementation of SSMD as both QC and hit selection metrics. While both of these tools offer high flexibility and powerful features for data normalization and annotation, their hit identification algorithms rely on frequently used techniques such as mean difference, percent activity, mean ± x SD, median ± x MAD, and *t*-test. For researchers who would like to have a hit identification strategy that has a statistical basis and a control on false hit rates, especially on screens with samples with extreme (large or moderate) effects, the SSMD technique is a more suitable option. However, as also stated by Birmingham et al. [28], it is not easy for nonprogrammers to carry out the calculations from scratch, and no open-access packages have implemented SSMD in their workflow yet. The user-friendly graphical interface of GUItars does not require any command entries from the end-user and makes the SSMD analyses achievable for researchers with no programming background. One of the major advantages of the SSMD-based hit selection metric over the other methods is the feasibility of ranking and classification of the RNAi hits by quantifying the size of RNAi effects. Therefore, GUItars not only selects the hits based on a user-defined hit cutoff, but also provides a list of gene counts based on their effect types (according to the thresholds introduced by Zhang [12]) to guide the researchers for more deliberated hit selection for validation studies. A table of classified siRNA counts from the screen shown is presented in Table 1.

Conclusions

A MATLAB-based open-access software tool, GUItars, for the analysis and illustration of RNAi screening data is described. The user-friendly graphical interface enables rapid analysis setup with

Table 1. siRNA counts classified by effect sizes.

Type	RNAi Effect Classes	RNAi Effect Cutoffs	Counts
upregulated	≥ 5	extremely strong	2
upregulated	$5 > \text{SSMD} \geq 3$	very strong	15
upregulated	$3 > \text{SSMD} \geq 2$	strong	39
upregulated	$2 > \text{SSMD} \geq 1.645$	fairly strong	34
upregulated	$1.645 > \text{SSMD} \geq 1.28$	moderate	110
upregulated	$1.28 > \text{SSMD} \geq 1$	fairly moderate	169
upregulated	$1 > \text{SSMD} \geq 0.75$	fairly weak	235
upregulated	$0.75 > \text{SSMD} \geq 0.5$	weak	338
upregulated	$0.5 > \text{SSMD} \geq 0.25$	very weak	417
upregulated	$0.25 > \text{SSMD} \geq 0$	extremely weak	561
downregulated	$0 > \text{SSMD} \geq -0.25$	extremely weak	542
downregulated	$-0.25 > \text{SSMD} \geq -0.5$	very weak	481
downregulated	$-0.5 > \text{SSMD} \geq -0.75$	weak	365
downregulated	$-0.75 > \text{SSMD} \geq -1$	fairly weak	238
downregulated	$-1 > \text{SSMD} \geq -1.28$	fairly moderate	144
downregulated	$-1.28 > \text{SSMD} \geq -1.645$	moderate	93
downregulated	$-1.645 > \text{SSMD} \geq -2$	fairly strong	34
downregulated	$-2 < \text{SSMD} \geq -3$	strong	21
downregulated	$-3 < \text{SSMD} \geq -5$	very strong	2
downregulated	> -5	extremely strong	0
zero	$= 0$	no effect	0

GUItars output with gene counts ranked based upon the criteria presented by Zhang [12]. Data is generated from a 12-plate luminescence-based assay with 3840 total genes.
doi:10.1371/journal.pone.0049386.t001

the aid of specially designed pop-up menus, push buttons, and a panel of plate configuration checkboxes instead of requiring command entries from the end-user. GUItars uses the UMVUE estimate of SSMD formulas for hit selection, which is preferred for its ability to diminish sample size effects and the false hit rate, making it superior to other widely used high-throughput screening analysis methods. The program can handle the analysis for screens with or without replicates in 96- or 384-well formats. With the demand for higher throughput screening formats, the use of 1536-well plates is gaining a foothold. Therefore, future versions of this software capable of handling this assay format can be made available. For comparison purposes, non-SSMD-based methods such as percent activity, z-score, and *t*-test are also provided as scoring options. In addition to the tab-delimited and Excel file outputs, the graphical outputs generated with GUItars display the most relevant information that is extracted from the input data sets and the analysis results. The general workflow of the program is demonstrated using an siRNA screen with luminescence as the readout. The software features will be further improved by the addition of various data normalization options for edge effect and systematic error corrections [29–32]. Since GUItars can handle data input files in tsv, tab-delimited txt, csv, xls and xlsx formats, it is capable to process the data directly exported from most microplate readers. Even though GUItars is designed for the analysis of RNAi screening data, it is also applicable for small molecule screens. In summary, this automated graphical analysis tool greatly reduces the time necessary to perform high-throughput screening analysis tasks manually, especially for SSMD-based analysis purposes.

Availability and Requirements

- Project name: GUItars
- Project home page: <http://sourceforge.net/projects/guitars/>
- Operating system(s): Microsoft Windows, Linux and Mac versions are available
- Programming language: MATLAB
- Other requirements: Web browser
- License: no license needed
- Any restrictions to use by non-academics: none

The source code, standalone executables and the exemplary dataset are provided in individual zip folders for each operating system along with the complete set of analysis results of the demonstrated screen.

References

1. Mahanthappa N (2005) Translating RNA interference into therapies for human disease. *Pharmacogenomics* 6: 879–883.
2. Brideau C, Gunter B, Pikounis B, Liaw A (2003) Improved statistical methods for hit selection in high-throughput screening. *J Biomol Screen* 8: 634–647.
3. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R (2006) Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 24: 167–175.
4. Gunter B, Brideau C, Pikounis B, Liaw A (2003) Statistical and graphical methods for quality control determination of high-throughput screening data. *J Biomol Screen* 8: 624–633.
5. Konig R, Chiang CY, Tu BP, Yan SF, DeJesus PD, et al. (2007) A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods* 4: 847–849.
6. Gribbon P, Lyons R, Laffin P, Bradley J, Chambers C, et al. (2005) Evaluating real-life high-throughput screening data. *J Biomol Screen* 10: 99–107.
7. Prummer M (2012) Hypothesis Testing in High-Throughput Screening for Drug Discovery. *J Biomol Screen* 17: 519–529.
8. Chung N, Zhang XD, Kreamer A, Locco L, Kuan PF, et al. (2008) Median absolute deviation to improve hit selection for genome-scale RNAi screens. *J Biomol Screen* 13: 149–158.
9. Zhang XD, Yang XC, Chung N, Gates A, Stec E, et al. (2006) Robust statistical methods for hit selection in RNA interference high-throughput screening experiments. *Pharmacogenomics* 7: 299–309.
10. Zhang XD (2007) A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* 89: 552–561.
11. Zhang XD (2007) A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. *J Biomol Screen* 12: 645–655.
12. Zhang XD (2009) A method for effectively comparing gene effects in multiple conditions in RNAi and expression-profiling research. *Pharmacogenomics* 10: 345–358.
13. Zhang XD (2010) Assessing the size of gene or RNAi effects in multifactor high-throughput experiments. *Pharmacogenomics* 11: 199–213.
14. Zhang XD, Lacson R, Yang R, Marine SD, McCampbell A, et al. (2010) The use of SSMD-based false discovery and false nondiscovery rates in genome-scale RNAi screens. *J Biomol Screen* 15: 1123–1131.
15. Zhang XD (2010) An effective method for controlling false discovery and false nondiscovery rates in genome-scale RNAi screens. *J Biomol Screen* 15: 1116–1122.
16. Zhang XD (2011) Optimal high-throughput screening: practical experimental design and data analysis for genome-scale RNAi research. Cambridge: Cambridge University Press. 203 p.

Acknowledgments

We thank Dr. Xiaohua Douglas Zhang of Merck Research Laboratories for many valuable suggestions and discussions; Jimmy Cui for providing automation support and other technical assistance as well as many valuable discussions; Dr. Sergio C. Chai and other members of the Department of Chemical Biology and Therapeutics, as well as collaborators at St. Jude Children's Research Hospital for valuable discussions; and David Galloway of the Department of Scientific Editing for editing the manuscript.

Author Contributions

Conceived and designed the experiments: ANG SSO TC. Performed the experiments: ANG SSO. Analyzed the data: ANG SSO TC. Contributed reagents/materials/analysis tools: ANG SSO TC. Wrote the paper: ANG SSO TC. Designed and developed the GUItars tool: ANG. Participated in screening implementation: ANG. Coordinated the project: SSO. Designed assays: SSO. Participated in the development of and beta tested the GUItars tool: SSO. Participated in the preparation of the manuscript: SSO. Participated in study design and coordination: TC. Read and approved the final manuscript: ANG SSO TC.

17. Boutros M, Bras LP, Huber W (2006) Analysis of cell-based RNAi screens. *Genome Biol* 7: R66.
18. Pelz O, Gilsdorf M, Boutros M (2010) web cellHTS2: a web-application for the analysis of high-throughput screening data. *BMC Bioinformatics* 11: 185–190.
19. Rieber N, Knapp B, Eils R, Kaderali L (2009) RNAiAther, an automated pipeline for the statistical analysis of high-throughput RNAi screens. *Bioinformatics* 25: 678–679.
20. Sui Y, Wu Z (2007) Alternative statistical parameter for high-throughput screening assay quality assessment. *J Biomol Screen* 12: 229–234.
21. Zhang JH, Chung TD, Oldenburg KR (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* 4: 67–73.
22. Zhang XD (2008) Novel analytic criteria and effective plate designs for quality control in genome-scale RNAi screens. *J Biomol Screen* 13: 363–377.
23. Kummel A, Gubler H, Gehin P, Beibel M, Gabriel D, et al. (2010) Integration of Multiple Readouts into the Z' Factor for Assay Quality Assessment. *J Biomol Screen* 15: 95–101.
24. Zhang XD (2011) Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. *J Biomol Screen* 16: 775–785.
25. Zhang XD, Ferrer M, Espeseth AS, Marine SD, Stec EM, et al. (2007) The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. *J Biomol Screen* 12: 497–509.
26. Rousseeuw PJ (1991) Tutorial to Robust Statistics. *Journal of Chemometrics* 5: 1–20.
27. Huber PJ (1981) Robust Statistics. New York: John Wiley & Sons. 301 p.
28. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, et al. (2009) Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* 6: 569–575.
29. Kevorkov D, Makarenkov V (2005) Statistical analysis of systematic errors in high-throughput screening. *J Biomol Screen* 10: 557–567.
30. Makarenkov V, Zentilli P, Kevorkov D, Gagarin A, Malo N, et al. (2007) An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics* 23: 1648–1657.
31. Carralot JP, Ogier A, Boese A, Genovesio A, Brodin P, et al. (2012) A novel specific edge effect correction method for RNA interference screenings. *Bioinformatics* 28: 261–268.
32. Tukey JW (1977) Exploratory data analysis. Massachusetts: Addison-Wesley Pub. Co. 688 p.