Drug Synergy User Guide

George Chen[1], Casey Shannon[1], and Dr. Raymond Ng[1]

[1]Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

August 7, 2020

1 Introduction & General Overview

This application outputs a user-specified number of synergistic drug combinations that are predicted to most reverse an input disease gene expression signature while minimizing drug-drug interactions.

First, the user inputs the Accession Code for their dataset of interest. Once the dataset is downloaded, the user enters other parameters to further specify their data of interest within the dataset. With this, the application performs an analysis to build a list of up- and down-regulated genes for the disease state relative to the control.

After confirming the gene analysis results, the user can submit these lists to Connectivity Map (CMap), a system that predicts individual drugs that, by themselves, most reverse the disease signature [1]. With these results and further user inputs (such as number of synergistic drugs desired), the application then produces a list of synergistic drugs that are predicted to perform better together than each drug on their own.

2 Guide

2.1 Inputting the dataset ID

When the application first opens, you will be prompted to enter your dataset Accession Code (Fig. 1). There are often multiple ways to identify a dataset (such as the GEO Accession Code, which begins with "GSE"). However, please follow the format given in the text input space (beginning with "E-GEOD-" and where the X's are numbers) to avoid errors. If your Accession code is not valid, you will get the following error: *Invalid Accession Code,*

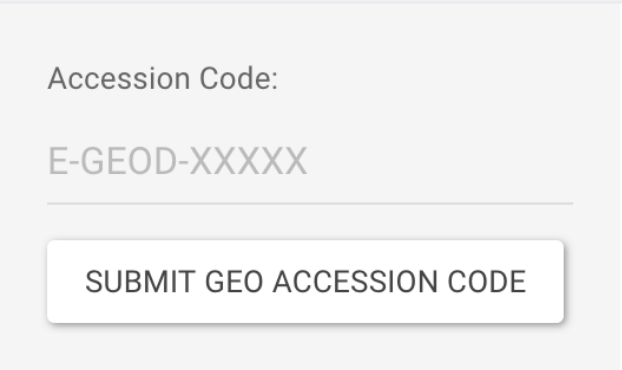

Figure 1: Input Field for Accession Code

Please check your spelling and follow the suggested format.

Once you hit the "SUBMIT GEO ACCESSION CODE" button, your data will begin downloading. This process may take a few minutes.

2.2 Entering Data-Unique Parameters

Once the data finishes downloading, you will be prompted with a few input fields (Fig. 2) and the Sample and Data Relationship Format (SDRF) table (Fig. 3). The SDRF table contains information on each sample (such as the sample name and condition) [2].

For the first drop-down menu select the name of the SDRF column that contains information on the sample names. Similarly, for the second drop-down menu, select the name of the SDRF column that contains information distinguishing the patient conditions (i.e. disease and control). For example, the sample column in the example SDRF would be "Source.Name"

Sometimes, each sample name will have a common suffix and/or prefix appended to them (e.g. **123**sample name**456**). If this is the case, enter the prefix and the suffix in their corresponding text fields. If there is no prefix, you may leave these fields blank.

Once you have entered all this information, click the "UPDATE SDRF" button.

2.3 Specifying Labels and Filters

Once the SDRF has been updated with the information you have inputted, you will be prompted with some more inputs (Fig. 4).

In the first set of check boxes, select the label(s) that you wish to use as the condition(s). If you select multiple, all of those data will be grouped together. For example, if you selected both "POST_SURGICAL" and "sepsis" from the screenshot, the post surgical and

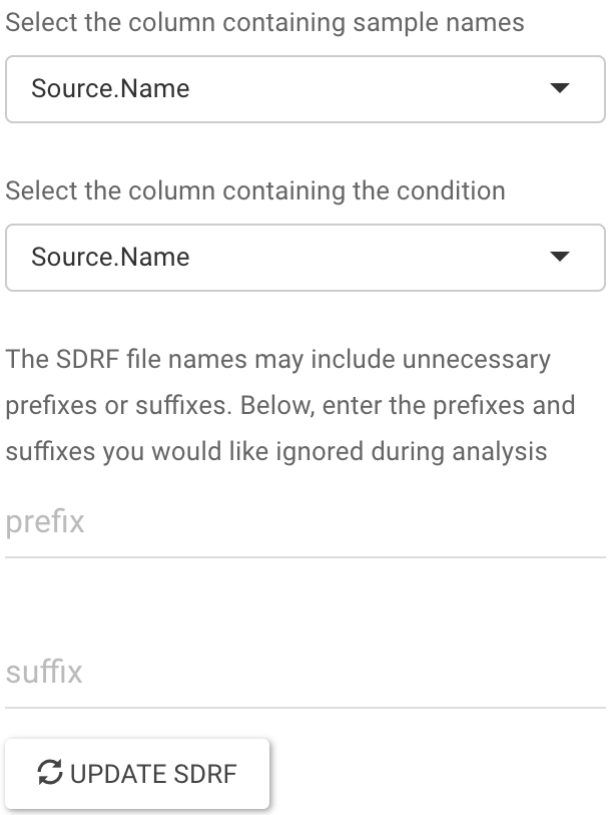

Figure 2: Input Area for SDRF information

	Source.Name	Comment..Sample_description.	Comment..Sample_source_name.	Characteristics.Organism.	Term.Source.REF	Term.Accession.Number	Characteristics.healt
1	GSM712518 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
2	GSM712517 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
3	GSM712516 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
4	GSM712515 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
5	GSM712514 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
6	GSM712513 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
7	GSM712512 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
8	GSM712511 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
9	GSM712510 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL
10	GSM712509 1	Collected within 24 hours post surgery, major open chest surgery	Whole blood	Homo sapiens	EFO	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_9606	POST_SURGICAL

Showing 1 to 10 of 41 entries

Previous 1 2 3 4 5 Next

Figure 3: Sample SDRF table

sepsis patient data will be grouped together and compared against the control. Do the same

thing in the second set of check boxes, except for the label(s) you wish to use as the control(s).

Next, select the number of up- and down-regulated genes you wish to include in your final result under the corresponding prompt. With custom, the maximum number of genes that you can select is 150. **Note:** this represents the *maximum* number of genes that will be in your final result. If you choose to add filters, the number of genes remaining may be less.

Lastly, you may add filters if you wish to filter the final up- and down-regulated gene tables. With filters, you can restrict the values of the log fold-change, average expression, t test statistic (t), probability value (p), adjusted p-value, and unstandardized beta (B) using inequalities ($<, >, <=, >=, =$).

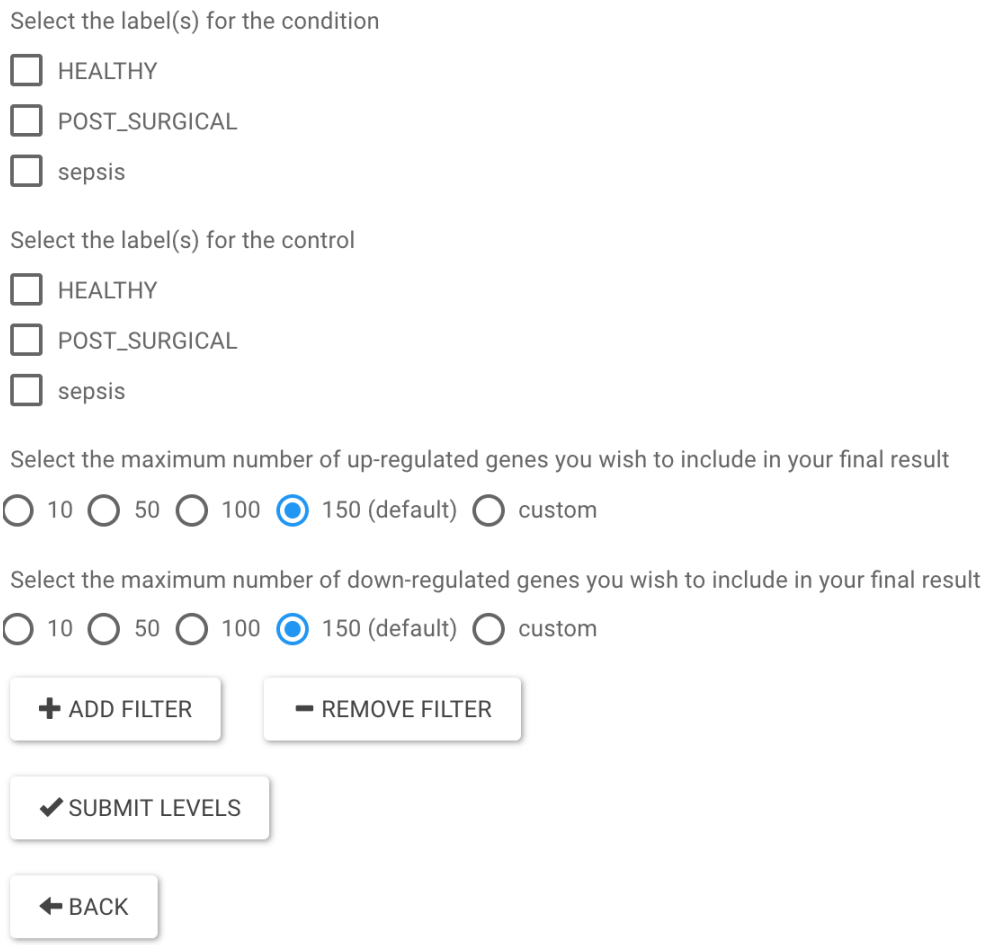

Figure 4: Input fields for differential gene expression analysis

2.4 Submitting the CMap Query

Once the gene expression signature is generated, you will be directed to a page that displays the top up- and down-regulated genes for your disease of interest given your selections

from section 2.3 (Fig. 5).

Once you have reviewed the tables, nothing further has to be done on this page other than clicking the "SUBMIT QUERY TO CMAP" button. This will submit the up- and down-regulated genes to CMap and poll for results until the analysis is complete. **Note:** this step will take *at least* 20 minutes.

Top 150 up-regulated genes

Show 10 entries Search:

	logFC	AveExpr	t	P.Value	adj.P.Val	B
199675	4.47523351955	9.35263299896667	19.3254577537725	5.80482763407077e-19	3.173789508928l9e-14	32.3672590417773
6283	3.37603684285	11.8172981394333	16.2513145257951	8.16567909691357e-17	2.23229252311875e-12	27.8875965124245
56729	4.84617347235	7.3521632407	14.3933763175231	2.3641989532889e-15	3.23156444427677e-11	24.7549176074103
79887	2.01241911	11.15531089	14.0852036341713	4.26406717587338e-15	4.66275745681754e-11	24.2000277681184
3240	4.88108222595	7.37779231245	13.6686088039662	9.61249738590205e-15	5.89283230369778e-11	23.4329097882201
3250	5.09785479475	7.53417216423333	13.6567452341754	9.83733615590289e-15	5.89283230369778e-11	23.4107813236028
6280	1.3629254335	13.8537125093333	13.6271060849584	1.043029786256e-14	5.89283230369778e-11	23.3554054500031
84418	2.7832690471	10.4055999389333	12.8210759618645	5.31100079363336e-14	2.00092621485237e-10	21.8096716916704
57126	5.72682771215	7.40641279156667	12.7349113680505	6.34695593253085e-14	2.16887384756953e-10	21.6397938026743
306	4.00738418235	8.776576752	12.6123542180038	8.18960552530708e-14	2.35666674787455e-10	21.3965873499827

Showing 1 to 10 of 150 entries Previous 1 2 3 4 5 … 15 Next

Figure 5: Example gene table. Top 150 up-regulated genes for sepsis patients in a particular study

2.5 Specifying Synergistic Drug Parameters

Once the CMap query completes, the page will automatically filter the results and display individual drugs predicted to reverse the disease signature in descending order (Fig. 6).

Below this table, you will be prompted with input fields to specify parameters for the synergy analysis (Fig. 7).

In the first field, input the number of synergistic drugs you would like to include in your result (not including the reference drug). You may choose a number from 1-5 (i.e. 2-6 total drugs).

In the next field, input your reference drug. This will be the drug against which the first synergistic drug will be evaluated. It does not need to be a drug in the CMap results table. If you leave this field blank, the program will automatically use the top drug as the reference drug.

In the last field, input your drug interaction threshold (0.000-1.000). Any drugs predicted to interact with a score higher than your interaction threshold will be ignored. A score of 1.000 means 'maximum similarity' and a score of 0.000 means 'maximum dissimilarity.'

Show 10 entries

Search:

	id	pert_iname	pert_type	TAG
1	BRD-K40530731	hyoscyamine	trt_cp	-96.6525726318359
2	BRD-K49456190	prima-1-met	trt_cp	-95.0669479370117
3	BRD-K67566344	KU-0063794	trt_cp	-94.2357177734375
4	BRD-A22897758	pinacidil	trt_cp	-93.9906311035156
5	BRD-K20285085	fostamatinib	trt_cp	-92.9521484375
6	BRD-K32821942	azathioprine	trt_cp	-92.5384521484375
7	BRD-K49865102	PD-0325901	trt_cp	-90.9398956298828
8	BRD-K77286328	reversine	trt_cp	-89.9127044677734
9	BRD-K69688083	mestinon	trt_cp	-89.7166290283203
10	BRD-K59197931	naproxen	trt_cp	-89.4552001953125

Showing 1 to 10 of 1,256 entries

Previous 1 2 3 4 5 … 126 Next

Figure 6: Example CMap query results. All drugs predicted to individually reverse the disease signature in descending order.

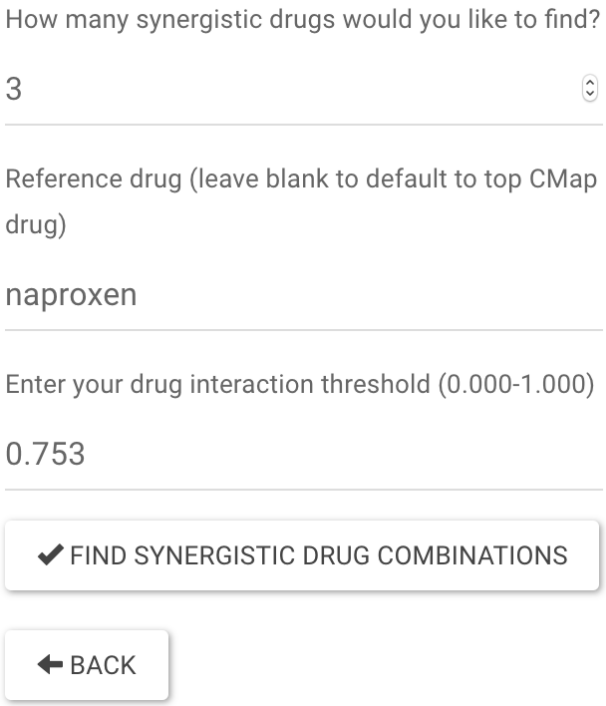

Figure 7: Example values for the synergy input fields.

Once you have inputted all your fields, click the "FIND SYNERGISTIC DRUG COMBINATIONS" button and your results will be processed. This may take a few minutes.

2.6 Interpreting Results

Once the synergy analysis completes, the results will be displayed on the next page as a menu with expandable items (Fig. 8). Each item represents one drug, and contains

information on the drug mechanisms, pubchem id, structure, synergy score, and interaction scores/comments with all the drugs above it.

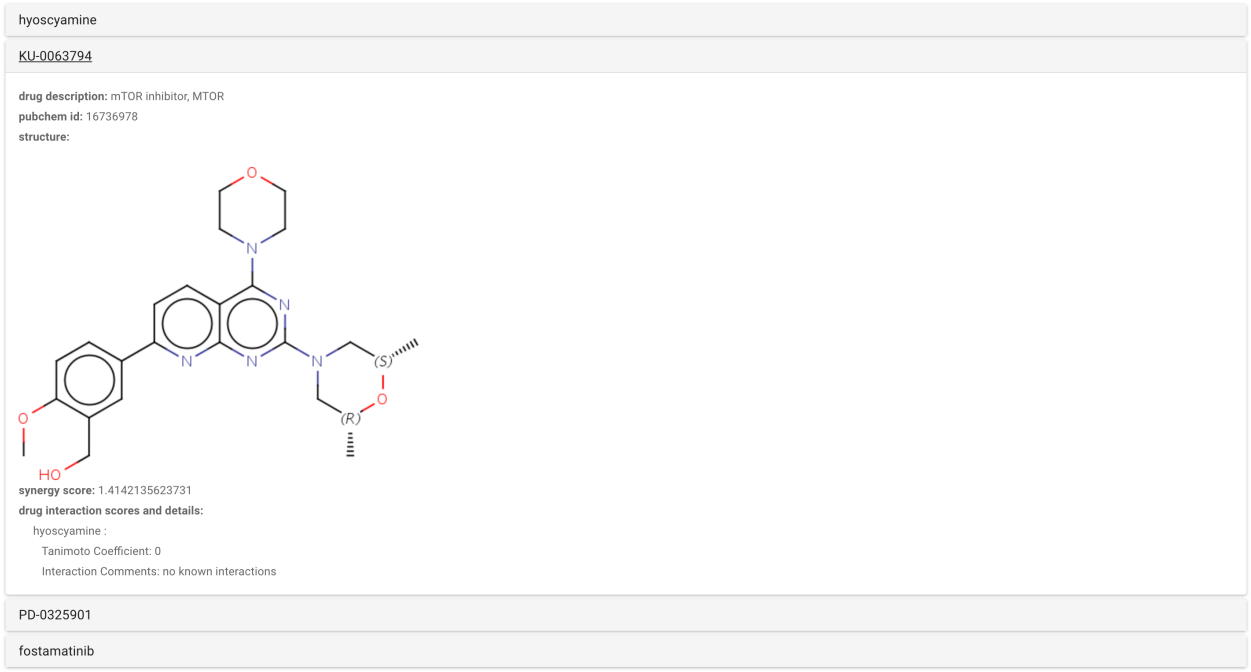

Figure 8: Example results for synergy analysis.

3 Background & Information for Developers

This section contains background information on how the analyses work in the application, as well as some details on where things are so that developers can expand the code and/or modify it to fit their specific needs if they wish to do so.

All genomics data is obtained from ArrayExpress and stored in a local temporary directory. You can find more information about downloading information in the *controller/AE_downloader.R* file. All patient files downloaded from ArrayExpress are loaded into the application in the *controller/gene_analyzer/file_path_loader.R* file. The top up- and down-regulated genes are computed in the *controller/gene_analyzer/tt_generator.R* file. During this process, all the Affymetrix probe IDs are converted into gene entrez IDs using the *hgu133plus2.db* data package.

Drugs are synergized via targeting orthogonal pathways. This method was adapted from the SynergySeq tool [3]. For each synergistic drug, its orthogonality score is computed with respect to the disease signature and all previous synergistic drugs. The orthogonality score is defined as: $\sqrt{(1-CR)^2+DR^2}$ where CR is the concordance ratio and DR is the discordance ratio. The drug with the largest orthogonality score is the next synergistic drug.

CR is calculated as the ratio of the number of genes that go in the same direction when comparing the potential synergistic drug with the list of already computed synergistic drugs to the number of genes that go in the opposite direction.

DR is calculated as the ratio of the number of genes that go in the opposite direction when comparing the potential synergistic drug with the disease signature and which are absent from the combined signature of already computed synergistic drugs to the number of genes that go in the same direction.

Drug synergy calculations can be found in the *controller/drug_synergizer.R* file.

Drug interactions are predicted using the dataset from Vilar's study containing current known Drugbank interactions as well as predicted interactions [4]. The drug interaction data is loaded into an R dataframe in *tools/interaction_table_loader.R* and the interactions are taken into account in *controller/drug_synergizer.R* as drug synergy is being computed.

4 Known Bugs & Future Features

1. Currently does not support the user's own datasets
2. Currently only supports datasets where each patient's data is split into separate files.
3. Currently only supports data obtained from Affymetrix GeneChip Human Genome U133 Plus 2.0 [HG-U133_Plus_2]

5 Releases

August 6, 2020: initial release

References

[1] Subramanian A, et al. (2017). A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 Profiles. *Cell. 171*(6):1437–1452.

[2] Embl-Ebi. (n.d.). Creating a SDRF. Retrieved August 07, 2020, from https://www.ebi.ac.uk/arrayexpress/help/creating_a_sdrf.html

[3] Stathias, V., Jermakowicz, A.M., Maloof, M.E., Forlin, M., Walters, W., Suter, R.K., Durante, M.A., Williams, S.L., Harbour, J.W., Volmar, C.H., et al. (2018). *Drug and disease signature integration identifies synergistic combinations in glioblastoma.* Nature Communications 9, 5315.

[4] Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., & Friedman, C. (2012). *Drug—drug interaction through molecular structure similarity analysis.* Journal of the American Medical Informatics Association, *19*(6), 1066-1074.