DIGREM web portal user manual

Method overview

DIGREM integrates three methods (DIGRE, IUPUI_CCBB and gene set-based methods) to predict drug combination effects by explicitly modeling the drug response dynamics and gene expression changes after individual drug treatments.

The DIGRE model developed by our group won the best performance in the National Cancer Institute's DREAM 7 Drug Combination Synergy Prediction Challenge, an international crowdsourcing-based computational challenge for predicting drug combination effects using transcriptome data. This challenge's blind-assessment of submitted computational models revealed that the prediction of drug pair activity from DIGRE was significantly consistent with the vast majority of the organizers' experimental validations. In addition, we further validated our DIGRE model using another experimental dataset.

IUPUI_CCBB is the second ranked method in DREAM 7 challenge. It has a similar workflow as DIGRE but does not consider gene-gene interaction and dose response curve. Statistical test is applied to identify commonly disturbed genes in a core gene set by two drugs, therefore replicates are required to ensure the statistical power.

Gene set-based method is another recently developed method aiming the same question. Rather than looking at individual genes or a defined gene set, it calculates a gene set enrichment score across all gene ontology and oncogenic signature gene sets, and ranks drug pairs by an average percentage of commonly disturbed genes with the co-enriched gene sets.

DIGREM could potentially be used for large-scale discovery of effective drug combinations for further experimental validation, possibly leading to the rapid identification of new therapies for complex diseases.

DIGRE model workflow

Step1:

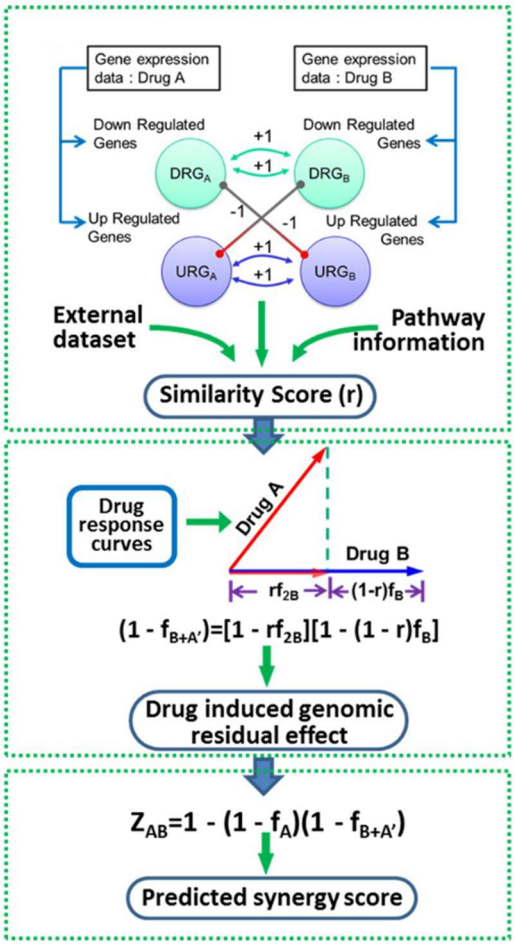
Show the derivation of the similarity score (r) between drug A and drug B from their genomic profiles.

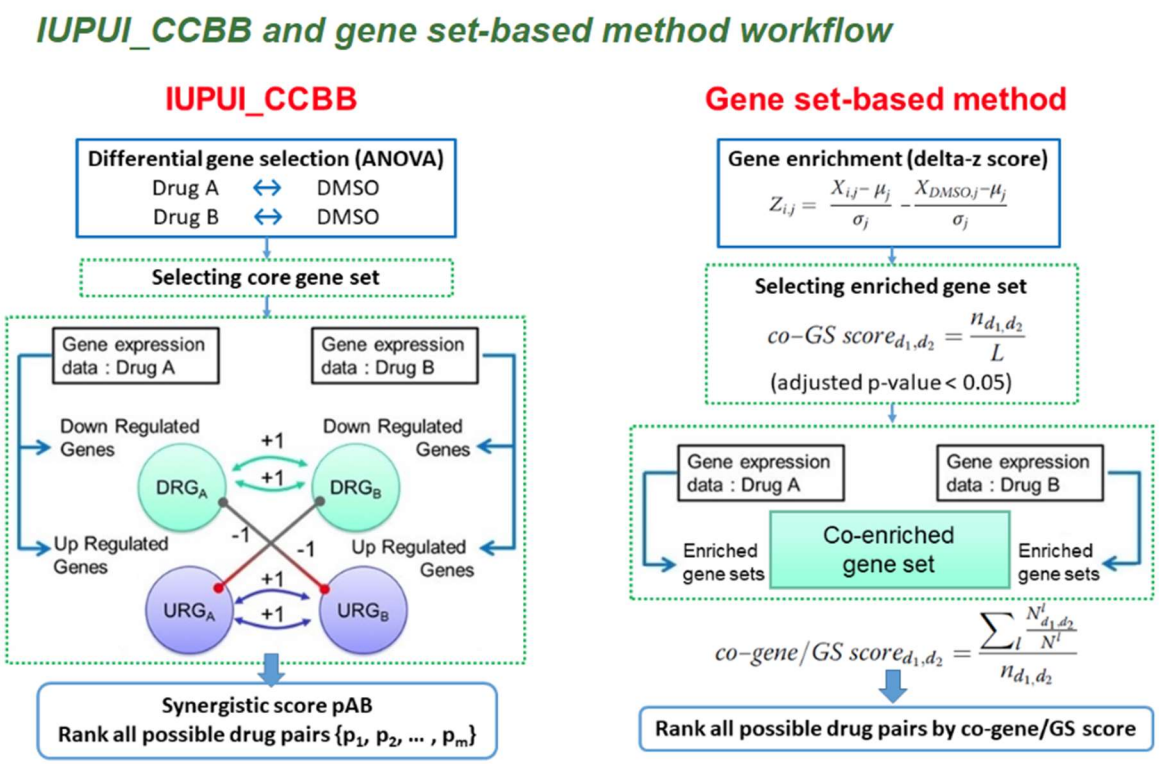
Step2:

Depict the estimation procedure of the drug-induced genomic residual effect from the similarity score.

Step3:

Show the model for the final synergy score.





Prepare input data

The DIGREM website takes three forms of input data to predict the compound synergistic effect: the drug treated gene expression data, the drug dose response data and the gene-gene interaction information.

1. Gene expression profile (required)

This is the gene expression profile (microarray or RNA-Seq) of cell after drug perturbation. User should prepare a table with each column representing one drug, and each row representing one gene. Negative control such as DMSO treatment should be included (the corresponding column name should be "Neg_control", otherwise DIGREM will not recognize it). The demo data we provided is from NCI-DREAM challenge competition for predicting drug pairs synergy. OCI-LY3 human diffuse large B-cell lymphoma (DLBCL) cell line was treated by 14 different drugs in its dose of IC20. Gene expression level was measured 24 hours after drug perturbation.

Home Insert Page Layout Formulas Data Review V

A26 fx AAK1

	A	B	C	D	E
1	Genename	Neg_control	Aclacinomyci	Blebbistatin	Blebbistatin
2	A1CF	3.91	3.81	3.76	3.69
3	A2BP1	4.26	4	4.22	4.03
4	A2LD1	5.49	5.54	5.55	5.55
5	A2M	4.16	3.91	4.2	4.45
6	A2ML1	3.38	3.47	3.39	3.33
7	A2ML1	3.48	3.54	3.5	3.42
8	A3GALT2	4.18	4.39	4.23	4.13
9	A4GALT	4.55	4.76	4.73	4.69

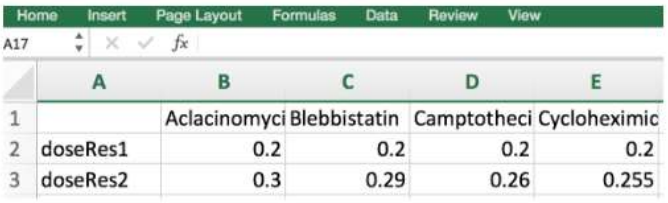
Only accept CSV fil. Maximum file size 10M.
The first column must be gene names. The first row of other columns should be compound name.
At least one column of negative control data should be provided. And the column name should be "Neg_control".
Same compound may have more than one column, in which case the average values will be use in calculation.

Example:

2. Dose response data (optional)

This is the dose response data of each drug. It is only required by DIGRE model (not for IUPUI_CCBB and gene set-based method), but if you do not provide dose response data, DIGRE will still predict based on drug similarity score.

To have a correct data format, user should prepare a table that contains the cell viability reduction values when cell line is treated with drug in two different dose. One drug dose is the IC20 of the drug, therefore the cell viability reduction is always 0.2 for all drugs. The other dose is double dose of IC20, This value is inferred from the dose response curve of each drug. (This step should be done by user himself, DIGREM only accepts the processed data not raw dose response curves; currently, the two drug doses are fixed: the IC20 and double of the IC20.) The demo data we provided contains dose response data of 14 drugs from NCI-DREAM challenge. (Notice: Check the drug name in the dose response data that match exactly the drug name in the gene expression data. The row names of the csv file should be “doseRes1” and “doseRes2”, otherwise DIGRE will not recognize it)

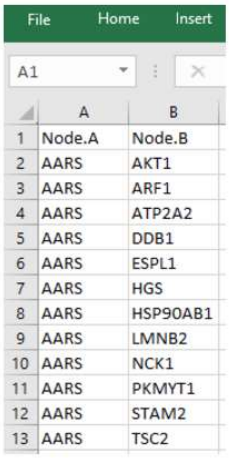


Only accept CSV fil. Maximum file size 60K.
The content should have three rows.
The first row should be compound names. Each compound should only have one column.
The second row should be the values of doseRes1. The first column should be labeled as "doseRes1".
The third row should be the values of doseRes2. The first column should be labeled as "doseRes2".

Example:

3. Gene-gene interaction information

This is the gene-gene interaction information needed for DIGRE to predict drug interaction effect. It is only required when user select “gene network” option in connectivity data (if KEGG pathway is selected, other additional input is required). It is recommended when user have knowledge about the gene network involved in the specific cell line he used in the experiment. The data table should contain two columns with gene SYMBOL names. Each raw represents two connected genes. The interaction is regarded as undirected. The demo data is a lymphoma-specific gene network refined from lymphoma patients gene expression data.



Only accept CSV file. Maximum file size 1M.
The file should have two columns to specify the two connected genes with their SYMBOL name.
Example: the following gene-network example files are refined from lymphoma patients gene expression data. If you are using lymphoma cell lines, you could use our gene network, otherwise we recommend you use KEGG pathway or upload your own gene-network file.

Gene network:

Requirement of data for DIGREM

The DIGREM will automatically call three methods to do prediction if the requirement is meet, and output the predicted result of each method.

1. DIGRE model

DIGRE is the method that has the least requirement, therefore is always run. It takes the above-mentioned three input to do prediction with tolerance of no dose response data.

2. IUPUI_CCBB model

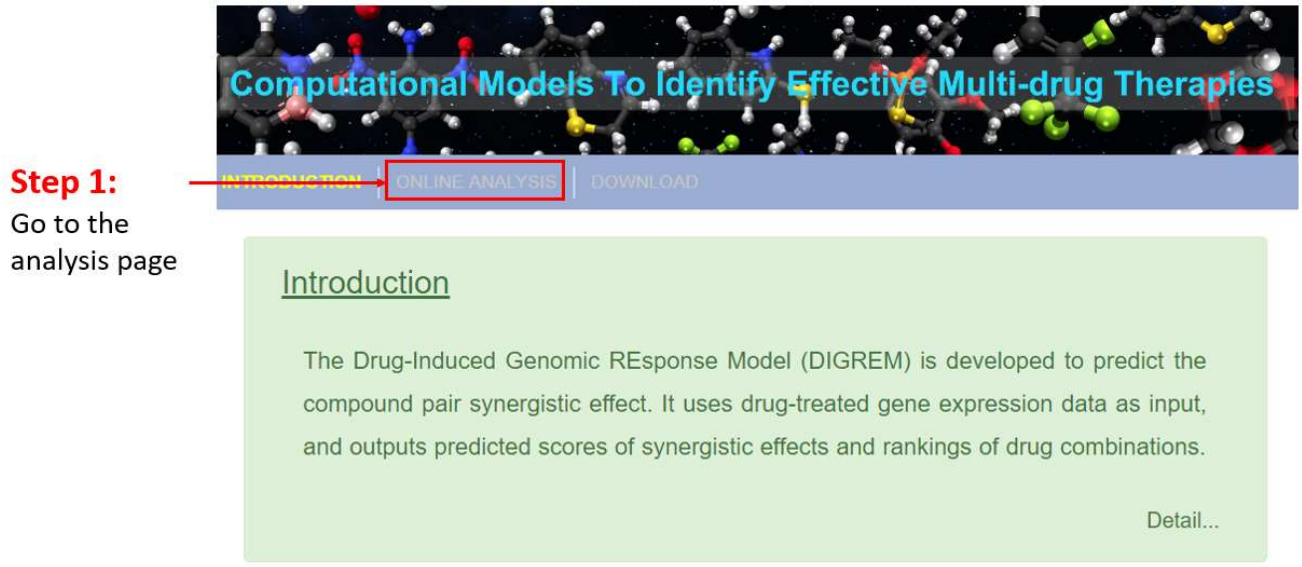
IUPUI_CCBB method requires the gene expression data to have at least three replicates for each drug (to ensure enough statistical power). If the requirement is meet, this method will be run, and prediction will be added to final output report, otherwise this method is skipped.

3. Gene set-based model

Gene set-based method require the gene expression data contains at least 16 columns to do normalization. If the requirement is meet, this method will be run, and prediction will be added to final output report, otherwise this method is skipped.

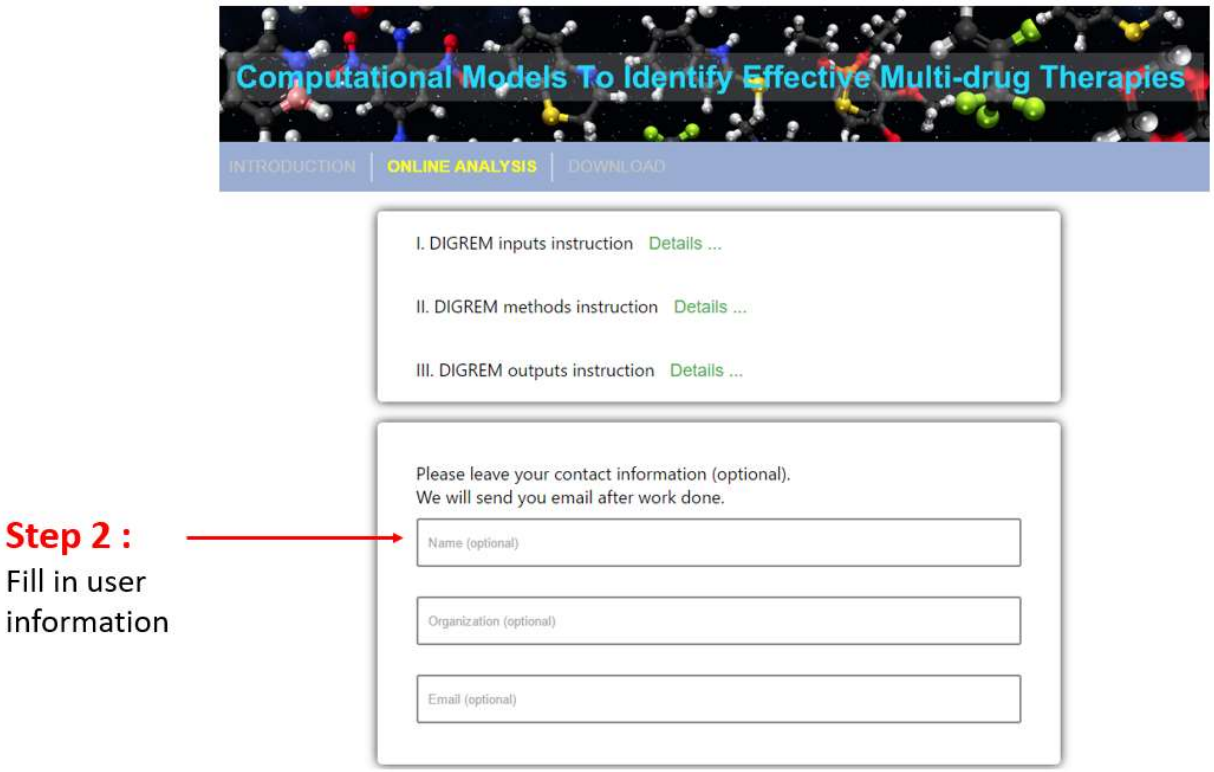
Workflow of using DIGREM website

1. Step1: Go to the analysis page



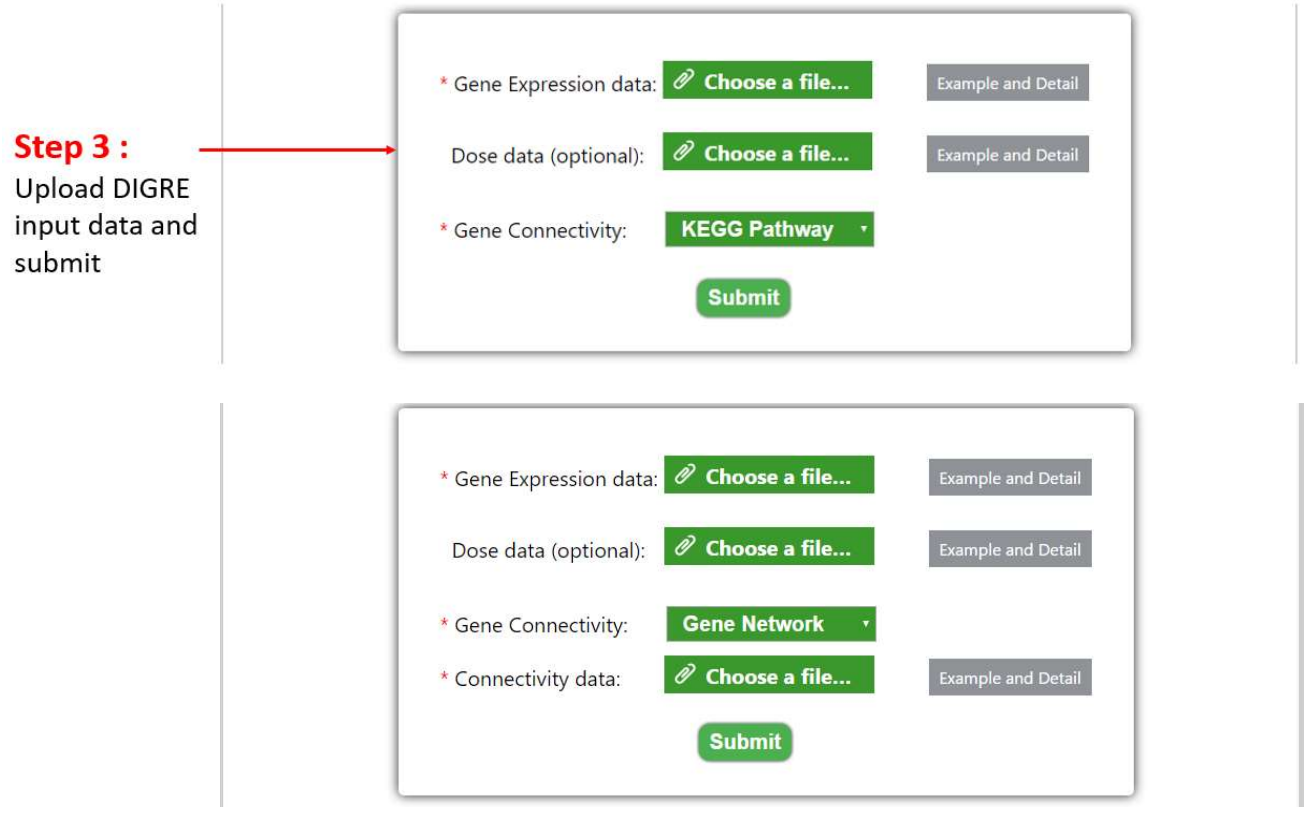
2. Step2: Fill in user information

Fill in the name, organization and e-mail information. This step is optional. If e-mail address is provided, an e-mail will be sent to the user once the job is done.



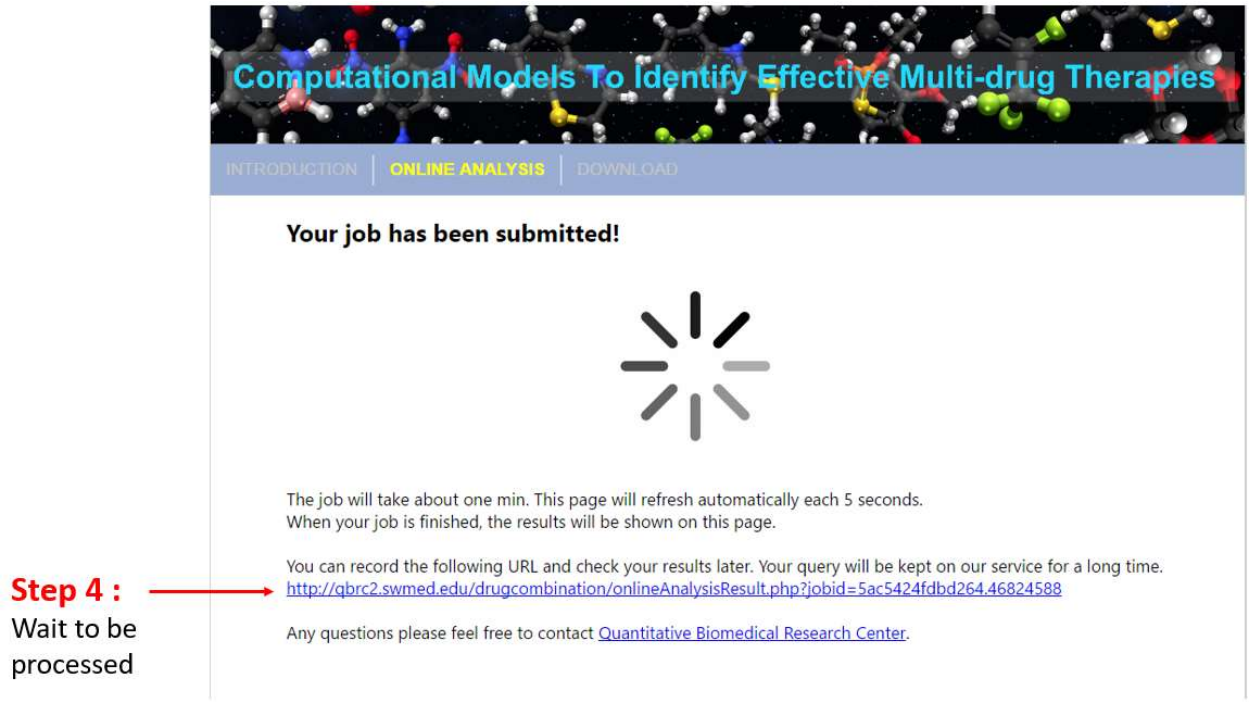
3. Step3: Upload input data and submit

Upload the gene expression data and dose response data. If user select the Gene Network method, the additional network input data should be given.



4. Step4: Wait to be processed

Once user submit his job, the website will be directed to a waiting page. User can record the url showed in the page, and come back to check the status of job.

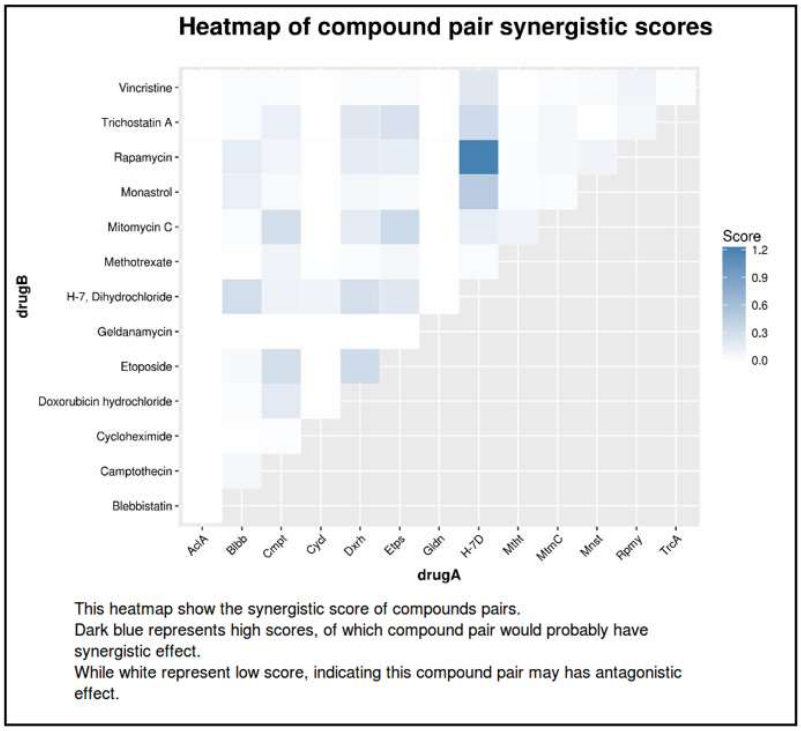


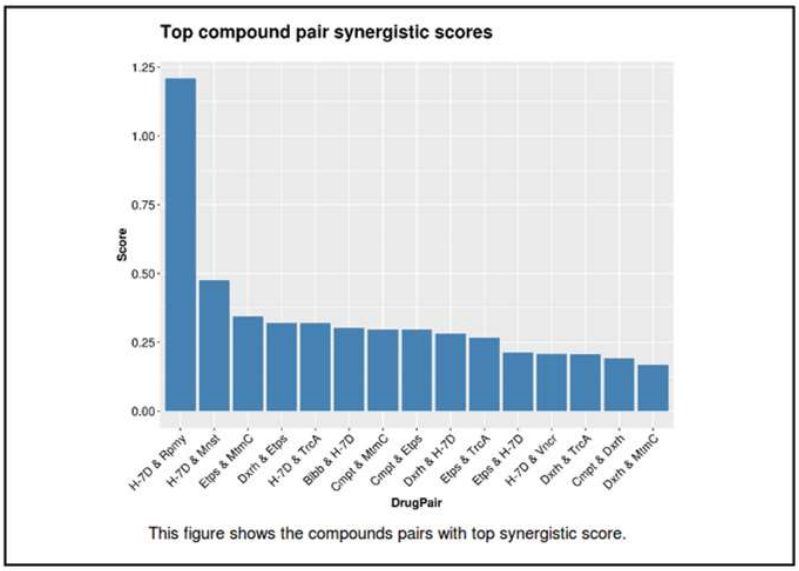
5. Step5: Visualize result

Once the job is done, the prediction result will be displayed automatically in the page. The heat map displays the predicted synergistic scores of all the compound pairs. Deep color indicates high synergistic score. The bar plot shows the top predicted synergistic drug pairs. The result will be stored in our server for a long time. User can record the link and go back at any time.

Step 5 :

Visualize result





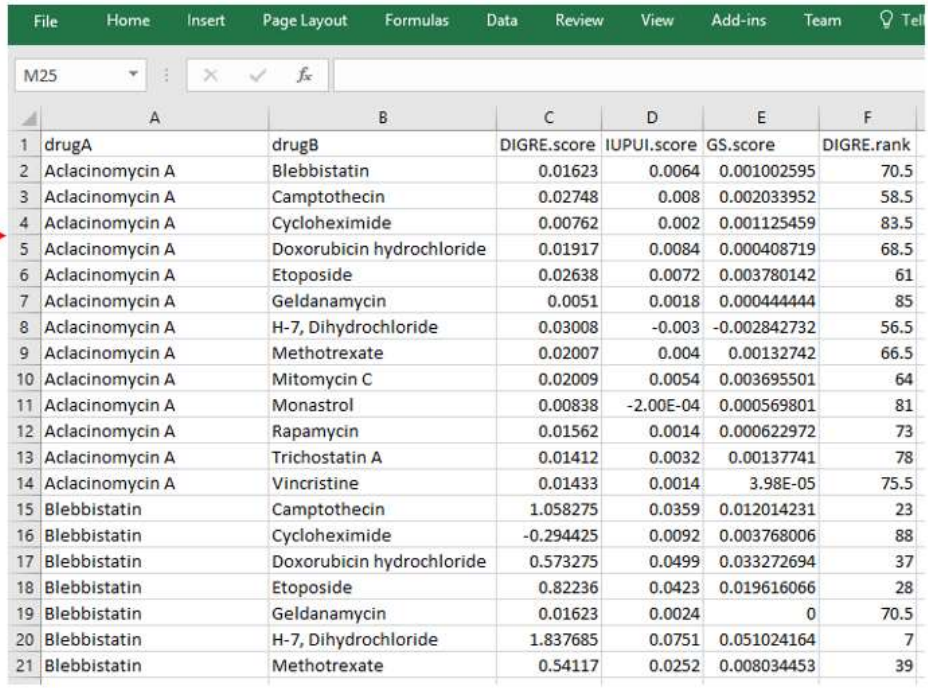
DOWNLOAD

6. Step6: Download result

User can click on the download button to download the result predicted by DIGREM.

Step 6 :

Download result



File Home Insert Page Layout Formulas Data Review View Add-ins Team Tel

M25

	A	B	C	D	E	F
1	drugA	drugB	DIGRE.score	IUPUI.score	GS.score	DIGRE.rank
2	Aclacinomycin A	Blebbistatin	0.01623	0.0064	0.001002595	70.5
3	Aclacinomycin A	Camptothecin	0.02748	0.008	0.002033952	58.5
4	Aclacinomycin A	Cycloheximide	0.00762	0.002	0.001125459	83.5
5	Aclacinomycin A	Doxorubicin hydrochloride	0.01917	0.0084	0.000408719	68.5
6	Aclacinomycin A	Etoposide	0.02638	0.0072	0.003780142	61
7	Aclacinomycin A	Geldanamycin	0.0051	0.0018	0.000444444	85
8	Aclacinomycin A	H-7, Dihydrochloride	0.03008	-0.003	-0.002842732	56.5
9	Aclacinomycin A	Methotrexate	0.02007	0.004	0.00132742	66.5
10	Aclacinomycin A	Mitomycin C	0.02009	0.0054	0.003695501	64
11	Aclacinomycin A	Monastrol	0.00838	-2.00E-04	0.000569801	81
12	Aclacinomycin A	Rapamycin	0.01562	0.0014	0.000622972	73
13	Aclacinomycin A	Trichostatin A	0.01412	0.0032	0.00137741	78
14	Aclacinomycin A	Vincristine	0.01433	0.0014	3.98E-05	75.5
15	Blebbistatin	Camptothecin	1.058275	0.0359	0.012014231	23
16	Blebbistatin	Cycloheximide	-0.294425	0.0092	0.003768006	88
17	Blebbistatin	Doxorubicin hydrochloride	0.573275	0.0499	0.033272694	37
18	Blebbistatin	Etoposide	0.82236	0.0423	0.019616066	28
19	Blebbistatin	Geldanamycin	0.01623	0.0024	0	70.5
20	Blebbistatin	H-7, Dihydrochloride	1.837685	0.0751	0.051024164	7
21	Blebbistatin	Methotrexate	0.54117	0.0252	0.008034453	39