TGx-DDI Biomarker for DNA Damage Classification User Guide

Page Navigation (Quick Links):

1) Description

The TGx-DDI Biomarker for DNA Damage Classification (TGx-DDI Classifier) enables users to predict the likelihood that a test article will induce DNA damage. The gene set (originally comprised of 64 gene transcripts) underlying the classifier was derived from human TK6 cells exposed to a training set of prototypical DNA damage-inducing chemicals: 13 DNA damage-inducing, 15 non-DNA-damage inducing. A user's test data gene signature is compared to that of the biomarker and the probability that the test article can cause DNA damage is calculated.

To open the application, select the "DNA Damage Classifier" button on the CEBS Homepage. The TGx-DDI Classifier Homepage provides links to the Classifier application and to pages that describe its development, use and example data files.

Publications: a list of publications describing development and use of the TGxDDI biomarker. Each publication has a PMID or PMCID link to the published abstract and a DOI number

Biomarker Description: describes the development of the biomarker, key points for consideration when using the application, and data file formats for each available microarray platform

Example Data Files: downloadable data files for each microarray platform and examples of required data formats are provided for users to test the application.

Microarray platforms supported by the application include:

- Agilent Human Genome 8X60K (text delimited files)
  - one dye
  - two dyes
  - dye-swapped
- Affymetrix Human Genome U133 Plus 2.0 array (CEL file)
- Generic arrays (using log2 transformed data for a single chemical and concentration in a text delimited file)

- Batch data (using log2 transformed data for multiple chemicals and/or concentrations in an Excel file)

2) Data submission

Study information related to the microarray data being uploaded for analysis is requested. Use the data fields displayed in the **Study Information** section for this purpose.

- Default entries in the first three fields should be changed to match the study for which data will be analyzed:
  - Cell line,
  - Sample time post-exposure, and
  - Microarray platform used
- Study information is optional for other data fields described in the table below

Description of Study Information Data Fields

Field	Description
Cell line	Name of cell line assessed (dropdown)
Sample time post-exposure	Time between last application of test chemical and sampling for microarray analysis (drop-down); other allows manual entry
Test chemical response in Ames test	Ames test response to test chemical (dropdown)
Test chemical response in chromosome aberration	Chromosome aberration test response to test chemical (dropdown)
Dose optimization performed	Check if dose optimization was performed
Description of dose optimization	Pop-up text box for describing the dose optimization procedure and/or results (optional) when dose optimization box is selected
Cytotoxicity observed in analysis sample	Check if cytotoxicity was observed
Description of cytotoxicity observed	Pop-up text box for describing the observed cytotoxicity (optional) when the cytotoxicity box is selected
S9 activation used	Check if S9 activation was used
Description of S9 activation	When the S9 activation check box is selected, a text box will be displayed in which to describe the S9 activation used (optional)
Cell line has intact p53	Check if the cell line has intact p53
Description of p53 status	When the p53 check box is selected, a text box will be displayed in which to add details regarding the p53 status (optional)

Field	Description
Microarray platform	Select the name of the microarray platform used in the study • Affymetrix Human Genome U133 Plus 2.0 Array • Agilent Whole Genome Microarray 8x60K Dye Swap • Agilent Whole Genome Microarray 8x60K Two Color Treated Cy5 • Agilent Whole Genome Microarray 8x60K Two Color Treated Cy3 • Agilent Whole Genome Microarray 8x60K One Color • Generic Array (probe set defined by user; log2 transformed data
Additional information	Any additional information that may be helpful such as institution, investigator, etc.

Enter information about the data being uploaded in the data fields of the **File Upload** section.

- Required information includes:
  - Test chemical name
  - Concentration of test articles (with unit)
  - Graph label name o Microarray data file(s)
- The CASRN is optional but preferable

Description of File Upload Fields

Field	Description
Test type	Select sample type from dropdown: test chemical or positive control
Test chemical name (full name)	Name of the test chemical or a blind ID can be used if desired
CASRN	Chemical Abstracts Registry Number
Concentration[1, 2]	Concentration of the test chemical with unit for data being analyzed
Graph label name[1, 2]	Short name for the test article to appear on the heat map and cluster analysis
Treatment file[3]	For Affymetrix U133 and Agilent one color arrays; select the CEL (Affymetrix) or txt (Agilent) file for the test article
Control file[3]	For Affymetrix U133 and Agilent one color arrays; select the CEL (Affymetrix) or txt (Agilent) file for the control
Treated (Cy5) to control (Cy3) [3]	For Agilent dye swap and two-color arrays; select the Cy5:Cy3 data file
Treated (Cy3) to control (Cy5) [3]	For Agilent dye swap and two-color arrays; select the Cy3:Cy5 data file
General file[3]	For Generic array or Batch data; select the text or Excel data file to be analyzed
Log(2) ratio column	For Generic array, select the column number containing log(2) transformed data; for Batch data enter "All"

[1] Required data field

[2] Not required for analysis of batch data

[3] File upload options are determined by the platform selected in **Study Information** section

Input data format requirements

- Agilent – delimited text file (.txt);
- Affymetrix – CEL data file (binary data output from Affymetrix);
- Generic – delimited text file (.txt) containing log2 transformed data; and
- Batch Data – Excel file (.xlsx) containing log2 transformed data

Probe IDs in Generic and Batch data files should use Agilent IDs. If they need to be converted to Agilent IDs, this can be achieved using a web based tool such as the Gene ID Conversion Tool from DAVID Bioinformatics Resources (https://david.ncifcrf.gov/conversion.jsp).

Generic Data File Format Requirements (text delimited)

Data must be normalized log2 transformed data saved as a tab-delimited text file. Probe ids should be in column one followed by the data in subsequent columns which are labeled according to data type (e.g., treated, control, chemical tested) and/or concentration.

Sample Generic array data file format

Probe	NTHi.treated.LR-CC	NTHi.treated.WT	OVA.treated.LR-CC	OVA.treated.WT
Aaas	-0.262378693599992	0.146432693658835	-0.157264050379592	-0.0651864466863898
Aacs	-1.53491245721088	1.10814149987023	-1.2157214508298	0.418492940073264
Aadac	-0.0586933881927107	0.1442209728927	0.0788273928233627	-0.0923946951609533
Aadacl1	-0.733289451992194	0.62914735709511	-0.632552945614945	-0.492177435976563
Aadat	-0.159316061603705	0.0613344138502185	-0.115660975691996	0.0753399558126331
Aak1	0.0904840090828394	0.0556915318859392	0.221756142818329	0.0216089565332158
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Zxda	2.45091984025087	0.120654237438437	3.34296788358073	0.320819430417137
Zyg11b	-0.0478395460279151	0.651829709531177	0.0557371282639751	0.0712314806347631
Zyx	0.374573669921801	0.407195369109495	0.205405113752109	0.163212746980367
Zzz3	0.0391893916866923	0.259348251462574	-0.17440149935954	0.0577744934435049

Batch Data Format Requirements

Input Data File (Excel format)

Like the generic text file format, the Excel file needs the probe ids listed in column one followed by normalized log2 data in the subsequent columns. These columns contain the data and are labeled according to data type (e.g., treated, control) and/or concentration.

Sample Batch data file format

	RG-ASPIRIN-9.5 nM	RG-ASPIRIN-95 nM	RG-ASPIRIN-300 nM	RGCAFE9.5 nM	RGCAFE95 nM	RGCAFE300 nM	RG-Sucrose-9.5 nM	RG-Sucrose-95 nM	RG-Sucrose-300 nM
ABCC5	-1.126164	-1.269833	-1.25312	1.0062	1.1522	1.0864	1.057812	1.104773	1.06714
ABCG1	1	1	1	1	1.0261	1	1	1	1
ADAT1	1.0169705	-1.00351	-1.00351	-1.0399	1.0068	-1.0399	-1.02030	1.041262	-1.02030
APOE	-1.106668	-1.373636	-1.17446	-1.188	-1.414	-1.4319	1.042606	1.151067	1.186988
ARNT2	-1.118083	-1.133342	-1.13439	-1.0381	1.0495	1.1557	1.105433	-1.11896	1.007305
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
ZMIZ1	-1.26888	-1.499692	-1.24535	-1.1992	-1.642	-1.4048	-1.44029	-1.22040	-1.01522
ZNF318	-1.003878	-1.003878	-1.00387	1.0028	1	1	1	1	1
ZNF395	1.2128864	-1.045602	-1.13918	-1.2072	-1.303	-1.245	-1.34889	1.021645	1.219635
ZNF451	-1.001233	-1.00131	-1.00131	1.0167	1	1	1.00185	1	1
ZNF586	-1.002545	-1.002545	-1.00254	1	1	1	-1.01337	-1.01431	-1.01720

User options after uploading data:

- [Submit] to start the analysis and generate the results table;
- [Reset] to clear the form and start over;
- [Show Results] to view the Results page after data are displayed from previous analyses; and

- Tool bar options to navigate to the respective webpages.

3) View results

Basic study information, classification, analytical results and links to output data files are available in the summary data table on the Results webpage. Options to view and/or download data are available as described in the table below starting with "Data Files".

Description of Results Table Fields

Field	Description
Edit	Option to remove a single data row from the table
Data Files	Links to view/download analytical data. Fold Change: fold change in gene expression for user data vs. the biomarker test chemical data; Gene Cluster and Chemical Cluster: Euclidian distance data from cluster analysis of (dis-)similarity of genes and/or chemicals
Plot Files	Link to download heat map and plots from cluster analysis of user data compared to the biomarker test chemical data
Class	Classification of the chemical analyzed as DNA damage inducing or non-DNA damage inducing
DDI Probability	Probability that the chemical induces DNA damage with a probability of 1 indicating that the chemical is likely to cause DNA damage
NDDI Probability	Probability that the chemical does not induce DNA damage; with a probability of 1 indicating that the chemical is not likely to cause DNA damage
Test Type	Test type identified in Study Information section
Chemical Name	Chemical name or blind ID identified in Study Information section
Concentration	Concentration with unit of chemical analyzed from File Upload section
Cell Line	Cell line identified in Study Information section
Sample Time	Sample time identified in Study Information section
Platform	Microarray platform selected in Study Information section

Field	Description
Submitted File(s)	Name of user-submitted data file associated with results displayed
Submission ID	Unique application-generated ID associated with data submitted and results for reference to archived data

Additional options:

In addition to the data view/download options available in the results table, other download and navigation buttons are available.

- Select [DOWNLOAD RESULTS TABLE] to download data in Results table columns "Test Type" through "Submission ID" as a text delimited file;
- Select [DOWNLOAD ALL FILES] to download all data and image files in the Results table in a zip file;
- Select [ADD NEW DATA] to return to the Data Submission page to analyze additional data; or
- Select [CLEAR ALL RESULTS] to clear the Results table and start compiling a new one;

4) Example data analysis

For this example, we will use fictitious data from a hypothetical study of benzo(a)pyrene (BaP) to demonstrate how the application works. For this study, TK6 cells were exposed to BaP for 72 hours then gene expression was measured using an Affymetrix Human Genome U133 Plus 2.0 array.

Dose optimization experiments were performed prior to running the test sample using a range of concentrations between 15 and 100 ug/ml. mRNA was isolated and used to assess transcriptional response in the stress response genes *ATF3*, *GADD45A*, and *CDKN1A* to determine the optimal dose for the microarray analysis.

Data in the Affymetrix example data files, available for download on the 'Example Data Files' webpage, will be used for the analysis. The files are Treatment (CEL) and Control (CEL) for Affymetrix Human Genome U133 Plus 2.0 Array files.

a) Enter Study Information

Based on the study scenario described above, enter study data in the Study information data fields.

Example Study Information Values

Data Requested	Example Value
Cell line	TK6
Sample time post-exposure	4 hr
Test chemical response in Ames test	Unknown
Test chemical response in chromosome aberration	Unknown
Dose optimization performed	Yes [*click check box*]
Description of dose optimization	*ATF3*, *GADD45A* and *CDKN1A* mRNA levels assessed in dose range 15-100 ug/ml
Cytotoxicity observed in analysis sample	Yes [*click check box*]
Description of cytotoxicity	No appreciable cytotoxicity observed at ≤ 75 ug/ml; 50% cytotoxicity observed at 100 ug/ml
S9 activation used	No [*don't click check box*]
Cell line has intact p53	Yes [*click check box*]
Microarray platform	Affymetrix Human Genome U133 Plus 2.0 Array
Additional information	Dr. I. M. Dunn, National Center for Risk Analysis

b) Upload Data File

Provide test chemical information and upload microarray data files to be analyzed.

Example Values for File Upload

Data Requested	Example Value
Test type	Test chemical
Test chemical name (full name)	Benzo[a]pyrene
CASRN	50-32-8
Concentration	50 ug/ml
Graph label name	B[a]P
Treatment file	Select: Affy_2.0_Array_Treated.CEL from Example Data download link "Treatment (CEL)"

Data Requested	Example Value
Control file	Select: Affy_2.0_Array_Control.CEL from Example Data download link "Control (CEL)"

Select [SUBMIT] at the bottom of the file upload section to begin the analysis. Note: depending on the file size this may take a few minutes

c) View Results

After each data analysis, the results are displayed in a summary table (see below). Data from each analysis are shown on separate lines. The data fields included in the summary table are described on page 5 in table "Description of Results Table Fields".

Review the results in the table to ensure the study information is correct (e.g., Chemical Name, Concentration, Cell Line, etc.). Use the links in the Data and Plot Files columns to view and download the results from the analysis. The Results table includes:

- The classification of the test agent, DNA damage inducing (DDI) or Non DNA damage inducing (NDDI);
- The probability of DNA Damage (a value of '1' indicates the chemical is likely to induce DNA damage) and probability of No DNA Damage;
- Data files with Fold Change, Gene Cluster, and Chemical Cluster Euclidean distance values; and
- Images and downloads of the Heatmap and Cluster plots showing B[a]P classification in relation to the biomarker chemical training set data

5) Additional Information

a) Citing CEBS data

1) From the CEBS Support page select How do I cite CEBS under FAQs
2) Follow the instructions displayed