

HARMLESS https://www.harmless-project.eu/

Automatic workflow for in vitro high-throughput screening data FAIRification, preprocessing and scoring



G. Tancheva^{a,b}, P. Nymark^c, V. Hongisto^d, K. Patyra^d, L. Iliev^a, N. Kochev^{a,b}, R. Grafström^{c,d}, N. Jeliazkova^a

^a Ideaconsult, Bulgaria

^b University of Plovdiv, Faculty of Chemistry, Department of Analytical Chemistry and Computer Chemistry, Bulgaria ^cInstitute of Environmental Medicine, Karolinska Insitutet, Sweden

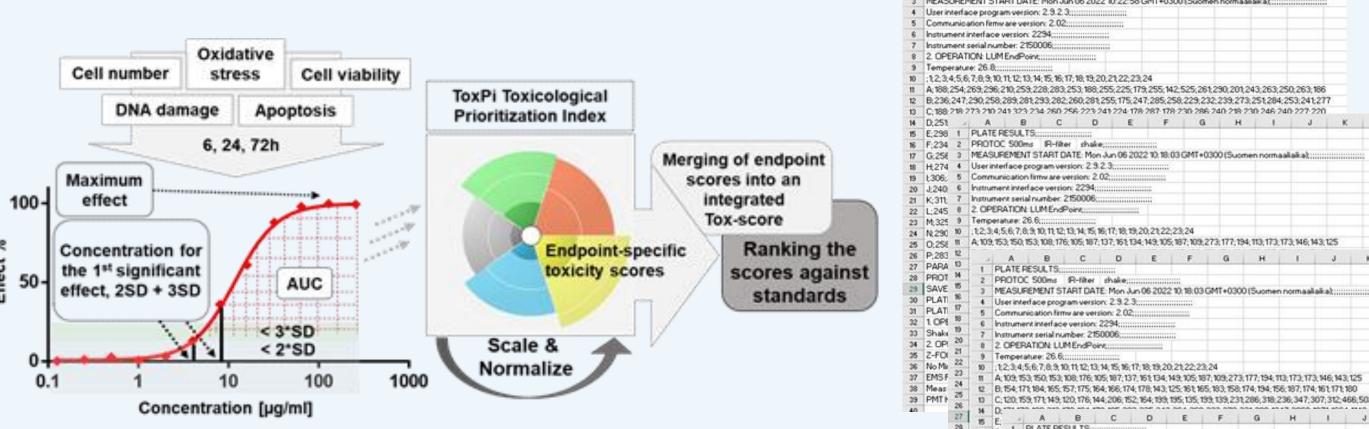
^dMisvik biology, Division of Toxicology Finland

misvik biology



BACKGROUND MOTIVATION AND OBJECTIVE

- Address safety challenges of new chemical substances and materials, including advanced nanomaterials.
- Regulatory agencies are interested in using \bullet safety data generated through New Approach Methodologies (NAM).
- Data management based on FAIR guiding principles helps with consistent curation



HTS generates vast amounts of data, which are difficult to process and analyze. Toxicological Priority Index [2] is a computational tool used in toxicology and environmental health sciences to assess and prioritize the potential health risks associated with exposure to multiple chemical stressors.

and reusing of accumulated data.

- Nanosafety, cheminformatics, and bioinformatics communities can benefit from this data management approach.
- The high-throughput screening (HTS) biological data is used for efficient clustering, ranking, prioritization of NMs and read across.

Tox5-score [1] in vitro toxicity scoring and ranking concept:

- normalization of the HTS metrics, separately in the range [0,1], for each time point and endpoint
 - combination of the normalized metric values to obtain final Tox5 endpoint scores

METHODS

Read HTS data/metadata - Orange

Read and annotate data from local directory

D:/PhD/projects/ToxPi/tox_data/vesa_files,

/PhD/projects/ToxPi/tox data/vesa files,

:/PhD/projects/ToxPi/tox_data/vesa_files

:/PhD/projects/ToxPi/tox_data/vesa_files

Set proper annotation dir

otation file

Recalculate dose

Recalculate Dose

nter well volume in µl:

ter plate growth area in cm2:

Dose per Cell Area (µg/cm²)

Dose based on Surface Area (cm²/cm²

Cell Delivered Dose (cell delivered cm²/cm

nter endpoints

g, dapi, casp

Set proper dir

16	(F;1		P 6P 16 P 6 5 6 6 1 S contraction of the second s
17	G	2	PROTOC S00ms IR-filter shake:
18	H	3	MEASUREMENT START DATE: Mon Jun 06 2022 10:18:03 GMT+0300 (Suomen normaaliaka)
19	ET	4	User interface program version: 2.9.2.3:
20	£	5	Communication firms are version: 2.02;
21	ĸ	6	Instrument interface version: 2294:
22	E.	7	Instrument serial number: 2150006;
23	M	8	2. OPERATION: LUMEndPoint.
24	N	9	Temperature: 26.6:
25	0	10	;12;3;4;5;6;7;8;9;10;11;12;13;14;15;16;17;16;19;20;21;22;23;24
28	P	11	A: 109: 153: 150: 153: 108: 176: 105: 187: 137: 161: 134: 149: 105: 187: 109: 273: 177: 194: 113: 173: 173: 146: 143: 125
27	P	12	B: 154; 171; 184; 165; 157; 175; 164; 166; 174; 178; 143; 125; 161; 165; 183; 158; 174; 194; 156; 187; 174; 161; 171; 180
28	PI	13	C; 120, 153, 171, 149, 120, 176, 144, 206, 152, 164, 139, 135, 135, 139, 133, 231, 286, 318, 236, 347, 307, 312, 466, 502
29	SI	34	D: 171; 170; 180; 213; 173; 184; 179; 195; 202; 225; 243; 264; 260; 333; 279; 331; 288; 1347; 3069; 1971; 1664; 1148; 1048; 546
30	PI	15	E, 160; 178; 169; 177; 151; 185; 144; 217; 145; 139; 151; 181; 147; 165; 143; 172; 169; 208; 163; 192; 176; 142; 184; 166
31	PL	16	F; 144; 155; 153; 136; 130; 146; 173; 180; 166; 169; 158; 153; 177; 135; 146; 161; 166; 184; 172; 174; 166; 182; 132; 163
32		17	G:140.141.191.140.165.129.169.183.139.182.172.174.138.179.171.145.172.167.167.167.166.177.185.147.172
30	s	18	H, 792; 458; 268; 217; 188; 199; 197; 178; 142; 165; 166; 142; 147; 172; 174; 183; 174; 178; 142; 164; 151; 174; 153; 138
34	2	19	L 153, 152, 175, 155, 174, 147, 186, 154, 176, 145, 158, 160, 161, 152, 185, 160, 175, 178, 163, 173, 159, 169, 170, 168
35	z.	20	J. 182, 155, 169, 148, 177, 148, 166, 180, 184, 141, 169, 158, 150, 175, 172, 174, 146, 159, 162, 170, 177, 178, 146, 164
36	N	21	K, 161, 142, 148, 154, 157, 146, 165, 145, 186, 180, 165, 168, 171, 168, 172, 136, 181, 151, 179, 184, 140, 172, 179, 151
37	EN	22	L: 144: 155: 124: 158: 142: 148: 150: 133: 157: 148: 176: 154: 157: 152: 178: 172: 149: 183: 180: 173: 156: 159: 159: 159: 166
38	M	23	M, 148, 172, 174, 154, 157, 143, 178, 177, 163, 151, 161, 160, 147, 203, 152, 183, 163, 166, 159, 160, 181, 158, 154, 156
39	PI	24	N; 145; 156; 153; 166; 156; 153; 143; 172; 170; 155; 156; 151; 160; 158; 130; 152; 148; 162; 161; 150; 147; 164; 166; 165
40	F1	25	0:153,165,160,162,166,173,150,160,150,158,132,154,153,164,152,132,158,161,136,155,176,167,172,172
		26	P; 156; 157; 171; 156; 177; 147; 148; 137; 152; 152; 173; 166; 165; 145; 148; 148; 153; 161; 144; 178; 147; 143; 164; 183
		27	PARAMETERS
		28	PROTOC S00ms IR-filter shake:
		29	SAVE TIME: Tue Apr 14 2020 13:41:20 GMT+0300 (Suomen normaaliaka)
		30	PLATE: C Flat bottom:
		31	PLATE FORMAT: 384 wells (16X24)
		32	1. OPERATION: Others Shake:
		33	Shake Time: 0 hours 0 minutes 10 seconds Shaking in normal mode Shaking in circular form
		34	2. OPERATION LUMEndPoint Top measurement.
		35	2-FOCUSImm 8.5
		36	No Mitror;
		37	EMS FILTER: IR Blocker (400-650) EMS SPOT SIZE: 2 (Default)
		38	Meastimelms: 500:
		39	PMT HV: 1000 V Disor: 47;
		40	

Challenges when using the **ToxPi** software

- requires time-consuming manual processing
- difficult to scale for larger NM datasets
- \blacktriangleright prone to errors
- challenging to implement FAIR principles

Results

A new Python module **TOX5** (tentative name) [3] has been developed for the collection and annotation of raw data, followed by normalization and the calculation of dose-response metrics. This module relies on the ToxpiR library as a key dependency and strictly the original Tox5-score follows approach. The **TOX5** module can be used as a library itself, but we also provide an Orange [4] add-on Tox5-

General steps in workflow for HTS data preprocessing and toxicity scoring

Read and annotate HTS data/metadata	Data preprocess	Dose-response parameters	Tox5-score
 Read and combine raw data from several file formats (e.gtxt, .csv) Annotate NM dose recalculation based on: Cell growth area NM surface area Cell delivered dose 	 ✓ Remove outliers ✓ Percent of control ✓ Baseline correction ✓ Clean DAPI data ✓ Normalize CASP to cell amount ✓ Mean and Median of replicates 	 ✓ Maximum effect ✓ Area under curve ✓ First significant dose (2SD, 3SD cutoff) 	 ✓ Scale data from 0 to 1 Data transformation → -log10(x)+6 ⇒ √(x) > Yeo–Johnson ✓ Slicing ✓ By individual time-point and endpoint ✓ Calculate TOX5 scores

cells

BEAS-2B

HEPG2

THP-1

A549

BEAS-2B

HEPG2

THP-1

A549

BEAS-2B

HEPG

THP-1

concentration nar

✓ The new Python module **TOX5** enables faster data preprocessing calculations and minimizes the possibility of errors through automation.

- ✓ The **TOX5** Python module and the **Tox5-scores** Orange workflow are extending the eNanoMapper FAIRification workflow [5] by facilitating FAIRification of HTS data.
- ✓ The resulting FAIR data includes both raw and interpreted data (scores) in machine readable format.
- \checkmark FAIR HTS data can be distributed as data archive and/or be integrated into the eNanoMapper database

scores [3] with user-friendly widgets for fine-tuning of the data processing.

Read data

Read raw data for each

entered endpoint and

annotate with metadata

from Template file with

View

🌣 HTS pre

Tox5 -

CTG

DAPI

CASP

≡ ?

option

doses.

data

endpoint

to recalculate

each

annotated

Tox5-scores automatic workflow

[6].

Tox5 - scores

				Tox5 Score - Orange						_		>	×	or
			~	Tox5-scores		F !	Slices			Manual slicing				for
				Select cell lines:		Se	et weight:			Load parameters				101
replicates	time	A1	_	THP-1						THP-1_24H_1st_2SD_CTG		_		Tra
54	24H	4802	496	BEAS-2B A549			Se	t weight		THP-1_6H_1st_2SD_CTG				пa
54	24H	10019	98f	HEPG2			Demous	selected slices		THP-1_72H_1st_2SD_CTG				cho
54	24H	4416	121				Remove	selected slices		BEAS-2B_24H_1st_3SD_CTG				CIIC
54	6H	12644	152				test1	test2	1	BEAS-2B_6H_1st_3SD_CTG BEAS-2B_72H_1st_3SD_CTG				
54	6H	3402	57(Select transforming functions:) BEAS-2B_24H	BEAS-2B_24H_1	BEAS-2	THP-1_24H_1st_3SD_CTG			·	spe
				1st significant dose					DEAC	THP-1_6H_1st_3SD_CTG				1
54	6H	8337	201	log10x_6 ~	 >		BEAS-2B_0H_A	BEAS-2B_6H_1s	BEA5-2	THP-1_72H_1st_3SD_CTG				sel
54	6H	9007	164	Area under curve	1	1	2 BEAS-2B_72H	BEAS-2B_72H_1	BEAS-2	BEAS-2B_24H_MAX_CTG BEAS-2B_6H_MAX_CTG				
54	72H	67286	10:	sqrt_x ~			3 THP-1_24H_AU	THP-1 24H 1ct	THP-1	BEAS-2B_72H_MAX_CTG				cal
54	72H	26575	528	Maximum dose		1	,	1111 1_2411_130		THP-1_24H_MAX_CTG				
34	72H	14113	266	log10x_6 ~		4	4 THP-1_6H_AUC	THP-1_6H_1st	THP-1	THP-1_6H_MAX_CTG				
54	72H	1763	854	Select slicing pattern			5 THP-1_72H_AU	THP-1 72H 1st	THP-1	THP-1_72H_MAX_CTG BEAS-2B_72H_AUC_DAPI				
				 Slice by time-endpoint]	BEAS-2B_24H_AUC_DAPI				
None	None	JRCNM04001a	XE	 Slice by endpoint 						Slice name:			- I F	Res
nan	nan	161016.2297818	161	Slice manually						test3				IC J
				Calculate tox5 scores	1					Create slice			5	SCO
Save as						L				Ci edite silice				
				≡ 🤋 🗗 46 46										
				1.00 m.										
										 Info 46 instances (no missing data) 		inde 1 33	ex Material 4NQO	toxpi_score 0.74319
										76 features No target variable.		2 13	MMC	0.49704
										No meta attributes.		3 37	Bleom	0.43958
										 Variables Show variable labels (if present) 	1	4 34	NANOFIL 9	0.39355
										Visualize numeric values		5 10	BENTONITE	0.35025
										Color by instance classes	-	6 17	C0304	0.34619
								_				7 21	NRCWE-058	0.33730
1	N							🖊 🖊		- Selection		8 45	XE2B	0.33503
		Data di	-Hom?	any //		ho.	x data → Ta	ble 🖊 🗸		Select full rows	9	9 30	JRCNM04001a	0.32601
All and the second		(3)34774 UU		ary a second to the		<u>1000</u>						10 3	NRCWE-056	0.31657

Create slices automatically create them manually cell selected lines. ansform with data functions, add osen ecific weight for metrics lected and Iculate Tox5 – scores.

sulting table with Tox5-

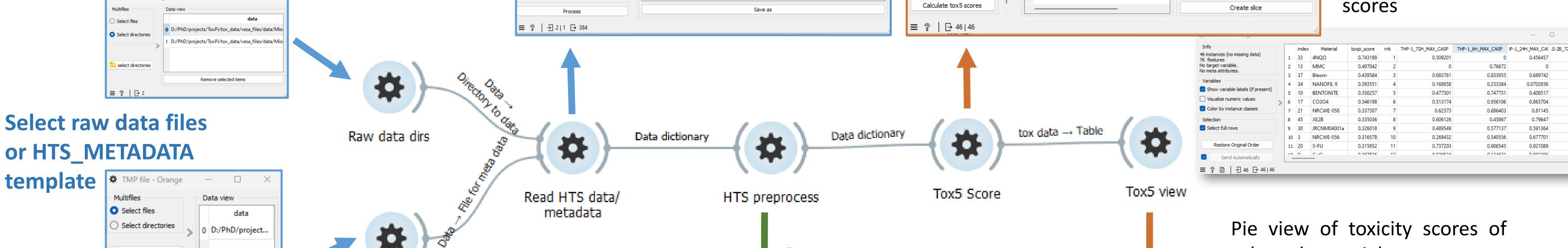
- 🗆 🗙

0.456457

0.81145

0.79647

0.591364



Raw data files

219 8.029E+1

Description Main/Cell populat

5.180E+1

A1 219 8.029E+1

A3 147 9.720E+1

7.034E+1

0.000E+0

222 9.836E+1

248 8.430E+1

5.628E+

l2 A12 265 9.234E+1 🕨 4.975E+1

A9 143 9.497E

11 A11 3 1.791E+2

3 A13 171 8.902E+1

14 A14 40 3.714E+1

15 A15 224 8.787E+1

16 A16 319 1.062E+2

10 A10 2

🌣 Raw data dirs - Orang

Description Main/Cell populat

7.034E+:

0.000E+0

5.622E+3

4.678E+3

5.735E+3

5.923E+3

6.033E+:

5.101E+3

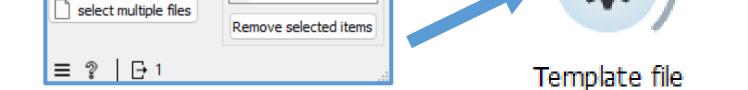
5.264E+1

1.847E+1

5 669E+1

5.068E+1

B C D E F G H I J K



HTS_METADATA template

	А	В		с	D
1	Partner				Study name
2	Operator				
3					
4			An	notation	
5	Well Address	Sample name		Sample code	Concentration (yg/ml)
6	A1	JRCNM04001a	-	1	256
7	A2	SIGMA LONG NT8 TIO2 CUBE		12	256
8	A3	TIO2 TUBE		2	256
9	A4	MKN-A15 MKN-A100		13	256
10	A5	JRCNM04001a	I.	3	256
11	A6	BIOBRAND 1 BIOBRAND 2		14	256
12	A7	NRCWE-040		4	256
13	A8	NN-ETCHED		15	256
14	A9	NRCWE-041		5	256

	А	В	С	D	E	F
1	Replicate	Time	Endpoint	Cell	Date	filename
2	S1	0h	Casp	A549	06062022	S1_0h_Casp_A549_06062022.csv
3	S1	0h	Casp	BEAS-2B	02052022	S1_0h_Casp_BEAS-2B_02052022.csv
4	S1	0h	Casp	HepG2	06062022	S1_0h_Casp_HepG2_06062022.csv
5	S1	0h	Casp	THP-1	02052022	S1_0h_Casp_THP-1_02052022.csv
6	S1	0h	CTG	A549	06062022	S1_0h_CTG_A549_06062022.csv
7	S1	0h	CTG	BEAS-2B	02052022	S1_0h_CTG_BEAS-2B_02052022.csv
8	S1	0h	CTG	HepG2	06062022	S1_0h_CTG_HepG2_06062022.csv
9	S1	0h	CTG	THP-1	02052022	S1_0h_CTG_THP-1_02052022.csv

			Data view						
	Normalizations		CTG						
	Remove outliers		Median						
	Percent of control								
	Baseline correction			A1	A2	A3			
	Normalize Casp to cell amount		A549_24H	-5.21211934073	11.055455955390393	11.32953001			
	Mean and Median of replicates		A549_6H	-6.15540169997	3.644084971886368	-1.29261214			
	Dose - response	>	A549_72H	-1.18400454644	6.24876510232507	-1.12053204			
			BEAS-2B_24H	-1.21882520491	25.956874869284373	21.57314248.			
			BEAS-2B_6H	-9.63077104406	10.241812479655716	2			
	Dose - response		BEAS-2B_72H	-2.83862204852	47.24186115569836	40.66741771.			
			HEPG2_24H	-23.3478931619	-18.09834356027081	8			
			HEPG2_6H	-19.2601374263	-6.870913078185542	-0.84564262			
Prep	rocess CASP								
View p	rocessed data			Si	ave as				

Data preprocess and dose-response

Normalize and calculate dose-response metrics for each endpoint separately. View resulting tables for normalized data, median dose response and ____ metrics, from drop-down menus.

selected materials.



1. Nymark, P; Hongisto, V et al. *Toxicology Letters, 314*, 2019, https://doi.org/10.1016/j.toxlet.2019.09.002

- 2. Marvel, S.W., et al. BMC Bioinformatics 19, 80, 2018 https://doi.org/10.1186/s12859-018-2089-2
- 3. https://github.com/ideaconsult/orange-tox5
- 4. Demsar, J et al, Journal of Machine Learning Research, 2013, 2349–2353.
- 5. Kochev, N et al. *Nanomaterials*, *10*, **2020**, <u>https://doi.org/10.3390/nano10101908</u>

6. Jeliazkova, N et al. Nat. Nanotechnol. 16, 2021, 644–654 https://doi.org/10.1038/s41565-021-00911-6

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 953183 HARMLESS.