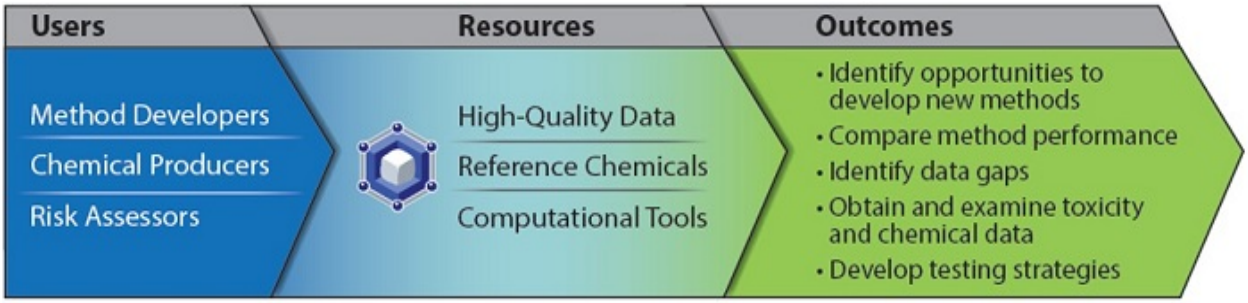
The Integrated Chemical Environment: Tools and Data to Support Toxicity Assessments

S Bell[1], J Phillips[2], N Cariello[1], P Ceger[1], X Chang[1], F Hermes[1], AL Karmaus[1], K Mansouri[1], E McAfee[2], R Rai[1], D Allen[1], W Casey[3], N Kleinstreuer[3]

[1]ILS, RTP, NC, USA; [2]Sciome, RTP, NC, USA; [3]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA

The Integrated Chemical Environment



ICE provides free online access to:

- Curated in vivo and in vitro data related to toxicity testing
- In silico toxicity predictions and chemical property data
- Curated lists of chemicals with defined assays (reference chemical lists)
- Computational tools related to chemical characterization and predicting toxicity

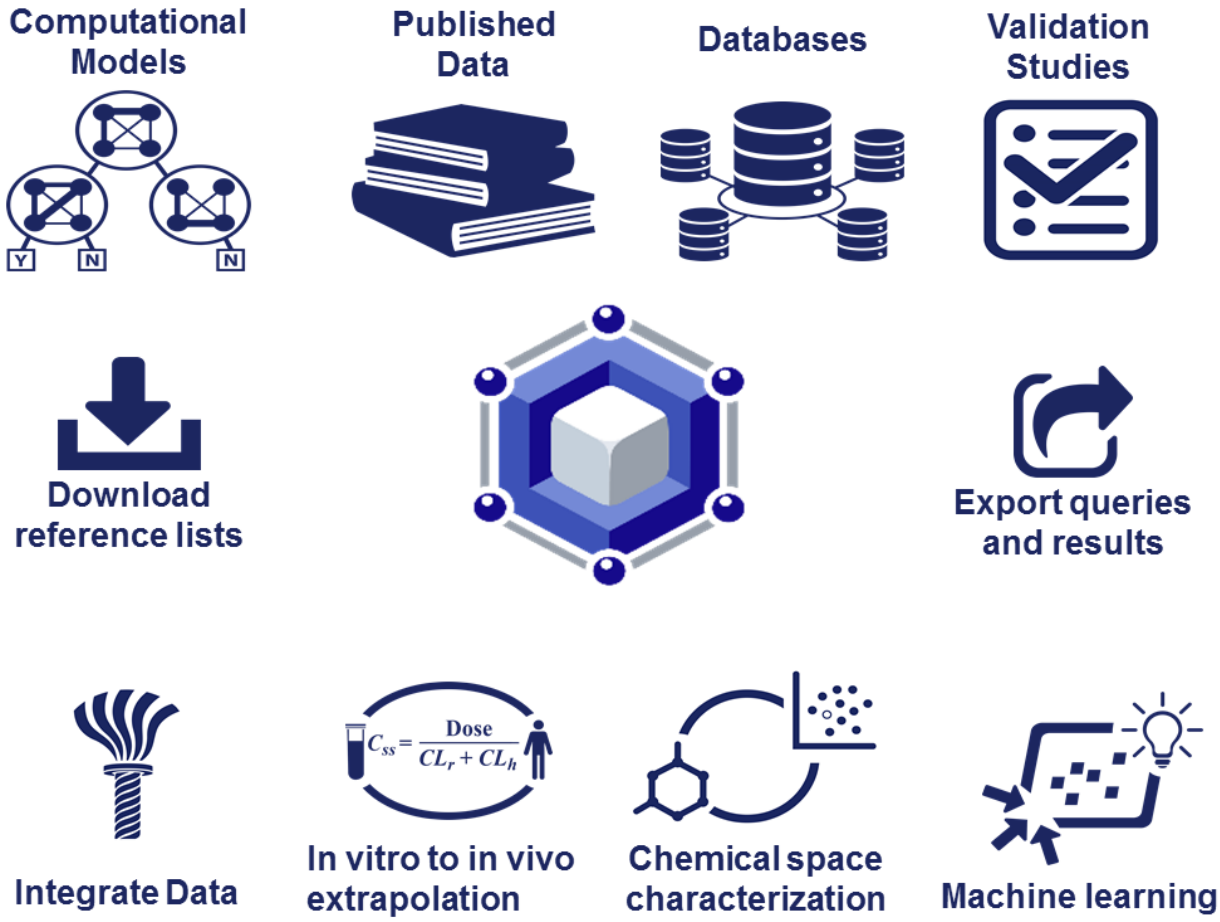
ICE supports:

- Data integration: brings together available data, including data on formulations
- Results exploration: enables dynamic, graphical exploration with publication-quality graphics
- Data analysis: allows characterization of data using online workflows
- FAIR (findable, accessible, interoperable, and reusable) data access

ICE 2.0

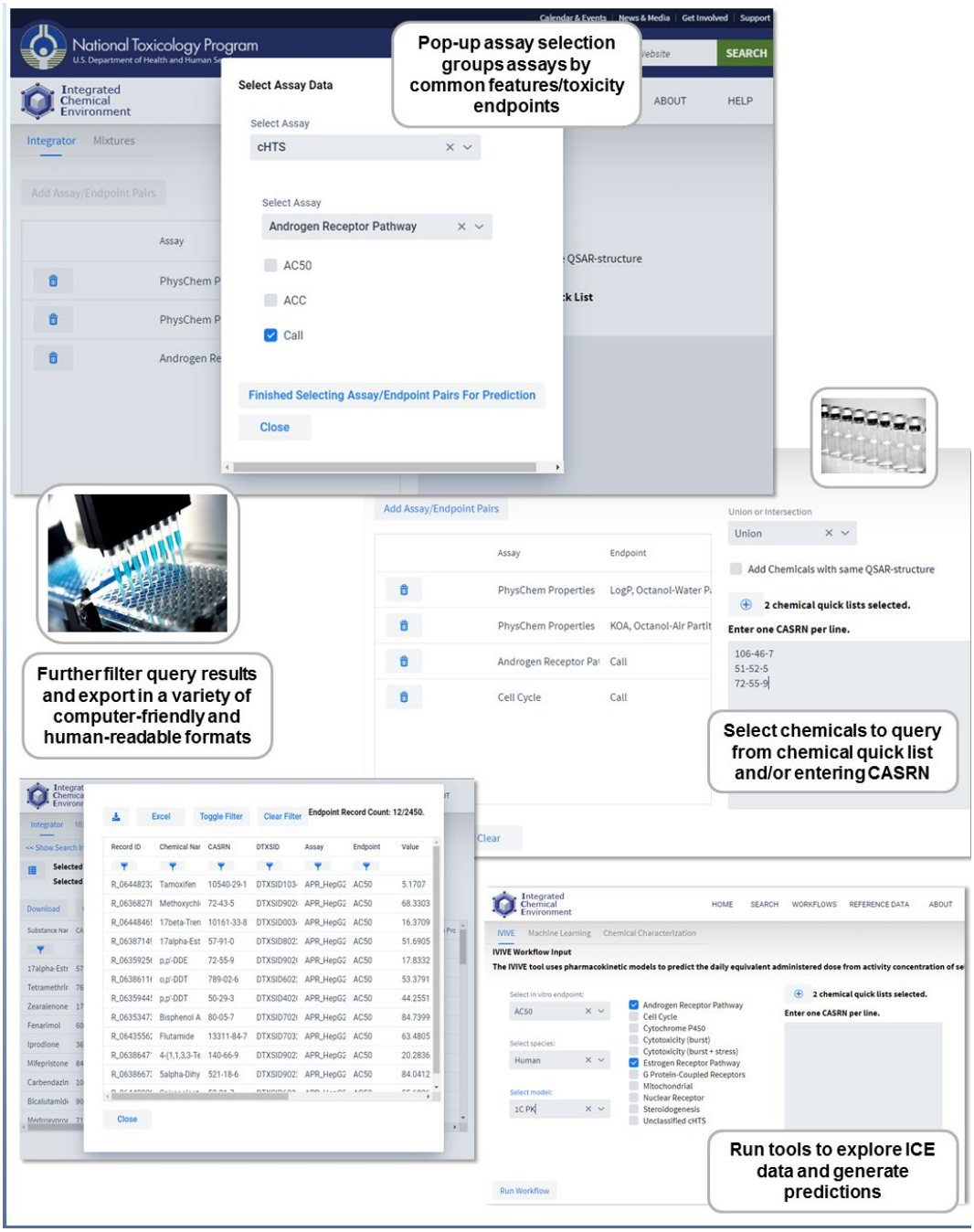
New features in ICE 2.0:

- Expand your search by adding chemicals in ICE with the same QSAR-ready structures as your chemical
- Simplified assay selection
- Updated tools

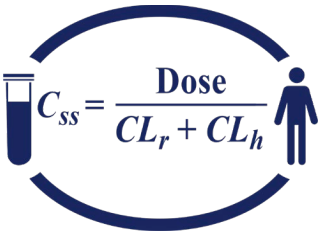


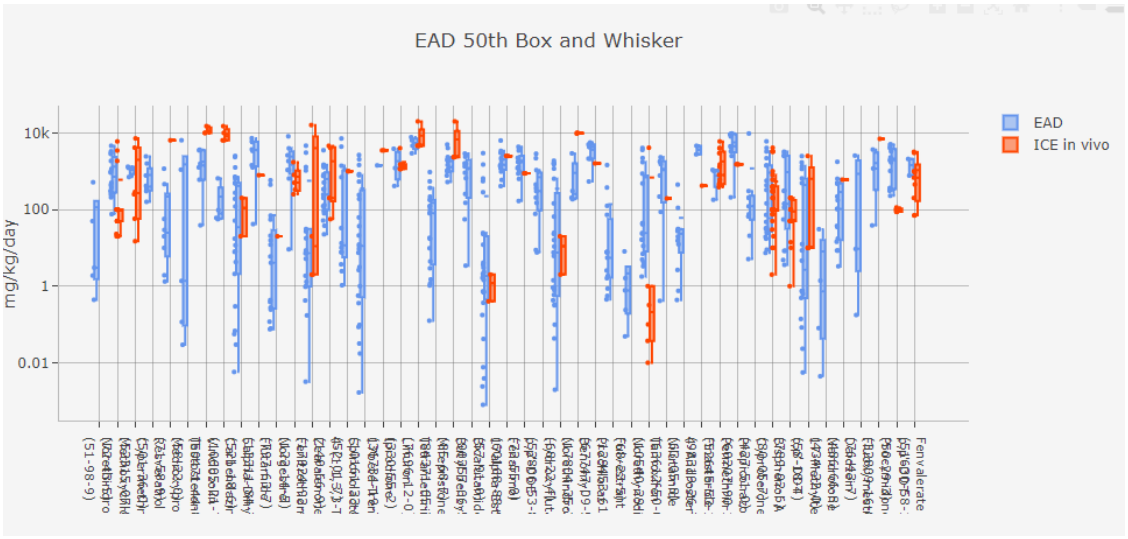
Overview of the Integrator

- Pop-up assay selection groups assays by common features/toxicity endpoints.
- Select chemicals to query from chemical quick list and/or entering CASRN.
- Further filter query results and export in a variety of computer-friendly and human-readable formats.
- Run tools to explore ICE data and generate predictions.



In Vitro to In Vivo Extrapolation



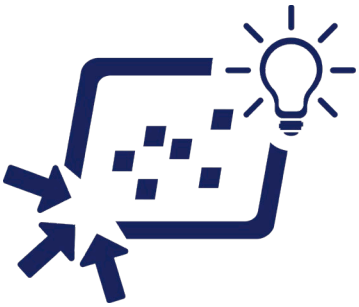


Use high-throughput in vitro data available from ICE to estimate external dose.

Choose from:

- One-compartment pharmacokinetic (1C PK) model including population simulation
- Two three-compartment physiologically based pharmacokinetic (PBPK) models:
  - 3C Glu: incorporates gut glucuronidation for BPA-family compounds
  - 3C HTTK: uses the httk package model
- Three-compartment models include gut, liver, and kidney
- Rat and human predictions
- A stand-alone version available for use with custom datasets: https://github.com/NIEHS/
- Find out more about the IVIVE workflow: Abstract 3138/Poster P886, Wednesday, March 13

Machine Learning



Use the machine learning tool for hypothesis generation and to explore different machine learning approaches using ICE data. Selected endpoints (for classification and regression modeling) and algorithms are available through ICE to facilitate the ease of use by those with limited background in computational toxicology.

Predicts endpoints for in vivo assays:

- Local lymph node assay (skin sensitization)
- Uterotrophic (estrogenic activity)
- Human skin sensitization potency

Machine learning methods available*

- cforest: conditional random forest
- rpart: recursive partitioning
- knn: k-nearest neighbor
- svmRadial: support vector machine with a radial kernel
- pls: partial least squares regression

*Machine learning tool uses imputation and/or removes sparse assays/chemicals to permit use of methods requiring complete cases.

Stand-alone version available for use with custom datasets: https://github.com/NIEHS/Machine-Learning-Pipeline

Performance statistics of the model including the confusion matrix (classification) and RMSE (regression) are available to compare method performance

```
Models: knn, rf
Metrics: ROC, Sens, Spec
Number of differences: 1
p-value adjustment: bonferroni
[1] --------------------------   knn   ----------
[1] Machine Learning Method: knn
Confusion Matrix and Statistics

          Reference
Prediction Active Inactive
  Active       61        1
  Inactive      4       38

               Accuracy : 0.9519
                 95% CI : (0.8914, 0.9842)
    No Information Rate : 0.625
    P-Value [Acc > NIR] : 4.602e-15

                  Kappa : 0.899
 Mcnemar's Test P-Value : 0.3711

            Sensitivity : 0.9385
            Specificity : 0.9744
         Pos Pred Value : 0.9839
         Neg Pred Value : 0.9048
             Prevalence : 0.6250
         Detection Rate : 0.5865
   Detection Prevalence : 0.5962
      Balanced Accuracy : 0.9564

       'Positive' Class : Active
```

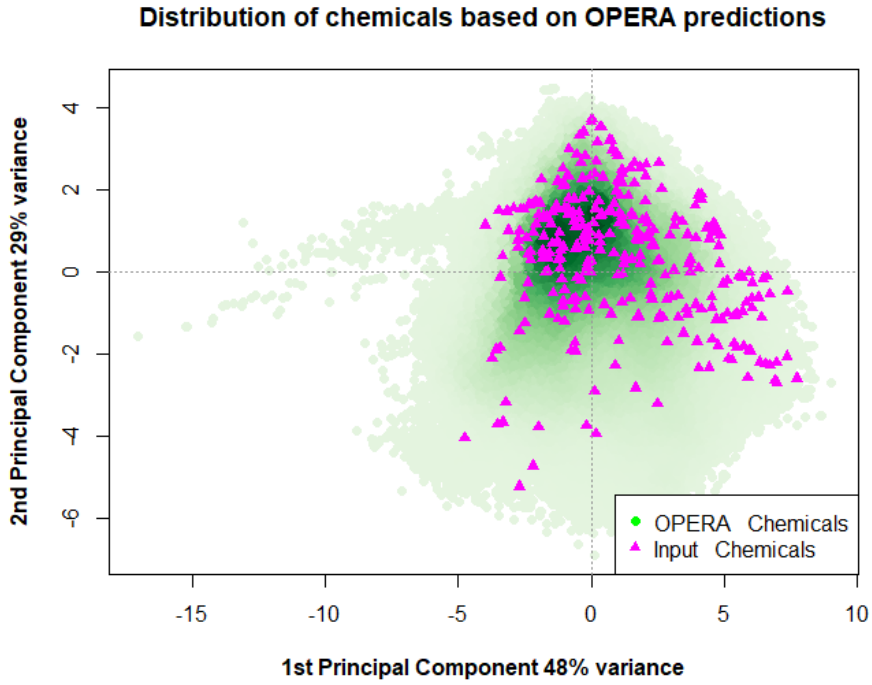
Machine learning tool outputs a table with the predictions from each model along with the data used for training the model

	A	B	C	D	E
1	CASRN	knn	svmRadial	toPredict	PhysChem
2	95-70-5	Active	Active	Active	0.355
3	26172-55-4	Active	Active	Active	1.219
4	81406-37-3	Active	Active	Active	4.539
5	119-36-8	Inactive	Inactive	Inactive	2.227
6	87-86-5	Active	Active	Active	4.78
7	862892-90	Active	Active	Active	2.604
8	531-95-3	Active	Active	Inactive	1.847
9	66422-95-5	Active	Active	Active	-0.816
10	122-57-6	Active	Active	Active	1.826
11	150114-71	Inactive	Inactive	Inactive	0.675
12	1154-59-2	Active	Active	Active	4.99
13	89-25-8	Active	Active	Active	0.929

Chemical Characterization

Leverage ICE models to characterize a user-supplied chemical list, getting information on the chemical space covered based on different physicochemical properties.

Graphical outputs highlight how representative the input chemical list (purple) is compared to the available chemical space in ICE (green).



Tabular summary comparing input chemicals to the >700,000 chemicals in ICE

```
[1] "----------------------- Summary of Input Data --------------------------------"
       MW               MP               BP               HL               VP
 Min.   : 46.04   Min.   :-109.62   Min.   : 79.62   Min.   :-10.988   Min.   :-10.213
 1st Qu.:220.06   1st Qu.:  93.95   1st Qu.:295.79   1st Qu.: -9.256   1st Qu.: -9.069
 Median :270.16   Median : 152.53   Median :343.22   Median : -7.804   Median : -7.678
 Mean   :278.46   Mean   : 145.91   Mean   :335.47   Mean   : -7.812   Mean   : -6.634
 3rd Qu.:317.59   3rd Qu.: 195.41   3rd Qu.:389.60   3rd Qu.: -6.532   3rd Qu.: -5.006
 Max.   :780.43   Max.   : 302.06   Max.   :517.60   Max.   : -4.034   Max.   :  1.548
       WS
 Min.   :-8.169
 1st Qu.:-4.338
 Median :-3.568
 Mean   :-3.162
 3rd Qu.:-2.253
 Max.   : 1.218
[1] "------------------ Summary of Available OPERA Data ----------------------------"
       MW               MP               BP               HL               VP
 Min.   :   0.0   Min.   :-177.46   Min.   :-139.2   Min.   :-12.880   Min.   :-12.744
 1st Qu.: 199.0   1st Qu.:  64.21   1st Qu.: 274.3   1st Qu.: -8.833   1st Qu.: -8.062
 Median : 253.0   Median : 120.53   Median : 314.0   Median : -7.498   Median : -5.981
 Mean   : 277.5   Mean   : 113.68   Mean   : 308.7   Mean   : -7.276   Mean   : -5.542
 3rd Qu.: 322.0   3rd Qu.: 172.31   3rd Qu.: 345.4   3rd Qu.: -6.002   3rd Qu.: -3.258
 Max.   :8435.5   Max.   : 482.74   Max.   : 541.5   Max.   :  1.374   Max.   :  4.843
 NA's   :232
       WS
 Min.   :-12.080
 1st Qu.: -4.354
 Median : -3.045
 Mean   : -3.154
 3rd Qu.: -1.834
 Max.   :  1.417
```

Future plans for Chemical Characterization tool:

- Generate physchem and other structure-based predictions from user-provided chemical lists
- Prediction of chemical parameters for use in modeling (example: fraction unbound, pka)
- Chemical use category overview provided by EPA's Consumer Products Database, cpDAT (https://www.epa.gov/chemical-research/chemical-and-products-database-cpdat)
- Integration with ChemMaps (http://www.chemmaps.com)

Contact Us

Visit ICE: https://ice.ntp.niehs.nih.gov/

To get announcements of ICE updates and other NICEATM activities, visit the NIH mailing list page for NICEATM News at https://list.nih.gov/cgi-bin/wa.exe?SUBED1=niceatm-l&A=1 and click “Subscribe.”

Acknowledgements

ICE has been funded in whole or in part with federal funds from the National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN273201500010C.

The views expressed above do not necessarily represent the official positions of any federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.