Altox
ALTERNATIVE TOXICOLOGY

2nd
PAN-AMERICAN
Conference for Alternative Methods
August 23-24, 2018
Rio de Janeiro

Best practices to develop artificial intelligence models for predicting multilevel effects in Adverse Outcome Pathways (AOP)

Rodolpho C. Braga
Altox Ltda

Senior Data Scientist,

Toxicology and Drug Discovery Specialist

rodolpho@altox.com.br

Rodolpho C. Braga
rodolpho@altox.com.br

Outline

- The “AOP-based *in silico* model” overview
- Transparency *vs* mechanistic interpretability *vs* predictivity
- AOP-based *in silico* model – Framework
  - How to ensure accuracy and mechanistic interpretability?
    - Chemical Representation / Description
    - STR continuous x categorical
    - Prediction of MIE and KEs
    - Model validation and development
    - Chemical Space and Coverage
- Final remarks

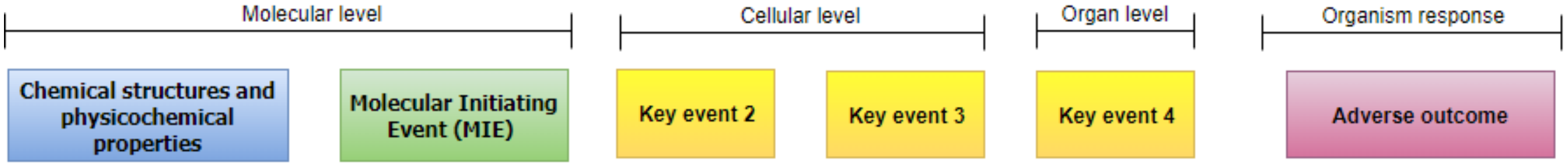
Rodolpho C. Braga
rodolpho@altox.com.br

Adverse outcome pathways (AOP) framework

An AOP is an analytical construct that describes a sequential chain of causally linked events at different levels of biological organization that lead to an adverse health or ecotoxicological effect (OECD, 2012).

A proposed "AOP-based in silico model" Concept (Altox)

An *in silico* framework used to identify chemicals that can **activate** the associated modular **AOP components** (MIE/KE) and based in these individual multilevel predictions, balanced by adjustments, relationships and weights, **to predict an adverse outcome.**



Rodolpho C. Braga
rodolpho@altox.com.br

Transparency *vs* mechanistic interpretability *vs* predictivity

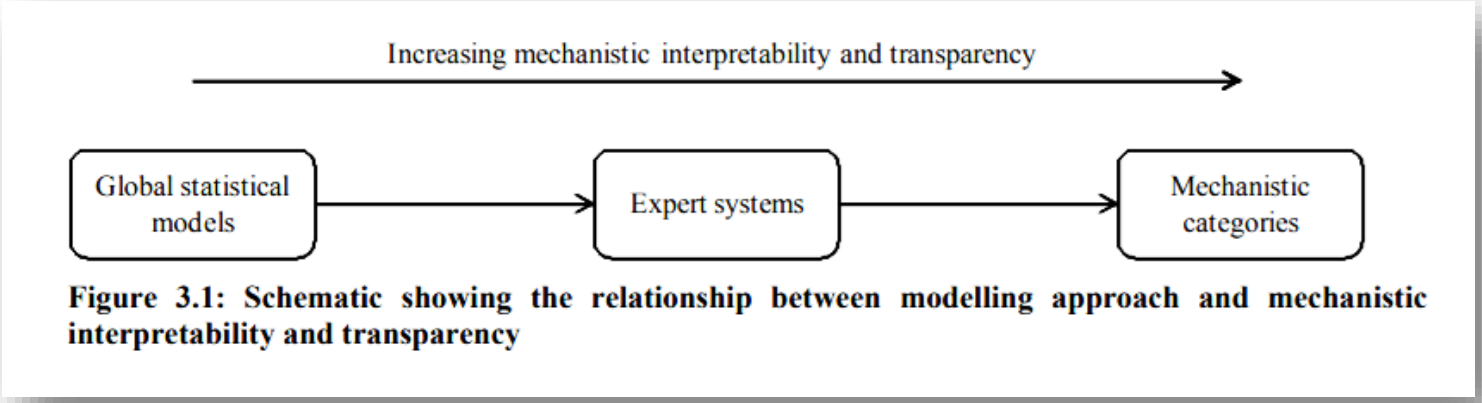


Figure 3.1: Schematic showing the relationship between modelling approach and mechanistic interpretability and transparency

Frequently discussed assumptions:

- Global statistical models
- ↑Accuracy
- ↓Transparency
- ↓Mechanistic interpretability

- Alert-based models
- ↓Accuracy
- ↑Transparency
- ↑ Mechanistic interpretability

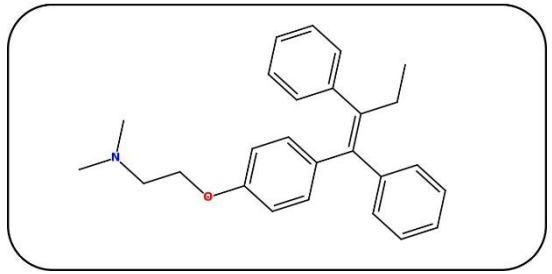
OECD - Organisation for Economic Co-operation and Development. *Report of the expert consultation on scientific and regulatory evaluation of organic chemistry mechanism-based structural alerts*. Series on Testing and Assessment, No. 120 PART 1, 2010. Available in: http://www.oecd.org/env/ehs/risk-assessment/45401393.pdf

Rodolpho C. Braga
rodolpho@altox.com.br

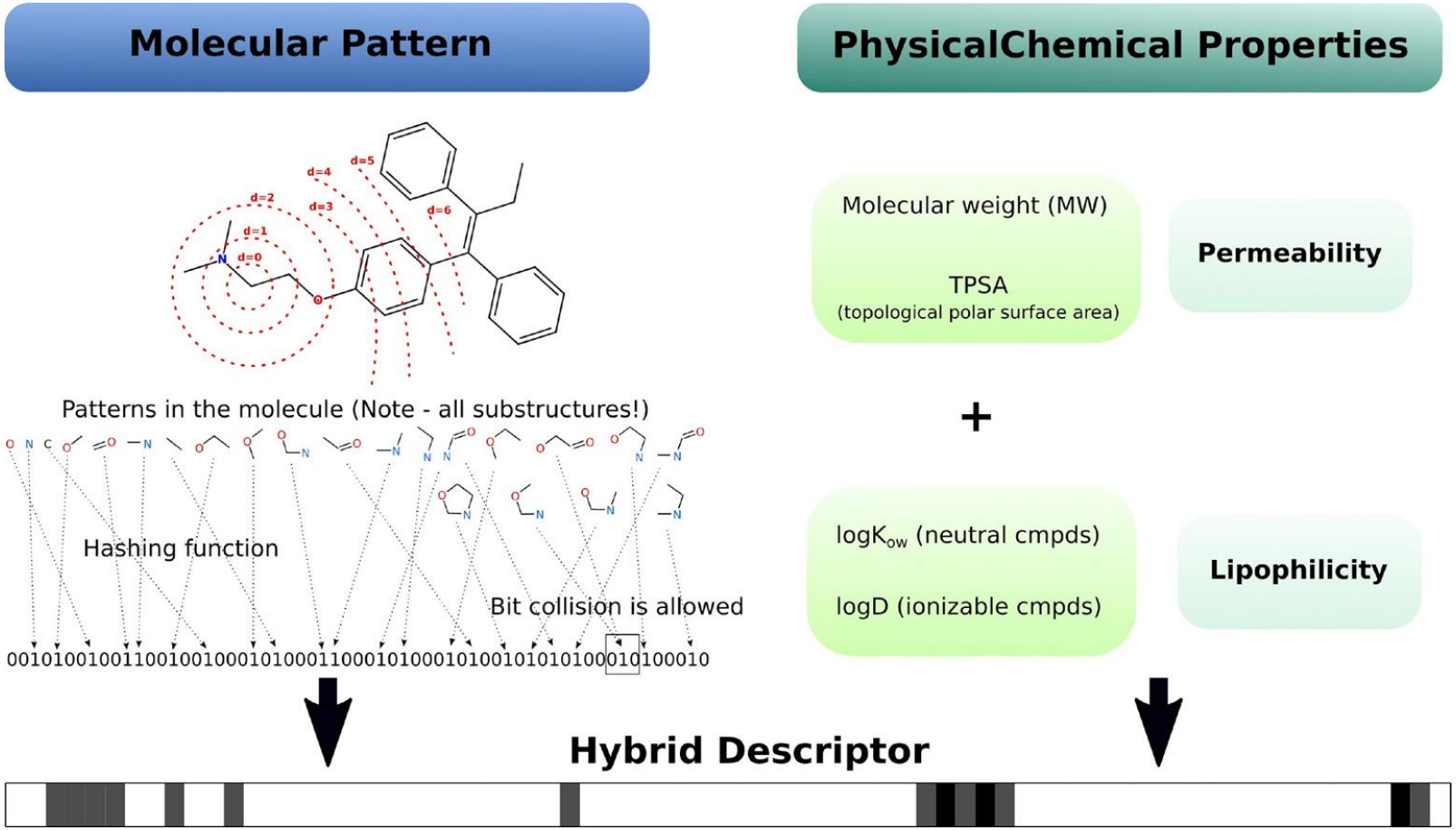
How to ensure accuracy and mechanistic interpretability?

chemical representation

Input Molecule

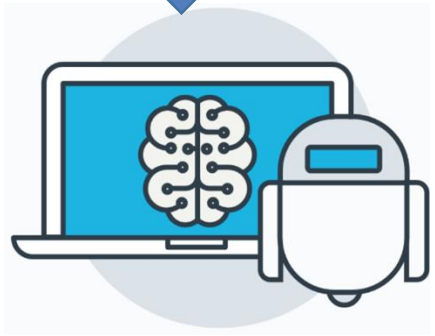


1



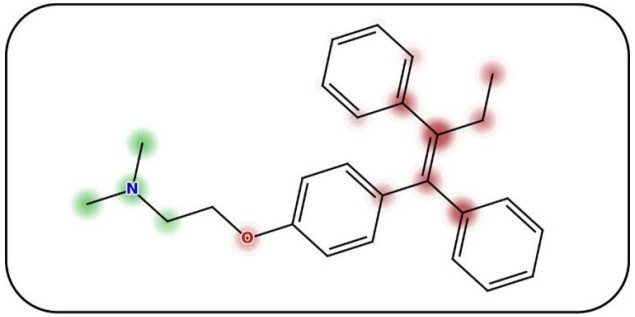
2

Artificial intelligence model



3

STR Probability Mapping



Transparency *vs* mechanistic interpretability *vs* predictivity

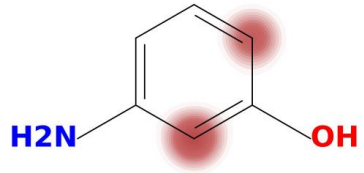
Rodolpho C. Braga
rodolpho@altox.com.br

Structure-toxicity relationship (STR)

"the fragments more related to the absence/decrease of toxicity (green) or presence/increase (red)"

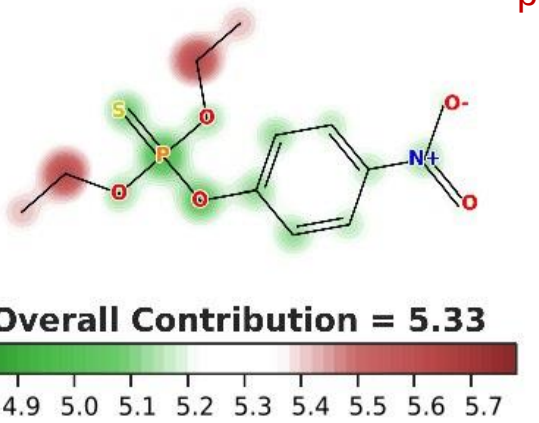
1 Categorical

AOP-Sens™

Predicted Outcome/Assay	Predicted class (Confidence)	STR Contribution Mapping
Local Lymph Node Assay (LLNA, OECD 429) Deep Learning categorical model implemented with hybrid descriptors (ECFP6 fingerprint and physicochemical properties: MW, TPSA, $logK_{ow}$, logD)	Sensitizer (+) (78.4%)	

2 Continuous

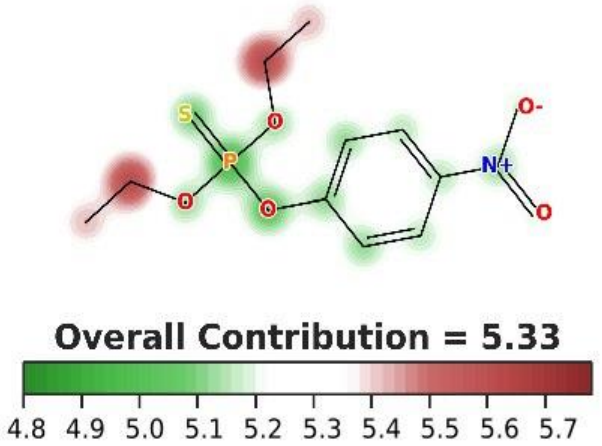
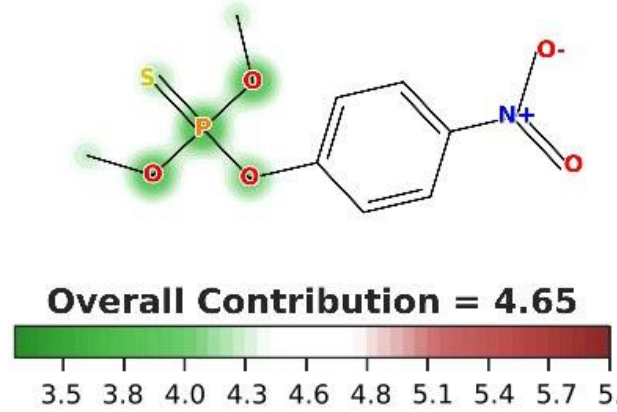
Pred-Ecotox™

Predicted endpoint/Method	Predicted Value (Confidence)	STR Contribution Mapping
LC_{50} (Fish, 96hrs) Deep Learning decision model implemented with hybrid descriptors	1.4 mg/L 4.7 µM (87.0%)	parathion 

Rodolpho C. Braga
rodolpho@altox.com.br

Structure-toxicity relationship (STR)

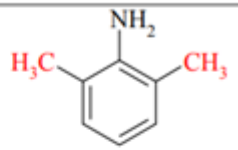
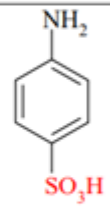
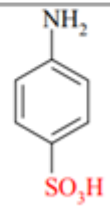
Pred-Ecotox™

	Predicted endpoint/Method	Predicted Value (Confidence)	STR Contribution Mapping
Reduce the Toxicity			
1	LC_{50} (Fish, 96hrs) Deep Learning decision model implemented with hybrid descriptors	1.4 mg/L 4.7 µM (87.0%)	parathion 
2	LC_{50} (Fish, 96hrs) Deep Learning decision model implemented with hybrid descriptors	6.0 mg/L 22.6 µM (88.0%)	parathion methyl 

How to ensure accuracy and mechanistic interpretability?

Detecting mitigating factors (steric, electronic, and detoxifying) by statistical models with a visual probability mapping

Theory

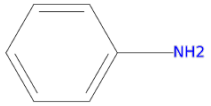
Mitigating factor	Example inactive chemical
Steric	
Electronic	
Detoxifying	

mitigating factors for aromatic amines (mitigating factor in red)

OECD - Organisation for Economic Co-operation and Development. Report of the expert consultation on scientific and regulatory evaluation of organic chemistry mechanism-based structural alerts for the Identification of DNA binding chemicals. Series on Testing and Assessment, No. 120 PART 1, 2010. Available in: http://www.oecd.org/env/ehs/risk-assessment/45401393.pdf

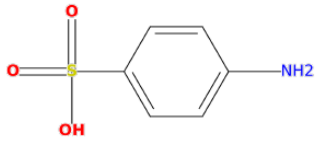
Alert model

Statistical models with probability map

1 

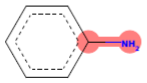
Examples: 

Genotox-iS™

2 

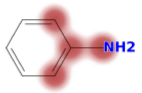
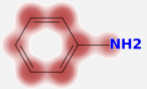
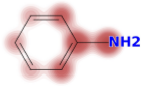
Result: (+) Positive

Alerts were found in the molecule. The results are in the table below and a description is provided at the end of the report.

Category	Alert	Alert ID	References
in vitro mutagenicity (Ames alert) alerts by ISS		Primary aromatic amine,hydroxyl amine and its derived esters	Benigni, R., Giuliani, A., Franke, R., and Gruska, A. (2000). Quantitative structure-activity relationships of mutagenic and carcinogenic aromatic amines. Chem.Revs. 100, 3697-3714. Woo, Y. T. and Lai, D. Y. (2001). Aromatic amino and nitro-amino compounds and their halogenated derivatives. In 'Patty's Toxicology. Vol. 4.' (Eds E. Bingham, B. Cohrssen, and C. H. Powell.) pp. 969-1105. (John Wiley and Sons, Inc: New York.)

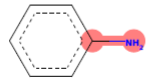
Probability Map

Non-Mutagen — Mutagen

Method	Prediction (Confidence)	Probability Mapping (SAR)
Random Forest Machine learning decision model implemented with the 2D MACCS fingerprint	Non-Mutagen (69.3%)	
kNN k-nearest neighbors decision model implemented with the 2D Extended Connectivity Fingerprint	Mutagen (71.4%)	
Deep Learning 3D Deep Learning decision model implemented with the 3D conformer fingerprint like Extended Connectivity Fingerprint	Mutagen (86.8%)	

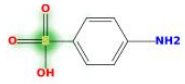
Result: (+) Positive

Alerts were found in the molecule. The results are in the table below and a description is provided at the end of the report.

Category	Alert	Alert ID	References
in vitro mutagenicity (Ames alert) alerts by ISS		Primary aromatic amine,hydroxyl amine and its derived esters	Benigni, R., Giuliani, A., Franke, R., and Gruska, A. (2000). Quantitative structure-activity relationships of mutagenic and carcinogenic aromatic amines. Chem.Revs. 100, 3697-3714. Woo, Y. T. and Lai, D. Y. (2001). Aromatic amino and nitro-amino compounds and their halogenated derivatives. In 'Patty's Toxicology. Vol. 4.' (Eds E. Bingham, B. Cohrssen, and C. H. Powell.) pp. 969-1105. (John Wiley and Sons, Inc: New York.)

Probability Map

Non-Mutagen — Mutagen

Method	Prediction (Confidence)	Probability Mapping (SAR)
Random Forest Machine learning decision model implemented with the 2D MACCS fingerprint	Non-Mutagen (98.2%)	
kNN k-nearest neighbors decision model implemented with the 2D Extended Connectivity Fingerprint	Non-Mutagen (85.7%)	
Deep Learning 3D Deep Learning decision model implemented with the 3D conformer fingerprint like Extended Connectivity Fingerprint	Non-Mutagen (95.0%)	

Rodolpho C. Braga
rodolpho@altox.com.br

AOP-based *in silico* model

Organism response

Adverse outcome

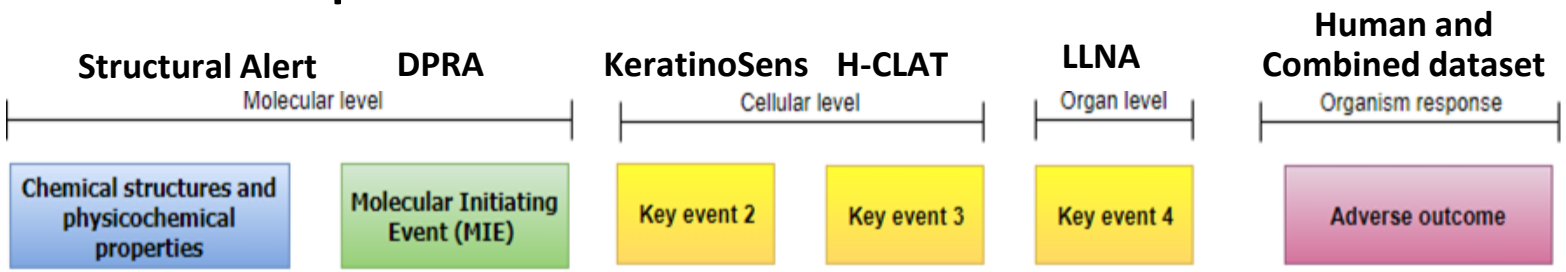
Global AOP-based *in silico* model balanced by KERs, adjustments, weight of evidence assessment in a logical framework

-To assess the results and predictivity measures in an **evidence assessment.**
- To integrate these results with *in chemico* and *in vitro* assessments, advancing in predictivity measures of an **IATA framework** for regulatory purposes.

1

Develop and Validate individual models

OECD Principles



2

key event relationships (KERs) integrating *in chemico, in vitro, ex vivo* and *in vivo* data in the *in silico* models

integrating all multilevel predictions balancing predictivity, key events relationships and WoE adjustments, and to predict an adverse outcome

Benchmark **Key event "Models" X**
AOP X

Combined dataset (GMTP, LLNA and human data)
6971 Chemicals

AOP-based algorithm = $\sum_{j=1}^{n} p_j\, c_j\, a_j\, w_j$

Benchmark - Combined dataset (GMTP, LLNA and human data)

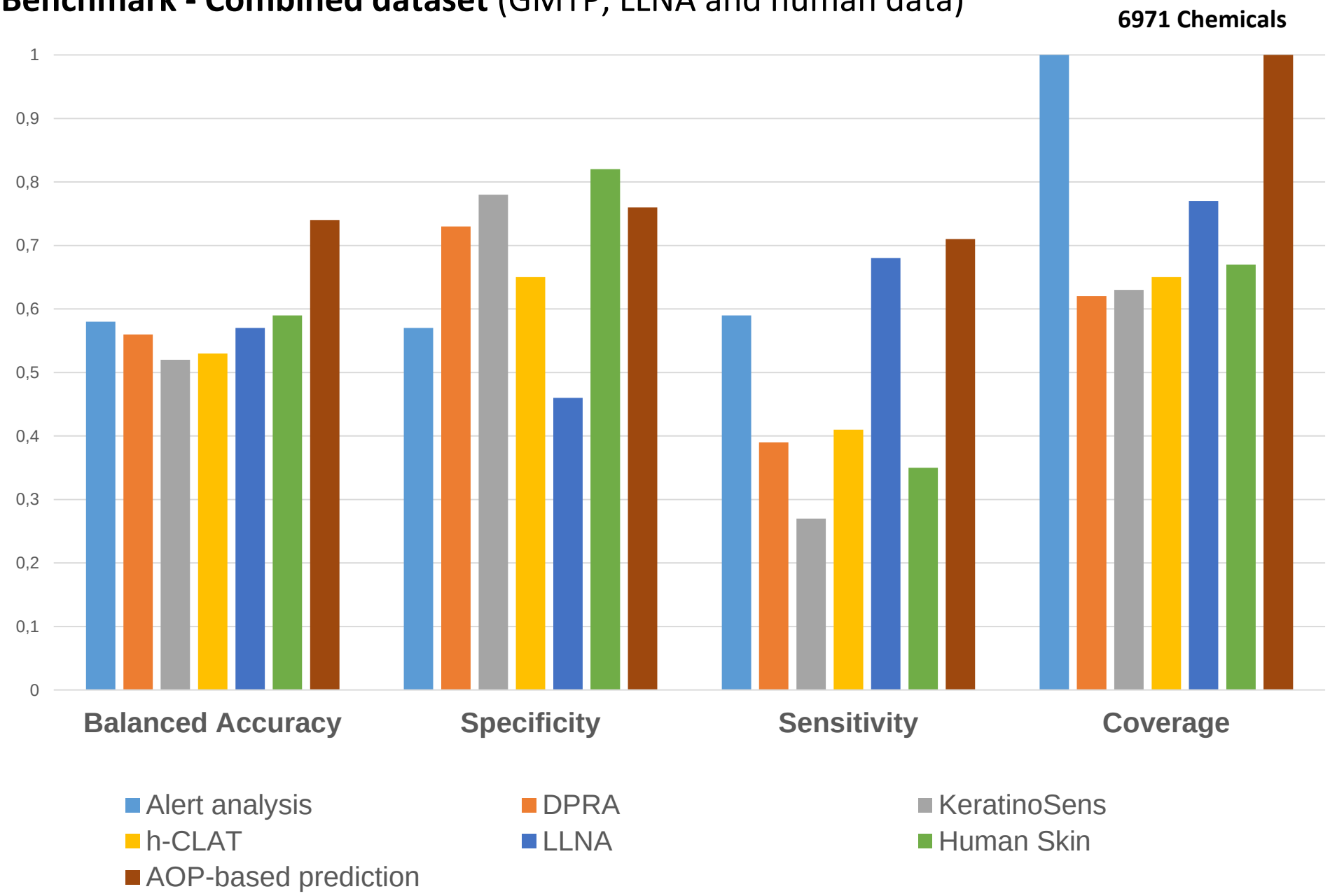
Model	Dataset	Specificity	Sensitivity	Accuracy	*Coverage*
Alert analysis	**197** 128: 1A 69: 1B	0.57	0.59	0.58	100%
DPRA	**195**	0.76	0.32	0.54	100%
DPRA AD		0.73	0.39	0.56	62%
KeratinoSens	**190**	0.74	0.27	0.51	100%
KeratinoSens AD		0.78	0.27	0.52	63%
h-CLAT	**161**	0.66	0.40	0.53	100%
h-CLAT AD		0.65	0.41	0.53	65%
LLNA	**997**	0.43	0.68	0.56	100%
LLNA AD		0.46	0.68	0.57	77%
Human Skin	**389**	0.83	0.33	0.57	100%
Human Skin AD		0.82	0.35	0.59	77%
Combined dataset (GMTP, LLNA and human data)*	**6971**	0.75	0.92	0.84	100%
Combined dataset (GMTP, LLNA and human data) AD*		0.75	0.92	0.84	93%
AOP-based prediction	-	0.76	0.71	0.74	100%

6971 Chemicals

Measures of goodness-of-fit, robustness and predictivity (External Validation)

Rodolph
rodolpho@

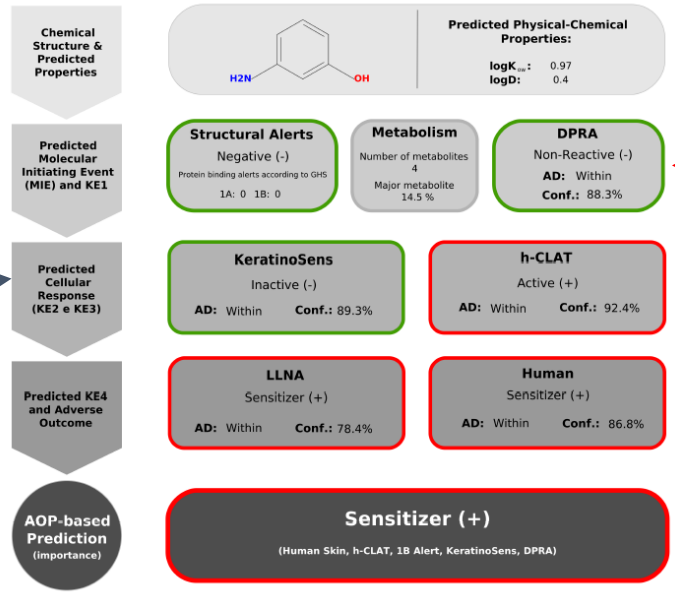
Benchmark - Combined dataset (GMTP, LLNA and human data)



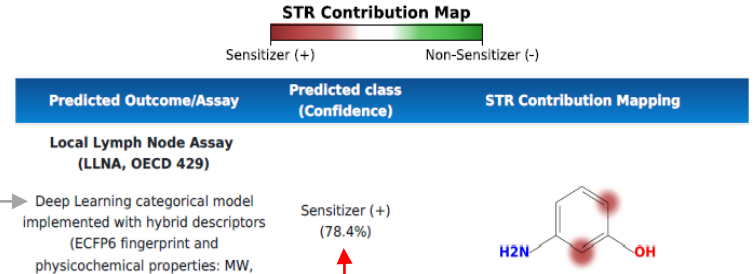
Measures of goodness-of-fit, robustness and predictivity (External Validation)

OECD Principles of (Q)SAR Validation for regulatory purposes

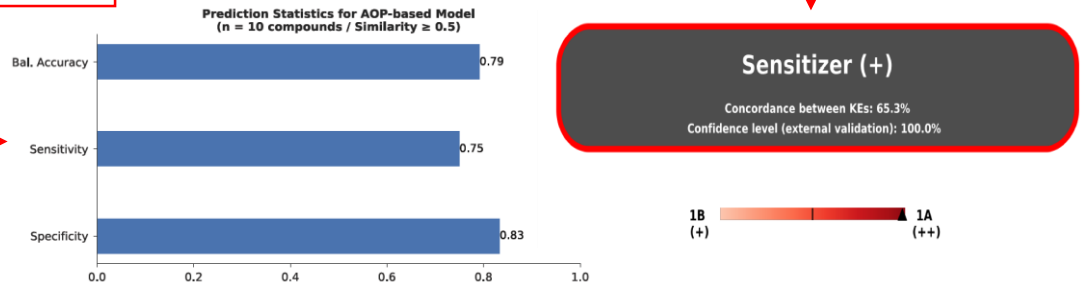
1. A defined endpoint;



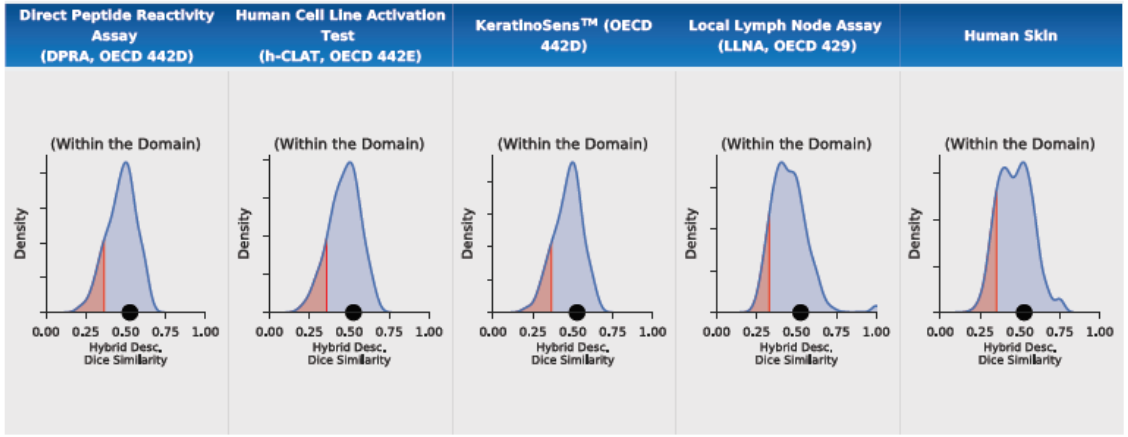
2. An unambiguous algorithm;



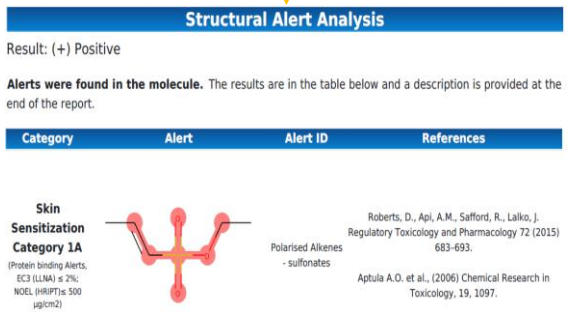
4. Appropriate measures of goodness-of-fit, robustness and predictivity;

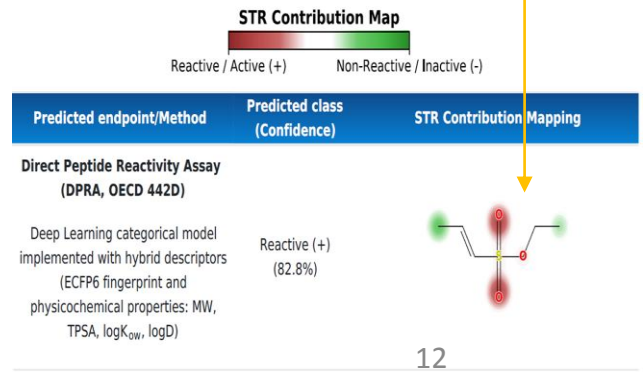


3. A defined domain of applicability;



5. A mechanistic interpretation, if possible.



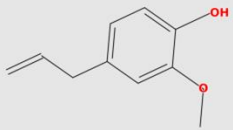


Rodolpho C. Braga
rodolpho@altox.com.br

Prehaptens and prohaptens - Activation of weak or non-sensitizing substances into sensitizers

1

Chemical Structure & Predicted Properties



Predicted Physical-Chemical Properties:

logK$_{ow}$: 2.13
logD: 1.97

Predicted Molecular Initiating Event (MIE)

Structural Alerts
Negative (-)
Protein binding alerts according to GHS
1A: 0 1B: 0

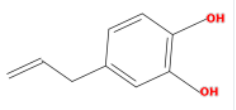
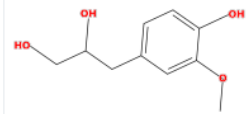
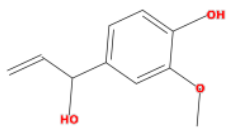
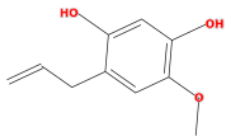
Metabolism
Number of metabolites
6
Major metabolite
27.7 %

DPRA
Reactive (+)
AD: Within
Confidence: 88.6%

2

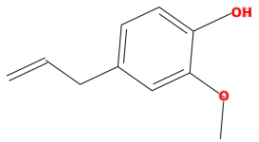
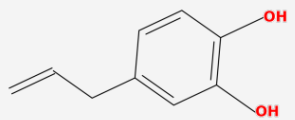
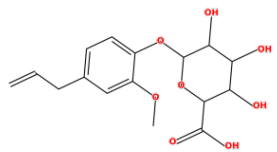
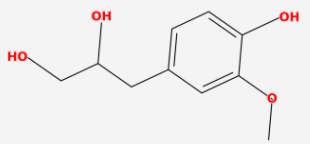
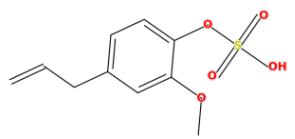
Metabolism prediction and potential for haptenation

To assess both direct and indirect haptens, this module predicts the potential for metabolic activation (pro-hapten formation) by known Phase I reactions, i.e., it can be used to identify potential skin sensitizers which require some type of metabolism to an active metabolite (pro-haptens) before initiation of the key event 1 (KE1) in a skin sensitization AOP (OECD Principle 5).

Metabolite (predicted structure)				
SMILES	C=CCc1ccc(O)c(O)c1	COc1cc(CC(O)CO)ccc1O	C=CC(O)c1ccc(O)c(OC)c1	C=CCc1cc(OC)c(O)cc1O
Reaction Rule	O-Demethylation	Vinyl Oxidation	Benzylic Hydroxylation	Aromatic Hydroxylation
Metabolite Score	27.7 %	20.0 %	7.3 %	5.6 %
AOP-based Prediction	**Sensitizer (+)**	**Sensitizer (+)**	**Sensitizer (+)**	**Sensitizer (+)**

3

Pred-CYP2D™

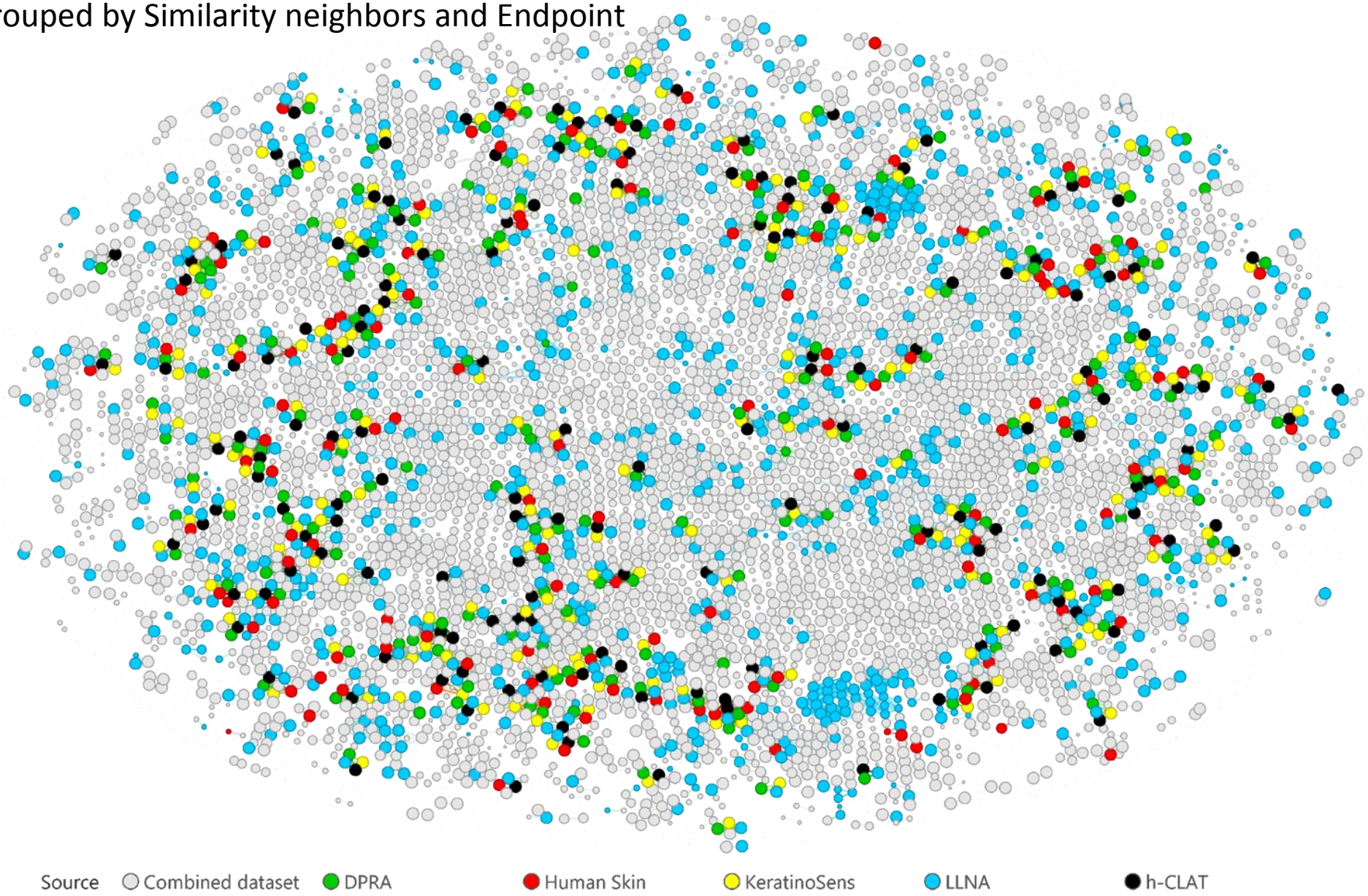
Metabolite (Chemical Name)	SMILES (Reaction Rule)	Metabolite Score	Metabolite Database	Bioassay Database
Parent Molecule (Query)  (Eugenol)	C=CCc1ccc(O)c(OC)c1	-	Human endogenous metabolite	Approved Drugs, Compounds that have been investigated (clinical trials)
 (4-(prop-2-en-1-yl) benzene-1,2-diol)	C=CCc1ccc(O)c(O)c1 (O-Demethylation)	27.7 %	Human endogenous metabolite	*in vitro* active at 10 uM
	C=CCc1ccc(OC2OC(C(=O)O)C(O)C(O)C2O)c(OC)c1 (O-Glucuronidation)	25.0 %	-	-
 (3-(4-Hydroxy-3-methoxyphenyl) -1,2-propanediol)	COc1cc(CC(O)CO)ccc1O (Vinyl Oxidation)	20.0 %	Human endogenous metabolite	Tested in man but not approved or in trials
 ([2-methoxy-4-(prop-2-en-1-yl) phenyl] oxidanesulfonic acid)	C=CCc1ccc(OS(=O)(=O)O)c(OC)c1 (Sulfation)	11.9 %	Human endogenous metabolite	*in vitro* active at 10 uM

http://ec.europa.eu/health/scientific_committees/opinions_layman/perfume-allergies/en/l-3/3-becoming-allergens.htm

Rodolpho C. Braga
rodolpho@altox.com.br

Chemical and Toxicological Space for Skin Sensitization

Grouped by Similarity neighbors and Endpoint



Diversity?

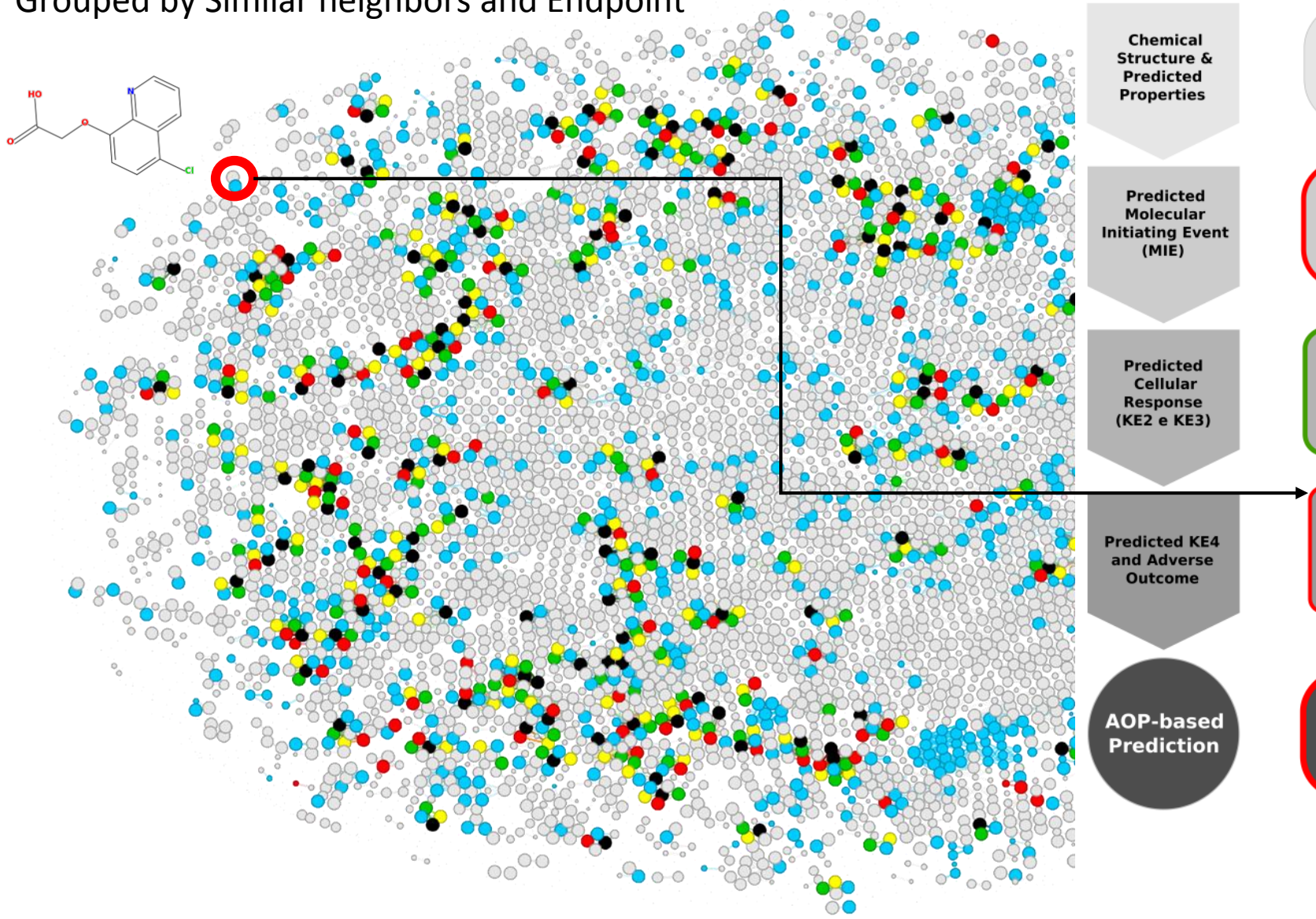
Data	
DPRA	**195**
KeratinoSens	**190**
h-CLAT	**161**
LLNA	**997**
Human Skin	**389**
Combined dataset (GMTP, LLNA and human data)*	**6971**

Rodolpho C. Braga
rodolpho@altox.com.br

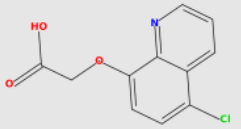
Chemical and Toxicological Space for Skin Sensitization

Grouped by Similar neighbors and Endpoint

AOP-Sens™



Chemical Structure & Predicted Properties



Predicted Physical-Chemical Properties:

$logK_{ow}$: 2.35
logD: -1.24

Predicted Molecular Initiating Event (MIE)

Structural Alerts
Positive (+)
Protein binding alerts according to GHS
1A: 0 1B: 1

Metabolism
Number of metabolites 8
Major metabolite 14.5 %

DPRA
Reactive (+)
AD: Outside
Confidence: 75.9%

Predicted Cellular Response (KE2 e KE3)

KeratinoSens
Inactive (-)
AD: Outside **Confidence:** 83.8%

h-CLAT
Active (+)
AD: Outside **Confidence:** 72.3%

Predicted KE4 and Adverse Outcome

***LLNA**
Sensitizer (+)
AD: Within
Confidence: 70.6%

Human
Non-Sensitizer (-)
AD: Outside
Confidence: 64.7%

***Combined Model**
(Human and Animal)
Sensitizer (+)
AD: Within
Confidence: 91.9%

AOP-based Prediction

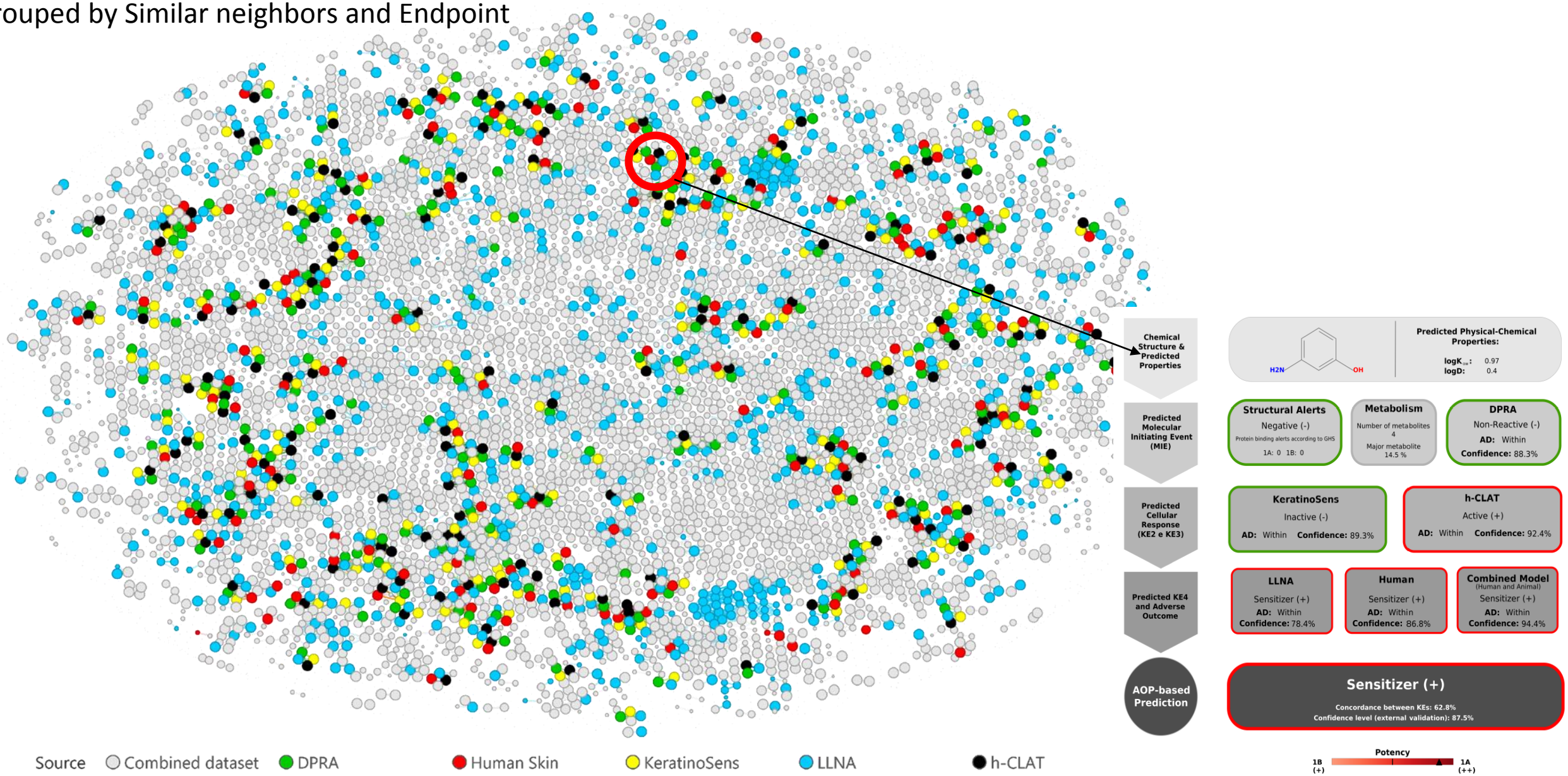
Sensitizer (+)
Concordance between KEs: 68.5%
Confidence level (external validation): 83.3%

Potency
1B (+) — 1A (++)

Rodolpho C. Braga
rodolpho@altox.com.br

Chemical and Toxicological Space for Skin Sensitization

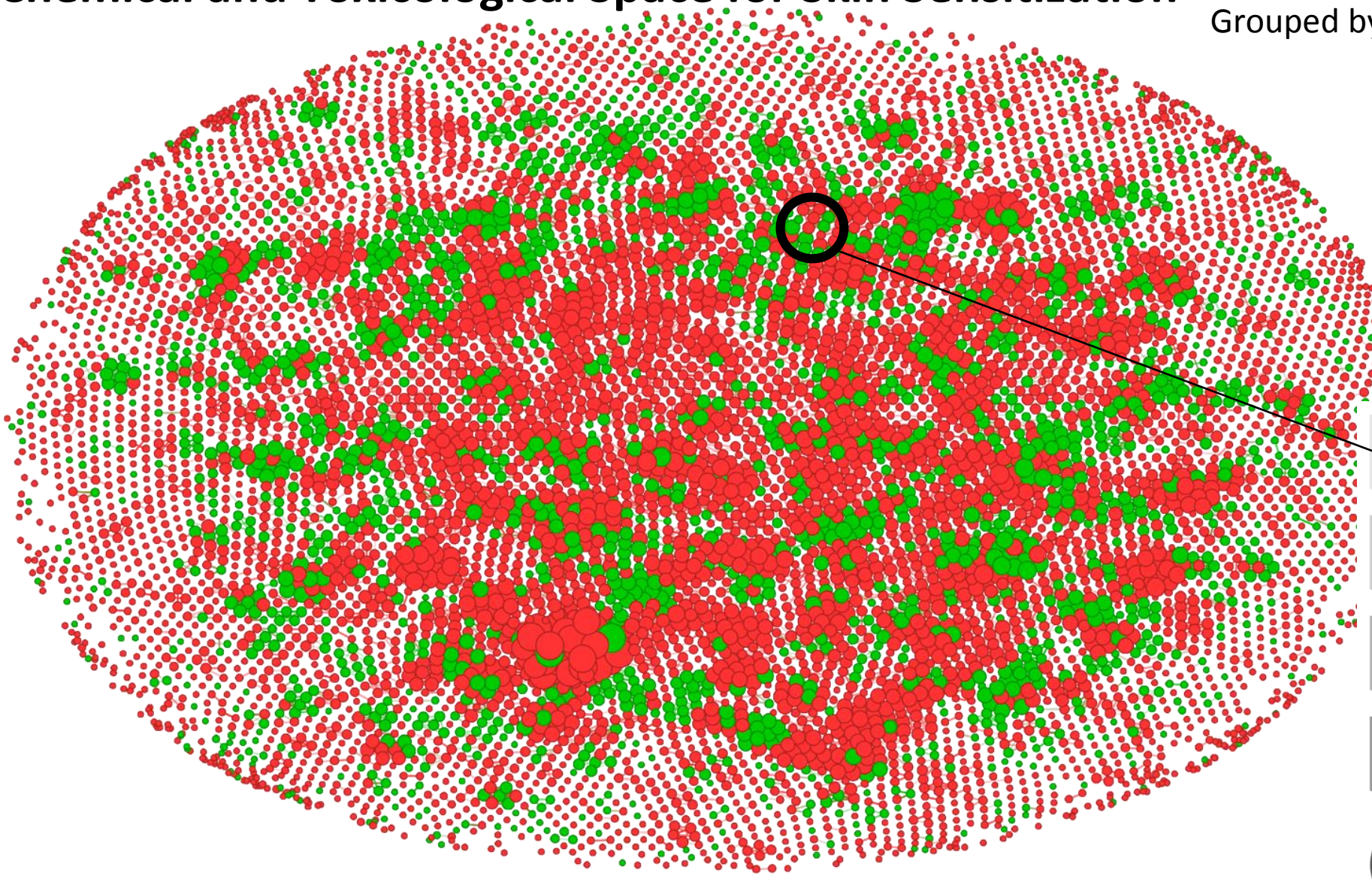
Grouped by Similar neighbors and Endpoint

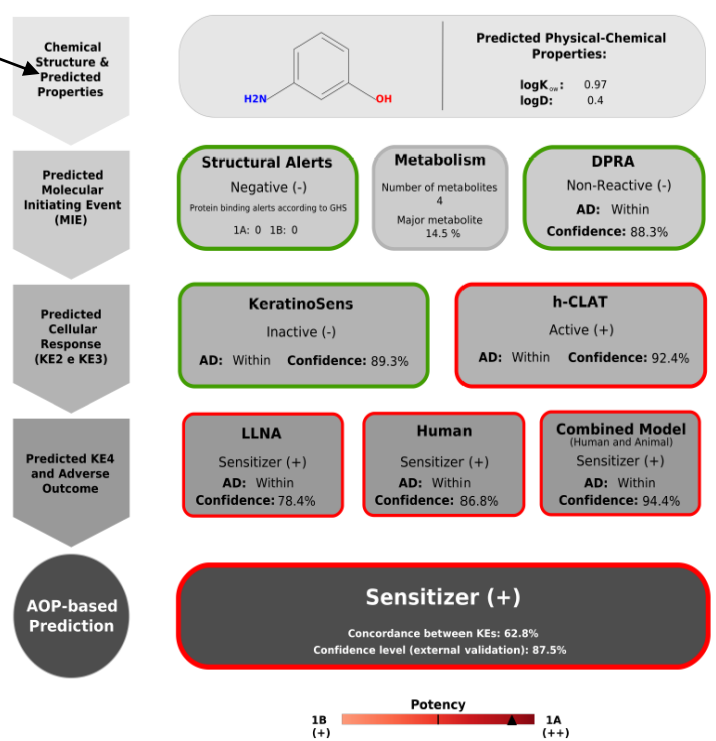


Rodolpho C. Braga
rodolpho@altox.com.br

Chemical and Toxicological Space for Skin Sensitization

Grouped by Similar neighbors and Endpoint





Rodolpho C. Braga
rodolpho@altox.com.br

Final Remarks

- A logical framework balancing transparency, mechanistic interpretability and predictivity in a sequential chain of causally linked events at different levels;

- To assess key event relationships (KERs) integrating *in chemico, in vitro, ex vivo* and *in vivo* data in the *in silico* models;

- To design new regulatory decision trees based in predictive Integrated Approaches to Testing and Assessment (IATA) containing *in silico* models ;

Thank you for your attention!

rodolpho@altox.com.br