us - eu
bridging nanoEHS research efforts
Arlington, 2-3 December 2013

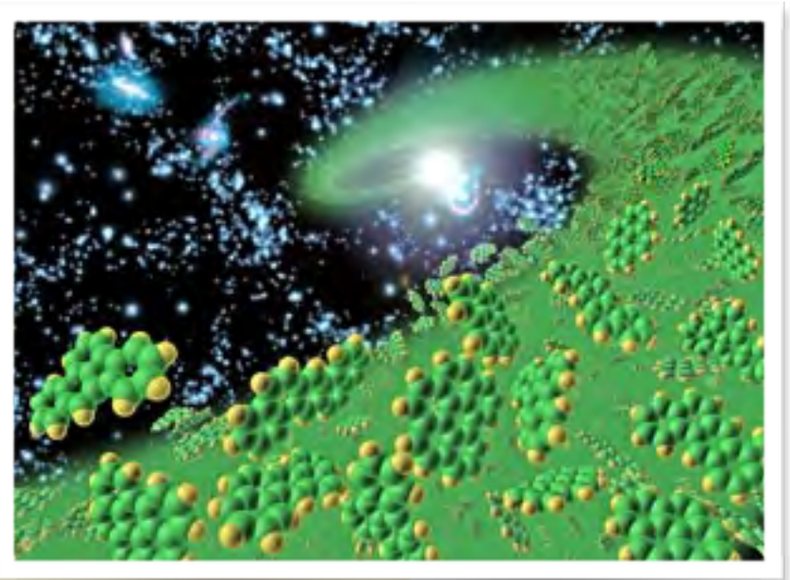
Structure-Activity Relationships for nanomaterials

Robert Rallo[1,3] and Yoram Cohen[2,3]

[1] BIOCENIT Research Lab, Universitat Rovira i Virgili, Catalunya
[2] Chemical and Biomolecular Engineering Department, UCLA, USA
[2] UC Center for Environmental Implications of Nanotechnology, USA

In silico toxicology

- Complement to in vitro and in vivo testing
- Fundamental component of Alternative Testing Strategies (ATS)
- Based on computational chemistry (and biology) principles
  - chemoinformatics, bioinformatics, computational systems biology
- (Quantitative) Structure-Activity Relationships (Q)SAR
  - chemical similarity principles
- **Challenges:**
  - Size and diversity of the chemical space
  - Regulatory use of models



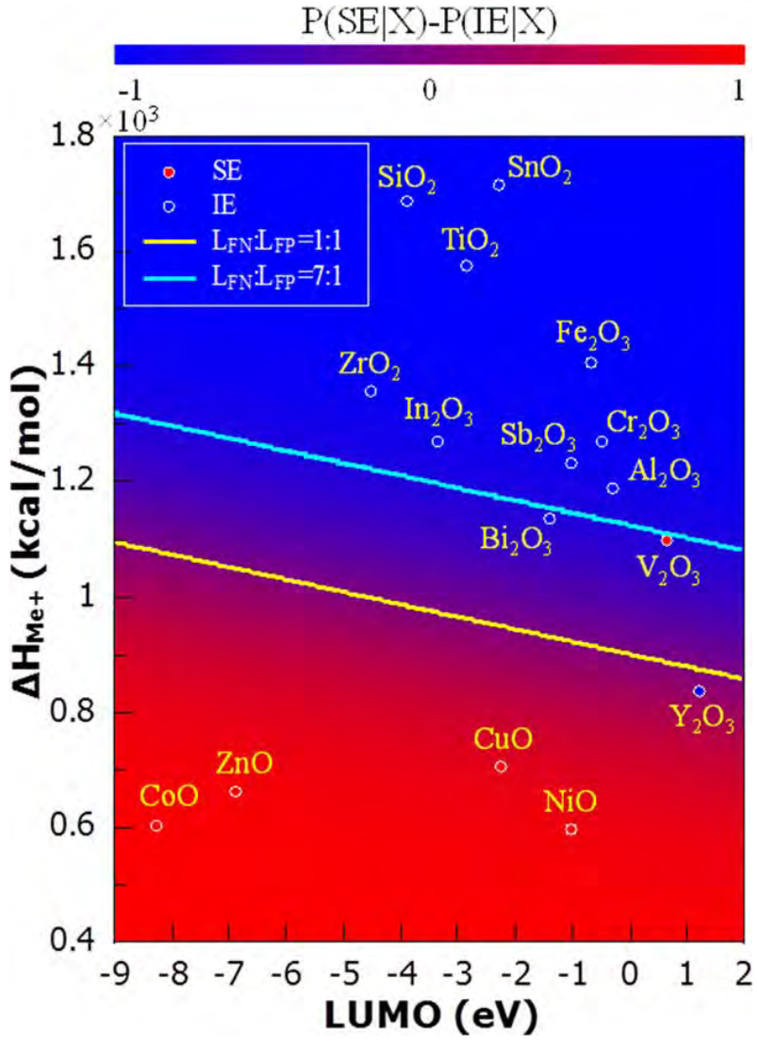
In silico (nano)toxicology

- Use of high-throughput assays for the rapid screening of NMs toxicity

- **Challenges/Needs:**
  - Common vocabulary for NMs
  - Development of nanodescriptors
  - Similarity metrics
  - Guidelines for model development and validation
  - Public, curated & quality controlled data

Development of structure-activity relationships

- Data exploration / data mining
- Identification of informative:
  - Features & samples
- Endpoint definition
- Model development
  - Performance assessment
  - Applicability domain
- Use of the model
  - Cost of errors



Cohen et al. *Accounts of Chemical Research*, 2012, **46**(3): 802-812

DATA EXPLORATION & DATA MINING

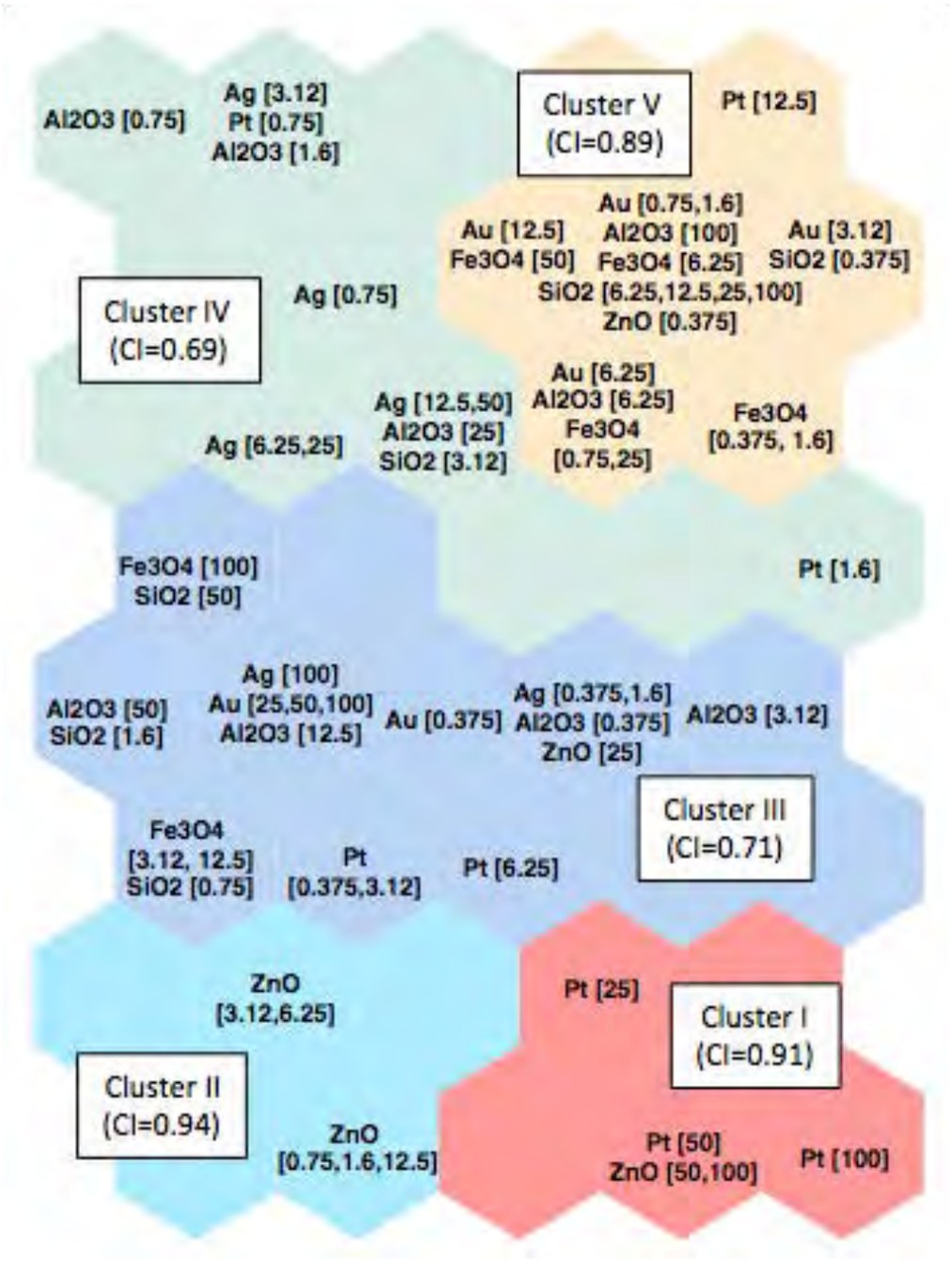
ENVIRONMENTAL Science & Technology

ARTICLE

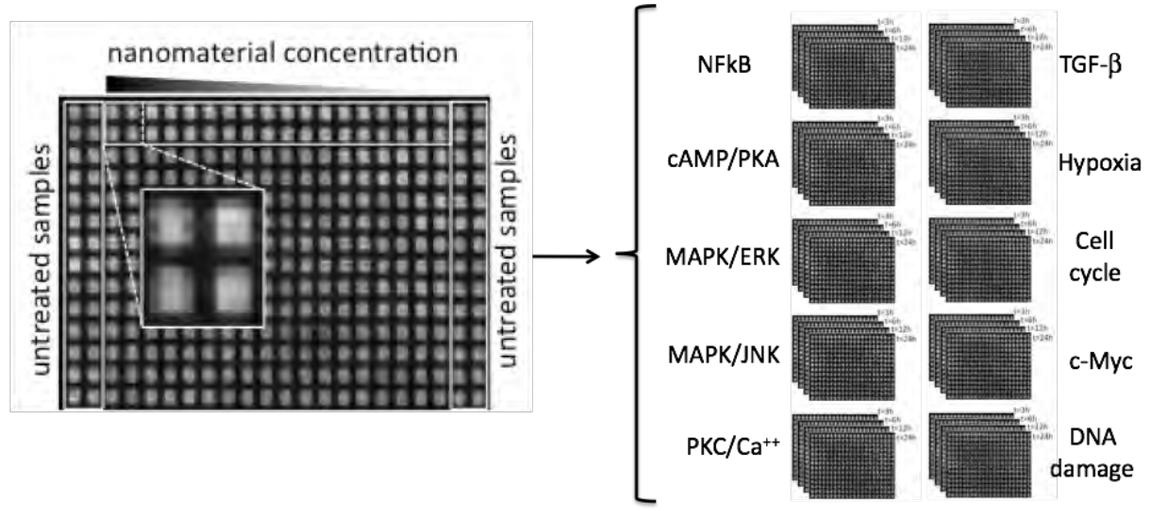
pubs.acs.org/est

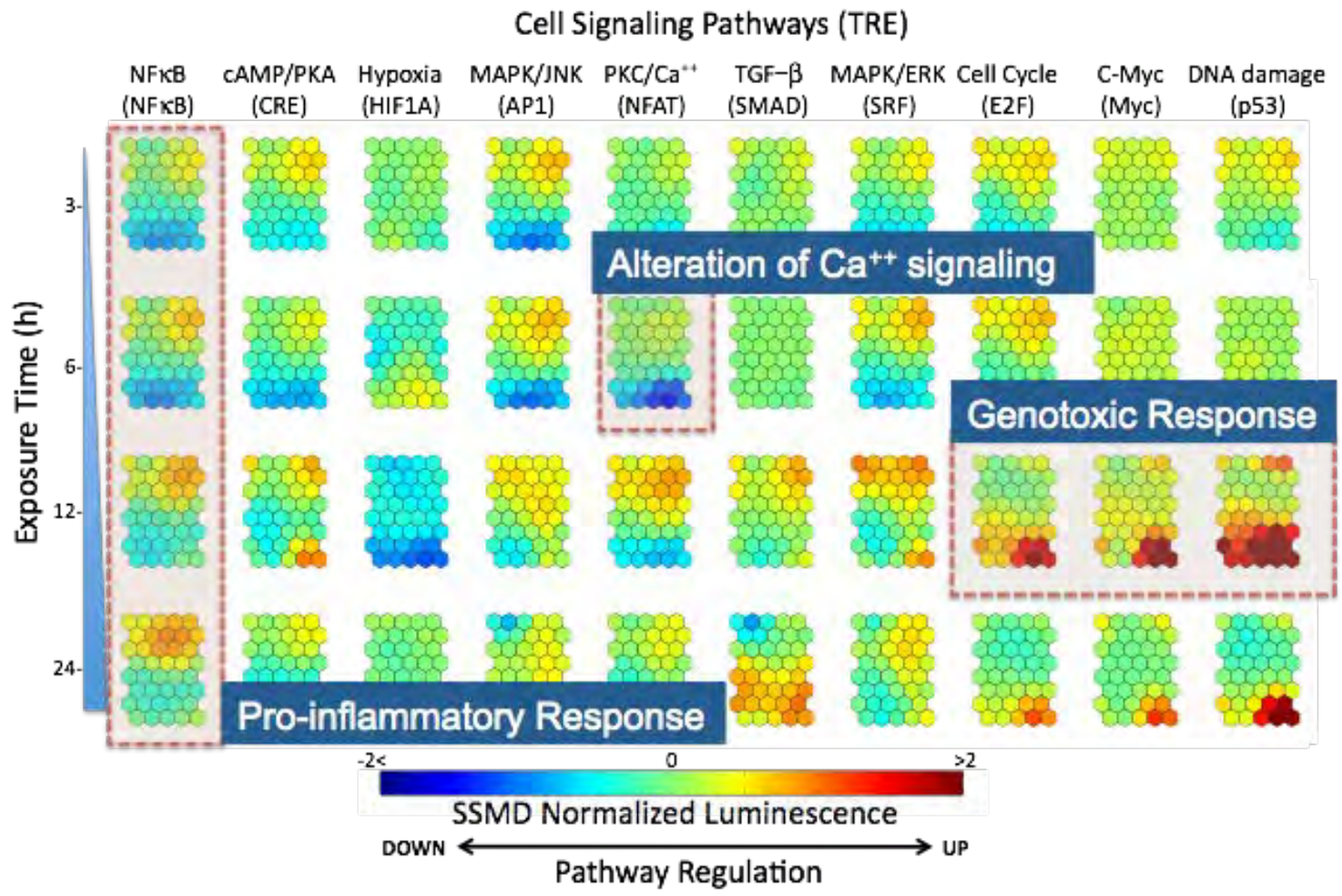
Self-Organizing Map Analysis of Toxicity-Related Cell Signaling Pathways for Metal and Metal Oxide Nanoparticles

Robert Rallo,[†,‡] Bryan France,[†] Rong Liu,[†] Sumitra Nair,[†] Saji George,[†,§] Robert Damoiseaux,[‡] Francesc Giralt,[⊥,‡] Andre Nel,[†,§] Kenneth Bradley,[†] and Yoram Cohen*[,∥,†]

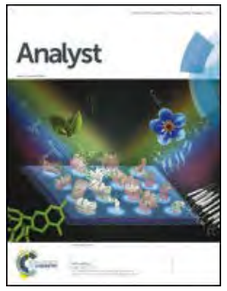


Clustering of eNPs with similar activity profiles



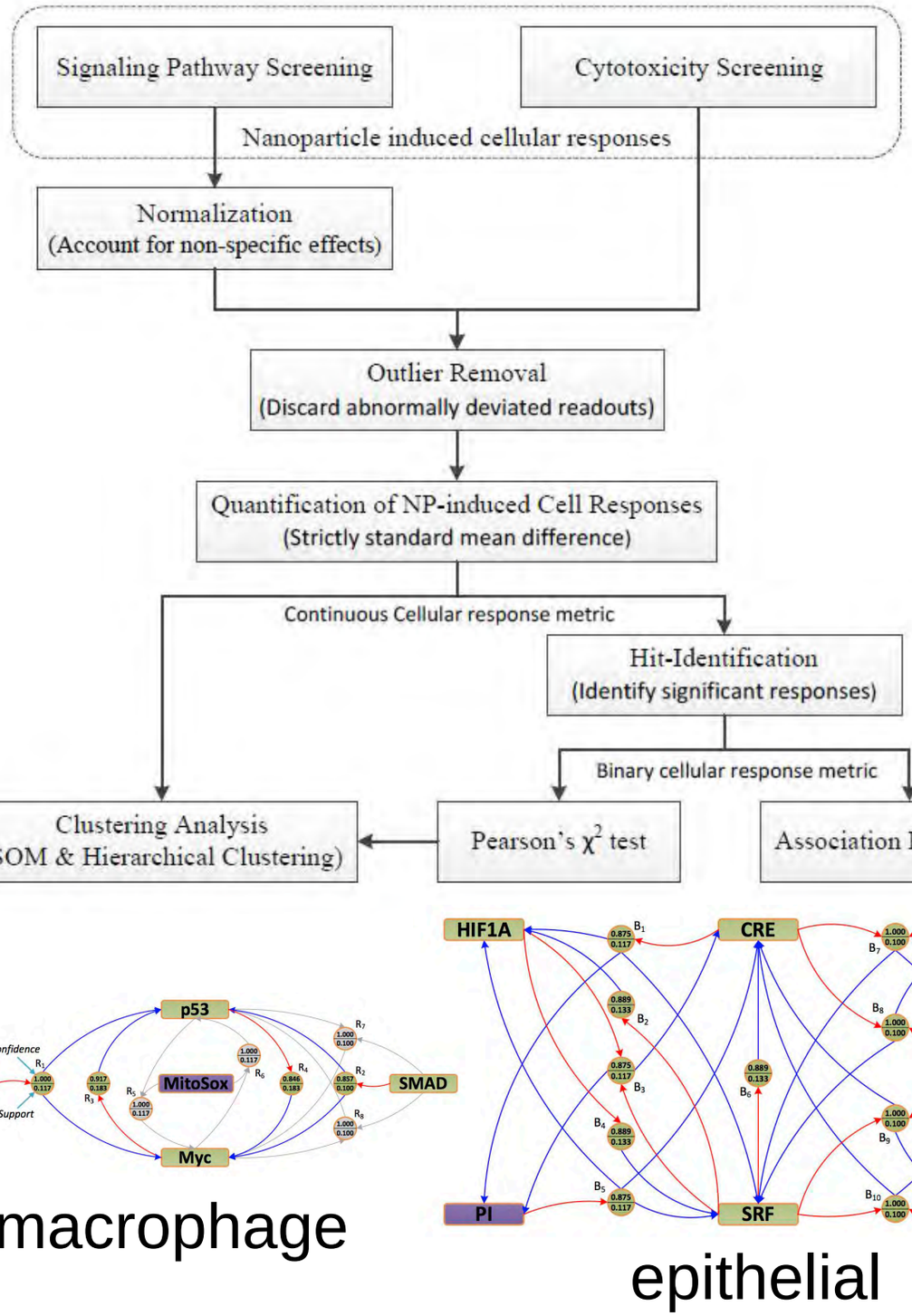


Visual data analysis



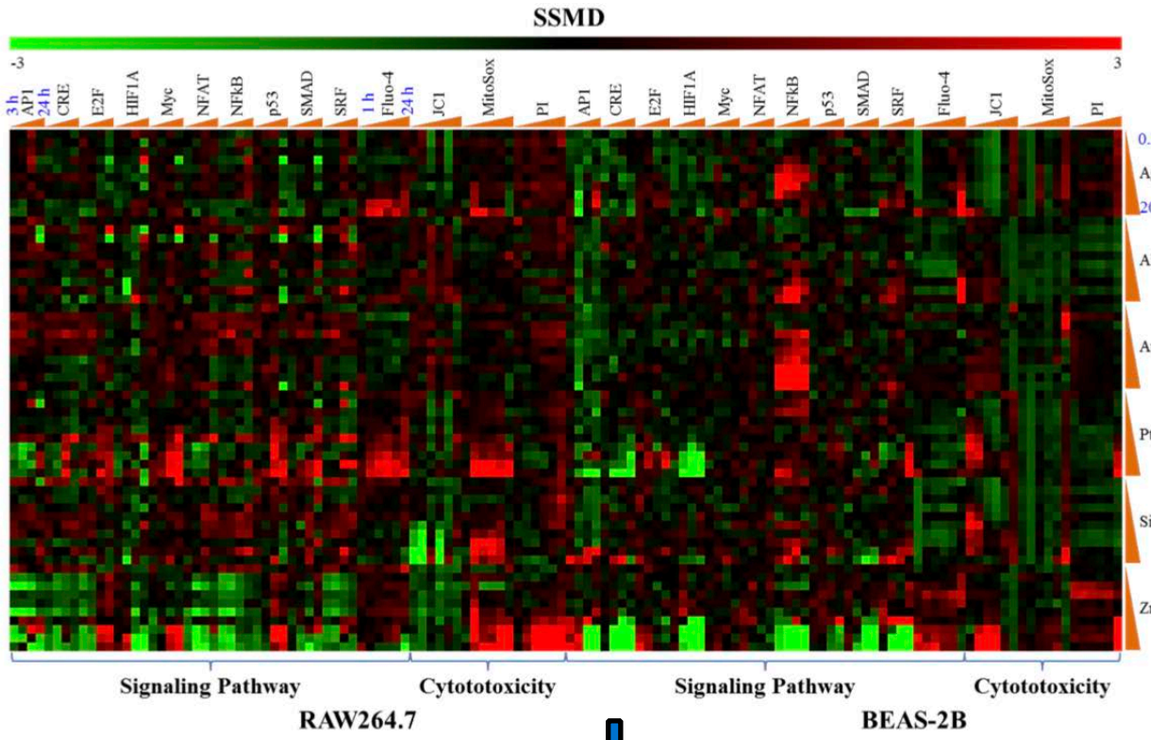
Association rule mining of cellular responses induced by metal and metal oxide nanoparticles

Rong Liu, Bryan France, Saji George, Robert Rallo, Haiyuan Zhang, Tian Xia, Andre E. Nel, Kenneth Bradley and Yoram Cohen. DOI: 10.1039/C3AN01409F

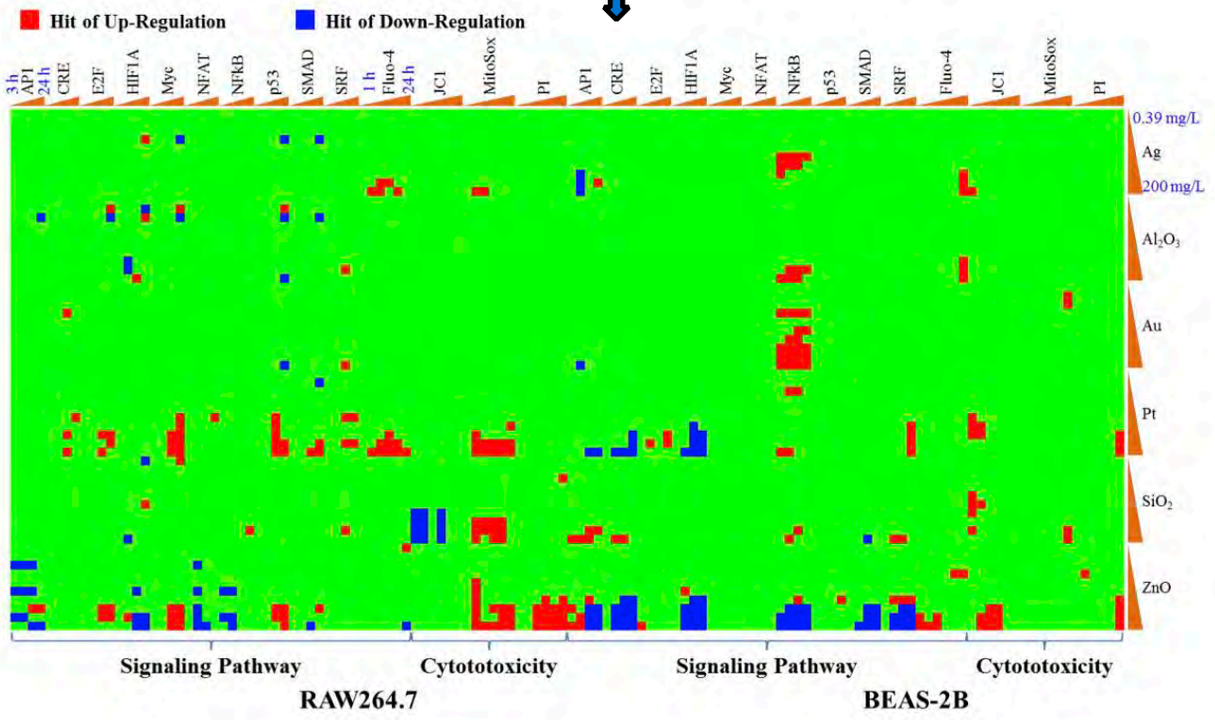


macrophage

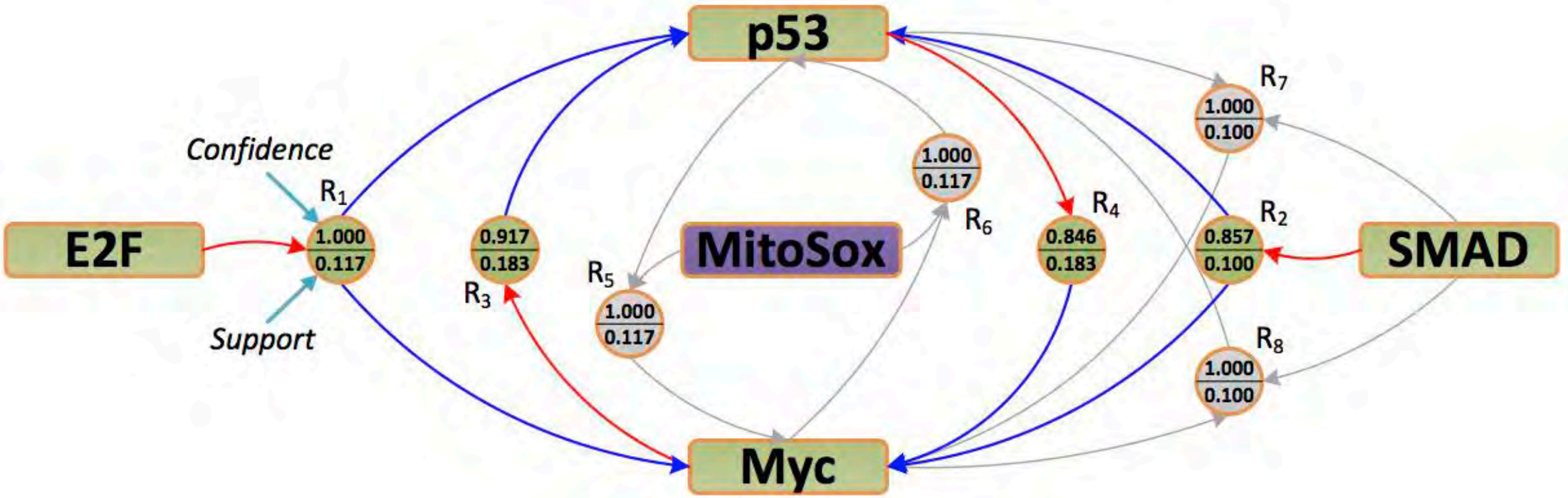
epithelial



Hit identification



Data-driven hypothesis generation

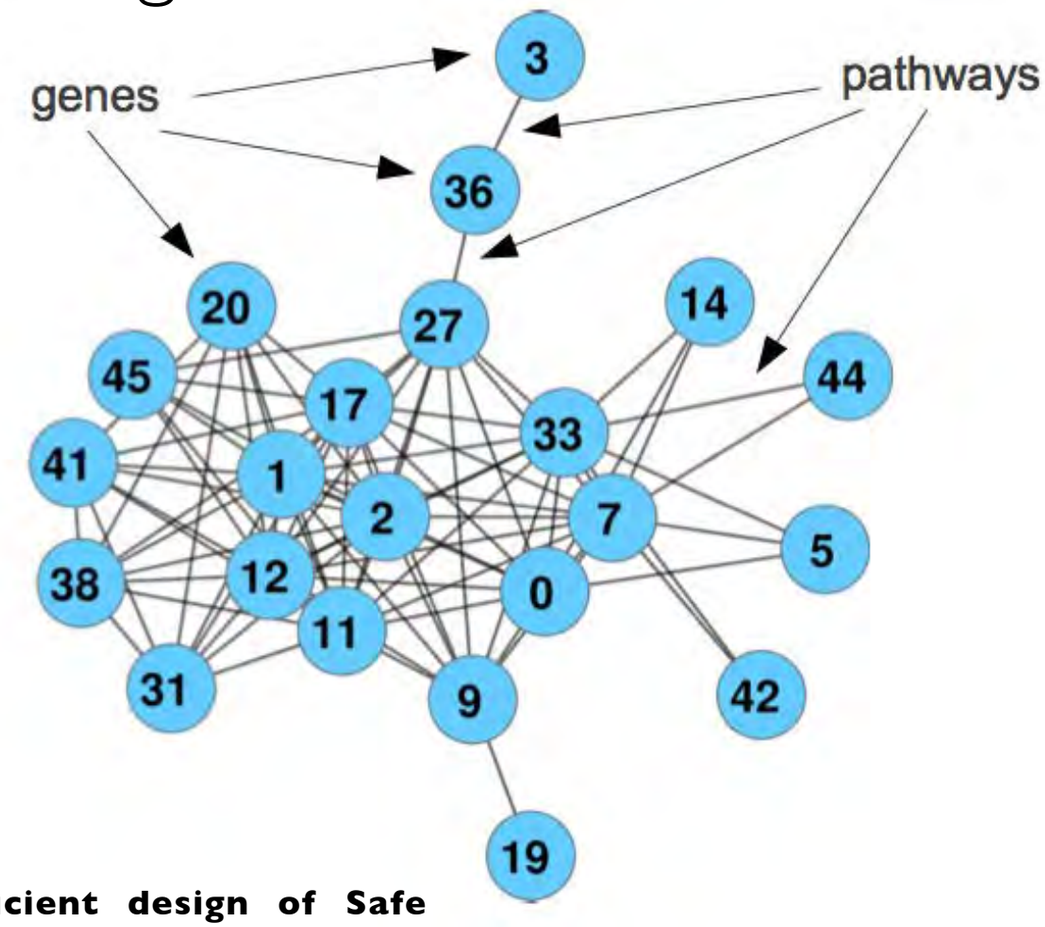


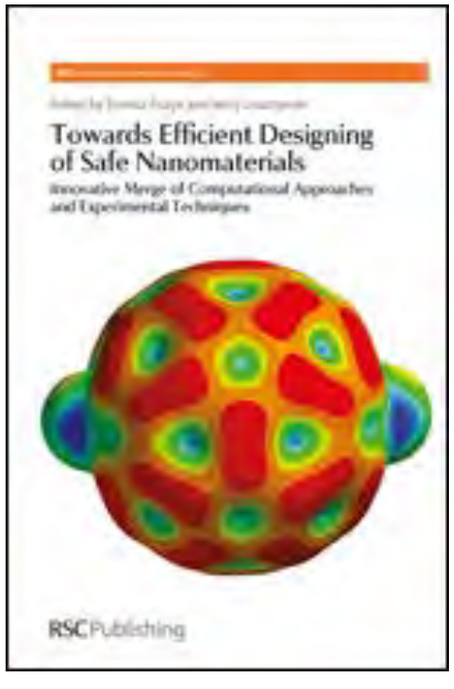
RAW264.7					
No.	*Antecedent*	*Consequent*	*Supp.*[a]	*Conf.*	Underlying NPs[b]
R_1	E2F	Myc, p53	0.117	1.000	ZnO[50-100], Pt[50-200], Al_2O_3[0.78-1.56]
R_2	SMAD	Myc, p53	0.100	0.857	ZnO[50, 200], Pt[100-200], Al_2O_3[1.56], Ag[3.13]
R_3	Myc	p53	0.183	0.917	ZnO[50-200], Pt[12.5-200], Al_2O_3[0.78-1.56], Ag[3.13]
R_4	p53	Myc	0.183	0.846	ZnO[50-200], Pt[12.5-200], Al_2O_3[0.78-1.56], Ag[3.13]
R_5	MitoSox, p53	Myc	0.117	1.000	ZnO[50-200], Pt[25-200]
R_6	MitoSx, Myc	p53	0.117	1.000	ZnO[50-200], Pt[25-200]
R_7	SMAD, p53	Myc	0.100	1.000	ZnO[50, 200], Pt[100-200], Al_2O_3[1.56], Ag[3.13]
R_8	SMAD, Myc	p53	0.100	1.000	ZnO[50, 200], Pt[100-200], Al_2O_3[1.56], Ag[3.13]

Identification of cause-effect relationships via confirmatory experiments

Computational Systems Biology

- network analysis of gene differential co-expression
- based on the complete genome information
- takes into account pathway cross-talking
- comparison with null case to compute statistical significance
- robust to noise in biological data



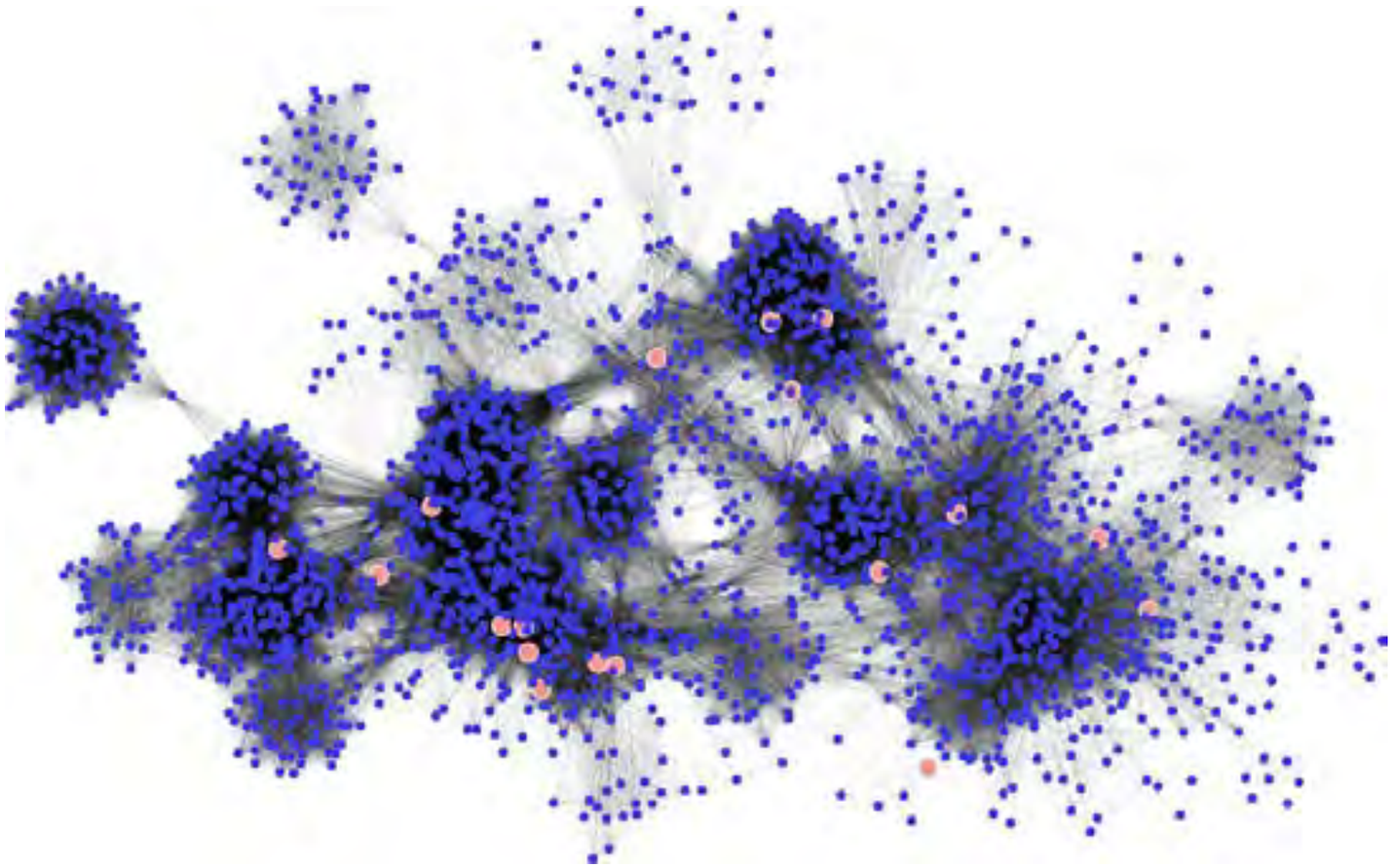


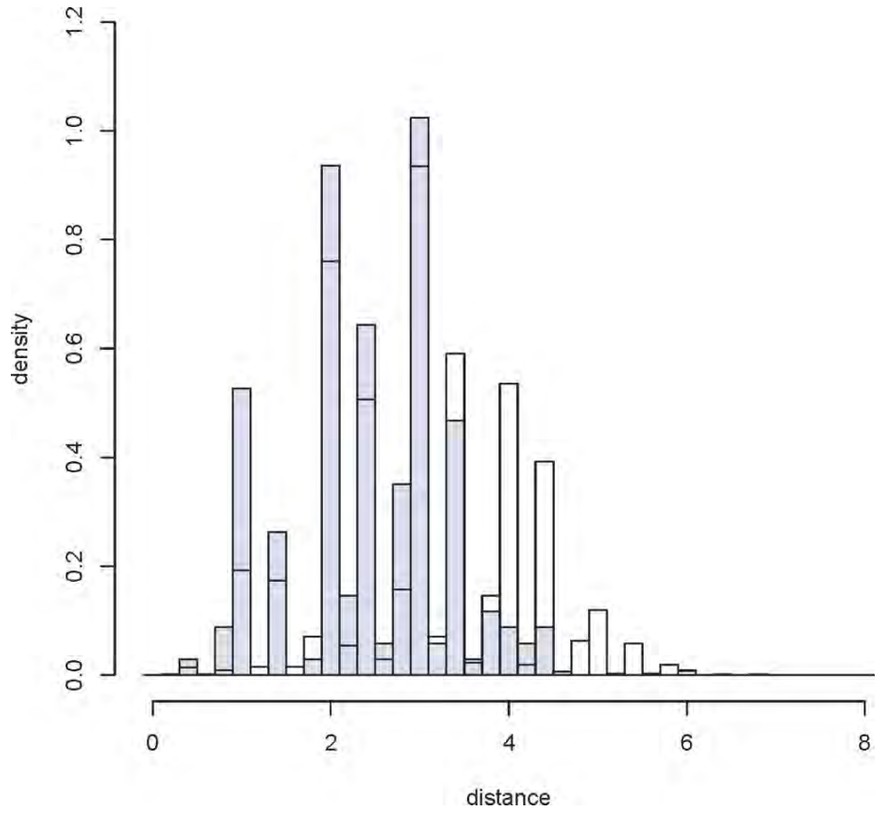
Roca et al. in Leszynski, Puzyn Eds. **Towards efficient design of Safe Nanomaterials: Innovative Merge of Computational Approaches and Experimental Techniques**, RSC Press, 2012

Roca et al. *Novel genomic approaches for environmental risk assessment of metal nanoparticles*, Nanotoxicology 2012, Beijing

C. elegans exposed to nano-Ag

- [nano-Ag]= 0.1 mg/L, exposure time=24h
- Affymetrix GeneChip C. elegans Genome Array
- Gene-pathway annotations obtained from KEGG
- Data size after preprocessing: 1857 genes and 123 pathways





Distance distribution between pairs of genes in the complete network (white) compared to the subset of 19 most differentially expressed genes (gray). Statistical significance of the difference p=0.027

- J.-y. Roh, S. J. Sim, J. Yi, K. Park, K. H. Chung, D.-y. Ryu and J. Choi, *Environ. Sci. Technol.*, 2009, **43**, 3933–3940.
- Roca et al. in Leszynski, Puzyn Eds. **Towards efficient design of Safe Nanomaterials: Innovative Merge of Computational Approaches and Experimental Techniques**, RSC Press, 2012

STRUCTURE-ACTIVITY RELATIONSHIPS

Classification NanoSAR Development for Cytotoxicity of Metal Oxide Nanoparticles

*Rong Liu, Robert Rallo, Saji George, Zhaoxia Ji, Sumitra Nair, André E. Nel, and Yoram Cohen**

Model: Logistic regression

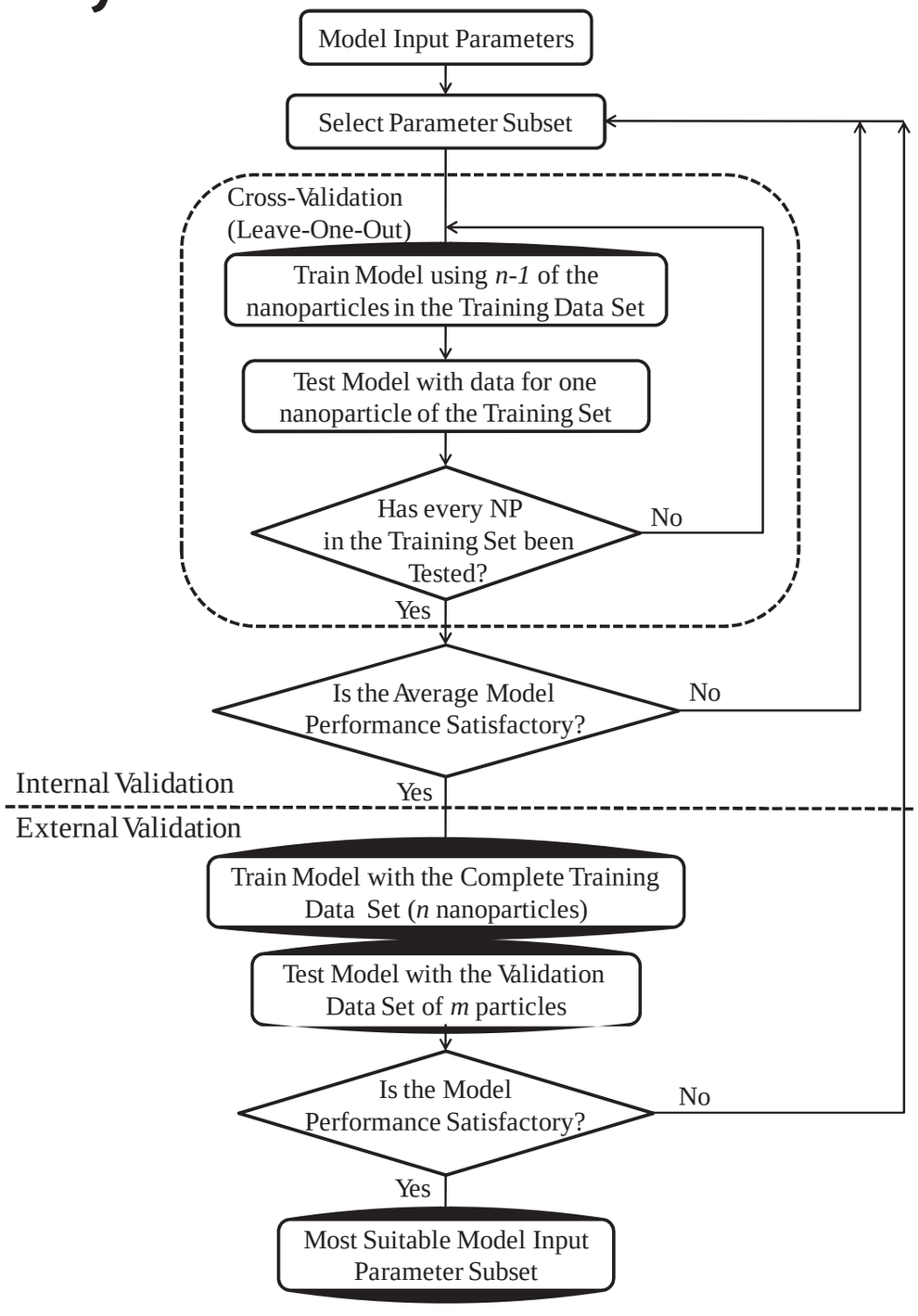
$$\ln\left(\frac{\mathrm{P}(NP\in T)}{\mathrm{P}(NP\in N)}\right) = 3600.6 + 103.5 \times d + 9.5 \times \theta_v + 97.6 \times P_{\mathrm{Me}} - 58.5 \times E_{\mathrm{MeO}}$$

Dataset: 9 metal oxide nanoparticles
Exposed sample: BEAS-2B cells
Endpoint: loss of plasma membrane integrity
Exposure conditions:

- 0.375-200 mg L^{-1}
- up to 24h

Model parameters:

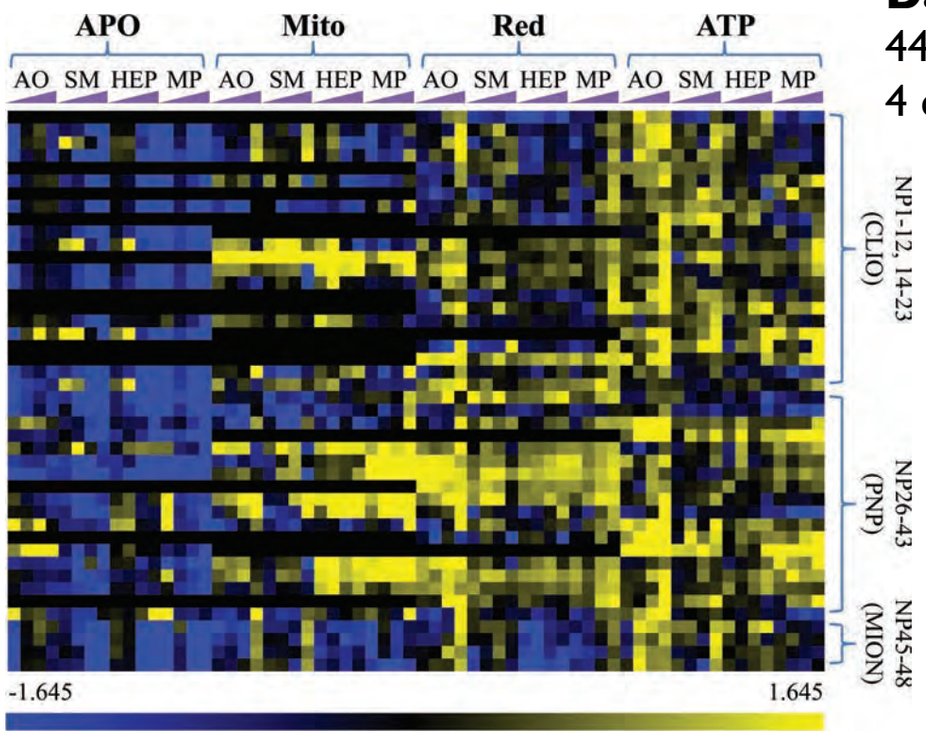
- Primary size
- Period of the metal
- Atomization energy of the metal oxide
- Nanoparticle Volume fraction



Model development pipeline

Nano-SAR Development for Bioactivity of Nanoparticles with Considerations of Decision Boundaries

*Rong Liu, Robert Rallo, Ralph Weissleder, Carlos Tassa, Stanley Shaw, and Yoram Cohen**

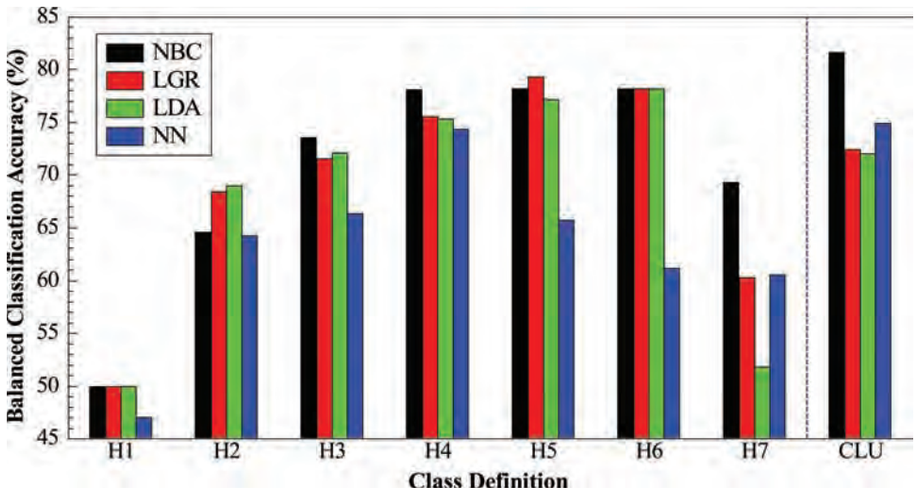


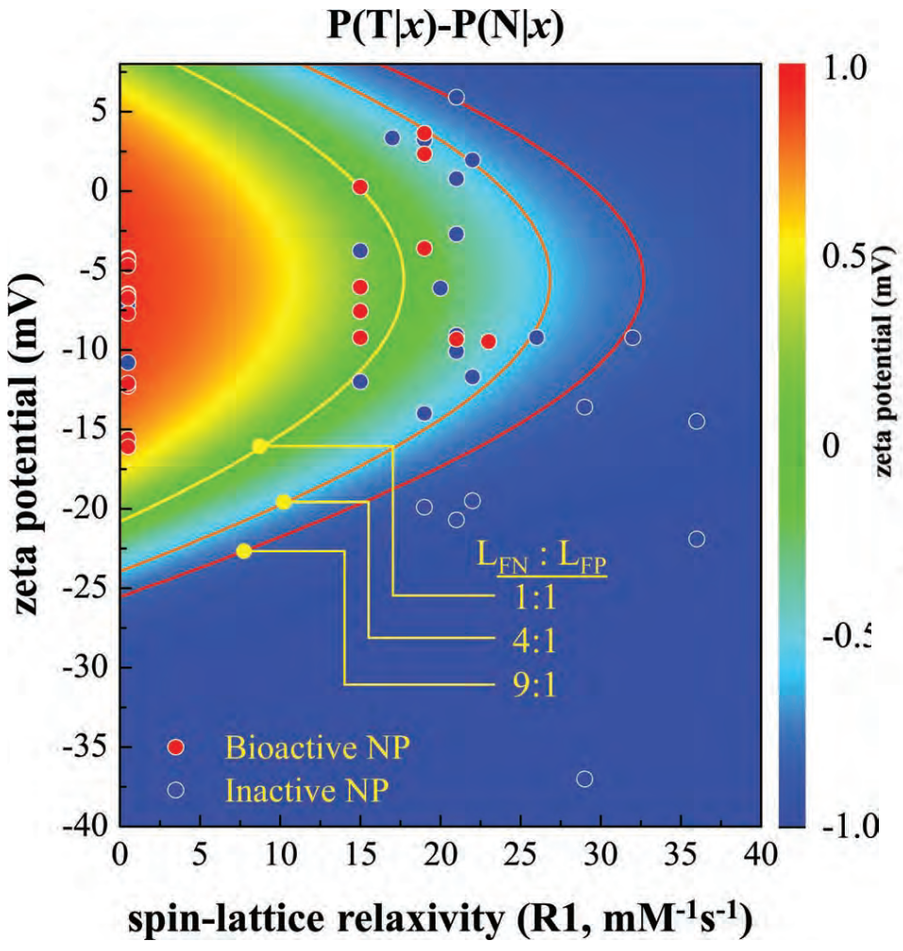
Dataset:
44 iron oxide core nanoparticles
4 cell types & 4 assays

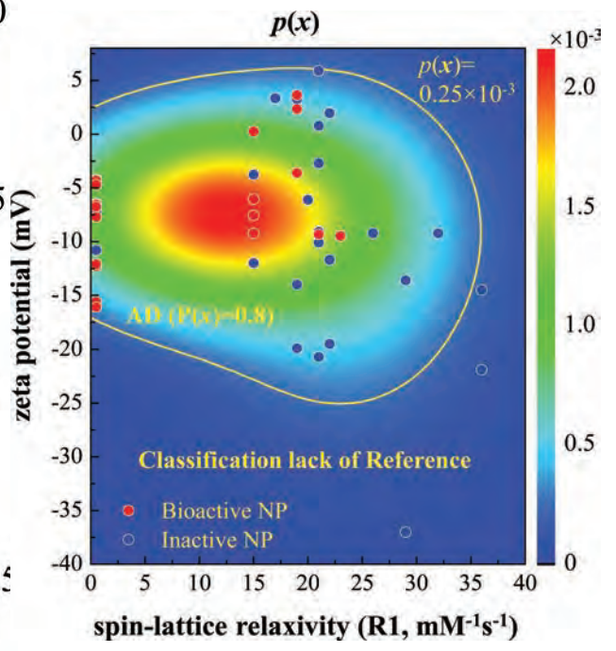
Model: Naive Bayes Classifier

$$p(x|\mathrm{T}) = 3.37 \times 10^{-3} \times \exp\left[-6.05 \times 10^{-3} \times (x_1 - 8.33)^2 - 18.50 \times 10^{-3} \times (x_2 + 6.79)^2\right]$$

$$p(x|\mathrm{N}) = 1.91 \times 10^{-3} \times \exp\left[-6.89 \times 10^{-3} \times (x_1 - 21.04)^2 - 5.20 \times 10^{-3} \times (x_2 + 9.91)^2\right]$$







Nanoscale 2013

RSC Publishing

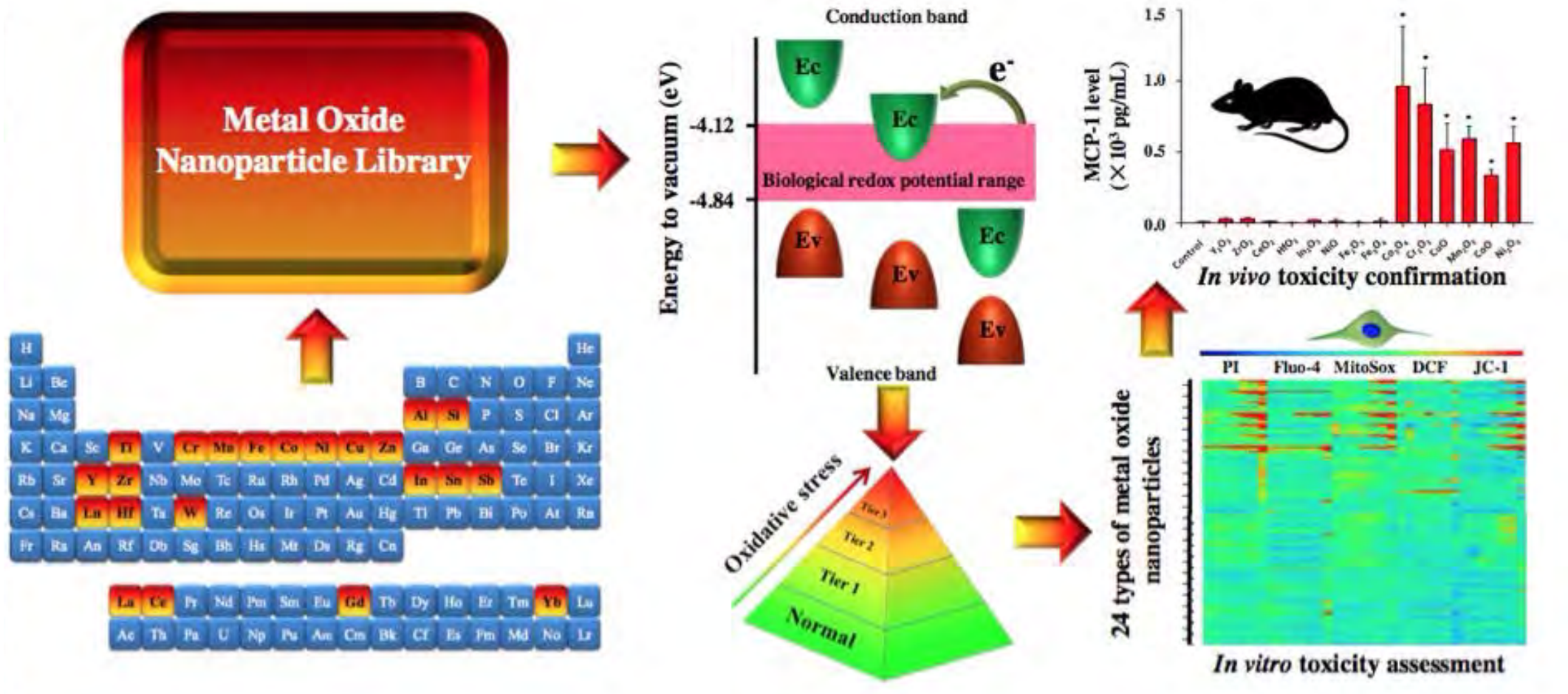
PAPER

View Article Online
View Journal

Development of structure–activity relationship for metal oxide nanoparticles†

Cite this: DOI: 10.1039/c3nr01533e

Rong Liu,[a] Hai Yuan Zhang,[a] Zhao Xia Ji,[a] Robert Rallo,[b] Tian Xia,[c] Chong Hyun Chang,[a] Andre Nel[c] and Yoram Cohen*[ad]

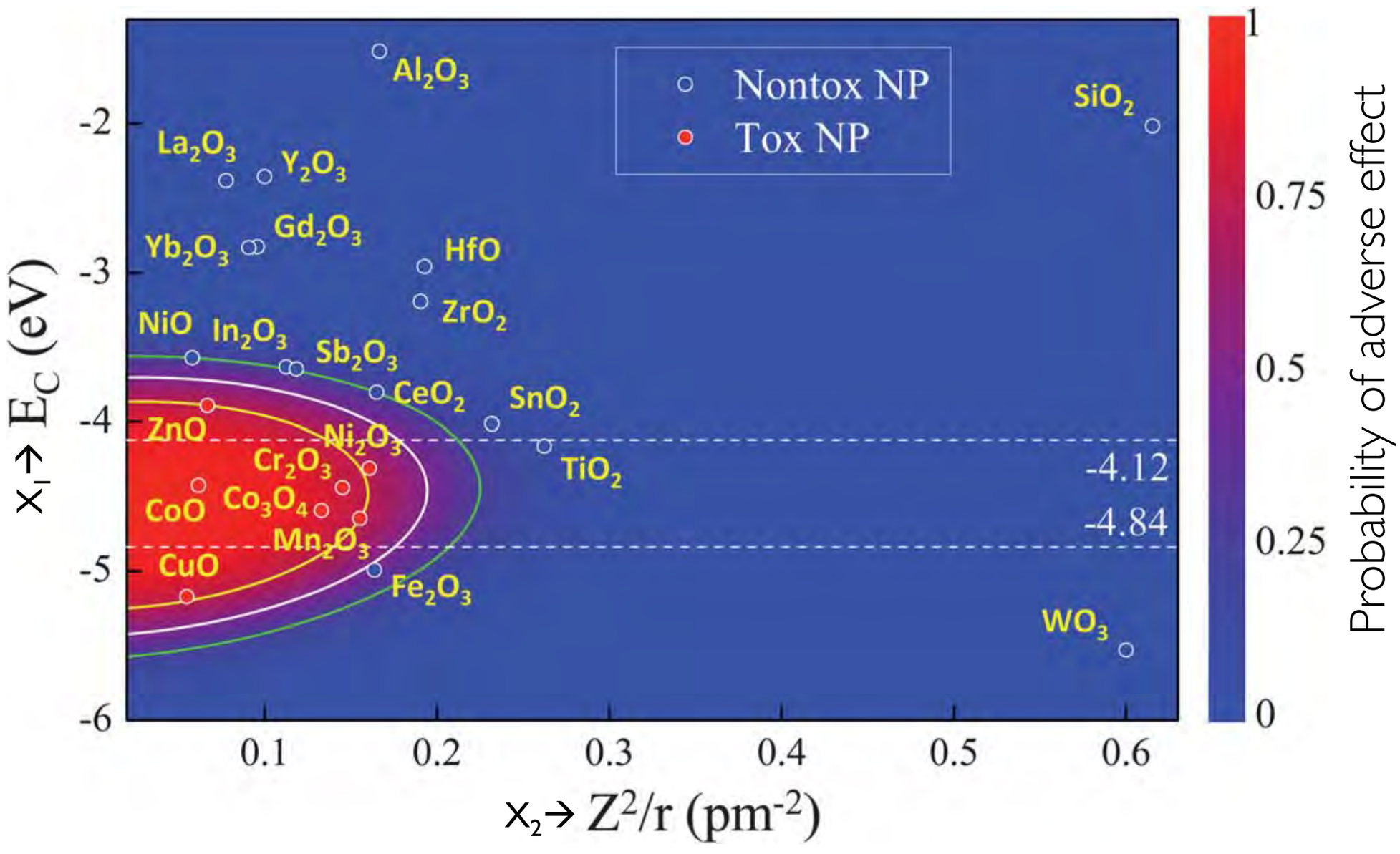


Zhang et al. *ACS Nano*, **2012**, 6(5):4349-4368

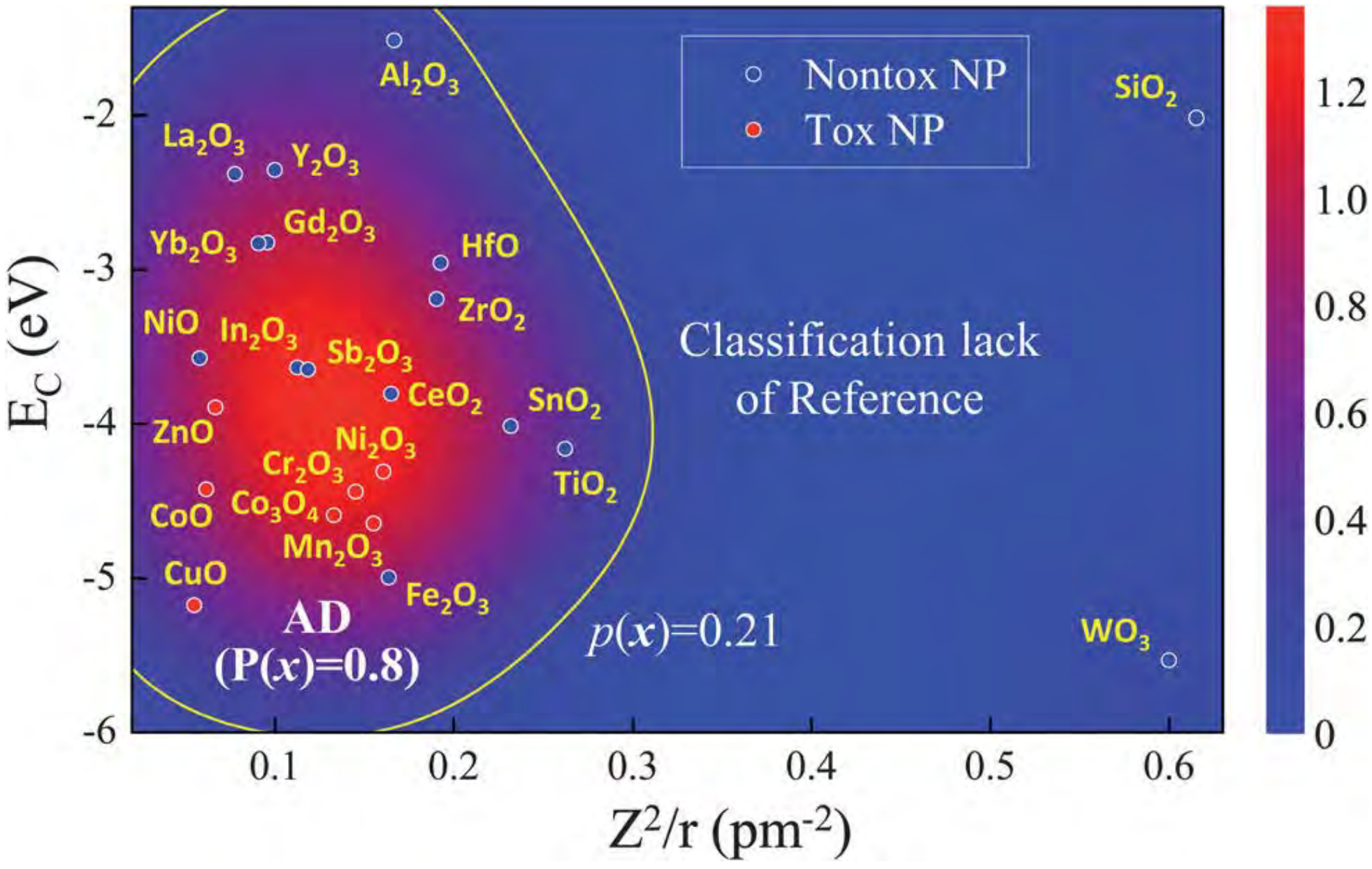
SVM-based Structure Activity Relationship

$$P(T|\boldsymbol{x}) = 1/(1 + \mathrm{e}^{-f(\boldsymbol{x})}) \qquad f(\boldsymbol{x}) = \sum_{i=1}^{6} \alpha_i \, \mathrm{e}^{-2\left[\left(x_{i,1}-x_1\right)^2+\left(x_{i,2}-x_2\right)^2\right]} + b$$

Support vectors:{ZnO, Ni_2O_3, Mn_2O_3, NiO, CeO_2, Fe_2O_3}



Application Domain Analysis
- Optimal Prediction Region -



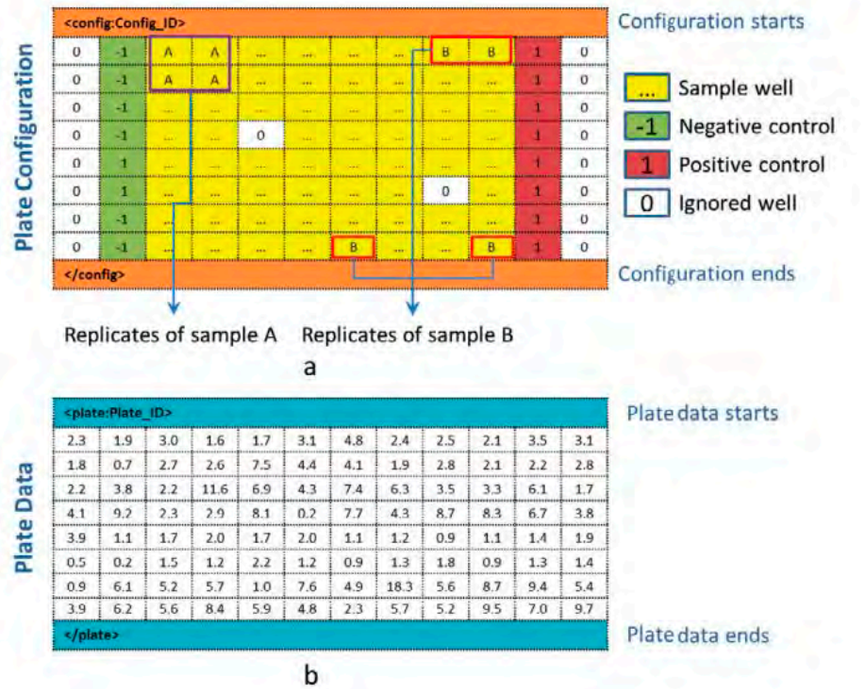
NANOINFORMATICS APPLICATIONS

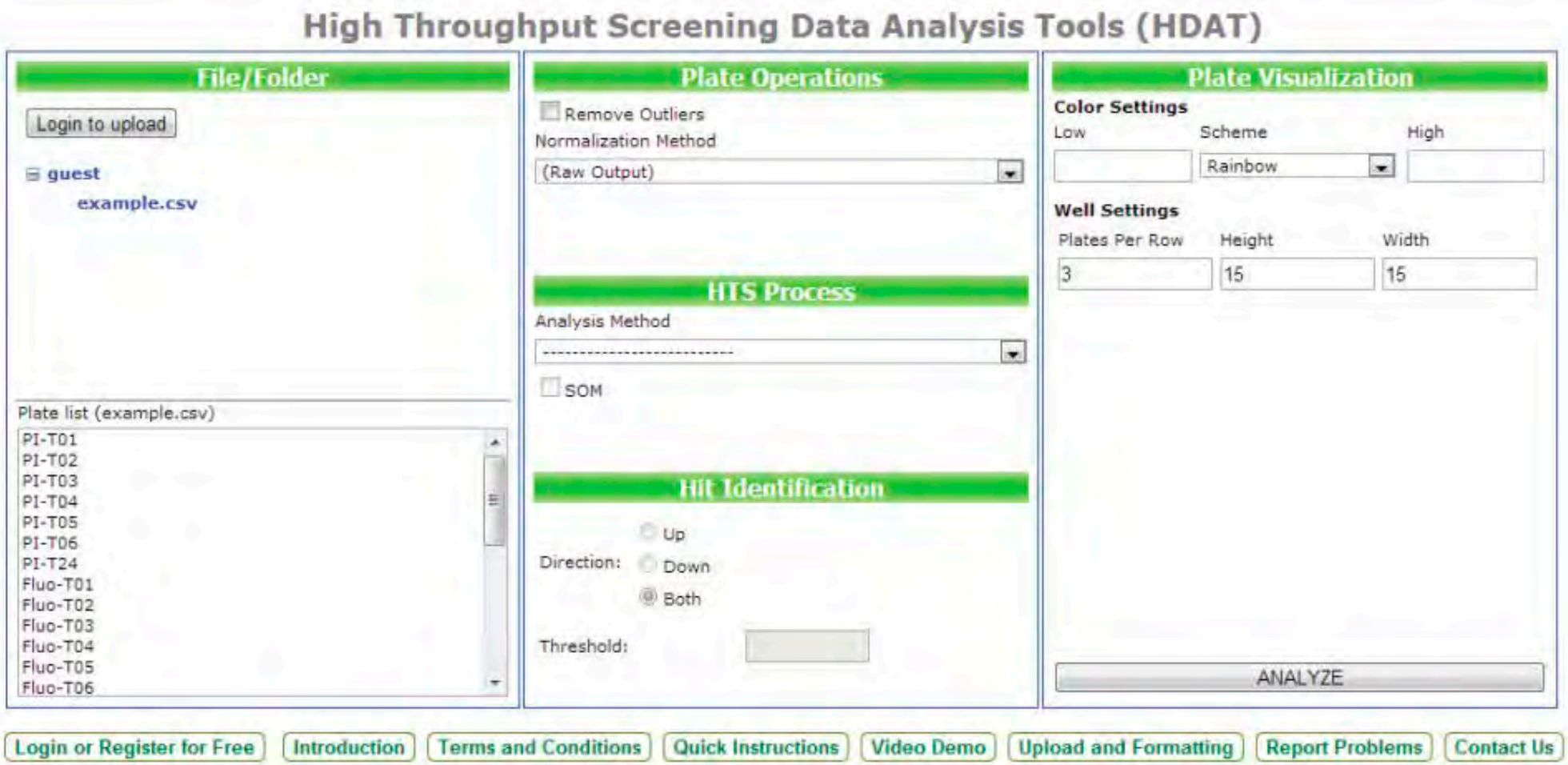
COMPUTATIONAL SCIENCE&DISCOVERY

HDAT: web-based high-throughput screening data analysis tools

Rong Liu[1], Taimur Hassan[1], Robert Rallo[2] and Yoram Cohen[3]

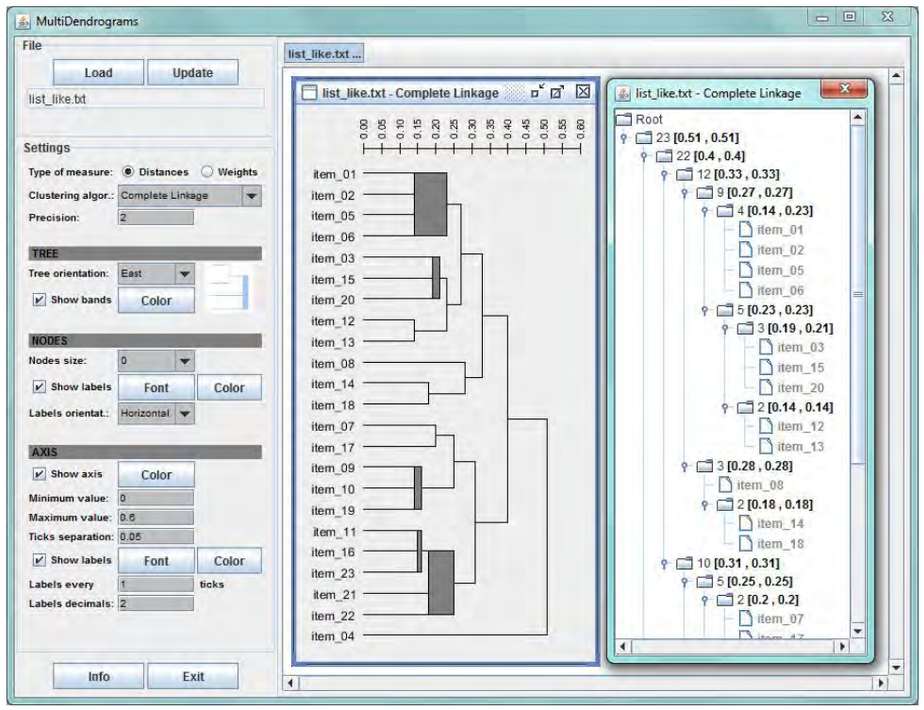
[1] California Nanosystems Institute, University of California, Los Angeles, CA 90095, USA
[2] Departament d'Enginyeria Informatica i Matematiques, Universitat Rovira i Virgili, Avinguda Paisos Catalans 26, E-43007 Tarragona, Catalunya, Spain
[3] Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095, USA
E-mail: yoram@ucla.edu





http://nanoinfo.cein.ucla.edu/public/hdat

Preprocessing and data analysis software



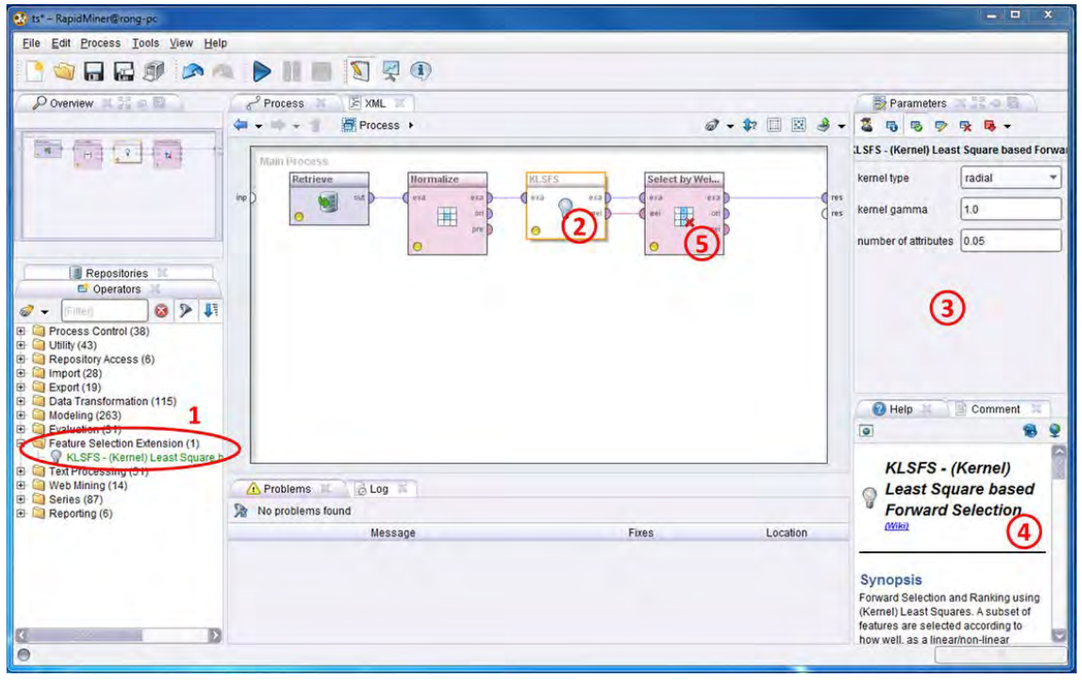
Multidendrograms

Journal of Classification
June 2008, Volume 25, Issue 1, pp 43-65

Solving Non-Uniqueness in Agglomerative Hierarchical Clustering Using Multidendrograms

Alberto Fernández, Sergio Gómez

Available from:
http://deim.urv.cat/~sgomez/multidendrograms.php



RapidMiner 5 Extension for Feature Selection

RONG LIU, ROBERT RALLO, and YORAM COHEN, *Int. J. Info. Tech. Dec. Mak.* **10**, 967 (2011). DOI: 10.1142/S0219622011004671

UNSUPERVISED FEATURE SELECTION USING INCREMENTAL LEAST SQUARES

RONG LIU
Center for the Environmental Implications of Nanotechnology and Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095, USA

ROBERT RALLO
Center for the Environmental Implications of Nanotechnology and Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095, USA
Departament d'Enginyeria Informatica i Matematiques, Universitat Rovira i Virgili, Av. Paisos Catalans, 26. 43007 Tarrragona, Catalunya, Spain

YORAM COHEN
Corresponding author.
Center for the Environmental Implications of Nanotechnology and Chemical and Biomolecular Engineering Department, University of California, Los Angeles, CA 90095, USA

Available from:
http://rongliu.weebly.com/software.html

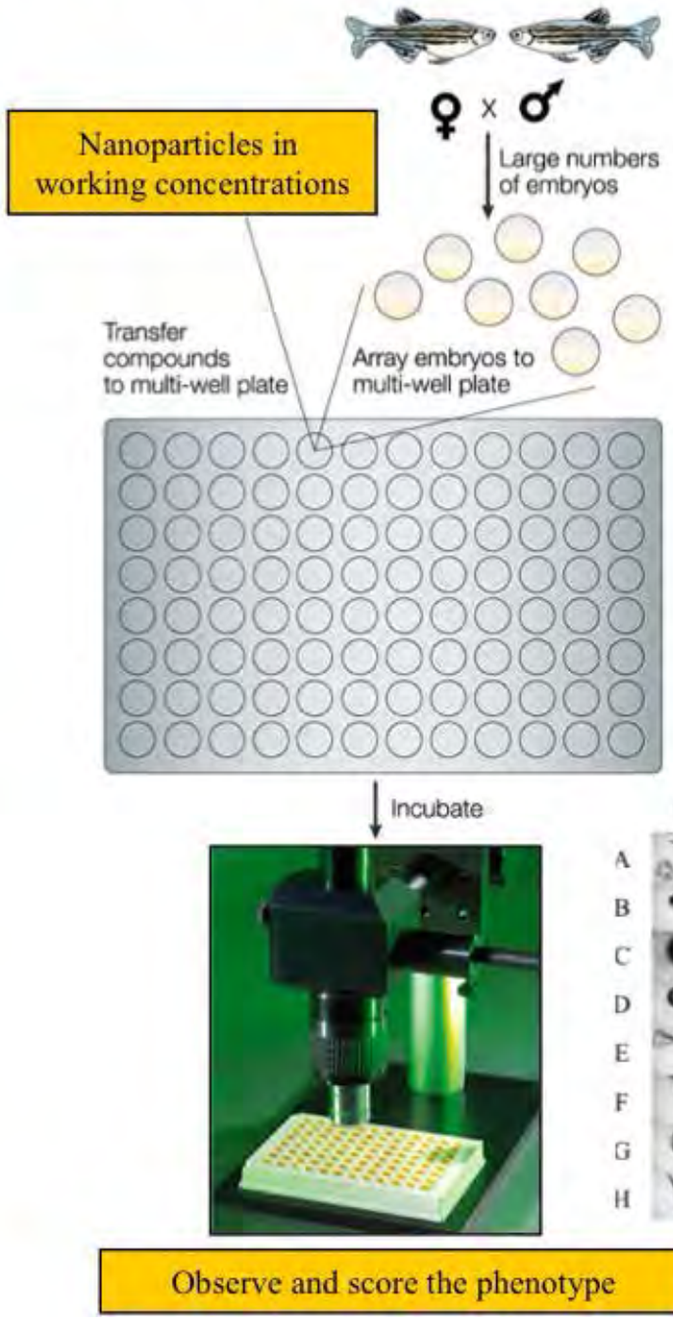
OPEN ACCESS Freely available online

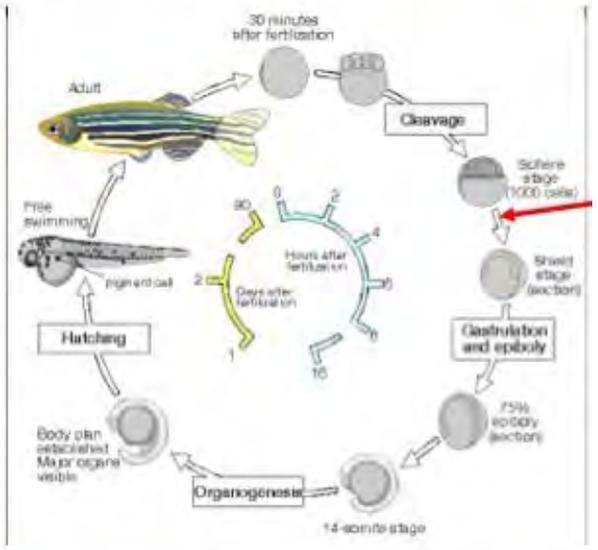
PLoS one

Automated Phenotype Recognition for Zebrafish Embryo Based *In Vivo* High Throughput Toxicity Screening of Engineered Nano-Materials

Rong Liu[1,2], Sijie Lin[1], Robert Rallo[1,3], Yan Zhao[4], Robert Damoiseaux[5], Tian Xia[1,6], Shuo Lin[4], Andre Nel[1,6], Yoram Cohen[1,2]*

Tool for automatic zebrafish embryo phenotyping

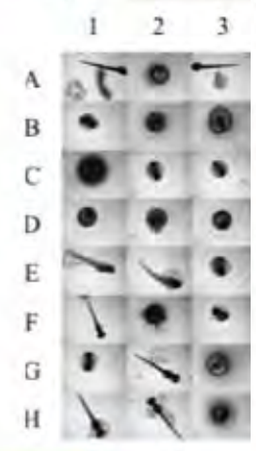


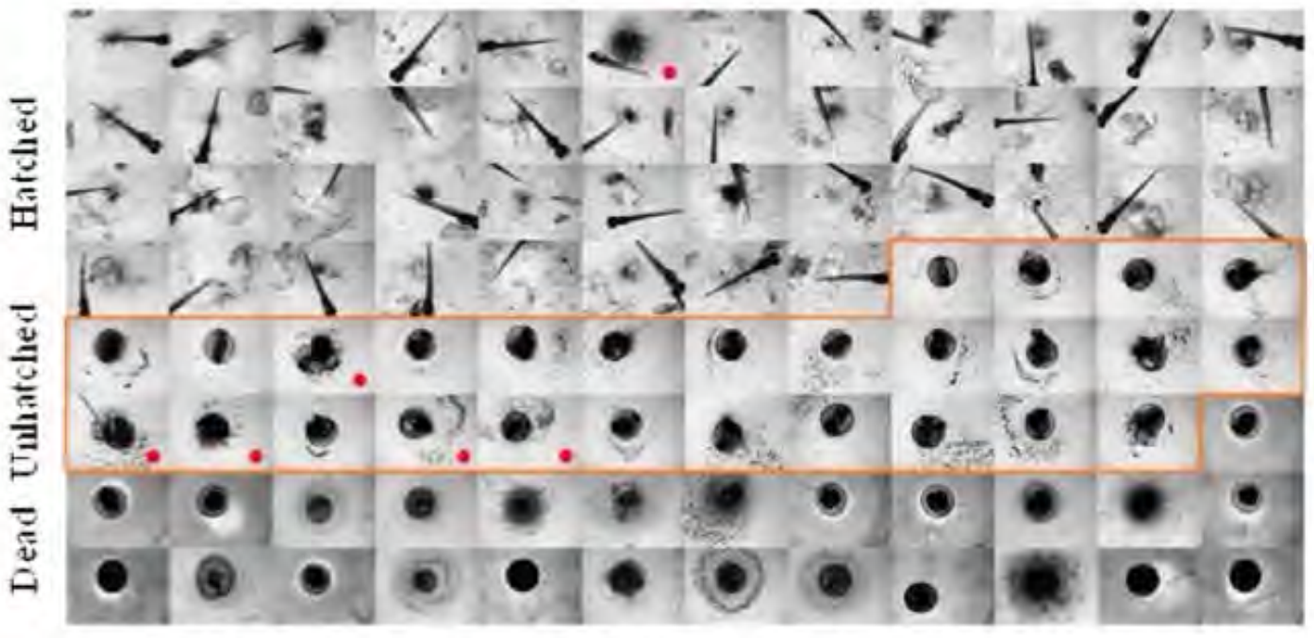


	true hatched	true unhatched	true dead	class precision*
pred. hatched	**524**	4	8	97.76%
pred. unhatched	4	**321**	12	95.25%
pred. dead	0	2	**278**	99.29%
class recall†	99.24%	98.17%	93.29%	

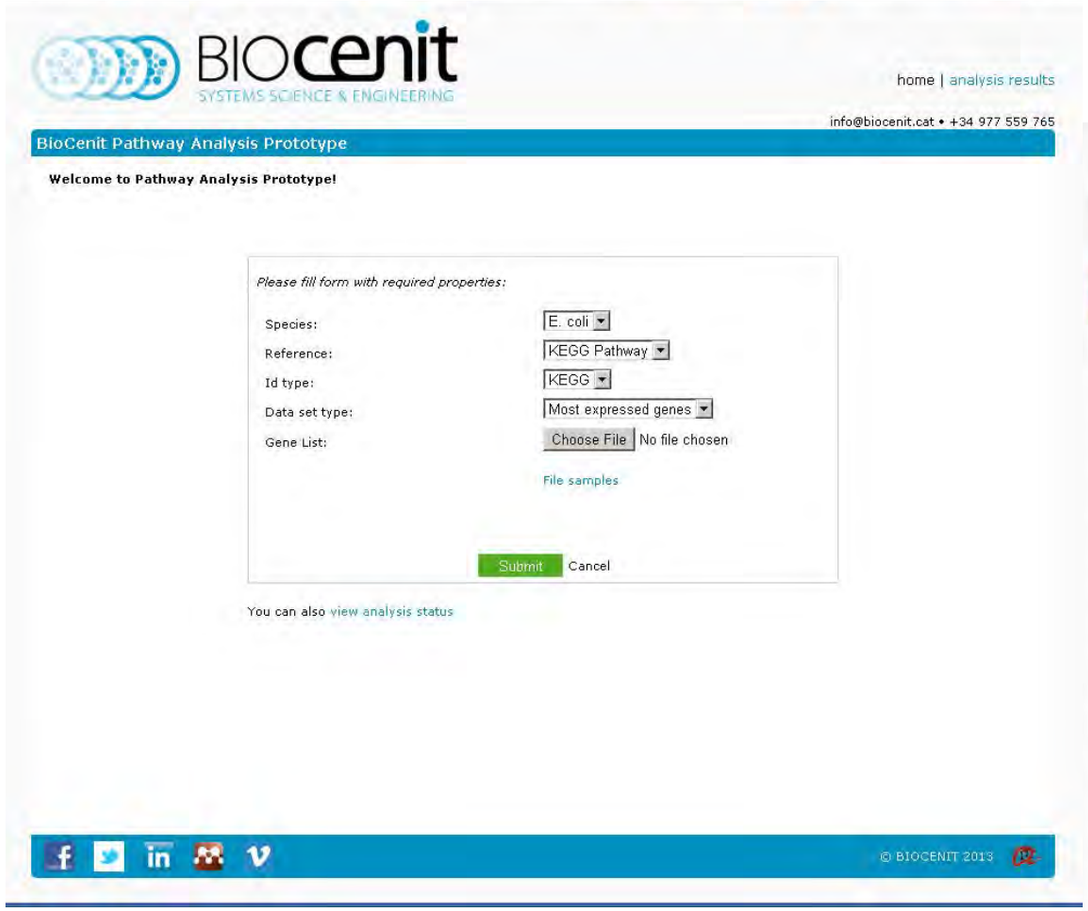
* - The class precision is the percentage of correct classified samples in a predicted class. For example the precision of the (predicted) hatched class is given by 524/(524+4+8) = 97.76%.

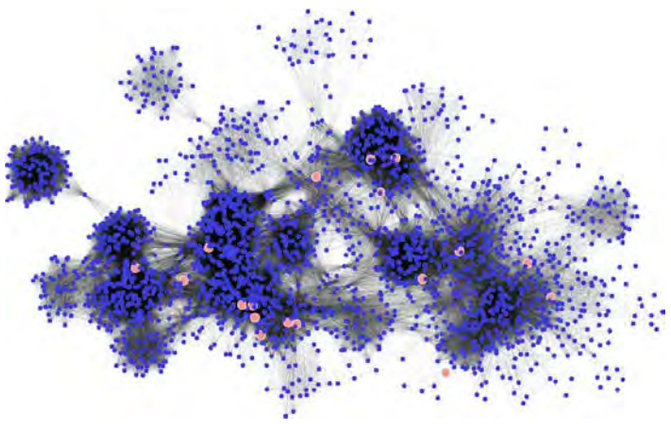
† - The class recall is the proportion of the samples in the class that were correct identified. For example the recall of (true) hatched class is 524/(524+4+0) = 99.24%.

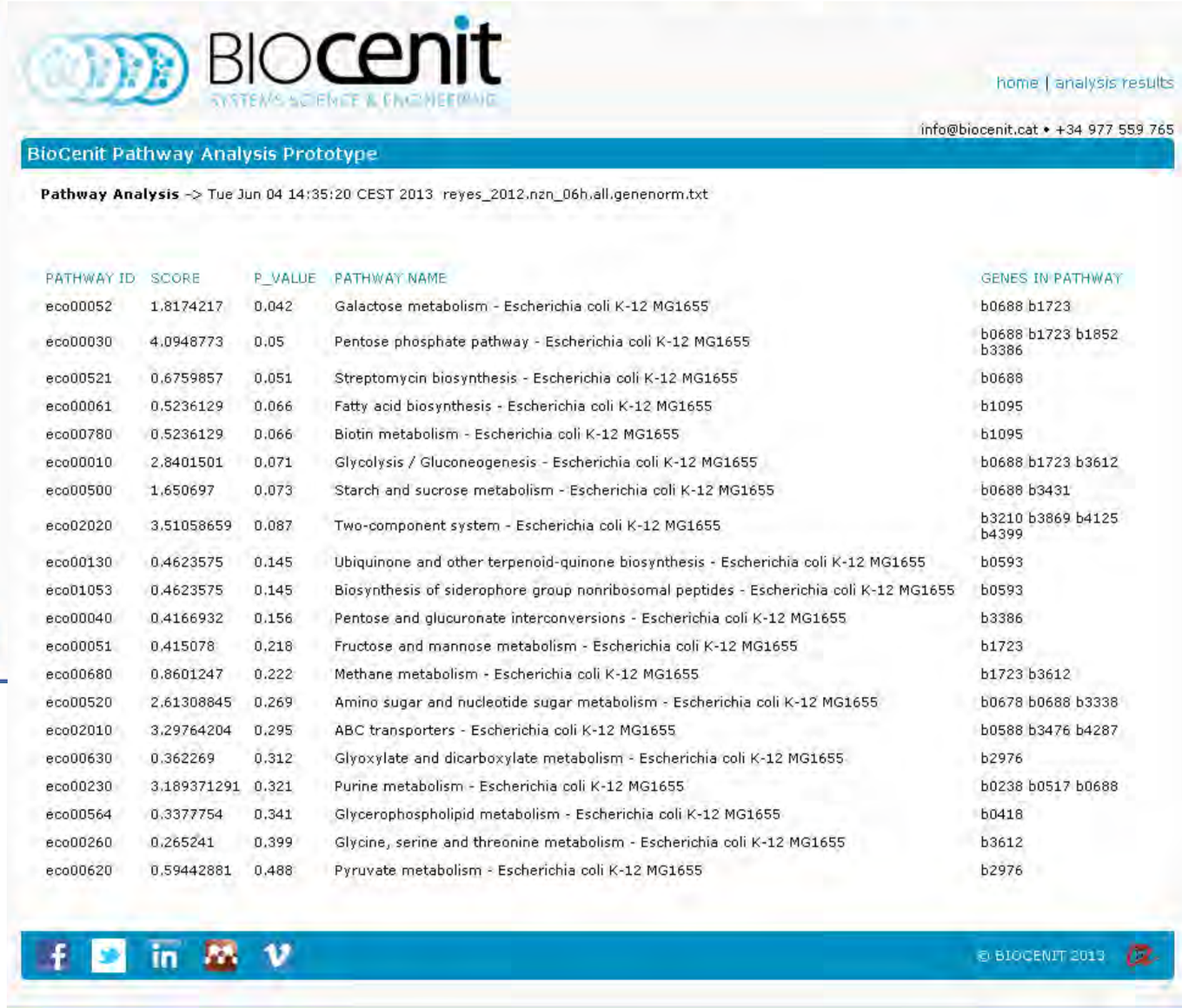




Web tool for the integrated analysis of molecular signatures



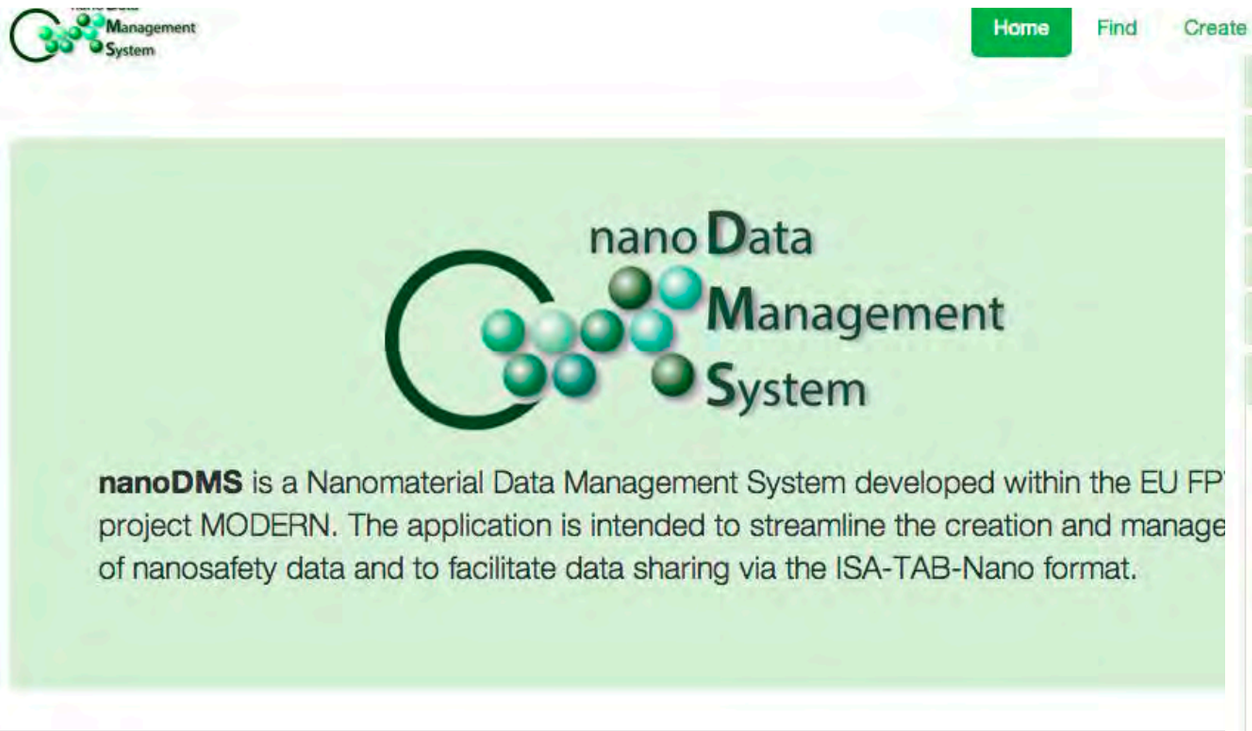


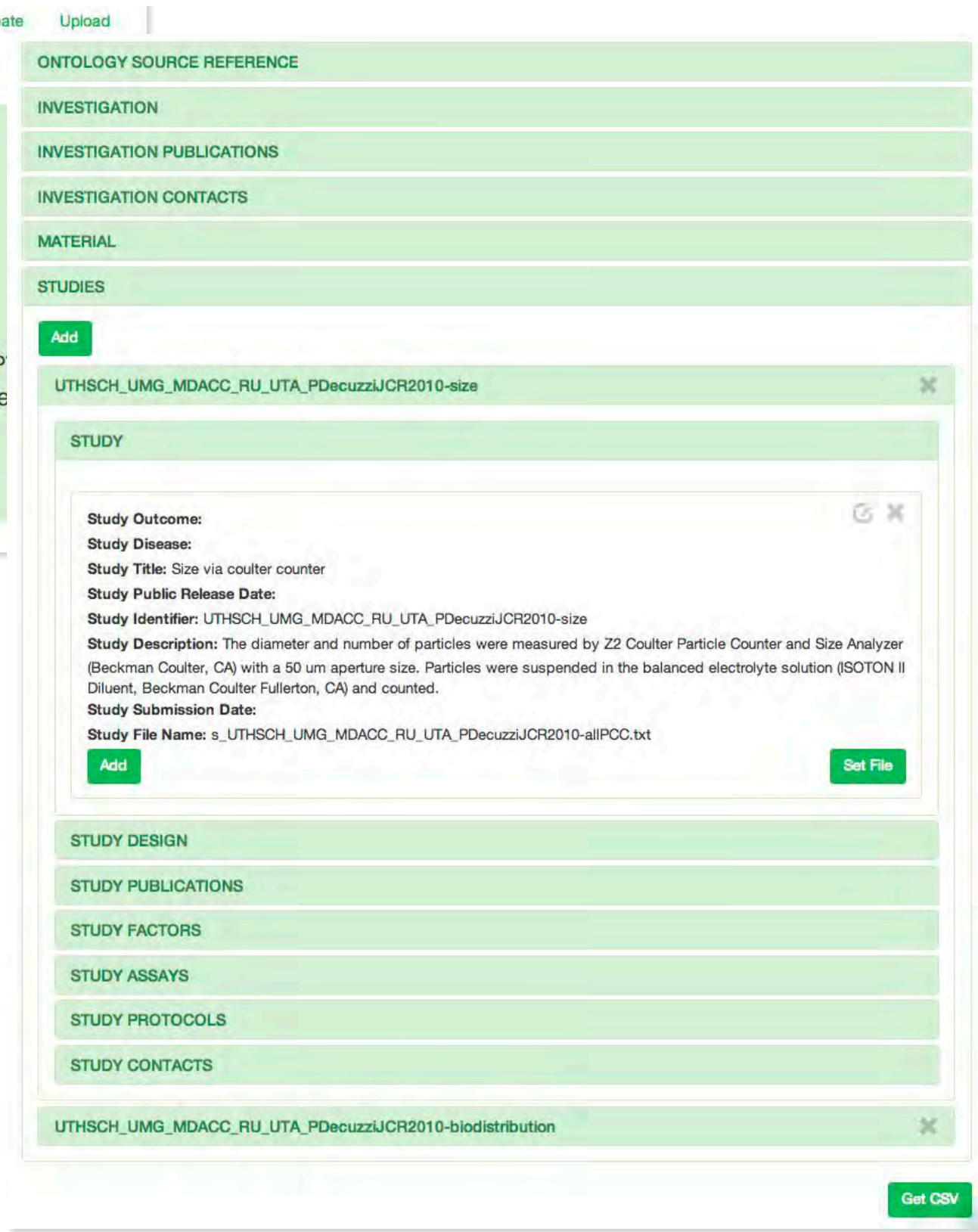


PATHWAY ID	SCORE	P_VALUE	PATHWAY NAME	GENES IN PATHWAY
eco00052	1.8174217	0.042	Galactose metabolism - Escherichia coli K-12 MG1655	b0688 b1723
eco00030	4.0948773	0.05	Pentose phosphate pathway - Escherichia coli K-12 MG1655	b0688 b1723 b1852 b3386
eco00521	0.6759857	0.051	Streptomycin biosynthesis - Escherichia coli K-12 MG1655	b0688
eco00061	0.5236129	0.066	Fatty acid biosynthesis - Escherichia coli K-12 MG1655	b1095
eco00780	0.5236129	0.066	Biotin metabolism - Escherichia coli K-12 MG1655	b1095
eco00010	2.8401501	0.071	Glycolysis / Gluconeogenesis - Escherichia coli K-12 MG1655	b0688 b1723 b3612
eco00500	1.650697	0.073	Starch and sucrose metabolism - Escherichia coli K-12 MG1655	b0688 b3431
eco02020	3.51058659	0.087	Two-component system - Escherichia coli K-12 MG1655	b3210 b3869 b4125 b4399
eco00130	0.4623575	0.145	Ubiquinone and other terpenoid-quinone biosynthesis - Escherichia coli K-12 MG1655	b0593
eco01053	0.4623575	0.145	Biosynthesis of siderophore group nonribosomal peptides - Escherichia coli K-12 MG1655	b0593
eco00040	0.4166932	0.156	Pentose and glucuronate interconversions - Escherichia coli K-12 MG1655	b3386
eco00051	0.415078	0.218	Fructose and mannose metabolism - Escherichia coli K-12 MG1655	b1723
eco00680	0.8601247	0.222	Methane metabolism - Escherichia coli K-12 MG1655	b1723 b3612
eco00520	2.61308845	0.269	Amino sugar and nucleotide sugar metabolism - Escherichia coli K-12 MG1655	b0678 b0688 b3338
eco02010	3.29764204	0.295	ABC transporters - Escherichia coli K-12 MG1655	b0588 b3476 b4287
eco00630	0.362269	0.312	Glyoxylate and dicarboxylate metabolism - Escherichia coli K-12 MG1655	b2976
eco00230	3.189371291	0.321	Purine metabolism - Escherichia coli K-12 MG1655	b0238 b0517 b0688
eco00564	0.3377754	0.341	Glycerophospholipid metabolism - Escherichia coli K-12 MG1655	b0418
eco00260	0.265241	0.399	Glycine, serine and threonine metabolism - Escherichia coli K-12 MG1655	b3612
eco00620	0.59442881	0.488	Pyruvate metabolism - Escherichia coli K-12 MG1655	b2976

Accesible from: http://www.biocenit.cat

Nanosafety Data Management System





Features:

- Web-based application
- ISA-TAB-Nano validation
- Import/Export
- Ontology support

Acknowledgment

- EU FP7 MODERN, Contract No. 309314
- UC CEIN, NSF/EPA (NSF DBI-0830117)

The members of UC CEIN and the members of BIOCENIT Research Lab at Universitat Rovira i Virgili