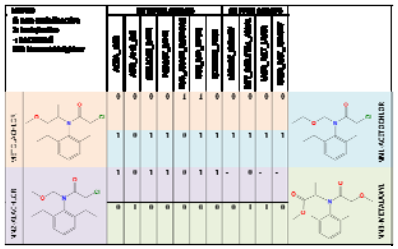
# Using ToxCast™ *in vitro* Assays in the Hierarchical Quantitative Structure-Activity Relationship (QSAR) Modeling for Predicting *in vivo* Toxicity of Chemicals

Zhu, Hao[1]; Zhang, Liying[1]; Sedykh, Alexander[1]; Tang, Hao[1]; Richard, Ann[2]; Rusyn, Ivan[1]; Tropsha, Alexander[1]
[1]Carolina Center for Computational Toxicology, University of North Carolina at Chapel Hill, Chapel Hill, NC U.S.A.
[2]NCCT, US EPA, Research Triangle Park, NC, USA.

## INTRODUCTION

- Phase I of the ToxCast™ program profiled 309 well-characterized chemicals (primarily pesticides) in over 600 HTS endpoints. Most of these compounds have also been tested in ca. 80 developmental toxicity, multi-generation reproductive studies, and sub-chronic and chronic bioassays.
- It is often challenging to rationalize differences between *in vivo* toxicological response to chemically similar molecules (Table 1). The analysis of ToxCast™ data requires new computational approaches to link chemical structures, *in vitro* responses and *in vivo* toxicity effects.

Table 1 Example of a hard-to-predict compound and its three structurally nearest neighbors. The *in vitro* and *in vivo* toxicity profiles are compared.



## DATA PREPARATION

- After data curation, 291 chemicals were kept for analysis. By removing highly correlated (R2 >0.95) and low-signal (<10 non-zero entries) assays, 284 in vitro assays were selected. In this study, we will focus on three reproductive toxicity endpoints (i.e., MGR_Rat_Kidney, MGR_Rat_Liver, and MGR_Rat_ViabilityPND4).

## METHODOLOGY

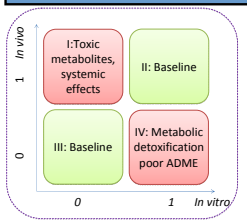



Figure 1 The relationship between *in vitro bioassay* and *in vivo* toxicity data. Class I compounds: toxic *in vivo* but non-toxic *in vitro*; Class II compounds: toxic *in vivo* and *in vitro*; Class III compounds: non-toxic *in vivo* but toxic *in vitro*; Class IV compounds: non-toxic *in vivo* and *in vitro*.

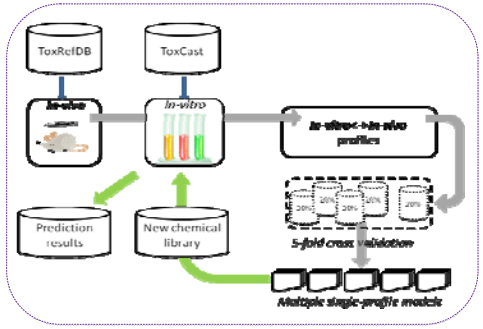



Figure 2. Scheme of the hierarchical QSAR modeling workflow on ToxCast™ *in vivo* toxicity data using chemical *in vitro-in vivo* correlation profiles. Grey arrows indicate modeling procedure and green arrows indicate prediction procedure.

## RESULTS

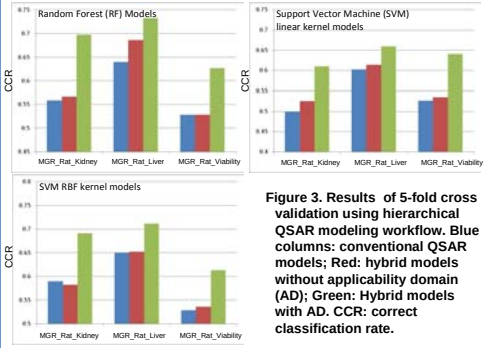



Figure 3. Results of 5-fold cross validation using hierarchical QSAR modeling workflow. Blue columns: conventional QSAR models; Red: hybrid models without applicability domain (AD); Green: Hybrid models with AD. CCR: correct classification rate.

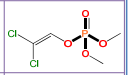
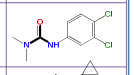
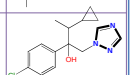
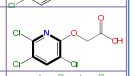
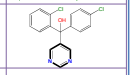
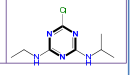
## TOXCAST™ BIOASSAYS ANALYSIS

Table 2. The top 20 ToxCast™ bioassays based on external predictivity for three reproductive toxicity endpoints. Colors represent data sources.

| MGR_Rat_Kidney | MGR_Rat_Liver | MGR_Rat_Viability |
|---|---|---|
| ATG_Ahr_CIS | ACEA_LOC3 | ATG_ERa_TRANS |
| ATG_EGR_CIS | ATG_Pax6_CIS | BSK_3C_hLADR_down |
| ATG_HSE_CIS | ATG_THRa1_TRANS | BSK_3C_Proliferation_down |
| ATG_LXRb_TRANS | BSK_3C_MCP1_up | BSK_BE3C_hLADR_up |
| ATG_Pax6_CIS | BSK_3C_SRB_down | BSK_hDFCGF_IP10_down |
| ATG_RARa_TRANS | BSK_4H_Pselectin_down | BSK_hDFCGF_MIG_down |
| ATG_VDRE_CIS | BSK_4H_VCAM1_down | BSK_LPS_PGE2_down |
| ATG_Xbp1_CIS | BSK_BE3C_uPAR_down | BSK_LPS_VCAM1_down |
| BSK_3C_MCP1_up | BSK_hDFCGF_CollagenIII_up | BSK_SM3C_VCAM_1_down |
| BSK_3C_Thrombomodulin | BSK_LPS_MPC1_down | CLM_Hepat_DNADmg_1hr |
| BSK_3C_Vis_down | CLM_MicrotubuleCSK_24hr | CLM_Hepat_DNADmg_24hr |
| BSK_4H_Eotaxin3_down | CLM_OxidativeStress_24hr | CLM_Hepat_DNADmg_48hr |
| BSK_BE3C_TGFb1_up | CLM_p53Act_1hr | CLM_MitoticArrest_24hr |
| BSK_hDFCGF_EGFR_up | CLZD_CYP1A1_6 | CLM_OxidativeStress_1hr |
| BSK_KF3CT_IL1a_down | CLZD_CYP2B6_48 | CLM_StressKinase_1hr |
| BSK_SAg_CD40_up | CLZD_GSTA2_24 | NCGC_AR_Antagonist |
| BSK_SAg_IL8_down | NCGC_ERalpha_Agonist | NVS_ADME_hCYP2C9 |
| CLM_p53Act_72hr | NVS_GPCR_g5HT4 | NVS_ADME_rCYP2A2 |
| CLZD_SULT2A1_24 | NVS_GPCR_hDRD1 | NVS_ADME_rCYP2C11 |
| NCGC_PXR_Agonist_rat | NVS_GPCR_hOpiate_mu | NVS_ADME_rCYP2C6 |

## CHEMICAL SCAFFOLD ANALYSIS

Table 3 Scaffold examples in ToxCast™ database.

| | Scaffold | Dragon Descriptor | Cluster Size | Example | |
|---|---|---|---|---|---|
| 1 | Phosphates and thio-phosphates | nPO4 | 24 | Dichlorvos CAS# 62-73-7 |  |
| 2 | Urea derivatives | nCONN | 20 | Diuron CAS# 330-54-1 |  |
| 3 | Triazoles | nTriazoles | 18 | Cyproconazole CAS# 94361-06-5 |  |
| 4 | Pyridines | nPyridines | 33 | Triclopyr CAS# 55335-06-3 |  |
| 5 | Pyrimidines | nPyrimidines | 23 | Fenarimol CAS# 60168-88-9 |  |
| 6 | 1,3,5-Triazines | n135-Triazines | 20 | Atrazine CAS# 1912-24-9 |  |

- QSAR models based on a subset of compounds with these scaffolds had better prediction accuracy than that on the entire ToxCast™ dataset. For example, models developed for MGR_Rat_Liver using this subset and RF/SVM methods increased $CCR_{ext}$ from 0.61~0.67 to 0.76~0.81.

## DISCUSSION

- This study used the concept of "*in vitro-in vivo* correlation", which has been successfully employed to correlate rodent acute toxicity of chemicals with cytotoxicity in our lab (Zhu, 2009), to develop predictive hierarchical QSAR models for three reproductive toxicity endpoints (i.e., MGR_Rat_Liver, MGR_Rat_Kidney, and MGR_Rat_Viability).
- Consensus prediction and applicability domain approaches can effectively increase the prediction power of QSAR models.
- Ranking of ToxCast™ bioassays for each toxicity endpoint based on the external predictivity helps to prioritize bioassays that are important for *in vivo* toxicity evaluation of chemicals.
- Improved models were developed on chemicals with six major scaffolds. This implies the possible relationships between these chemical scaffolds and specific toxicological mechanisms. Future study will focus on finding the possible roles of these structure features in *in vivo* toxicological process of chemicals.

## CONCLUSIONS

- We have developed predictive QSAR models for three *in vivo* reproductive toxicity endpoints by taking into account the entire chemical structure – *in vitro* – *in vivo* data continuum.
- Our resulting models could be used to guide the future toxicity studies on the EPA-10K compounds by selecting *in vitro* assays and prioritizing compounds for *in vivo* toxicity evaluation.

**REFERENCES:**
- Martin et al. *Environmental Health Perspectives* 2009 117(3):392–399
- Dix et al. *Toxicological Sciences* 2007 95(1):5-12
- Hao et al. Environ Health Perspect.2009 17:1257-1264

**ACKNOWLEDGEMENTS:**
These studies have been supported in part by the U.S. EPA Center grants RD832720 and RD833825 and NIH grants GM076059 and ES005948) awarded to UNC collaborators.

**CONTACT**: Alexander Tropsha (alex_tropsha@unc.edu)