CoMPARA: Collaborative Modeling Project for Androgen Receptor Activity

Kamel Mansouri[1], Nicole Kleinstreuer[2], Eric Watt[1], Jason Harris[1], Richard S. Judson[3]

[1]ORISE Fellow, NCCT, ORD/U.S. EPA, RTP, NC, USA

[2]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA

[3]National Center for Computational Toxicology, ORD/U.S. EPA, RTP, NC, USA

In order to protect human health from chemicals that can mimic natural hormones, the U. S. Congress mandated the U.S. EPA to screen chemicals for their potential to be endocrine disruptors through the Endocrine Disruptor Screening Program (EDSP). However, the number of chemicals to which humans are exposed is too large (tens of thousands) to be accommodated by the EDSP Tier 1 battery, so combinations of in vitro high-throughput screening (HTS) assays and computational models are being developed to help prioritize chemicals for more detailed testing. Previously, CERAPP (Collaborative Estrogen Receptor Activity Prediction Project) demonstrated the effectiveness of combining many QSAR models trained on HTS data to prioritize a large chemical list for estrogen receptor activity. The limitations of single models were overcome by combining all models built by the consortium into consensus predictions. CoMPARA is a larger scale collaboration between 35 international groups, following the steps of CERAPP to model androgen receptor activity using a common training set of 1746 compounds provided by U.S. EPA. Eleven HTS ToxCast/Tox21 *in vitro* assays were integrated into a computational network model to detect true AR activity. Bootstrap uncertainty quantification was used to remove potential false positives/negatives. Reference chemicals (158) from the literature were used to validate the model, which showed 95.2% and 97.5% balanced accuracies for AR agonists and antagonists respectively. A library of ~80k bioactivities, representing ~11k chemicals curated from PubChem literature data using ScrubChem tools was integrated with CoMPARA's consensus predictions that combined several structure-based and QSAR modeling approaches. The results of this project will be used to prioritize a large set of more than 50k chemicals for further

testing over the next phases of ToxCast/Tox21, among other projects. *This work does not reflect the official policy of any federal agency.*