Supporting Information

An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library

Authors: Nisha S. Sipes[1]*, John F. Wambaugh[2], Robert Pearce[2], Scott S. Auerbach[1], Barbara A. Wetmore[4], Jui-Hua Hsieh[1], Andrew J. Shapiro[1], Daniel Svoboda[3], Michael J. DeVito[1], Stephen S. Ferguson[1]

[1]NTP/NIEHS/NIH, RTP, NC; [2]NCCT/ORD/EPA, RTP, NC, [3]Sciome, RTP, NC; [4]NERL/ORD/EPA, RTP, NC

*Corresponding Author: Nisha S. Sipes, 111 T.W. Alexander Drive, PO Box 12233, MD: K2-17, Research Triangle Park, NC 27709, Telephone: 919-316-4603, Email: nisha.sipes@nih.gov

9 pages including Supporting Information Cover Sheet, Supplemental Methods, Supplemental Results, and Figures S1-S5

Figures S1-S5

Tables S1-S5

SUPPLEMENTAL METHODS

Tox21/ToxCast HTS Data.

Figure S1 provides a detailed workflow of important considerations between the Tox21 and ToxCast pipelines, how the data were treated post-processing, and detailed numbers on the active calls.

Tox21 pipeline. Tox21-only HTS data were obtained from a pipeline specifically designed to analyze the Tox21 HTS data.[1, 2] It is important to note that in addition to statistically calculating parameters such as AC_{50} and efficacy, this pipeline accounts for autofluorescence, weak signals, and conflicts between the ratio and the readout channel to ensure there is no signal interference simply due to the background channel changing. Additionally, most Tox21 antagonist assays have a cytotoxicity readout in the same well, which as standard protocol, this pipeline directly compares to the AC_{50} for close proximity to rule out falsely identifying an antagonist. Since the EPA's TCPL pipeline does not account for these factors, we chose to use this fitting method for the Tox21 data. The current pipeline indicates 44,881 active sample id (Tox21ID)-assay calls.

ToxCast pipeline. ToxCast-only HTS data were obtained from the US EPA's TCPL pipeline[3, 4] at a sample id (spid)-assay data level (LEVEL 5). Background measurements and assays with >80% noise threshold were excluded. The pipeline uses a hill, gain-loss, or constant model to fit the data and determines noise and activity cutoff thresholds for determining active chemical-assay pairs. The pipeline does not filter data based on poor quality fits, but notes flags when there might be a concern. As is, the pipeline gives 97,424 active sample id-assay ToxCast pairs.

Post pipeline processing. For the purposes of this work, further limiting the activity calls to the best quality fits enabled us to focus on the most promising chemical-assay pairs for prioritization. For the Tox21 data, this included filtering out "marginally active" calls[1, 2] and calls with good quality fits, indicated by a curve class of ± (1.1, 1.2, 2.1, or 2.2).[5] 18,143 active calls by sample id remained. For the ToxCast data, AC_{50} values were first adjusted to the lowest concentration tested when the AC_{50}<lowest concentration tested. This occurred for 3,469 out of 97,424 sample-assay pairs. The ToxCast model fitting flags listed on this set of data were, "#6 Only highest conc above baseline, active", "# 7 Only one conc above baseline, active", "# 8 Multiple points above baseline, inactive", "# 10 Noisy data", "# 11 Borderline active", "# 12 Borderline inactive", "#15 Gain AC50 < lowest conc & loss AC50 < mean conc", "# 16 Hit-call potentially confounded by overfitting", and "# 17 Biochemical assay with < 50% efficacy". We chose to remove data with any flag, except "# 16 Hit-call potentially confounded by overfitting", which indicated that this activity call would get changed to a different model with a small change in the AIC value and "# 17 Biochemical assay with < 50% efficacy", since we would be filtering out <40% efficacy in subsequent steps. The filtered ToxCast data revealed 54,688 active sample id-assay calls. The data were then collapsed to CASRN giving 65,039 total active Tox21/ToxCast CASRN-assay pairs Table S4. Finally, removing low efficacies (<40% or 2-fold) gave 56,135 active CASRN-assay pairs. Additional columns in Table S4 include the US EPA's cytotoxicity z-score concentration range, which is the cytotoxicity concentration range for a given chemical estimated from ~35 different cytotoxicity assays.[6] This z-score is a working hypothesis with several assumptions, including the cytotoxic effect in a given cell type is always at a higher concentration than a non-cytotoxic effect in another cell type and is not dependent on factors such as serum concentration, length of experiment, cell type, and plating conditions. We did not choose to filter the data by this method because we did not want to rule out true activity in this work, but have provided flag columns 07-10 in Tables S4 & S5. Table S4 contains 65,039 chemical-assay pairs (4,706 unique chemicals and 838 unique assays), with 56,135 deemed active by filtering out efficacies<40% or <2-fold, depending on the assay response. The active pairs represent 1.4%

of the possible data. Additionally, 2,302,283 pairs are inactive (58.4%) and 1,585,210 (40.2%) have not been tested. Table S5 contains chemicals within Table S4 and have ExpoCast exposure predictions.

C_{max}/AC_{50} Glucocorticoid Receptor (GR) and Peroxisome Proliferator-Activated Receptor Gamma (PPARγ) Case Studies.

For the GR and PPARγ agonist case studies, all chemicals affecting these pathways without regard to having *in vivo* human C_{max} available can be found in Table S4. The specific GR agonist assays consists of three GR agonist assays: 1)Attagene(ATG); 2)NovaScreen(NVS) cell-free human GR binding; and3)Tox21. Specifically, these assays are: "ATG_GR_TRANS_up", "NVS_NR_hGR", and "tox21-gr-hela-bla-agonist-p1". The specific PPARγ assays used in the manuscript include, "ATG_PPARg_TRANS_up", "tox21-pparg-bla-agonist -p1", "OT_PPARg_PPARgSRC1_0480", "OT_PPARg_PPARgSRC1_1440", and "NVS_NR_hPPARg".

Uncertainty Analysis.

The performance of the random forest model describes how well it can bin the C_{max} values as under-predicted, over-predicted, or within 10-fold. Specifically, the random forest model for binning C_{max} values as under-predicted by >10-fold had 82% accuracy, 82% specificity, 85% sensitivity; as within 10-fold had 66% accuracy, 68% specificity, 66% sensitivity; and as over-predicted by>10-fold had 83% accuracy, 87% specificity, 40% sensitivity.

Environmental Exposure Data.

In the comparison with the ExpoCast median daily dose exposure estimates, the data were limited to 49,789 active CASRN-assay calls, due to not all Tox21 chemicals with a median daily dose exposure estimate. Table S5 gives these data (49,789 active CASRN-assay calls) plus 7,932 active chemical-assay pairs with low efficacies for comparison.

SUPPLEMENTAL RESULTS

C_{max}/AC_{50} Ratios for Two Pharmaceutical Case Studies.

For the GR pathway, 12 and 15 of the 16 known GR modulators were classified as at least 'possible'(C_{max}/AC_{50} ≥0.1) to affect this pathway in humans at therapeutic dosing scenarios using the *in vivo* and *in silico* C_{max} values, respectively(Figure 3A). Prednisone was classified as 'remote' in both cases, but its active metabolite prednisolone was classified as 'likely'. Filtering out efficacies ≤ 40% and using *in vivo* or *in silico* C_{max} values, 10 and 11 additional compounds, respectively, were identified to have 'possible' interactions. These chemicals (e.g., antibiotics and NSAIDs) are known to target various mechanisms, which may be pleiotropic.

REFERENCES

(1) Hsieh, J. H., Accounting Artifacts in High-Throughput Toxicity Assays. *Methods Mol. Biol.* **2016,** *1473*, 143-52.

(2) Hsieh, J. H.; Sedykh, A.; Huang, R.; Xia, M.; Tice, R. R., A Data Analysis Pipeline Accounting for Artifacts in Tox21 Quantitative High-Throughput Screening Assays. *J. Biomol. Screen.* **2015,** *20* (7), 887-97.

(3) Filer, D.; Kothiya, P.; Setzer, R. W.; Judson, R.; Martin, M. T. The ToxCast Analysis Pipeline: An R Package for Processing And Modeling Chemical Screening Data. https://www.epa.gov/sites/production/files/2015-08/documents/pipeline_overview.pdf

(4) Filer, D. L.; Kothiya, P.; Setzer, R. W.; Judson, R. S.; Martin, M. T., tcpl: the ToxCast pipeline for high-throughput screening data. *Bioinformatics* **2017,** *33* (4), 618-620.

(5) Huang, R.; Xia, M.; Cho, M. H.; Sakamuru, S.; Shinn, P.; Houck, K. A.; Dix, D. J.; Judson, R. S.; Witt, K. L.; Kavlock, R. J.; Tice, R. R.; Austin, C. P., Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. *Environ. Health Perspect.* **2011,** *119* (8), 1142-8.
(6) Judson, R. S.; Magpantay, F. M.; Chickarmane, V.; Haskell, C.; Tania, N.; Taylor, J.; Xia, M.; Huang, R.; Rotroff, D. M.; Filer, D. L.; Houck, K. A.; Martin, M. T.; Sipes, N.; Richard, A. M.; Mansouri, K.; Setzer, R. W.; Knudsen, T. B.; Crofton, K. M.; Thomas, R. S., Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor. *Toxicol. Sci.* **2015,** *148* (1), 137-54.

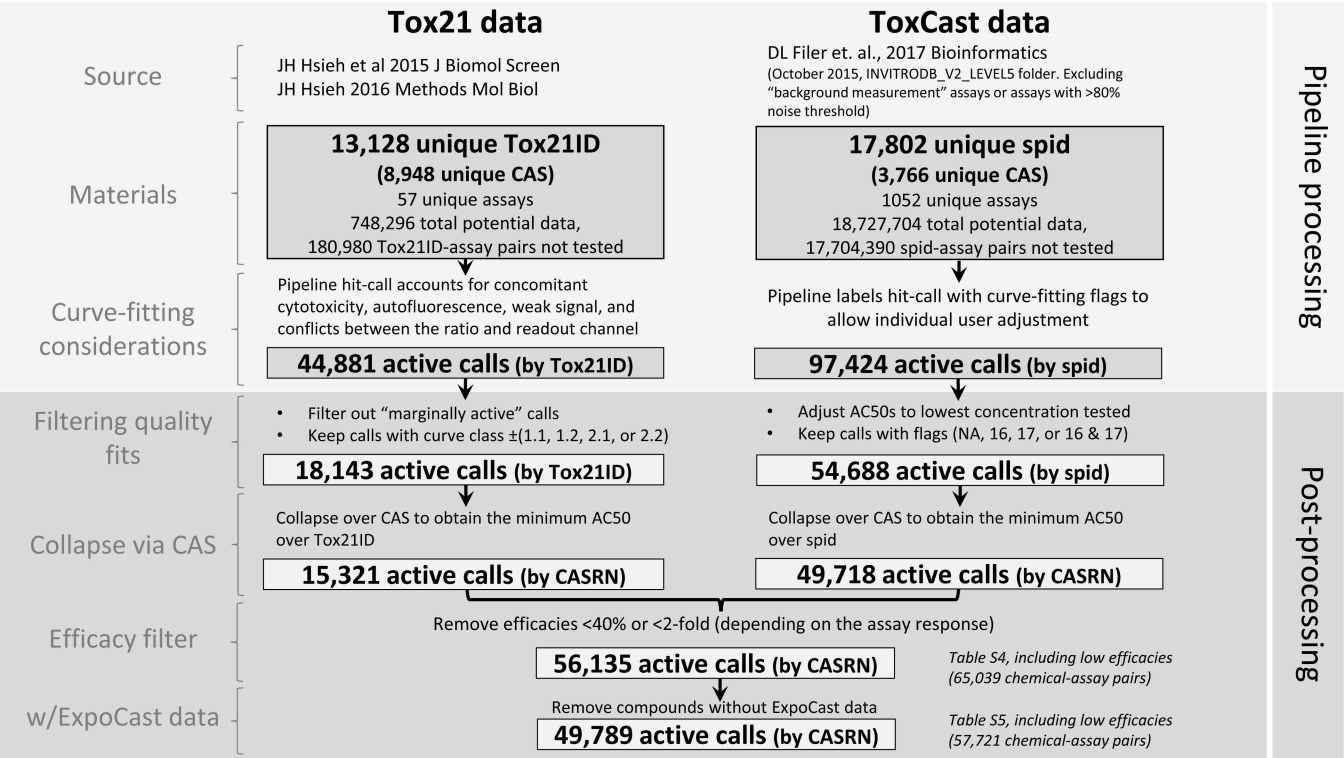

Figure S1. Detailed Tox21/ToxCast data processing workflow.

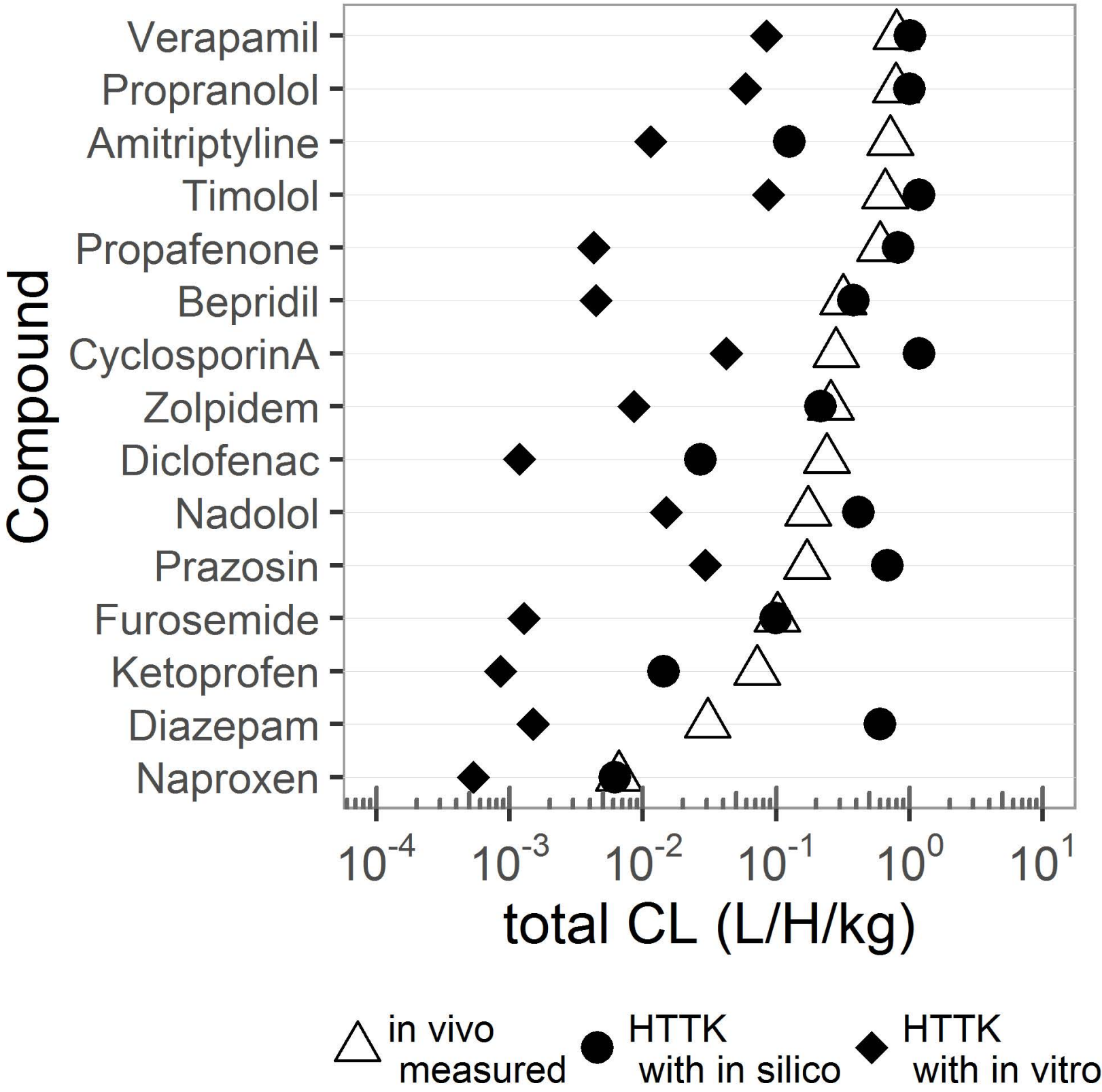

Figure S2. Total clearance prediction comparisons for chemicals with >10x difference between *in silico* and *in vitro*.

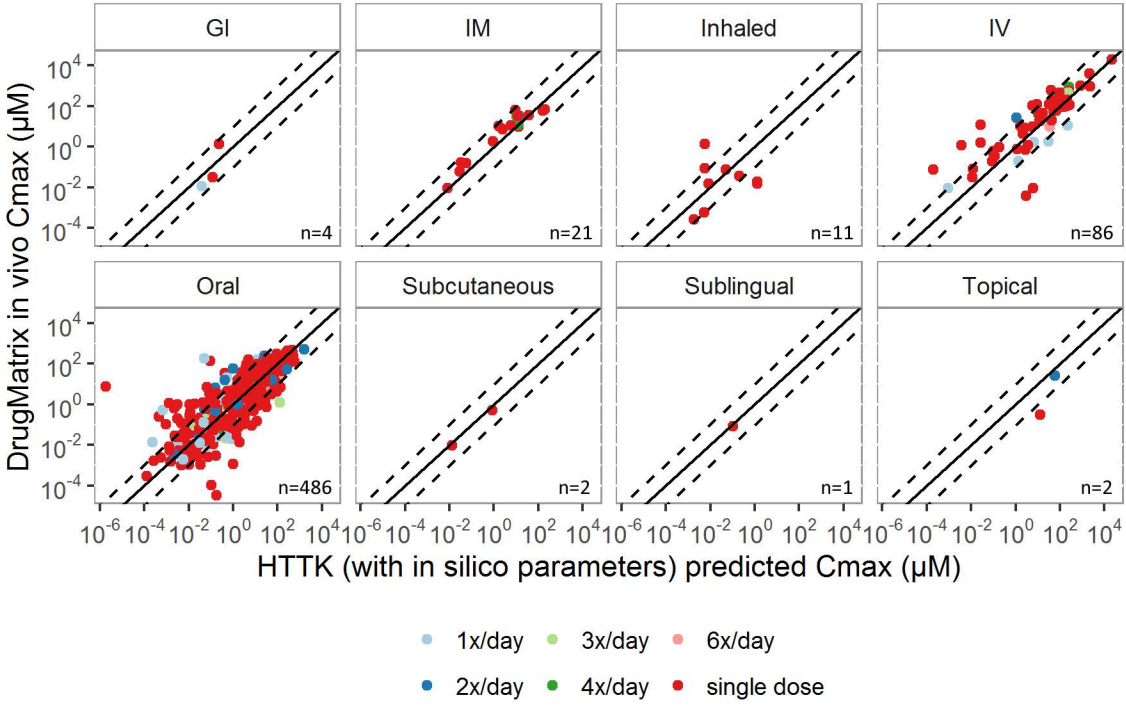

Figure S3. Maximum human plasma concentration prediction comparison over administration route. Separated by administration route and colored by dose frequency. GI=gastro-intestinal, IM=intramuscular, IV=intravenous. Solid line is 1:1, dotted lines are 1 log10 difference.

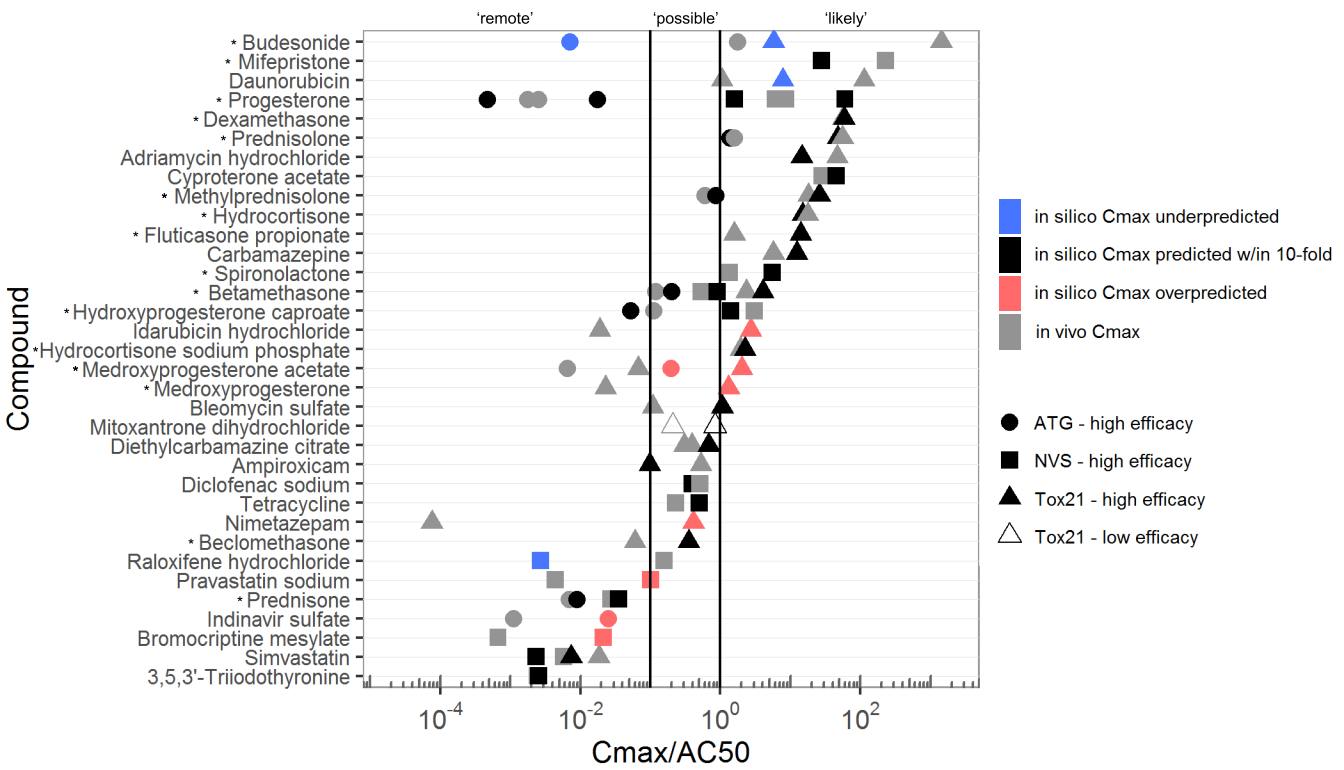

Figure S4: Cmax/AC50 ratios for the GR pathway at pharmacological doses. Compounds with in vivo dosing scenarios and human Cmax values as well as in vitro AC50 values in GR Tox21/ToxCast assays were evaluated for likelihood of in vivo interactions using Cmax/AC50 ratios. For 34 active compounds in GR assays, Cmax /AC50 ratios were plotted using the in vivo Cmax from DrugMatrix and in silico estimated Cmax using the 3-compartment model within the HTTK-package, based on therapeutic external dose and assuming a 70kg adult. Assays included: 1) Attagene (ATG), 2) NovaScreen (NVS) cell-free human GR binding, 3) Tox21. Activity across multiple assays are indicated by multiple data points on the same row. Colored data points indicate the accuracy of in silico predicted Cmax values (blue, black, red) vs. in vivo measured Cmax value (grey). Vertical lines at 0.1 and 1 show regions of 'remote' (<0.1), 'possible' (0.1<X<1) and 'likely' (>1) in vivo interaction. Efficacies greater than 40% or 2-fold change are considered high efficacy, and all others are considered low efficacy. (*) indicate known modulators.

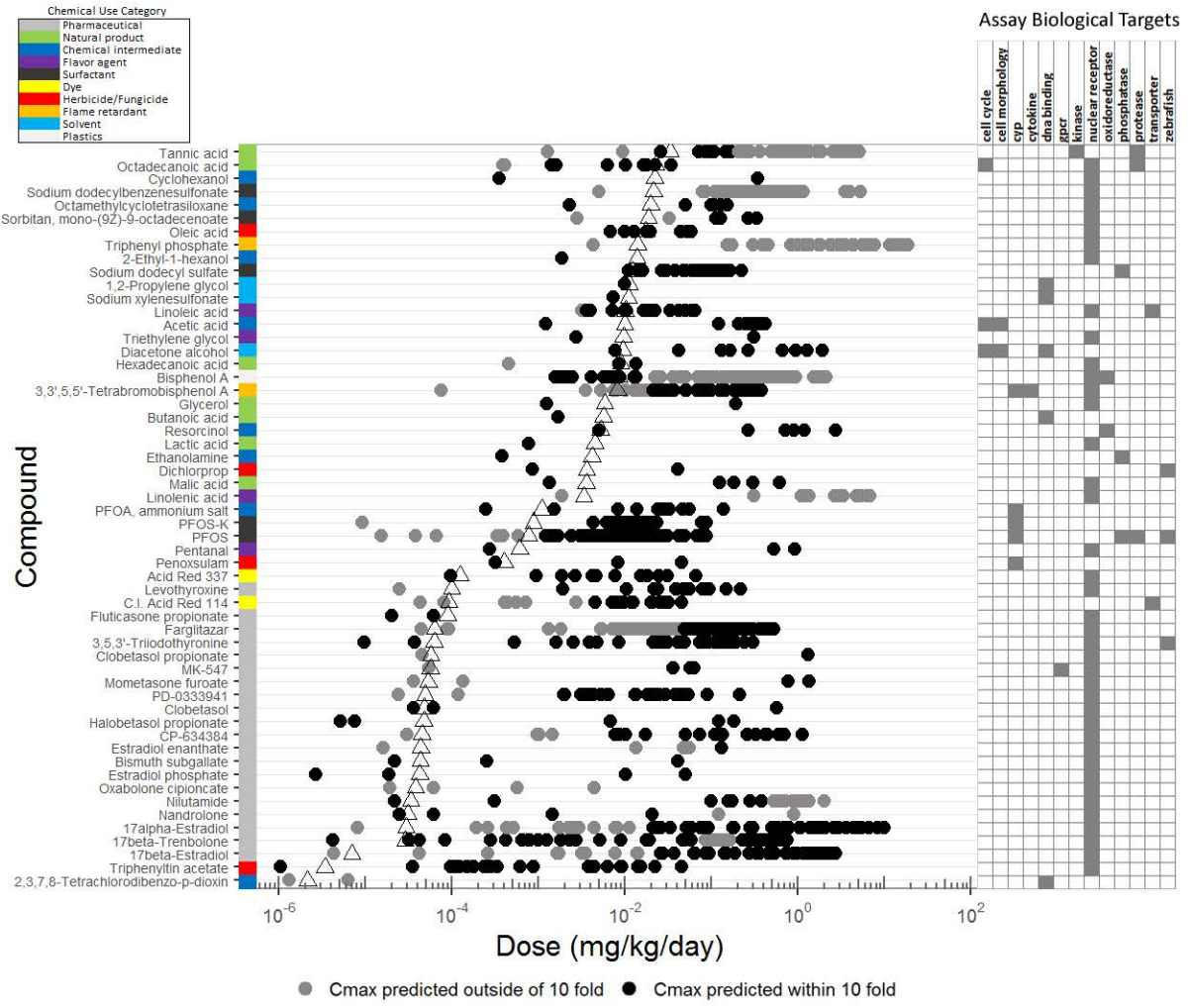

Figure S5. Doses of compounds eliciting a 'possible' human *in vivo* interaction (Cmax/AC50>0.1) which are lower than estimated daily exposure. AC50's for active Tox21/ToxCast compounds with efficacies >40% or 2-fold in the HTS assays were converted to equivalent human *in vivo* doses using exposure scenarios where *in vivo* Cmax concentrations were equal to 0.1*AC50 to represent 'possible' *in vivo* interactions. Exposure estimates in terms of dose per day were calculated from a model built using NHANES biomonitoring data and physicochemical properties and use categories[21](open triangles). Predicted doses for chemical-assay pairs are shown for all active chemical-assay pairs (dots) for chemicals with predicted doses below exposure estimates. Biological targets for compounds where doses are lower than estimated exposures (i.e., data points to the left of the triangles) are shown in the right panel. Gray dots indicate lower confidence and black dots indicate higher confidence in calculating the doses from the estimated Cmax values based on the random forest model.