Supporting Information

Trust, But Verify II: A Practical Guide to Chemogenomics Data Curation

Denis Fourches[1,*], Eugene Muratov[2], & Alexander Tropsha[2,*]

[1] *Department of Chemistry, Bioinformatics Research Center, North Carolina State University, Raleigh, NC, 27695, USA.*

[2] *Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA.*

*please address the correspondence to these authors; emails: dfourch@ncsu.edu or alex_tropsha@unc.edu

Table S1. Distribution of 2D structural duplicates for the NCGC CYP dataset. Out of 1,280 duplicate couples identified among the 17,000 screened chemicals, 874 pairs of chemicals were reported with biological profile differences (ΔlogAC50 $\geq$ 1). A total of 1,535 discrepancies were found for the 874 couples of duplicates.

	CYP2C9	CYP1A2	CYP3A4	CYP2D6	CYP2C19
# of discrepancies	154	363	426	422	170