

SUPPORTING INFORMATION

In Silico Models to Predict the Perturbation of Molecular Initiating Events Related to Thyroid Hormone Homeostasis

*Marina Garcia de Lomana,^{1,2} Andreas Georg Weber,¹ Barbara Birk,¹ Robert Landsiedel,¹
Janosch Achenbach,^{1,†} Klaus-Juergen Schleifer,¹ Miriam Mathea^{1*} and Johannes Kirchmair^{2*}*

¹ BASF SE, 67063 Ludwigshafen am Rhein, Germany

² Department of Pharmaceutical Chemistry, Faculty of Life Sciences, University of Vienna, 1090
Vienna, Austria

* miriam.mathea@basf.com; Tel.: +49 621 60-29054;

johannes.kirchmair@univie.ac.at; Tel.: +43 1-4277-55104

Table S1. Flags Available in the ToxCast Database for Tagging Potential Errors in Class Labeling.

ToxCast flags

Only one concentration above baseline (active)

Multiple points above baseline (inactive)

Noisy data

Borderline active

Borderline inactive

Gain AC_{50} lower than the lowest concentration and loss AC_{50} lower than the mean concentration

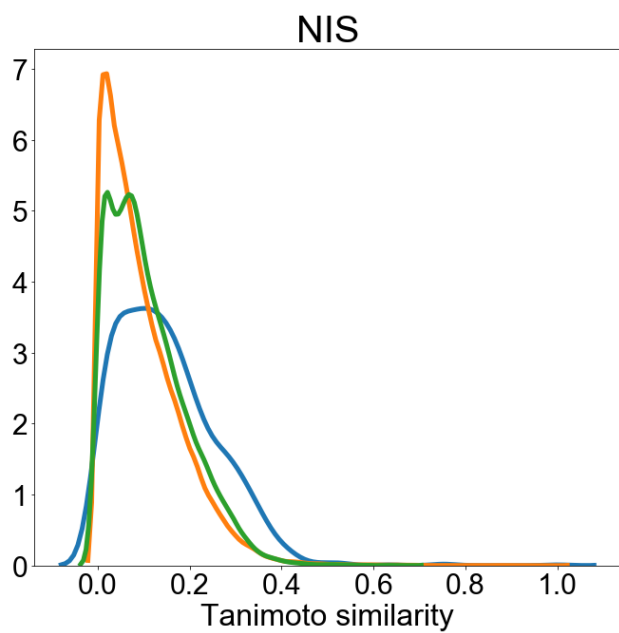
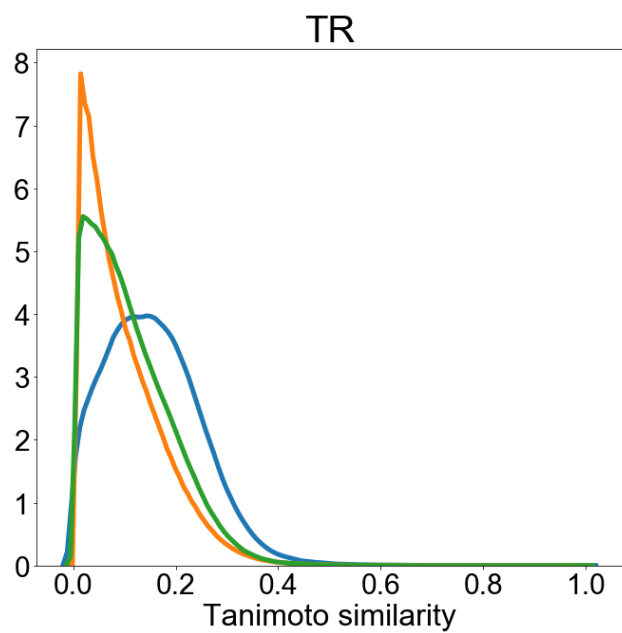
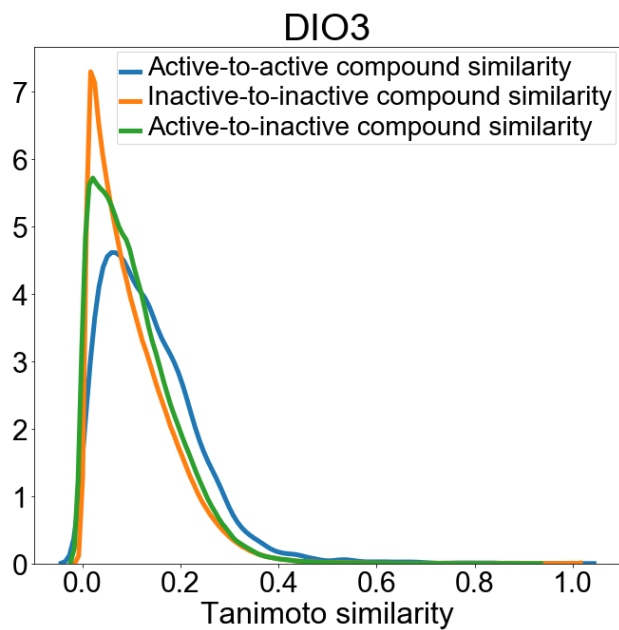
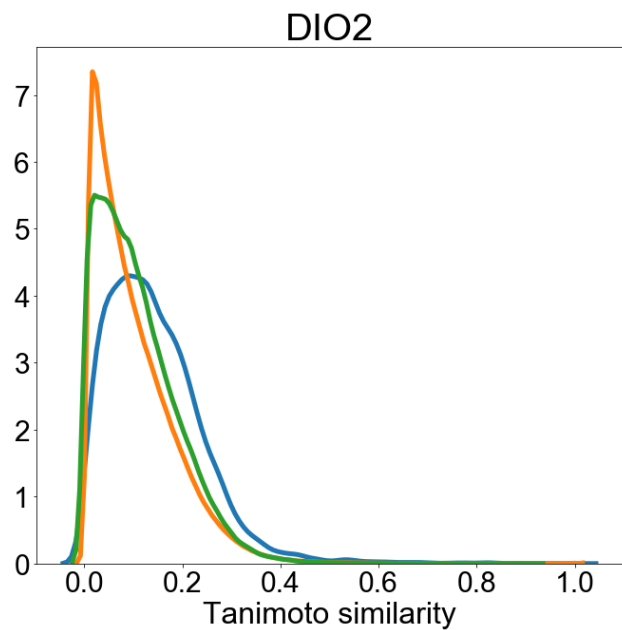
Gain AC_{50} lower than the lowest concentration and loss AC_{50} lower than the mean concentration

Hit-call potentially confounded by overfitting

Biochemical assay with less than 50% efficacy

Table S2. List of Molecular Descriptors Used in Principal Component Analysis.

Descriptors
SlogP
TPSA
ExactMW
NumLipinskiHBA
NumLipinskiHBD
NumRotatableBonds
NumHBD
NumHBA
NumAmideBonds
NumHeteroAtoms
NumHeavyAtoms
NumAtoms
NumStereocenters
NumUnspecifiedStereocenters
NumRings
NumAromaticRings
NumSaturatedRings
NumAliphaticRings
NumAromaticHeterocycles
NumSaturatedHeterocycles
NumAliphaticHeterocycles
NumAromaticCarbocycles
NumSaturatedCarbocycles
NumAliphaticCarbocycles
FractionCSP3



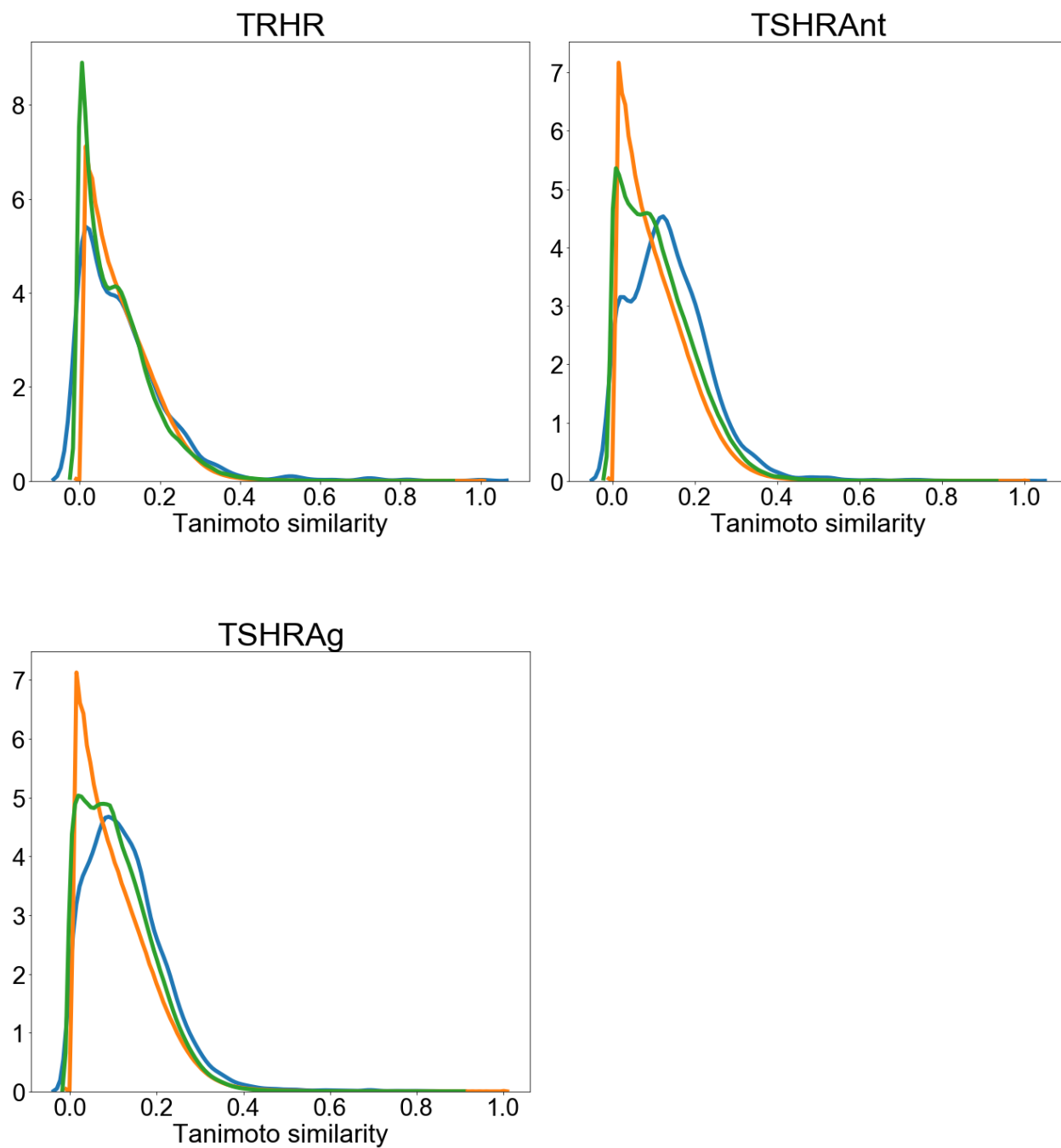
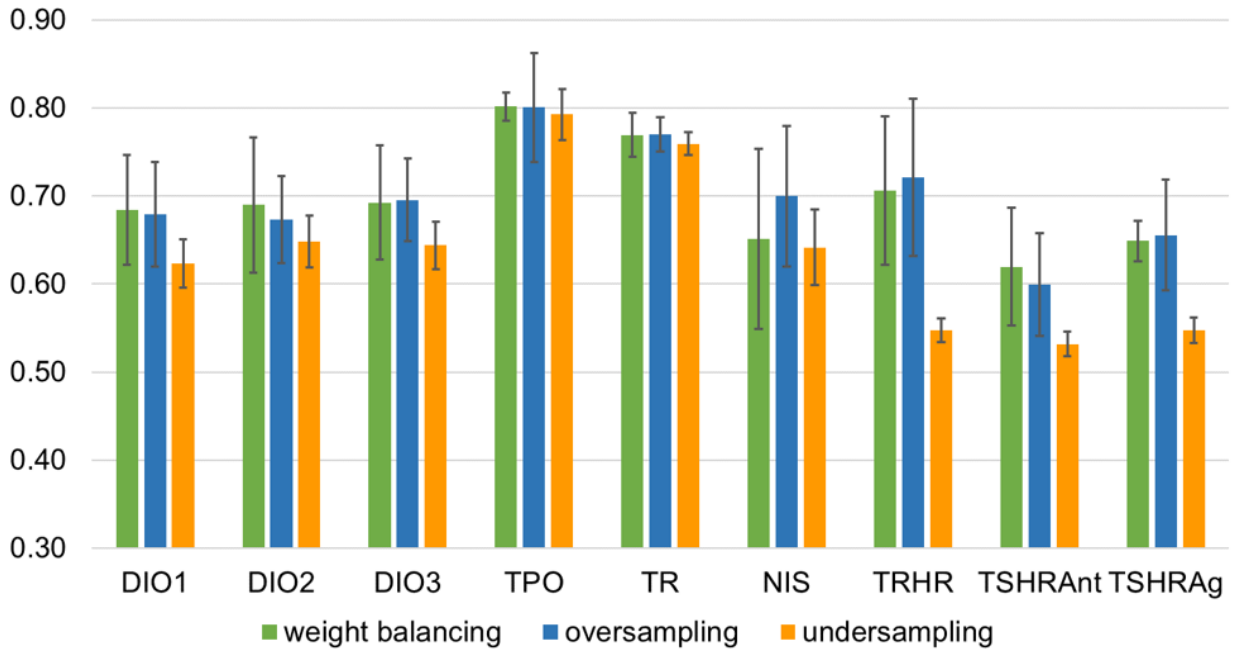


Figure S1. Distribution of pairwise Tanimoto similarities based on atom-pair fingerprints, for DIO2, DIO3, TR, NIS, TRHR, TSHRAg and TSHRant and three types of compound pairs: a) active-to-active, b) inactive-to-inactive and c) active-to-inactive.

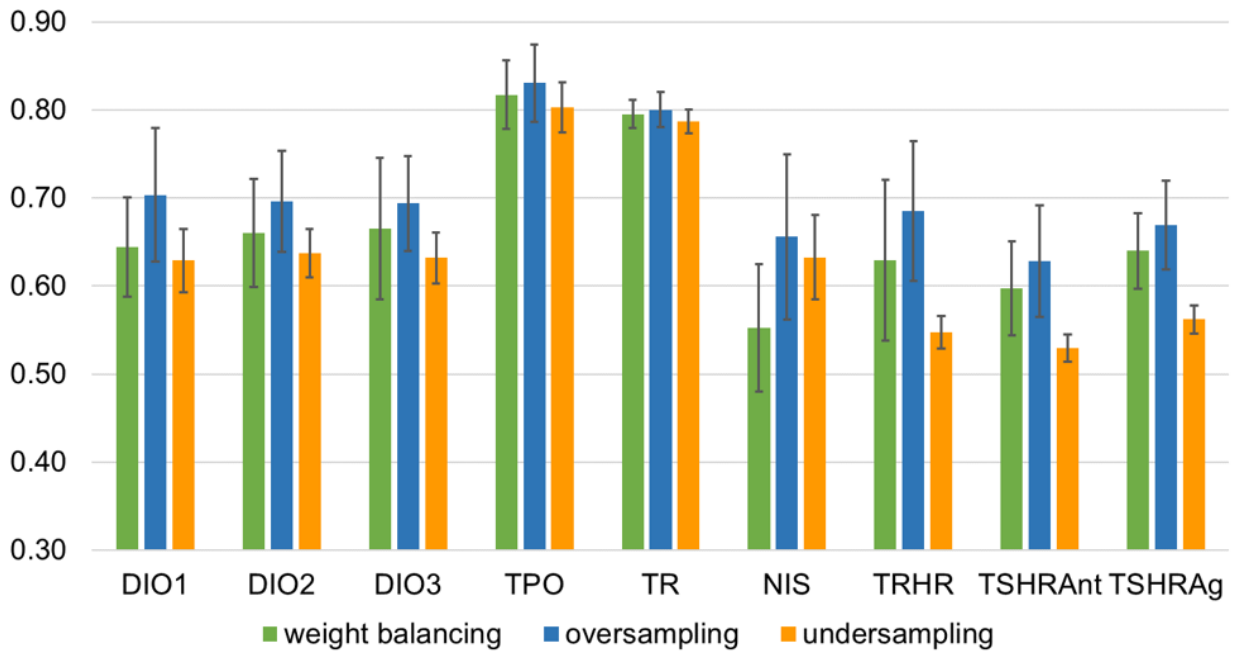
a)

F1 score LR



b)

F1 score XGB



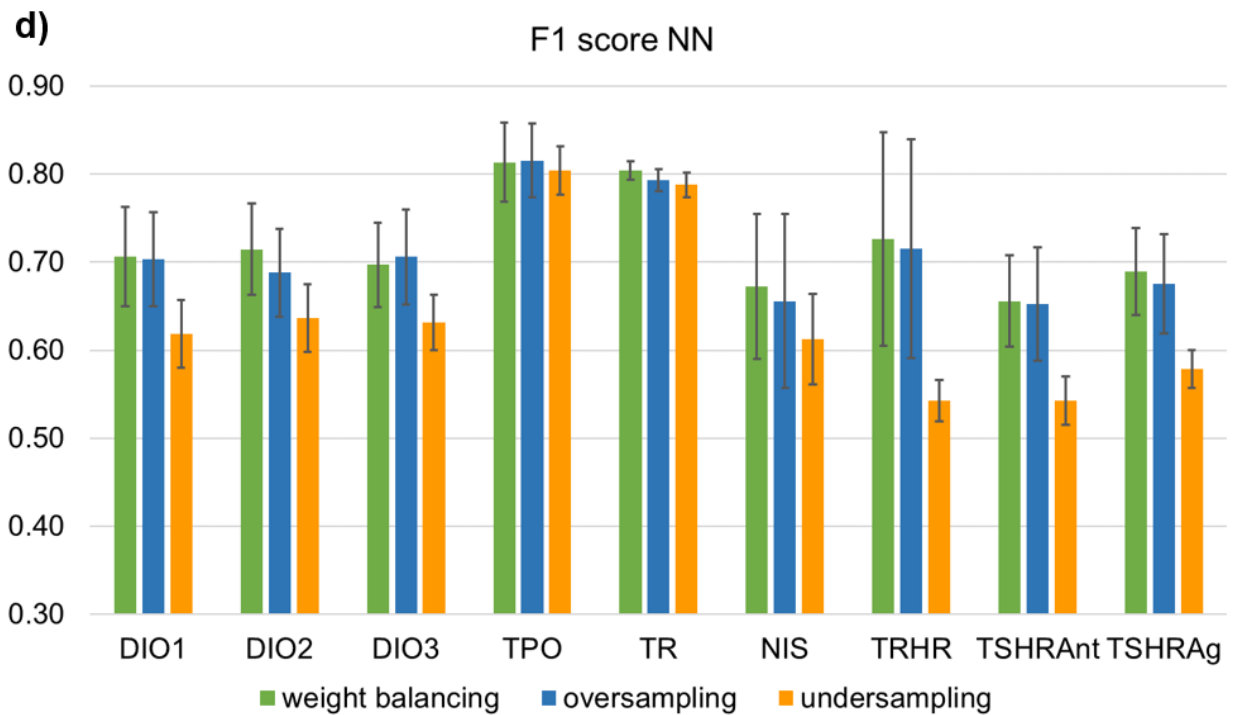
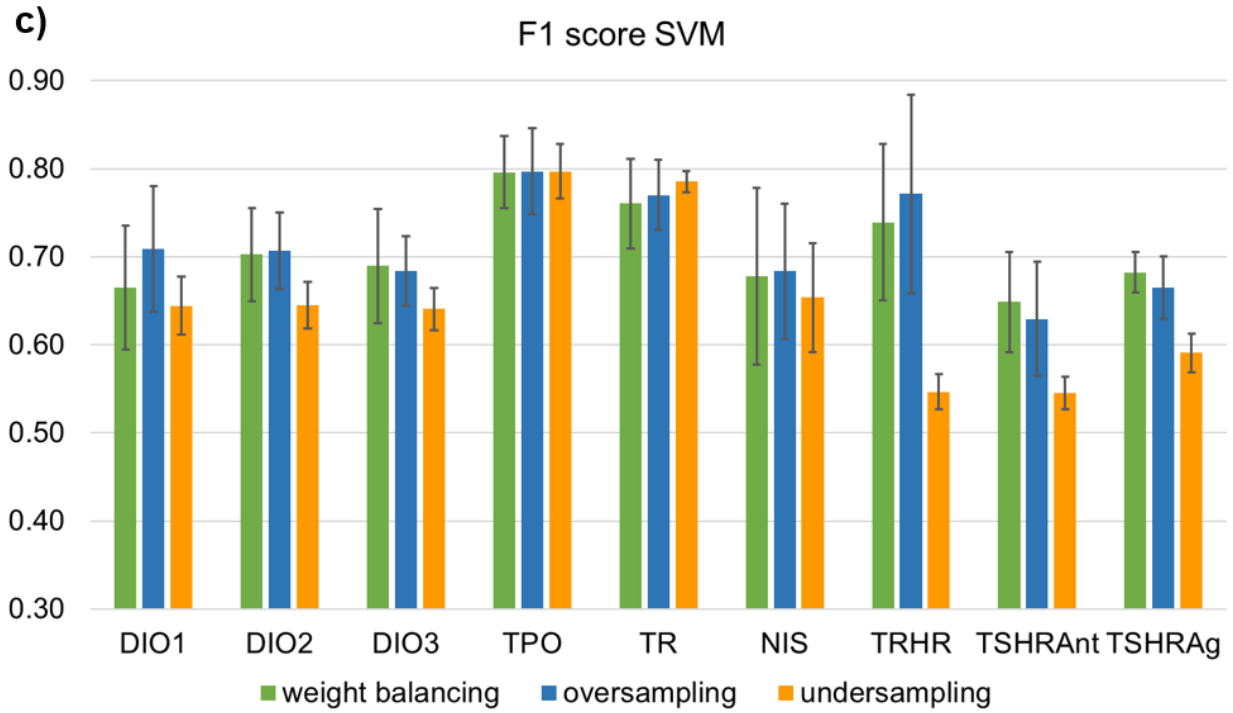


Figure S2. Comparison of the mean F1 score obtained for the nine thyroid end points with (a) LR, (b) XGB, (c) SVM and (d) NN in combination with the three data sampling techniques.