Supplemental Information for

# MetSim: Integrated Programmatic Access and Pathway Management for Xenobiotic Metabolism Simulators

Louis Groff[†], Antony Williams[†], Imran Shah[†], Grace Patlewicz[†*]

[†]Center for Computational Toxicology and Exposure (CCTE), Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

ORCiDs:

LG: https://orcid.org/0000-0002-6357-0508
AW: https://orcid.org/0000-0002-2668-4821
IS: https://orcid.org/0000-0003-0808-0140
GP: https://orcid.org/0000-0003-3863-9689

*__Correspondence to be addressed to__:

Grace Patlewicz
Center for Computational Toxicology and Exposure (CCTE)
Office of Research and Development
U.S. Environmental Protection Agency
109 TW Alexander Drive
Research Triangle Park, NC 27711
Telephone: (919) 541-1540
E-mail: patlewicz.grace@epa.gov.

**List of Supplemental Material**

Figure S6: Average log($K_{ow}$) values queried from the from the Toolbox API (right, negative values in blue shades, positive values in red shades, zero is white, empty cells returned no log($K_{ow}$) data) sorted with the chemical class and recall rate hierarchical clustering analysis results from the Drug Dataset (left).

Table S1. Comparison of performance metrics for BioTransformer 2019 validation dataset to current BioTransformer predictions for human tissue metabolism in absence of gut microbiome predictions.

Table S2. Comparison of performance metrics for BioTransformer validation dataset to current BioTransformer predictions for single-step Phase I metabolism predictions.

Table S3. Comparison of TIMES performance metrics against training set observed maps with parameters set to match current prediction parameters.

Table S4. Predictive performance characteristics for the Phase I metabolite relationships in the Literature Dataset.

Table S5. Predictive performance characteristics for the Phase I and Phase II metabolite relationships in the SMPDB Dataset.

Table S6. Spearman Correlation analysis of log($K_{ow}$) on Recall Rate with P-values.

Supplementary File 1. MetSim_Tool_Details.xlsx. Data table of the metabolism simulators considered, their availability, implementation language, operating system, chemical input format, metabolite outputs, model parameter details and options, batch execution options, and tool execution times per chemical.

Supplementary File 2. MetSim_Performance_Tables.xlsx Input data for parent chemicals (Chemical name, QSAR-Ready SMILES, InChIKey, CASRN), ClassyFire chemical class assignment, Log Kow values, Performance metrics for each tool (True positives, false negatives, false positives, recall).

Supplementary File 3. JSON file of observed metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

Supplementary File 4. JSON file of BioTransformer predicted metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

Supplementary File 5. JSON file of CTS predicted metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

Supplementary File 6. JSON file of TIMES In Vitro Rat Liver S9 predicted metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

Supplementary File 7. JSON file of TIMES In Vivo Rat Simulator predicted metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

Supplementary File 8. JSON file of Toolbox API In Vitro Rat Liver S9 predicted metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

Supplementary File 9. JSON file of Toolbox API In Vivo Rat Simulator predicted metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

Supplementary File 10. JSON file of aggregate model predicted metabolism pathways for the 112 parent chemicals in the Drug Dataset, in MetSim format.

**Performance Analyses on Individual Datasets**

*Literature Dataset Performance*

A summary of the performance metrics calculated from the predictions from each MetSim tool on the Literature Dataset of 59 drug and NSAID parent chemicals and their 179 reported phase I metabolites, is given in Table S4.

BioTransformer predictions on the Literature Dataset using b.ec.1.cyp450.2.phase2.1 generated a total of 11202 metabolite predictions. Of those predictions, 111 TP predictions of the 179 total reported metabolites were correctly identified. Conversely, there were 68 FN predictions of the 179 reported metabolites, yielding a recall of 0.62. There were a total of 11091 FP predictions, yielding a precision of 0.010.

TIMES predictions on Drug Dataset were generated via the tm.vitro_rat.6 and tm.vivo_rat.6 models, where tm.vitro_rat.6 generated a total of 662 metabolite predictions. Of those predictions, 83 TP predictions of the 179 total reported metabolites were correctly identified. Conversely, there were 96 FN predictions of the 179 reported metabolites, yielding a recall of 0.46. There were a total of 579 FP predictions, yielding a precision of 0.125. The tm.vivo_rat.6 model generated a total of 625 metabolite predictions. Of those predictions, 75 TP predictions of the 179 total reported metabolites were correctly identified. Conversely, there were 104 FN predictions of the 179 reported metabolites, yielding a recall of 0.42, which was the lowest individual recall rate across all four tools. There were a total of 550 FP predictions, yielding a precision of 0.120.

Toolbox API predictions on the Drug Dataset were generated from the tb.vitro_rat.3 and tb.vivo_rat.6 models, where the tb.vitro_rat.3 model generated a total of 539 metabolite predictions. Of those predictions, 93 TP predictions of the 179 total reported metabolites were correctly identified. Conversely, there were 86 FN predictions of the 179 reported metabolites, yielding a recall of 0.52. There were a total of 446 FP predictions, yielding a precision of 0.173. The tb.vivo_rat.6 model generated a total of 623 metabolite predictions. Of those predictions, 86 TP predictions of the 179 total reported metabolites were correctly identified. Conversely, there were 93 FN predictions of the 179 reported metabolites, yielding a recall of 0.48. There were a total of 537 FP predictions, yielding a precision of 0.138.

CTS predictions on the Drug Dataset using cts.chemaxon.3 generated a total of 5713 metabolite predictions. Of those predictions, 131 TP predictions of the 179 total reported metabolites were correctly identified. Conversely, there were 48 FN predictions of the 179 reported metabolites, yielding a recall of 0.73, which was the highest individual recall rate across all four tools. There were a total of 5582 FP predictions, yielding a precision of 0.023.

When the predictions from all four of the tools are aggregated together, a total of 15497 unique metabolite predictions are generated. Of those predictions, 154 TP predictions of the 179 total reported metabolites were correctly identified. Conversely, there were 25 FN predictions of the 179 reported metabolites, yielding a recall of 0.86, which yields a 13% increase in recall rate compared to CTS, the highest performing individual tool. There were a total of 15343 FP predictions, yielding a precision of 0.010.

*Literature Dataset Clustering Analysis*

Hierarchical clustering analysis of recall against the chemical classes within both datasets was performed and compared across all tools. The result of this analysis for the Literature Dataset is shown in Figure S4. The highest recall rates across all tools were observed for phenanthrenes, diazines, phenylpropanoic acids, azoles, lactams, benzothiazenes, steroids, and piperadines, comprising ~17% of the dataset (10 compounds), where most tools yield high recall rates. Conversely, most tools performed poorly on nucleoside and nucleotide analogues, carboxylic acids, pyrrolines, and organonitrogen compounds, comprising ~12% of the dataset (7 compounds). The remaining ~71% of the dataset (42 compounds) varied greatly in their recall rates, depending on the choice of tool and model. However, in all cases, the aggregated recall rate over all models yielded either matches the highest recall rate among available tools, or improved recall compared across all tools. It is additionally worth noting that, of the 17/28 ClassyFire assigned chemical classes where a mean recall rate of one was achieved for that chemical class, all but one case are classes that account for a single parent chemical in the dataset of 59. The singular case is indoles, which account for 2/59 of the parent chemicals. Thus, recall rates equal to one were attained for ~31% of the dataset, whether by individual models, or the aggregation of all model predictions. Examples of consensus recall rate improvement compared to the highest individual model recall rate for specific chemical class groupings include pyrroles (+20% recall), pyrrolines (+20% recall),

benzodiazepines (+17% recall), benzene and substituted derivatives (+6% recall), organooxygen compounds (+7% recall), and naphthalenes (+5% recall).

*SMPDB Dataset*

A summary of the performance metrics calculated from the predictions from each MetSim tool on the SMPDB Dataset of 59 Drug parent chemicals and their 259 reported phase I and phase II metabolites, is given in Table S5.

BioTransformer predictions on the Literature Dataset using bt.ec.1.cyp450.2.phase2.1 generated a total of 16534 metabolite predictions. Of those predictions, 110 TP predictions of the 259 total reported metabolites were correctly identified. Conversely, there were 149 FN predictions of the 259 reported metabolites, yielding a recall of 0.42. There were a total of 16424 FP predictions, yielding a precision of 0.007.

TIMES predictions on Drug Dataset were generated via the tm.vitro_rat.6 and tm.vivo_rat.6 models, where tm.vitro_rat.6 generated a total of 734 metabolite predictions. Of those predictions, 93 TP predictions of the 259 total reported metabolites were correctly identified. Conversely, there were 166 FN predictions of the 259 reported metabolites, yielding a recall of 0.36. There were a total of 641 FP predictions, yielding a precision of 0.127. The tm.vivo_rat.6 model generated a total of 721 metabolite predictions. Of those predictions, 103 TP predictions of the 259 total reported metabolites were correctly identified. Conversely, there were 156 FN predictions of the 259 reported metabolites, yielding a recall of 0.40. There were a total of 618 FP predictions, yielding a precision of 0.143.

Toolbox API predictions on the Drug Dataset were generated from the tb.vitro_rat.3 and tb.vivo_rat.6 models, where the tb.vitro_rat.3 model generated a total of 712 metabolite predictions. Of those predictions, 86 TP predictions of the 259 total reported metabolites were correctly identified. Conversely, there were 173 FN predictions of the 259 reported metabolites, yielding a recall of 0.33, which was the lowest individual recall rate across all four tools. There were a total of 626 FP predictions, yielding a precision of 0.121. The tb.vivo_rat.6 model generated a total of 916 metabolite predictions. Of those predictions, 95 TP predictions of the 259 total reported metabolites were correctly identified. Conversely, there were 164 FN predictions of the

259 reported metabolites, yielding a recall of 0.37. There were a total of 821 FP predictions, yielding a precision of 0.104.

CTS predictions on the Drug Dataset using cts.chemaxon.3 generated a total of 8248 metabolite predictions. Of those predictions, 111 TP predictions of the 259 total reported metabolites were correctly identified. Conversely, there were 148 FN predictions of the 259 reported metabolites, yielding a recall of 0.43, which was the highest individual recall rate across all four tools. There were a total of 8137 FP predictions, yielding a precision of 0.013.

When the predictions from all four of the tools are aggregated together, a total of 23821 unique metabolite predictions are generated. Of those predictions, 166 TP predictions of the 259 total reported metabolites were correctly identified. Conversely, there were 93 FN predictions of the 259 reported metabolites, yielding a recall of 0.64, which yields a 21% increase in recall rate compared to CTS, the highest performing individual tool. There were a total of 23655 FP predictions, yielding a precision of 0.007.

*SMPDB Dataset Clustering Analysis*

Hierarchical clustering analysis of recall against the chemical classes within both datasets was performed and compared across all tools. The result of this analysis for the SMPDB Dataset is shown in Figure S5. The highest recall rates across most tools were observed for carboxylic acids, phenylpropanoic acids, and morphinans, comprising ~26% of the dataset (15 compounds). Conversely, most tools performed poorly on nucleoside and nucleotide analogues, fatty acyls, benzimidazoles, 5'-deoxyribonucleosides, imidazopyramidines, and diazines, comprising ~20% of the dataset (12 compounds). The remaining ~54% of the dataset (32 compounds) varied greatly in their recall rates, depending on the choice of tool and model. However, in all cases, the aggregated recall rate over all models yielded either matches the highest recall rate among available tools, or improved recall compared across all tools. It is additionally worth noting that, of the 8/27 ClassyFire assigned chemical classes where a mean recall rate of one was achieved for that chemical class, all but one case are classes that account for a single parent chemical in the dataset of 59. The singular case is morphinans, which account for 3/59 of the parent chemicals. Thus, recall rates equal to one were attained for ~17% of the dataset, whether by individual models, or the aggregation of all model predictions. Examples of consensus recall rate improvement

compared to the highest individual model recall rate for specific chemical class groupings include pyridines (+36% recall), anthracyclines (+33% recall), phenol ethers (+16% recall), fatty acyls (+13% recall), benzazepines (+11% recall), stilbenes (+6% recall), organonitrogen compounds (+5% recall), and carboxylic acids (+4% recall).

Figure S1. Truncated YAML hierarchically structured output for Aripiprazole using results from TIMES In Vivo Rat Simulator model.

aripiprazole.yaml

```yaml
datetime: 2022-04-25_16h56m25s
params:
  depth: 3
  model:
  - In Vivo Rat Simulator v.08.14
  organism: Rat
  site_of_metabolism: false
software: OASIS TIMES
version: 2.31.2.82
input:
  casrn: 129722-12-9
  chem_name: Aripiprazole
  dtxsid: DTXSID3046083
  hcd_smiles: O=C1CCC2=CC=C(C=C2N1)OCCCCN1CCN(CC1)C1=CC=CC(Cl)=C1Cl
  inchikey: CEUORZQYGODEFX-HXTKINSTNA-N
  smiles: O=C1NC2C(=CC=C(C=2)OCCCCN2CCN(C3C(Cl)=C(Cl)C=CC=3)CC2)CC1
output:
- precursor:
    casrn: 129722-12-9
    chem_name: Aripiprazole
    dtxsid: DTXSID3046083
    hcd_smiles: O=C1CCC2=CC=C(C=C2N1)OCCCCN1CCN(CC1)C1=CC=CC(Cl)=C1Cl
    inchikey: CEUORZQYGODEFX-HXTKINSTNA-N
    smiles: O=C1NC2C(=CC=C(C=2)OCCCCN2CCN(C3C(Cl)=C(Cl)C=CC=3)CC2)CC1
  successors:
  - enzyme: '[phase I]'
    generation: 1
    mechanism: Oxidative O-Dealkylation|Alkylphenyl Ether Oxidative O-Dealkylation
    metabolite:
      casrn: 22246-18-0
      chem_name: 7-Hydroxy-3,4-dihydroquinolin-2(1H)-one
      dtxsid: DTXSID70382941
      hcd_smiles: OC1=CC=C2CCC(=O)NC2=C1
      inchikey: LKLSFDWYIBUGNT-KZFATGLANA-N
      smiles: OC1C=C2C(CCC(N2)=O)=CC=1
  - enzyme: '[phase I]'
    generation: 1
    mechanism: Aromatic C-Hydroxylation|Aromatic Amine C-Hydroxylation
    metabolite:
      casrn: null
      chem_name: null
      dtxsid: null
      hcd_smiles: OC1=CC2CCC(=O)NC=2C=C1OCCCCN1CCN(CC1)C1=CC=CC(Cl)=C1Cl
      inchikey: ANUISDKMSVXQOY-HXTKINSTNA-N
      smiles: OC1C(OCCCCN2CCN(C3C(Cl)=C(Cl)C=CC=3)CC2)=CC2=C(CCC(N2)=O)C=1
```

Figure S2. Default TIMES model parameters for the In Vitro Rat Liver S9 model.

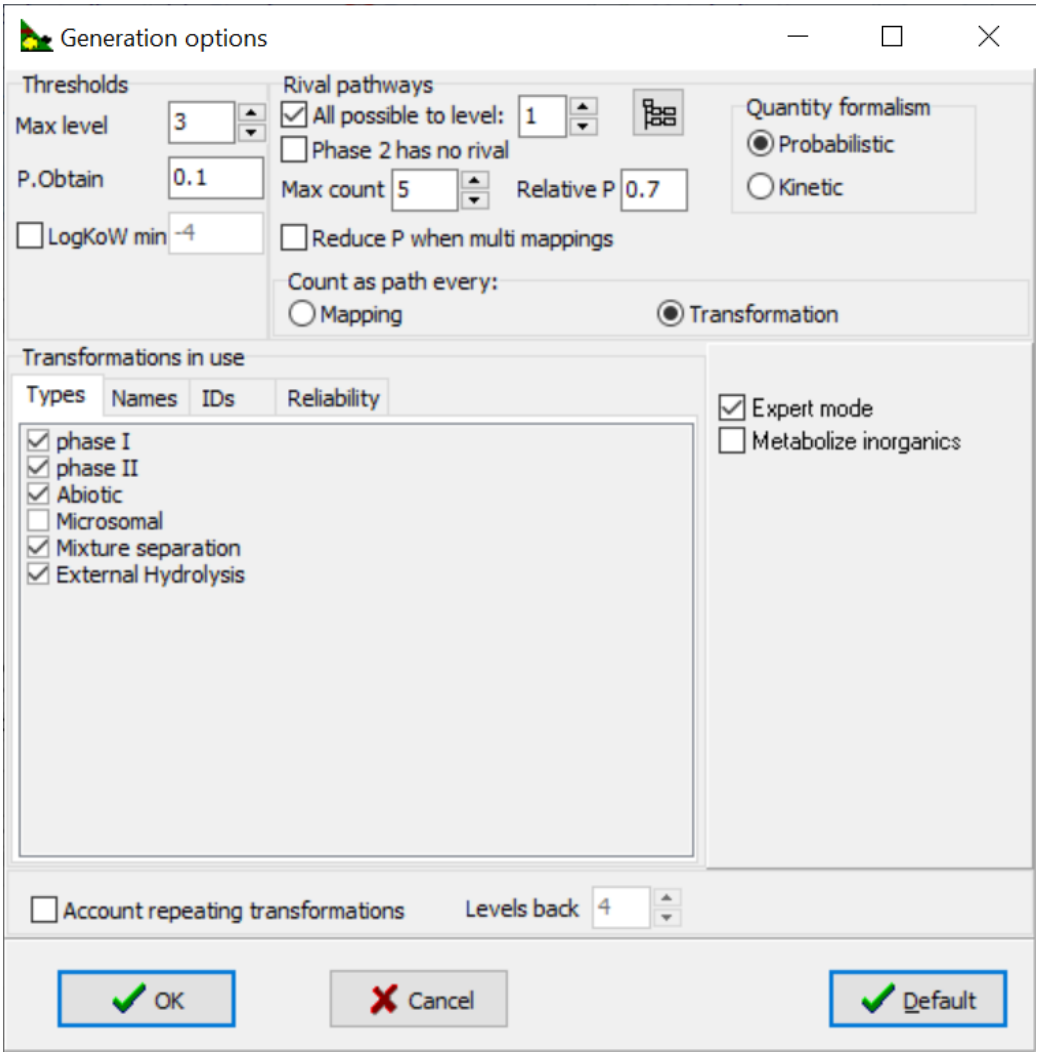
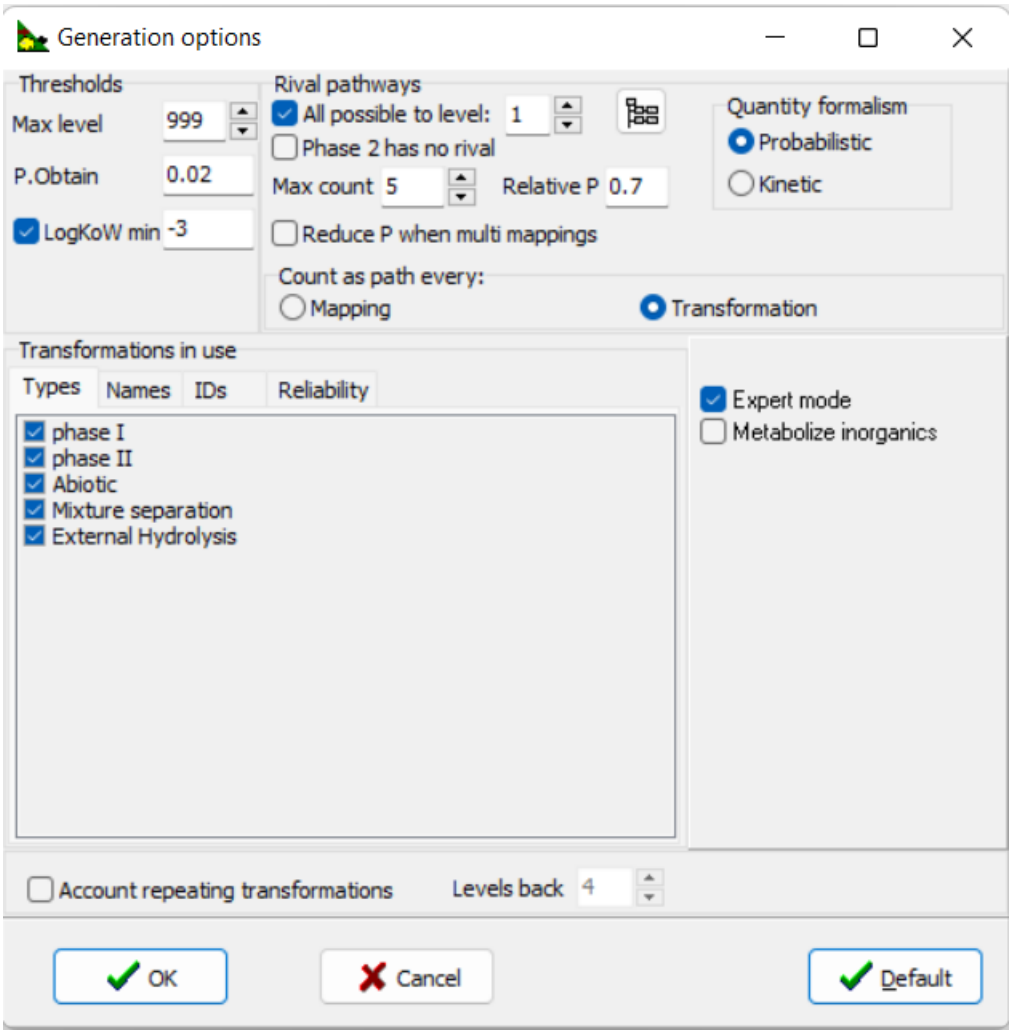


Figure S3. Default TIMES model parameters for In Vivo Rat Simulator model.




Parameter adjustments in TIMES take place after loading the desired metabolism simulator (either In Vivo Rat Simulator or In Vitro Rat Liver S9). Within the tab corresponding to the simulator, input parameters are accessed via the "Metabolization" button within the Model Options toolbar. In this study, both metabolism simulators are set to the same Max Level of 6, with the probability of obtaining a metabolite (P. Obtain) thresholded at 0.1. For both simulators, the "Max count" parameter that thresholds the maximum number of metabolites per cycle was increased to 10 per cycle. All other parameters were left at their default settings. It is worth noting as well that increasing the "Rival Pathways" parameter to greater than one will typically increase the breadth of a transformation tree out to the selected transformation depth, and that often, transformation

depth is maximized at 5 to 6 cycles at most, given phase II elimination rules that terminate transformation pathways within TIMES.

Fig. S4 Hierarchically clustered heatmap of Recall Rate clustered on chemical class as designated by ClassyFire for each choice of metabolism simulator and model applied to the Literature Dataset containing 59 parent chemicals. Model selections are indicated at the bottom of each column of the clustered heatmap as one cycle of the "ecbased", two cycles of "Cyp450", and one cycle of "phaseII" metabolism in BioTransformer (BioTransformer), Three cycles of ChemAxon Human Phase I within CTS (CTS), Toolbox In Vitro Rat Liver S9 (Toolbox Vitro) or In Vivo Rat Simulator (Toolbox Vivo), and TIMES In Vitro Rat Liver S9 (TIMES Vitro) or In Vivo Rat Simulator (TIMES Vivo). Average recall rate for a given chemical class is illustrated by an increasingly dark gradient from zero (white) to unity (dark red) with the actual mean recall rate given in each heatmap cell. (Right) Bar chart of chemical class versus log2 scaled occurrence frequency in the dataset.

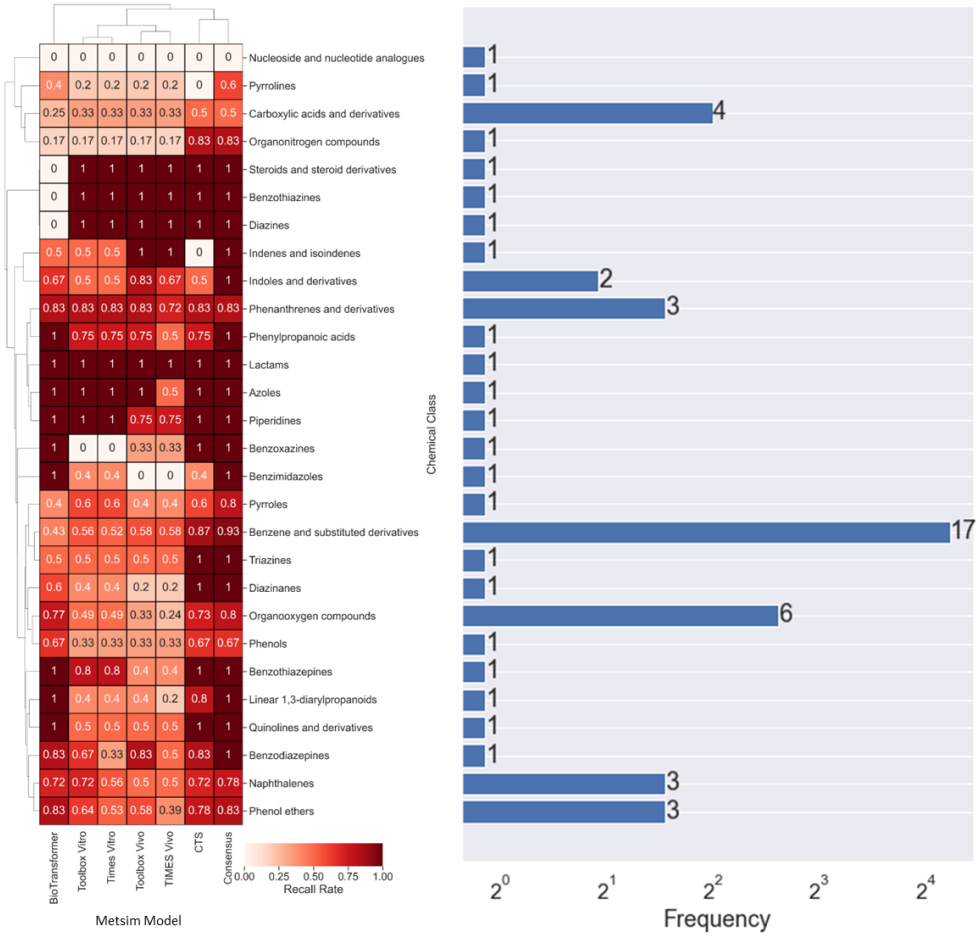


Fig. S5 Hierarchically clustered heatmap of Recall Rate clustered on chemical class as designated by ClassyFire for each choice of metabolism simulator and model applied to the SMPDB Dataset containing 59 parent chemicals. Model selections are indicated at the bottom of each column of the clustered heatmap as one cycle of the "ecbased", two cycles of "Cyp450", and one cycle of "phaseII" metabolism in BioTransformer (BioTransformer), Three cycles of ChemAxon Human Phase I within CTS (CTS), Toolbox In Vitro Rat Liver S9 (Toolbox Vitro) or In Vivo Rat Simulator (Toolbox Vivo), and TIMES In Vitro Rat Liver S9 (TIMES Vitro) or In Vivo Rat Simulator (TIMES Vivo). Average recall rate for a given chemical class is illustrated by an increasingly dark gradient from zero (white) to unity (dark red) with the actual mean recall rate given in each heatmap cell. (Right) Bar chart of chemical class versus log2 scaled occurrence frequency in the dataset.

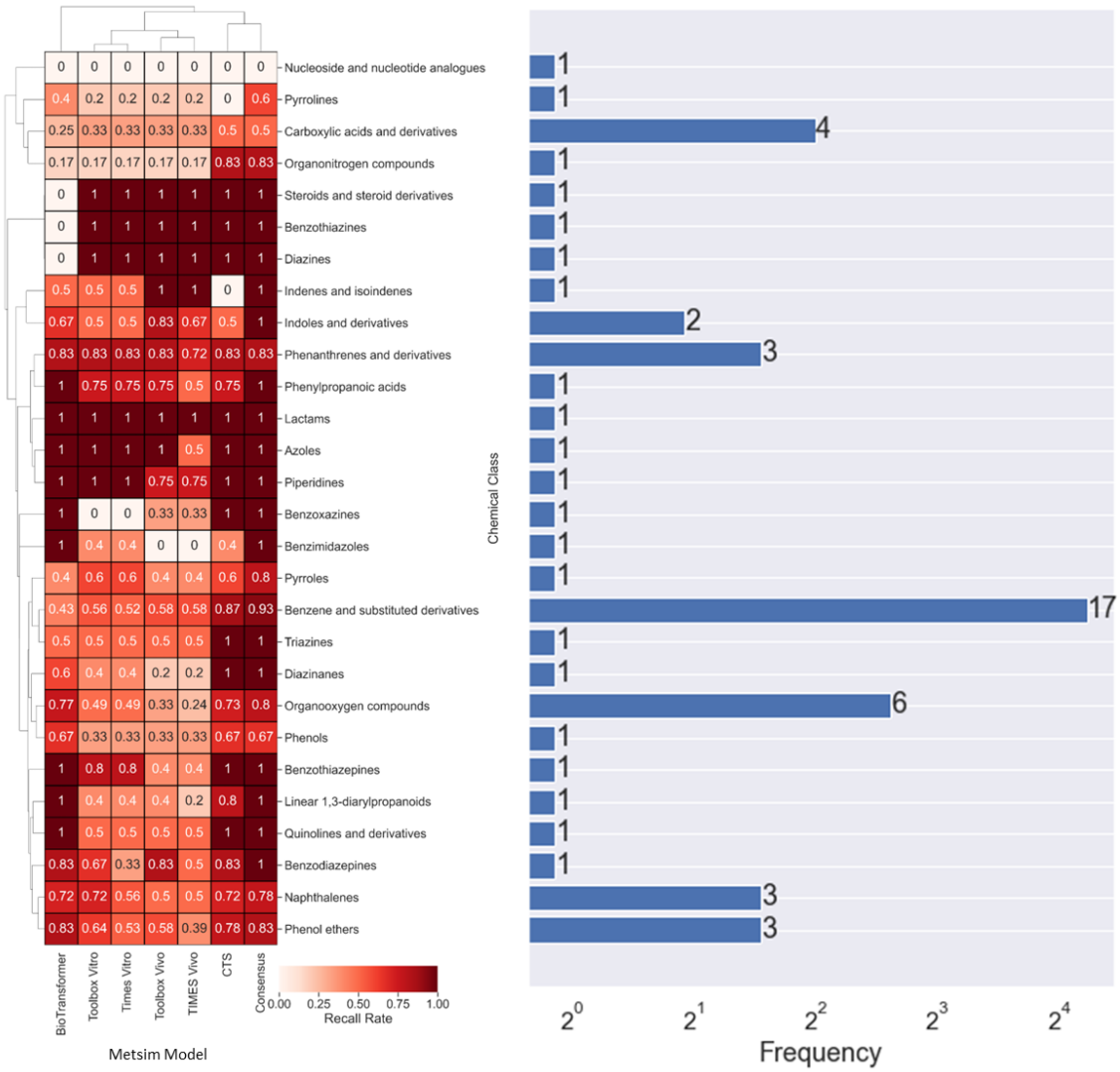

Nucleoside and nucleotide analogues
Pyrrolines
Carboxylic acids and derivatives
Organonitrogen compounds
Steroids and steroid derivatives
Benzothiazines
Diazines
Indenes and isoindenes
Indoles and derivatives
Phenanthrenes and derivatives
Phenylpropanoic acids
Lactams
Azoles
Piperidines
Benzoxazines
Benzimidazoles
Pyrroles
Benzene and substituted derivatives
Triazines
Diazinanes
Organooxygen compounds
Phenols
Benzothiazepines
Linear 1,3-diarylpropanoids
Quinolines and derivatives
Benzodiazepines
Naphthalenes
Phenol ethers
BioTransformer
Toolbox Vitro
Times Vitro
Toolbox Vivo
TIMES Vivo
CTS
Consensus
Metsim Model
0.00 0.25 0.50 0.75 1.00
Recall Rate
Chemical Class
Frequency

Figure S6: Average log($K_{ow}$) values queried from the from the Toolbox API (right, negative values in blue shades, positive values in red shades, zero is white, empty cells returned no log($K_{ow}$) data) sorted with the chemical class and recall rate hierarchical clustering analysis results from the Drug Dataset (left).

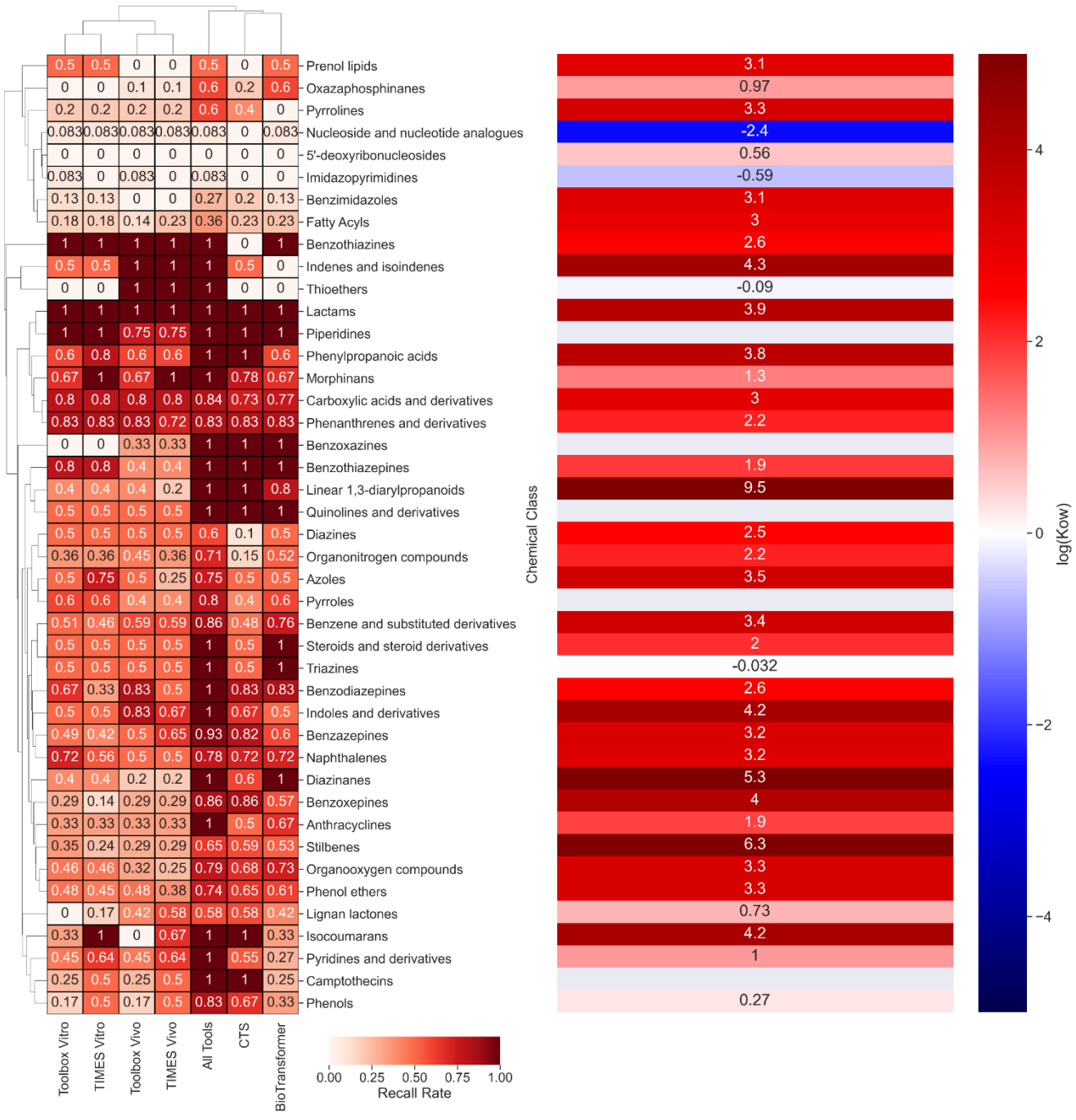


Table S1. Comparison of performance metrics for BioTransformer 2019 validation dataset to current BioTransformer predictions for human tissue metabolism in absence of gut microbiome predictions.

| **BioTransformer Single-Step Human Metabolism No Gut (hgut)** | **Literature Values** | **Combined Individual Runs 1x cyp450 (CypReact) 1x EC-based 1x Phase II** | **Sequential Runs 1x Cyp450 (CypReact) 1x EC-Based 1x Phase II** | **Sequential Runs 1x EC-Based 1x Cyp450 (CypReact) 1x Phase II** |
|---|---|---|---|---|
| **True Positives** | 188 | 150 | 151 | 158 |
| **False Positives** | 198 | 184 | 1765 | 1271 |
| **False Negatives** | 26 | 63 | 62 | 55 |
| **Total No. of Predictions** | 386 | 334 | 1916 | 1429 |
| **Total Precision** | 0.49 | 0.45 | 0.08 | 0.11 |
| **Total Recall** | 0.88 | 0.70 | 0.71 | 0.74 |
| **Total No. of Reported Metabolites** | 224 | 213* | 213* | 213* |

** Metabolites for the literature SDF were counted manually before automated processing in Python, and after processing in Python. Counted 213 metabolites instead of 224.*

Table S2. Comparison of performance metrics for BioTransformer validation dataset to current BioTransformer predictions for single-step phase I metabolism predictions.

| **BioTransformer Single-Step Cyp450 (CypReact)** | **Literature Values** | **1x Cyp450 (CypReact)** |
|---|---|---|
| **True Positives** | 162 | 141 |
| **False Positives** | 188 | 186 |
| **False Negatives** | 18 | 45 |
| **Total No. of Predictions** | 350 | 327 |
| **Total Precision** | 0.46 | 0.43 |
| **Total Recall** | 0.90 | 0.76 |
| **Total No. of Reported Metabolites** | 180 | 186* |

** Metabolites for the literature SDF were counted manually before automated processing in Python, and after processing in Python. Counted 186 metabolites instead of 180.*

Table S3. Comparison of TIMES performance metrics against training set observed maps with parameters set to match current prediction parameters.

| **TIMES**<br>**Version 2.31.2.82** | **In Vitro Rat Liver S9**<br>**Version 12.18**<br>**438 Parents in Training Set**<br>**112 Out of Domain**<br>**Documented Recall**<br>**0.81** | **In Vivo Rat Simulator**<br>**Version 08.14**<br>**701 Parents In Training Set**<br>**204 Out of Domain**<br>**Documented Recall**<br>**0.77** |
|---|---|---|
| **True Positives** | 682 | 1409 |
| **False Positives** | 623 | 1253 |
| **False Negatives** | 89 | 640 |
| **Total No. of Predictions** | 1305 | 2662 |
| **Total Precision** | 0.52 | 0.77 |
| **Total Recall** | 0.88 | 0.69 |
| **Total No. of Reported Metabolites** | 771 | 2049 |

Table S4. Predictive performance characteristics for the Phase I metabolite relationships in the Literature Dataset.

| **JCIM**<br>**59 Drugs & NSAIDs** | **CTS ChemAxon**<br>**Phase I**<br>**3 cycles** | **BioTransformer**<br>**1x EC-Based,**<br>**2x Cyp450,**<br>**1x Phase II**<br>**4 cycles** | **Toolbox API**<br>**In Vivo Rat Simulator**<br>**Phase I** | **Toolbox API**<br>**In Vitro Rat Liver S9**<br>**Phase I** | **TIMES**<br>**In Vivo Rat Simulator**<br>**6 cycles** | **TIMES**<br>**In Vitro Rat Liver S9**<br>**6 cycles** | **Aggregate of All Models** |
|---|---|---|---|---|---|---|---|
| **True Positives** | 131 | 111 | 86 | 93 | 75 | 83 | 154 |
| **False Positives** | 5582 | 11091 | 537 | 446 | 550 | 579 | 15343 |
| **False Negatives** | 48 | 68 | 93 | 86 | 104 | 96 | 25 |
| **Total No. of Predictions** | 5713 | 11202 | 623 | 539 | 625 | 662 | 15497 |
| **Total Precision** | 0.023 | 0.010 | 0.138 | 0.173 | 0.120 | 0.125 | 0.010 |
| **Total Recall** | 0.73 | 0.62 | 0.48 | 0.52 | 0.42 | 0.46 | 0.86 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Total No. of Reported Metabolites** | 179 | 179 | 179 | 179 | 179 | 179 | 179 |

Table S5. Predictive performance characteristics for the Phase I and Phase II metabolite relationships in the SMPDB Dataset.

| **SMPDB 59 Drugs** | **CTS ChemAxon Phase I 3 cycles** | **BioTransformer 1x EC-Based, 2x Cyp450, 1x Phase II 4 cycles** | **Toolbox API In Vivo Rat Simulator Phase I** | **Toolbox API In Vitro Rat Liver S9 Phase I** | **TIMES In Vivo Rat Simulator 6 cycles** | **TIMES In Vitro Rat Liver S9 6 cycles** | **Aggregate of All Models** |
|---|---|---|---|---|---|---|---|
| **True Positives** | 111 | 110 | 95 | 86 | 103 | 93 | 166 |
| **False Positives** | 8137 | 16424 | 821 | 821 | 618 | 641 | 23655 |
| **False Negatives** | 148 | 149 | 164 | 173 | 156 | 166 | 93 |
| **Total No. of Predictions** | 8248 | 16534 | 916 | 712 | 721 | 734 | 23821 |
| **Total Precision** | 0.013 | 0.007 | 0.104 | 0.121 | 0.143 | 0.127 | 0.007 |
| **Total Recall** | 0.43 | 0.42 | 0.37 | 0.33 | 0.40 | 0.36 | 0.64 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Total No. of Reported Metabolites** | 259 | 259 | 259 | 259 | 259 | 259 | 259 |

Table S6. Spearman Correlation analysis of log($K_{ow}$) on Recall Rate with P-values.

| **Tool** | **Spearman Coefficient (Log$K_{ow}$ & Recall Rate)** | **P-Value** |
|---|---|---|
| **BioTransformer** | 0.26 | 0.009 |
| **Toolbox In Vitro Rat Liver S9** | 0.19 | 0.004 |
| **Toolbox In Vivo Rat Simulator** | 0.20 | 0.044 |
| **TIMES In Vitro Rat Liver S9** | 0.19 | 0.059 |
| **TIMES In Vivo Rat Simulator** | 0.15 | 0.144 |
| **CTS** | 0.31 | 0.002 |
| **Ensemble** | 0.27 | 0.006 |

## References


1. Boyce, M.; Meyer, B.; Grulke, C.; Lizarraga, L.; Patlewicz, G., Comparing the performance and coverage of selected in silico (liver) metabolism tools relative to reported studies in the literature to inform analogue selection in read-across: A case study. *Comput Toxicol* **2022,** *21*, 1-15.
2. Djoumbou-Feunang, Y.; Fiamoncini, J.; Gil-de-la-Fuente, A.; Greiner, R.; Manach, C.; Wishart, D. S., BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. *J Cheminformatics* **2019,** *11*.
3. Tian, S. Y.; Cao, X.; Greiner, R.; Li, C.; Guo, A. C.; Wishart, D. S., CyProduct: A Software Tool for Accurately Predicting the Byproducts of Human Cytochrome P450 Metabolism. *J Chem Inf Model* **2021,** *61* (6), 3128-3140.
4. Tian, S. Y.; Djoumbou-Feunang, Y.; Greiner, R.; Wishart, D. S., CypReact: A Software Tool for in Silico Reactant Prediction for Human Cytochrome P450 Enzymes. *J Chem Inf Model* **2018,** *58* (6), 1282-1291.
5. McKerns, M. M.; Aivazis, M. A. G. Pathos: A Framework for Heterogeneous Computing. https://uqfoundation.github.io/project/pathos.
6. McKerns, M. M.; Strand, L.; Sullivan, T.; Fang, A.; Aivazis, M. A. G., Building a Framework for Predictive Science. *Proceedings of the 10th Python in Science Conference* **2011**.
7. Mekenyan, O. G.; Dimitrov, S. D.; Pavlov, T. S.; Veith, G. D., A systematic approach to simulating metabolism in computational toxicology. I. The TIMES heuristic modelling framework. *Curr Pharm Design* **2004,** *10* (11), 1273-1293.
8. Mekenyan, O.; Dimitrov, S.; Pavlov, T.; Dimitrova, G.; Todorov, M.; Petkov, P.; Kotov, S., Simulation of chemical metabolism for fate and hazard assessment. V. Mammalian hazard assessment. *Sar Qsar Environ Res* **2012,** *23* (5-6), 553-606.
9. Yordanova, D.; Kuseva, C.; Tankova, K.; Pavlov, T. S.; Chankov, G.; Chapkanov, A.; Gissi, A.; Sobanski, T.; Schultz, T. W.; Mekenyan, O. G., Using metabolic information for categorization and read-across in the OECD QSAR Toolbox. *Comput Toxicol* **2019,** *12*, 100102.
10. Piechota, P.; Cronin, M. T. D.; Hewitt, M.; Madden, J. C., Pragmatic Approaches to Using Computational Methods To Predict Xenobiotic Metabolism. *J Chem Inf Model* **2013,** *53* (6), 1282-1293.
11. Jewison, T.; Su, Y. L.; Disfany, F. M.; Liang, Y. J.; Knox, C.; Maciejewski, A.; Poelzer, J.; Huynh, J.; Zhou, Y.; Arndt, D.; Djoumbou, Y.; Liu, Y. F.; Deng, L.; Guo, A. C.; Han, B.; Pon, A.; Wilson, M.; Rafatnia, S.; Liu, P.; Wishart, D. S., SMPDB 2.0: Big Improvements to the Small Molecule Pathway Database. *Nucleic Acids Res* **2014,** *42* (D1), D478-D484.
12. Frolkis, A.; Knox, C.; Lim, E.; Jewison, T.; Law, V.; Hau, D. D.; Liu, P.; Gautam, B.; Ly, S.; Guo, A. C.; Xia, J. G.; Liang, Y. J.; Shrivastava, S.; Wishart, D. S., SMPDB: The Small Molecule Pathway Database. *Nucleic Acids Res* **2010,** *38*, D480-D487.
13. Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. N.

H.; Pedregosa, F.; van Mulbregt, P.; Contributors, S., SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **2020,** *17* (3), 261-272.