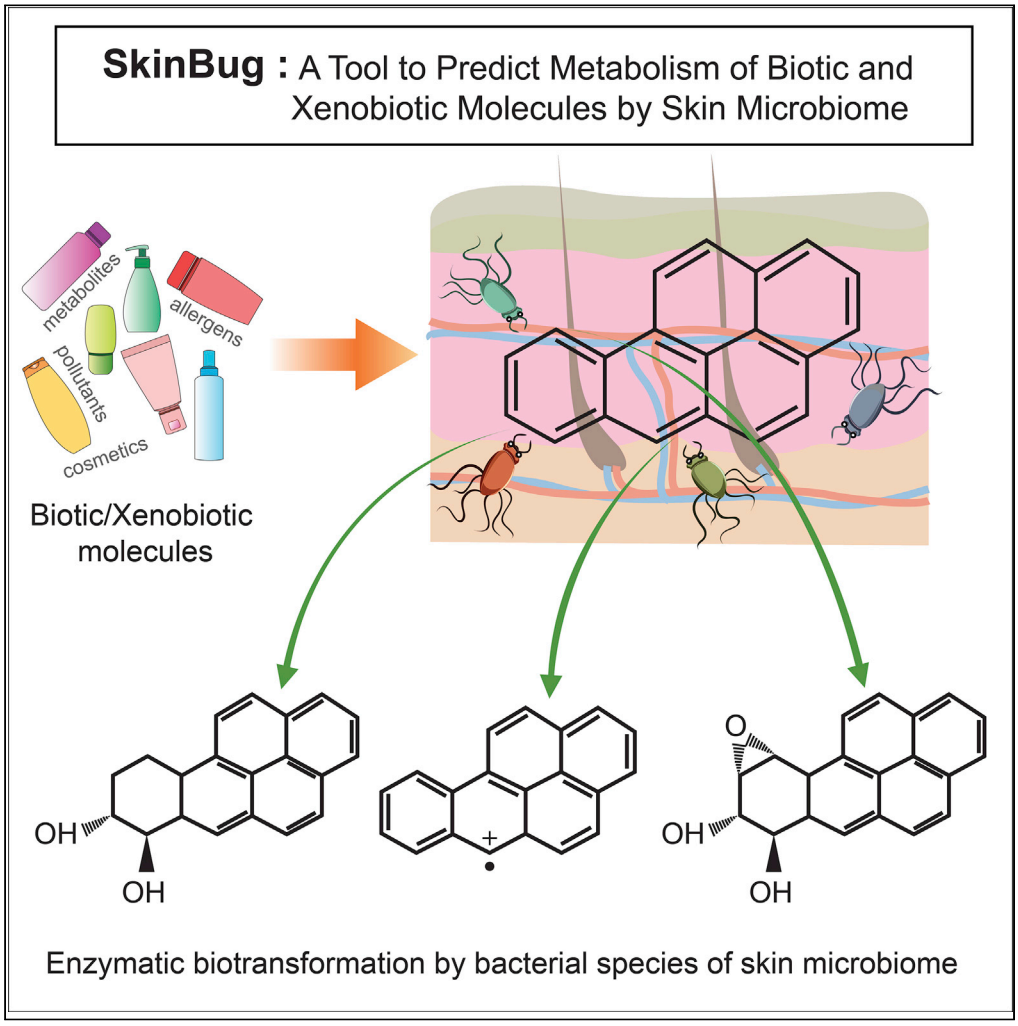
iScience

CellPress
OPEN ACCESS

Article

SkinBug: an artificial intelligence approach to predict human skin microbiome-mediated metabolism of biotics and xenobiotics



Shubham K. Jaiswal, Shitij Manojkumar Agarwal, Parikshit Thodum, Vineet K. Sharma

vineetks@iiserb.ac.in

HIGHLIGHTS

SkinBug is AI/ML-based tool to predict metabolism of molecules by Skin microbiome

Database of 1,094,153 metabolic enzymes from 897 pangenomes of skin microbiome

Predicts enzymes, bacterial species, and skin sites for the predicted reactions

82.4% multilabel and 90.0% binary accuracy, and validated on 28 diverse real cases

Jaiswal et al., iScience 24, 101925
January 22, 2021 © 2020
https://doi.org/10.1016/j.isci.2020.101925

iScience

CellPress
OPEN ACCESS

Article

SkinBug: an artificial intelligence approach to predict human skin microbiome-mediated metabolism of biotics and xenobiotics

Shubham K. Jaiswal,[1] Shitij Manojkumar Agarwal,[1] Parikshit Thodum,[1] and Vineet K. Sharma[1,2,*]

Summary

In addition to being pivotal for the host health, the skin microbiome possesses a large reservoir of metabolic enzymes, which can metabolize molecules (cosmetics, medicines, pollutants, etc.) that form a major part of the skin exposome. Therefore, to predict the complete metabolism of any molecule by skin microbiome, a curated database of metabolic enzymes (1,094,153), reactions, and substrates from ~900 bacterial species from 19 different skin sites were used to develop "SkinBug." It integrates machine learning, neural networks, and chemoinformatics methods, and displays a multiclass multilabel accuracy of up to 82.4% and binary accuracy of up to 90.0%. SkinBug predicts all possible metabolic reactions and associated enzymes, reaction centers, skin microbiome species harboring the enzyme, and the respective skin sites. Thus, SkinBug will be an indispensable tool to predict xenobiotic/biotic metabolism by skin microbiome and will find applications in exposome and microbiome studies, dermatology, and skin cancer research.

Introduction

After the human gut, it is the skin that harbors the largest and most exposed human-associated microbiome known as the skin microbiome that plays a pivotal role in both health and disease (Grice et al., 2009; Grice and Segre, 2011; Kong and Segre, 2012). Skin is composed of three layers, among which the outermost layers (epidermis and dermis) serve as an elaborate host for building an ecosystem with more than a trillion microbial cells belonging to more than 1,000 microbial species from 19 different phyla (Edwards and Marks, 1995; Grice et al., 2009; Grice and Segre, 2011; Nakatsuji et al., 2013).

Several specialized niches such as sebaceous, moist, and dry are found on the skin with differences in moisture level; aerobic, anaerobic, and semi-aerobic conditions; carbon and sulfur availability; secretions; and exogenous environmental factors (Grice and Segre, 2011). The aerobic niches include the skin surface and the epidermis layer, whereas the dermis layer and specifically the invaginations that form folliculo-sebaceous units or roots of the hair follicles and sweat glands form the anaerobic and lipid-rich niches for anaerobic species (Bay et al., 2020; Grice and Segre, 2011). The skin site-specific niches support the growth of diverse bacterial species and result in enormous compositional differences in skin microbiome (Grice and Segre, 2011).

The commensal skin microbiome influences the physical characteristics of the epidermis layer and regulates the development of immune system, and a dysbiosis of this community is associated with several skin diseases such as acne, atopic dermatitis, and psoriatic lesions (Byrd et al., 2018; Grice et al., 2009; Kong et al., 2012; Naik et al., 2012; Picardo and Ottaviani, 2014; SanMiguel and Grice, 2015; Zeeuwen et al., 2013; Saxena, 2018). In addition to being crucial for the host health, the skin microbiome also possesses a large and diverse reservoir of metabolic enzymes that play a key role in metabolizing numerous biomolecules naturally produced by either host skin cells or other commensal microorganisms. One such example of a common skin commensal is *Staphylococcus epidermidis* that can ferment the glycerol present in the skin and can inhibit the growth of *Propionibacterium acnes*, which causes acne vulgaris (Wang et al., 2014).

This gigantic pool of metabolic enzymes can also potentially metabolize the xenobiotic molecules that come in regular contact with our skin such as those present in cosmetics, pharmacological formulations,

[1]MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research, Bhopal, Madhya Pradesh 462066, India

[2]Lead contact

*Correspondence: vineetks@iiserb.ac.in

https://doi.org/10.1016/j.isci.2020.101925

and the environmental pollutants, which together comprise a large part of the skin exposome (Stingley et al., 2010). Furthermore, due to the presence of both aerobic and anaerobic microbial species in the aerobic and anaerobic niches, the skin microbiome can perform the aerobic or anaerobic metabolism of any biotic/xenobiotic molecule.

For example, the metabolism of synthetic azo dyes such as methyl red (MR) is reported by both aerobic and anaerobic bacterial species of skin microbiome that reduce the dye using their azoreductase enzyme (Stingley et al., 2010). The azo dyes are regularly used in cosmetics, tattoo inks, and other products, and the reduction of these dyes due to their undesirable metabolism by the skin microbiome-harbored metabolic enzymes can produce carcinogenic aromatic amines, which poses significant health risks (Chung, 1983; Nakayama et al., 1983). Another example is the aerobic oxidative metabolism of an abundant environment pollutant BaP by monooxygenase enzymes from different bacterial species of skin microbiome (Sowada et al., 2014). The anaerobic metabolism of different skin secretion substances such as triglyceride lipids and secretory proteins by lipases and proteases has also been reported by experimental studies (Byrd et al., 2018). Similarly, the metabolism of glycolic acid, cholesterol, and glycerol by aerobic bacterial species; the metabolism of arginine, triglyceride lipids, propylene glycol, and palmitic acid by anaerobic bacterial species; and the metabolism of alpha-tocopherol, uric acid, lactic acid, ethanolamine, linolenic acid by both aerobic and anaerobic bacterial species of skin microbiome have been reported recently by an experimental study (Timm et al., 2020). However, in most of these cases, the exact metabolic reactions and the metabolic enzymes responsible for the metabolism of these molecules are yet unknown.

The skin microbiome-mediated undesired metabolism of commonly used cosmetic products and therapeutic drugs applied on the skin can significantly alter the bioavailability and therapeutic bioactivity of these molecules. Therefore, exploration of the metabolic potential of the enzymes present in the skin microbiome is much needed to understand their impact and role in human health and disease, in exposome studies, to determine the efficacy and bioavailability of the therapeutic and cosmetic molecules applied on the skin, and also to predict their potential toxicity due to the promiscuous metabolism by the skin microbiome. However, the comprehensive experimental metabolic investigation of each molecule separately by skin microbiome through experimental approaches is a very challenging and tedious task because skin microbiome shows the highest longitudinal variability and largest phylogenetic diversity (Grice and Segre, 2011).

In this scenario, the development of an efficient computational method for the prediction of metabolism of chemical substances by microbial species present in skin microbiome appears as a promising approach. At present, there exists no tool for the prediction of metabolism of cosmetics or any xenobiotic molecule solely by the skin microbial species. Therefore, we developed a tool named "SkinBug" by integrating machine learning, neural networks, and chemoinformatics methods to predict the metabolic reactions, corresponding reaction centers, metabolic enzymes that can catalyze the predicted reactions, species containing predicted enzymes, and skin sites harboring predicted species for any given biotic/xenobiotic molecule by the skin microbiome. As the metabolic enzymes show promiscuity and are capable of metabolizing structurally similar substrates, the structural and chemical properties of substrate molecules from the known reactions were used as features for predicting the metabolism of biotic or xenobiotic/therapeutic molecules by skin microbiome (Babtie et al., 2010; Hult and Berglund, 2007; Khersonsky et al., 2006; Pandya et al., 2014; Khersonsky and Tawfik, 2010; Sharma, 2017). The predictions from the tool can help to determine the metabolic potential of the skin microbiome, in exposome studies, and to design more efficient skin therapeutic molecules and cosmetic agents by considering their metabolism by skin microbiome.

Results

In this study, we have developed a computational tool named "SkinBug" to predict the metabolic reaction, enzymes, species, and skin sites of the skin microbiome that can potentially metabolize the biotic/xenobiotic molecules by integrating chemoinformatics, machine learning, and neural network methodologies.

Construction of pangenomes for skin microbiome

Several metagenomic studies have explored the bacterial diversity of different skin sites; however, until now there is no single resource available that provides information on the bacterial species present at different skin sites. Therefore, the data and text mining of the available literature was performed to construct the database of bacterial species present at different skin sites. A total of 1,616 unique bacterial species were identified from the 19 different skin sites from literature. Of the 1,616 identified bacterial

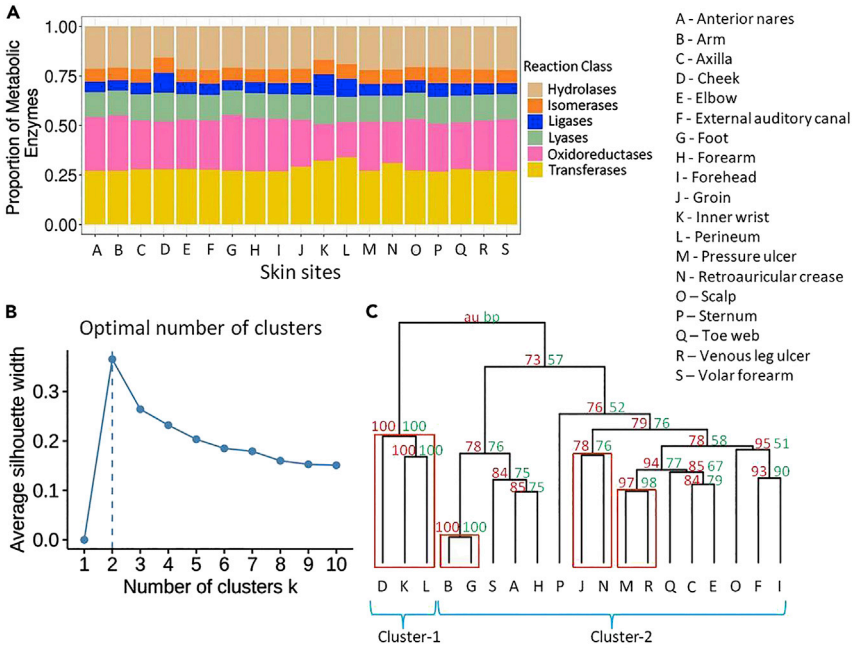


Figure 1. Diversity of metabolic enzymes on skin sites

(A) Proportions of different metabolic enzymes from six types of reaction classes ("Oxidoreductases," "Hydrolases," "Transferases," "Lyases," "Isomerases," and "Ligases") plotted as stacked bar plots across the 19 different skin sites.

(B) Silhouette plot for determining the optimum number of clusters for clustering different skin sites. The y axis is average silhouette width value, and the x axis is the number of clusters. The k-value at which the maximum of average silhouette width is achieved is the optimum number of clusters.

(C) The dendrogram showing the results of hierarchical clustering of the skin sites based on the 2,523 unique reactions that can be performed by the metabolic enzymes of species present in skin microbiome. The approximate unbiased p values (AUp) and the bootstrap probability (BP) values for each branch/cluster were calculated using multiscale bootstrap resampling and using normal bootstrap resampling, respectively. The AUp values are mentioned in red, and the BP values are mentioned in green.

See also Figures S1–S4.

species, the complete genomes of 897 could be identified at the "NCBI RefSeq" database. The genomes of all the strains were used to construct the pangenome of a particular species. A total of 897 pangenomes were constructed for the human skin microbiome and were used for further analysis.

Transferases are the most abundant enzymes in skin microbiome

Using the ExPASy enzyme database a total of 1,094,153 metabolic enzymes were identified from the 897 pangenomes. Each of these metabolic enzymes was linked to a four-digit EC number based on the associated EC number of its closest homolog. From the distribution of metabolic enzymes across six reaction classes, it was apparent that the enzymes belonging to "Transferases" class were the most abundant in the skin microbiome, whereas "Isomerases" enzymes were the least abundant (Figure S1).

The proportions of different reaction class were identified for each of the 19 skin sites (Figure 1A). For all the sites, we observed that the "Transferases" reactions were most abundant followed by the "Oxidoreductases" and "Hydrolases" reactions, perhaps due to the higher abundance of these classes of enzymes in the bacterial pangenomes. The inner wrist, perineum, and cheek had the highest proportions of the "Ligases"_underreactions and the lowest proportions of "Hydrolysis" reactions when compared with the other skin sites. The inner wrist and perineum also have the least proportion of "Oxidoreductases" reactions when compared with the other skin sites.

Skin sites are highly variable in enzymatic reactions

Of the 5,430 reactions on KEGG (Kyoto Encyclopedia of Genes and Genomes) database, the complete skin microbiome had enzymes for 2,523 reactions (Kanehisa and Goto, 2000). To identify the similarities and differences

in skin sites with respect to the enzymatic reactions, hierarchical clustering was performed and the approximate unbiased p values (AUp) and the bootstrap probability (BP) values for each branch/cluster were calculated. The optimum number of clusters was identified to be two based on the average silhouette method (Figure 1B). Among the two most optimum clusters, the sites cheek, inner wrist, and perineum clustered together, and the rest of the sites formed another cluster (Figure 1C). Moreover, four clades were found to be statistically significant in the hierarchical clustering of skin sites. The first clade was formed by the cheek, inner wrist, and perineum; the second clade was formed by the arm and foot; the third clade was formed by the groin and retroauricular crease; and the fourth clade was formed by pressure ulcer and venous leg ulcer (Figure 1C). It is apparent that the sites with similar secretions, humidity, environmental exposure, or disease status clustered together.

Also the total reactions present at each site and the reactions common to the different sites were identified using the matrix layout analysis (Supplementary Text S5). Foot and arm had the highest number of enzymatic reactions, whereas the perineum, inner writs, and cheek had the lowest number of enzymatic reactions (Figure S2). These observations are supported by a previous finding, which showed that dry niches of skin microbiome such as foot and arm have higher phylogenetic diversity than the other niches of skin microbiome (Grice and Segre, 2011). Only 277 reactions of 2,523 were common to all the 19 skin microbiome sites, and many reactions were unique to specific groups of sites. For example, 134 reactions were unique to foot and arm and occur only at these two sites (Figure S2). This suggests that the metabolic potential of different skin microbiome sites is variable, and site-specific metabolism should be considered while developing the tool to predict the metabolism of molecules by skin microbiome.

Metabolic complexity of microbiome

The principal-component analysis (PCA) with 2,322 variables was performed for the 3,769 substrates (Supplementary Text S6 and S7). In the PCA analysis, PC-1 (14.7%) and PC-2 (5.7%), collectively could explain only 20.4% of the variance present in the input dataset (Figure 2A). Furthermore, from the scree and cumulative scree plot it is apparent that more than 1,000 dimensions (PCs) were needed to explain the complete variance (>95%) present in the dataset (Figure 2A). This suggests that the dataset is very diverse and most of the variables add significant variance to the dataset (Figures 2A, S3, and S4).

Furthermore, to identify if the different reaction classes are separable from each other or cluster separately, the density-based unsupervised clustering was performed on the PCA results using PC-1 and PC-2 (Kriegel et al., 2011). The density-based clustering is resistant to noise, outliers, and inherent irregular shapes of the clusters present in the complex datasets when compared with the other methods such as k-means, PAM (partition around medoids), and hierarchical clustering. The two parameters that affect the quality of clustering are epsilon (eps) and minimum points (MinPts). Larger datasets need a larger "MinPts" value, thus, a value of 20 was used. The epsilon value of 3.25 was chosen based on the k-distance graph, the best value being the knee point in the graph (Figure S5). In the density-based clustering we observed that the dataset formed three clusters with a lot of outliers marked as black points (Figure 2B). Moreover, each of these clusters had substrate molecules from all the six reaction classes; cluster-1 had the highest number of substrates for Transferases reaction, whereas cluster-2 and cluster-3 had the highest number of substrates for Oxidoreductase reaction (Figure 2C). Furthermore, the PCA analysis of pure substrates that can perform only one type of reaction was performed to check if any pattern of separation exists among these pure substrates. For these substrates also, more than 1,000 dimensions were required for explaining the complete variance present in the dataset (Figure S6), and no clear separation between different reactions classes was observed (Figure S6).

Collectively, it is evident that the input dataset is very diverse and complex and there is no clear linear or non-linear separation between different reaction classes. Thus, robust machine learning and/or deep learning methods are required to build the models that can perform the multiclass multilabel classification for these reaction classes.

Prediction of metabolism by skin microbiome

The complete workflow of the construction of "SkinBug" is mentioned in Figure 3. "SkinBug" provides a comprehensive prediction of the possible metabolic reactions and their respective reaction centers, corresponding metabolic enzymes, species having those enzymes, and skin sites carrying those species. The five key steps involved in the prediction are classified into five modules. The first module predicts the reaction class, the second module predicts reaction subclass, the third module predicts complete

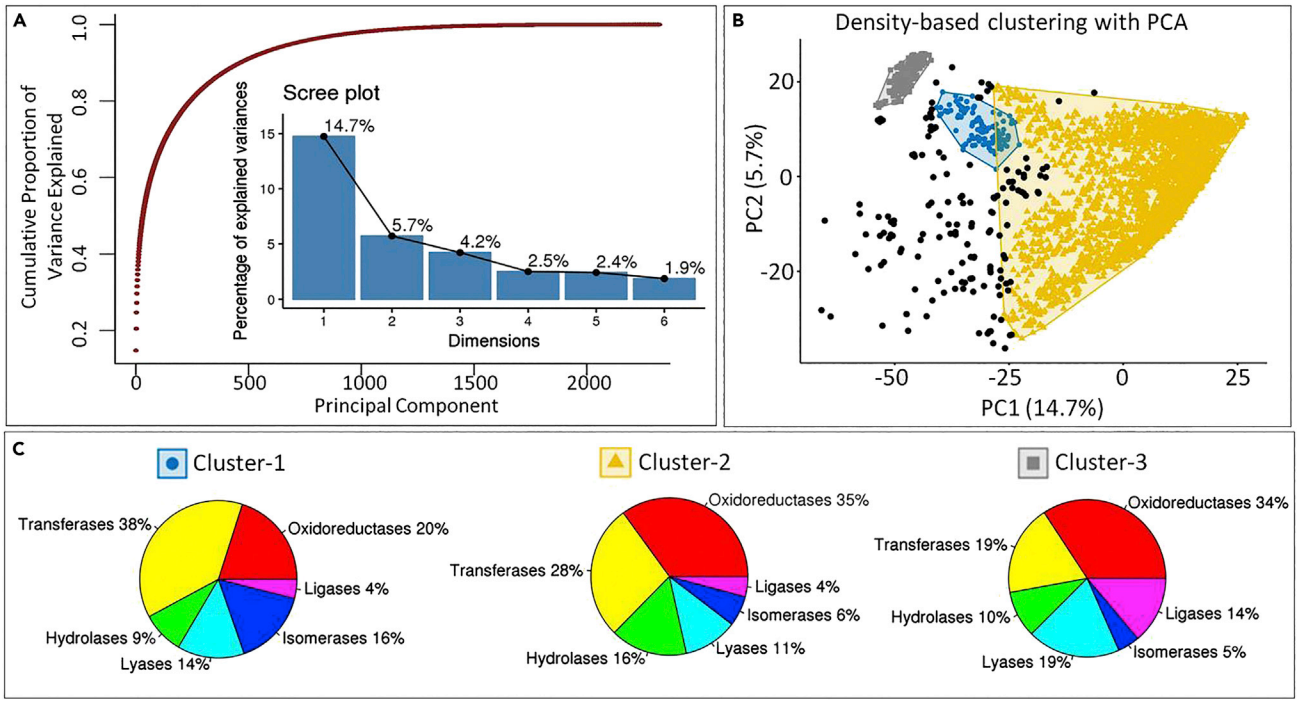


Figure 2. Evaluating the diversity and complexity of metabolic dataset

(A) The cumulative scree plot and a normal scree plot from the PCA analysis of all the substrate molecules from the complete input dataset based on the 2,322 selected features. The x axis is the principal component number, y axis for the dot plot is the cumulative variance explained by the individual principal components, and the y axis for the bar plot is the percentage of variance explained by the individual principal components.

(B) The density-based clustering of substrate molecules using the principal component PC1 and PC2 from the PCA analysis. The density-based clustering was performed with MinPts value 20 and epsilon value 3.25. Three dense clusters were identified, which are colored as blue for cluster-1, yellow for cluster-2, and gray for cluster-3.

(C) The proportion of different reaction classes in each cluster is shown as pie chart with percentage value labeled for each reaction class. The proportion is calculated based on the number of substrate molecules that can undergo a particular kind of reaction class in that cluster. The pie chart is mentioned for all three above identified clusters: cluster-1, cluster-2, and cluster-3.

See also Figures S5 and S6.

reactions, the fourth module predicts the reaction centers in the molecule for the predicted reactions, and the fifth module provides information on the metabolic enzymes, species, and skin sites.

Module-1: Construction of reaction class-specific prediction models

Two kinds of approaches were used; one is machine learning that will be efficient at learning the class correlations and class discrimination, and the other is artificial neural network (ANN) models that will be efficient at learning the class-specific patterns.

For the machine learning-based multiclass multilabel predictions, both the problem transformation (five types) and algorithm adaptation methods were used. For each type of problem transformation seven types of best suited core learners were used. Thus, the performances of a total of 34 different problem transformation models and the two available algorithm adaptation models were compared. To obtain a comprehensive and reliable comparison of the models, the comparison was performed using three types of datasets: (1) boruta selected fingerprints, boruta selected descriptor, and ECFP, (2) boruta selected fingerprints, boruta selected descriptor, and FCFP; and (3) boruta selected fingerprints, boruta selected descriptor, ECFP, and FCFP. The values of multilabel accuracy, multilabel precision, F1 score, and hamming loss for all the models for all the three datasets are plotted as line plot and are mentioned in Figures S7–S9. From the plots it is evident, that only Random forest survival regression classification (RFSRC) algorithm consistently performed better on all the three datasets, and thus, this algorithm was selected for further modeling.

The complete dataset of 3,769 substrates was split into a working dataset and a blind dataset with a split ratio of 95:5. Furthermore, the working dataset was split into training and testing dataset using a

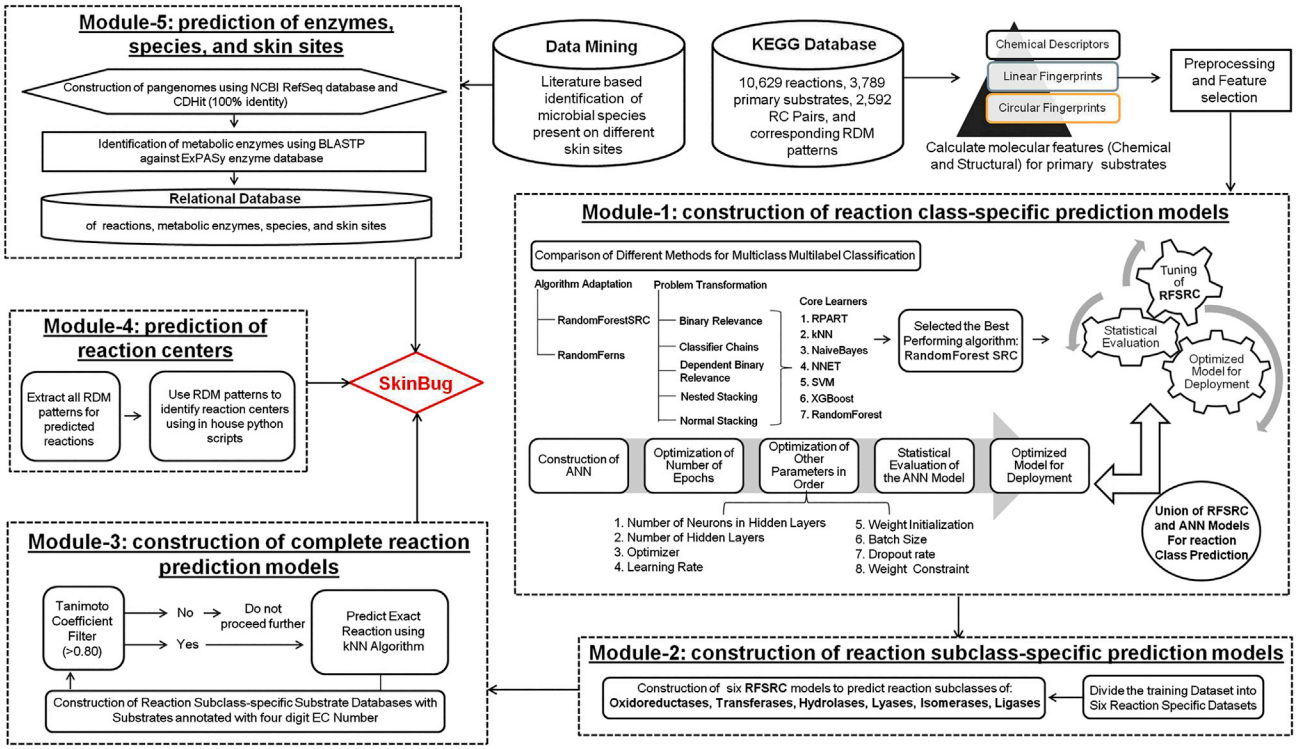


Figure 3. The algorithm and complete workflow for the construction of SkinBug

customized stratified random sampling method (Methods). The RandomForestSRC model was trained on training dataset, and its performance was evaluated on the testing dataset. From the multilabel performance, it is apparent that the model showed a multilabel accuracy of up to 82.4%, multilabel sensitivity of up to 91.8%, multilabel precision of up to 83.3%, F1 score of up to 85.5%, and hamming loss of up to 0.067 in the three types of statistical testing methods (Figure 4A).

However, this performance could also be a result of over-fitting; thus, examining the over-fitting of the model is required. One way to examine over-fitting is cross validation. If the performance varies too much among the different folds of the cross-validation, it indicates the case of over-fitting. Hence for the RFSRC model, the performance on each fold of the 5-fold cross-validation was calculated and plotted as box plot to evaluate the over-fitting of the model (Figure 4B). From the box plot it is clear that the performance of the model across the folds of cross-validation is similar for all the five measures of multiclass multilabel performance. Thus, it is apparent that the RFSRC model does not show any sign of over-fitting.

Additionally, the binary performance of the RFSRC model on test dataset was also evaluated for each of the reaction class and is shown as a heatmap (Figure 4C). The binary performance measures the quality of learning for each of the target class present in the multiclass multilabel dataset. The performance for each class is measured using the 11 different binary performance matrices. In the binary performance, the model could achieve the binary accuracy of up to 98.8%, area under the curve of up to 95.5%, and Matthews correlation coefficient of up to 0.78. The binary performance of the RFSRC model on the 5-fold cross-validation and blind dataset was also calculated and is mentioned in the Tables S1 and S2.

For the ANN model, the final selected architecture had a total of three layers: input layer with size of 2,322 neurons based on the number of features selected, hidden layer with size of 1,162 neurons, and output layer with size of six neurons based on six types of reaction classes. As ANN model was to be used for the multilabel classification, the three evaluation matrices were categorical accuracy, binary accuracy, and log loss. The change in the three matrices along the epochs is measured for the training and test dataset to obtain the most optimum epoch value. The value of 1,500 epochs was found to be the most optimum using all the three matrices as all the three matrices showed a plateau after this value (Figures 5A–5C).

A

Testing type / Evaluation matrices	Accuracy (%)	Sensitivity (%)	Precision (%)	F1 Measure (%)	Hamming loss
Stratified test dataset	82.4	91.8	83.3	85.5	0.067
5-fold Cross Validation	63.6	74.5	77.1	72.9	0.195
Blind dataset	68.3	78.7	80.8	75.7	0.131

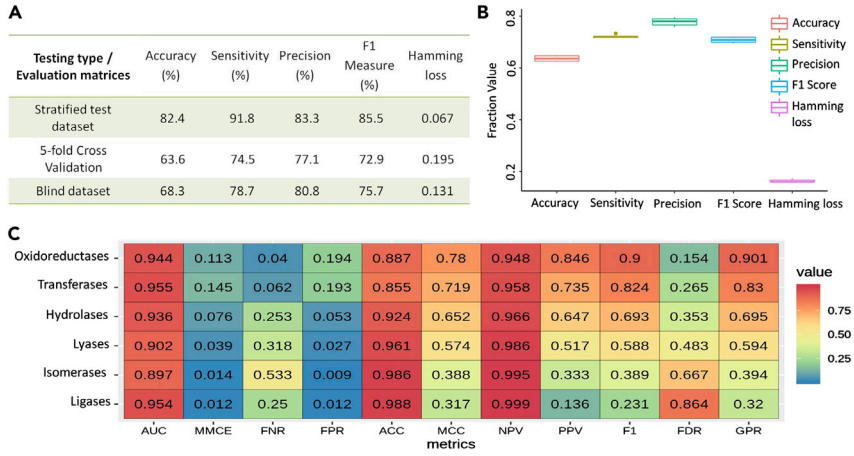


Figure 4. The multilabel and binary performance of reaction class-specific RFSRC prediction model

(A) The multilabel accuracy, multilabel sensitivity, multilabel precision, multilabel F1 score, and hamming loss mentioned for three type of statistical testing of the RFSRC model: stratified randomly sampled test dataset, 5-fold cross-validation, and blind dataset.

(B) The box plot for the performance across each fold of the 5-fold cross-validation for the five different performance measures: multilabel accuracy, multilabel sensitivity, multilabel precision, multilabel F1 score, and hamming loss. The y axis shows the fraction value for each of the performance measure. Fraction values were chosen so that all the measures including hamming loss could be plotted on the same y axis scale.

(C) The binary performance of the reaction class prediction RFSRC model for each of six different reaction classes ("Oxidoreductases," "Hydrolases," "Transferases," "Lyases," "Isomerases," and "Ligases"). Fraction values were chosen so that all the measures could be shown on the same scale.

AUC = area under the curve, MMCE = binary mean misclassification error, FNR = binary false-negative rate, FPR = binary false-positive rate, ACC = binary accuracy, MCC = Matthews correlation coefficient, NPV = binary negative predictive value, PPV = binary positive predicted value, F1 = binary F1 score, FDR = binary false discovery rate, GPR = geometric mean of binary precision and binary recall.

See also Figures S7–S9.

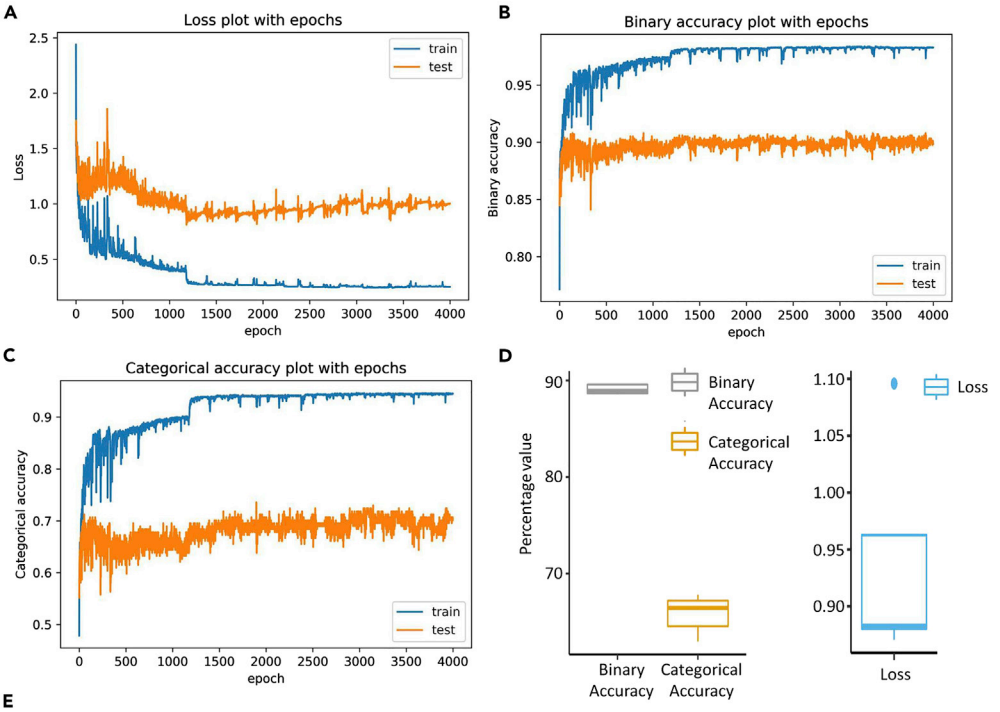
The specific order for the hyperparameters tuning as suggested by Greff et al., 2016, was performed using grid search to achieve the most optimum value of all the parameters (Greff et al., 2016). The results of the grid search with 5-fold cross-validation-based hyperparameters tuning are mentioned in Figures S10–S14. The final ANN model was trained with the hyperparameter values: epochs = 1,500, optimizer algorithm = RMSprop, learning rate = 0.001, weight initialization method = lecun_uniform, batch size = 150, dropout rate = 0.4, and weight constraint = 4. As it was a multiclass multilabel classification problem, for the hidden layer "rectified linear unit" activation function was used, whereas for the output layer the "sigmoid" activation function was used.

The final ANN model on the optimized hyperparameters was tested using the 5-fold cross-validation on the training dataset. The values of all the three evaluation matrices were calculated for each fold of the 5-fold cross-validation and were plotted as box plot to evaluate for any kind of over-fitting (Figure 5D). From the box plot it is apparent that the values across the folds are very similar, thus conveying no over-fitting in the final ANN model. Furthermore, the performance of the ANN model was also evaluated on the test dataset. The ANN model could achieve a categorical accuracy of up to 70.7%, binary accuracy of up to 90.0%, and log loss of up to 0.813 (Figure 5E).

As the learning capabilities of the RFSRC model and the ANN models are complementary to each other, the union of the predictions of multiclass multilabel RFSRC and ANN models was used to make the prediction of the reaction class.

Module-2: Construction of reaction subclass-specific prediction models

As one molecule could undergo multiple types of reactions belonging to different reaction subclasses, the multiclass multilabel prediction models were constructed for each type of reaction class to predict the reaction subclasses. The working dataset was split into six datasets, one for each reaction class,



E

Model	CV-5 Fold			Test Set		
	Log loss / Binary Cross Entropy	Categorical Accuracy (%)	Binary Accuracy (%)	Log loss / Binary Cross Entropy	Categorical Accuracy (%)	Binary Accuracy (%)
Artificial Neural Network	0.938	65.8	89.074	0.813	70.7	90.0

Figure 5. The optimization of epochs and performance evaluation of ANN model

(A) The plot of log loss or binary cross-entropy for different values of epochs for the training and testing dataset.

(B) The plot of binary accuracy for different values of epochs for the training and testing datasets.

(C) The plot of categorical accuracy for different values of epochs for the training and testing datasets.

(D and E) (D) The box plot for the performance across each fold of the 5-fold cross-validation for the three different performance measures of ANN model: categorical accuracy, binary accuracy, and log loss or binary cross-entropy. The y axis for the binary and categorical accuracy shows the percentage value, whereas the y axis for the log loss or binary cross-entropy shows the actual value. (E) The performance of ANN model measured using three evaluation measures, categorical accuracy, binary accuracy, and log loss or binary cross-entropy for two types of statistical testing methods: 5-fold cross-validation and testing on stratified randomly sampled testing dataset.

See also Figures S10–S14.

and these datasets were utilized for the construction of reaction subclass-specific prediction models (Methods). The previously optimized RFSRC algorithm was used to construct six different multilabel prediction models specific to the six types of reaction classes. The multiclass multilabel performance on testing dataset for RFSRC model for each reaction class is mentioned in Table 1, and for 5-fold cross-validation on training dataset is mentioned in Table 2. The RFSRC models showed the multilabel accuracy of 61.0–74.4%, multilabel sensitivity of 63.7%–77.7%, multilabel precision of 75.0%–91.4%, F1 score of 63.9%–76.5%, and hamming loss of 0.093–0.019 on the test dataset. The binary performance of RFSRC models for each of the corresponding reaction subclass for each reaction class for 5-fold cross-validation is mentioned in Tables S3–S8, and on test dataset, is mentioned in Tables S9–S14. After the different reaction classes are predicted by the reaction class prediction models, the prediction of reaction subclass was performed for each of the predicted reaction class using the respective reaction class-specific RFSRC models.

Module-3: Construction of complete reaction prediction models

The molecular similarity search was performed against all the predicted reaction subclasses from each of the predicted reaction class. The molecular similarity search measure known as Tanimoto coefficient

Table 1. Multiclass multilabel performance metrics for the reaction subclass-specific models calculated using stratified random sampling

Reaction class	Accuracy (%)	Sensitivity (%)	Precision (%)	F1 score (%)	Hamming loss
Oxidoreductases	73.9	77.7	88.6	76.5	0.019
Transferases	61.0	63.7	83.3	63.9	0.072
Hydrolases	72.4	76.8	79.2	74.8	0.04
Lyases	74.4	76.7	85.6	76.2	0.058
Isomerases	72.2	73.6	91.4	73.1	0.06
Ligases	72.0	72.0	75.0	72.0	0.093

All the parameters mentioned above were calculated using multilabel evaluation methodology (see Tables S9–S14).

was calculated, and the subclasses that had substrates with Tanimoto coefficient value of >0.80 were considered further. This step was essential to filter out the false-positive predictions from the previous steps. The k-nearest neighbors (KNN) were identified in the reaction subclasses that qualify the aforementioned criteria using the KNN method. In this approach, the molecular similarity searching and KNN methods together allow for identifying the best structural and chemical homolog of the input molecule. Thus, the complete reactions of all the molecules that qualify the molecular similarity searching and KNN method were assigned to the given molecule. This way all the complete metabolic reactions that could potentially occur to a given molecule were identified.

Module-4: Prediction of reaction centers

The reaction centers on the given molecule for all complete reactions predicted in the abovementioned steps are identified using the RDM (Reaction center, Difference region, and Matched region) patterns database. The RDM patterns are derived from the structure alignments of the substrates and contain the information about the KEGG atom type changes at the reaction center, matched region of the molecules, and the dissimilar region of the molecule (Yamanishi et al., 2009). The KEGG reaction-class pairs (RC pairs) were identified for each of the predicted complete reactions, and the corresponding RDM patterns were extracted from the RDM pattern database constructed in this study. Each of the RDM pattern is applied to the given molecule, and the respective reaction center is identified using the in-house python script while taking into account the atoms types, their valancies, and bonding information about the molecule.

Module-5: Prediction of metabolic enzymes, species, and skin sites

In this module, for each of the predicted complete reaction, the metabolic enzymes that could potentially perform those reactions, the corresponding species that harbors those metabolic enzymes, and all the skin sites that carry those species are retrieved from the in-house constructed skin microbiome specific metabolic database that contains information on 1,094,153 metabolic enzymes from 897 species pangenomes of skin microbiome from 19 skin sites.

The development of SkinBug web server

For easy and user-friendly usage of the SkinBug approach, a user-friendly web server tool was developed and is available at http://metagenomics.iiserb.ac.in/skinbug. On the web server, a user can upload a mol/sdf file of a molecule, or can provide the PubChem ID of a molecule. For a novel molecule, these files can also be prepared by drawing its structure in any of the molecular editor software, e.g., ChemDraw. The complete workflow of the web server-based predictions using SkinBug is mentioned in Figure 6, and the other details of the web server tool are mentioned in Supplementary Text S8.

The results from the SkinBug tool are displayed as results sections: R1 to R6. All the possible reaction classes are provided in R1; reaction subclass in R2; complete reactions in R3; the RC pair, RDM pattern, and predicted reaction centers in R4; and the predicted skin sites in R5. The complete reaction annotated as a four-digit EC number, corresponding metabolic enzyme, species harboring the enzyme, and the homology parameters for enzyme annotations (percent identity, e-value, query coverage, subject coverage) are mentioned in R6. All the results could also be downloaded as a single text file by clicking on the "Download Results" icon.

Table 2. Multiclass multilabel performance metrics for the reaction subclass-specific models calculated using 5-fold cross-validation

Reaction class	Accuracy (%)	Sensitivity (%)	Precision (%)	F1 score (%)	Hamming loss
Oxidoreductases	61.5	64.3	82.8	64.9	0.036
Transferases	56.3	59.6	80.8	60.0	0.09
Hydrolases	63.1	66.3	80.5	66.3	0.052
Lyases	64.4	67.1	77.7	67.1	0.084
Isomerases	65.2	67.7	82.5	67.0	0.089
Ligases	77.7	80.0	84.5	79.9	0.075

All the parameters mentioned above were calculated using multilabel evaluation methodology (see also Tables S3–S8).

Validation of the SkinBug tool

This is the first tool that can predict the enzymatic biotransformation of any biotic or xenobiotic molecule by the skin microbiome, and thus a direct comparison with any existing state-of-the-art tool was not possible. The biological validation of SkinBug was performed on diverse molecules including natural molecules present on skin, cosmetics, pharmaceuticals, etc. that come in regular contact with our skin, and their metabolism is known from the experimental studies (Table 3). A total of 28 diverse molecules selected for which information about their metabolism by human host or microbial species of skin microbiome through experimental studies is available in literature were evaluated by SkinBug and presented as case studies. This validation set also included the examples of molecules that can undergo the aerobic and/or anaerobic metabolism. The comparison of SkinBug predictions was performed with the metabolism information known from previous experimental studies, and information available from three reference databases, namely, EAWAG-BBD (The University of Minnesota Biocatalysis/Biodegradation Database), Transformer (Metabolism of Xenobiotics Database), and SMPDB (The Small Molecule Pathway Database) (Gao et al., 2010; Hoffmann et al., 2014; Jewison et al., 2014). The prediction results for these 28 molecules along with the information from literature are mentioned in Table 3. Some examples from the comparative analysis with the reference databases are provided below.

For the metabolism of hydroquinone molecule (used as a demelanizing agent) SkinBug predicted monooxygenases, dioxygenases, hydroxylases, and arylesterases, and the reference databases also indicated the same four classes of enzymes for its metabolism. Likewise, for terbinafine (used to treat fungal infections on skin) and fluorouracil (used in actinic keratosis and skin warts conditions), the reference databases suggested their metabolism by human liver cytochromes, and SkinBug also predicted metabolic enzymes from skin microbiome with similar activity such as monooxygenases. For para-aminobenzoic acid, which is used as melanizing agent and for the treatment of different skin disorders, both SkinBug and reference databases suggested benzoate dioxygenases for its metabolism. Similarly, for methylparaben (used in cosmetic products), which is metabolized by 4-hydroxybenzoate 1-hydroxylase and monooxygenases as per the reference databases, SkinBug also predicted the same enzymes from skin microbiome for its metabolism. It is noteworthy that for some molecules such as cinnamyl alcohol, BaP, fluorouracil, and 7,12-dimethylbenz(a)anthracene molecules, SkinBug only predicted those reactions that are known in literature without any false-positives confirming to its robustness. However, the strength of SkinBug lies in the fact that for the other molecules as mentioned in Table 3, in addition to correctly predicting the known reactions it also predicted additional metabolic reactions, which seems correct as per the reactive functional groups present on those molecules, along with the implicated microbial enzymes and species. Thus, the predictions of SkinBug corroborated with the metabolic information available on these reference databases and experimental studies (Table 3), confirming to the accuracy and reliability of the predictions.

Metabolism of Benzo(a)pyreme by skin microbiome

BaP is an abundant environmental pollutant found in almost all types of soot and smoke generated by the incomplete combustion of fossil fuel, coal, and other biomass including tobacco. The metabolism of this polycyclic aromatic hydrocarbon molecule by oxidation reaction through cytochrome P450 enzyme from human host and monooxygenases and dioxygenase enzymes from bacterial species of skin microbiome is known from experimental studies, and reference databases such as the University of Minnesota

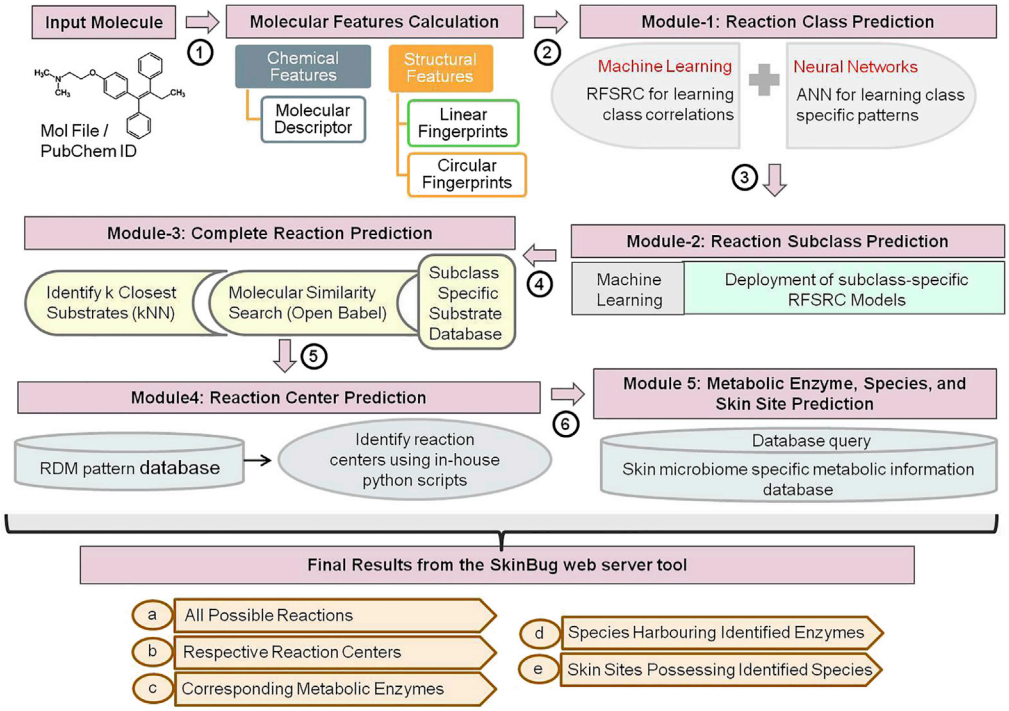


Figure 6. The prediction steps and complete workflow for processing of any input molecule by the SkinBug web server

Biocatalysis/Biodegradation Database (Gelboin, 1980; Gibson et al., 1975; Jiang et al., 2007; Schwarz et al., 2001; Sowada et al., 2014). For this molecule, SkinBug correctly predicted its metabolism by skin microbiome as shown by an experimental study from Sowada et al., 2014, and predicted its oxidation by naphthalene 1,2-dioxygenase and monooxygenase enzymes from multiple species of *Burkholderia*, *Polaromonas*, *Pseudomonas*, and *Ralstonia* genera from the skin microbiome. This case study further supports the validity, accuracy, and utility of SkinBug tool. The ingestion of BaP is known to be toxic due to its metabolism by human cytochrome P450 enzyme and the products of its oxidation are known to cause carcinogenicity and other adverse health effects by reacting with the host DNA (Gibson et al., 1975; Vo-Dinh et al., 1987; Zhou et al., 2017). Thus considering the previous findings and prediction results, it is apparent that its contact with skin could also be toxic due to its undesired metabolism by the skin microbiome.

Similarly, the prediction of aniline metabolism that is largely present in the tobacco smoke reveals that this can also be metabolized by the microbial species present in the skin microbiome.

Metabolism of Azo dyes by skin microbiome

Azo dyes such as MR are widely used in cosmetics and other products that regularly come in contact with our skin in the form of tattoo ink, hair colors, and textile colors. These dyes are experimentally known to be reduced by azoreductase (Azo1) enzyme found in several species of the skin microbiome (Stingley et al., 2010). SkinBug also predicted the reduction of these dyes by the same metabolic enzyme and species and also predicted some other metabolic enzymes from different bacterial species of skin microbiome with azoreductase activity that could also potentially metabolize the azo dyes.

Discussion

The skin microbiome is capable of undertaking numerous metabolic reactions in addition to that of the human genome (Stingley et al., 2010). The regular contact of skin microbiome with cosmetic agents, pollutants, and topical substances like skincare products and medical ointments, etc., which are part of the skin exposome, creates unprecedented possibilities of their promiscuous metabolism that results in modulation of their efficacy, and occasional toxicity associated with skin rashes and cancer (Chung, 1983; Stingley et al., 2010). To predict all the possible metabolic reactions that can occur to such chemical substances by

Table 3. The prediction of metabolism of the 28 selected molecules by SkinBug along with their known metabolism from literature

Sr. No.	Compound	Function	Reaction subclass from literature	Reaction subclass from SkinBug	Enzyme from literature	Enzyme from SkinBug	Skin microbial genus from literature	Skin microbes genus from SkinBug	References
1	Cinnamyl alcohol	Perfuming, Masking	Acting on the CH-OH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on the CH-OH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen; acyltransferases	Alcohol dehydrogenase; cytochrome P450	Cinnamyl alcohol dehydrogenase; acetyl CoA benzyl alcohol acetyltransferase	NA	No genus predicted	(Jäckh et al., 2012; Walker et al., 2013)
2	Leucine	Antistatic, hair conditioning, skin conditioning	Transferring nitrogenous groups	Acting on the CH-NH2 group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen; acyltransferases; transferring nitrogenous groups; acting on carbon-nitrogen bonds, other than peptide bonds; intramolecular transferases; forming carbon-oxygen bonds	Branched chain amino acid aminotransferase	Isoleucine N-monooxygenase; valine N-monooxygenase; leucine N-acetyltransferase; branched chain amino acid transferase; leucine transaminase; leucine 2,3-aminomutase; leucine tRNA ligase; etc.	*Corynebacterium*; *Staphylococcus*	*Zymomonas; Pseudomonas; Burkholderia; Streptomyces; Staphylococcus; Yersinia; Corynebacterium*; etc.	(Fredrich et al., 2013; Holeček, 2018)
3	Glycerol	Perfuming, denaturant, humectant, solvent, hair conditioning, viscosity controlling	Acting on the CH-OH group of donors; transferring phosphorus-containing groups	Acting on the CH-OH group of donors; glycosyltransferases; transferring phosphorus-containing groups; acting on ester bonds; glycosylases; carbon-carbon lyases; carbon-oxygen lyases	Glycerol kinase; glycerol dehydrogenase	Glycerol kinase; glycerol dehydrogenase; 1,2-alpha-glucosylglycerol phosphorylase; acylglycerol lipase; glycerol phosphatase; dihydroxy acid dehydratase; etc.	*Staphylococcus*	*Zymomonas; Xylella; Xanthomonas; Gluconobacter; Geobacillus; Staphylococcus; Streptococcus*; etc.	(Fredrich et al., 2013; Ruzheinikov et al., 2001; Xue et al., 2017)
4	Benzo(a)pyrene[b]	Environmental pollutant found in soot, tobacco smoke, diesel exhaust etc.	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen	Cytochrome P450: CYP1A1, monooxygenases, dioxygenases	Naphthalene 1,2-dioxygenase; unspecific monooxygenases; steroid 21-monooxygenase; etc.	*Pseudomonas; Micrococcus; Bacillus*; etc.	*Pseudomonas; Bacillus; Burkholderia; Polaromonas; Ralstonia*	(Ahmad and Mukhtar, 2004)
5	Retinoic acid	Antiseborrheic	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the aldehyde or oxo group of donors; Glycosyltransferases	Cytochrome P450: CYP1A1	Unspecific monooxygenase; alkane monooxygenase; retinal dehydrogenase; aldehyde oxidase; glucuronosyl transferase; etc.	NA	*Escherichia; Pseudomonas; Alcanivorax; Methylobacterium; Jonesia*; etc.	(Ahmad and Mukhtar, 2004)

(Continued on next page)

Table 3. *Continued*

Sr. No.	Compound	Function	Reaction subclass from literature	Reaction subclass from SkinBug	Enzyme from literature	Enzyme from SkinBug	Skin microbial genus from literature	Skin microbes genus from SkinBug	References
6	7,12-Dimethylbenz (a)anthracene	Environmental pollutant found in tobacco smoke	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen	Cytochrome P450: CYP1A1, cytochrome P450: CYP1B1	Unspecific monooxygenase; aromatase; estradiol 6-beta monooxygenase	NA	No genus predicted	(Ahmad and Mukhtar, 2004)
7	Vitamin D3	Skin conditioning	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on the CH-CH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen	Cytochrome P450	Vitamin D3,24-hydroxylase; calcidiol 1-monooxygenase; vitamin D1, 25-hydroxylase	NA	*Nocardia; Verrucosispora; Stackebrandtia*	(Ahmad and Mukhtar, 2004)
8	4-Allylanisole	Perfuming	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on the CH-OH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen; transferring one-carbon groups	Cytochrome P450	4-(hydroxymethyl) benzenesulfonate dehydrogenase; (iso) eugenol O-methyltransferase; trans-anol O-methyltransferase	NA	No genus predicted	(Jäckh et al., 2012)
9	Isoeugenol	Perfuming, masking, drug: local antiseptic and analgesic	Acting on the CH-OH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on the CH-OH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen; transferring one-carbon groups	Cytochrome P450, cytochrome P450: CYP2E1, alcohol dehydrogenase	Alcohol dehydrogenase; isoeugenol synthase; salicylate 1-monooxygenase; 4-hydroxypheny lacetate 3-monooxygenase; catechol O-methyltransferase; etc.	NA	*Thermus; Zymomonas; Staphylococcus; Pseudomonas; Bacillus; Acinetobacter*, etc.	(Jäckh et al., 2012)
10	4-Phenylenediamine	Hair dyeing	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on other nitrogenous compounds as donors; acting on paired donors, with incorporation or reduction of molecular oxygen; carbon-carbon lyases	Cytochrome P450	Nitrobenzene nitroreductase; histidine decarboxylase	NA	*Vibrio; Raoultella; Pseudomonas; Ralstonia; Variovorax*; etc.	(Jäckh et al., 2012)

(*Continued on next page*)

Table 3. *Continued*

Sr. No.	Compound	Function	Reaction subclass from literature	Reaction subclass from SkinBug	Enzyme from literature	Enzyme from SkinBug	Skin microbial genus from literature	Skin microbes genus from SkinBug	References
11	Aniline[b]	Outdoor air, tobacco smoke	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on other nitrogenous compounds as donors; acting on paired donors, with incorporation or reduction of molecular oxygen; transferring one-carbon groups; acyltransferases; glycosyltransferases; transferring sulfur-containing groups; acting on carbon-nitrogen bonds, other than peptide bonds; carbon-carbon lyases; forming carbon-nitrogen bonds	Cytochrome P450	Phenol 2-monooxygenase; azobenzene reductase; arylamine N-acetyltransferase; arylamine glucosyltransferase; amine sulfotransferase; aryl acylamidase; aminobenzoate decarboxylase; gamma-glutamylanilide synthase; etc.	NA	*Bacillus; Staphylococcus; Mycobacterium*	(Jäckh et al., 2012)
12	Methyl red or azo dye[a]	Textile dyes, tattoo inks, and cosmetic colorant	Acting on other nitrogenous compounds as donors	Acting on the CH-CH group of donors; acting on other nitrogenous compounds as donors; acting on single donors with incorporation of molecular oxygen (oxygenases); acting on paired donors, with incorporation or reduction of molecular oxygen	Azoreductase (Azo1)	FMN-dependent NADPH azoreductase (Azo1); NAD(P)H-dependent oxidoreductase; 4-(dimethylamino) phenylazoxybenzene reductase; azobenzene reductase	*Staphylococcus; Corynebacterium; Micrococcus; Dermacoccus; Kocuria*	*Staphylococcus; Bacillus*	(Stingley et al., 2010)
13	Fluorouracil	Used in actinic keratosis and skin warts conditions	Glycosyltransferases; acting on the CH-CH group of donors	Glycosyltransferase; acting on the CH-CH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen; acting on CH or CH_2 groups; glycosylases; acting on carbon-nitrogen bonds, other than peptide bonds	Orotate phosphoribosyl transferase; uridine phosphorylase; thymidine phosphorylase; dihydropyrimidine dehydrogenase	Orotate phosphoribosy ltransferase; uridine phosphorylase; thymidine phosphorylase; dihydropyrimidine dehydrogenase; unspecific monooxygenase	NA	*Yersinia; Wolinella; Vibrio; Tolumonas; Shigella; Aeromonas; Xylella*; etc.	(Amirfallah et al., 2018; Longley et al., 2003)

(Continued on next page)

Table 3. *Continued*

Sr. No.	Compound	Function	Reaction subclass from literature	Reaction subclass from SkinBug	Enzyme from literature	Enzyme from SkinBug	Skin microbial genus from literature	Skin microbes genus from SkinBug	References
14	Propylene glycol	Used as demulcent and in skin lotions and ointments	Acting on the CH-OH group of donors	Acting on the CH-OH group of donors; carbon-carbon lyases; carbon-oxygen lyases	Alcohol dehydrogenases	L-glycol dehydrogenase; aldehyde reductase; glycerol dehydrogenase; lactaldehyde reductase; propanediol dehydratase; etc.	NA	*Salmonella; Escherichia; Geobacillus; Citrobacter; Klebsiella*; etc.	(Ewaschuk et al., 2005; Kraut and Kurtz, 2008)
15	Hydroquinone	Demelanizing agent	Acting on single donors with incorporation of molecular oxygen; glycosyltransferases	Acting on single donors with incorporation of molecular oxygen (oxygenases); acting on paired donors, with incorporation or reduction of molecular oxygen; glycosyltransferases; transferring alkyl or aryl groups, other than methyl groups; glycosylases; carbon-carbon lyases; etc.	Dioxygenases; quinone oxidoreductase	Catechol 1 2-dioxygenase; hydroquinone 1 2-dioxygenase; toluene 4-monooxygenase; p-benzoquinone reductase; 4-hydroxybenzoate 1-hydroxylase; arylesterase; hydroquinone glucosyltransferase; etc.	NA	*Pseudomonas; Bacillus; Escherichia; Acinetobacter*; etc.	(McDonald et al., 2001; Zhang et al., 2012)
16	Para-aminobenzoic acid	Used as sunscreen and melanizing agent. Also used in treatment of fibrotic skin disorders	Transferring alkyl or aryl groups, other than methyl groups	Transferring alkyl or aryl groups, other than methyl groups; acting on paired donors, with incorporation or reduction of molecular oxygen; transferring one-carbon groups; carbon-carbon lyases; etc.	Dihydropteroate synthase	Benzoate 1 2-dioxygenase; phenylethanolamine N-methyltransferase; dihydropteroate synthase; aminobenzoate decarboxylase; etc.	NA	*Yersinia; Vibrio; Streptococcus; Shigella; Streptomyces*; etc.	(Wegkamp et al., 2007)
17	Benzophenone	Used as organic sunscreen and melanizing agent. Also used as a fragrance enhancer	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the CH-OH group of donors	Cytochrome P450 (CYPs)	Nitrilotriacetate monooxygenase	NA	*Sorangium; Anabaena; Burkholderia; Sorangium*; etc.	(Watanabe et al., 2015)

(*Continued on next page*)

Table 3. *Continued*

Sr. No.	Compound	Function	Reaction subclass from literature	Reaction subclass from SkinBug	Enzyme from literature	Enzyme from SkinBug	Skin microbial genus from literature	Skin microbes genus from SkinBug	References
18	Lindane	Used in scabies and pediculosis skin conditions	Acting on halide bonds; acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on halide bonds; acting on paired donors, with incorporation or reduction of molecular oxygen; etc.	Dehydrochlorinase enzyme	Haloalkane dehalogenase	NA	*Sphingobium; Psychrobacter; Phenylobacterium; Caulobacter*; etc.	(Macholz and Kujawa, 1985; Nagata et al., 1993; Tanaka et al., 1979)
19	Ethinyl estradiol	Used for the treatment of moderate acne vulgaris (common acne) in females	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the CH-OH group of donors; glycosyltransferases; etc.	Cytochrome P450 (CYPs)	Unspecific monooxygenase; aromatase; 3-beta-hydroxy-delta-5-steroid dehydrogenase; etc.	NA	No genus predicted	(Stimmel et al., 1951; Wang et al., 2004)
20	Docosanol	Used in the treatment of herpes virus infection-caused recurring episodes of small, painful, fluid-filled blisters on the skin	Acting on the CH-OH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on the CH-OH group of donors; acting on paired donors, with incorporation or reduction of molecular oxygen	NA	Hexadecanol dehydrogenase; long-chain-alcohol dehydrogenase; naphthalene 1 2-dioxygenase; etc.	NA	*Geobacillus; Zymomonas; Staphylococcus; Bacillus*; etc.	(Pope et al., 1996)
21	Lauric acid	Used on skin for its antibacterial properties and ability to effectively combat acne	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the aldehyde or oxo group of donors	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the aldehyde or oxo group of donors; acting on ester bonds	Fatty acid beta-oxidation enzymes such as acyl CoA dehydrogenase, hydrolase, etc.	Trimethyllysine dioxygenase; aldehyde dehydrogenase; oleoyl-(acyl-carrier-protein) hydrolase; clavaminate synthase; etc.	NA	*Bacillus; Mycobacterium; Escherichia; Thermus*; etc.	(Dayrit, 2015)
22	Methyl lactate[b]	Used as a soothing and cooling agent for skin. Also to provide relief against itching and irritation on skin	Acting on the CH-OH group of donors	Acting on the CH-OH group of donors; acting on the aldehyde or oxo group of donors; acyltransferases; etc.	Lactate dehydrogenase	L-lactate dehydrogenase; glyoxylate reductase; oxaloglycolate reductase (decarboxylating); pyruvate synthase; formate C-acetyltransferase; etc.	*Staphylococcus; Micrococcus*; etc.	*Staphylococcus; Vibrio; Micrococcus; Shewanella*; etc.	(Gladden, 2004; Lam et al., 2018)

(Continued on next page)

Table 3. *Continued*

Sr. No.	Compound	Function	Reaction subclass from literature	Reaction subclass from SkinBug	Enzyme from literature	Enzyme from SkinBug	Skin microbial genus from literature	Skin microbes genus from SkinBug	References
23	Methylparaben[b]	Used in cosmetic products as preservative to give products a longer shelf life	Glycosyltransferases; acting on ester bonds; acting on paired donors, with incorporation or reduction of molecular oxygen	Glycosyltransferases; acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the CH-OH group of donors; etc.	Glucuronosyltransferase	Cyanohydrin beta-glucosyltransferase; 4-hydroxypheny lacetaldehyde oxime monooxygenase; 4-hydroxybenzoate 1-hydroxylase; 3-phenylpropanoate dioxygenase; etc.	NA	*Pseudomonas; Escherichia; Bacillus; Mycobacterium*; etc.	(Abbas et al., 2010; Moos et al., 2016)
24	Triclosan	Used as antiseptic and antibacterial on the skin	Acting on paired donors, with incorporation or reduction of molecular oxygen; glycosyltransferases	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on single donors with incorporation of molecular oxygen (oxygenases); acting on the CH-CH group of donors	Sulfur oxygenase/reductase; cytochrome P450 (CYPs); glucuronosyltransferase	Sulfur oxygenase/ reductase; carbazole 1,9a-dioxygenase; naphthalene 1,2-dioxygenase; persulfide dioxygenase; etc.	NA	*Pseudomonas; Bacillus; Streptococcus; Yersinia*; etc.	(Fang et al., 2016; Wang et al., 2018)
25	Terbinafine[b]	Used to treat the fungal infections on skin	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen	Cytochromes (CYPs)	Ammonia monooxygenase	NA	*Nitrosomonas; Nitrosospira*	(Vickers et al., 1999)
26	Alpha-tocopherol[b]	Used in cosmetic products to prevent UV damage to the skin	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the CH-OH group of donors; transferring one-carbon groups	Cytochromes (CYPs)	Validamycin A dioxygenases; aurachin C monooxygenase/ isomerase; calcidiol 1-monooxygenase; thymidylate synthase; etc.	*Deinococcus; Stenotrophomonas*	*Deinococcus; Stenotrophomonas; Zymomonas; Yersinia*; etc.	(Johnson et al., 2013; Timm et al., 2020)
27	Cholesterol[b]	Used as an emollient in cosmetic products such as eye makeup, face makeup, skin lotions, creams, and hair care formulations	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on the CH-OH group of donors; acting on the CH-CH group of donors; intramolecular oxidoreductases	Sterol 27-hydroxylase	Cholesterol 25-hydroxylase; vitamin D 25-hydroxylase; 11-beta-hydroxysteroid dehydrogenase; 3-beta-hydroxysteroid 3-dehydrogenase; etc.	*Streptococcus; Bacillus; Aerococcus*	*Streptococcus; Bacillus; Mycobacterium; Aeromonas*; etc.	(Iuliano, 2011; Timm et al., 2020)

(*Continued on next page*)

Table 3. ***Continued***

Sr. No.	Compound	Function	Reaction subclass from literature	Reaction subclass from SkinBug	Enzyme from literature	Enzyme from SkinBug	Skin microbial genus from literature	Skin microbes genus from SkinBug	References
28	Linoleic acid	Used in the conditions of skin irritation and to reduce acne breakouts	Acting on paired donors, with incorporation or reduction of molecular oxygen	Acting on paired donors, with incorporation or reduction of molecular oxygen; acting on single donors with incorporation of molecular oxygen (oxygenases); acting on ester bonds	Lipoxygenase; delta-6-desaturase	Linoleate 13S-lipoxygenase; linoleate 11-lipoxygenase; acyl-CoA 6-desaturase; palmitoyl-CoA hydrolase; etc.	*Paracoccus; Stenotrophomonas; Brevundimonas; Staphylococcus*; etc.	*Paracoccus; Stenotrophomonas; Brevundimonas; Staphylococcus*; etc.	(Brown et al., 2000; Gardner, 1970; Timm et al., 2020)

[a]Threshold for the subclass prediction models of "Oxidoreductases" was set from 0.5 to 0.1, and for "Transferases," "Hydrolases," and "Lyases" the thresholds were set from 0.5 to 0.2. The Tanimoto coefficient threshold was set to 0.5.
[b]Tanimoto coefficient threshold was set to 0.5 or 0.3.

our skin microbiome, along with the information on respective reaction centers, metabolic enzymes, microbial species carrying these enzymes, and also the skin sites harboring these species, "SkinBug" is proposed to be a unique, reliable and user-friendly tool.

One of the key contributions of this study is the construction of the skin microbiome-specific metabolic database using the pangenomes instead of individual genomes that helped in incorporating the metabolic potential of all the strains of a given species. Furthermore, we used only the high-quality pangenomes of 897 species that helped in the construction of the first comprehensive metabolic enzyme database of skin microbiome, which is the integral part of the SkinBug tool for making reliable metabolic predictions. The inclusion of information on bacterial species from 19 different skin sites and various skin-specific niches including anaerobic/aerobic niches and carbon- and sulfur-rich/-limiting niches in the skin microbiome database was very important for comprehensive training and achieving higher prediction accuracy. It was observed that the skin microbiome database had a higher representation of aerobic species because the metagenomic studies so far have been carried out majorly for the aerobic niches than the anaerobic niches due to the difficulty associated with anaerobic niches in sampling, isolation, and cultivation (Sfriso et al., 2020). Another key aspect of SkinBug is the inclusion of all the well-annotated and manually curated reactions of the KEGG database and their primary substrates for the construction of training set, analysis, and modeling. Likewise, the manual curation of database entries for the selected bacterial species and inclusion of only the complete bacterial genomes from NCBI RefSeq database, along with the usage of very strict thresholds, helped in the exclusion of any false-positives that often lead to errors in the predictions. Furthermore, SkinBug exploits the structural and chemical properties of substrates using chemical descriptors, linear fingerprints, and circular fingerprints, and thus is able to predict all the possible reactions for the given molecule. Therefore, it is also well-equipped to address the cases of enzymatic moonlighting and promiscuity because any additional substrate (for additional metabolic reaction) in such cases will also need to have similar structural and chemical properties as the original substrate. These features ensure the accuracy and wide applicability of the SkinBug tool for metabolic prediction of diverse biotic and xenobiotic molecules. The tool has a modular structure, which helps in easy updates of databases and models.

Due to the different levels of secretions, environmental exposure, and topography of skin sites, microbiome variability and the consequential metabolic variability across sites are expected that was also observed in this study. Of the 2,523 unique reactions that were found to be occurring in skin microbiome, only 277 were common to all the sites and the rest were specific to the different skin sites. The results pointed toward the need for site-specific metabolic prediction for skin microbiome, and this is one of the key features of SkinBug.

It was apparent from the hierarchical clustering of skin sites that the sites with similar physical and physiological properties clustered together. The exposed, dry, and desiccated skin sites such as the arm, foot, and forearm clustered together as they have similar physical and physiological properties such as moisture level, pH, and environmental exposure. Similarly, the sites with high amount of secretions such as diseased sites, face, and skin surface invaginations clustered together. Likewise, the groin and retroauricular crease sites are similar in terms of moisture levels and secretions, and pressure ulcer and venous leg ulcer are both the diseased sites with similar physiology, thus their clustering was also expected and observed in this study. Furthermore, it was intriguing to observe that cheek, perineum, and inner wrist clustered together. The clustering of cheek and perineum was expected because both sites have a lot of secretions by glands such as sebaceous, sweat, and hepatoid, and thus may harbor similar microbiome, which was also reported by other studies (Grice and Segre, 2011). However, the clustering of inner wrist, which is a desiccated and dry site, with check and perineum was surprising. The plausible explanation for it emerges from the recent studies which reported that the dry and desiccated sites harbor a highly dynamic flora, and a phylogenetic diversity that is even higher than the diversity of gut and oral cavity microbiomes of the same individual (Costello et al., 2009; Grice and Segre, 2011). In summary, it was evident that the bacterial species and the corresponding metabolic reactions observed at the different skin sites associates with the physiological and physical properties of the skin sites such as topographical location, different secretions, moisture levels, and environmental exposure (Grice et al., 2009; Grice and Segre, 2011).

However, there were some key challenges during the development of SkinBug. The first one was to train the multiclass multilabel classification models from the limited dataset of well-annotated reactions with sufficient accuracy. To overcome this challenge, the integration of both the machine learning and neural

network methods, selection of optimum algorithms, and rigorous optimization of selected models helped to achieve the high multiclass multilabel classification accuracy. The highly skewed distribution of metabolic reactions in the six reaction classes and their subclasses constituted the second challenge for optimal training of models. To tackle this challenge, the usage of a modified version of the stratified random sampling that included down-sampling, in addition to the standard stratified random sampling method, helped in effectively splitting of the data into training and testing for efficient predictive modeling.

The third challenge was the hierarchical nature of the complete reactions annotated as the four-digit EC number, for which the prediction strategies were also needed to be hierarchical and accurate at each step of the prediction to achieve accurate four-digit EC number prediction. To overcome this challenge, an integrated approach employing machine learning and neural network models to predict the reaction class, followed by construction of six different machine learning models to predict the reaction subclasses, and a molecular similarity search-guided k-nearest neigbors method to predict the complete reactions were incorporated into SkinBug.

Comprehensive validation of a prediction tool is important to ensure reliable and accurate predictions, which in the case of SkinBug was carried out using a diverse set of molecules (cosmetic, pharmaceutical, pollutants, natural molecules present on skin, etc.) that includes the cases of aerobic and anaerobic metabolism at various skin sites including the dry, moist, and sebaceous niches. The results of validation were well supported by the experimental studies and reference databases. The accuracy of the predictions was also exemplified by the selected case studies of BaP and azo dyes where SkinBug predicted the correct metabolic reactions, enzymes, and microbial species as known from the experimental studies and reference databases (Gibson et al., 1975; Jiang et al., 2007; Schwarz et al., 2001; Sowada et al., 2014; Zhou et al., 2017).

Taken together, the robust methodology, training, comprehensive validations, and their biological significance make SkinBug a very useful, accurate, and reliable method to predict the metabolic reaction, enzyme, microbial species, and the specific skin site for any given molecule. However, for the occurrence of a metabolic reaction, the xenobiotic/biotic molecule needs to come in close vicinity of the active enzyme along with other favorable reaction conditions, and it is tempting to question if the predicted reaction will really occur given these conditions. The plausible biological reasoning to answer this emerges from the fact that bacterial species have an abundance of transporters such as outer membrane-associated β-barrel-containing proteins or porins, which may allow for the transport of biotic/xenobiotic molecules, and thus will facilitate their metabolism by the bacteria. One such example is the TonB-dependent transport system where the outer membrane-associated TonB-dependent transporter (TBDT), and other similar ATP-driven influx transporters, facilitate the active transport and subsequent metabolism of xenobiotic compounds by bacterial cells (Jindal et al., 2019; Samantarrai et al., 2020). The experimental supports to these processes are provided by a few recent studies that showed the metabolism of different biotic/xenobiotic molecules such as BaP, glycoholic acid, cholesterol, glycerol, azo dyes, arginine, triglyceride lipids, propylene glycol, palmitic acid, alpha-tocopherol, uric acid, lactic acid, ethanol amine, and linolenic acid on their incubation with the bacterial species of the skin microbiome (Sowada et al., 2014; Stingley et al., 2010; Timm et al., 2020).

The case studies along with biological validations confirm the accuracy and reliability of SkinBug with strong control over false-positives and false-negatives, highlight the potential of SkinBug in revealing the metabolic consequences of undesired metabolism by our skin microbiome, and provide leads for further experimental studies. It is also anticipated to have applications in skin microbiome and exposome studies, in the development of novel diagnostic and therapeutic approaches in dermatology, and in cosmetics industry for developing safer, more effective and population-specific products.

Limitations of the study

The availability of complete and annotated genomes of bacterial species for skin microbiome is one of the limiting factors that determines the size and comprehensiveness of the metabolic database for the skin microbiome. Similarly, machine learning and neural network models are trained on available metabolic reaction information, which currently has a lower representation of Isomerases and Ligases reaction classes. Thus, the availability of more information on genomes and metabolic reactions in the reference databases will further strengthen and broaden the applicability of SkinBug.

Conclusion

Our skin comes in regular contact with many chemical molecules due to pollution or usage of skin care products, thus the prediction and evaluation of metabolism of these molecules by skin microbiome species was much needed for which SkinBug is a valuable contribution. The integrated approach of using machine learning, neural networks, and chemoinformatics along with the database of metabolic enzymes for skin microbiome species constructed in this study helped in prediction of all the possible metabolic reactions that can occur to a given molecule, their respective reaction centers, the metabolic enzymes that can perform the predict reactions, species that harbors these metabolic enzymes, and the skin sites that carry these species. The case studies along with biological validations presented here attest to the accuracy, applicability, and potential of SkinBug. This first state-of-the-art tool to predict the metabolism of a given molecule by the skin microbiome will be very useful for the development of novel therapeutic approaches in dermatology and cosmetics and will also provide leads for future experimental studies.

Resource availability

Lead contact

Further information, requests, and inquiries should be directed to the Lead Contact, Vineet K. Sharma (vineetks@iiserb.ac.in).

Materials availability

This study did not generate new data and only used the data available in public databases mentioned in the manuscript text.

Data and code availability

The published article includes all data used or analyzed during this study.

Methods

All methods can be found in the accompanying Transparent methods supplemental file.

Supplemental information

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101925.

Acknowledgments

The author SK.J. thanks Department of Science and Technology for the DST-INSPIRE fellowship. We thank the research grant (BT/PR34239/AI/133/23/2019) received from the Department of Biotechnology, India, and the intramural research funds provided by IISER Bhopal. We acknowledge the help of Mr. Vishnu Prasoodanan PK in preparing the header image for SkinBug web server and Graphical Abstract.

Author contributions

V.K.S. conceived and coordinated the project. S.K.J. and V.K.S. designed the computational framework of the study. S.K.J., S.M.A., and P.T. constructed the site-specific species database. S.K.J. constructed the skin microbiome-specific metabolic information database and substrate database, performed exploratory data analysis, and optimized the machine learning models. S.K.J. and S.M.A. constructed the machine learning models. S.K.J. optimized and constructed the ANN models, KNN models, molecular similarity search modules, and reaction center prediction modules; performed the statistical analysis; constructed the web server; and constructed the manuscript figures. S.K.J. and V.K.S. interpreted the results. S.K.J. and V.K.S. wrote the manuscript. All the authors read and approved the final version of the manuscript.

Declaration of interests

The authors declare no competing interests.

Received: July 31, 2020
Revised: November 8, 2020
Accepted: December 7, 2020
Published: January 22, 2021

References

Abbas, S., Greige-Gerges, H., Karam, N., Piet, M.-H., Netter, P., and Magdalou, J. (2010). Metabolism of parabens (4-hydroxybenzoic acid esters) by hepatic esterases and UDP-glucuronosyltransferases in man. Drug Metab. Pharmacokinet. *25*, 568–577.

Ahmad, N., and Mukhtar, H. (2004). Cytochrome p450: a target for drug development for skin diseases. J. Invest. Dermatol. *123*, 417–425.

Amirfallah, A., Calibasi Kocal, G., Unal, O.U., Ellidokuz, H., Oztop, I., and Basbinar, Y. (2018). DPYD, TYMS and MTHFR genes polymorphism frequencies in a series of Turkish colorectal cancer patients. J. personalized Med. *8*, 45.

Babtie, A., Tokuriki, N., and Hollfelder, F. (2010). What makes an enzyme promiscuous? Curr. Opin. Chem. Biol. *14*, 200–207.

Bay, L., Barnes, C.J., Fritz, B.G., Thorsen, J., Restrup, M.E.M., Rasmussen, L., Sørensen, J.K., Hesselvig, A.B., Odgaard, A., and Hansen, A.J. (2020). Universal dermal microbiome in human skin. Mbio *11*, e02945-19.

Brown, J., Lindsay, R., and Riemersma, R. (2000). Linoleic acid metabolism in the spontaneously diabetic rat: Δ6-desaturase activity vs. product/precursor ratios. Lipids *35*, 1319–1323.

Byrd, A.L., Belkaid, Y., and Segre, J.A. (2018). The human skin microbiome. Nat. Rev. Microbiol. *16*, 143.

Chung, K.-T. (1983). The significance of azo-reduction in the mutagenesis and carcinogenesis of azo dyes. Mutat. Res. Rev. Genet. Toxicol. *114*, 269–281.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. Science *326*, 1694–1697.

Dayrit, F.M. (2015). The properties of lauric acid and their significance in coconut oil. J. Am. Oil Chem. Soc. *92*, 1–15.

Edwards, C., and Marks, R. (1995). Evaluation of biomechanical properties of human skin. Clin. Dermatol. *13*, 375–380.

Ewaschuk, J.B., Naylor, J.M., and Zello, G.A. (2005). D-lactate in human and ruminant metabolism. J. Nutr. *135*, 1619–1625.

Fang, J.L., Vanlandingham, M., da Costa, G.G., and Beland, F.A. (2016). Absorption and metabolism of triclosan after application to the skin of B 6 C 3 F 1 mice. Environ. Toxicol. *31*, 609–623.

Fredrich, E., Barzantny, H., Brune, I., and Tauch, A. (2013). Daily battle against body odor: towards the activity of the axillary microbiota. Trends Microbiol. *21*, 305–312.

Gao, J., Ellis, L.B., and Wackett, L.P. (2010). The University of Minnesota biocatalysis/biodegradation database: improving public access. Nucleic Acids Res. *38*, D488–D491.

Gardner, H. (1970). Sequential enzymes of linoleic acid oxidation in corn germ: lipoxygenase and linoleate hydroperoxide isomerase. J. Lipid Res. *11*, 311–321.

Gelboin, H.V. (1980). Benzo [alpha] pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes. Physiol. Rev. *60*, 1107–1166.

Gibson, D.T., Mahadevan, V., Jerina, D.M., Yogi, H., and Yeh, H. (1975). Oxidation of the carcinogens benzo [a] pyrene and benzo [a] anthracene to dihydrodiols by a bacterium. Science *189*, 295–297.

Gladden, L. (2004). Lactate metabolism: a new paradigm for the third millennium. J. Physiol. *558*, 5–30.

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., and Schmidhuber, J. (2016). LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. *28*, 2222–2232.

Grice, E.A., and Segre, J.A. (2011). The skin microbiome. Nat. Rev. Microbiol. *9*, 244.

Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C., Bouffard, G.G., Blakesley, R.W., Murray, P.R., and Green, E.D. (2009). Topographical and temporal diversity of the human skin microbiome. Science *324*, 1190–1192.

Hoffmann, M.F., Preissner, S.C., Nickel, J., Dunkel, M., Preissner, R., and Preissner, S. (2014). The Transformer database: biotransformation of xenobiotics. Nucleic Acids Res. *42*, D1113–D1117.

Holeček, M. (2018). Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. Nutr. Metab. *15*, 33.

Hult, K., and Berglund, P. (2007). Enzyme promiscuity: mechanism and applications. Trends Biotechnol. *25*, 231–238.

Iuliano, L. (2011). Pathways of cholesterol oxidation via non-enzymatic mechanisms. Chem. Phys. Lipids *164*, 457–468.

Jäckh, C., Fabian, E., van Ravenzwaay, B., and Landsiedel, R. (2012). Relevance of xenobiotic enzymes in human skin in vitro models to activate pro-sensitizers. J. Immunotoxicol. *9*, 426–438.

Jewison, T., Su, Y., Disfany, F.M., Liang, Y., Knox, C., Maciejewski, A., Poelzer, J., Huynh, J., Zhou, Y., and Arndt, D. (2014). SMPDB 2.0: big improvements to the Small molecule pathway database. Nucleic Acids Res. *42*, D478–D484.

Jiang, H., Gelhaus, S.L., Mangal, D., Harvey, R.G., Blair, I.A., and Penning, T.M. (2007). Metabolism of benzo [a] pyrene in human bronchoalveolar H358 cells using liquid chromatography–mass spectrometry. Chem. Res. Toxicol. *20*, 1331–1341.

Jindal, S., Yang, L., Day, P.J., and Kell, D.B. (2019). Involvement of multiple influx and efflux transporters in the accumulation of cationic fluorescent dyes by Escherichia coli. BMC Microbiol. *19*, 1–16.

Johnson, C.H., Bonzo, J.A., Cheng, J., Krausz, K.W., Kang, D.W., Luecke, H., Idle, J.R., and Gonzalez, F.J. (2013). Cytochrome P450 regulation by α-tocopherol in Pxr-null and PXR-humanized mice. Drug Metab. Dispos. *41*, 406–413.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Khersonsky, O., and Tawfik, D.S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu. Rev. Biochem. *79*, 471–505.

Khersonsky, O., Roodveldt, C., and Tawfik, D.S. (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. Curr. Opin. Chem. Biol. *10*, 498–508.

Kong, H.H., and Segre, J.A. (2012). Skin microbiome: looking back to move forward. J. Invest. Dermatol. *132*, 933–939.

Kong, H.H., Oh, J., Deming, C., Conlan, S., Grice, E.A., Beatson, M.A., Nomicos, E., Polley, E.C., Komarow, H.D., and Murray, P.R. (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. Genome Res. *22*, 850–859.

Kraut, J.A., and Kurtz, I. (2008). Toxic alcohol ingestions: clinical features, diagnosis, and management. Clin. J. Am. Soc. Nephrol. *3*, 208–225.

Kriegel, H.P., Kröger, P., Sander, J., and Zimek, A. (2011). Density-based clustering. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. *1*, 231–240.

Lam, T.H., Verzotto, D., Brahma, P., Ng, A.H.Q., Hu, P., Schnell, D., Tiesman, J., Kong, R., Ton, T.M.U., and Li, J. (2018). Understanding the microbial basis of body odor in pre-pubescent children and teenagers. Microbiome *6*, 1–14.

Longley, D.B., Harkin, D.P., and Johnston, P.G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. Nat. Rev. Cancer *3*, 330–338.

Macholz, R., and Kujawa, M. (1985). Recent state of lindane metabolism. Part III. In Residue Reviews, *94*, F.A. Gunther, ed (Springer), pp. 119–149.

McDonald, T., Holland, N., Skibola, C., Duramad, P., and Smith, M. (2001). Hypothesis: phenol and hydroquinone derived mainly from diet and gastrointestinal flora activity are causal factors in leukemia. Leukemia *15*, 10–20.

Moos, R.K., Angerer, J., Dierkes, G., Brüning, T., and Koch, H.M. (2016). Metabolism and elimination of methyl, iso-and n-butyl paraben in human urine after single oral dosage. Arch. Toxicol. *90*, 2699–2709.

Nagata, Y., Imai, R., Sakai, A., Fukuda, M., Yano, K., and Takagi, M. (1993). Isolation and characterization of Tn5-induced mutants of Pseudomonas paucimobilis UT26 defective in γ-hexachlorocyclohexane dehydrochlorinase (LinA). Biosci. Biotechnol. Biochem. *57*, 703–709.

Naik, S., Bouladoux, N., Wilhelm, C., Molloy, M.J., Salcedo, R., Kastenmuller, W., Deming, C., Quinones, M., Koo, L., and Conlan, S. (2012).

Compartmentalized control of skin immunity by resident commensals. Science *337*, 1115–1119.

Nakatsuji, T., Chiang, H.-I., Jiang, S.B., Nagarajan, H., Zengler, K., and Gallo, R.L. (2013). The microbiome extends to subepidermal compartments of normal skin. Nat. Commun. *4*, 1–8.

Nakayama, T., Kimura, T., Kodama, M., and Nagata, C. (1983). Generation of hydrogen peroxide and superoxide anion from active metabolites of naphthylamines and aminoazo dyes: its possible role in carcinogenesis. Carcinogenesis *4*, 765–769.

Pandya, C., Farelli, J.D., Dunaway-Mariano, D., and Allen, K.N. (2014). Enzyme promiscuity: engine of evolutionary innovation. J. Biol. Chem. *289*, 30229–30236.

Picardo, M., and Ottaviani, M. (2014). Skin microbiome and skin disease: the example of rosacea. J. Clin. Gastroenterol. *48*, S85–S86.

Pope, L., Marcelletti, J., Katz, L., and Katz, D. (1996). Anti-herpes simplex virus activity of n-docosanol correlates with intracellular metabolic conversion of the drug. J. Lipid Res. *37*, 2167–2178.

Ruzheinikov, S.N., Burke, J., Sedelnikova, S., Baker, P.J., Taylor, R., Bullough, P.A., Muir, N.M., Gore, M.G., and Rice, D.W. (2001). Glycerol dehydrogenase: structure, specificity, and mechanism of a family III polyol dehydrogenase. Structure *9*, 789–802.

Samantarrai, D., Lakshman Sagar, A., Gudla, R., and Siddavattam, D. (2020). TonB-dependent transporters in Sphingomonads: unraveling their distribution and function in environmental adaptation. Microorganisms *8*, 359.

SanMiguel, A., and Grice, E.A. (2015). Interactions between host factors and the skin microbiome. Cell Mol. Life Sci. *72*, 1499–1515.

Saxena. (2018). Comparison of healthy and dandruff **scalp microbiome** reveals the role of commensals in **scalp** health. Frontiers in cellular and infection microbiology. https://doi.org/10.3389/fcimb.2018.00346.

Schwarz, D., Kisselev, P., Cascorbi, I., Schunck, W.-H., and Roots, I. (2001). Differential metabolism of benzo [a] pyrene and benzo [a] pyrene-7, 8-dihydrodiol by human CYP1A1 variants. Carcinogenesis *22*, 453–459.

Sfriso, R., Egert, M., Gempeler, M., Voegeli, R., and Campiche, R. (2020). Revealing the secret life of skin-with the microbiome you never walk alone. Int. J. Cosmet. Sci. *42*, 116–126.

Sharma, A. (2017). A **novel approach** for the prediction of species-specific biotransformation of xenobiotic/drug molecules by the human gut microbiota. Scientific reports. https://doi.org/10.1038/s41598-017-10203-6.

Sowada, J., Schmalenberger, A., Ebner, I., Luch, A., and Tralau, T. (2014). Degradation of benzo [a] pyrene by bacterial isolates from human skin. FEMS Microbiol. Ecol. *88*, 129–139.

Stimmel, B.F., May, J.A., Randolph, J.A., and Conn, W.M. (1951). The metabolism of ethinyl estradiol in man. J. Clin. Endocrinol. *11*, 408–415.

Stingley, R.L., Zou, W., Heinze, T.M., Chen, H., and Cerniglia, C.E. (2010). Metabolism of azo dyes by human skin microbiota. J. Med. Microbiol. *59*, 108.

Tanaka, K., Kurihara, N., and Nakajima, M. (1979). Oxidative metabolism of lindane and its isomers with microsomes from rat liver and house fly abdomen. Pestic. Biochem. Physiol. *10*, 96–103.

Timm, C.M., Loomis, K., Stone, W., Mehoke, T., Brensinger, B., Pellicore, M., Staniczenko, P.P., Charles, C., Nayak, S., and Karig, D.K. (2020). Isolation and characterization of diverse microbial representatives from the human skin microbiome. Microbiome *8*, 1–12.

Vickers, A.E., Sinclair, J.R., Zollinger, M., Heitz, F., Glänzel, U., Johanson, L., and Fischer, V. (1999). Multiple cytochrome P-450s involved in the metabolism of terbinafine suggest a limited potential for drug-drug interactions. Drug Metab. Dispos. *27*, 1029–1038.

Vo-Dinh, T., Tromberg, B., Griffin, G., Ambrose, K., Sepaniak, M., and Gardenhire, E. (1987). Antibody-based fiberoptics biosensor for the carcinogen benzo (a) pyrene. Appl. Spectrosc. *41*, 735–738.

Walker, A.M., Hayes, R.P., Youn, B., Vermerris, W., Sattler, S.E., and Kang, C. (2013). Elucidation of the structure and reaction mechanism of sorghum hydroxycinnamoyltransferase and its structural relationship to other coenzyme A-dependent transferases and synthases. Plant Physiol. *162*, 640–651.

Wang, B., Sanchez, R.I., Franklin, R.B., Evans, D.C., and Huskey, S.-E.W. (2004). The involvement of CYP3A4 and CYP2C9 in the metabolism of 17α-ethinylestradiol. Drug Metab. Dispos. *32*, 1209–1212.

Wang, Y., Kuo, S., Shu, M., Yu, J., Huang, S., Dai, A., Two, A., Gallo, R.L., and Huang, C.-M. (2014). Staphylococcus epidermidis in the human skin microbiome mediates fermentation to inhibit the growth of Propionibacterium acnes: implications of probiotics in acne vulgaris. Appl. Microbiol. Biotechnol. *98*, 411–424.

Wang, S., Poon, K., and Cai, Z. (2018). Removal and metabolism of triclosan by three different microalgal species in aquatic environment. J. Hazard. Mater. *342*, 643–650.

Watanabe, Y., Kojima, H., Takeuchi, S., Uramaru, N., Sanoh, S., Sugihara, K., Kitamura, S., and Ohta, S. (2015). Metabolism of UV-filter benzophenone-3 by rat and human liver microsomes and its effect on endocrine-disrupting activity. Toxicol. Appl. Pharmacol. *282*, 119–128.

Wegkamp, A., van Oorschot, W., de Vos, W.M., and Smid, E.J. (2007). Characterization of the role of para-aminobenzoic acid biosynthesis in folate production by Lactococcus lactis. Appl. Environ. Microbiol. *73*, 2673–2681.

Xue, L.-L., Chen, H.-H., and Jiang, J.-G. (2017). Implications of glycerol metabolism for lipid production. Prog. Lipid Res. *68*, 12–25.

Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M. (2009). E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. Bioinformatics *25*, i179–i186.

Zeeuwen, P.L., Kleerebezem, M., Timmerman, H.M., and Schalkwijk, J. (2013). Microbiome and skin diseases. Curr. Opin. Allergy Clin. Immunol. *13*, 514–520.

Zhang, S., Sun, W., Xu, L., Zheng, X., Chu, X., Tian, J., Wu, N., and Fan, Y. (2012). Identification of the para-nitrophenol catabolic pathway, and characterization of three enzymes involved in the hydroquinone pathway, in pseudomonas sp. 1-7. BMC Microbiol. *12*, 1–11.

Zhou, S., Yeung, L.W., Forbes, M.W., Mabury, S., and Abbatt, J.P. (2017). Epoxide formation from heterogeneous oxidation of benzo [a] pyrene with gas-phase ozone and indoor air. Environ. Sci. Process. Impacts *19*, 1292–1299.

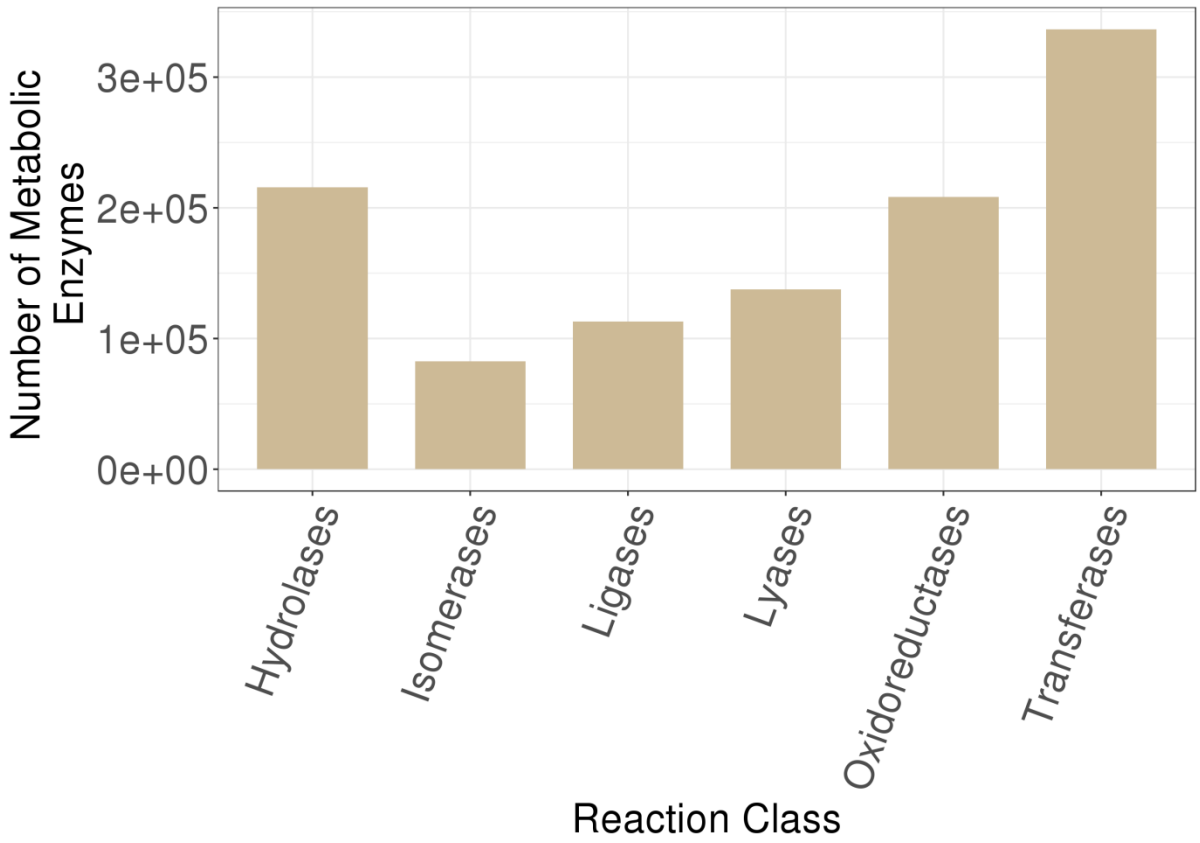
iScience, Volume *24*

Supplemental Information

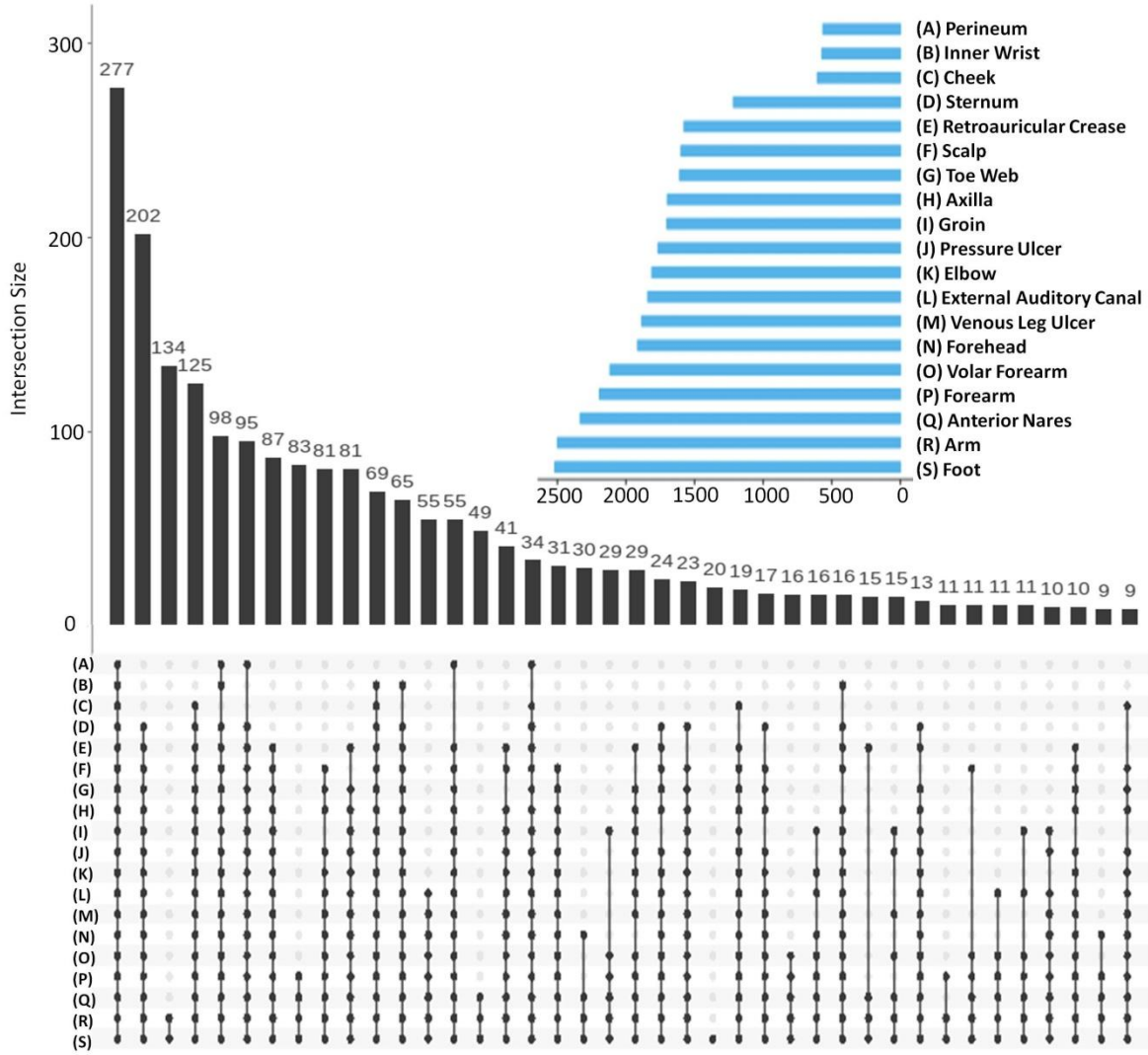
SkinBug: an artificial intelligence approach to predict human skin microbiome-mediated metabolism of biotics and xenobiotics

Shubham K. Jaiswal, Shitij Manojkumar Agarwal, Parikshit Thodum, and Vineet K. Sharma

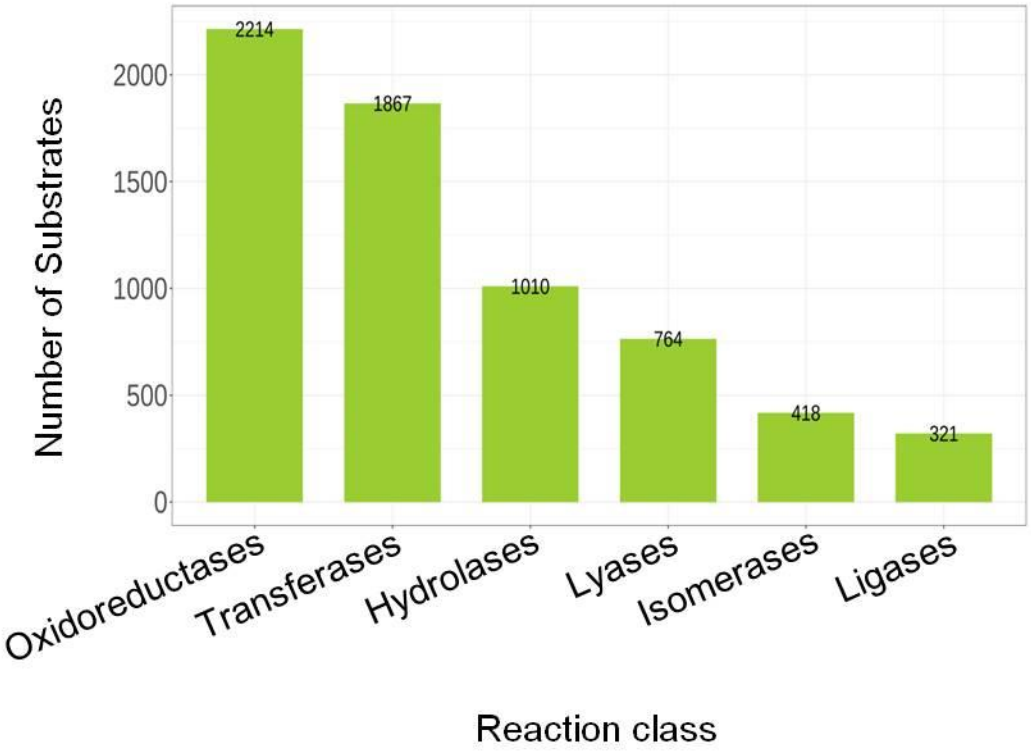
SUPPLEMENTARY FIGURES



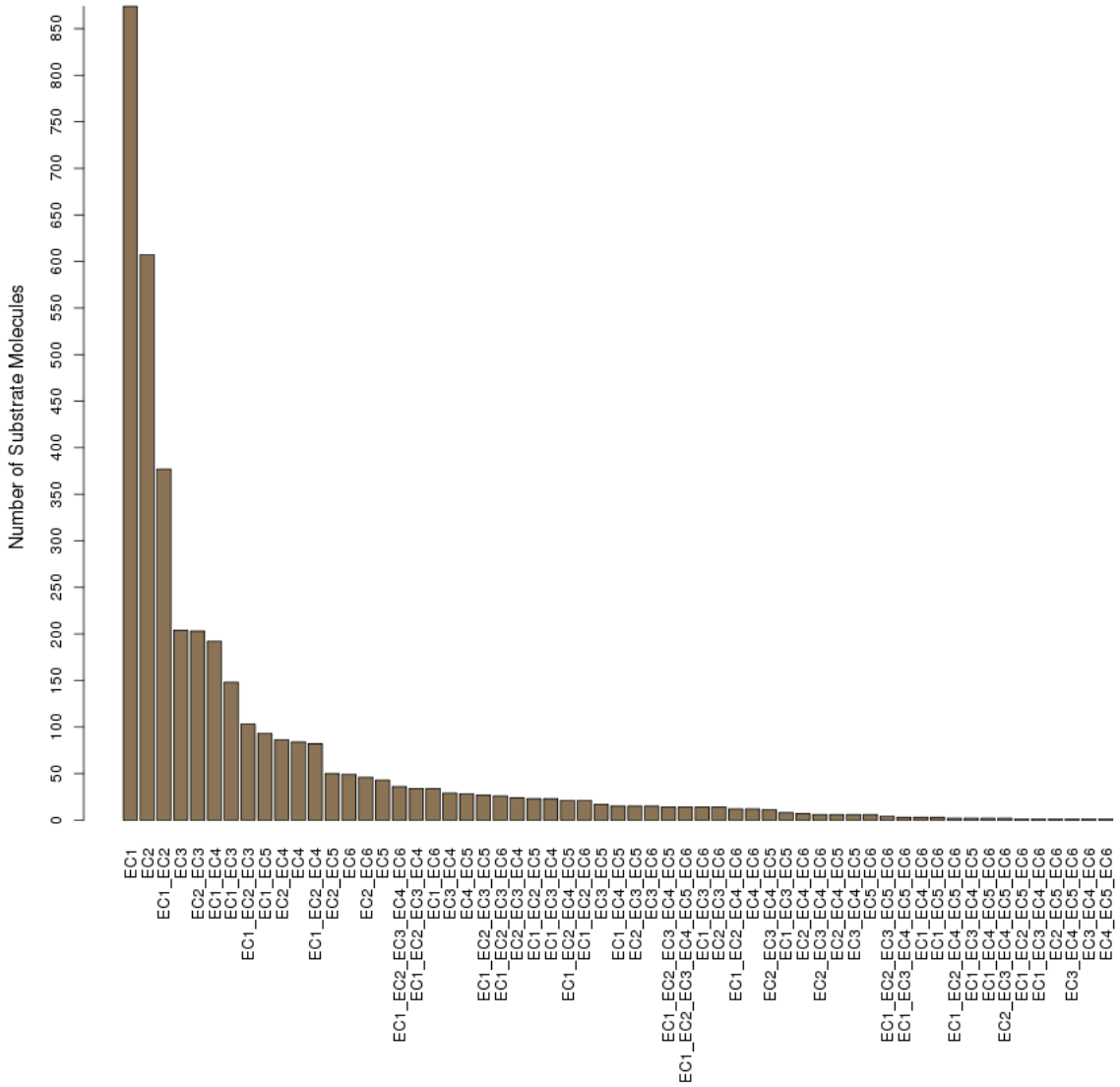
Supplementary Figure S1: Bar plot of number of metabolic enzymes from different reaction classes present in the human skin microbiome (Related to Figure 1)



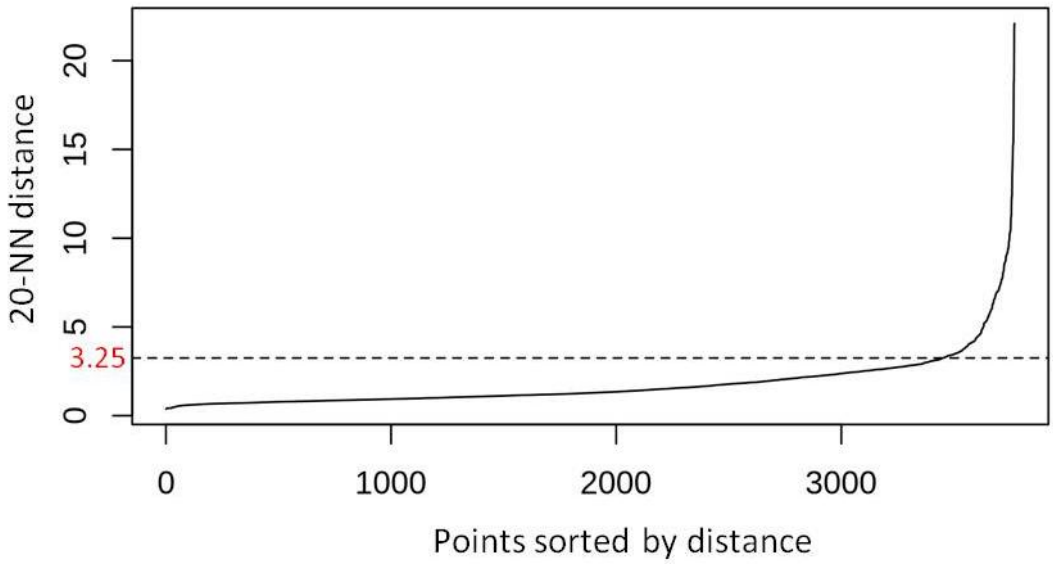
Supplementary Figure S2: Matrix layout for the intersection of unique enzymatic reactions from 19 different skin sites. The blue horizontal bar depicts the absolute number of unique enzymatic reactions that can occur on the each skin site. The vertical bar plot represents the number of unique enzymatic reactions (top of bars) shared by the different skin sites. The sites that share a particular number are shown as the intersection of filled ellipsoids at x-axis. The ellipsoids are placeholders for individual skin site. (Related to Figure 1)



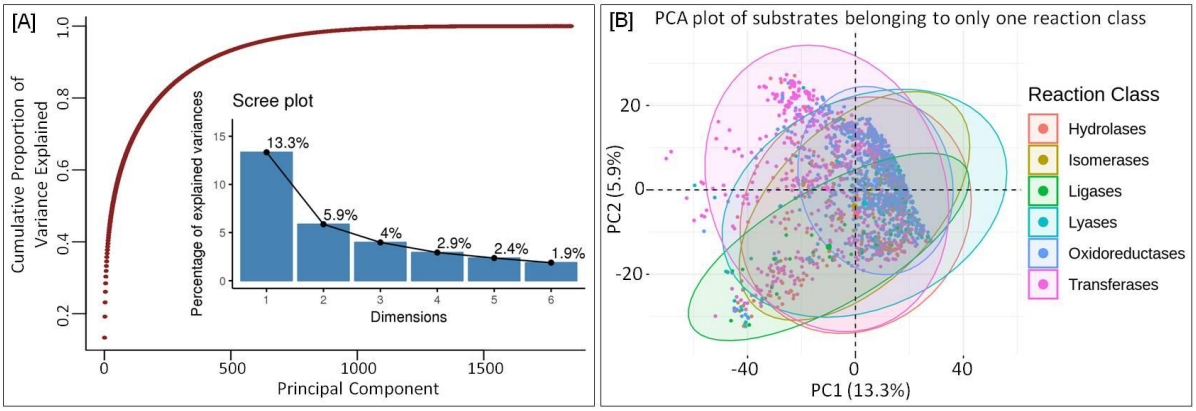
Supplementary Figure S3: Distribution of substrate molecules across different reactions classes (annotated as one-digit EC number). Here all inclusive approach has been used where if one molecule can undergo reactions from different reaction claseses the count is incremented in all of them. (Related to Figure 1)



Supplementary Figure S4: Distribution of 3,769 substrate molecules across different reactions classes (annotated as one-digit EC number) unique combinations. (Related to Figure 1)



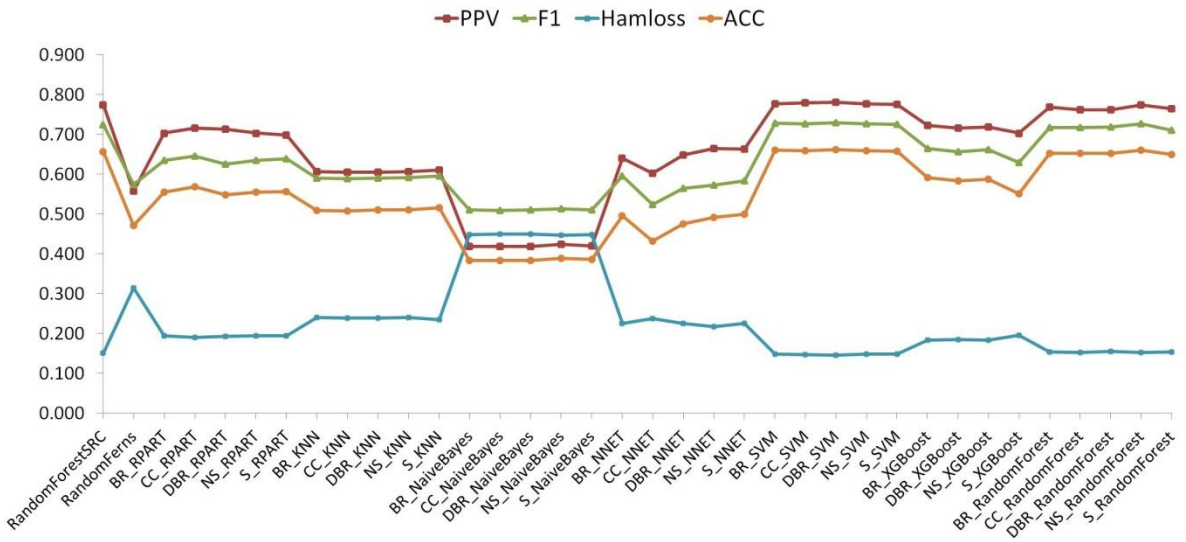
Supplementary Figure S5: The distribution of 20-nearest neighbours distances across the dataset of 3,769 substrate molecules based on the selected 2,322 variables. The knee point at 3.25 is the most suitable epsilon value for the density based clustering using DBSCAN. (Related to Figure 2)



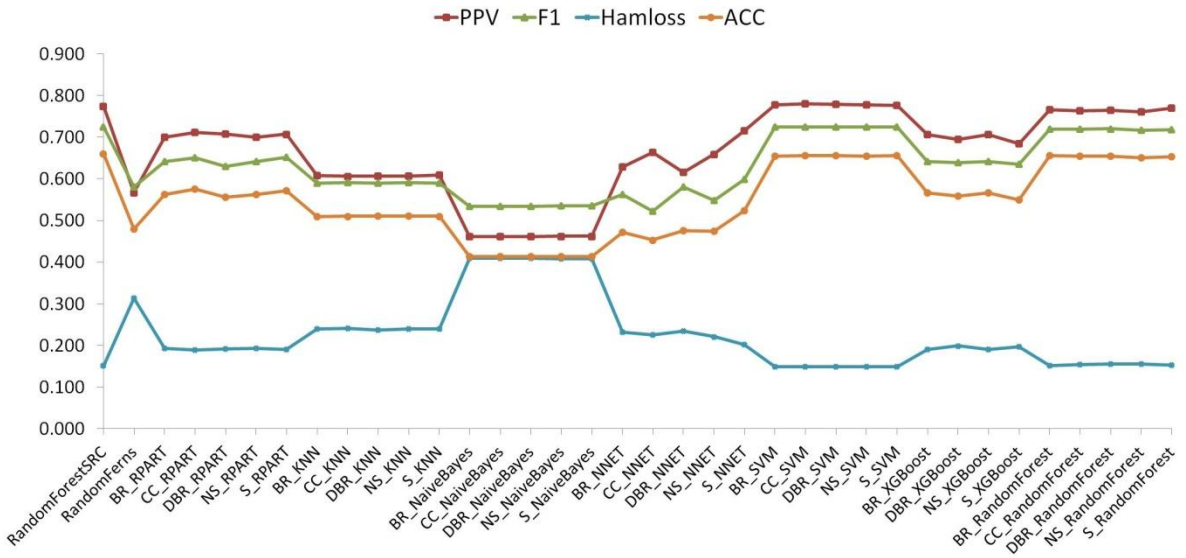
Supplementary Figure S6: Evaluating the diversity and complexity of a subset dataset where one substrate can undergo only one type of reaction (Related to Figure 2)

[A] The cumulative scree plot and a normal scree plot from the PCA analysis of all the substrate molecules from the dataset of substrates that can undergo only one type of reaction class. All the selected 2,322 features were used for performing the PCA analysis. The x-axis is the principal component number, y-axis for the dot plot is the cumulative variance explained by the individual principal components, and the y-axis for the bar plot is the percentage of variance explained by the individual principal components.

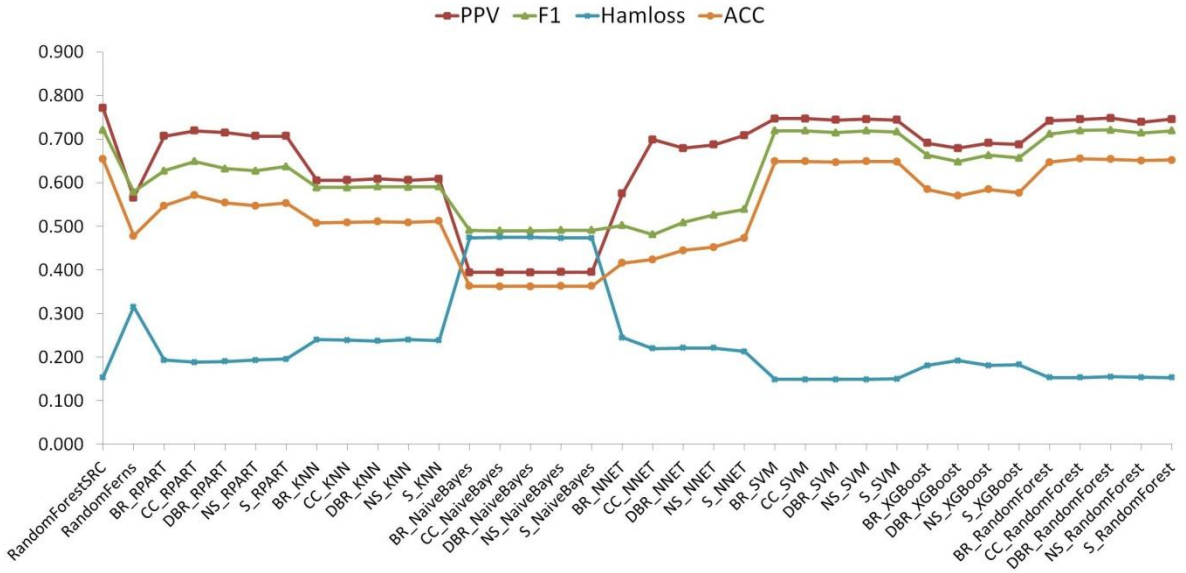
[B] The PCA plot of the substrate molecules that can undergo only one type of reaction class using the principal component PC-1 and PC-2 from the PCA analysis. Different reaction classes are coloured differently and the ellipsoids are drawn for each reaction class based on the distribution of substrate molecules in the plot.



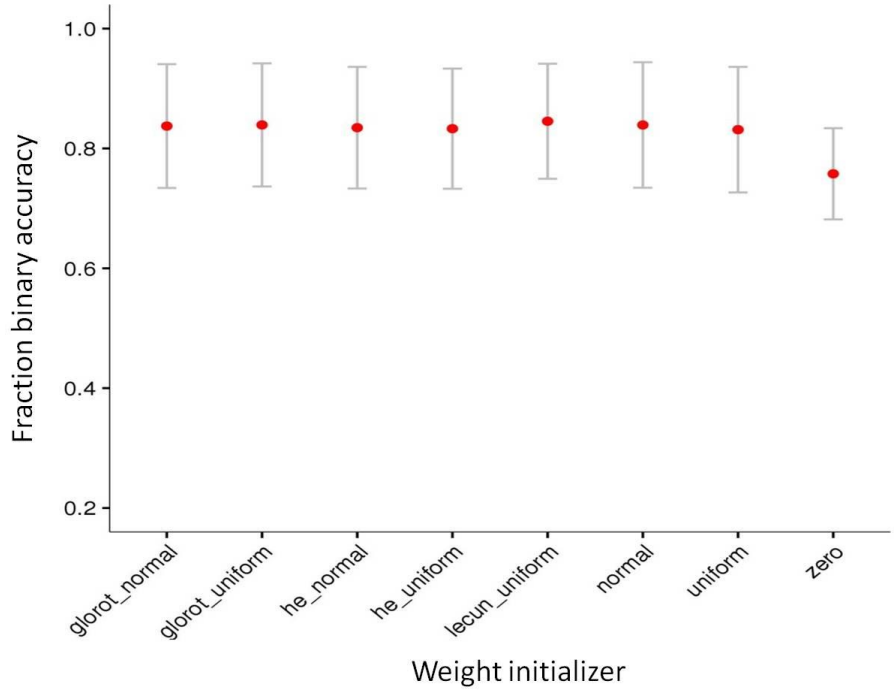
Supplementary Figure S7: The performance of different multiclass multilabel classification algorithms on the multilabel dataset with ECFP fingerprints, FCFP fingerprints, boruta selected descriptors, and boruta selected fingerprints. ACC – Multilabel accuracy, PPV – Multilabel precision or Multilabel positive predicted value, Hamloss – Hamming loss, F1 – Multilabel F1 score. Binary relevance – BR, Classifier chains – CC, Nested stacking – NS, Dependent binary relevance – DBR, Stacking – S. (Related to Figure 4)



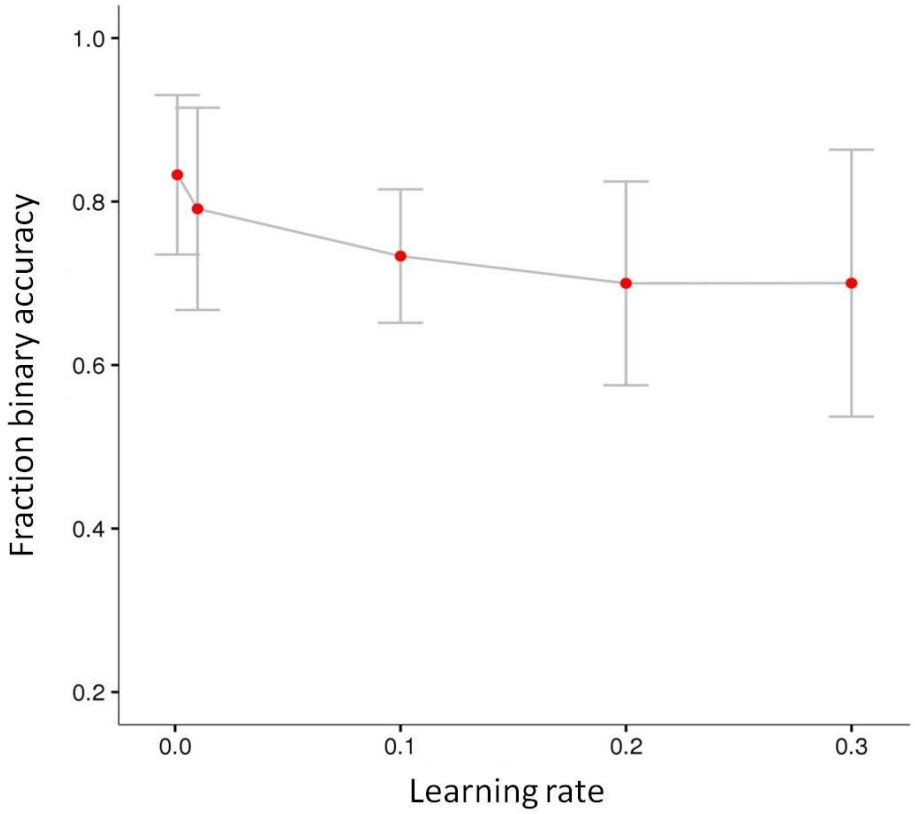
Supplementary Figure S8: The performance of different multiclass multilabel classification algorithms on the multilabel dataset with ECFP fingerprints, boruta selected descriptors, and boruta selected fingerprints. ACC – Multilabel accuracy, PPV – Multilabel precision or Multilabel positive predicted value, Hamloss – Hamming loss, F1 – Multilabel F1 score. Binary relevance – BR, Classifier chains – CC, Nested stacking – NS, Dependent binary relevance – DBR, Stacking – S. (Related to Figure 4)



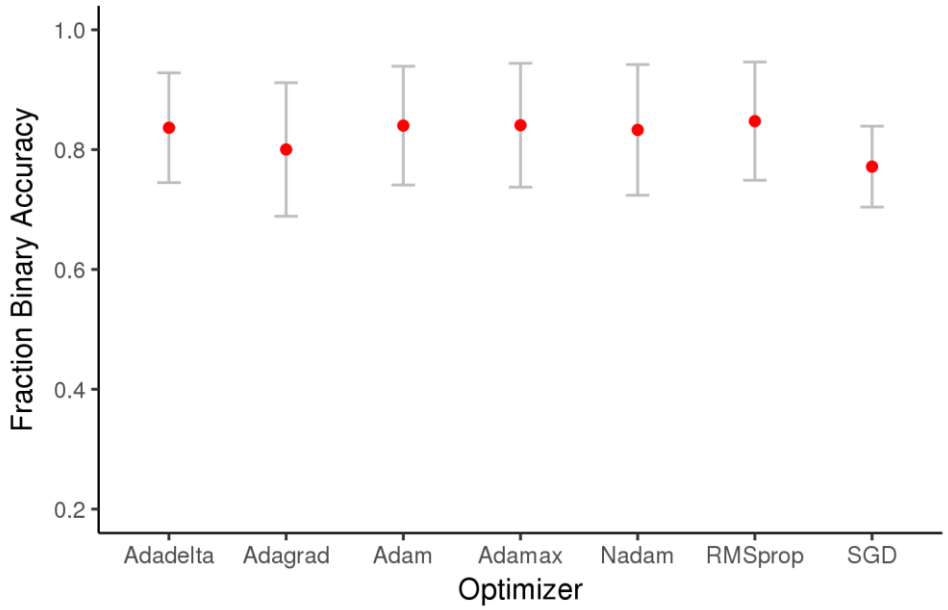
Supplementary Figure S9: The performance of different multiclass multilabel classification algorithms on the multilabel dataset with FCFP fingerprints, boruta selected descriptors, and boruta selected fingerprints. ACC – Multilabel accuracy, PPV – Multilabel precision or Multilabel positive predicted value, Hamloss – Hamming loss, F1 – Multilabel F1 score. Binary relevance – BR, Classifier chains – CC, Nested stacking – NS, Dependent binary relevance – DBR, Stacking – S. (Related to Figure 4)



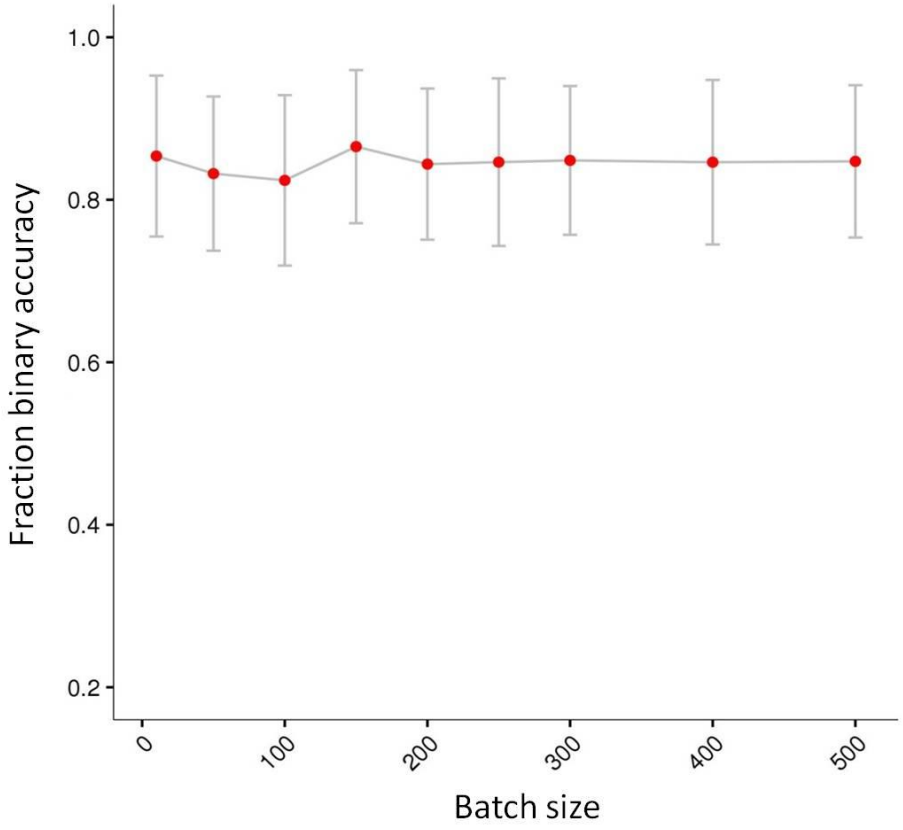
Supplementary Figure S10: Selecting the most optimum weight initializer for ANN models using grid search method using 5-fold cross validation. Different weight initializers are plotted against their respective fraction binary accuracies. The most optimum weight initializer was “lecun_uniform”. (Related to Figure 5)



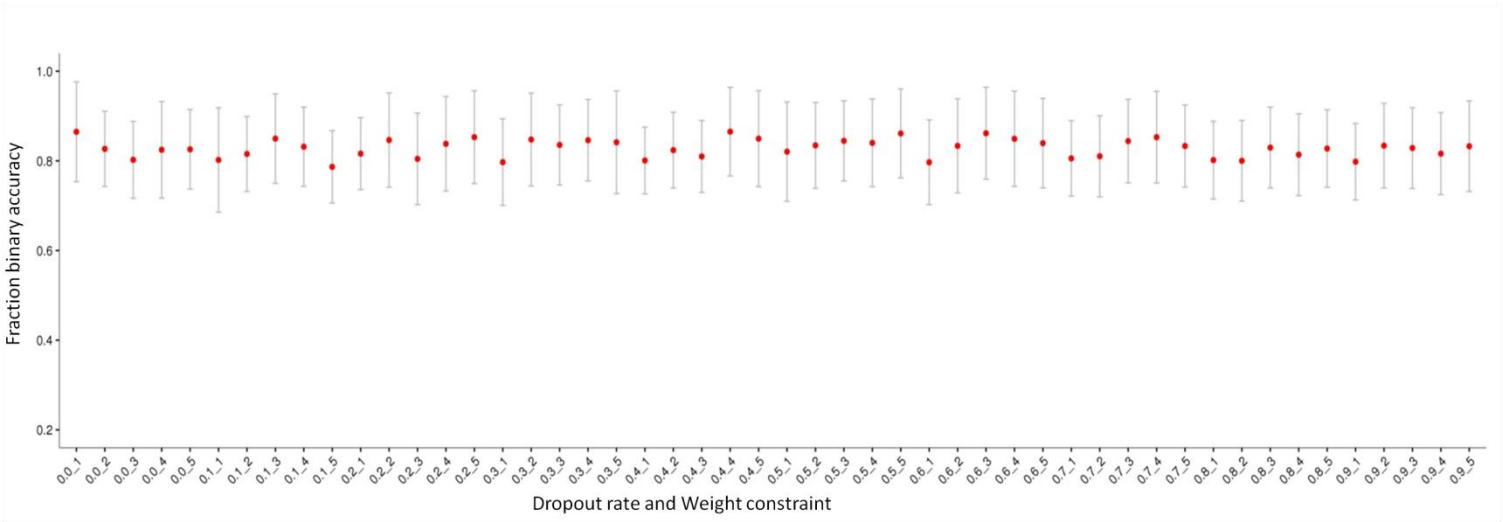
Supplementary Figure S11: Selecting the most optimum value of learning rate for ANN models using grid search method using 5-fold cross validation. Different values of learning rates are plotted against their respective fraction binary accuracies. The most optimum value for the learning rate was "0.001". (Related to Figure 5)



Supplementary Figure S12: Selecting the best performing optimizer for ANN models using grid search method using 5-fold cross validation. Different optimizers are plotted against their respective fraction binary accuracies. The best performing optimizer was “RMSprop”. (Related to Figure 5)



Supplementary Figure S13: Selecting the most optimum value of batch size for ANN models using grid search method using 5-fold cross validation. Different values of batch size are plotted against their respective fraction binary accuracies. The most optimum value for the batch size was “150”. (Related to Figure 5)



Supplementary Figure S14: Selecting the most optimum value of dropout rate and weight constraint for ANN models for performing dropout regularization using grid search method using 5-fold cross validation. Different combinations of values of dropout rate and weight constraint are plotted against their respective fraction binary accuracies. The most optimum value for the dropout rate was “0.4” and most optimum value of weight constraint was “4”. (Related to Figure 5)

SUPPLEMETARY TABLES

Table S1: The binary performance of multiclass multilabel model for predicting the reaction class on 5-fold cross validation testing. (Related to Figure 4)

Reaction Class	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1 Score	FDR	GPR
Oxidoreductases	0.855	0.207	0.111	0.345	0.793	0.568	0.806	0.786	0.834	0.214	0.836
Transferases	0.841	0.235	0.225	0.244	0.765	0.531	0.774	0.757	0.766	0.243	0.766
Hydrolases	0.824	0.210	0.538	0.089	0.790	0.422	0.821	0.656	0.543	0.344	0.551
Lyases	0.844	0.164	0.620	0.047	0.836	0.420	0.857	0.673	0.486	0.327	0.506
Isomerases	0.900	0.082	0.571	0.020	0.918	0.522	0.931	0.734	0.541	0.266	0.561
Ligases	0.886	0.082	0.761	0.017	0.918	0.336	0.931	0.577	0.338	0.423	0.371

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

Table S2: The binary performance of multiclass multi-label model for predicting the reaction class on blind set testing. (Related to Figure 4)

Reaction Class	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1 Score	FDR	GPR
Oxidoreductases	0.871	0.174	0.081	0.309	0.826	0.638	0.855	0.813	0.863	0.188	0.864
Transferases	0.884	0.222	0.217	0.226	0.778	0.557	0.783	0.774	0.778	0.226	0.778
Hydrolases	0.810	0.186	0.487	0.094	0.814	0.450	0.859	0.625	0.563	0.375	0.566
Lyases	0.896	0.120	0.536	0.036	0.880	0.516	0.899	0.722	0.565	0.278	0.579
Isomerases	0.943	0.060	0.417	0.032	0.940	0.551	0.968	0.583	0.583	0.417	0.583
Ligases	0.878	0.030	0.429	0.013	0.970	0.602	0.981	0.667	0.615	0.333	0.617

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

Table S3: The binary performance of multiclass multilabel model for predicting the subclasses of "Oxidoreductases" class on 5-fold cross validation testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Acting on the CH-OH group of donors	0.944	0.115	0.237	0.056	0.885	0.733	0.892	0.868	0.812	0.132	0.814
Acting on the aldehyde or oxo group of donors	0.921	0.079	0.611	0.013	0.921	0.520	0.929	0.791	0.521	0.209	0.555
Acting on the CH-CH group of donors	0.914	0.099	0.503	0.022	0.901	0.586	0.911	0.812	0.617	0.188	0.635
Acting on the CH-NH2 group of donors	0.945	0.046	0.686	0.008	0.954	0.449	0.961	0.698	0.433	0.302	0.468
Acting on the CH-NH group of donors	0.936	0.036	0.723	0.004	0.964	0.440	0.967	0.743	0.403	0.257	0.453
Acting on NADH or NADPH	0.899	0.007	0.714	0.000	0.993	0.533	0.993	1.000	0.444	0.000	0.535
Acting on other nitrogenous compounds as donors	0.952	0.017	0.800	0.000	0.983	0.420	0.983	0.900	0.327	0.100	0.424
Acting on a sulfur group of donors	0.920	0.026	0.742	0.002	0.974	0.437	0.977	0.773	0.386	0.227	0.446
Acting on a heme group of donors	0.744	0.001	1.000	0.000	0.999	0.000	0.999	0.000	0.000	NA	NA
Acting on diphenols and related substances as donors	0.881	0.014	1.000	0.000	0.986	0.000	0.986	0.000	0.000	NA	NA
Acting on a peroxide as acceptor	0.927	0.018	0.739	0.002	0.982	0.436	0.984	0.750	0.387	0.250	0.442
Acting on hydrogen as donor	0.877	0.004	0.692	0.000	0.996	0.554	0.996	1.000	0.471	0.000	0.555
Acting on single donors with incorporation of molecular oxygen (oxygenases)	0.890	0.071	0.659	0.008	0.929	0.503	0.933	0.824	0.483	0.176	0.530
Acting on paired donors, with incorporation or reduction of molecular oxygen	0.894	0.197	0.223	0.175	0.803	0.602	0.819	0.784	0.780	0.216	0.780
Oxidizing metal ions	0.949	0.002	0.357	0.000	0.998	0.801	0.998	1.000	0.783	0.000	0.802
Acting on CH or CH2 groups	0.930	0.026	0.653	0.002	0.974	0.530	0.976	0.839	0.491	0.161	0.539
Acting on iron-sulfur proteins as donors	0.997	0.003	0.700	0.000	0.997	0.547	0.997	1.000	0.462	0.000	0.548
Acting on reduced flavodoxin as donor	0.993	0.002	1.000	0.000	0.998	0.000	0.998	0.000	0.000	NA	NA
Acting on phosphorus or arsenic in donors	0.880	0.003	1.000	0.000	0.997	0.000	0.997	0.000	0.000	NA	NA
Catalysing the reaction X-H + Y-H = X-Y	0.916	0.017	0.814	0.001	0.983	0.381	0.983	0.800	0.302	0.200	0.386
Reducing C-O-C group as acceptor	0.813	0.004	1.000	0.000	0.996	0.000	0.996	0.000	0.000	NA	NA
Other oxidoreductases	0.798	0.007	0.737	0.000	0.993	0.466	0.993	0.833	0.400	0.167	0.468

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary

negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S4: The binary performance of the multiclass multi-label model for predicting the subclasses of "Transferases" class on 5-fold cross validation testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Transferring one-carbon groups	0.896	0.150	0.360	0.069	0.850	0.610	0.870	0.781	0.704	0.219	0.707
Transferring aldehyde or ketonic groups	0.939	0.017	1.000	0.001	0.983	-0.004	0.984	0.000	0.000	1.000	0.000
Acyltransferases	0.853	0.169	0.502	0.057	0.831	0.512	0.848	0.747	0.597	0.253	0.610
Glycosyltransferases	0.890	0.144	0.384	0.059	0.856	0.608	0.873	0.790	0.692	0.210	0.698
Transferring alkyl or aryl groups, other than methyl groups	0.855	0.090	0.754	0.012	0.910	0.383	0.918	0.708	0.365	0.292	0.417
Transferring nitrogenous groups	0.956	0.062	0.465	0.011	0.938	0.650	0.944	0.863	0.660	0.137	0.679
Transferring phosphorus-containing groups	0.943	0.106	0.278	0.053	0.894	0.697	0.917	0.809	0.763	0.191	0.764
Transferring sulfur-containing groups	0.891	0.073	0.745	0.014	0.927	0.366	0.937	0.617	0.361	0.383	0.397
Transferring selenium-containing groups	0.412	0.001	1.000	0.000	0.999	0.000	0.999	0.000	0.000	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S5: The binary performance of the multiclass multi-label model for predicting the subclasses of "Hydrolases" class on 5-fold cross validation testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Acting on ester bonds	0.878	0.204	0.237	0.177	0.796	0.586	0.814	0.774	0.768	0.226	0.768
Acting on sulfur-nitrogen bonds	0.336	0.002	1.000	0.000	0.998	0.000	0.998	0.000	0.000	NA	NA
Acting on carbon-phosphorus bonds	0.844	0.005	1.000	0.000	0.995	0.000	0.995	0.000	0.000	NA	NA
Acting on sulfur-sulfur bonds	0.426	0.001	1.000	0.000	0.999	0.000	0.999	0.000	0.000	NA	NA
Acting on carbon-sulfur bonds	0.766	0.009	1.000	0.000	0.991	0.000	0.991	0.000	0.000	NA	NA
Glycosylases	0.890	0.092	0.452	0.023	0.908	0.622	0.919	0.817	0.656	0.183	0.669
Acting on ether bonds	0.944	0.032	0.475	0.002	0.968	0.690	0.969	0.941	0.674	0.059	0.703
Acting on peptide bonds (peptidases)	0.906	0.032	1.000	0.000	0.968	0.000	0.968	1.000	0.000	NA	NA
Acting on carbon-nitrogen bonds, other than peptide bonds	0.903	0.156	0.283	0.094	0.844	0.638	0.870	0.784	0.749	0.216	0.750
Acting on acid anhydrides	0.952	0.058	0.383	0.023	0.942	0.646	0.960	0.744	0.674	0.256	0.677
Acting on carbon-carbon bonds	0.928	0.060	0.907	0.010	0.940	0.159	0.949	0.357	0.147	0.643	0.182
Acting on halide bonds	0.990	0.021	0.606	0.000	0.979	0.621	0.979	1.000	0.565	0.000	0.628
Acting on phosphorus-nitrogen bonds	0.355	0.003	1.000	0.000	0.997	0.000	0.997	0.000	0.000	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S6: The binary performance of the multiclass multi-label model for predicting the subclasses of "Lyases" class on 5-fold cross validation testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Carbon-carbon lyases	0.851	0.224	0.259	0.201	0.776	0.537	0.823	0.710	0.725	0.290	0.726
Carbon-oxygen lyases	0.839	0.223	0.231	0.214	0.777	0.554	0.742	0.810	0.789	0.190	0.789
Carbon-nitrogen lyases	0.781	0.109	0.961	0.011	0.891	0.076	0.899	0.300	0.070	0.700	0.109
Carbon-sulfur lyases	0.922	0.053	0.691	0.001	0.947	0.524	0.947	0.944	0.466	0.056	0.540
Carbon-halide lyases	0.834	0.011	0.583	0.001	0.989	0.585	0.990	0.833	0.556	0.167	0.589
Phosphorus-oxygen lyases	0.968	0.023	0.941	0.001	0.977	0.166	0.978	0.500	0.105	0.500	0.171
carbon-phosphorus lyases	0.760	0.004	1.000	0.000	0.996	0.000	0.996	0.000	0.000	NA	NA
Other lyases	0.953	0.026	0.833	0.006	0.974	0.256	0.979	0.429	0.240	0.571	0.267

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S7: The binary performance of the multiclass multi-label model for predicting the subclasses of "Isomerases" class on 5-fold cross validation testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Racemases and epimerases	0.946	0.108	0.133	0.095	0.892	0.765	0.926	0.832	0.849	0.168	0.849
cis-trans-Isomerases	0.822	0.027	0.529	0.005	0.973	0.601	0.977	0.800	0.593	0.200	0.614
Intramolecular oxidoreductases	0.924	0.167	0.356	0.077	0.833	0.605	0.843	0.802	0.714	0.198	0.719
Intramolecular transferases	0.827	0.165	0.646	0.049	0.835	0.389	0.859	0.636	0.455	0.364	0.475
Intramolecular lyases	0.932	0.059	0.292	0.015	0.941	0.767	0.946	0.902	0.793	0.098	0.799
Other isomerases	0.277	0.005	1.000	0.000	0.995	0.000	0.995	0.000	0.000	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S8: The binary performance of the multiclass multi-label model for predicting the subclasses of "Ligases" class on 5-fold cross validation testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Forming carbon-oxygen bonds	0.902	0.086	0.463	0.029	0.914	0.581	0.933	0.733	0.620	0.267	0.627
Forming carbon-sulfur bonds	0.939	0.134	0.167	0.112	0.866	0.722	0.888	0.833	0.833	0.167	0.833
Forming carbon-nitrogen bonds	0.933	0.124	0.099	0.151	0.876	0.752	0.890	0.864	0.882	0.136	0.882
Forming carbon-carbon bonds	0.892	0.067	0.864	0.007	0.933	0.264	0.939	0.600	0.222	0.400	0.286
Forming phosphoric-ester bonds	0.981	0.019	0.500	0.007	0.981	0.568	0.987	0.667	0.571	0.333	0.577
Forming nitrogen-D-metal bonds	0.969	0.019	1.000	0.000	0.981	0.000	0.981	0.000	0.000	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S9: The binary performance of the multiclass multi-label model for predicting the subclasses of "Oxidoreductases" class on stratified random sampling split testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Acting on the CH-OH group of donors	0.973	0.099	0.158	0.070	0.901	0.776	0.922	0.857	0.850	0.143	0.850
Acting on the aldehyde or oxo group of donors	0.975	0.047	0.462	0.013	0.953	0.624	0.963	0.778	0.636	0.222	0.647
Acting on the CH-CH group of donors	0.932	0.058	0.320	0.014	0.942	0.749	0.948	0.895	0.773	0.105	0.780
Acting on the CH-NH2 group of donors	0.964	0.012	0.500	0.000	0.988	0.703	0.988	1.000	0.667	0.000	0.707
Acting on the CH-NH group of donors	0.996	0.012	0.333	0.006	0.988	0.661	0.994	0.667	0.667	0.333	0.667
Acting on NADH or NADPH	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on other nitrogenous compounds as donors	1.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
Acting on a sulfur group of donors	1.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
Acting on a heme group of donors	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on diphenols and related substances as donors	0.994	0.006	1.000	0.000	0.994	0.000	0.994	0.000	0.000	NA	NA
Acting on a peroxide as acceptor	0.994	0.006	1.000	0.000	0.994	0.000	0.994	0.000	0.000	NA	NA
Acting on hydrogen as donor	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on single donors with incorporation of molecular oxygen (oxygenases)	0.854	0.052	0.583	0.013	0.948	0.521	0.958	0.714	0.526	0.286	0.546
Acting on paired donors, with incorporation or reduction of molecular oxygen	0.941	0.122	0.125	0.120	0.878	0.755	0.890	0.864	0.870	0.136	0.870
Oxidizing metal ions	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on CH or CH2 groups	0.994	0.006	0.500	0.000	0.994	0.705	0.994	1.000	0.667	0.000	0.707
Acting on iron-sulfur proteins as donors	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on reduced flavodoxin as donor	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on phosphorus or arsenic in donors	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Catalysing the reaction X-H + Y-H = X-Y	1.000	0.006	1.000	0.000	0.994	0.000	0.994	1.000	0.000	NA	NA
Reducing C-O-C group as acceptor	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Other oxidoreductases	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary

negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S10: The binary performance of the multiclass multi-label model for predicting the subclasses of "Transferases" class on stratified random sampling split testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Transferring one-carbon groups	0.924	0.133	0.302	0.070	0.867	0.654	0.892	0.789	0.741	0.211	0.742
Transferring aldehyde or ketonic groups	NA	0.006	NA	0.006	0.994	0.000	1.000	0.000	0.000	1.000	NA
Acyltransferases	0.862	0.165	0.541	0.050	0.835	0.492	0.852	0.739	0.567	0.261	0.583
Glycosyltransferases	0.961	0.089	0.220	0.043	0.911	0.764	0.926	0.865	0.821	0.135	0.822
Transferring alkyl or aryl groups, other than methyl groups	0.945	0.057	0.538	0.014	0.943	0.561	0.953	0.750	0.571	0.250	0.588
Transferring nitrogenous groups	0.968	0.051	0.500	0.007	0.949	0.639	0.953	0.875	0.636	0.125	0.661
Transferring phosphorus-containing groups	0.937	0.089	0.286	0.033	0.911	0.731	0.922	0.862	0.781	0.138	0.785
Transferring sulfur-containing groups	0.967	0.063	0.667	0.027	0.937	0.345	0.960	0.429	0.375	0.571	0.378
Transferring selenium-containing groups	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S11: The binary performance of the multiclass multi-label model for predicting the subclasses of "Hydrolases" class on stratified random sampling split testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Acting on ester bonds	0.895	0.202	0.216	0.191	0.798	0.591	0.826	0.763	0.773	0.237	0.773
Acting on sulfur-nitrogen bonds	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on carbon-phosphorus bonds	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on sulfur-sulfur bonds	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Acting on carbon-sulfur bonds	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Glycosylases	0.968	0.083	0.250	0.056	0.917	0.672	0.958	0.692	0.720	0.308	0.721
Acting on ether bonds	0.997	0.024	0.400	0.000	0.976	0.765	0.975	1.000	0.750	0.000	0.775
Acting on peptide bonds (peptidases)	0.964	0.012	1.000	0.000	0.988	0.000	0.988	0.000	0.000	NA	NA
Acting on carbon-nitrogen bonds, other than peptide bonds	0.931	0.119	0.200	0.085	0.881	0.715	0.915	0.800	0.800	0.200	0.800
Acting on acid anhydrides	0.963	0.048	0.143	0.039	0.952	0.731	0.987	0.667	0.750	0.333	0.756
Acting on carbon-carbon bonds	0.959	0.036	1.000	0.000	0.964	0.000	0.964	0.000	0.000	NA	NA
Acting on halide bonds	1.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
Acting on phosphorus-nitrogen bonds	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S12: The binary performance of the multiclass multi-label model for predicting the subclasses of "Lyases" class on stratified random sampling split testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Carbon-carbon lyases	0.895	0.169	0.192	0.154	0.831	0.650	0.868	0.778	0.792	0.222	0.793
Carbon-oxygen lyases	0.907	0.169	0.194	0.138	0.831	0.664	0.781	0.879	0.841	0.121	0.841
Carbon-nitrogen lyases	0.850	0.077	1.000	0.000	0.923	0.000	0.923	1.000	0.000	NA	NA
Carbon-sulfur lyases	0.968	0.031	0.333	0.016	0.969	0.651	0.984	0.667	0.667	0.333	0.667
Carbon-halide lyases	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Phosphorus-oxygen lyases	1.000	0.015	1.000	0.000	0.985	0.000	0.985	1.000	0.000	NA	NA
carbon-phosphorus lyases	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Other lyases	1.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S13: The binary performance of the multiclass multi-label model for predicting the subclasses of "Isomerases" class on stratified random sampling split testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Racemases and epimerases	0.973	0.083	0.077	0.087	0.917	0.824	0.955	0.857	0.889	0.143	0.889
cis-trans-Isomerases	1.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
Intramolecular oxidoreductases	0.931	0.083	0.273	0.000	0.917	0.806	0.893	1.000	0.842	0.000	0.853
Intramolecular transferases	0.842	0.139	0.714	0.000	0.861	0.494	0.853	1.000	0.444	0.000	0.535
Intramolecular lyases	0.961	0.056	0.200	0.032	0.944	0.768	0.968	0.800	0.800	0.200	0.800
Other isomerases	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

Table S14: The binary performance of the multiclass multi-label model for predicting the subclasses of "Ligases" class on stratified random sampling split testing. (Related to Table 1 and 2)

Reaction subclass	AUC	MMCE	FNR	FPR	ACC	MCC	NPV	PPV	F1	FDR	GPR
Forming carbon-oxygen bonds	0.957	0.040	0.500	0.000	0.960	0.692	0.958	1.000	0.667	0.000	0.707
Forming carbon-sulfur bonds	0.887	0.200	0.200	0.200	0.800	0.592	0.857	0.727	0.762	0.273	0.763
Forming carbon-nitrogen bonds	0.910	0.240	0.231	0.250	0.760	0.519	0.750	0.769	0.769	0.231	0.769
Forming carbon-carbon bonds	0.958	0.080	1.000	0.042	0.920	-0.042	0.958	0.000	0.000	1.000	0.000
Forming phosphoric-ester bonds	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA
Forming nitrogen-D-metal bonds	NA	0.000	NA	0.000	1.000	0.000	1.000	0.000	NA	NA	NA

AUC = Area under the curve, MMCE = Binary mean misclassification error, FNR = Binary false negative rate, FPR = Binary false positive rate, ACC = Binary accuracy, MCC = Matthews correlation coefficient, NPV = Binary negative predictive value, PPV = Binary positive predicted value, F1 = Binary F1 score, FDR = Binary false discovery rate, GPR = Geometric mean of binary precision and binary recall

NA - Although, stratified random sampling was used to split the six datasets into the training and testing datasets, yet, some of the reaction subclasses had no representation in the respective test datasets, thus few of the binary matrices could not be calculated for these reaction subclasses. These are represented as NA values

TRANSPARENT METHODS

Construction of microbial species database

The data and text mining of the available literature was performed to construct a manually curated database of bacterial species present at different skin sites. The protein sequences of genomes of the bacterial species from the constructed database were retrieved from NCBI Reference Sequence Database (RefSeq) (O'Leary et al., 2015). Only complete genome assemblies were used from the RefSeq database. The genomes of all the available strains of a species were used to construct the pangenome of that species which helped to compile the metabolic potential of all the strains for a particular species. The pangenome of a species includes all the genes from all the different strains of that species. The pangenomes were constructed for all the bacterial species that are experimentally known to be a part of skin microbiome, and for which the complete genomes were available on the NCBI RefSeq database. The information on different bacterial species for which the pangenomes were constructed, and the skin sites harbouring these species is provided in **Supplementary Data Sheet 1**. The information on 19 different sites primarily including the sebaceous, moist, and dry skin niches was retrieved from literature, manually curated, and was used for further analysis. For the construction of pangenome, the protein sequences of all genomes of a species were merged and clustered at 100% identity using CD-HIT v4.6 (Li and Godzik, 2006).

Construction of skin microbiome specific metabolic information database

The ExPASy enzyme database was used to find the Uniprot/SwissProt IDs of all the annotated enzymes that belong a particular metabolic reaction annotated as four-digit EC number (Gasteiger et al., 2003). The protein sequences for these enzymes were downloaded from the Uniprot database (Consortium, 2014). The homology search of these enzyme sequences was performed against each pangenome to identify all the metabolic enzymes present in that pangenome using the NCBI BLASTP program (Altschul et al., 1990). The hits were filtered using the cut-off criteria of identity >50%, bit-score >100, query coverage >50%, subject coverage >50%, E-value $<10^{-10}$, mismatch percentage <50%, and gap percentage <50%. Finally, a database of complete reactions annotated as four-digit EC number and corresponding metabolic enzymes from all the pangenomes was constructed. Each of the metabolic enzymes was tagged with the bacterial species pangenome containing the enzyme. Further, the metabolic enzymes were also tagged with the skin sites that harbour the bacterial species with those enzymes.

Construction of reaction, RDM pattern, and substrate database

All the enzymatic reactions and their corresponding reactions IDs were retrieved from KEGG database (Kanehisa and Goto, 2000). For each reaction ID, the corresponding reactions pairs and respective RDM patterns were also retrieved from the KEGG database (Kanehisa and

Goto, 2000). From this data the databases of reactions, reaction pairs, and RDM patterns were constructed. From the reactions, the primary substrates were identified and a database of primary substrates and their respective reactions annotated as four-digit EC number was constructed.

Calculation of molecular features of substrates

The structural and chemical features were calculated for each of the substrate molecule in the substrate database. Thus, the molecular information of substrates was translated into machine-readable features that include chemical properties parameters, linear structural fingerprints, and circular molecular connectivity information. The chemical features were calculated using the PaDEL software (Yap, 2011). These chemical features included different types of chemical descriptions such as acidic atom count, aromatic atom count, aromatic bonds count, carbon types, molecular distance edge etc. encoded into 240 different values. Two types of structural fingerprints were calculated: linear and circular. The linear fingerprints were calculated using the PaDEL software (Yap, 2011). A total of 12 different types of linear fingerprints (Fingerprinter, Pubchem, MACCS, Atom pairs 2D, KlekotaRoth etc.) were calculated that were represented as 10,208 bits (values either 0 or 1). The two types of circular/topological fingerprints, Morgan FCFP - 512 bits and Morgan ECFP - 512 bits, were calculated using RDkit software (Landrum, 2016).

Feature selection

The Boruta algorithm implemented in R as the “Boruta” package was used to extract the important features among all the above calculated molecular features (Kursa and Rudnicki, 2010). Boruta is a wrapper algorithm for feature selection that uses “Random Forest” algorithm, and scores each feature and marks them as important, unimportant or tentative. The tentative features were then finalized as important or unimportant using “TentativeRoughFix” function of Boruta package in R. The variable importance was calculated for each EC reaction (EC1 to EC6) class separately. Finally, the important features for each EC were merged and unique sorted to obtain the final set of important features.

Principal component and cluster analysis

Principal component analysis was performed using the “prcomp” function from “stats” package in R v3.4.4. This function performs the principal component analysis (PCA) by performing the singular value decomposition of the input data (Mankin, 2003). This method is the preferred method for better numeric accuracy. The PCA and scree plots were generated using the “factoextra” and “ggfortify” package in R v3.4.4 (Kassambara and Mundt, 2017). The density-based clustering was performed using the “fpc” and “dbscan” package in R v3.4.4. The kNN distance plot was generated using the “kNNdistplot” function

from “dbscan” package in R v3.4.4 (Tran et al., 2013). The density cluster plot was generated using the “factoextra” package in R v3.4.4 (Kassambara and Mundt, 2017).

Hierarchical clustering

The hierarchical clustering was performed using the ‘hclust’ function of ‘stats’ package in R v3.4.4. The approximate unbiased p-values (AUp) and the bootstrap probability (BP) values for each branch/cluster were calculated using multiscale bootstrap resampling and using normal bootstrap resampling, respectively. The optimum number of clusters was identified to be two based on the average silhouette method .

Construction of machine learning models

Dataset construction

The dataset of 3,769 substrate molecules was randomly split into a working and blind dataset with a ratio of 95:5. Thus, the working dataset had 3,602 molecules and the blind dataset had 167 molecules. The working dataset was utilized for the training and statistical evaluation of the machine learning model, and the blind dataset was used for the independent evaluation of the model. The dataset was highly skewed with a higher number of substrate molecules for “Oxidoreductases” and “Transferases” in comparison to other reaction classes. Also the abundances of substrate molecules belonging to different combinations of reaction classes were also highly variable. Thus, a modified strategy of stratified random sampling approach was used to divide the working dataset into the training and testing dataset for modeling. The details of the dataset construction are mentioned in **Supplementary Text S1**.

Training and evaluation

The prediction of reaction class is a multiclass multilabel problem because one substrate molecule can undergo more than one type of reaction among the six types of reactions classes. In machine learning, there exists two methods to model the multiclass multi-label problem, one is problem transformation method where the multiclass multi-label problem is divided into several multiclass or binary problems, and another is algorithm adaptation method where the algorithms are adapted to perform the multiclass multi-label predictions. For the problem transformation method all the algorithms used for binary or multiclass classification can be used, whereas for algorithm adaptation method the algorithms need to be changed before using them for the multiclass multi-label classification. In the problem transformation method, a learner known as “wrapped multilabel learner” is employed on the “core learner”. The function of wrapped learner is to manage and combine several core learners so that they can work in synchronization to achieve the multilabel classification. The core learner is any traditional algorithm for binary or multiclass classification. We used five different wrapper methods: (1) binary relevance (BR) method, (2) classifier chains (CC)

method, (3) Nested stacking (NS) method, (4) Dependent binary relevance (DBR) method, and (5) Stacking method. We used the seven core learners for each of the above mentioned wrapper methods, these are: k-Nearest Neighbors (kNN), Recursive Partitioning (RPART), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost),Perceptive Neural Network (NNET), Naive Bayes, Random Forests (RF). In the algorithm adaptation method, we used two methods, randomForestSRC (RFSRC) and random ferns (RFerns).

The performance of the models was evaluated using two types of matrices, multilabel - to assess the capability of the model to perform the multilabel classification, and binary - to assess the capability of the model to perform the binary classification for each label. Five matrices were used in the multilabel case namely: Multilabel Accuracy, Multilabel Sensitivity or Recall or True Positive Rate, Multilabel Precision or Positive Prediction Value (PPV), Multilabel F1 measure (F1), and Hamming loss (Charte and Charte, 2015). The formulas for these matrices for multiclass multilabel classification are mentioned below (Charte and Charte, 2015):

$$Multilabel\ Accuracy = \frac{1}{|D|}\sum_{i=1}^{i=|D|}\frac{|Pi\ \cap\ Ti|}{|Pi\ \cup\ Ti|};\ Multilabel\ Precision\ = \frac{1}{|D|}\sum_{i=1}^{i=|D|}\frac{|Pi\ \cap\ Ti|}{|Ti|}$$

$$Multilabel\ Recall\ or\ Sensitivity\ = \frac{1}{|D|}\sum_{i=1}^{i=|D|}\frac{|Pi\ \cap\ Ti|}{|Pi|}$$

$$Multlabel\ F1\ measure = \ 2 * \frac{Precision * Recall}{Precision + Recall};\ Hamming\ loss = \frac{1}{|D|}\sum_{i=1}^{i=|D|}\frac{|Pi\ \Delta\ Ti|}{|C|}$$

Where, D is the total number of instances in the multiclass multilabel dataset, C is the complete set of labels present in the multiclass multilabel dataset, *Pi* is the predicted labels for the i^{th} instance, and Ti is the true labels for the i^{th} instance. The set operations used were: ∩ meaning intersection, ∪ meaning union, and *Δ* meaning symmetric difference.

We used eight matrices to evaluate the binary performance: Binary Accuracy, Mean Misclassification Error (MMCE), Matthews Correlation Coefficient (MCC), Binary Precision or Positive Predicted Value (PPV), Area under the curve (AUC), Binary False Negative Rate (FNR), Binary False Positive Rate (FPR), Binary Sensitivity or Recall or True Positive Rate (TPR), Binary Specificity or True Negative Rate (TNR), Binary Negative predictive value (NPV), Binary False discovery rate (FDR), and Binary Geometric Mean of binary precision and binary recall (GPR). The formulas for these binary performance matrices are mentioned in **Supplementary Text S2**. The final model for reaction class and subclass prediction was constructed with the method that showed the best multilabel and binary performance.

Construction of artificial neural network (ANN) models

Dataset construction

The aim of constructing the ANN model was to improve upon the learning about the class-specific patterns, thus, only the substrates where the molecule could undergo the reactions of only one type of reaction class were extracted from the working dataset (as mentioned above) and were used for the construction of ANN models. This dataset had a total of 1,758 substrate molecules with the distribution of molecules across different reactions classes: “Oxidoreductases” - 832, “Transferases” - 573, “Hydrolases” - 195, “Lyases” - 79, “Isomerases” – 41, and “Ligases”- 36. It is evident from the distribution that the dataset is very biased and imbalanced, thus, the stratified random sampling was performed to split this dataset into training and testing dataset. For stratified random sampling, this dataset was first divided into six parts, one for each reaction class, and then each reaction class dataset was splitted separately into training and testing dataset using random sampling with the split ration of approximately 90:10. Now all the six training sets were merged to create the final training dataset and all the six testing datasets were merged to create the final testing dataset.

Training and evaluation

The ANN network was constructed in Python using libraries tensorflow v1.4.1 and keras v2.2.4. Based on the nature of the problem, the best suited multilayer perceptron model that is based on the backpropagation method for training is used. In the backpropagation method, the error rate is provided as feedback to the whole neural network that is known as back propagating the error, which is then used by an optimizer algorithm to optimize the parameters of artificial neural network.

Three different matrices were used to evaluate the performance of the ANN model: categorical accuracy, binary accuracy, and log loss/binary cross entropy. Since it is a multiclass classification problem the target variable here is one hot encoded. The categorical accuracy checks if the maxima in the true values and the maxima in the predicted values have the same index, if yes, it is considered a true prediction, else it is considered a wrong prediction. This is performed on all test dataset instances, and the fraction of correct predictions out of total predictions on test dataset gives the categorical accuracy. In contrast, for calculating the binary accuracy, at first all the probabilities are converted into values with the threshold of 0.5 (if <0.5 means 0, and if >0.5 means 1), then all the true values of each instance are compared with the predicted values. If the true value is equal to the predicted value then it is considered as correct prediction, else it is considered a wrong prediction. This was also performed on all the values of each of the test dataset instance, and the fraction of correct predictions out of the total predictions gives the binary accuracy. The formula to calculate the log loss/binary cross entropy is:

$$-\sum_{c=1}^{N} Y(i,c)\log\left(P(i,c)\right)$$

Where, N is the number of different classes present in the dataset, log is the natural logarithm, Y(i,c) is the indicator if the classification is correct (1 if yes and 0 if no) for i^{th} observation for c class, and P(i,c) is the probability predicted by the ANN model for i^{th} observation for c class

The hypermeters of the ANN models were also optimized based on the three evaluation matrices mentioned above to obtain the best performance from the ANN model. To calculate the optimum number of neurons in the hidden layer, the values close to the average of the size of input and output layers were tried and the best value was selected while keeping the number of hidden layer as one. Different number of hidden layers were tried to select the best performing ANN model with the most optimum number of hidden layers. A range of epoch values from 1 to 4000 were tried and based on the plateau in the performance an optimum value was selected. The other parameters of the ANN models were optimized using the grid search method with 5-fold cross validation the details are mentioned in **Supplementary Text S3**. The parameters optimized were: Weight initializer, Learning rate, Optimizer, Batch size, Dropout rate, and Weight constraint.

Statistical evaluation of the machine learning and ANN models

We used three methods to statistically evaluate the performance of the machine learning and ANN models. These three methods are split testing, cross validation, and blind set testing. The details of these methods are mentioned in **Supplementary Text S4**.

Molecular similarity search

The open source chemoinformatics tool Open Babel v2.3.2 was used for performing the molecular similarity search using the inbuilt default fingerprint FP2 which is a path-based fingerprint. The complete substrate molecule database was divided into several reaction subclass specific databases, depending on the type of reaction subclass a particular substrate can undergo. Once the reaction class and subclass are predicted by the machine learning and ANN models, the molecular similarity search against the predicted reactions subclass specific database is performed and Tanimoto Coefficient or Jaccard Index was calculated. The formula for calculating the Tanimoto Coefficient or Jaccard Index is:

$$Tanimotto\ coefficient\ or\ jaccard\ index\ T(a,b) = \frac{Nc}{Na + Nb - Nc}$$

Where, T(a,b) is the tanimoto coefficient for molecule a and b, Na is number of bits that are 1 in the fingerprints of molecule a, Nb is number of bits that are 1 in the fingerprints of

molecule b, and Nc is the number of bits that are 1 in the intersection of fingerprints of molecule a and b.

K-nearest neighbour (KNN) model construction or lazy learning

KNN is a preferred method for the identification of structurally and chemically similar molecules to the input molecule in the search against a heterogeneous database (Soucy and Mineau, 2001). The KNN algorithm was implemented using the R package "FNN" (Beygelzimer et al., 2015). The k-nearest neighbours for any given molecule were extracted using the function "get.knnx" from the "FNN" package that uses "Euclidean distance" as the measure of similarity between molecules.

Identification of reaction center

The reaction centers were identified by using the RDM pattern information that is associated with each of the substrate-product pair of an enzyme catalyzed reaction in KEGG database (Kanehisa, 2002). In the RDM pattern database constructed in this study, all the complete metabolic reactions are associated with the respective Reaction Class (RC) pairs, and all the RC pairs were tagged with corresponding RDM patterns. For a given biochemical reaction available in KEGG, the KEGG-defined RDM (Reaction center, Difference region, Matching region) patterns contain the information on the KEGG atom type changes at the reaction center, matched region of the molecule, and the difference region of the molecule (Kotera et al., 2013). Here a reaction center is the atom where the reaction occurs, a matched region is the region common between substrate and product that remained unchanged after the reaction, and a difference region is the part of molecule that changed after the reaction. The RDM patterns are derived from the structural alignments of the substrates and products which identifies the reaction center, matched and difference regions (Yamanishi et al., 2009). To identify the reaction center in a molecule for each of the predicted metabolic reaction, all the RC pairs and corresponding RDM patterns were extracted. Using these RDM patterns, the reaction centers were identified by in-house python scripts. Thus, this computational approach is similar to the biochemical approach in which the primary substrate and product are compared to identify the reaction center where the biochemical reaction has occurred in the enzyme active site.

SUPPLEMENTARY TEXT

Supplementary Text S1: (Related to Figure 4 and 5)

The dataset of 3,769 substrate molecules was randomly split into a working and blind dataset with a ratio of 95:5, the working dataset had 3602 molecules and the blind dataset had 167 molecules. The working dataset was utilized further for the training and statistical evaluation of the machine learning model and the blind dataset was used for the independent evaluation of the model. Since the dataset was much skewed with a very higher number of substrate molecules for “Oxidoreductases” and “Transferases” in comparison to other reaction classes and also abundance of substrate molecules with different combinations of reaction classes was very variable, thus, a modified strategy of stratified random sampling approach was used to divide the working dataset into the training and testing dataset for modeling.

In this approach, to account for the differences in substrates belonging to different combinations of reaction classes the working dataset was divided into pure (contains substrates that can undergo only one type of reaction among different reaction classes) and mixed datasets (contains substrates that can undergo multiple reactions among different reaction classes). The pure dataset which had 1756 substrate was statistically down-sampled to randomly select the same number (lowest in the sample = 36) of substrates for each reaction class. Thus, the down-sampled pure had a total of 216 substrates (36 of each reaction class). The mixed dataset had 1,846 substrate molecules which was split into two datasets large and small with the ratio of 95:5, the large part had 1,774 substrates, whereas the small part had 72 substrate molecules. The training dataset was constructed by merging the down-sampled pure dataset (216 substrates) and the large part of mixed dataset (1,774 substrates), and had a total of 1,990 substrate molecules. The testing dataset was constructed by merging the remaining of pure dataset after down-sampling (1,540 substrates) and the small part of the mixed dataset (72 substrates), and has a total of 1,612 substrate molecules. These final training and testing datasets corresponded to an approximate ratio of 55:45 of the working dataset of 3602 substrate molecules.

Similarly, for the training of machine learning models for reaction subclass prediction the working dataset was divided into six parts, one for each reaction class. The same substrate could belong to multiple parts if it can undergo reactions from multiple reaction classes. The numbers of reaction sub-classes in each reaction class were: “Oxidoreductases” - 22, “Transferases” - 9, “Hydrolases” -13, “Lyases” -8, “Isomerases” – 6, and “Ligases”- 6. For each dataset the stratified random sampling was performed to split the input dataset into training and testing dataset with the split ration of 90:10.

Supplementary Text S2: (Related to Figure 4)

To evaluate the binary performance we used eight matrices, Binary Accuracy, Mean Misclassification Error (MMCE), Matthews Correlation Coefficient (MCC), Binary Precision or Positive Predicted Value (PPV), Area under the curve (AUC), Binary False Negative Rate (FNR), Binary False Positive Rate (FPR), Binary Sensitivity or Recall or True Positive Rate (TPR), Binary Specificity or True Negative Rate (TNR), Binary Negative predictive value (NPV), Binary False discovery rate (FDR), and Binary Geometric Mean of binary precision and binary recall (GPR). The formulas for these binary performance matrices are mentioned below:

$$Binary\ Accuracy = \frac{TP + TN}{P + N}$$

$$MMCE = \frac{FP + FN}{P + N};\ MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$Binary\ Precision = \frac{TP}{FP + TP};\ Binary\ FNR = \frac{FN}{TP + FN};\ Binary\ FPR = \frac{FP}{TN + FP}$$

$$Binary\ Sensitivity\ or\ Recall = \frac{TP}{FN + TP};\ Binary\ Specificity\ = \frac{TN}{FP + TN}$$

$$Binary\ NPV = \frac{FP}{TN + FP};\ Binary\ FDR = \frac{FP}{TP + FP};\ Binary\ GPR = \sqrt{Precision * Recall}$$

Where, TP is true positives, FP is false positives, TN is true negatives, FN is false negatives, P is the total number of positives, and N is the total number of negatives in the input dataset.

Supplementary Text S3: (Related to Figure 5)

The other parameters of the ANN models were optimized using the grid search method with 5-fold cross validation, the parameters along with the values tested are: (1) Weight initializer – Zero, Normal, Uniform, Glorot normal, Glorot uniform, He normal, He uniform, and Lecun uniform (2) Learning rate – 0.0, 0.02, 0.1, 0.2, and 0.3 (3) Optimizer – Adadelta, Adagrad, Adam, Adamax, Nadam, RMSprop, and SGD (4) Batch size – 0, 50, 100, 150, 200, 250, 300, 400, and 500 (5) Dropout rate and Weight constraint – [0.0, 1], [0.0, 2], [0.0, 3], [0.0, 4], [0.0, 5], [0.1, 1], [0.1, 2], [0.1, 3], [0.1, 4], [0.1, 5], [0.2, 1], [0.2, 2], [0.2, 3], [0.2, 4], [0.2, 5], [0.3, 1], [0.3, 2], [0.3, 3], [0.3, 4], [0.3, 5], [0.4, 1], [0.4, 2], [0.4, 3], [0.4, 4], [0.4, 5], [0.5, 1], [0.5, 2], [0.5, 3], [0.5, 4], [0.5, 5], [0.6, 1], [0.6, 2], [0.6, 3], [0.6, 4], [0.6, 5], [0.7, 1], [0.7, 2], [0.7, 3], [0.7, 4], [0.7, 5], [0.8, 1], [0.8, 2], [0.8, 3], [0.8, 4], [0.8, 5], [0.9, 1], [0.9, 2], [0.9, 3], [0.9, 4], and [0.9, 5]. The final model was constructed using the most optimum parameters selected based on the gird search method.

Supplementary Text S4: (Related to Figure 4 and 5)

a) *Split testing*: As mentioned in the dataset construction part the complete working dataset was divided into training and testing dataset using a specific splitting approach. The models were trained on the training dataset and evaluated in the test dataset.
b) *Cross validation*: In this study, we used 5-fold cross validation for machine learning models and ANN models. In this method, during the process of training the dataset was randomly divided into five equal parts and five iterations of training and testing are performed. In each of the iteration four parts are used for training and the rest one part is used for the testing. This way in five iterations each of the training instances is used for testing the model and thus, avoiding any bias in the evaluation of the performance matrices. Finally, the mean/median and standard deviation value of performance matrices across five iterations is used to evaluate any bias in the model such as over-fitting or under-fitting.
c) *Blind set testing*: Approximately 5% of the randomly selected instances are kept aside before starting the training and testing process of model and, the model never sees these instances at any stage of its training and testing, hence called a blind dataset to model. Therefore, the performance of the model on this blind dataset is considered to be a real or unbiased performance of the model.

Supplementary Text S5: (Related to Figure 1)

To further evaluate the variability in skin sites in terms of enzymatic reactions it is critical to know the reactions that are common to the different sites. To identify the number of reactions that are common to different sites the matrix layout analysis performed using the 'UpSetR' package in R (Conway et al., 2017; Lex and Gehlenborg, 2014). It generates a matrix layout diagram for visualizing the set intersections.

Supplementary Text S6: (Related to Figure 1)

A skin microbiome specific metabolic enzyme database of four-digit EC number and corresponding metabolic enzymes from all the pangenomes was constructed. Each metabolic enzyme in this database is tagged with the bacterial species if their pangenome harbors this enzyme. Also the metabolic enzymes were tagged with the skin sites based on the presence and absence of the bacterial species harboring the enzyme on that particular skin site. All the well-annotated enzymatic reactions were extracted from KEGG database

and corresponding reaction, primary substrates, RC pair and RDM pattern databases were constructed. A total of 10,629 reactions, 3,769 primary substrates, and 2,592 RC pairs and RDM patterns were extracted from the KEGG database. Also each of the primary substrate in the database was tagged with the reaction class (EC-one digit), reaction subclass (EC-two digit), and complete reaction category (EC-four digit). The four types of molecular features were calculated for each of the primary substrate: chemical descriptors, linear fingerprints, Morgan ECFP fingerprints, and Morgan FCFP fingerprints. All this data was used for the training of the prediction models and for making the final metabolism predictions.

Supplementary Text S7: (Related to Figure 4 and 5)

Feature selection was performed using the Boruta package on chemical descriptor and linear fingerprints (Kursa and Rudnicki, 2010). Boruta selected 194 features out of 240 chemical descriptors and 1,104 features out of 10,208 linear fingerprints across different reaction classes. All features of Morgan FCFP and Morgan ECFP were included and were not subjected to feature selection as these features are elementary and all the bits are needed to adequately describe a substrate molecule. The total number of features used for further analysis was 2,322: 194 from chemical descriptors, 1104 linear fingerprints, 512 Morgan ECFP fingerprints, and 512 Morgan FCFP fingerprints.

Supplementary Text S8: (Related to Figure 6)

A previous study have reported that the adjustment in the threshold of multiclass multilabel classification model could significantly improve on their performance (Al-Otaibi et al., 2014; Fan and Lin, 2007). Thus, although we evaluated the performance of our models using the threshold value of 0.5 for all the machine learning and ANN models so that we do not overestimate the sensitivity of our models, the prediction threshold of the models deployed in the web server was lowered for the reaction subclass prediction models of “Oxidoreductases”, “Transferases”, “Hydrolases”, and “Lyases” classes from 0.5 to 0.2 because they had a range of 8 to 22 different subclasses and a high threshold could lead to miss out on some possible reaction subclasses.

REFERENCES

Al-Otaibi, R., Flach, P., and Kull, M. (2014). Multi-label classification: A comparative study on threshold selection methods. Paper presented at: First International Workshop on Learning over Multiple Contexts (LMCE) at ECML-PKDD 2014.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.
Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., Li, S., and Li, M.S. (2015). Package 'FNN'. Accessed June *1*.
Charte, F., and Charte, D. (2015). Working with Multilabel Datasets in R: The mldr Package. R Journal *7*.
Consortium, U. (2014). UniProt: a hub for protein information. Nucleic acids research *43*, D204-D212.
Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938-2940.
Fan, R.-E., and Lin, C.-J. (2007). A study on threshold selection for multi-label classification. Department of Computer Science, National Taiwan University, 1-23.
Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research *31*, 3784-3788.
Kanehisa, M. (2002). The KEGG database. Paper presented at: Novartis Foundation Symposium (Wiley Online Library).
Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research *28*, 27-30.
Kassambara, A., and Mundt, F. (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses *76*.
Kotera, M., Tabei, Y., Yamanishi, Y., Moriya, Y., Tokimatsu, T., Kanehisa, M., and Goto, S. (2013). KCF-S: KEGG Chemical Function and Substructure for improved interpretability and prediction in chemical bioinformatics. BMC systems biology *7*, S2.
Kursa, M.B., and Rudnicki, W.R. (2010). Feature selection with the Boruta package. J Stat Softw *36*, 1-13.
Landrum, G. (2016). RDKit: open-source cheminformatics software.
Lex, A., and Gehlenborg, N. (2014). Points of view: Sets and intersections. Nature Methods *11*, 779-779.
Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658-1659.
Mankin, E. (2003). Principal Components Analysis: A How-To Manual for R. Desde http://psych colorado edu/wiki/lib/exe/fetch php.
O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., and Ako-Adjei, D. (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research *44*, D733-D745.
Soucy, P., and Mineau, G.W. (2001). A simple KNN algorithm for text categorization. Paper presented at: Proceedings 2001 IEEE International Conference on Data Mining (IEEE).
Tran, T.N., Drab, K., and Daszykowski, M. (2013). Revised DBSCAN algorithm to cluster data with dense adjacent clusters. Chemometrics and Intelligent Laboratory Systems *120*, 92-96.
Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M. (2009). E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. Bioinformatics *25*, i179-i186.
Yap, C.W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of computational chemistry *32*, 1466-1474.