IRASA International Scientific Conference
SCIENCE, EDUCATION, TECHNOLOGY AND INNOVATION
SETI VI 2024
Book of Proceedings

DOI: `https://doi.org/10.62982/seti06.boje.13

# APPLICATION AND EFFECT OF ARTIFICIAL INTELLIGENCE FOR THE FUNCTIONING OF METABOLIC BIOENGINEERING IN SYSTEMS BIOLOGY

*Jelena Bošković[46], Jelena Mladenović[47], Vera Popović[48], Aleksandar Stevanović[49], Ljubica Šarčević-Todosijević[50]*

## Abstract

Emerging synthetic biology capabilities have the potential to dramatically improve our ability to engineer biological systems. However, a fundamental obstacle to realizing this potential is the inability to accurately predict biological behavior following modification of the corresponding genotype. This method systematically leverages arbitrary amounts of new data to improve predictions, and does not assume any particular interactions, but rather implicitly chooses the most predictive ones. Metabolic bioengineering aims to maximize the production of bio-economically important substances (compounds, enzymes, or other proteins) through the optimization of the genetics, cellular processes and growth conditions of microorganisms. To achieve this goal, a large system of experimental techniques, compound libraries, computational methods and data resources, including multi-omics data, are used. The recent advent of multi-omics systems biology approaches significantly impacted the field by opening new avenues to perform dynamic and large-scale analyses that deepen our knowledge on the manipulations. However, given the great potential of transcriptomics, proteomics, and metabolomics, integrating the data to gain a more comprehensive understanding is not a trivial task.

**Key words:** *Artificial intelligence, systems biology, machine learning, metabolic bioengineering*.

[46] Full Professor Jelena Bošković, PhD, Academician of IRASA, Metropolitan University, Belgrade, Republic of Serbia, jelena.boskovic@metroplitan.ac.rs
[47] Jelena Mladenović, Bsci in ecology, University Business Academy in Novi Sad, Faculty of Economics and Engineering Management, Novi Sad, Republic of Serbia, jelenamladenovic3110@gmail.com
[48] Vera Popović, PhD, Principal Research Fellow, Full Professor, Academician of IRASA, Institute of Field and Vegetable Crops, National Institute for the Republic of Serbia, Novi Sad, Republic of Serbia, vera.popovic@ifvcns.ns.ac.rs vera.popovic@nsseme.com
[49] Lecturer Aleksandar Stevanović, PhD, Academician of IRASA, Academy of Applied Technical Studies, Belgrade, Republic of Serbia, organic.polj.72@gmail.com
[50] Lecturer Ljubica Šarčević-Todosijević, MA, Candidate for IRASA Academician, High Medical-Sanitary School of Professional Studies "Visan", Belgrade, Republic of Serbia, ljsarcevic@gmail.com

## Introduction

With the global population on the rise, food security is a pressing issue, particularly for countries with limited agricultural land like those in the Middle East, Japan, and Singapore (Laborde et al., 2020). The COVID-19 pandemic has further emphasized the importance of food security as borders closed and food trade was disrupted. Some countries even halted food exports to ensure domestic supply. In times of crises like pandemics, it is crucial to explore alternative sources of food and ingredients (Laborde et al., 2020). One innovative approach is using engineered microorganisms like bacteria and yeast to produce food compounds. Interdisciplinary collaborations between mathematics, computational science, physics, and metabolic engineering could offer promising solutions to enhance food security in the future.

## Modeling strategies in metabolic engineering

Computational modelling in biological research involves using mathematical and statistical algorithms in computer programs to analyze experimental data, providing insights into biological systems and making predictions that guide further lab work(Helmy et al., 2009; Piras et al., 2014; Selvarajoo, 2017, 2018).It has become essential in studying cellular networks and metabolic engineering, where it is used to manipulate microorganisms to optimize protein or substance production(Kim et al., 2018). When natural organisms do not produce sufficient amounts of desired compounds, metabolic engineering can be employed to modify key pathways and enzymes for better production (Nozzi et al., 2014; Xiao et al., 2015). This can involve controlling transcription and translation, engineering enzymes, and optimizing growth conditions (Kallscheuer, 2018). The goal is to maximize production yield for economically viable outcomes, especially for rare and valuable products like taste substances and cosmetics (García-Granados et al., 2019; Shukal et al., 2019). Computational models play a crucial role in achieving this goal, providing new insights and solutions to challenges in metabolic engineering. By maximizing titres, rates, and yields, these modelling approaches aim to make biological production methods competitive with chemical synthesis and extraction from natural sources (Zhang et al., 2020a; Comba et al., 2012). The use of computational modelling in metabolic engineering continues to advance and improve industrial production processes (Saa and Nielsen, 2017).

### Dynamic and constraint-based metabolic modeling

There are two main types of modelling approaches used today: parametric and non-parametric (Kim et al., 2018). Parametric approaches involve dynamic modeling using ordinary differential equations, while non-parametric models utilize Boolean logics, stoichiometric matrix, and Bayesian inference algorithms (John et al., 2019; Toya and Shimizu, 2013). Dynamic models built using differential equations allow for the

construction of an organism's metabolism based on known biochemical reactions and reaction kinetics derived from genomic, enzymatic, and biochemical information (Selvarajoo et al., 2008; McCloskey et al., 2013). These models can predict metabolic outcomes for different perturbations and identify key regulatory mechanisms and flux distributions (Saa and Nielsen, 2017).However, for large-scale modeling like genome-scale modeling, the absence of large-scale experimentally measured kinetics poses a challenge for dynamic modeling. As a solution, scientists use constraint-based modeling approaches, such as flux balance analysis (FBA) (Bordbar et al., 2014). FBA models thousands of metabolites and reactions with reasonable computational cost and can determine the contribution of each gene to a certain trait. It is used for analyzing, optimizing, and designing metabolic pathways.Constraint-based models have constraints for each decision, representing the minimum and maximum values of the decision, such as reaction rates (Orth et al., 2010). FBA is a widely used constraint-based modeling approach that provides valuable insights into metabolic engineering. By combining dynamic and constraint-based modeling approaches, scientists can gain a comprehensive understanding of metabolism and design more efficient pathways for biotechnological applications (Skraly et al., 2018).

**Transcriptional control and ensemble modeling for metabolic pathway analysis**

Recent research has shown that transcriptional and translation control can lead to significant increases in the yield of targeted metabolites compared to traditional metabolic regulation methods (Shukal et al., 2019; Curran et al., 2014). Transcriptional control involves modifying the promoter region of the gene of interest, such as mutating ribosomal binding sites, transcription factor binding sites, or designing artificial promoter regions. Understanding how genes are regulated and the genomic structure around binding sites is crucial for effective transcriptional control (Curran et al., 2014). Statistical approaches like position weight matrix modeling have shown promise in understanding the impact of mutations on transcriptional regulation in mammalian cells. Ensemble modeling is another strategy used in metabolic engineering to predict pathway outcomes by combining multiple models with different algorithms or training sets (Sharon et al., 2012). This approach allows for the simulation of network changes in response to perturbations, without the need for detailed kinetic parameters. However, ensemble modeling requires accurate perturbation data and faces challenges in building models with different algorithms and interpreting results (Yiu Chan et al., 2019; Ji et al., 2018). An example of ensemble modeling success was seen in a study involving non-native central pathways for carbon conservation, where ensemble modeling robustness analysis helped determine probabilities of system failure and identify areas for flux improvement (Kotu and Deshpande, 2015). Ensemble modeling has also been used in developing a strain of E. coli that produces L-lysine. Despite its potential, ensemble modeling has limitations including the need for reliable data and the complexity of interpreting results, which may hinder its widespread use in metabolic engineering(Tran et al., 2008). Overall, the combination of transcriptional and

translation control, along with ensemble modeling strategies, shows promise in improving the efficiency of metabolic engineering outcomes (Contador et al., 2009). However, further research is needed to overcome challenges and fully utilize these innovative approaches in the field of metabolic engineering.

**Protein Modeling for Metabolic Engineering**

In silico three-dimensional molecular modeling is a widely used approach for metabolic engineering, specifically for studying receptor/enzyme-ligand docking and protein homology design (Wang et al., 2018). This modeling technique has various applications in drug design, metabolism research, therapeutic antibodies design, and molecular interactions research (Fisher et al., 2014). In metabolic engineering, 3D modeling is crucial for designing and simulating engineered enzymes to optimize the microorganism's metabolism. In cases where structural data is unavailable, molecular modeling is used to predict the 3D structures of enzymes and perform enzyme-substrate docking studies to enhance specificity, activity, and stability in a specific environment. This approach has been successful in designing single enzymes for industrial biocatalysts (e.g. sitagliptin (Savile et al., 2010)), as well as for entire enzyme cascades (e.g. islatravir (Huffman et al., 2019)) for the production of active pharmaceutical ingredients.

## Limitations of current modeling strategies

Dynamic modeling strategies rely heavily on model parameters, such as reaction kinetics or flux ranges, which can be determined using bottom-up or top-down approaches (Cuperlovic-Culf, 2018). The bottom-up approach relies on experimental data, such as enzymatic assays, to gather information on reaction kinetics, which can be challenging to obtain for all enzymes in a pathway. In vitro data is often significantly different from in vivo conditions (Selvarajoo et al., 2009). Additionally, modeling requires data for multiple conditions or time points to train and test the model accurately, leading to iterative experimental work (Helmy et al., 2009). While bottom-up modeling uses optimization algorithms like genetic algorithms to estimate model parameters, the complex and non-linear relationships between metabolites can limit the effectiveness of these algorithms. Overall, dynamic modeling strategies can be complex and require careful consideration of experimental data and parameters to accurately represent biological systems (Cuperlovic-Culf, 2018; Srinivasan et al., 2015).

**Another limiting aspect is the scale of the model.**

The bottom-up approach in modeling biological systems is more suitable for small-scale models due to the detailed experimental measurements required. However, extending the model size leads to higher costs and longer time requirements or increased reliance on computational estimation, resulting in lower accuracy (Andreozzi et al., 2016). Ensemble modeling can help build large-scale models but

comes with limitations. In contrast, top-down approaches use metabolomic data to infer kinetics, flux rates, and metabolite concentrations through correlation and causation networks (Cuperlovic-Culf, 2018). These networks establish cause-effect relationships and probabilistic relations between enzymes and metabolites. Top-down methods often use optimization algorithms to estimate model parameters, but the complex and nonlinear relationships within metabolic models limit the effectiveness of fitting algorithms (Srinivasan et al., 2015). Despite this, top-down approaches have been successful in analyzing cellular pathways with simple linear or mass-action kinetic models(Selvarajoo, 2011, Selvarajoo et al. 2009).Comparative 3D protein modeling is commonly performed using template-based methods, where homologous protein structures are used to generate models (Cuperlovic-Culf, 2018). However, the availability of good templates can affect the quality of the models generated. These modeling strategies require high-quality experimental data, including metabolite concentrations, chemical structures, genomic sequences, and gene expression data (Sali and Blundell, 1993). Fortunately, there are many bioinformatics databases and servers available that aggregate and collect this data, such as KEGG and MetaCyc (Kanehisa, 2004; Caspi et al., 2020). Despite the benefits of these resources, the challenge lies in finding the correct dataset and modeling approaches to utilize this data effectively, highlighting the need for novel data mining and analytics approaches like artificial intelligence.Overall, various modeling strategies exist for understanding biological systems, each with its advantages and limitations. The availability of bioinformatics resources and the potential of artificial intelligence are key factors in addressing the challenges of modeling complex biological systems accurately.

## Integration of artificial intelligence into metabolic engineering research

Artificial intelligence (AI) is a powerful tool that allows computers to make decisions by analyzing data and following predetermined rules or patterns. In the fields of biomedicine and biotechnology, AI is being increasingly employed to address challenges in drug discovery, vaccine development, genomics, and cancer diagnosis (Smith, 2020a). Over 40 pharmaceutical companies and 230 startup companies are utilizing AI in drug discovery, leading to the development of over one hundred drugs in various stages of development. AI has also been instrumental in the rapid development of drugs and vaccines for COVID-19(Smith, 2020b, 2020c).While AI has seen significant success in these areas, it is still underutilized in fields such as metabolomics and metabolic engineering for food applications(Regulatory Affairs Professionals Society, 2020; Ledford, 2020). Combining AI with systems biology can greatly enhance our understanding of metabolism, although this area remains relatively unexplored. Machine learning (ML), a subset of AI, is particularly interested in developing programs that can learn and improve performance based on experience without explicit programming (Alipanahi et al., 2015; Hui et al., 2013; Poplin et al., 2018). Recent advancements in ML have enabled its application in identifying biomarkers for weight loss, discovering food identity markers, and studying farm

animal metabolism.In the realm of metabolic engineering, ML is being utilized for biosystems design, microbial bio-manufacturing, pathways discovery, essential enzyme identification, metabolism modeling, genome annotation, multi-omics dataset analysis, and 3D protein modeling(Zelezniak et al., 2018). By integrating ML with systems biology, researchers are able to enhance their understanding of complex biological systems and improve the efficiency of biotechnological processes (Volk et al., 2020; Choi et al., 2019). As technology continues to advance and produce larger datasets, the potential for AI and ML in the fields of metabolomics and metabolic engineering is vast, promising innovative solutions to current challenges.

**ML for genome annotation**

Pathway identification and analysis are essential for metabolic engineering when the biochemical pathway of a desired substance is not well-known. Transferring the responsible gene or gene cluster to a model organism for optimization is often necessary. However, traditional modeling techniques have limitations, and combining omics data with standard approaches for pathway analysis results in uncertainty (García-Granados et al., 2019). Machine learning (ML) can be used to identify pathways upstream of target substances, as demonstrated by models using naive Bayes, decision trees, and logistic regression in MetaCyc(Cheng et al., 2015). These ML models outperform traditional mathematical and statistical methods in predicting the presence of novel metabolic pathways in newly-sequenced organisms. The analysis highlights the importance of the number of reactions along the pathway as a key feature for prediction accuracy. Despite the success of ML in pathway prediction, current methods still heavily rely on traditional approaches like gene sequence similarity (Cuperlovic-Culf, 2018). Therefore, there is a need for improved ML algorithms to enhance pathway discovery in metabolic engineering.
Machine Learning (ML) is a valuable tool for identifying important genes or enzymes in metabolic pathways. Various ML classifiers, such as support vector machines, logistic regression, and decision tree-based models, have been crucial in predicting gene essentiality within pathways by using labeled data of essential and non-essential genes (Quest et al., 2010). This method has also been used to discover new drug targets by determining essential enzymes in metabolic networks based on local network topology, co-expression, gene homologies, and flux balance analyses. An ML model created by Plaimas et al. was trained to differentiate between essential and non-essential reactions, and then validated experimentally using single knockout mutants of E. coli from the KEIO collection. This model also helped detect errors in experimental data when predictions did not align with the collection. In another study, the side effects of drugs on metabolic networks were investigated through an ML model that predicted enzyme inhibitory effects (Nandi et al., 2017). The model utilized network topology, functional classes of inhibitors and enzymes, logic-based representation, abduction, and induction methods to predict drug inhibitory side effects. Simulations showed that with sufficient training data, non-ground hypotheses had better predictive accuracy (Plaimas et al., 2008). Overall, ML plays a crucial role

in identifying genes, enzymes, and drug targets within metabolic pathways and predicting drug inhibitory effects (Tamaddoni-Nezhad et al., 2006).

## ML for genome annotation

Newly sequenced genomes undergo both structural annotation and functional annotation processes. Structural annotation involves identifying genome components such as genes and their structures, while functional annotation focuses on determining gene functions and their products. These annotations are crucial for metabolic engineering research as they help in finding alternative organisms with similar genes or pathways (Bradbury et al., 2013; Ikeda et al., 2014). Various methods, including comparative genomics and network biology, are used for annotation. Advancements in genome sequencing technologies have led to the development of faster and more accurate annotation methods that analyze genomes from different platforms (El-Metwally et al., 2014).Despite progress, challenges remain, such as missing genes and errors in annotation. To address these issues, new methods have been introduced that incorporate multi-omics data, particularly proteomic and transcriptomic data (Armstrong et al., 2019; Li et al., 2019). Machine learning algorithms have been increasingly utilized in genome annotation, with models trained on annotated genomes to identify structures in new sequences. ML methods have been developed to identify various components like protein-coding genes and regulatory elements in genomes.
Some recent approaches include the use of deep learning in genome annotation, with tools like DeepAnnotator outperforming traditional pipelines (Khodabandelou et al., 2020). ML algorithms are being integrated into annotation tools like GeneMarks, improving gene identification in prokaryotic and eukaryotic genomes. Additionally, deep convolutional neural networks have been employed to annotate gene-start sites across species (Tetko et al., 2005). While ML has been more widely used in structural annotation, efforts are being made to enhance its role in functional annotation processes. ML-based methods have been applied to predict gene functions, control false discovery rates, and annotate splice variants in eukaryotes (Panwar et al., 2016).

## ML of multi-omics datasets

The advancements in -omics technologies have led to a massive accumulation of data, including genomics, transcriptomics, proteomics, and metabolomics (Stephens et al., 2015). This data is expected to grow exponentially by 2025. As a result, there has been a shift towards data-driven approaches such as machine learning in scientific research (Cuperlovic-Culf, 2018). By combining machine learning methods with omics data, researchers are able to address various biomedical challenges.One example of utilizing machine learning in omics data analysis is in estimating metabolite concentrations over time (Costello and Martin, 2018). By combining machine learning models with proteomic and metabolomic time series data, researchers were able to make qualitative and quantitative predictions that outperformed traditional kinetic

models (Zelezniak et al., 2018). Additionally, machine learning has been used to predict the yeast metabolome under different perturbation conditions by analyzing enzyme expression proteome data from kinase knockouts. Furthermore, the availability of transcriptome data has enabled the development of genome-scale methods for predicting cell phenotypes using machine learning models (Guo et al., 2017). DeepMetabolism, a biology-guided deep learning system, uses transcriptomics data to predict cell phenotypes with high accuracy and speed. Machine learning algorithms have also been utilized to model bacterial ribosome binding sites and predict optimal high-producers, which has broad applications in metabolic engineering.In addition to transcriptomic and proteomic data, machine learning methods have also been applied to fluxomic data, which describes the complete set of metabolic fluxes in an organism (Jervis et al., 2019). Tools like MFlux use machine learning to identify relationships between environmental and genetic factors and metabolic fluxes, providing predictive models that accelerate flux quantification (Wu et al., 2016a). Machine learning approaches have also been integrated with genome-scale modeling to evaluate microbial factory performance, producing accurate predictions through ensemble modeling techniques.Overall, machine learning has become an essential tool in analyzing and interpreting the vast amounts of omics data generated by modern technologies. By combining machine learning with omics data, researchers are able to make more accurate predictions and gain deeper insights into complex biological systems (Oyetunde et al., 2019).

## ML in protein modeling

Recent advances in 3D protein modeling using AI have been demonstrated at the Critical Assessment of protein Structure Prediction (CASP) meeting in 2018. The program AlphaFold utilized neural networks to extract covariant residue pairs from sequence alignments and estimated distances to fold proteins using the ROSETTA energy function (Senior et al., 2020). This method outperformed other approaches, producing high-accuracy models for 24 out of 43 domains. ProSPr, a lab-based version of AlphaFold, has since been developed. Additionally, trROSETTA, developed by Yang et al, improved predictions further by estimating relative residue orientations (Alford et al., 2017). These AI-based 3D modeling methods have the potential to be integrated into comprehensive metabolic engineering platforms (Billings et al., 2019). One area for improvement is the inclusion of cofactors, essential molecules for enzyme folding and function (Yang et al., 2020). Template-based modeling often overlooks cofactors, resulting in the need for additional manipulation to incorporate them back into the structure. By surveying the Protein Data Bank and utilizing ML methods to identify cofactor binding determinants, a combined sequence-and-template-based optimization protocol can be developed to account for these structural features (Higgins et al., 2005). This approach could also be extended to identify substrates for enzymes in metabolic pathways, aiding in the design of synthetic biosynthetic pathways.Enzyme sequence alignments and structural information can be used in combination with neural networks to identify common binding pocket residues for

different substrates across enzyme families (Berman et al., 2002). This enables the prediction of potential sequences suitable for inclusion in metabolic pathways, considering ease of integration into heterologous expression systems. Additionally, it may be possible to predict mutations to binding pocket residues to produce specific products when no suitable sequence is available. These predictions can be experimentally validated, with results contributing to model refinement.

### Application of ML models for engineering bio-economy strains

The importance of utilizing natural and food ingredients from diverse sources, such as engineered microbes or synthetically derived, for creating a sustainable bio-based economy is highlighted. The complexity of living systems poses challenges in optimizing bioengineering processes, leading to suboptimal results. Efforts in metabolic engineering focus on enhancing production strains for economically viable large-scale production of microbial-derived compounds (Yadav et al., 2012; Lim and Kim, 2019). Brunk et al engineered E. coli lab strains to produce biofuels and identified key regulatory bottlenecks through systems biology approaches (Brunk et al., 2016). They found time-dependent regulation of metabolic pathways affecting biofuel production and identified a crucial gene for improving isopentenol production. Costello and Martin developed a machine learning model to predict pathway dynamics using proteomics and metabolomics data, which outperformed traditional kinetic models. Their model provided accurate qualitative and quantitative predictions without detailed regulatory knowledge (Brunk et al., 2016). ML modeling has the potential to guide bioengineering strains effectively without complete understanding of metabolic regulatory processes. In the food and consumer care industries, terpenes and terpenoids are industrially relevant metabolic engineering products that can be produced using microbes (Costello and Martin, 2018). While significant progress has been made at the laboratory scale, challenges remain in scaling up production in industrial bioreactors. ML models can help optimize production processes by identifying targeted steps for improving output in bioreactor scale production (Caputi and Aprea, 2011; Zhang et al., 2020b). Although evidence is currently lacking, the future prospects for utilizing ML models in metabolic engineering are promising, provided significant investments are made to generate the biological data required for dynamic modeling (Czajka et al., 2018).

## Challenges and future projections

Integrating systems biology and machine learning (ML) shows promise in improving the study and understanding of metabolism, as well as in enhancing the development of healthier, affordable, and nutritious alternative food sources. However, challenges and limitations exist in fully utilizing the power of both systems biology and ML. One major challenge is the lack of data, especially high throughput data from various omics levels, which is necessary for a holistic study of microorganisms (García-Granados et al., 2019). This data scarcity hinders the discovery of new pathways,

proteins, and enzymes with better production rates, as well as the selection of suitable organisms for engineering projects (Khoury et al., 2014). Additionally, the availability of sufficient information about both donor organisms and chassis is crucial to avoid unexpected qualities or complications in metabolic engineering projects.In the context of ML, training models for metabolic engineering requires quantitative data for multiple conditions, such as knockouts, perturbations, and growth conditions (Zhang et al., 2015). The differences observed in various cultivation scales further complicate the training process, as the behavior of microorganisms can vary between lab setups and industrial bioreactors due to factors like lighting systems, mixing, and gas transfer(Brennan and Owende, 2010; Johnson et al., 2018). Therefore, data generated from lab setups may not be suitable for modeling the dynamics of large-scale cultivation setups. High-quality quantitative data is essential for developing predictive ML models related to translation control, transcription factor binding sites, enzyme engineering, and growth optimization in various conditions (Winkler et al., 2015; Birkel et al., 2017; Arkin et al., 2018). Overall, addressing these data challenges is crucial for advancing the integration of systems biology and ML in metabolic engineering research and realizing their full potential in improving food sources and understanding metabolism. The integration of systems biology and machine learning in metabolic engineering research aims to utilize multi-omics data through the development of mathematical and statistical models (Helmy et al., 2016). While several databases exist for AI applications, there is a need for high-quality quantitative data and online resources to support the building of AI models for metabolic engineering. The lack of details in current online resources hinders the application of data mining and machine learning methods (Rieder and Simon, 2017; MT Ribeiro, 2016; Burrell, 2016). The black box problem in AI techniques presents challenges in understanding and improving the quality of output, particularly in fields such as structural modeling and enzyme-substrate identification. Explainable artificial intelligence (XAI) methods are being developed to make AI results more understandable to humans and address the black box problem. Genome annotation, especially in food-grade microorganisms, significantly impacts metabolic engineering and applications in the food industry. Improved ML-based genome annotation methods are needed to enhance the annotation of food-safe genomes and facilitate research in this area (Zednik, 2019). Pathway prediction in the absence of genome sequence or annotation is another challenge, particularly for food-safe microorganisms. ML methods can predict pathways using omics data, offering a solution to traditional pathway prediction approaches.Despite challenges and limitations, there is great potential in using AI and ML techniques to explore the complexities of living systems. The combination of physical or biochemical theories with heuristic approaches and ML can support future systems biology efforts. Improvements in data collection from omics technologies will help narrow the gap of uncertainty and ambiguity in systems biology and ML integration for optimizing metabolic engineering strategies.Overall, the integration of systems biology and machine learning in metabolic engineering research has the potential to advance the field by utilizing multi-omics data, developing mathematical and statistical models,

and addressing challenges such as the black box problem in AI techniques. Improved genome annotation methods and pathway prediction using ML can enhance research in food-grade microorganisms, supporting the development of novel metabolic engineering strategies. Continued advancements in AI and ML techniques, along with improvements in omics data collection, will contribute to the future success of integrating systems biology and machine learning in metabolic engineering research.

**Acknowledgements**

The organizer gratefully acknowledges the work done by Scientific Committee of the SETI V 2024 International Scientific Conference for efforts done for the success of this event.

## References

[1] Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., Labonte, J.W., Pacella, M.S., Bonneau, R., Bradley, P., Dunbrack, R.L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., Gray, J.J., 2017. The rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theor. Comput*. 13, 3031–3048.
[2] Contador, C.A., Rizk, M.L., Asenjo, J.A., Liao, J.C., 2009. Ensemble modeling for strain development of l-lysine-producing Escherichia coli. Metab. Eng. 11, 221–233.
[3] Costello, Z., Martin, H.G., 2018. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *Npj Syst. Biol. Appl*. 4, 19.
[4] Cuperlovic-Culf, M., 2018. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* 8, 4. https://doi.org/10.3390/metabo8010004.
[5] Curran, K.A., Crook, N.C., Karim, A.S., Gupta, A., Wagman, A.M., Alper, H.S., 2014. Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun*. 5, 1–8.
[6] Czajka, J.J., Nathenson, J.A., Benites, V.T., Baidoo, E.E.K., Cheng, Q., Wang, Y., Tang, Y.J., 2018. Engineering the oleaginous yeast Yarrowia lipolytica to produce the aroma compound β-ionone. *Microb. Cell Factories* 17, 136.
[7] Dias-Audibert, F.L., Navarro, L.C., de Oliveira, D.N., Delafiori, J., Melo, C.F.O.R., Guerreiro, T.M., Rosa, F.T., Petenuci, D.L., Watanabe, M.A.E., Velloso, L.A., Rocha, A.R., Catharino, R.R., 2020. Combining machine learning and metabolomics to identify weight gain biomarkers. *Front. Bioeng. Biotechnol*. 8, 6. https://doi.org/10.3389/fbioe.2020.00006.
[8] El-Metwally, S., Ouda, O.M., Helmy, M., 2014. Challenges in the Next-Generation Sequencing Field. Springer, New York, NY, 45–49.
[9] Erban, A., Fehrle, I., Martinez-Seidel, F., Brigante, F., Mas, A.L., Baroni, V., Wunderlin, D., Kopka, J., 2019. Discovery of food identity markers by metabolomics and machine learning technology. *Sci. Re*p. 9, 1–19.

[10] Fisher, A.K., Freedman, B.G., Bevan, D.R., Senger, R.S., 2014. A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *Comput. Struct. Biotechnol. J.* 11, 91–99.
[11] García-Granados, R., Lerma-Escalera, J.A., Morones-Ramírez, J.R., 2019. Metabolic engineering and synthetic biology: synergies, future, and challenges. *Front. Bioeng. Biotechnol*. 7, 36.
[12] Ghaffari, M.H., Jahanbekam, A., Sadri, H., Schuh, K., Dusel, G., Prehn, C., Adamski, J., Koch, C., Sauerwein, H., 2019. Metabolomics meets machine learning: longitudinal metabolite profiling in serum of normal versus overconditioned cows and pathway analysis. *J. Dairy Sci.* 102, 11561–11585.
[13] Guo, W., Xu, Y., Feng, X., 2017. DeepMetabolism: A Deep Learning System to Predict Phenotype from Genome Sequencing. http://arxiv.org/abs/1705.03094.
[14] Heinemann, J., 2019. Machine learning in untargeted metabolomics experiments. In: *Methods Mol. Biol. Humana Press Inc*., 287–299.
[15] Helmy, M., Gohda, J., Inoue, J., Tomita, M., Tsuchiya, M., Selvarajoo, K., 2009. Predicting novel features of toll-like receptor 3 signaling in macrophages. PloS One 4, e4661.
[16] Helmy, M., Tomita, M., Ishihama, Y., 2012. Peptide identification by searching large-scale tandem mass spectra against large databases: bioinformatics methods in proteogenomics Metabolomics View project chi sequence View project. *Genes, Genomes Genomics* 6, 76–85.
[17] Helmy, M., Crits-Christoph, A., Bader, G.D., 2016. Ten simple rules for developing public biological databases. *PLoS Comput. Biol*. 12, e1005128 https://doi.org/10.1371/ journal.pcbi.1005128.
[18] Higgins, C., Muralidhara, B., Wittung-Stafshede, P., 2005. How do cofactors modulate protein folding*? Protein Pept. Lett*. 12, 165–170.
[19] Hong, J., Luo, Y., Zhang, Y., Ying, J., Xue, W., Xie, T., Tao, L., Zhu, F., 2020. Protein functional annotation of simultaneously improved stability, accuracy and false discovery rate achieved by a sequence-based deep learning. *Brief. Bioinform*. 21, 1437–1447.
[20] Huffman, M.A., Fryszkowska, A., Alvizo, O., Borra-Garske, M., Campos, K.R., Canada, K.A., Devine, P.N., Duan, D., Forstater, J.H., Grosser, S.T., Halsey, H.M., Hughes, G.J., Jo, J., Joyce, L.A., Kolev, J.N., Liang, J., Maloney, K.M., Mann, B.F., Marshall, N.M., McLaughlin, M., Moore, J.C., Murphy, G.S., Nawrat, C.C., Nazor, J., Novick, S., Patel, N.R., Rodriguez-Granillo, A., Robaire, S.A., Sherer, E.C., Truppo, M.D., Whittaker, A.M., Verma, D., Xiao, L., Xu, Y., Yang, H., 2019. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* 366, 1255–1259.
[21] Hui, S., Xing, X., Bader, G.D., 2013. Predicting PDZ domain mediated protein interactions from structure. BMC Bioinformatics 14. https://doi.org/10.1186/1471-2105-14-27. Ikeda, H., Shin-Ya, K., Omura, S., 2014. Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J. Ind. Microbiol. Biotechnol*. 41, 233–250.

[22] Jervis, A.J., Carbonell, P., Vinaixa, M., Dunstan, M.S., Hollywood, K.A., Robinson, C.J., Rattray, N.J.W., Yan, C., Swainston, N., Currin, A., Sung, R., Toogood, H., Taylor, S., Faulon, J.L., Breitling, R., Takano, E., Scrutton, N.S., 2019. Machine learning of designed translational control allows predictive pathway optimization in Escherichia coli. *ACS Synth. Biol*. 8, 127–136.
[23] Ji, Z., He, L., Rotem, A., Janzer, A., Cheng, C.S., Regev, A., Struhl, K., 2018. Genome-scale identification of transcription factors that mediate an inflammatory network during breast cellular transformation. *Nat. Commun*. 9
[24] John, P.C. St, Strutz, J., Broadbelt, L.J., Tyo, K.E.J., Bomble, Y.J., 2019. Bayesian inference of metabolic kinetics from genome-scale multiomics data. *PLoS Comput. Biol*. 15, e1007424.
[25] Johnson, T.J., Katuwal, S., Anderson, G.A., Gu, L., Zhou, R., Gibbons, W.R., 2018. Photobioreactor cultivation strategies for microalgae and cyanobacteria. *Biotechnol. Prog*. 34, 811–827.
[26] Kallscheuer, N., 2018. Engineered microorganisms for the production of food additives approved by the European Union-A systematic analysis. *Front. Microbiol*. 9, 1746.
[27] Kanehisa, M., 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 32, 277D–280. https://doi.org/10.1093/nar/gkh063. Khodabandelou, G., Routhier, E., Mozziconacci, J., 2020.
[28] Genome annotation across species using deep convolutional neural networks. *PeerJ Comput. Sci*. 6, e278.
[29] Khoury, G.A., Smadbeck, J., Kieslich, C.A., Floudas, C.A., 2014. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol*. 32, 99–109.
[30] Kim, O.D., Rocha, M., Maia, P., 2018. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. *Front. Microbiol*. 9, 1690.
[31] Kiritchenko, S., Matwin, S., Famili, A.F. Functional annotation of genes using hierarchical text categorization. n.d. http://www.pdg.cnb.uam.es/BioLINK/. (Accessed 9 August 2020). Kotu, V., Deshpande, B., 2015. Data mining process. In: *Predict. Anal. Data Min. Elsevier*, 17–36.
[32] Laborde, D., Martin, W., Swinnen, J., Vos, R., 2020. COVID-19 risks to global food security. Science 369, 500–502, 80-. Ledford, H., 2020. Dozens of coronavirus drugs are in development - what happens next? *Nature* 581, 247–248.
[33] Lee, Y., Lafontaine Rivera, J.G., Liao, J.C., 2014. Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metab. Eng*. 25, 63–71.
[34] Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., Walters, J.R., 2019. Insect genomes: progress and challenges. *Insect Mol. Biol*. 28, 739–758. https://doi.org/ 10.1111/imb.12599.
[35] Liebal, U.W., Phan, A.N.T., Sudhakar, M., Raman, K., Blank, L.M., 2020. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites* 10, 243.

[36] Lim, H.J., Kim, D.-M., 2019. Cell-free metabolic engineering: recent developments and future prospects. *Methods Protoc* 2, 33.
[37] Lomsadze, A., Gemayel, K., Tang, S., Borodovsky, M., 2018. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Res.* 28, 1079–1089.
[38] Mahood, E.H., Kruse, L.H., Moghe, G.D., 2020. Machine learning: a powerful tool for gene function prediction in plants. *Appl. Plant Sci*. 8.
[39] McCloskey, D., Palsson, B., Feist, A.M., 2013. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Mol. Syst. Biol*. 9.
[40] Mt Ribeiro, S.S.C.G., 2016. Why Should I Trust You? Explaining the Predictions of Any Classifier, *ArXiv*. 1602, 04938v3.
[41] Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., Thomas, A.D., Ali, R., Sharma, K., Kyrpides, N.C., Reddy, T.B.K., 2017. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res*. 45, D446–D456.
[42] Nakano, F.K., Lietaert, M., Vens, C., 2019. Machine learning for discovering missing or wrong protein function annotations. *BMC Bioinformatics* 20, 1–32.
[43] Nandi, S., Subramanian, A., Sarkar, R.R., 2017. An integrative machine learning strategy for improved prediction of essential genes in Escherichia coli metabolism using fluxcoupled features. *Mol. Biosyst*. 13, 1584–1596.
[44] Nozzi, N.E., Desai, S.H., Case, A.E., Atsumi, S., 2014. Metabolic engineering for higher alcohol production. *Metab. Eng.* 25, 174–182. https://doi.org/10.1016/j.ymben.2014.07.007.
[45] Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? *Nat. Biotechnol*. 28, 245–248.
[46] Oyetunde, T., Liu, D., Martin, H.G., Tang, Y.J., 2019. Machine learning framework for assessment of microbial factory performance. *PloS One 14*, e0210558.
[47] Panwar, B., Menon, R., Eksi, R., Li, H.D., Omenn, G.S., Guan, Y., 2016. Genome-wide functional annotation of human protein-coding splice variants using multiple instance learning. *J. Proteome Res*. 15, 1747–1753.
[48] Piras, V., Tomita, M., Selvarajoo, K., 2014. Transcriptome-wide variability in single embryonic development cells. *Sci. Rep*. 4, 1–9. https://doi.org/10.1038/srep07137.
[49] Plaimas, K., Mallm, J.P., Oswald, M., Svara, F., Sourjik, V., Eils, R., Konig, R., 2008. € Machine learning based analyses on metabolic networks supports high-throughput knockout screens. *BMC Syst. Biol*. 2
[50] Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., Gross, S.S., Dorfman, L., McLean, C.Y., Depristo, M.A., 2018. A universal snp and small-indel variant caller using deep neural networks. *Nat. Biotechnol*. 36, 983.
[51] Quest, D.J., Land, M.L., Brettin, T.S., Cottingham, R.W., 2Contador, C.A., Rizk, M.L., Asenjo, J.A., Liao, J.C., 2009. Ensemble modeling for strain development of l-lysine-producing Escherichia coli. *Metab. Eng*. 11, 221–233.

[52] Costello, Z., Martin, H.G., 2018. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *Npj Syst. Biol. Appl.* 4, 19.
[53] Cuperlovic-Culf, M., 2018. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* 8, 4.
[54] Curran, K.A., Crook, N.C., Karim, A.S., Gupta, A., Wagman, A.M., Alper, H.S., 2014. Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun*. 5, 1–8.
[55] Czajka, J.J., Nathenson, J.A., Benites, V.T., Baidoo, E.E.K., Cheng, Q., Wang, Y., Tang, Y.J., 2018. Engineering the oleaginous yeast Yarrowia lipolytica to produce the aroma compound β-ionone. *Microb. Cell Factories* 17, 136. https://doi.org/10.1186/s12934-018-0984-x.
[56] Dias-Audibert, F.L., Navarro, L.C., de Oliveira, D.N., Delafiori, J., Melo, C.F.O.R., Guerreiro, T.M., Rosa, F.T., Petenuci, D.L., Watanabe, M.A.E., Velloso, L.A., Rocha, A.R., Catharino, R.R., 2020. Combining machine learning and metabolomics to identify weight gain biomarkers. *Front. Bioeng. Biotechn*ol. 8, 6.
[57] El-Metwally, S., Ouda, O.M., Helmy, M., 2014. Challenges in the Next-Generation Sequencing Field. Springer, New York, NY, 45–49.
[58] Erban, A., Fehrle, I., Martinez-Seidel, F., Brigante, F., Mas, A.L., Baroni, V., Wunderlin, D., Kopka, J., 2019. Discovery of food identity markers by metabolomics and machine learning technology. *Sci. Rep*. 9, 1–19.
[59] Fisher, A.K., Freedman, B.G., Bevan, D.R., Senger, R.S., 2014. A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *Comput. Struct. Biotechnol. J.* 11, 91–99.
[60] García-Granados, R., Lerma-Escalera, J.A., Morones-Ramírez, J.R., 2019. Metabolic engineering and synthetic biology: synergies, future, and challenges. *Front. Bioeng. Biotechnol*. 7, 36.
[61] Ghaffari, M.H., Jahanbekam, A., Sadri, H., Schuh, K., Dusel, G., Prehn, C., Adamski, J., Koch, C., Sauerwein, H., 2019. Metabolomics meets machine learning: longitudinal metabolite profiling in serum of normal versus overconditioned cows and pathway analysis. *J. Dairy Sci*. 102, 11561–11585.
[62] Guo, W., Xu, Y., Feng, X., 2017. DeepMetabolism: A Deep Learning System to Predict Phenotype from Genome Sequencing.
[63] Heinemann, J., 2019. Machine learning in untargeted metabolomics experiments. In: *Methods Mol. Biol. Humana Press Inc*., 287–299.
[64] Helmy, M., Gohda, J., Inoue, J., Tomita, M., Tsuchiya, M., Selvarajoo, K., 2009. Predicting novel features of toll-like receptor 3 signaling in macrophages. *PloS One* 4, e4661.
[65] Helmy, M., Tomita, M., Ishihama, Y., 2012. Peptide identification by searching large-scale tandem mass spectra against large databases: bioinformatics methods in proteogenomics Metabolomics View project chi sequence View project. Genes, *Genomes Genomics* 6, 76–85.
[66] Helmy, M., Crits-Christoph, A., Bader, G.D., 2016. Ten simple rules for developing public biological databases. *PLoS Comput. Biol.* 12, e1005128.

[67] Higgins, C., Muralidhara, B., Wittung-Stafshede, P., 2005. How do cofactors modulate protein folding? *Protein Pept. Lett*. 12, 165–170.
[68] Hong, J., Luo, Y., Zhang, Y., Ying, J., Xue, W., Xie, T., Tao, L., Zhu, F., 2020. Protein functional annotation of simultaneously improved stability, accuracy and false discovery rate achieved by a sequence-based deep learning. *Brief. Bioinform*. 21, 1437–1447.
[69] Huffman, M.A., Fryszkowska, A., Alvizo, O., Borra-Garske, M., Campos, K.R., Canada, K.A., Devine, P.N., Duan, D., Forstater, J.H., Grosser, S.T., Halsey, H.M., Hughes, G.J., Jo, J., Joyce, L.A., Kolev, J.N., Liang, J., Maloney, K.M., Mann, B.F., Marshall, N.M., McLaughlin, M., Moore, J.C., Murphy, G.S., Nawrat, C.C., Nazor, J., Novick, S., Patel, N.R., Rodriguez-Granillo, A., Robaire, S.A., Sherer, E.C., Truppo, M.D., Whittaker, A.M., Verma, D., Xiao, L., Xu, Y., Yang, H., 2019. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* 366, 1255–1259.
[70] Hui, S., Xing, X., Bader, G.D., 2013. Predicting PDZ domain mediated protein interactions from structure. *BMC Bioinformatics* 14.
[71] Ikeda, H., Shin-Ya, K., Omura, S., 2014. Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J. Ind. Microbiol. Biotechnol*. 41, 233–250.
[72] Jervis, A.J., Carbonell, P., Vinaixa, M., Dunstan, M.S., Hollywood, K.A., Robinson, C.J., Rattray, N.J.W., Yan, C., Swainston, N., Currin, A., Sung, R., Toogood, H., Taylor, S., Faulon, J.L., Breitling, R., Takano, E., Scrutton, N.S., 2019. Machine learning of designed translational control allows predictive pathway optimization in Escherichia coli. *ACS Synth. Biol.* 8, 127–136.
[73] Ji, Z., He, L., Rotem, A., Janzer, A., Cheng, C.S., Regev, A., Struhl, K., 2018. Genome-scale identification of transcription factors that mediate an inflammatory network during breast cellular transformation. *Nat. Commun*.
[74] John, P.C. St, Strutz, J., Broadbelt, L.J., Tyo, K.E.J., Bomble, Y.J., 2019. Bayesian inference of metabolic kinetics from genome-scale multiomics data. *PLoS Comput. Biol*. 15, e1007424
[75] Johnson, T.J., Katuwal, S., Anderson, G.A., Gu, L., Zhou, R., Gibbons, W.R., 2018. Photobioreactor cultivation strategies for microalgae and cyanobacteria. *Biotechnol. Prog*. 34, 811–827.
[76] Kallscheuer, N., 2018. Engineered microorganisms for the production of food additives approved by the European Union-A systematic analysis. Front. *Microbiol*. 9, 1746.
[77] Kanehisa, M., 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 32, 277D–280.
[78] Khodabandelou, G., Routhier, E., Mozziconacci, J., 2020. Genome annotation across species using deep convolutional neural networks. *PeerJ Comput. Sci*. 6, e278.
[79] Khoury, G.A., Smadbeck, J., Kieslich, C.A., Floudas, C.A., 2014. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol*. 32, 99–109.

[80] Kim, O.D., Rocha, M., Maia, P., 2018. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. *Front. Microbiol*. 9, 1690.
[81] Kotu, V., Deshpande, B., 2015. Data mining process. In: *Predict. Anal. Data Min*. Elsevier, pp. 17–36.
[82] Laborde, D., Martin, W., Swinnen, J., Vos, R., 2020. COVID-19 risks to global food security. *Science* 369, 500–502.
[83] Ledford, H., 2020. Dozens of coronavirus drugs are in development - what happens next? *Nature* 581, 247–248.
[84] Lee, Y., Lafontaine Rivera, J.G., Liao, J.C., 2014. Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metab. Eng*. 25, 63–71.
[85] Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., Walters, J.R., 2019. Insect genomes: progress and challenges. *Insect Mol. Biol*. 28, 739–758.
[86] Liebal, U.W., Phan, A.N.T., Sudhakar, M., Raman, K., Blank, L.M., 2020. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites* 10, 243.
[87] Lim, H.J., Kim, D.-M., 2019. Cell-free metabolic engineering: recent developments and future prospects. *Methods Protoc* 2, 33.
[88] Lomsadze, A., Gemayel, K., Tang, S., Borodovsky, M., 2018. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Res.* 28, 1079–1089.
[89] Mahood, E.H., Kruse, L.H., Moghe, G.D., 2020. Machine learning: a powerful tool for gene function prediction in plants. *Appl. Plant S*ci. 8.
[90] McCloskey, D., Palsson, B., Feist, A.M., 2013. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Mol. Syst. Biol.* 9.
[91] Mt Ribeiro, S.S.C.G., 2016. Why Should I Trust You? Explaining the Predictions of Any Classifier, *ArXiv*. 1602, 04938v3.
[92] Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., Thomas, A.D., Ali, R., Sharma, K., Kyrpides, N.C., Reddy, T.B.K., 2017. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res*. 45, D446–D456.
[93] Nakano, F.K., Lietaert, M., Vens, C., 2019. Machine learning for discovering missing or wrong protein function annotations. *BMC Bioinformatics* 20, 1–32.
[94] Nandi, S., Subramanian, A., Sarkar, R.R., 2017. An integrative machine learning strategy for improved prediction of essential genes in Escherichia coli metabolism using fluxcoupled features. *Mol. Biosyst.* 13, 1584–1596.
[95] Nozzi, N.E., Desai, S.H., Case, A.E., Atsumi, S., 2014. Metabolic engineering for higher alcohol production. *Metab. Eng*. 25, 174–182.
[96] Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? *Nat. Biotechnol*. 28, 245–248.
[97] Oyetunde, T., Liu, D., Martin, H.G., Tang, Y.J., 2019. Machine learning framework for assessment of microbial factory performance. *PloS One 14,* e0210558.

[98] Panwar, B., Menon, R., Eksi, R., Li, H.D., Omenn, G.S., Guan, Y., 2016. Genome-wide functional annotation of human protein-coding splice variants using multiple instance learning. *J. Proteome Res*. 15, 1747–1753.
[99] Piras, V., Tomita, M., Selvarajoo, K., 2014. Transcriptome-wide variability in single embryonic development cells. *Sci. Rep*. 4, 1–9.
[100] Plaimas, K., Mallm, J.P., Oswald, M., Svara, F., Sourjik, V., Eils, R., Konig, R., 2008. € Machine learning based analyses on metabolic networks supports high-throughput knockout screens. *BMC Syst. Biol*. 2.
[101] Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., Gross, S.S., Dorfman, L., McLean, C.Y., Depristo, M.A., 2018. A universal snp and small-indel variant caller using deep neural networks. *Nat. Biotechnol*. 36, 983.