Systems and synthetic biology tools for advanced bioproduction hosts

Yan Chen,[1,2,3,‡] Deepanwita Banerjee,[1,3,‡] Aindrila Mukhopadhyay,[1,3,4] and Christopher J. Petzold[1,2,3,]*

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA, USA
[2]Agile BioFoundry, Lawrence Berkeley National Laboratory, Emeryville, CA, USA
[3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

*Corresponding author: Petzold, Christopher J. (cjpetzold@lbl.gov)
[‡]These authors contributed equally to the manuscript.

Keywords: Synthetic biology, Systems biology, Machine learning, Metabolic engineering.

Introduction

The genomic revolution ushered in an era of discovery and characterization of enzymes from novel organisms that fueled engineering of microbes to produce commodity and high-value compounds. Over the past decade advances in synthetic biology tools in recent years contributed to significant progress in metabolic engineering efforts to produce both natural and non-natural biofuels and bioproducts [1,2] resulting in several bioproducts being brought to market [3*,4]. These successes represent a burgeoning bioeconomy, however, significant resources and time are still necessary to progress a system from proof-of-concept to market. In most cases, production of biofuels and bioproducts at economically-feasible titers, rates, and yields (TRY) rely on a systemic understanding of the metabolism of the organism and a suite of synthetic biology tools to

engineer them. Yet, examples of metabolic engineering efforts powered by a deep, systems-level knowledge of an organism and a toolbox filled with a diverse set of synthetic biology tools remain scattered. In order to fully utilize this potential, methods that examine biological systems in a comprehensive, systematic and high-throughput manner are essential [3*]. Recent success in metabolic engineering and synthetic biology efforts has coincided with the development of systems biology and analytical approaches that have kept pace and scaled with technology development. Here, we review a selection of systems biology methods and their potential use in synthetic biology approaches for developing microbial biotechnology platforms.

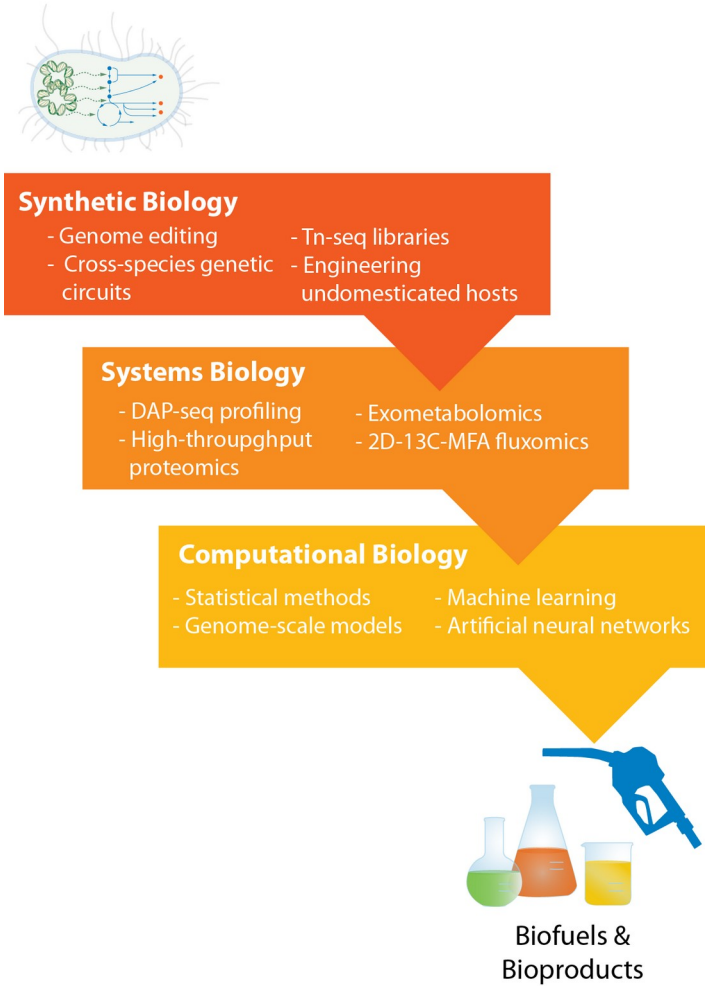
Recent developments in systems and synthetic biology

Increasingly, a diverse set of host organisms, beyond *Saccharomyces cerevisiae* and *Escherichia coli*, are the focus of biomanufacturing research (Table 1). Development of genome editing tools, such as CRISPR-Cas9, make it possible to engineer organisms that historically have been intractable to metabolic engineering [5] despite having characteristics that are attractive to biomanufacturing applications. Complementing these tools are methods to transfer DNA into cells of undomesticated organisms based on the integrative and conjugative element from *Bacillus subtilis* (ICEBs1) [6], expression systems for cross-species genetic circuits [7–9], and modular design principles for metabolic pathways [10]. Likewise, a parts-based synthetic biology strategy for constructing mutant libraries enabled the genome-wide mutant fitness screening of 32 diverse bacteria across dozens of growth conditions [11,12**]. These synthetic biology developments give researchers new control of the organisms they want to engineer with respect to tunability, orthogonality and composability of the parts they use.

Similarly, there has been significant recent progress in systems biology tools. DNA affinity purification sequencing (DAP-Seq or DAP-chip) enables low-cost and high-throughput profiling genome-wide DNA-binding motifs of individual transcription factors (TFs) [13,14] to facilitate characterization of regulatory architecture of transcriptional networks [15,16]. Likewise, high-throughput proteomic analysis capable of analyzing hundreds of samples per day [17], tools to shorten method development time [18], and automated proteomic sample preparation procedures [19] make these types of experiments routine. As microorganisms uptake and secrete metabolites from and into their environment, exometabolomics analysis, supported by various high-throughput analytical methodologies, provides a rich and valuable

phenotypic data, and gains popularity in bioprocessing and biofuel development [20,21]. Additionally, systems biology approaches contributed to better understanding of metabolism and increased the quality of genome-scale metabolic model (GSM) in various organisms [22]. More development of systems biology tools and strategies for metabolic engineering efforts in model and non-model microorganisms could be found in recent reviews and the references cited therein [3*,23,24]. Applied in the context of biomanufacturing, these tools and workflows reduce the time and effort required to complete an iteration of a DBTL cycle, speeding progress toward economic biofuels and bioproducts.

Figure 1



Integration of synthetic and systems biology to improve biomanufacturing research

Rewiring cellular metabolism to produce valuable chemicals and materials at economically competitive level benefits from an integrated approach, where tools and strategies of synthetic, systems and computational biology are used to inform changes to cellular metabolism to optimize TRY of the desired product (Figure 1). The recent breakthroughs in omics technologies enable in-depth studies of cellular metabolic networks and regulation (Table 1). The data-driven systems biology approach provides valuable insights that lead researchers identifying metabolic pathway bottlenecks and conducting fine-tuned changes to meet industrial TRY requirements. One common bottleneck is the toxicity to the

Figure 1. Overview of Systems, Synthetic and Computational biology methods that allow the use of microbial hosts for the production of biofuels and bioproducts. Several selected approaches in each category are shown.

host organism of the final product or some of the intermediates present and/or generated during conversion. George *et al*. [25*] took an integrated multi-omics approach to understand the response of *E. coli* to the accumulation of isopentenyl pyrophosphate (IPP), a toxic intermediate, in isoprenoid-producing strains. Poudel *et al.* applied integrated omics analysis to reveal metabolic adaptations of *Clostridium thermocellum*'s specific response to cytotoxic compounds derived from hydrolysis of hemicellulose and lignin, which provided systems biological information to potentially improve *C. thermocellum*'s industrial efficacy [26]. Likewise, Ohtake *et al*. [27] took a metabolomic-driven approach to identify limitations in 1-butanol production in *E. coli*. The metabolomic analysis revealed an accumulation of butanoyl-CoA suggesting that aldehyde-alcohol dehydrogenase (Adhe2) was the cause of the bottleneck. Based on this knowledge they rationally engineered *E. coli* to remove the imbalance and increased the 1-butanol titer to 18.3 g/L [27]. Characterization of the effect of membrane viscosity on cellular respiratory metabolism was probed by an integrated systemic engineering approach to manipulate lipid composition in *E. coli* and *S. cerevisiae* [28]. Systematic engineering of industrial brewing yeast cell metabolism based on metabolomic and proteomic data was used to optimize the amount of hops flavors for beer [29], and in *Corynebacterium glutamicum* to develop a robust production system for the biogasoline target, isopentenol [30]. Lim *et al*. [31**] isolated *Vibrio sp. dhg*, a fast-growing bacterium, that is naturally capable of simultaneously assimilating mannitol and alginate, the most prominent sugars in brown macroalgae. Through genome analysis and the development of a synthetic biology toolbox for this strain the team engineered it to produce ethanol from these sugars at 25.7 g/L, 1.1 g/L/h and 64% of the theoretical maximum yield.

Carbon-13 metabolic flux analysis (^{13}C-MFA) is another approach that helps trace the relative contributions of different metabolic routes. ^{13}C-MFA-guided cofactor engineering, metabolomic analysis, and gene expression profiling were applied to direct *E. coli* strain development resulting in increased acetol titer from 0.91 g/L to 2.81 g/L [32]. Data from ^{13}C-MFA was used to identify a newly discovered fructokinase enzyme (FruA) in sucrose metabolism of *Basfia succiniciproducens* [33]. By knocking out the *fruA* gene, *B. succiniciproducens* produced 71 g/L of succinate with a yield of 2.5 mol/mol from sucrose under fed-batch conditions. ^{13}C-MFA can also be used to identify engineering targets that are unintuitive to most researchers. For example, ^{13}C-MFA was used to identify energy limitation in a strain of fatty acid producing *E. coli*. The limitation was removed by expression of

Vitreoscilla hemoglobin (VHb) a membrane protein facilitating O2 transport to promote cell respiration [34].

While these studies supplied information for rational engineering approaches, construction of a strain more tolerant to various stresses often requires non-rational engineering approaches. Adaptive laboratory evolution (ALE) coupled with systems biology analyses is an inverse engineering strategy for obtaining desirable host traits and characterizing the underlying molecular mechanisms of these traits. It has been applied successfully to gain novel insights into molecular mechanisms used by *E. coli* to overcome the metabolic burden or elevated cytotoxicity due to the deletion of major metabolic genes [35,36]. Laboratory evolution can also help identify non-intuitive gene deletion candidates to develop host organisms such as a *cydC* mutant in *E. coli* that not only conferred increased tolerance to inhibitory residual reagents in the carbon source, but also restored production of the bio-jetfuel target to optimal levels [37].

Metabolizing different nutrition sources are desirable features of cell factories to efficiently utilize lignocellulosic biomass. Quantitative metabolomics analysis identified pyruvate kinase as a major bottleneck of arabinose metabolism suppression by glucose in *C. glutamicum*, which directed rational engineering of manipulating only two genes to achieve simultaneous utilization of D-glucose and L-arabinose at the same rate [38]. A systems biology analysis of *Rhodococcus opacus* PD630 provided insight into lignocellulose hydrolysate utilization and subsequent evolved strains showed improved growth rate on phenol [39]. Multi-omic analysis *P. putida cultured* with glucose and benzoate simultaneously provided evidence that metabolic segregation represented an efficient strategy in *P. putida* to meet biosynthetic flux [40**]. Similarly, this approach was applied to study the molecular mechanisms that influence and regulate lipogenesis in the oleaginous yeast *Rhodosporidium toruloides* [41] and *Yarrowia lipolytica* [42], and identify bottlenecks in the bioconversion of glycerol into acetol in *E. coli* [32]. The non-targeted ALE approach is also a useful tool to engineer carbon source assimilation pathways and bypass catabolite repression regulatory systems [43*].

There are now several examples for improving biomanufacturing research using systems and synthetic biology via semi- or fully-automated metabolic engineering Design-Build-Test-Learn (DBTL) pipelines. For instance, Carbonell *et al*. [44**] developed an automated pipeline that uses a

retrobiosynthesis tool, RetroPath [45], to propose pathway selection and derives common design rules from machine learning (ML) techniques. Another computational retrobiosynthesis tool, BNICE.ch, was used to assess more than 3.6 million biosynthetic pathways from central carbon metabolites of *E. coli* to methyl ethyl ketone [46]. A Method for Metabolite Annotation and Gene Integration (MAGI) [47] is another workflow that integrates metabolomics data with genomic predictions to improve biological interpretations made by both. Discovery of novel enzymes is enabled by global mutant fitness data generated from RB-Tn-Seq experiments [48]. Likewise, literature evidence of more than one DBTL cycle utilizing omics data to inform subsequent designs in the field of biomanufacturing research is still rare [44**,49*,50]. Opgenorth *et al*. [49*] used their first DBTL cycle data to train machine learning algorithms that suggested protein profiles for the second DBTL cycle. It led to a 21% increase in dodecanol titer in cycle two.

Computational approaches to systems biology data analysis

Integrative multi-omics analyses attempt to draw connections between disparate omics data sources, either with or without biological knowledge. Yet, extracting actionable information from omics datasets is quite challenging. Consequently, there has been significant effort directed into the development of computational tools that integrate different omics datasets acquired during the Test step of a typical engineering DBTL cycle to assist with the Learning step [51]. Several toolboxes exist to integrate transcriptomics, proteomics, fluxomics and metabolomics data. Tn-Core [52] is a recent example that can automate integration of Tn-Seq data in addition to RNA-seq data to generate context specific models that may provide a “systems-level” view for metabolic engineering purposes. Traditionally, statistical methods have been used to evaluate the significance of different design parameters (promoters and plasmid copy numbers) on product titers [44**]. Principal component analysis (PCA) was recently utilized to evaluate different engineered strains for metabolites that contribute to acetol production using metabolomics datasets [32] that helped identify NADPH regeneration as the bottleneck for efficient acetol biosynthesis. PCA has also been used on extracellular metabolomics datasets to identify metabolites that are the main drivers of variation in IPP production engineered strains [25*].

Machine learning approaches

Machine learning has the capability to extract patterns from high throughput biological data and shows great potential in systems metabolic engineering [53]. It has been used to propose pathway selection [44**,54] as well as build kinetic models using proteomic and metabolomic time-series data from isopentenol-producing and limonene-producing *E. coli* strains [55]. These models qualitatively predict pathway dynamics and outperformed a classical Michaelis-Menten kinetic model. Artificial Neural Networks (ANN) are another tool used for integration of omics datasets independently or as a component of a ML workflow. Recently, improvement in β-carotene production was assessed in *S. cerevisiae* through the use of an ANN model in conjunction with a YeastFab Assembly strategy (MiYA) [56]. Variable expression levels of three heterologous β-carotene genes and product titers were used to train an ANN model. This model was then used to predict the optimal level of expression for these enzymes in the next iteration of strain construction. After two cycles of ANN training and evaluation, it was found that 7 out of 10 top predicted engineered strains had improved product titers compared to the highest-producing strain of the initial library. This workflow was extended to improve violacein production (five enzyme pathway) that led to successful prediction of a strain that showed a 2.42-fold titer improvement in violacein production among 3125 possible promoter combinatorial designs. Going forward, ML algorithms need high-quality experimental data, larger training data sets, more pathway independent data, and more libraries to improve their prediction power.

Biology-informed approaches

Genome-scale metabolic models (GSM) and constraint-based modeling (CBM) methods [57*] assist in integrating disparate omics data onto the metabolic map of the host and improve phenotype predictions. Yet, constructing GSMs requires comprehensive networks of enzymatic reactions and metabolism of the host organism. Recently omics data has been integrated with a GSM for *E. coli* to identify key factors for bioproduction and prediction of cell factory performance [58*]. GSMs are also important inputs to computational strain optimization methods (CSOM) (*e.g.*, Optknock [59], constrained minimal cut sets (cMCS) [60]) that are used to identify genetic knockouts, over/under expression of genes for strain optimization efforts [2]. A multi-omics integrative analysis based on GSM of *A. niger* was performed to understand the mechanisms supporting a high yield of glucoamylase production and global metabolic regulation under oxygen limitation, an

industrially relevant scenario [61]. Recently a workflow to improve the accuracy of predictions of large scale kinetic models in metabolic engineering studies was proposed [62]. Likewise, a kinetic model of metabolism at a genome scale containing 457 model reactions, 337 metabolites and 295 substrate-level regulatory interactions was developed and applied to multiple mutant strains [63]. Omics data of optimally grown *E. coli* was integrated into a stoichiometric model and used to construct populations of non-linear large-scale kinetic models consistent with the physiology of the *E. coli* aerobic metabolism [62]. Genome-scale ^{13}C-MFA is computationally expensive method that requires knowledge of all of the carbon transitions in the network. Two-scale ^{13}C-MFA (2S-^{13}C-MFA) is an alternative that overcomes the problem associated with genome scale ^{13}C-MFA by constraining fluxes in the genome-scale model simultaneously using two resolution scales: for core reactions, both stoichiometric and ^{13}C labeling constraints are used, whereas, for non-core reactions, only stoichiometric constraints are used [64].

Concluding remarks

To date, most of the successes we see in the biomanufacturing space are the result of an Edisonian 'trial-and-error' approach which are rarely generalizable across products or microbial platforms. Current designs continue to rely on historical knowledge to design pathways and engineer microbial hosts in an artisanal manner. However, the studies outlined in this review indicate that the increasing availability of high-throughput synthetic biology methods, coupled with lowering costs of systems biology experiments to inform computational biology enable predictive approaches to aid our biofuel and bioproduct research and development. Further, there is an emerging need for open source (e.g. Python based) frameworks such as IMPACT (Integrated Microbial Physiology: Analysis, Characterization and Translation) for bioengineers working with big biological data to interpret, model, and visualize data [65]. Using such pipelines, metabolic engineers can opt for such scalable data integration and analysis workflows, alleviating a bottleneck limiting the throughput of bioengineering research. As these efforts mature, they can be extended to improve automation and real-time learning capabilities of global biofoundries [66,67] in academic, research and commercial institutions in the biomanufacturing research and application space.

Conflict of interest statement

Nothing declared.

Acknowledgements

This was funded as part of the Joint BioEnergy Institute (JBEI; http://www.jbei.org) and part of the Agile BioFoundry (ABF; http://agilebiofoundry.org) supported through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Figure and Table Captions

Table 1. Selected examples of tools to integrate systems biology data for biomanufacturing research

Figure 1. Overview of Systems, Synthetic and Computational biology methods that allow the use of microbial hosts for the production of biofuels and bioproducts. Several selected approaches in each category are shown.

* of special interest
** of outstanding interest

References

X1. Ko JK, Lee S-M: **Advances in cellulosic conversion to fuels: engineering yeasts for cellulosic bioethanol and biodiesel production.** *Curr. Opin. Biotechnol.* 2018, **50**:72–80.

2. Biz A, Proulx S, Xu Z, Siddartha K, Mulet Indrayanti A, Mahadevan R: **Systems biology based metabolic engineering for non-natural chemicals.** *Biotechnol. Adv.* 2019, doi:10.1016/j.biotechadv.2019.04.001.

3*. Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY: **Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering.** *Trends Biotechnol.* 2019, doi:10.1016/j.tibtech.2019.01.003.

An in-depth review that provides extensive background on systems biology and synthetic biology tools and strategies for systems metabolic engineering.

4. Wehrs M, Tanjore D, Eng T, Lievense J, Pray TR, Mukhopadhyay A: **Engineering Robust Production Microbes for Large-Scale Cultivation.** *Trends Microbiol.* 2019, **27**:524–537.

5. Wang G, Zhao Z, Ke J, Engel Y, Shi Y-M, Robinson D, Bingol K, Zhang Z, Bowen B, Louie K, et al.: **CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria.** *Nat. Microbiol.* 2019, doi:10.1038/s41564-019-0573-8.

6. Brophy JAN, Triassi AJ, Adams BL, Renberg RL, Stratis-Cullum DN, Grossman AD, Voigt CA: **Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria.** *Nat. Microbiol.* 2018, **3**:1043–1053.

7. Kushwaha M, Salis HM: **A portable expression resource for engineering cross-species genetic circuits and pathways.** *Nat. Commun.* 2015, **6**:7832.

8. Johns NI, Gomes ALC, Yim SS, Yang A, Blazejewski T, Smillie CS, Smith

MB, Alm EJ, Kosuri S, Wang HH: **Metagenomic mining of regulatory elements enables programmable species-selective gene expression.** *Nat. Methods* 2018, **15**:323–329.

9. Xia P-F, Ling H, Foo JL, Chang MW: **Synthetic genetic circuits for programmable biological functionalities.** *Biotechnol. Adv.* 2019, **37**:107393.
10. Chaves JE, Presley GN, Michener JK: **Modular engineering of biomass degradation pathways**. *Processes* 2019, **7**:230.
11. Liu H, Price MN, Waters RJ, Ray J, Carlson HK, Lamson JS, Chakraborty R, Arkin AP, Deutschbauer AM: **Magic pools: parallel assessment of transposon delivery vectors in bacteria.** *mSystems* 2018, **3**.
12**. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, et al.: **Mutant phenotypes for thousands of bacterial genes of unknown function.** *Nature* 2018, **557**:503–509.

Combining genome-scale mutagenesis and high-throughput mutant fitness profiling led to functional annotation of 11,779 protein-coding genes across 32 different bacteria. This valuable resource is accessible through an open source Fitness Browser platform.

13. Rajeev L, Luning EG, Dehal PS, Price MN, Arkin AP, Mukhopadhyay A: **Systematic mapping of two component response regulators to gene targets in a model sulfate reducing bacterium.** *Genome Biol.* 2011, **12**:R99.
14. Bartlett A, O'Malley RC, Huang S-SC, Galli M, Nery JR, Gallavotti A, Ecker JR: **Mapping genome-wide transcription-factor binding sites using DAP-seq.** *Nat. Protoc.* 2017, **12**:1659–1672.
15. Brown K, Takawira LT, O'Neill MM, Mizrachi E, Myburg AA, Hussey SG: **Identification and functional evaluation of accessible chromatin associated with wood formation in Eucalyptus grandis.** *New Phytol.* 2019, **223**:1937–1951.
16. Hussey SG, Grima-Pettenati J, Myburg AA, Mizrachi E, Brady SM, Yoshikuni Y, Deutsch S: **A Standardized Synthetic Eucalyptus Transcription Factor and Promoter Panel for Re-engineering Secondary Cell Wall Regulation in Biomass and Bioenergy Crops.** *ACS Synth. Biol.* 2019, **8**:463–465.
17. Vowinckel J, Zelezniak A, Bruderer R, Mülleder M, Reiter L, Ralser M: **Cost-effective generation of precise label-free quantitative proteomes in high-throughput by microLC and data-independent acquisition.** *Sci. Rep.* 2018, **8**:4346.

18. Chen Y, Vu J, Thompson MG, Sharpless WA, Chan LJG, Gin JW, Keasling JD, Adams PD, Petzold CJ: **A rapid methods development workflow for high-throughput quantitative proteomic applications.** *PLoS One* 2019, **14**:e0211582.
19. Chen Y, Guenther JM, Gin JW, Chan LJG, Costello Z, Ogorzalek TL, Tran HM, Blake-Hedges JM, Keasling JD, Adams PD, et al.: **Automated "Cells-To-Peptides" Sample Preparation Workflow for High-Throughput, Quantitative Proteomic Assays of Microbes.** *J. Proteome Res.* 2019, **18**:3752–3761.
20. Silva LP, Northen TR: **Exometabolomics and MSI: deconstructing how cells interact to transform their small molecule environment.** *Curr. Opin. Biotechnol.* 2015, **34**:209–216.
21. Pinu FR, Villas-Boas SG: **Extracellular microbial metabolomics: the state of the art.** *Metabolites* 2017, **7**.
22. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY: **Current status and applications of genome-scale metabolic models.** *Genome Biol.* 2019, **20**:121.
23. Yu T, Dabirian Y, Liu Q, Siewers V, Nielsen J: **Challenges and strategies for metabolic rewiring**. *Current Opinion in Systems Biology* 2019, **15**:30–38.
24. Yu R, Nielsen J: **Big data in yeast systems biology.** *FEMS Yeast Res* 2019, doi:10.1093/femsyr/foz070.
25*. George KW, Thompson MG, Kim J, Baidoo EEK, Wang G, Benites VT, Petzold CJ, Chan LJG, Yilmaz S, Turhanen P, et al.: **Integrated analysis of isopentenyl pyrophosphate (IPP) toxicity in isoprenoid-producing Escherichia coli.** *Metab. Eng.* 2018, **47**:60–72.
Multi-omic analyses of isoprenoid-producing *E. coli* revealed toxicity evidence and probable recovery mechanism for dealing with isopentenyl pyrophosphate (IPP) accumulation.
26. Poudel S, Giannone RJ, Rodriguez M, Raman B, Martin MZ, Engle NL, Mielenz JR, Nookaew I, Brown SD, Tschaplinski TJ, et al.: **Integrated omics analyses reveal the details of metabolic adaptation of Clostridium thermocellum to lignocellulose-derived growth inhibitors released during the deconstruction of switchgrass.** *Biotechnol Biofuels* 2017, **10**:14.
27. Ohtake T, Pontrelli S, Laviña WA, Liao JC, Putri SP, Fukusaki E: **Metabolomics-driven approach to solving a CoA imbalance for improved 1-butanol production in Escherichia coli.** *Metab. Eng.* 2017, **41**:135–143.

28. Budin I, de Rond T, Chen Y, Chan LJG, Petzold CJ, Keasling JD: **Viscous control of cellular respiration by membrane lipid composition.** *Science* 2018, **362**:1186–1189.
29. Denby CM, Li RA, Vu VT, Costello Z, Lin W, Chan LJG, Williams J, Donaldson B, Bamforth CW, Petzold CJ, et al.: **Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer.** *Nat. Commun.* 2018, **9**:965.
30. Sasaki Y, Eng T, Herbert RA, Trinh J, Chen Y, Rodriguez A, Gladden J, Simmons BA, Petzold CJ, Mukhopadhyay A: **Engineering Corynebacterium glutamicum to produce the biogasoline isopentenol from plant biomass hydrolysates.** *Biotechnol Biofuels* 2019, **12**:41.
31**. Lim HG, Kwak DH, Park S, Woo S, Yang J-S, Kang CW, Kim B, Noh MH, Seo SW, Jung GY: **Vibrio sp. dhg as a platform for the biorefinery of brown macroalgae.** *Nat. Commun.* 2019, **10**:2486.
The identification and characterization of a fast-growing and novel host *Vibrio* sp dhg as a biomanufacturing platform for efficient conversion of non-traditional carbon sources into value-added bioproducts.

32. Yao R, Li J, Feng L, Zhang X, Hu H: **13C metabolic flux analysis-guided metabolic engineering of Escherichia coli for improved acetol production from glycerol.** *Biotechnol Biofuels* 2019, **12**:29.
33. Lange A, Becker J, Schulze D, Cahoreau E, Portais J-C, Haefner S, Schröder H, Krawczyk J, Zelder O, Wittmann C: **Bio-based succinate from sucrose: High-resolution 13C metabolic flux analysis and metabolic engineering of the rumen bacterium Basfia succiniciproducens.** *Metab. Eng.* 2017, **44**:198–212.
34. Liu D, Wan N, Zhang F, Tang YJ, Wu SG: **Enhancing fatty acid production in Escherichia coli by Vitreoscilla hemoglobin overexpression.** *Biotechnol. Bioeng.* 2017, **114**:463–467.
35. McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO: **Adaptation to the coupling of glycolysis to toxic methylglyoxal production in tpiA deletion strains of Escherichia coli requires synchronized and counterintuitive genetic changes.** *Metab. Eng.* 2018, **48**:82–93.
36. McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO: **Multiple Optimal Phenotypes Overcome Redox and Glycolytic Intermediate Metabolite Imbalances in Escherichia coli pgi Knockout Evolutions.** *Appl. Environ. Microbiol.* 2018, **84**.
37. Eng T, Demling P, Herbert RA, Chen Y, Benites V, Martin J, Lipzen A,

Baidoo EEK, Blank LM, Petzold CJ, et al.: **Restoration of biofuel production levels and increased tolerance under ionic liquid stress is enabled by a mutation in the essential Escherichia coli gene cydC.** *Microb. Cell Fact.* 2018, **17**:159.

38. Kawaguchi H, Yoshihara K, Hara KY, Hasunuma T, Ogino C, Kondo A: **Metabolome analysis-based design and engineering of a metabolic pathway in Corynebacterium glutamicum to match rates of simultaneous utilization of D-glucose and L-arabinose.** *Microb. Cell Fact.* 2018, **17**:76.
39. Roell GW, Carr RR, Campbell T, Shang Z, Henson WR, Czajka JJ, Martín HG, Zhang F, Foston M, Dantas G, et al.: **A concerted systems biology analysis of phenol metabolism in Rhodococcus opacus PD630.** *Metab. Eng.* 2019, **55**:120–130.
40**. Kukurugya MA, Mendonca CM, Solhtalab M, Wilkes RA, Thannhauser TW, Aristilde L: **Multi-omics analysis unravels a segregated metabolic flux network that tunes co-utilization of sugar and aromatic carbons in Pseudomonas putida.** *J. Biol. Chem.* 2019, **294**:8464–8479.

Integrated omics analysis revealed metabolic architecture and highlighted regulatory points in *P. putida* that enable metabolic segregation of the substrate carbons to achieve optimally sustained biosynthetic flux demands and redox balance.

41. Wang Y, Zhang S, Zhu Z, Shen H, Lin X, Jin X, Jiao X, Zhao ZK: **Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast Rhodosporidium toruloides.** *Biotechnol Biofuels* 2018, **11**:148.
42. Kerkhoven EJ, Kim Y-M, Wei S, Nicora CD, Fillmore TL, Purvine SO, Webb-Robertson B-J, Smith RD, Baker SE, Metz TO, et al.: **Leucine Biosynthesis Is Involved in Regulating High Lipid Accumulation in Yarrowia lipolytica.** *MBio* 2017, **8**.
43*. Strucko T, Zirngibl K, Pereira F, Kafkia E, Mohamed ET, Rettel M, Stein F, Feist AM, Jouhten P, Patil KR, et al.: **Laboratory evolution reveals regulatory and metabolic trade-offs of glycerol utilization in Saccharomyces cerevisiae.** *Metab. Eng.* 2018, **47**:73–82.

A recent example of ALE based improvement of *S. cerevisiae* to utilize glycerol. They also re-engineered industrial strains to test the causal mutations responsible for improved growth on glycerol.

44**. Carbonell P, Jervis AJ, Robinson CJ, Yan C, Dunstan M, Swainston N, Vinaixa M, Hollywood KA, Currin A, Rattray NJW, et al.: **An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals.** *Commun. Biol.* 2018, **1**:66.
An example of an automated and iterative DBTL pipeline mainly focused at metabolic engineering from the synthetic biology perspective that is target agnostic and can be extended efficiently to the discovery and rapid optimization of high-performance pathways towards automated biomanufacturing.

45. Delépine B, Duigou T, Carbonell P, Faulon J-L: **RetroPath2.0: A retrosynthesis workflow for metabolic engineers.** *Metab. Eng.* 2018, **45**:158–170.
46. Tokic M, Hadadi N, Ataman M, Neves D, Ebert BE, Blank LM, Miskovic L, Hatzimanikatis V: **Discovery and evaluation of biosynthetic pathways for the production of five methyl ethyl ketone precursors.** *ACS Synth. Biol.* 2018, **7**:1858–1873.
47. Erbilgin O, Rübel O, Louie KB, Trinh M, Raad M de, Wildish T, Udwary D, Hoover C, Deutsch S, Northen TR, et al.: **MAGI: A method for metabolite annotation and gene integration.** *ACS Chem. Biol.* 2019, **14**:704–714.
48. Price MN, Zane GM, Kuehl JV, Melnyk RA, Wall JD, Deutschbauer AM, Arkin AP: **Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics.** *PLoS Genet.* 2018, **14**:e1007147.
49*. Opgenorth P, Costello Z, Okada T, Goyal G, Chen Y, Gin J, Benites V, de Raad M, Northen TR, Deng K, et al.: **Lessons from Two Design-Build-Test-Learn Cycles of Dodecanol Production in Escherichia coli Aided by Machine Learning.** *ACS Synth. Biol.* 2019, **8**:1337–1351.
An example of more than one DBTL cycle and utilization of proteomics, metabolomics data to train machine learning to inform subsequent designs for improved production titers.

50. Liu S, Xiao H, Zhang F, Lu Z, Zhang Y, Deng A, Li Z, Yang C, Wen T: **A seamless and iterative DNA assembly method named PS-Brick and its assisted metabolic engineering for threonine and 1-propanol production.** *Biotechnol Biofuels* 2019, **12**:180.
51. St John PC, Bomble YJ: **Approaches to computational strain design in the multiomics era.** *Front. Microbiol.* 2019, **10**:597.

52. diCenzo GC, Mengoni A, Fondi M: **Tn-Core: A Toolbox for Integrating Tn-seq Gene Essentiality Data and Constraint-Based Metabolic Modeling.** *ACS Synth. Biol.* 2019, **8**:158–169.
53. Presnell KV, Alper HS: **Systems Metabolic Engineering Meets Machine Learning: A New Era for Data-Driven Metabolic Engineering.** *Biotechnol. J.* 2019, **14**:e1800416.
54. Jervis AJ, Carbonell P, Vinaixa M, Dunstan MS, Hollywood KA, Robinson CJ, Rattray NJW, Yan C, Swainston N, Currin A, et al.: **Machine Learning of Designed Translational Control Allows Predictive Pathway Optimization in Escherichia coli.** *ACS Synth. Biol.* 2019, **8**:127–136.
55. Costello Z, Martin HG: **A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data.** *npj Syst. Biol. Appl.* 2018, **4**:19.
56. Zhou Y, Li G, Dong J, Xing X-H, Dai J, Zhang C: **MiYA, an efficient machine-learning workflow in conjunction with the YeastFab assembly strategy for combinatorial optimization of heterologous metabolic pathways in Saccharomyces cerevisiae.** *Metab. Eng.* 2018, **47**:294–302.
57*. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, et al.: **Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0.** *Nat. Protoc.* 2019, **14**:639–702.
A widely used open source computational tool to integrate heterogeneous omics data onto genome scale metabolic networks for metabolic engineering of a host.

58*. Oyetunde T, Liu D, Martin HG, Tang YJ: **Machine learning framework for assessment of microbial factory performance.** *PLoS One* 2019, **14**:e0210558.
A useful machine learning framework that integrates genome-scale metabolic modeling with 1200 experimental datasets of *E.coli* cell factories to provide additional features to train machine learning models to assess influential factors for bioproduction (TRY) and predict cell factory performance.

59. Pharkya P, Burgard AP, Maranas CD: **Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock.** *Biotechnol. Bioeng.* 2003, **84**:887–899.
60. Hädicke O, Klamt S: **Computing complex metabolic intervention**

strategies using constrained minimal cut sets. *Metab. Eng.* 2011, **13**:204–213.
61. Lu H, Cao W, Liu X, Sui Y, Ouyang L, Xia J, Huang M, Zhuang Y, Zhang S, Noorman H, et al.: **Multi-omics integrative analysis with genome-scale metabolic model simulation reveals global cellular adaptation of Aspergillus niger under industrial enzyme production condition.** *Sci. Rep.* 2018, **8**:14404.
62. Hameri T, Fengos G, Ataman M, Miskovic L, Hatzimanikatis V: **Kinetic models of metabolism that consider alternative steady-state solutions of intracellular fluxes and concentrations.** *Metab. Eng.* 2019, **52**:29–41.
63. Khodayari A, Maranas CD: **A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains.** *Nat. Commun.* 2016, **7**:13806.
64. Backman TWH, Ando D, Singh J, Keasling JD, García Martín H: **Constraining Genome-Scale Models to Represent the Bow Tie Structure of Metabolism for 13C Metabolic Flux Analysis.** *Metabolites* 2018, **8**.
65. Venayak N, Raj K, Mahadevan R: **Impact framework: A python package for writing data analysis workflows to interpret microbial physiology.** *Metab. Eng. Commun.* 2019, **9**:e00089.
66. Hillson N, Caddick M, Cai Y, Carrasco JA, Chang MW, Curach NC, Bell DJ, Le Feuvre R, Friedman DC, Fu X, et al.: **Building a global alliance of biofoundries.** *Nat. Commun.* 2019, **10**:2040.
67. Chao R, Mishra S, Si T, Zhao H: **Engineering biological systems using automated biofoundries.** *Metab. Eng.* 2017, **42**:98–108.

Table 1: Selected examples of tools to integrate systems biology data for biomanufacturing research					
Type of Data	**Analytical Technique(s)**	**Research Interest**	**Computational Tool(s) and Database(s)**	**Host**	**Reference**
Single-Omics					
Genome-scale phenotyping	RB-TnSeq, High-throughput phenotyping	Functional annotation of unknown genes in diverse bacteria	Pfam, TIGRFam, SEED/RAST and KEGG	32 different bacteria	[12**]
Genome-scale phenotyping	RB-TnSeq, High-throughput phenotyping	Gap Filling in Amino acid metabolic pathways	Pfam, TIGRFam, SEED/RAST and KEGG	10 different bacteria	[48]
Transcriptomics	RNA-Seq	Host engineering of a fast growing bacterium that utilizes non traditional carbon sources	Machine learning	*Vibrio sp. dhg*	[31**]
Proteomics	LC-MS/MS	Chassis-independent recombinase-assisted genome engineering	Diffusion-coupled reaction modeling	*E. coli*	[28]
Proteomics	LC-MS/MS	Tolerance and Metabolic engineering for improved production	-	*C. glutamicum*	[30]
Metabolomics	LC-MS/MS, GC-FID	Metabolic engineerin for improved production	Principal component analysis	*E. coli*	[27]
Metabolomics	LC-MS/MS,	Metabolic engineering for improved production	openBIS software	*E. coli*	[44**]
Metabolomics	LC-MS/MS	Identifying bottlenecks limiting carbon source utilization	-	*C. glutamicum*	[38]
Fluxomics	13C-MFA, GC-MS	Metabolic engineering for improved production	Constraint-based modeling and Metabolic flux analysis using OPENFLUX	*B. succiniciproducens*	[33]
Phenomics	LC-UV/VIS/DAD/RID	Optimization of heterologous pathway	Machine learning using artificial neural network	*S. cerevesiae*	[55]
Multi-Omics					
Genomics + Proteomics	NGS, LC-MS/MS	Toxicity Tolerance and Metabolic engineering for improved production	EggNOG database	*E. coli*	[37]
Genomics + Metabolomics	NGS, LC-MS/MS	Metabolite identification, gene annotation and biochemical predictions for poorly annotated genes	Metabolite Annotation and Gene Integration (MAGI) database	*S. coelicolor*	[47]
Transcriptomics + Metabolomics	LC-UV, RT-qPCR, GC-MS	Metabolic engineering for improved production	-	*E. coli*	[50]
Proteomics + Metabolomics	LC-MS/MS, GC-MS, NIMS, MSI	Metabolic engineering for improved production	Machine learning, OpenMSI Arrayed Analysis Toolkit	*E. coli*	[49**]
Proteomics + Metabolomics	LC-MS/MS, LC-RID, GC-MS	Metabolic engineering for improved hoppy flavored beer	Machine learning	*S. cerevesiae*	[29]
Fluxomics + Metabolomics +Transcriptomics	LC-MS/MS, LC-RID, 13C-MFA, RT-qPCR	Metabolic engineering for improved production	Principal component analysis, Constraint-based modeling and Metabolic flux analysis	*E. coli*	[32]
Trancriptomics + Proteomics + Metabolomics	LC-MS/MS, DNA Microarray, GC-MS	Host response to toxicity during deconstruction and utilization of bioenergy relevant feedstock	Functional enrichment using PIANO package and Cytoscape	*C. thermocellum*	[27]
Proteomics + Metabolomics + Fluxomics	LC-MS/MS, 13C-MFA	Metabolic segregation during co-utilization of different carbon sources	Constraint-based modeling and Metabolic flux analysis using 13CFLUX2	*P. putida*	[40**]
Transcriptomics + Proteomics + Metabolomics	LC-MS/MS, RNA-Seq	Toxicity tolerance	Principal component analysis, Functional enrichment analysis using clusterProfiler and Cytoscape	*E. coli*	[25*]
Transcriptomics + Metabolomics + Phenomics	LC-MS/MS, GC-MS, RNA-Seq	Global Metabolic Regulation under industrially relevant condition for enzyme production	Principal component analysis, Functional enrichment, Constraint-based modeling using COBRA Toolbox	*A. niger*	[60]
Transcriptomics + Proteomics + Metabolomics	LC-MS/MS, RNA-Seq, RT-qPCR	Effect of phosphate limitation on lipogenesis	Differential gene expression analysis, Enrichment analysis using WEGO software, Principal component analysis	*R. toruloides*	[41]
Transcriptomics + Proteomics + Metabolomics	LC-MS/MS, RNA-Seq, GC-MS	Effect of carbon or nitrogen limitation on lipogenesis	Gene set analysis using PIANO package, Constraint-based modeling	*Y. lipolytica*	[42]
Genomics + Transcriptomics + Proteomics + Metabolomics	NGS, LC-MS/MS, RNA-Seq, GC-MS	Metabolic trade-offs of glycerol utilization	Differential expression analysis using DEseq package, HTSeq python library, Constraint-based modeling	*S. cerevesiae*	[43*]
Genomics + Transcriptomics + Metabolomics + Fluxomics + Phenomics	NGS, LC-MS/MS,13C-MFA, RNA-Seq	Mechanism to overcome cytotoxicity due to loss of a major metabolic gene	Constraint-based modeling and Metabolic flux analysis using INCA	*E. coli*	[35]
Genomics + Transcriptomics + Metabolomics + Fluxomics + Phenomics	NGS, LC-MS/MS,13C-MFA, RNA-Seq	Mechanism to overcome metabolic imbalance due to loss of a major metabolic gene	Constraint-based modeling and Metabolic flux analysis using INCA, Biopython	*E. coli*	[36]

652