Review

MOLECULAR METABOLISM

Toward modeling metabolic state from single-cell transcriptomics

Karin Hrovatin [1,2], David S. Fischer [1,2], Fabian J. Theis [1,2,3,*]

ABSTRACT

Background: Single-cell metabolic studies bring new insights into cellular function, which can often not be captured on other omics layers. Metabolic information has wide applicability, such as for the study of cellular heterogeneity or for the understanding of drug mechanisms and biomarker development. However, metabolic measurements on single-cell level are limited by insufficient scalability and sensitivity, as well as resource intensiveness, and are currently not possible in parallel with measuring transcript state, commonly used to identify cell types. Nevertheless, because omics layers are strongly intertwined, it is possible to make metabolic predictions based on measured data of more easily measurable omics layers together with prior metabolic network knowledge.

Scope of Review: We summarize the current state of single-cell metabolic measurement and modeling approaches, motivating the use of computational techniques. We review three main classes of computational methods used for prediction of single-cell metabolism: pathway-level analysis, constraint-based modeling, and kinetic modeling. We describe the unique challenges arising when transitioning from bulk to single-cell modeling. Finally, we propose potential model extensions and computational methods that could be leveraged to achieve these goals.

Major Conclusions: Single-cell metabolic modeling is a rising field that provides a new perspective for understanding cellular functions. The presented modeling approaches vary in terms of input requirements and assumptions, scalability, modeled metabolic layers, and newly gained insights. We believe that the use of prior metabolic knowledge will lead to more robust predictions and will pave the way for mechanistic and interpretable machine-learning models.

Keywords Metabolic modeling; Single-cell RNA-seq; Constraint-based modeling; Kinetic modeling; Pathway analysis

1. INTRODUCTION

Multimodal single-cell analysis enables robust studying of single cells across different omics levels that characterize specific cellular processes: DNA sequencing can reveal lineage structure, chromatin accessibility measurements are used to study epigenomic events, and RNA sequencing and protein measurements uncover molecular states of cells that arise during stimulation and development. All of these molecular views of cells are directly observable in high throughput by using specific chemical procedures; therefore, it is not surprising that much of the recent advances in single-cell biology have been centered on these omics layers [1,2]. We want to focus on a less studied area, observing single-cell metabolism. It consists of metabolic reactions necessary for production and degradation of cellular building components and provision of energy [3]. Metabolism lies downstream of other omics and thus summarizes the effects from upstream layers of epigenome, genome, and transcriptome, as well as extracellular effects, such as nutrient depletion. Thus, it provides a very meaningful readout of the cellular phenotype [4,5] and has been shown to be closely related to disease states or used for developing efficient biomarkers in bulk assays [6–8].

Single-cell metabolic analysis largely consists of the study of metabolite concentrations in single cells or compartments thereof (metabolome), the study of metabolite fluxes (fluxome), which describe the rates of change of metabolite concentrations, and fluxes through individual reactions governed by enzyme activity. All of them are relevant for modern single-cell biology, which is centered around cell type diversity [9] and cell state transitions [10]. First, metabolic differences may identify cell state heterogeneity, which is key for understanding tissue function and microbial populations [11–16] as metabolic changes may not be observed in other omics layers. Second, metabolic measurements can be used to decipher metabolic mechanisms of individual cell types, refining the currently available organism-wide metabolic networks. This may be applied to understanding drug target- and side-effects across cell populations, similarly as was done before on bulk-tissue level [17,18], or understanding the emergence of

[1]Institute of Computational Biology, Helmholtz Center Munich, Ingolstaedter Landstraße 1, Neuherberg, 85764, Germany [2]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Alte Akademie 8, Freising, 85354, Germany [3]Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, Garching bei München, 85748, Germany

*Corresponding author. Institute of Computational Biology, Helmholtz Center Munich, Ingolstaedter Landstraße 1, Neuherberg, 85764, Germany.

E-mails: karin.hrovatin@helmholtz-muenchen.de (K. Hrovatin), david.fischer@helmholtz-muenchen.de (D.S. Fischer), fabian.theis@helmholtz-muenchen.de (F.J. Theis).

Abbreviations: GEM, genome-scale metabolic model; FBA, flux balance analysis; DNN, deep neural network; scRNA-seq, single-cell RNA-sequencing

Received August 5, 2021 • Revision received October 21, 2021 • Accepted November 9, 2021 • Available online 14 November 2021

https://doi.org/10.1016/j.molmet.2021.101396

drug-resistance in individual cells [19]. Third, metabolite exchange is a key mechanism of cell—cell dependence in tissues [20] and interactions between neighboring cells play a vital role in tissue formation and function [21].

Recent advances in metabolomics techniques have enabled measurements on the single-cell level, as had been recently reviewed [4,22], also with focus on experimental applications [3], spatial metabolomics and proteomics [23], and multi-omics analysis [24]. However, single-cell metabolomics techniques are still limited, requiring further improvements before becoming widely applicable. Metabolomic measurements are expensive, skill intensive, and need additional improvements in terms of measurement sensitivity and throughput of samples and metabolites [4,20,22]. In comparison, the widely-used single-cell RNA-sequencing (scRNA-seq) datasets usually consist of thousands of cells and genes [2], while the metabolomic datasets usually consist of a couple of dozen to a couple of hundred metabolites and cells [3,4,22,25,26]. Thus, they cover only a small proportion of cellular metabolism, may miss rare cell types, and suffer from lower robustness due to the small number of cells, each of which provides a noisy view of the metabolism. Lack of established techniques for simultaneous measurement of metabolome and transcriptome may complicate contextualization in terms of previous transcriptomic studies and thwart cell type annotation, which is usually done on transcriptomic data. Nevertheless, methodological improvements are forthcoming [3,4,22,25,26].

Because omics layers are strongly intertwined, the knowledge about omics layers that can be more easily measured in singles cells, such as mRNA, together with prior knowledge about metabolism, enable the prediction of cellular fluxome, metabolome, and reaction fluxes, providing a promising alternative to single-cell metabolomic measurements. This can be used to identify changes in metabolic mechanisms that cause different metabolic states. The modeling of metabolism depends on well-established metabolic network models that provide prior information on feasible metabolic conversions [27], described in section "Resources for metabolic modeling", and wealth of single-cell sequencing measurements [28—30] used to define constraints specific for the biological context [31]. Concept-wise, modeling single-cell metabolism rather than bulk modeling (Figure 1) poses additional challenges that have led to the development of new metabolic modeling paradigms. First, models need to scale to a large number of cells. Second, the joint modeling of multiple similar or potentially nearby and communicating cells enables sharing of parameters across cells. This may increase robustness, which is important when using noisy scRNA-seq data; however, cells are known to have different metabolic functionality [3], which must be accounted for in the modelling assumptions. Third, metabolite exchange plays an important role within cell populations [32], which results in metabolite availability dependencies between cells and violates common independence assumptions used in other branches of single-cell modelling. We review and compare approaches for computational single-cell modeling of cellular metabolism based predominantly on transcriptomic data.

2. RESOURCES FOR METABOLIC MODELING

Prior knowledge for metabolic modeling includes genome-scale metabolic models (GEM) and metabolic pathway collections, which are available in large, curated databases. Metabolic pathway collections summarize reactions into pathways [33]. Genome-scale metabolic models provide systematic encoding of gene-protein-reaction-metabolite and gene-protein-transport-metabolite associations that describes the entire metabolism of an organism [27,34—37]. While pathway databases are mainly used for interpretation, the GEMs allow the prediction of metabolism on a system-wide level. However, both usually encompass the whole spectrum of possible metabolic conversions [27,38—40] that do not jointly occur in individual cells.

The most widely used pathway repositories are KEGG [41], Reactome [42], and MetaCyc [43], as reviewed in [33]. Genome-scale metabolic models can be obtained from BiGG [44], BioModels, and other repositories listed in [27]. The modelling of human metabolism often relies on GEMs [27,45—47] from Recon [37] and Human Metabolic Reaction series [48] and the resources are being constantly improved [49]. GEMs can be also constructed de novo from KEGG or other resources and are often updated with manual curation [20,46,50]. To enable comparison and knowledge integration across resources, standardization efforts are needed [44,51]. To leverage the available resources the metabolic modelling tools should be able to accept as input different models in standard format, such as MAT (mat) or SBML (xml) for GEMs.

Measured metabolite concentrations from specific reference tissues may be used as validation data for computational modeling. Such datasets can be obtained from specialized databases, including MetaboLights [52] and Metabolomics Workbench [53]. Because of potential difficulties of measuring cell level heterogeneity of metabolism, another validation option are also bulk expression or metabolic datasets from systems with environmental or genetic perturbations with known effects on metabolism.

3. MODELLING APPROACHES

We identified three main classes of methods used for prediction of metabolism on the single-cell level (Figure 2): (A) pathway-level analysis, which derives pathway phenotype-associations or activity from gene expression information of associated genes and in some

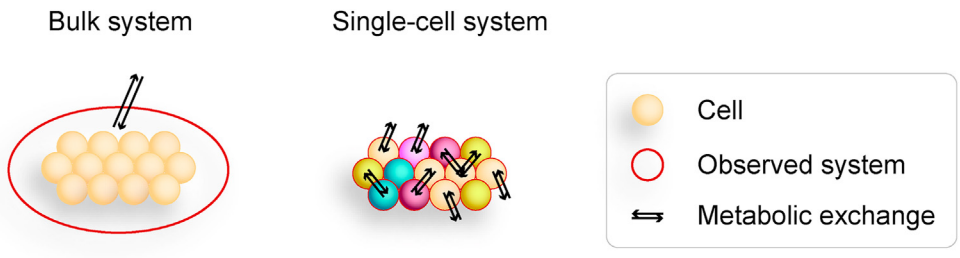

Figure 1: Bulk versus single-cell analysis. Bulk analysis assumes that cells are identical and can model the exchange between cells and environment. Single-cell analysis accounts for cellular heterogeneity and can model exchange between different cells and between individual cells and environment.

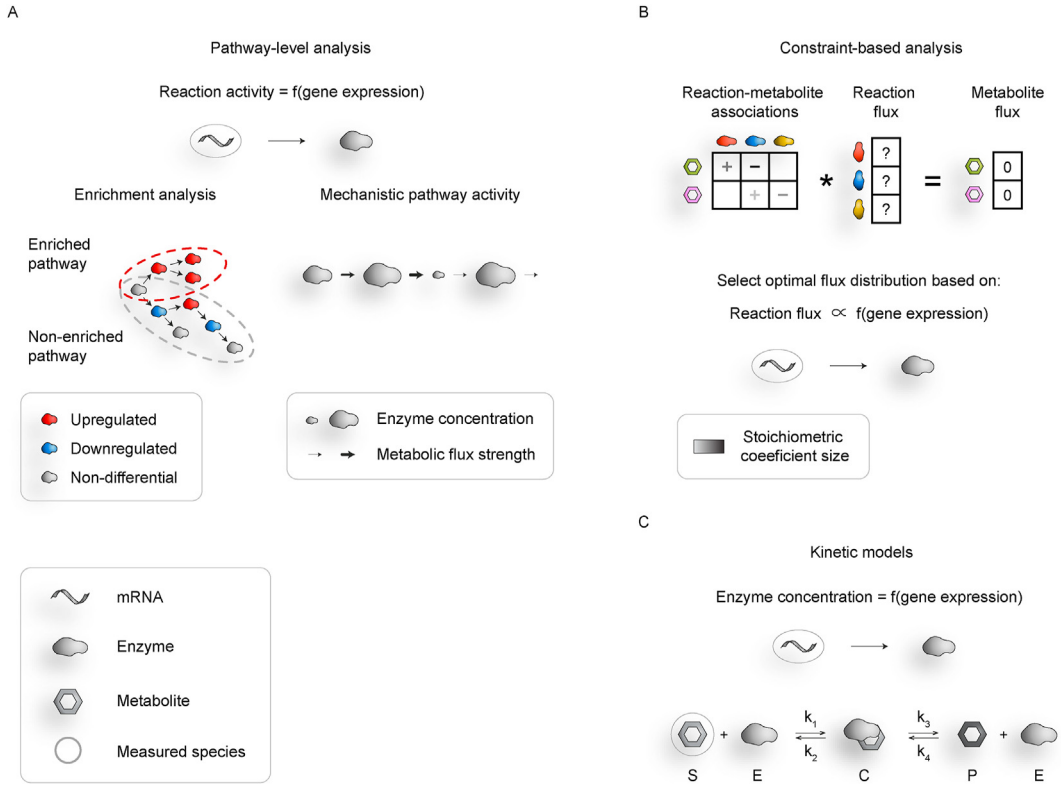

Figure 2: Metabolic modelling approaches. (A) Pathway-level analysis assumes that gene expression maps directly to enzyme concentration or reaction activity. It can be used for defining pathways enriched in differentially expressed genes or for calculating the activity of individual pathways by propagating activity through the pathway (mechanistic pathway analysis). (B) Constraint-based analysis assumes that metabolite concentrations are constant and based on this defines possible fluxes that would result in such a system, as further described in the main text. Namely, metabolite fluxes, which are constrained to be zero, are obtained by multiplying reaction fluxes with the stoichiometric matrix that defines metabolite conversions involved in each reaction. This gives an undetermined system of equations that defines possible reaction flux distributions. Optimal flux configuration is then selected based on optimization objectives, such as correspondence between gene expression and enzyme activity. (C) Kinetic models predict metabolic changes based on detailed prior metabolic knowledge and information on metabolite and enzyme availability. Gene expression can be used for the prediction of enzyme concentration.

cases includes pathway topology, (B) constraint-based modelling that aims to predict complete cellular metabolome and reaction activities based on gene expression and reaction stoichiometry information from GEMs, and (C) kinetic models, which use highly parameterized systems of differential equations informed with measured data to predict metabolite concentrations of well characterized systems.

3.1. Pathway-level analysis

Pathway-level analysis (Figure 2A) predicts pathway activity or phenotype-associations directly from differential expression or occasionally normalized expression information of pathway-associated metabolic genes, sometimes including pathway topology. It is focused on individual pathways rather than the whole metabolism, and it does not account for reaction mechanisms, such as stoichiometry or metabolite abundance. There are many available tools for pathway analysis, as the methods designed for bulk can be in most cases directly applied to single-cell data. To rely on bulk data workflows and to reduce the sparsity the analysis is often performed on pseudobulk data [54]. In this review we focus on approaches that were developed for scRNA-seq data.

Currently, differential expression analysis followed by pathway enrichment is the most commonly used method for metabolic analysis based on scRNA-seq data [20,55], and uses general purpose enrichment methods [56–58]. Alternatively, pathway activity can be inferred from gene expression of pathway-associated genes [14,59,60].

Pathway-level analysis, as described above, has a few shortcomings. First, the curated pathway definitions aim to capture the whole biological process under different biological contexts, disregarding complex regulatory interplay, and thus contain more genes than are involved in any specific cellular state or type [39,40]. Predefined pathways may be composed of genes responding to perturbation in different directions, leading to false negatives in enrichment analysis, which could be solved by defining context-specific pathways [45]. Second, pathway enrichment does not account for biological changes spanning multiple pathways [40] and does not leverage the information about regulatory and reaction interactions. For example, the activity of the whole pathway depends on the activity of all enzymes in the chain with downregulation of intermediate enzymes potentially resulting in metabolic bottlenecks that lead to lower pathway activity even when other enzymes are highly active. Methods that incorporate pathway topology information or enable data-driven pathway discovery had been developed [61–66]. For example, to pinpoint critical genes and metabolites scMetNet [39] constructs a metabolic network based on KEGG and then finds submodules with significant differential expression across cell populations. Similarly, scPPIN was developed for identification of differentially active modules in protein–protein interaction networks based on scRNA-seq data [67], but may be also applicable to metabolic networks. Another alternative is reporter metabolite analysis, such as perturb-Met, which aims to identify

affected metabolites by analyzing changes in expression of metabolite-associated genes [40].

3.2. Constraint-based modelling

3.2.1. Background

Steady-state constraint-based methods are the most widely used group of metabolic modeling approaches for large-scale systems [68–70]. They assume, due to the separation of omics layers time scales, that metabolism is in a quasi-steady-state versus expression and other longer-term processes, and then search across reaction flux configurations that all yield this steady-state system, selecting one that maximizes some objective function based on prior biological knowledge about the system (Figure 2B). The quasi steady-state assumption is supported by the following reasoning: The traditional view on cellular metabolic systems is that metabolites are in mass balance, which means that metabolites are converted into each other such that all molecular mass is accounted for. To maintain this system, cells require external sources of energy and starting products and disposition mechanisms for end products. Because metabolic reactions are very fast, occurring in milliseconds, compared to enzyme gene expression, operating on the scale of tens of minutes [71], the reactions reach equilibrium relatively quickly, leading to metabolic steady-state in most situations.

Constraint-based methods rely on system-wide information about stoichiometry, thermodynamics, and gene-associations of reactions, which are provided by GEMs [27]. Among the most widely used constraint-based metabolic modeling methods is flux balance analysis (FBA), in which the metabolic network is encoded by stoichiometric matrix (S) of metabolites (rows) and reactions (columns). The reaction fluxes (v) are obtained by assuming a steady-state of the system where metabolite concentrations (dx/dt) do not change, namely:

$$Sv = dx/dt = 0$$

As the system of $Sv = 0$ contains fewer equations (number of metabolites, over 8,000 in human [49]) than variables (number of reaction fluxes, with over 13,000 reactions in human [49]) this approach results in an underdetermined system with a whole subspace of possible flux distributions. In many cases, a single solution can be recovered with linear programming based on an optimization objective, which relies on prior knowledge about the system [68,69,72,73].

The modeling objectives, used for selection of a single reaction flux configuration, fall into two categories. On the one hand, some objectives are based on general assumptions about cellular purpose, such as maximal growth rate, ATP production, or combination thereof that may better correspond to the cell behavior [38,46,68]. On the other hand, some models rely on system-specific data, such as RNA-sequencing, partial metabolomics measurements, or information from literature. Measurement-based objectives are of great use when theory-based biological objective is unknown or simple objectives, such as growth maximization, are not applicable due to the complexity of the system [45,46,69,72,74]. Integration of measured data, such as from scRNA-seq, enables context-specific predictions, such as for specific cell type, developmental stage, or patient [27,38]. These constraints can be accounted for in linear programming in different ways, including by adjusting GEMs or reaction rate boundaries, by enforcing fluxes to follow omics measurements, or by redistributing fluxes from reactions with low to reactions with high omics-activity support [50,69,72].

3.2.2. Constraint-based modeling applied to scRNA-seq data

There are currently two main approaches for scRNA-seq constraint-based metabolic modeling, namely FBA and GEM-informed deep neural network (DNN) models. Multiple methods have been proposed to adapt FBA to single-cell inference. The simplest approach is the construction of context-specific GEMs based on scRNA-seq data, followed by classical FBA on each context-specific GEM, as proposed by [75], using specific metabolite production as the optimization objective. Another option is MERGE, which optimizes for reactions with lowly expressed genes carrying less flux and vice versa as well as for low overall flux [50]. Moreover, Compass [45] first computes maximal theoretically possible flux through each reaction and then selects flux configuration that reduces reaction penalties inversely proportional to gene expression. scFBA computes maximal theoretically possible fluxes, uses scRNA-seq data to add an additional constraint on possible flux distributions, and selects optimal flux distribution based on maximal biomass production [46].

A complementary approach was implemented in scFEA [20], which models the metabolic map with a graph neural network. The GEM is encoded into a directed factor graph with reaction modules as variables (unknown quantity or random variable), intermediate metabolites as factors (function of variables), and directed edges indicating substrate and product relationships. scFEA trains the network to learn reaction fluxes from scRNA-seq input so that the metabolite flux imbalance is minimized. Additional loss components are added to discourage negative reaction fluxes and to promote correspondence between reaction fluxes and gene expression in order to evade the trivial solution of zero-reaction-fluxes. The interpretable DNN structure and predictions can be directly used for reasoning about cellular metabolism. Higher weights of a gene indicate stronger impact of gene expression variation onto predicted reaction flux, corresponding to rate limiting reactions. Furthermore, pathways with high metabolite flux imbalance may indicate metabolic stress.

The presented models make different assumptions about cellular metabolism. For example, some methods use RNA-seq data as a constraint and optimize for maximization of cellular growth, while others optimize solely based on the agreement with scRNA-seq data. In FBA-based methods the metabolic steady-state assumption is enforced., while in scFEA it is used as a component of the optimization objective. Furthermore, scFEA and scFBA assume that cells exchange metabolites, sitting in a well-mixed environment, and thus solve the metabolic model jointly for all cells [20,46].

One of the key challenges is the mapping from gene expression to enzyme activity, which is handled differently across the methods. First, the mapping should ideally be gene-specific due to different gene translation and enzyme kinetic rates [46] and cell state-specific due to the influence of the intracellular environment on enzyme kinetics [20]. Both these challenges may be resolved by DNN models that can learn reaction-specific mapping, as done in scFEA, and by integrating additional cell state specific prior knowledge [20]. Second, the mapping should be modeled as a nonlinear function in accordance with Michaelis–Menten kinetics, which was approximated with multiple neural networks layers in scFEA [20]. The mapping is complicated by enzyme isoforms (OR relationship) and complexes (AND relationship) [45,46]. Third, gene expression and enzyme activity may not correspond both due to technical artefacts and differences in half-lives across omics-layers [76]. Thus, the minimization of the flux through reactions with lowly-expressed genes is less prone to false negatives than pruning of the GEM based on a fixed expression threshold. One should not penalize highly expressed genes if they have low flux in reactions with low substrate availability [50].

An important aspect of single-cell metabolic modeling in contrast to bulk is to overcome noise and sparsity of the scRNA-seq input, which can be achieved in multiple ways. The most common approach is to share information across cells, either by using pseudobulk, by sharing flux distribution information across similar cells, or by jointly modelling metabolism of all cells [20,45,75]. Data can be also pooled across genes by modeling pooled reaction modules rather than individual reactions or by assuming that neighboring genes affect each other and thus share information on expression variation among them. The definition of unexpressed genes used for GEM pruning is affected by false positives due to dropout in scRNA-seq data. This can be mitigated using statistical methods or matched bulk expression data [20,46].

Due to the growing dataset sizes [28–30], the methods must be compute-efficient. Two core paradigms of computational efficiency appear. First, the number of metabolic states may be reduced from the full number of metabolites and compartments to a more coarse metabolic network by reaction pooling, removing reactions predicted to be inactive based on expression data, or modelling only a reaction subset, such as core metabolism [20,45,46]. Second, inference procedures have to scale to many observations as there are often thousands or millions of cells, which lies many orders of magnitude above what is typically considered in bulk. A common sped-up technique is the pooling and per-group-modelling of similar cells [20,46,50,75].

Overall, the field of single-cell constraint-based modeling is quickly expanding and systematic benchmarks will be needed to recommend the best modelling choices. Among the FBA-based methods, Compass stands out due to its modelling choices and user friendliness. DNN methods promise greater model flexibility and desirable scaling properties, therefore, scFEA represents an interesting alternative.

3.3. Kinetic models

Kinetic models have different applications. First, kinetic models can be used for quantitative instead of relative predictions of the effects of environmental and parameter value perturbations. They solve a set of differential equations with detailed kinetic and regulatory information starting from a predefined initial metabolic state, characterized by defined enzyme and metabolite concentrations (Figure 2C). This stands in contrast to FBA for example, which does not require any metabolite concentration input. However, concentration and kinetic parameter information is scarce, which makes kinetic models largely inapplicable for modeling of larger systems [68,74,77,78]. This is further complicated in single-cell systems as kinetic parameters can differ between cells due to different intracellular environments, such as pH and cofactor availability [79,80]. Second, model selection and covariate modelling on top of kinetic models can be used to understand metabolic mechanisms that lead to a certain metabolic state [68]. For example, genetic variations that cause subtle changes in enzyme activity can be accounted for in such kinetic models but go undetected in constraint-based models [81]. Third, the modeled time-resolved metabolite concentrations in response to stimuli are of interest in many cases. In contrast to temporal insight that can be obtained from single-cell time-series measurements, that are usually spaced hours or days, kinetic modeling enables prediction on a much quicker timescale of metabolism, starting from an initial measured metabolic state. Furthermore, kinetic models solve differential equation systems in time and therefore do not rely on the steady-state assumption. Thus, they present an alternative to constraint-based modeling when steady-state assumption does not hold [68,82], as described in [83].

A kinetic model that can leverage scRNA-seq data for estimation of enzyme concentrations was proposed by [78]. Their method enables efficient prediction of metabolic changes based on changes in kinetic parameters. They model the product concentration using gaussian mixture models. The model consists of reaction catalysis terms, which are conditioned on enzyme concentration that can be informed with single-cell measurements. The combination of kinetic models and measured data can also be used for inference of sample-specific kinetic parameters that determine reaction dynamics [81]. Such sample-specific kinetic parameters can be used to understand kinetic heterogeneity across tissues, thus extending efforts to characterize tissue heterogeneity through single-cell phenotypes in cellular atlases. An example for such a tissue-level phenotype is the exchange of metabolites between cells that mitigates the diseased phenotype of affected cells in X-inactivation-linked enzyme heterozygosity in females [84]. A kinetic model of RNA splicing has recently been directly fit on RNA-seq data in the context of RNA velocities [85]. This RNA velocity model infers the kinetic parameters of transcription, splicing, and degradation from measured un-spliced and spliced read abundances of single cells in different gene expression states.

3.4. Discussion

3.4.1. Method comparison

The three classes of metabolic modeling approaches aim to answer different questions and thus use different techniques (Table 1). Their predictions range in size, from a single reaction to the whole metabolism, and in coarseness, from direct mapping between measured gene expression and predicted enzyme activity to detailed metabolic characterization of the whole system. Kinetic modelling is focused on robust characterization of small systems, enabling mechanistic, quantitative and often dynamic predictions. In contrast, pathway-level analysis compares activities of multiple pathways across conditions. It does not account for reaction mechanisms, such as stoichiometry or metabolite abundance, and is thus often regarded as less robust. The analysis of reaction networks can also serve as an interpretation method for results from other metabolic modeling methods, such as constraint-based models. Constraint-based models provide characterization of complete cellular metabolism based on GEMs.

The methods also differ in the resource requirements (Figure 3) and assumptions. In terms of increasing complexity, the approaches are usually viewed in the following order: pathway-level analysis, followed by constraint-based, and lastly kinetic modelling. First, pathway level analysis relies on curated pathway-gene association sets or networks. Thus, it is often applicable to less characterized organisms, where complete GEMs or kinetic parameters are not available, and might also require less compute-resources. Pathway-analysis methods developed for bulk level can be often applied to single-cell data, resulting in a large set of available methods. Second, constraint-based modelling requires GEMs, whose number is however growing [27]. We expect that constraint-based deep learning models will, in the future, enable scaling to large datasets, integration of additional metabolic and multi-omics constraints, and extension to dynamic or tissue-level models. Constraint-based models rely on the steady-state assumption, making them inappropriate when this assumption does not hold. Third, kinetic-modeling requires detailed kinetic information, making them inappropriate for large systems. Methods for constraint-based and kinetic modeling can usually not be directly transferred from bulk to single-cell level, requiring the development of new approaches or adaptation of existing ones.

Table 1 — Insights gained from different metabolic modelling methods. The methods differ in modeled metabolic layers and scales, cellular resolutions, modelling assumptions, and results interpretability. The differences in method inputs are described in Figure 3. The here summarized assumptions are general ones and methods may have additional assumptions.

	Gained insight	Applicable to whole metabolism	Applicable to individual pathways or reactions	Metabolic layer	Assumptions	Results require post-processing summarization	Applied to individual cells
Pathway-level analysis	List of enriched pathways or data-defined pathways; pathway activity	Yes	Yes	Reactions, rarely metabolites	Often disregards reaction associations within pathways or assumes that pathways are independent; sometimes approximately accounts for mass balance	Usually not as pathway-level information is already available	Usually not
Constraint-based analysis	Whole-system metabolic predictions of reaction fluxes and sometimes metabolite concentrations	Yes, but often simplified to reduce the necessary compute resources	No	Reactions and metabolites	Metabolite steady-state and mass balance; may assume availability of all substrates	Yes; in exploratory analysis as results are on the level of reactions or metabolites and thus pathway-level analysis is often applied	Yes, but often pooled to reduce the necessary compute resources
Kinetic models	Metabolite and reaction fluxes predictions through time or under different parameter setting; model selection	Usually not	Yes	Reactions and metabolites	Mass balance and Michaelis—Menten kinetics; may assume constant availability of all substrates	No, due to small systems	Yes

3.4.2. Pros and cons of metabolic modeling

The main reason for the popularity of metabolic modeling is its greater accessibility compared to metabolic measurements as well as the clear availability of biochemical knowledge versus, for example, gene regulation models. For example, single-cell metabolomics are focused on measuring metabolite concentrations with large-scale fluxome measurements being mainly out of reach [46,86,87]. However, computational methods enable prediction of metabolic fluxes, which, in contrast to metabolite concentrations information, enable identification of cellular mechanisms that lead to specific metabolomic phenotypes [87]. With increasing availability of single-cell datasets the scientific community has realized the importance of cross-study analysis, thus shifting from analysis of individual datasets to analyses of integrated cellular atlases [30,88,89]. However, metabolomic measurements are still poorly comparable across studies due to the lack of firmly-established analysis methods [22]. In contrast, metabolic modelling mainly relies on scRNA-seq data, for which normalization and integration techniques are well-established [88,90], increasing the comparability across studies. Generation of new single-cell metabolomic datasets will lead to the development of analysis techniques, similarly as had happened in bulk metabolomics [91].

The integration of prior knowledge from curated metabolic resources used in computational modeling can bring new insights into metabolic mechanisms and causal relationships, compared to raw metabolic measurements, as described below [20,27,35,38,65,77,92—94]. First, metabolic modeling can provide mechanistic insights into perturbation effects. In terms of single-cell applications this can be used for understanding of molecular mechanisms that lead to disease development. This information can be used for the design of drugs targeting specific metabolic steps. Moreover, the understanding of single-cell metabolism across cell types and tissues can be useful for prediction of drug side effects. Metabolic modeling can be also used to identify genes that can be targeted by metabolic engineering for increased production. Second, metabolic modelling can be used for prediction of interactions between cells, organisms, and the environment. This includes both availability and exchange of nutrients as well as production of signaling molecules. Third, metabolic modeling can be used for prediction of essential enzymes and functional annotation of genomes. It can provide insight into metabolic adaptations and capacities of cells, strains, and organisms. Similar reasoning as outlined above also applies to the understanding of measurements of higher omics layers, whose changes may be hard to interpret in terms of cellular phenotype as they affect distant parts of metabolic network or since alterations in different genes lead to the same phenotype. These effects can be again better understood by integrating prior information on metabolic interactions from metabolic networks [20,45].

However, metabolic modeling is limited by inconsistencies between transcriptomic level, used as input, and metabolic level, which the methods aim to predict. For example, metabolic changes may not be captured by RNA-sequencing due to their shorter half-lives compared to transcriptional regulation or due to posttranscriptional stochasticity. Similarly, metabolic function may be affected by post-translational or post-transcriptional regulation, effects of degradation, and exogenous factors, such as nutrient availability, that are not reflected on transcriptomic level [20,31,45,46,78]. Nevertheless, the integration of prior information from metabolic networks may mitigate these negative effects compared to transcriptomic analysis alone. For example, scRNA-seq sparsity may lead to stochastic capture of lowly expressed genes, resulting in false positives. However, RNA-sequencing count-noise may be overcome by integration of prior knowledge of metabolic networks, providing a denoised view of the metabolism [20,31,45].

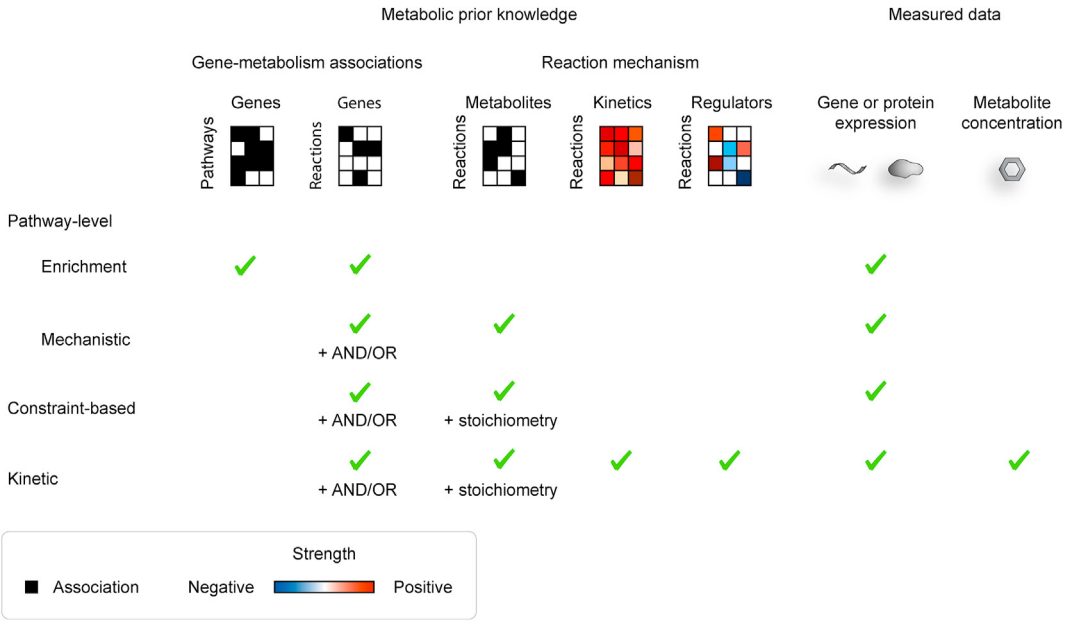

Figure 3: Inputs across modelling approaches. Modelling approaches require different information about gene—reaction associations, reaction mechanisms, and measured data. This figure shows only the most commonly used inputs, although different input information may be also incorporated in the models.

4. PERSPECTIVES

While a range of approaches for modeling cellular metabolic state have been proposed, a set of limitations restricts larger applicability. Here, we outline a series of potential extensions to address these.

4.1. Multi-omics integration

Currently, single-cell metabolic models are based mainly on mRNA concentration. The measurement and modeling of molecular entities that are chemically more directly related to metabolite concentrations could improve current models. For example, this information could be used to make RNA to enzyme activity mapping more reliable or to inform reaction activity constraints and kinetics [27,31,46]. Single-cell proteomics measurements could be used instead of transcriptomic data as they are more informative for prediction of enzyme activity. They directly correspond to enzyme availability and enable detection of posttranslational modifications that are key in determining enzyme function [95—97]. However, despite recent advances in single-cell proteomics there are still limitations in terms of availability, throughput, and sensitivity [95,98—100], which largely prohibit the use of single-cell proteomics in single-cell metabolic modeling. The used input data modality types dictate suitability of different modeling approaches. Namely, constraint-based systems are relevant if one can quantify reaction rates via enzyme concentration, for example from gene expression data. On the other hand, direct observations of metabolite species may result in models that are more akin to data imputation algorithms. More complex prior knowledge may require strong algorithmic changes of the models. Flux balance analysis, for example, centrally rests on a linear program. Deviations from mathematical assumptions inherent in this linear program would require a shift to a completely different parametric system and optimizer.

Similarly, inclusion of regulatory mechanisms would result in more reliable predictions [31,69,72,101]. This is of especial interest due to increasing availability of paired multi-omics datasets [1], such as joint scRNA-seq and scATAC-seq. To integrate regulatory information the models may need to account for complex interactions between entities, such as feedback loops [46]. As long as GEMs containing additional regulatory information are not yet available, one could try to leverage machine-learning to identify interactions between genes that may approximate the complex regulatory patterns in a data-driven manner.

4.2. Extensions of metabolic models across scales

Another direction for development of novel constraint-based models is adaptation to different systems. For example, eukaryotic cells are partitioned into metabolically different compartments with distinct metabolite exchange pathways between them, which could be included in the models. The necessary information on compartment-specific and exchange reactions is already included in some GEMs [37,102]. Furthermore, proteomic measurements with subcellular resolution [23,98] could further aid in improving predictions on subcellular level. Similarly, cells are not independent and exchange metabolites within tissues or microbial communities. Thus, the models could be extended to tissue-level, microbial-community-wide, and environment-aware systems [27,31]. Some currently available methods have already tried to go into this direction by modeling the availability of extracellular metabolites or assuming metabolite-sharing across cells [20,46,50,102]. In this respect, attention-DNNs, or in case of spatial data graph-DNNs, could be used for passing information across cells [103]. However, most deep learning frameworks developed for other single-cell data modalities treat cells independently because the models' scalability often crucially depends on this independence assumption. Many common deep learning approaches are limited in their ability to reflect dependencies between cells.

Changes of cellular metabolism through time are of great interest as they are involved, among others, in development, disease progression, environmental responses, and cyclic behavior. In terms of metabolic changes, these phenotypes are realized on vastly different time scales. First, development and disease progression occur over hours or months, resulting in changes in cell state or cell type composition. While these processes match the temporal resolution of common

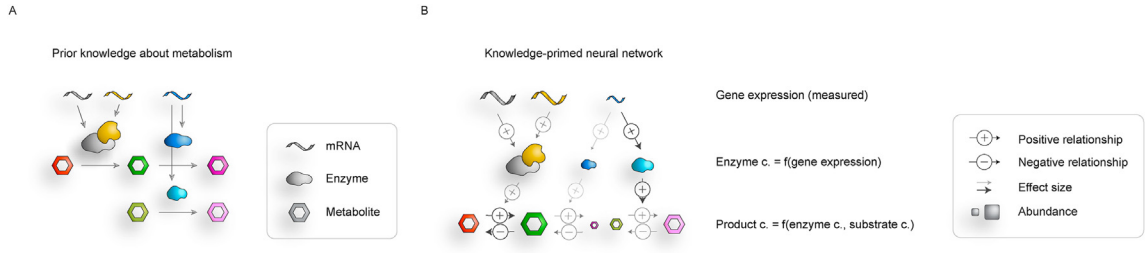

Figure 4: Knowledge-primed neural networks. (A) Prior knowledge about genes encoding enzymes and enzymes catalyzing metabolite conversions. (B) Neural network architecture can be based on metabolic knowledge. Different omics layers can be represented by different components or layers of the neural network so that connections between nodes correspond to prior metabolic knowledge.

single-cell experiments, time-series modelling in single-cell data poses a challenge since individual cells are not measured through time. Instead, data is generated by obtaining separate population snapshots across time-points, which are usually used to infer a typical trajectory of cellular progression and matched pseudotime. However, this trajectory does not reveal dynamic behavior of individual cells [10,104,105]. This static characteristic of the time-trajectory data corresponds well to dynamic FBA modelling, which solves a steady-state system for each time point based on time point-specific measurements [68,82,106], thus generating a transcriptome-matched metabolic snapshot. In contrast, kinetic models may be used for prediction of cell fate of individual cells, similarly as is done in the dynamic RNA-velocity model [85]. Finally, DNN architectures, such as recurrent-DNNs, could be used for time-dependent models [107]. Second, response to stimuli on metabolic level can occur almost instantaneously, which is not captured by commonly used single-cell methods such as scRNA-seq. In this respect, kinetic models may be used for prediction of short-term metabolic changes of individual cells starting from an initial measured state [68,82].

4.3. Deep-learning based models

Metabolic concentrations and fluxes can be thought of as latent states of a multi-omics molecular system of a cell (Figure 4). In terms of latent variable modelling, DNNs have recently emerged as a set of models that can flexibly deal with constraints, integration of different data modalities, and quantitative predictions, making them a promising candidate for development of new metabolic modelling tools. First, DNNs can scale to large datasets currently generated in the single-cell field [88,108,109]. Moreover, the use of large datasets will contribute to model regularization and thus more reliable predictions [30]. Second, the flexibility of the DNN structures enables the integration of diverse biological prior knowledge [108,109], which is necessary for the above-described extensions of metabolic modelling approaches. Third, DNNs are capable of learning patterns from data [109,110], which may lead to improved models even when biological mechanisms that could be used for constraining the model are not completely defined. Lastly, deep learning frameworks provide a convenient way for construction of new DNN models by abstracting mathematical and technical details [109], significantly improving over currently used linear programming libraries used in FBA for example.

4.4. Interpretation of metabolic data

The advent of single-cell metabolomic measuring techniques and computational prediction methods will require development of downstream analysis and interpretation approaches for metabolomic and fluxomic data. Many analysis ideas could be transferred from current single-cell omics analysis workflows, such as the ones developed for scRNA-seq [55,90,111–113]. However, prior knowledge of metabolic networks - a unique characteristic of metabolic data compared to other omics - should be leveraged for interpretation, leading to development of new analysis approaches [31,65]. Currently the cross-study contextualization of metabolic activity by tissue-domain specialists is often challenging due to the lack of metabolic characterization of tissues in single-cell resolved studies. Thus, shifting focus from genes to reactions would aid in metabolic interpretation.

5. CONCLUSION

In many signaling cascades, metabolism is the last omics layer that pools effects from higher omics layers and thus directly defines the cellular phenotype. Metabolic analysis has a large set of applications in all fields of biology, from medicine to bioprocess engineering. Because metabolic measurements are not yet widely available other approaches for understanding cellular metabolism are required. Here, metabolic modeling infers latent molecular states from widely available information about reaction networks and scRNA-seq data to predict reaction activities and metabolic concentrations.

Since single-cell data have different characteristics than bulk and provide additional insights into context-specific behaviors, new approaches for single-cell metabolomic modelling are to be anticipated. A diverse set of metabolic modelling approaches, as reviewed here, has been applied to single-cell data. We believe that new computational approaches, such as deep learning, may enable efficient modelling at single-cell and multi-omics level.

Single-cell metabolomic modeling will lead to new insights from existing single-cell sequencing data. First, the metabolic network information that is already integrated in the modelling approaches enables interpretation of results in terms of metabolic pathways or molecular function. Second, the use of prior knowledge about molecular networks increases robustness of predictions, which may be thus used to generate more reliable single-cell embeddings. Third, kinetic models could be used for prediction of swift perturbation responses that cannot be directly captured with single-cell sequencing techniques.

Computational methods used for metabolic analysis are informed by extensive and high-fidelity prior knowledge of metabolic networks, which is a unique feature of the metabolic modality. Machine-learning models developed for the analysis of single-cell metabolism could lay the foundation for the development of further interpretable machine-learning methods for molecular data.

AUTHOR CONTRIBUTIONS

Karin Hrovatin: Conceptualization; Writing - Original Draft, Review, and Editing; Visualization.

David S. Fischer: Conceptualization; Writing - Review and Editing. **Fabian J. Theis**: Conceptualization; Writing - Review; Supervision; Funding acquisition.

ACKNOWLEDGEMENTS

This work was supported by the Helmholtz Association under the joint research school Munich School for Data Science - MUDS [for K.H.], Add-on Fellowships for Interdisciplinary Life Science from Joachim Herz Foundation [to K.H.], German Research Foundation (DFG) fellowship through the Graduate School of Quantitative Biosciences Munich (QBM) [GSC 1006 to D.S.F.], and Helmholtz AI.

CONFLICT OF INTEREST

Fabian J. Theis reports receiving consulting fees from Immunai and ownership interest in Dermagnostix GmbH and Cellarity.

REFERENCES

[1] Zhu, C., Preissl, S., Ren, B., 2020. Single-cell multimodal omics: the power of many. Nature Methods 17(1):11–14.

[2] Kashima, Y., Sakamoto, Y., Kaneko, K., Seki, M., Suzuki, Y., Suzuki, A., 2020. Single-cell sequencing techniques from individual to multiomics analyses. Experimental & Molecular Medicine 52(9):1419–1427.

[3] Evers, T.M.J., Hochane, M., Tans, S.J., Heeren, R.M.A., Semrau, S., Nemes, P., et al., 2019. Deciphering metabolic heterogeneity by single-cell analysis. Analytical Chemistry 91(21):13314–13323.

[4] Kumar, R., Ghosh, M., Kumar, S., Prasad, M., 2020. Single cell metabolomics: a future tool to unmask cellular heterogeneity and virus-host interaction in context of emerging viral diseases. Frontiers in Microbiology 11. https://doi.org/10.3389/fmicb.2020.01152.

[5] Newgard, C.B., 2017. Metabolomics and metabolic diseases: where do we stand? Cell Metabolism 25(1):43–56.

[6] Petersen, A.-K., Krumsiek, J., Wägele, B., Theis, F.J., Wichmann, H.E., Gieger, C., et al., 2012. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. BMC Bioinformatics 13(1):120.

[7] Krumsiek, J., Suhre, K., Evans, A.M., Mitchell, M.W., Mohney, R.P., Milburn, M.V., et al., 2012. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genetics 8(10). https://doi.org/10.1371/journal.pgen.1003005.

[8] Mittelstrass, K., Ried, J.S., Yu, Z.H., Krumsiek, J., Gieger, C., Prehn, C., et al., 2011. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PLoS Genetics 7(8). https://doi.org/10.1371/journal.pgen.1002215.

[9] Lindeboom, R.G.H., Regev, A., Teichmann, S.A., 2021. Towards a human cell atlas: taking notes from the past. Trends in Genetics: TIG 37(7):625–630.

[10] Tritschler, S., Büttner, M., Fischer, D.S., Lange, M., Bergen, V., Lickert, H., et al., 2019. Concepts and limitations for learning developmental trajectories from single cell genomics. Development 146(12). https://doi.org/10.1242/dev.170506.

[11] Dal Co, A., Ackermann, M., van Vliet, S., 2019. Metabolic activity affects the response of single cells to a nutrient switch in structured populations. Journal of the Royal Society, Interface / the Royal Society(156):20190182.

[12] Tritschler, S., Theis, F.J., Lickert, H., Bottcher, A., 2017. Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas. Mol Metab 6(9):974–990.

[13] Sachs, S., Bastidas-Ponce, A., Tritschler, S., Bakhti, M., Böttcher, A., Sánchez-Garrido, M.A., et al., 2020. Targeted pharmacological therapy restores β-cell function for diabetes remission. Nature Metabolism 2(2):192–209.

[14] Xiao, Z., Dai, Z., Locasale, J.W., 2019. Metabolic landscape of the tumor microenvironment at single cell resolution. Nature Communications 10(1):3763.

[15] Artyomov, M.N., Van den Bossche, J., 2020. Immunometabolism in the single-cell era. Cell Metabolism 32(5):710–725.

[16] Levine, L.S., Hiam-Galvez, K.J., Marquez, D.M., Tenvooren, I., Madden, M.Z., Contreras, D.C., et al., 2021. Single-cell analysis by mass cytometry reveals metabolic states of early-activated CD8+ T cells during the primary immune response. Immunity 54(4):829–844 e5.

[17] Yao, Y., Zhang, P., Wang, J., Chen, J., Wang, Y., Huang, Y., et al., 2017. Dissecting target toxic tissue and tissue specific responses of irinotecan in rats using metabolomics approach. Frontiers in Pharmacology 8(122). https://doi.org/10.3389/fphar.2017.00122.

[18] Hao, Y., Quinnies, K., Realubit, R., Karan, C., Tatonetti, N.P., 2018. Tissue-specific analysis of pharmacological pathways. CPT: Pharmacometrics & Systems Pharmacology 7(7):453–463.

[19] Su, Y., Ko, M.E., Cheng, H., Zhu, R., Xue, M., Wang, J., et al., 2020. Multi-omic single-cell snapshots reveal multiple independent trajectories to drug tolerance in a melanoma cell line. Nature Communications 11(1):2345.

[20] Alghamdi, N., Chang, W., Dang, P., Lu, X., Wan, C., Gampala, S., et al., 2021. A graph neural network model to estimate cell-wise metabolic flux using single cell RNA-seq data. bioRxiv, 2020.9.23.310656.

[21] Bechtel, T.J., Reyes-Robles, T., Fadeyi, O.O., Oslund, R.C., 2021. Strategies for monitoring cell–cell interactions. Nature Chemical Biology 17(6):641–652.

[22] Liu, R., Yang, Z., 2021. Single cell metabolomics using mass spectrometry: techniques and data analysis. Analytica Chimica Acta 1143:124–134.

[23] Taylor, M.J., Lukowski, J.K., Anderton, C.R., 2021. Spatially resolved mass spectrometry at the single cell: recent innovations in proteomics and metabolomics. Journal of the American Society for Mass Spectrometry 32(4):872–894.

[24] Pinu, F.R., Beale, D.J., Paten, A.M., Kouremenos, K., Swarup, S., Schirra, H.J., et al., 2019. Systems biology and multi-omics integration: viewpoints from the metabolomics research community. Metabolites 9(4):31.

[25] Feng, D., Xu, T., Li, H., Shi, X., Xu, G., 2020. Single-cell metabolomics analysis by microfluidics and mass spectrometry: recent new advances. Journal of Analysis and Testing 4(3):198–209.

[26] Yao, H., Zhao, H., Zhao, X., Pan, X., Feng, J., Xu, F., et al., 2019. Label-free mass cytometry for unveiling cellular metabolic heterogeneity. Analytical Chemistry 91(15):9777–9783.

[27] Gu, C.D., Kim, G.B., Kim, W.J., Kim, H.U., Lee, S.Y., 2019. Current status and applications of genome-scale metabolic models. Genome Biology 20:18.

[28] Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., et al., 2020. Eleven grand challenges in single-cell data science. Genome Biology 21(1):31.

[29] Angerer, P., Simon, L., Tritschler, S., Wolf, F.A., Fischer, D., Theis, F.J., 2017. Single cells make big data: new challenges and opportunities in transcriptomics. Current Opinion in Systems Biology 4:85–91.

[30] Fischer, D.S., Dony, L., König, M., Moeed, A., Zappia, L., Tritschler, S., et al., 2020. Sfaira accelerates data and model reuse in single cell genomics. bioRxiv, 2020.12.16.419036.

[31] Zampieri, G., Vijayakumar, S., Yaneske, E., Angione, C., 2019. Machine and deep learning meet genome-scale metabolic modeling. PLoS Computational Biology 15(7):24.

[32] Richter, F.C., Obba, S., Simon, A.K., 2018. Local exchange of metabolites shapes immunity. Immunology 155(3):309–319.

[33] Labena, A.A., Gao, Y.-Z., Dong, C., Hua, H.-L., Guo, F.-B., 2018. Metabolic pathway databases and model repositories. Quantitative Biology 6(1):30–39.

[34] Fondi, M., 2018. Metabolic network reconstruction and modeling: methods and protocols, vol. 1716. Totowa: Humana Press Inc, 1–410.

[35] Faria, J.P., Rocha, M., Rocha, I., Henry, C.S., 2018. Methods for automated genome-scale metabolic model reconstruction. Biochemical Society Transactions 46:931–936.

[36] Bernstein, D.B., Sulheim, S., Almaas, E., Segre, D., 2021. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. Genome Biology 22(1):22.

[37] Brunk, E., Sahoo, S., Zielinski, D.C., Altunkaya, A., Dräger, A., Mih, N., et al., 2018. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nature Biotechnology 36(3):272–281.

[38] Angione, C., 2019. Human systems biology and metabolic modelling: a review—from disease metabolism to precision medicine. BioMed Research International 2019:8304260.

[39] Jatav, S., Malhotra, S., Miller, F.D., Jha, A., Goyal, S., 2020. Inferring metabolic rewiring in embryonic neural development using single cell data. bioRxiv, 2020.09.03.282442.

[40] Wang, Y., 2021. Identifying neuron subtype-specific metabolic network changes in single cell transcriptomics of Alzheimer's disease using perturbMet. bioRxiv, 2021.01.18.427154.

[41] Kanehisa, M., 2019. Toward understanding the origin and evolution of cellular organisms. Protein Science: A Publication of the Protein Society 28(11):1947–1951.

[42] Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al., 2020. The reactome pathway knowledgebase. Nucleic Acids Research 48(D1):D498–D503.

[43] Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., et al., 2017. The BioCyc collection of microbial genomes and metabolic pathways. Briefings in Bioinformatics 20(4):1085–1093.

[44] Norsigian, C.J., Pusarla, N., McConn, J.L., Yurkovich, J.T., Dräger, A., Palsson, B.O., et al., 2019. BiGG models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Research 48(D1):D402–D406.

[45] Wagner, A., Wang, C., Fessler, J., DeTomaso, D., Avila-Pacheco, J., Kaminski, J., et al., 2021. Metabolic modeling of single Th17 cells reveals regulators of autoimmunity. Cell. https://doi.org/10.1016/j.cell.2021.05.045.

[46] Damiani, C., Maspero, D., Di Filippo, M., Colombo, R., Pescini, D., Graudenzi, A., et al., 2019. Integration of single-cell RNA-seq data into population models to characterize cancer metabolism. PLoS Computational Biology 15(2):e1006733.

[47] Di Filippo, M., Colombo, R., Damiani, C., Pescini, D., Gaglio, D., Vanoni, M., et al., 2016. Zooming-in on cancer metabolic rewiring with tissue specific constraint-based models. Computational Biology and Chemistry 62:60–69.

[48] Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., Nielsen, J., 2012. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS Computational Biology 8(5):e1002518.

[49] Robinson, J.L., Kocabaş, P., Wang, H., Cholley, P.-E., Cook, D., Nilsson, A., et al., 2020. An atlas of human metabolism. Science Signaling 13(624):eaaz1482.

[50] Yilmaz, L.S., Li, X., Nanda, S., Fox, B., Schroeder, F., Walhout, A.J.M., 2020. Modeling tissue-relevant Caenorhabditis elegans metabolism at network, pathway, reaction, and metabolite levels. Molecular Systems Biology 16(10):e9649.

[51] Moretti, S., Tran, V.D.T., Mehl, F., Ibberson, M., Pagni, M., 2020. MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. Nucleic Acids Research 49(D1):D570–D574.

[52] Steinbeck, C., Conesa, P., Haug, K., Mahendraker, T., Williams, M., Maguire, E., et al., 2012. MetaboLights: towards a new COSMOS of metabolomics data management. Metabolomics: Official Journal of the Metabolomic Society 8(5):757–760.

[53] Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., et al., 2015. Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Research 44(D1):D463–D470.

[54] Crowell, H.L., Soneson, C., Germain, P.-L., Calini, D., Collin, L., Raposo, C., et al., 2020. Muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. Nature Communications 11(1):6077.

[55] Ma, Y., Sun, S., Shang, X., Keller, E.T., Chen, M., Zhou, X., 2020. Integrative differential expression and gene set enrichment analysis using summary statistics for scRNA-seq studies. Nature Communications 11(1):1585.

[56] Maleki, F., Ovens, K., Hogan, D.J., Kusalik, A.J., 2020. Gene set analysis: challenges, opportunities, and future research. Frontiers in Genetics 11(654). https://doi.org/10.3389/fgene.2020.00654.

[57] Das, S., McClain, C.J., Rai, S.N., 2020. Fifteen years of gene set analysis for high-throughput genomic data: a review of statistical approaches and future challenges. Entropy 22(4):427.

[58] Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al., 2019. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, cytoscape and enrichmentmap. Nature Protocols 14(2):482–517.

[59] Zhang, Y., Ma, Y., Huang, Y., Zhang, Y., Jiang, Q., Zhou, M., et al., 2020. Benchmarking algorithms for pathway activity transformation of single-cell RNA-seq data. Computational and Structural Biotechnology Journal 18:2953–2961.

[60] Richelle, A., Kellman, B.P., Wenzel, A.T., Chiang, A.W.T., Reagan, T., Gutierrez, J.M., et al., 2021. Model-based assessment of mammalian cell metabolic functionalities using omics data. Cell Reports Methods 1(3):100040.

[61] Ma, J., Shojaie, A., Michailidis, G., 2019. A comparative study of topology-based pathway enrichment analysis methods. BMC Bioinformatics 20(1):546.

[62] Ihnatova, I., Popovici, V., Budinska, E., 2018. A critical comparison of topology-based pathway analysis methods. PLoS One 13(1):e0191154.

[63] Amadoz, A., Hidalgo, M.R., Çubuk, C., Carbonell-Caballero, J., Dopazo, J., 2018. A comparison of mechanistic signaling pathway activity analysis methods. Briefings in Bioinformatics 20(5):1655–1668.

[64] Çubuk, C., Hidalgo, M.R., Amadoz, A., Rian, K., Salavert, F., Pujana, M.A., et al., 2019. Differential metabolic activity and discovery of therapeutic targets using summarized metabolic pathway models. Npj Systems Biology and Applications 5(1):7.

[65] Dusad, V., Thiel, D., Barahona, M., Keun, H.C., Oyarzun, D.A., 2021. Opportunities at the interface of network science and metabolic modeling. Frontiers in Bioengineering and Biotechnology 8:9.

[66] Novoa-del-Toro, E.M., Mezura-Montes, E., Vignes, M., Térézol, M., Magdinier, F., Tichit, L., et al., 2021. A multi-objective genetic algorithm to find active modules in multiplex biological networks. PLoS Computational Biology 17(8):e1009263.

[67] Klimm, F., Toledo, E.M., Monfeuga, T., Zhang, F., Deane, C.M., Reinert, G., 2020. Functional module detection through integration of single-cell RNA sequencing data with protein–protein interaction networks. BMC Genomics 21(1):756.

[68] Volkova, S., Matos, M.R.A., Mattanovich, M., Marín de Mas, I., 2020. Metabolic modelling as a framework for metabolomics data integration and analysis. Metabolites 10(8). https://doi.org/10.3390/metabo10080303.

[69] Vijayakumar, S., Conway, M., Lió, P., Angione, C., 2018. Seeing the wood for the trees: a forest of methods for optimization and omic-network integration in metabolic modelling. Briefings in Bioinformatics 19(6):1218–1235.

[70] Pinzon, W., Vega, H., Gonzalez, J., Pinzon, A., 2019. Mathematical framework behind the reconstruction and analysis of genome scale metabolic models. Archives of Computational Methods in Engineering. State of the Art Reviews 26(5):1593–1606.

[71] Tonn, M.K., Thomas, P., Barahona, M., Oyarzún, D.A., 2019. Stochastic modelling reveals mechanisms of metabolic heterogeneity. Communications Biology 2(1):108.

[72] Machado, D., Herrgård, M., 2014. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Computational Biology 10(4):e1003580.

[73] Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nature Biotechnology 28(3):245—248.

[74] Kim, O.D., Rocha, M., Maia, P., 2018. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. Frontiers in Microbiology 9:1690. https://doi.org/10.3389/fmicb.2018.01690.

[75] Zhang, Y., Kim, M.S., Nguyen, E., Taylor, D.M., 2020. Modeling metabolic variation with single-cell expression data. bioRxiv, 2020.01.28.923680.

[76] Moss Bendtsen, K., Jensen, M.H., Krishna, S., Semsey, S., 2015. The role of mRNA and protein stability in the function of coupled positive and negative feedback systems in eukaryotic cells. Scientific Reports 5, 13910—13910.

[77] Yasemi, M., Jolicoeur, M., 2021. Modelling cell metabolism: a review on constraint-based steady-state and kinetic approaches. Processes 9(2):322.

[78] Tonn, M.K., Thomas, P., Barahona, M., Oyarzun, D.A., 2020. Computation of single-cell metabolite distributions using mixture models. Frontiers in Cell and Developmental Biology 8. https://doi.org/10.3389/fcell.2020.614832.

[79] Persi, E., Duran-Frigola, M., Damaghi, M., Roush, W.R., Aloy, P., Cleveland, J.L., et al., 2018. Systems analysis of intracellular pH vulnerabilities for cancer therapy. Nature Communications 9(1):2997.

[80] Ferland, G., Doucet, I., Mainville, D., 2016. Phylloquinone and menaquinone-4 tissue distribution at different life stages in male and female sprague-dawley rats fed different vk levels since weaning or subjected to a 40% calorie restriction since adulthood. Nutrients 8(3):141.

[81] Bordbar, A., McCloskey, D., Zielinski, D.C., Sonnenschein, N., Jamshidi, N., Palsson, B.O., 2015. Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics. Cell Systems 1(4):283—292.

[82] Antoniewicz, M.R., 2013. Dynamic metabolic flux analysis—tools for probing transient states of metabolic networks. Current Opinion in Biotechnology 24(6):973—978.

[83] Song, H.-S., Ramkrishna, D., 2009. When is the quasi-steady-state approximation admissible in metabolic modeling? when admissible, what models are desirable? Industrial & Engineering Chemistry Research 48(17):7976—7985.

[84] Migeon, B.R., 2020. X-linked diseases: susceptible females. Genetics in Medicine: Official Journal of the American College of Medical Genetics 22(7):1156—1174.

[85] Bergen, V., Lange, M., Peidli, S., Wolf, F.A., Theis, F.J., 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. Nature Biotechnology 38(12):1408—1414.

[86] Rich, L.J., Bagga, P., Wilson, N.E., Schnall, M.D., Detre, J.A., Haris, M., et al., 2020. 1H magnetic resonance spectroscopy of 2H-to-1H exchange quantifies the dynamics of cellular metabolism in vivo. Nature Biomedical Engineering 4(3):335—342.

[87] Balcells, C., Foguet, C., Tarrago-Celada, J., de Atauri, P., Marin, S., Cascante, M., 2019. Tracing metabolic fluxes using mass spectrometry: stable isotope-resolved metabolomics in health and disease. Trac-Trends in Analytical Chemistry 120:13.

[88] Luecken, Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., et al., 2020. Benchmarking atlas-level data integration in single-cell genomics. bioRxiv, 2020.05.22.111161.

[89] Lotfollahi, M., Naghipourfar, M., Luecken, M.D., Khajavi, M., Büttner, M., Avsec, Z., et al., 2020. Query to reference single-cell integration with transfer learning. bioRxiv, 2020.07.16.205997.

[90] Luecken, M.D., Theis, F.J., 2019. Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular Systems Biology 15(6):e8746.

[91] Bartel, J., Krumsiek, J., Theis, F., 2013. Statistical methods for the analysis of high-throughput metabolomics data. Computational and Structural Biotechnology Journal 4.

[92] Chowdhury, S., Fong, S.S., 2020. Leveraging genome-scale metabolic models for human health applications. Current Opinion in Biotechnology 66:267—276.

[93] Sertbas, M., Ulgen, K.O., 2020. Genome-scale metabolic modeling for unraveling molecular mechanisms of high threat pathogens. Frontiers in Cell and Developmental Biology 8:22.

[94] Altamirano, A., Saa, P.A., Garrido, D., 2020. Inferring composition and function of the human gut microbiome in time and space: a review of genome-scale metabolic modelling tools. Computational and Structural Biotechnology Journal 18:3897—3904.

[95] Kelly, R.T., 2020. Single-cell proteomics: progress and prospects. Molecular & Cellular Proteomics: MCP 19(11):1739—1748.

[96] Großeholz, R., Koh, C.-C., Veith, N., Fiedler, T., Strauss, M., Olivier, B., et al., 2016. Integrating highly quantitative proteomics and genome-scale metabolic modeling to study pH adaptation in the human pathogen Enterococcus faecalis. NPJ Systems Biology and Applications 2:16017.

[97] Rai, A., Saito, K., 2016. Omics data input for metabolic modeling. Current Opinion in Biotechnology 37:127—134.

[98] Paul, I., White, C., Turcinovic, I., Emili, A., 2021. Imaging the future: the emerging era of single-cell spatial proteomics. The FEBS Journal.. https://doi.org/10.1111/febs.15685.

[99] Vistain, L.F., Tay, S., 2021. Single-cell proteomics. Trends in Biochemical Sciences 46(8):661—672.

[100] Schoof, E.M., Furtwängler, B., Üresin, N., Rapin, N., Savickas, S., Gentil, C., et al., 2021. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. Nature Communications 12(1):3341.

[101] Jensen, P.A., 2018. Coupling fluxes, enzymes, and regulation in genome-scale metabolic models. Methods in Molecular Biology 1716:337—351.

[102] Thiele, I., Sahoo, S., Heinken, A., Hertel, J., Heirendt, L., Aurich, M.K., et al., 2020. Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. Molecular Systems Biology 16(5):e8982.

[103] Ha, S., Jeong, H., 2021. Unraveling hidden interactions in complex systems with deep learning. Scientific Reports 11(1):12804.

[104] Fischer, D.S., Fiedler, A.K., Kernfeld, E.M., Genga, R.M.J., Bastidas-Ponce, A., Bakhti, M., et al., 2019. Inferring population dynamics from single-cell RNA-sequencing time series data. Nature Biotechnology 37(4):461—468.

[105] Weinreb, C., Wolock, S., Tusi, B.K., Socolovsky, M., Klein, A.M., 2018. Fundamental limits on dynamic inference from single-cell snapshots. Proceedings of the National Academy of Sciences 115(10):E2467—E2476.

[106] Chandrasekaran, S., Zhang, J., Sun, Z., Zhang, L., Ross, C.A., Huang, Y.-C., et al., 2017. Comprehensive mapping of pluripotent stem cell metabolism using dynamic genome-scale network modeling. Cell Reports 21(10):2965—2977.

[107] Rajendra, P., Brahmajirao, V., 2020. Modeling of dynamical systems through deep learning. Biophysical Reviews 12(6):1311—1320.

[108] Oller-Moreno, S., Kloiber, K., Machart, P., Bonn, S., 2021. Algorithmic advances in machine learning for single-cell expression analysis. Current Opinion in Systems Biology 25:27—33.

[109] Eraslan, G., Avsec, Ž., Gagneur, J., Theis, F.J., 2019. Deep learning: new computational modelling techniques for genomics. Nature Reviews Genetics 20(7):389—403.

[110] Lopez, R., Li, B., Keren-Shaul, H., Boyeau, P., Kedmi, M., Pilzer, D., et al., 2021. Multi-resolution deconvolution of spatial transcriptomics data reveals continuous patterns of inflammation. bioRxiv, 2021.05.10.443517.

[111] Wolf, F.A., Angerer, P., Theis, F.J., 2018. SCANPY: large-scale single-cell gene expression data analysis. Genome Biology 19(1):15.

[112] Zappia, L., Phipson, B., Oshlack, A., 2018. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. PLoS Computational Biology 14(6):e1006245.

[113] Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., Regev, A., 2015. Spatial reconstruction of single-cell gene expression data. Nature Biotechnology 33(5):495—502.

