iScience

CellPress
OPEN ACCESS

Review

From sampling to simulating: Single-cell multiomics in systems pathophysiological modeling

Alexandra Manchel,[1] Michelle Gee,[1,2] and Rajanikanth Vadigepalli[1,*]

[1]Daniel Baugh Institute of Functional Genomics/Computational Biology, Department of Pathology and Genomic Medicine, Thomas Jefferson University, Philadelphia, PA, USA

[2]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA

*Correspondence: rajanikanth.vadigepalli@jefferson.edu

https://doi.org/10.1016/j.isci.2024.111322

SUMMARY

As single-cell omics data sampling and acquisition methods have accumulated at an unprecedented rate, various data analysis pipelines have been developed for the inference of cell types, cell states and their distribution, state transitions, state trajectories, and state interactions. This presents a new opportunity in which single-cell omics data can be utilized to generate high-resolution, high-fidelity computational models. In this review, we discuss how single-cell omics data can be used to build computational models to simulate biological systems at various scales. We propose that single-cell data can be integrated with physiological information to generate organ-specific models, which can then be assembled to generate multi-organ systems pathophysiological models. Finally, we discuss how generic multi-organ models can be brought to the patient-specific level thus permitting their use in the clinical setting.

INTRODUCTION

Single-cell sequencing, including transcriptomics, spatial transcriptomics, proteomics, and most recently, metabolomics, provides high-resolution insight into the biological mechanisms and processes occurring within individual cells and their microenvironments. The recent single-cell ''omics'' revolution initiated a cascade of computational efforts for downstream analysis and understanding of the underlying biological systems. The current state-of-the-art single-cell analysis pipelines are used to elucidate cellular states within a greater cell type population such that the relative distribution and proportion of these cell states within the cell type population can then be studied. Here, we lay out the opportunities and a conceptual framework for using this rich information toward building high-fidelity computational models (i.e., ODE-based, Boolean, neural networks) and simulations of biological systems at multiple scales. As a step toward recognizing these opportunities, it is instructive to note how the data generation enabled by the single-cell omics technologies represents a transformative shift from previous practices of data collection and integration for use in modeling and simulation at a substantively lower scale and throughput. Computational modeling has been essential and long standing in the field of biology as a means to study and predict complex biological systems and their dynamics. For decades these models have been built and parameterized using data from very targeted and biased experiments, thereby limiting the scale and resolution of these models. In the present review, we refer to these data as ''Not So Big'' as it is disparate and, therefore must be obtained individually from various literature sources, collated together, and then integrated for the development of mechanistic models (Figure 1). Some classical examples of this type of modeling include the Hodgkin Huxley conductance model,[1] the Michaelis-Menten kinetic model,[2] and the Tyson (1991) cell cycle gene regulatory network (GRN) model.[3]

Our use of the term ''Not So Big'' data to refer to conventional datasets is to draw a direct contrast to ''Big'' data, which is also not a precisely defined term but is widely understood to denote the large volume of high dimensional datasets, such as transcriptomics, proteomics, metabolomics, etc. Another way to refer to the available spectrum of datasets is to differentiate them as small, medium/mesoscale, and large. The different collections of many conventional small-scale datasets can be considered as forming a mesoscale dataset that informs computational modeling. In the text below, we chose to use the terms ''Big'' and ''Not So Big'' data based on the relatively popular use of the term ''Big'' data to refer to omics datasets. ''Big'' datasets can thus provide high dimensional unbiased quantification of >10,000 genes (transcriptomics) and proteins (proteomics) and >100,000 metabolites (metabolomics) at single-cell, single-molecule, and spatial resolution within a tissue microenvironment. With the unprecedented rise of ''Big'' data, there is a wide landscape of new opportunity for analyzing such data for further and deeper investigation of the inner workings of various biological systems. In this review, we propose a systematic framework for ''Big'' data transformation and feature extraction, which can be further informed by compendia of conventional ''Not So Big'' data, to bridge the gap between the collection and analysis of ''Big'' and ''Not So Big'' data and the development and simulation of high-fidelity, multiscale models of tissues at various spatiotemporal levels.

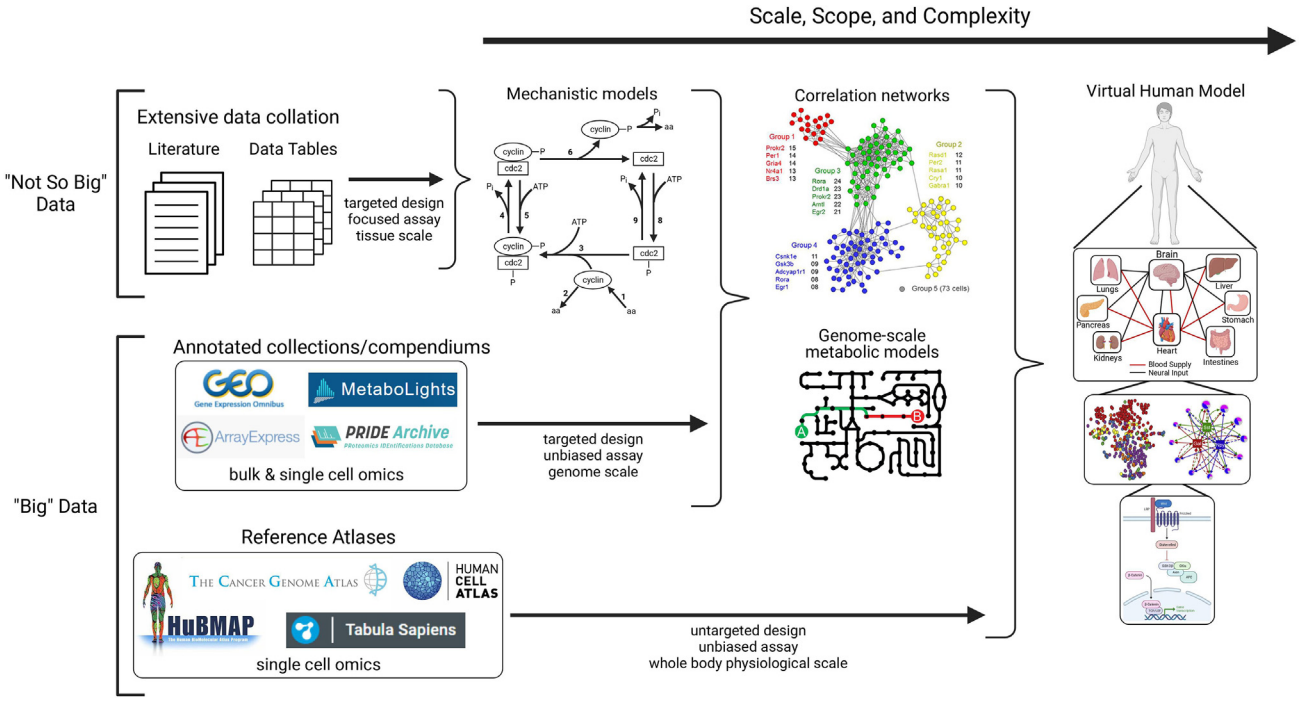

Figure 1. Data sources for advancing computational modeling

"Not So Big" and "Big" data sources can be utilized collectively to build computational models of varying complexity. "Not So Big" data are usually disparate, require extensive data collation, and must be obtained individually from various literature sources in the form of data tables present in the supplementary material. The "Not So Big" data are derived from targeted and focused experiments and provides tissue-level detail for mechanistic models such as the first cell cycle gene regulatory network from Tyson (1991).[3] Bulk and single-cell "Big" data are derived from targeted and unbiased assays, and are usually stored in annotated collections and compendiums such as GEO[4] (transcriptomics), ArrayExpress[5] (transcriptomics), MetaboLights[6] (metabolomics), and PRIDE[7] (proteomics). This "Big" data provides genome-scale detail for informing correlation networks and genome scale metabolic models. Single-cell "Big" data from reference atlases, including The Cancer Genome Atlas,[8] the Human Cell Atlas,[9] HuBMAP,[10] and Tabula Sapiens,[11] provide untargeted and unbiased assays at the whole-body physiological scale. These data can be utilized to inform future virtual human models at various scales, including the molecular level (i.e., Wnt/B-Catenin signaling pathway), single-cell level (i.e., gene correlation networks), and physiological level (i.e., multi-organ interactions). Figure 1 adapted from Tyson (1991),[3] Gustafsson et al. (2023),[12] Park et al. (2016),[13] Moss et al. (2021).[14] Created using biorender.com.

The data analytical approach for "Big" data are significantly different from that of the "Not So Big" data as there has been a significant push for data processing and sharing standards. For example, the Gene Expression Omnibus (GEO)[4] was developed to standardize storing and sharing policies for bulk and single-cell transcriptomics data, among other omics data. Additional repositories exist for storing transcriptomics, proteomics and metabolomics data, including ArrayExpress,[5] PRIDE,[7] and MetaboLights,[6] respectively. Databases for "Big" data storing then enable automated analysis algorithms and workflows for feature extraction and downstream use for computational model building (Figure 1). While experiments yielding "Not So Big" data are targeted and focused, resulting in highly specified data derived from the biological question of interest, "Big" data sources, such as those hosted on databases like GEO, are derived from targeted, unbiased experimental designs, which cast a larger net for which biological questions can be formulated and answered by repurposing the data. The evolution from biased "Not So Big" data to unbiased "Big" data have facilitated new modeling efforts such as "Big" data informed genome scale metabolic models (GEMs)[12] and correlation network models[13] (Figure 1). However, we seek to explore the unique opportunities of computational modeling in biology that explicitly incorporate and are driven by single-cell omics datasets such that high-fidelity, high resolution, dynamic models of pathophysiological systems can be developed and simulated.

The conceptual and technological advances from "Not So Big" data toward "Big" data standards have also initiated the development of untargeted, unbiased, whole-body physiological scale data sources such as those from single-cell reference atlases. These data are not specific to any single study or disease state but rather encapsulate information from nearly all single-cell studies. The reference databases contain large amounts of data that can be used to query the functionality of a microenvironment or tissue at single-cell resolution. The growing interest in single-cell technology led to the formulation of The Human BioMolecular Atlas Program (HuBMAP), which developed state-of-the-art and publicly available platforms for mapping healthy cells in the human body to determine cellular function and relationships.[10] Other cell atlas initiatives, including The Human Cell Atlas,[9] The Tabula Sapiens,[11] and The Cancer Genome Atlas,[8] have been developed for similar purposes. Additional single-cell atlases specific to tissues, including liver (The Liver Atlas)[15] and white adipose tissue (Single-cell atlas of human

Table 1. Molecularly targeted methods for single-cell and spatial transcriptomics

Single-cell transcriptomics method	Spatial?	Cell capture efficiency	Resolution	Cells captured per sample	Gene detection per cell or spot
InDrop[26]	No	~70–80%	Single-cell	>40,000	500-5000 genes/cell
Drop-seq[25]	No	~10%	Single-cell	>50,000	~300 genes/cell
Chromium (10X Genomics)[27]	No	~50–70%	Single-cell	>80,000	500-5000 genes/cell
Visium (10X Genomics)	Yes	N/A	55 μm spot (multicellular)	~5,000 spots	~150 genes/spot
Slide-Seq[28,29]	Yes	N/A	10 μm spot (single-cell)	~70,000	~150 genes/spot
DBiT-seq[30]	Yes	N/A	10 μm spot (single-cell)	~15,000	~2,000 genes/spot
Stereo-seq[31]	Yes	N/A	220 nm spot (subcellular)	~280,000	~500 genes/cell
CosMx[32]	Yes	N/A	Subcellular	>100,000	~100 genes/cell
Xenium (10X Genomics)	Yes	N/A	Subcellular	>100,000	~100 genes/cell

and mouse white adipose tissue),[16] have been developed to better understand tissue-specific cellular states and their respective function in health and disease. The Broad Institute also developed a single-cell portal for the exploration and acceleration of single-cell research for individuals unfamiliar with data analysis and computational biology.[17] These reference atlases, therefore, provide an unbiased foundational platform for computational model building of a virtual human model with physiological,[18] cellular,[14] and molecular signaling details (Figure 1).

Several efforts have also been made toward generating platforms for large-scale multiscale model building. For instance, the NIH funded the "Simmune" project, which enables comprehensive modeling by providing a suite of software tools that guide the user through the multiple hierarchical scales of cellular behavior.[19] Such initiatives and software tools can be helpful for those who want to build computational models using experimental data but lack the skills necessary for model building, development, and simulation. An additional software developed for cell-based simulation of multicellular systems, CellSys, implements agent-based modeling that supports simulation and analysis of virtual tissues.[20] A large-scale, agent-based model of granuloma formation and function in lungs during *M. tuberculosis* infection, GranSim, was developed by incorporating cellular and molecular details and can be utilized for predicting the pharmacokinetic and dynamics profiles of various drugs, among many other functionalities.[21] While these software platforms provide a framework for large-scale modeling and simulation, the growth in the availability of single-cell data have not been matched by the growth in predictive knowledge generated by computational models. Additional tools for model development are necessary for integrating "Big" data with "Not So Big" data such that computational models can be built and simulated at a pace that matches data generation.

In the remainder of the review, we first discuss the range of technological developments for sampling at the single-cell scale. Next, we explore the available methods for extracting features from single-cell omics data for use in computational models. Then, we detail how the features extracted from the "Big" and "Not So Big" data can be utilized collectively to build and test high-resolution, high-fidelity computational models. We then propose that these computational models can be expanded to multi-organ systems pathophysiological models and consider the implications for patient-specific modeling and simulation and in silico clinical trials. Lastly, we comment on the possible opportunities and implications for computational models informed by single-cell data.

SAMPLING METHODS FOR SINGLE-CELL OMICS DATA ACQUISITION

Sampling for single-cell and spatial transcriptomics

As the need for single-cell transcriptomics data have rapidly increased, various sampling methods have been developed to meet such demands. Prior to droplet-based single-cell RNA-seq (scRNA-seq) methods, single-cell data were acquired using methods including fluorescence-activated cell sorting (FACS),[22] magnetic-activated cell sorting (MACS),[23] and laser capture microdissection (LCM).[24] These methods selectively isolate specific cell populations based on regions of interest (LCM) or cell markers (FACS and MACS), which have high sensitivity and accuracy. The throughput of single-cell sampling methods varies; while LCM throughput is low, FACS and droplet-based microfluidics throughput is high.

Several untargeted methods of single-cell sampling for transcriptomics exist, including the microfluidic-based Drop-seq,[25] InDrop,[26] and 10X Chromium[27] methods (Table 1). These three isolation methods use microfluidics to tag individual cells with a unique molecular identifier (UMI), resulting in a matrix containing the absolute number of counts for each transcript in each cell. The droplet-based microfluidic single-cell isolation methods are advantageous as they have high sensitivity and specificity. These sampling methods have varying cell capture efficiencies and gene detection rates (Table 1).

While each method of single-cell sampling for transcriptomics has its benefits and limitations, they all produce high-resolution sequencing data for single-cells but lack spatial information. Therefore, sequencing-based and multiplexed immunohistochemistry (IHC)/immunofluorescence (IF)-based technologies have been developed to spatially profile the transcriptomes of tissue slices. Fluorescent *in situ* sequencing (FISSEQ) was the first available platform for spatial transcriptomics.[33] This method initiated a cascade of other spatial barcoding and *in situ* sequencing methods with single-cell and subcellular resolution. One of these methods, Visium (10x Genomics), produces lower spatially resolved sequencing data of cell spots (50 μm) from whole tissue slides (Table 1). Methods that have adopted

Table 2. Mass-spectrometry-based methods of single-cell proteomics sampling

Proteomics methods	Cell Isolation Method	Cells captured per run	Protein or metabolite detection per cell
CyTOF[44]	Microfluidics	>100,000	~50 proteins/cell
SCoPE-MS[37,38]	Microfluidics or manual picking	~200	~1,000 proteins/cell
nanoPOTS[39]	Microfluidics or manual picking	10–100	~200–300 proteins/cell
OAD chip[40]	Microfluidics	1 (per chip)	~50 proteins/cell
iPAD[41]	Syringe pump	1 (24 cells/day)	~100–200 proteins/cell
FAIMS[42]	Microfluidics or manual picking	1 (200 cells/day)	>1,000 proteins/cell

higher resolution include Slide-seq,[28,29] DBiT-seq (deterministic barcoding in tissue for spatial omics sequencing),[30] and most recently, Stereo-seq[31] (Table 1). Furthermore, NanoString's CosMx[32] and 10X Genomics' Xenium probe-based platforms for spatial transcriptomics sampling achieve subcellular resolution (Table 1).

Sampling for single-cell proteomics and metabolomics

Fewer widely utilized sampling methods exist for single-cell proteomics and metabolomics compared to transcriptomics, as the fields are only just emerging. This is due to the increased difficulty of protein and metabolite quantification at the single-cell level.[34,35] However, mass spectrometry (MS) is very popular in the proteomics and metabolomics fields, and innovations within MS have propelled research and our current understanding of various biological systems. Several methods exist for proteomics sequencing of single-cells and are discussed in more detail in a recent review by Kelly (2020).[36] Briefly, we highlight a few single-cell proteomics methods, including single-cell ProtEomics by mass spectrometry (SCoPE-MS)[37,38] nanodroplet processing in one pot for trace samples (nanoPOTS),[39] the oil-air-droplet (OAD) chip,[40] the integrated proteome analysis device (iPAD),[41] and high field asymmetric ion mobility spectrometry (FAIMS)[42] (Table 2). These single-cell protein quantification methods have been utilized alone or in conjunction as part of a larger workflow to quantify over 1000 protein groups per single.[43]

Several single-cell sampling methods exist for metabolite analysis based on ion beams, lasers, probes, or microfluidic devices.[45,46] Ion beam-based techniques include cytometry by time of flight (CyTOF),[44] time-of-flight secondary ion MS (TOF-SIMS)[47] and nanoscale SIMS (nanoSIMS),[48] which can achieve resolution as high as 100 nm and have shown significant advancements over flow cytometry for single-cell proteomic quantification (Table 3). Since SIMS-based methods are more energetic and result in a larger number of fragments produced, matrix-assisted laser desorption-ionization (MALDI)-based methods have been more commonly utilized for single-cell MS experiments, as this method provides high sensitivity and throughput.[45] Some variations of MALDI-based methods of single-cell metabolomics quantification include MALDI-MS,[49] laser ablation electrospray ionization MS (LAESI-MS),[50] and nanopost array-laser desorption ionization (NAPA-LDI)[51] (Table 3). A popular probe-based method is nano-desorption electrospray ionization (nano-DESI)[52] (Table 3), while several microfluidic-based approaches have been introduced by the Yang research group,[53,54] which were designed to allow for higher throughput analysis.

Sampling for simultaneous transcriptomics-proteomics collection

Approaches for simultaneous single-cell sampling and transcriptomics-proteomics analysis have been developed. Some of the discussed methods for transcriptome-proteome collection methods include proximity extension assay/specific RNA target amplification (PEA/STA),[55] proximity ligation assay for RNA (PLAYR),[56] cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq),[57] and RNA expression and protein sequencing assay (REAP-seq),[58] which permit sensitive and specific integrative analysis of the transcriptome and proteome (Table 4).

EXTRACTING SINGLE-CELL SCALE FEATURES FOR COMPUTATIONAL MODELING

Single-cell omics enables cellular gene expression quantification with higher resolution, such that cellular heterogeneity and differential expression across cell states can be revealed.

Table 3. Popular methods for single-cell metabolomics sampling

Metabolomics Method	Type	Resolution	Molecule size detection limit	Limit of Detection
TOF-SIMS[47]	Ion beam-based	100 nm to 1 μm	0–10,000 Da	Ppm - ppb
nanoSIMS[48]	Ion beam-based	50 nm to 1 μm	0–400 Da	Ppm - ppb
MALDI-MS[49]	Laser-based	30–200 μm	0–200,000 Da	fmol
LAESI-MS[50]	Laser-based	10–100 μm	0 - > 100,000 Da	fmol
NAPA-LDI-MS[51]	Laser-based	40 um	0–2,000 Da	zmol
nano-DESI[52]	Probe-based	10–15 μm	0–2000 Da	pmol

Table 4. Examples of transcriptome-proteome multiomics technologies

Multiomics Method	Method of single-cell isolation	Method of RNA/protein separation	RNA/protein measured	Cellular Throughput	RNA detection rate	Protein detection rate
PEA/STA[55]	Microfluidic channels	Reverse transcription of PEA probe and RNA, then amplify	Targeted RNA and targeted proteins	96 cells	~96 RNAs/cell	~38 proteins/cell
PLAYR[56]	Flow or mass cytometry	Amplify PLAYR probe pair product and perform antibody staining	Targeted cytosolic RNA and targeted proteins	100-1000's of cells	~20 RNAs/cell	~20 proteins/cell
CITE-seq[57]	Drop-seq and 10X Genomics	Reverse transcription of mRNA and antibody-derived oligonucleotides, then separate libraries	Polyadenylated RNA and targeted cell surface proteins	100,000's of cells	1,000–5,000 genes/cell	100's of proteins
REAP-seq[58]	10X Genomics	Reverse transcription of mRNA and antibody-derived oligonucleotides, then separate libraries	Polyadenylated RNA and targeted cell surface proteins	100,000's of cells	1,000–5,000 genes/cell	100's of proteins

Technological advances in multiomics data collection have been matched with the advancement in computational tools for data integration and analysis, including those using generative artificial intelligence and large language models for cell type annotation and statistical modeling.[59–62] These emerging methods differ from previous methods as they utilize high dimensional single-cell atlases and reference datasets to scale by building data-driven models of the underlying gene annotation relationships in ways that are not captured by previous approaches. For example, scGPT is a single-cell foundation model which is pre-trained on over 33 million cells from various cell atlases, enabling learnings of cell and gene embeddings, thereby facilitating the modeling of various aspects of cellular processes in addition to cell type annotation, batch correction, multiomic data integration, genetic perturbation prediction, and gene network inference.[61]

State-of-the-art pipelines for cell state identification from single-cell transcriptomics data involve aligning reads to the genome, assessing RNA quality, identifying variable features, and performing dimensionality reduction and clustering (Figure 2A). For example, the output for the 10X Genomics single-cell transcriptomics method is transcript reads, which is stored in a raw FASTQ file. The transcript reads within the FASTQ file must then be aligned to the genome using tools such as STAR[63] or TopHat2.[64] The quality of the data should then be assessed, after which the data can be normalized and utilized for downstream analysis. A widely used R package for single-cell and spatial transcriptomics analysis is Seurat,[65] which was designed for quality control and data exploration. The Seurat package provides additional functions for identifying sources of heterogeneity within the data and integrating multiple datasets from various sources. Despite different sampling methodologies and experimental output, the data must be normalized, after which data clustering and visualization methods can be applied (Figure 2A).

Single-cell analysis methods have mainly focused on cell type and cell state identification by high dimensional data reduction in the form of t-distributed stochastic neighbor embedding (tSNE) or Uniform Manifold Approximation and Projection (UMAP) plots. tSNE and UMAP plots are two methods for visualizing high-dimensional omics data in two dimensions, such that a given sample (individual cell) is represented as a point in space and cells that are quantitatively similar across the omics profile group together. Communities of quantitatively similar cell groups can be detected and then classified by cell type and further by cell state based on functional similarities within the cell type cluster. Following cell state identification of each cluster within the UMAP/tSNE plot, several tools have been recently developed for predicting cell state interactions, transitions, and trajectories, which can serve as the substrates for building computational models of cellular interaction networks (Figure 2B). For instance, information about cell state interactions or the communication between and across cell states via receptor-ligands pairs within a given cell type can be extracted using a variety of publicly available tools and algorithms. One popular tool for this type of data inference is NicheNet, which models and predicts ligand-target links between interacting cells by combining their gene expression with prior knowledge on signaling and gene regulatory

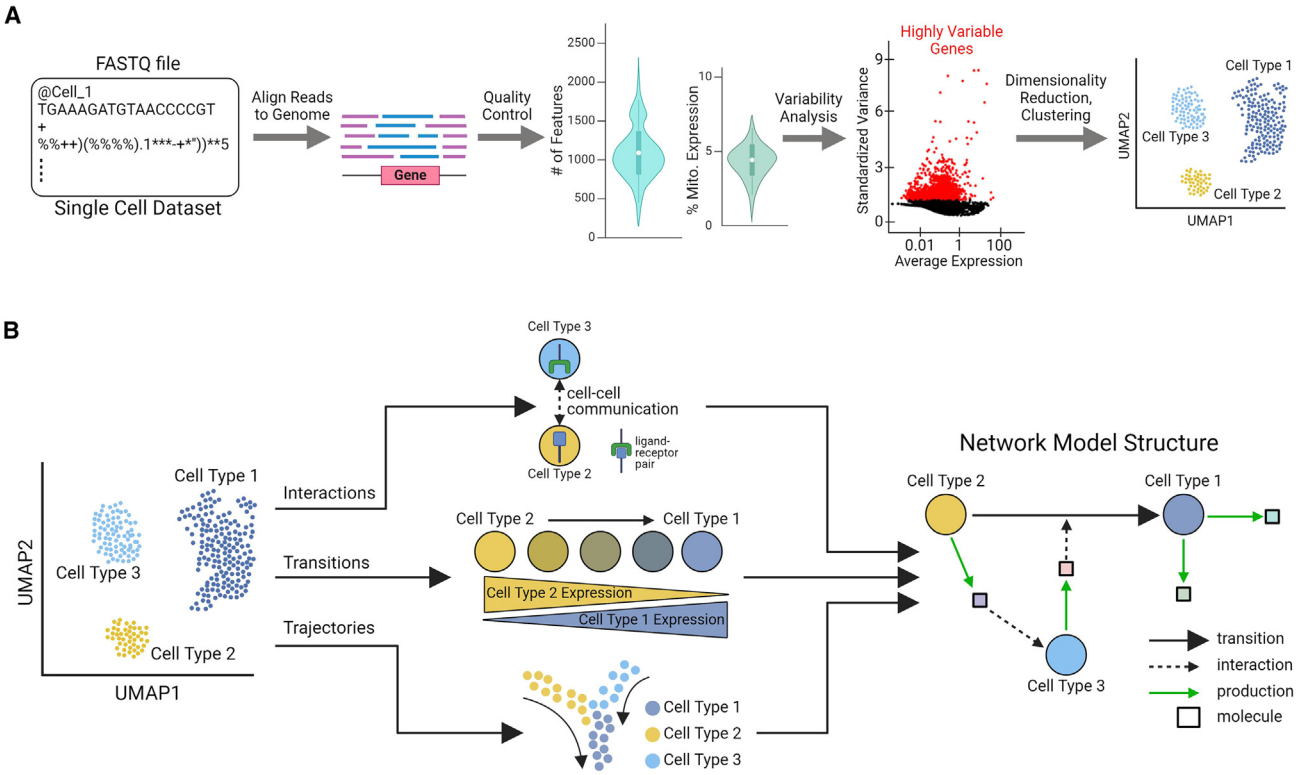

Figure 2. Extracting single-cell features from omics data for network modeling

(A) Raw single-cell datasets may be output in FASTQ file format, which can be fed though data analysis pipelines such as STAR[63] or TopHat2[64] to align reads to the reference genome. The quality of the data can then be assessed, and variability analysis can be performed to identify highly variable genes in the dataset. Dimensionality reduction and clustering algorithms can then be applied to visualize the data in two-dimensions in the form of a UMAP plot and identify cell types. (B) Following cell state identification of each cluster within the UMAP plot, cell state interactions, transitions, and trajectories can be predicted, which may serve as the substrates for building and simulating computational models of cellular interaction networks. Created using biorender.com.

networks.[66] Cells may also transition from one state to another, which can be vital in a variety of biological processes including tissue development and cancer. MuTrans (multiscale method for transient cells) is one method for investigating and identifying the underlying stochastic dynamics within transitioning cells.[67] Since single-cell omics data are static and only captures a snapshot of the underlying biology at a single point in time. Therefore, computational methods have been developed to determine the progression of cells along a trajectory in pseudotime and space. A popular method for inferring cell lineage and pseudotime from single-cell transcriptomics data are Slingshot.[68] Each of these computational methods discussed for inference of cell state transitions, interactions, and trajectories can be used collectively to extract features from the single-cell omics data for structuring and parameterizing computational models that can be simulated.

UTILIZING "BIG" AND "NOT SO BIG" DATA FOR MODEL BUILDING

Single-cell omics methods for inference of cell types, cell states and their distributions, state transitions, state trajectories, and interactions provide a perspective into biological processes occurring at various scales including the tissue, microenvironment, cellular, and subcellular level. This opens up a unique opportunity for biologists to utilize the single-cell information for structuring and parameterizing dynamic computational models of various biological systems. Conventionally, sophisticated computational models have been built using "Not So Big" data sources. An early example of this type of modeling is the classical GRN model of the sea urchin developmental process.[69] The model was generated using specific targeted and directed experiments to perturb the regulatory genes and alter the signaling process of the developmental pathway. Subsequent quantitative PCR (qPCR) was performed and utilized for constructing the GRN. Additional early GRN models that utilized "Not So Big" data sources for model construction and parameterization include that from Novak and Tyson (1993), who constructed a model of cell cycle control in *Xenopus* oocytes,[70] and Forger and Peskin (2003), who generated a model of the mammalian circadian clock.[71] The process of constructing and generating GRN's has drastically changed with the unprecedented rise in omics technologies and the availability of publicly available omics datasets. Now, correlative GRN models have been generated using omics data as many algorithms have streamlined the process of single-cell data analysis to GRN

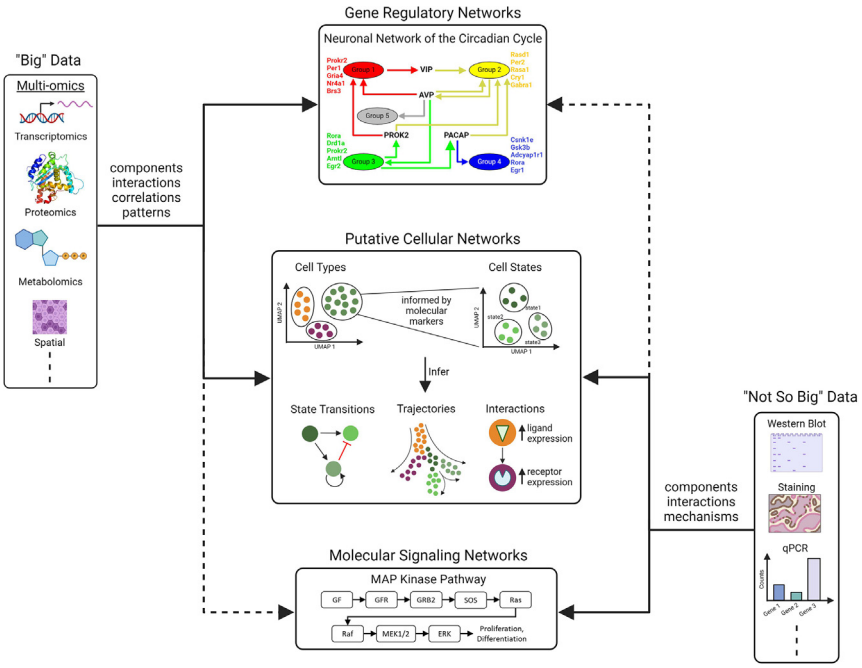

Figure 3. Computational models informed by experimental data

The components, interactions, correlations, and patterns extracted from "Big" data (multi-omics data including transcriptomics, proteomics, metabolomics, and spatial omics) and the components, interactions and mechanisms extracted from "Not So Big" data (i.e., western blots, immuno-staining, and qPCR) can be utilized to generate and inform molecular signaling networks, putative cellular networks, and gene regulatory networks. For instance, while the MAP kinase pathway was discovered using "Not So Big" data sources (solid line) many "Big" data sources (dashed line) have confirmed and further explained and complemented these initial findings. Similarly, while gene regulatory networks have been mainly developed using "Big" data (solid line), "Not So Big" data (dashed line) can also be informative when generating such networks. For example, Park et al., (2016) modeled neurons during the circadian cycle.[13] First, five neuronal groups were identified according to their unique transcriptional landscapes with marker genes shown for each of the groups. A gene regulatory network was then developed based on the major molecular interactions between key neuropeptides (VIP, AVP, PROK2, and PACAP) and the neuronal groups.[13] "Big" and "Not So Big" data (solid lines) have be analyzed in combination to identify putative cellular networks. Cell types can be identified within the "Big" data by using information from "Not So Big" data. Then, cell states within each cell type community can be determined by molecular markers. The cell types and states can then be used to infer cell state transitions, trajectories, and interactions. A greater influence of "Big" and "Not So Big" data on developing the various networks is shown with solid lines with lesser influence shown by dashed lines. GF: growth factor, GFR: growth factor receptor, VIP: Vasoactive Intestinal Peptide, AVP: Arginine Vasopressin, PROK2: Prokineticin 2, PACAP: Pituitary Adenylate Cyclase-Activating Polypeptide. Fig. adapted from Park et al., (2016).[13] Created using biorender.com.

model construction and simulation[72–74]; a subset of which are discussed in more detail futher.

While molecular signaling networks, such as the MAP kinase pathway, have classically been informed by the components, interactions, and mechanisms elucidated from "Not So Big" data, the components, interactions, correlations, and patterns from "Big" data have provided an additional cell-type specific level of detail to such network models (Figure 3). Both "Big" and "Not So Big" data can be collectively analyzed to extract and identify putative cellular networks. These cellular networks can be elucidated by identifying cell types and cell states, which are informed by molecular markers. Cell state transitions, trajectories, and interactions can then be inferred from the cell types and cell states to further understand cellular heterogeneity and the organization of the tissue's microenvironment (Figure 3). GRN's are also being developed using "Big" data in which gene-gene interactions are inferred from correlations and patterns within the omics data. However, these networks only contain structural details and are not utilized for dynamic simulations. Therefore, we instead choose to focus on GRN's that have been generated for simulating the dynamic behavior of a biological system. For instance, a neuronal GRN of the circadian cycle was developed in Park et al., (2016) to explore how neurons with asynchronous behavior and transcriptional heterogeneity generate a coordinated response that synchronizes the body to the circadian cycle[13] (Figure 3). We propose that the information gained from the molecular signaling networks, putative cellular networks, and GRN's can then be utilized collectively to parameterize and structure multiscale pathophysiological computational models with spatial resolution.

BUILDING COMPUTATIONAL MODELS INFORMED BY SINGLE-CELL OMICS DATA

Computational models based on ODE formalizations are commonly utilized to describe the dynamic behavior of biological systems and consist of a system of equations used to describe the quantity or concentration of different biological species over time.[75,76] While ODE-based models can capture the detailed dynamics of biological species, the models have been classically limited by the breadth of knowledge and available literature detailing, which biological species interact, proliferate, and transition. Further, many assumptions are made, and simplifications are introduced when parameterizing the model with specific kinetic descriptions. Boolean modeling techniques are also frequently utilized in the biological context.[75,77] Boolean models are constructed by using binarized variables in either an on or off state representing gene/protein/metabolite activation or inhibition. The variables can be combined to form logic rules such that simulations of the transitions between the on and off state are possible. Boolean models require less parameterization, making them simpler to construct and analyze, however, this modeling technique is limited as it neglects to model the intermediate states between on and off. Additionally, neural

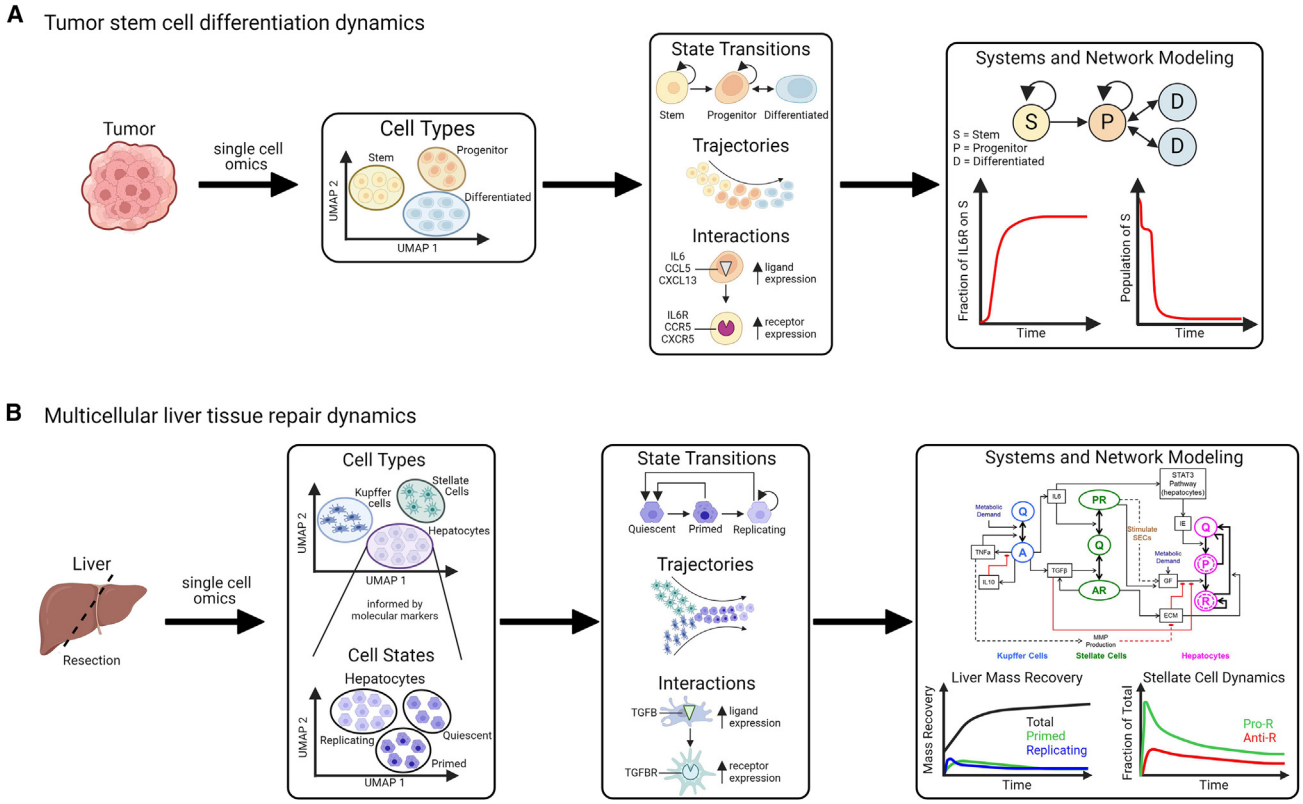

Figure 4. Computational models informed by single-cell omics

(A) Single-cell omics, including transcriptomics, proteomics, and metabolomics can be used for modeling tumor cell differentiation dynamics. The specific cell types of interest that were identified within the tumor tissue include stem, progenitor, and differentiated cell types. State transitions, trajectories, and interactions between these cell types can then be inferred such that a network model can be generated. The tumor cell differentiation model can then be simulated to determine how the individual cell populations within the tumor change over time.

(B) Single-cell omics experiments can be performed on the liver following resection to elucidate liver-specific cell types including Kupffer cells, Stellate cells and hepatocytes. For simplicity, we only show the hepatocyte cell states (replicating, quiescent, and primed), which are informed by molecular markers from the single-cell data. State transitions, trajectories and interactions can then be inferred from the cell states. A systems network model of liver regeneration can then be developed using the features extracted from the single-cell data and the model can be simulated for liver mass recovery and cellular dynamics. The total mass recovery as well as the populations of primed and replicating hepatocytes populations during regeneration are shown. Additionally, the populations of pro- and anti-regenerative stellate cell populations during regeneration are shown. Figure 4A adapted from Nazari et al., (2018).[80] Figure 4B adapted from Cook et al., (2018).[81] Created using biorender.com.

networks and other machine learning algorithms have been developed for analyzing, modeling, and simulating various biological processes.[78,79] Neural networks consist of an input layer, one or more hidden layers, and an output layer where each neuron is connected to one another and holds a particular weight and threshold. The network learns by examining individual records from a training set, generating a prediction for each record, and then adjusting the weights when a prediction is incorrect. This process is repeated until the network improves its predictions to meet the desired stopping criteria. While neural networks can be highly automated and adaptable to various data types, they require an abundance of high-quality data for training and risk being overfit i.e., the model performs well on the training data but poorly on the testing data. In all these cases, computational modeling efforts may use single-cell data in a multitude of ways to elucidate biological mechanisms at various levels, including the subcellular, cellular, microenvironment, tissue, and overall physiological scale. In each case, the model structure and parameterization can be informed and constrained by both "Big" and "Not So Big" Data (Figure 3). In this section, we discuss how "Big" data can be utilized in computational modeling of pathophysiology.

Computational models informed by single-cell transcriptomic data are most common, as data sampling and analysis methods are the most widely utilized and accepted. However, high-dimensional proteomics and metabolomics data sources can also be utilized for model building and simulation. For example, single-cell omics, including transcriptomics, proteomics, and metabolomics can be used for modeling and simulation of tumor cell differentiation dynamics, as presented in Nazari et al. (2018)[80] (Figure 4A). First, the various cell types and states are identified in the data such that cell state transitions, trajectories, and interactions can be extracted. In this example, the specific cell types of interest that were identified within the tumor tissue

include stem, progenitor, and differentiated cell types. In this ODE-based computational model, stem cells can transition to progenitor cells which can then transition to differentiated cells. Additionally, both stem and progenitor cells can replicate, and the transition from a progenitor to differentiated cell is reversible. The predicted cellular trajectory begins with stem cells progressing to progenitor cells and finally to differentiated cells. Increased expression of ligands such as IL6, CCL5, and CXCL13 in progenitor cells and increased expression of their receptors IL6R, CCR5, and CXCR5 in stem cells would allude to ligand-receptor interactions between the stem and progenitor cells. A network model can then be constructed based on the analyzed single-cell data. Specifically, a single stem cell can transition to a progenitor cell, which can transition to two differentiated cells. The model can then be simulated to determine how the individual cell populations within the tumor change over time. For example, the fraction of IL6R on stem cells increases over time until a steady state is reached, while the total population of stem cells decreases toward a steady state of zero. The network model can be extended in multiple ways based on results from additional single-cell transcriptomics, for example, toward a more comprehensive characterization of tumor microenvironment with multiple cell types and their states, state transitions and interactions.

In a recent study, a group utilized a similar pipeline to that presented in Figure 4A to generate a multiscale model of epithelial-to-mesenchymal transition (EMT) using scRNA-seq data.[82] Prior to modeling, the researchers first detected transition cells and then subsequently identified intermediate cell states (ICS) such that transition trajectories could be constructed. The multiscale agent-based model of EMT was extended from previous models to include the heterogeneous ICS populations identified from the single-cell data to determine how the fate of this intermediate cell state may be affected. Following the EMT model simulation, the authors uncovered the roles of ICS on adaptation, noise attenuation, and transition efficiency in EMT.

Single-cell omics can also be utilized to develop more specific computational models of biological processes such as that of the liver during regeneration (Figure 4B).[81,83] For example, scRNA-seq time series data have been collected following 70% liver resection in mice as well as from mice exposed to acetaminophen (acute liver injury).[83] The data on cell types and states, as well as predicted cell state transitions from such time series omics experiments can provide rich information on the dynamics of the molecular regulatory networks operating in the liver tissue microenvironment during liver repair and regeneration. Integration of such high-resolution single-cell data and statistically predicted state distributions and transitions into an already established model of liver regeneration[81] can be performed to add more detail on the cell states, their transition kinetics, and cell-cell interactions within the tissue-scale network model. For instance, in the previously published ODE-based network model of liver regeneration,[81] Kupffer cells are considered as distributed between quiescent and active states, whereas hepatic stellate cells are distributed between quiescent, pro-regenerative, and anti-regenerative states, and hepatocytes are modeled across quiescent, primed, and replicating states. For illustrative purposes, we focus on the hepatocyte cell states and their inferred transitions in Figure 4B. Hepatocyte state transitions within the systems network model include quiescent to primed to replicating states. While replicating hepatocytes can proliferate and transition back to the quiescent state, primed hepatocytes can only transition back to the quiescent state. The Kupffer and stellate cell populations show separate trajectories toward the hepatocyte cell population, which exhibit a progression from the quiescent to primed to replicating state. As an example of receptor-ligand interactions within the liver during regeneration, Kupffer cells with high expression of TGFB ligand and hepatic stellate cells with high expression of the TGFB receptor are likely to interact within a tissue microenvironment. The resulting systems network model would then include the various Kupffer, stellate, and hepatocyte cell states, their state transitions as well as the molecular interactions between these cell states. The model can then be simulated to elucidate the proportion of relative cell types (i.e., stellate cells that are pro vs. anti-regenerative) and their effect on overall liver mass recovery and repair (Figure 4B).

While this section focused on computational models informed by single-cell omics data, the next section discusses the advantageous implications of incorporating spatially resolved omics data into computational models.

INCORPORATING SPATIAL INFORMATION INTO SINGLE-CELL OMICS-INFORMED MODELS

Single-cell spatial omics data can provide computational models with further details on tissue microenvironments, cellular proximities, and paracrine interactions in additional to the cell state transitions, interactions and trajectories, which can be inferred from non-spatial data. Single-cell and spatial multi-omics data from heterogeneous tumor samples can be utilized collectively to develop multiscale cell-biology-based tumor models (Figure 5A). As in Figure 4, cell types, cell states, state transitions, trajectories, and interactions can be elucidated from the single-cell data. However, with the addition of spatial omics data, individual spatial features can be extracted, and the heterogeneity of the cell population within the tumor can be further uncovered. Interactions within the tumor microenvironment, such as cellular proximities and paracrine interactions, can be elucidated within a spatial region of interest. These details can then aid in the parameterization and structuring of either graph-based or agent/partial differential equation (PDE)-based network models. Graph-based models implicitly incorporate the spatial organization of a tissue and depend on an adjacency matrix that determines which cells interact with one another. Then, a system of ODEs can be solved for the state of an individual cell with respect to its interactions with adjacent cells. An example of a spatial, graph-based network model is that of resistant tumor clones, as presented in Waclaw et al., (2015),[84] which shows that the size of the tumor increases over time until treatment begins (Figure 5A). Once treatment has been initiated, the tumor size significantly decreases, but resistant clones are left behind. The resistance clones can then continue to proliferate despite the patient continuing treatment, resulting in a growing tumor that may increase in size beyond that of the tumor before treatment. The

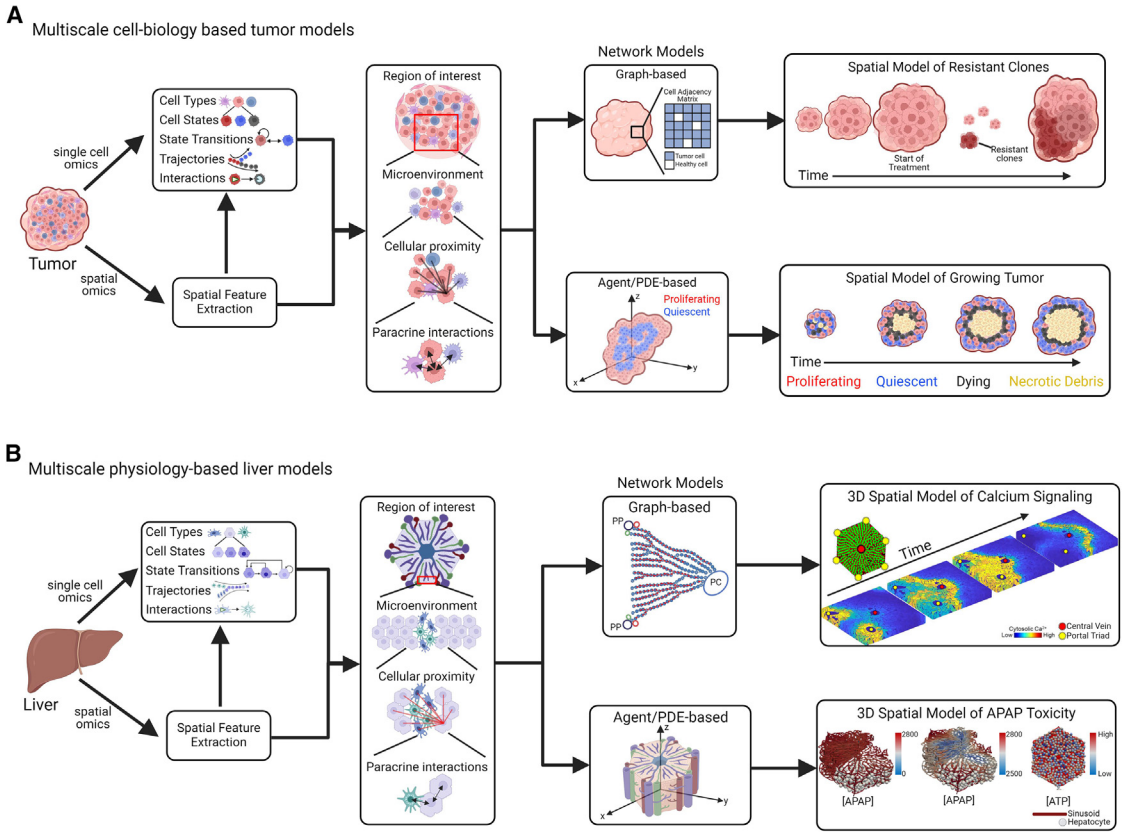

Figure 5. Computational models informed by single-cell and spatial omics

(A) Cell types, cell states, state transitions, trajectories, and interactions can be inferred from single-cell omics data from a tumor sample. Additional spatial omics profiling of the tumor sample allows for extraction of spatial features that provide an additional level of detail to the model. Spatial information from a region of interest in a tissue sample can elucidate information about the microenvironment, cellular proximities, and paracrine interactions. Graph-based modeling can be utilized for instance, when modeling resistant clones within the tumor over time.[84] Agent or PDE-based modeling techniques also may be used for spatial modeling, such as that of heterogeneous growing tumor over time.[85]

(B) Single-cell and spatial omics data of the liver also permits extraction of cellular and spatial features as in Figure 5A. Graph-based models may then be developed for simulating a model of hepatic calcium signaling,[86] while agent/PDE-based models have been developed for modeling acetaminophen (APAP) toxicity in the liver lobule.[87] Figure 5A adapted from Waclaw et al., (2015)[84] and Jagiella et al., (2016).[85] Figure 5B adapted from Dichamp et al., (2023).[87] Created using biorender.com.

spatial information from the omics data can provide a level of detail to the model such that the tumor microenvironment and interactions between resistant and non-resistant cells can be simulated. Agent/PDE-based modeling explicitly accounts for the spatial organization of the cells (or agents) within the three-dimensional model. An example of a spatial, agent-based model is that of a growing tumor, as in Jagiella et al. (2016).[85] As the size of the tumor increases, the distribution of various cell populations changes; there is an increase in necrotic debris and quiescent cells compared to proliferating cells, which contribute more to the tumor in the initial stages.

A similar methodology can be utilized to generate multiscale physiology-based models of the liver. For instance, spatial features on the microenvironment, cellular proximities, and paracrine interactions can be extracted from liver-specific spatial omics data, while cell types, states, transitions, trajectories, and interactions can be obtained from the liver-specific single-cell data. Then, graph-based and agent/PDE-based network models can be generated and parameterized using the features extracted from the data. An example of graph-based modeling is present in Verma et al. (2018), in which a 2D ODE-based model of calcium propagation across a liver lobule was developed by utilizing an adjacency matrix.[86] In this model, calcium molecules form waves, which begin at the portal triad and terminate at the central vein of the liver lobule. This model was also integrated with a model of glucose metabolism in Verma, Manchel et al. (2021).[88] Since hepatic glycogenolysis is governed by metabolic zonation of the liver lobule, single-cell omics data can be utilized for modeling the molecular interactions between cells. This model can be extended into three dimensions using spatial omics data such that the microenvironment and cellular interactions along the zonated liver lobule can be simulated. Model simulations are consistent with those from the 2D model in that the calcium waves still progress from the portal triad to the central vein (Figure 5B). An example of a PDE-based model utilizing spatial omics data are that of hepatic acetaminophen (APAP)

toxicity. Figure 5B shows the distribution of the toxin throughout the liver lobule after a 281 mg/kg dose of APAP.[87] The APAP concentration is highest in the portal region and lowest in the central region of the sinusoid. The portal hepatocytes, however, already begin to eliminate a fraction of the APAP before reaching the central region. Additional spatial models of the liver have been generated for several functions, including regeneration following CCl4 intoxication,[89] regeneration following resection,[90] and various disease states, including non-alcoholic steatohepatitis.[91] Future implementations of these models may utilize spatial transcriptomics to incorporate additional detail of the liver microenvironment.

As an illustrative example of spatial modeling, consider PDE-based spatially resolved model of liver fibrosis introduced by Friedman and Hao 2017.[92] The model was utilized to explore the efficacy of currently available drugs for treating liver fibrosis. PDE-based spatially resolved models are beneficial for explicitly modeling and simulating the 3D interactions between cells and molecules in a user-defined space, while graph-based methods are limited by the bounded space of the adjacency matrix. A model that utilized the graph-based spatial modeling approach previously discussed but applied to a model of embryonic development is present in Cang et al. (2021).[93] They developed a multiscale spatial model of the mammalian embryo using scRNA-seq data and single-cell qPCR data to identify robust patterning mechanisms early in development.[93] Their model utilized gene regulatory, cell-cell communication, and physical interaction information from single-cell transcriptomics data. Using the generated model, the authors identified two processes critical to the formation of the later epiblast and primitive endoderm, namely, a cell-cell adhesion mechanism and a temporal attenuation mechanism. In an additional model of the liver, ADViSOR, the spatiotemporal molecular underpinnings of hepatic regeneration were studied using time-interval principal-component analysis and sliding dynamic hypergraphs.[89] This novel modeling methodology was applied to spatial transcriptomics data assayed serially through the liver regeneration process post partial hepatectomy in mice and was utilized to identify key functional gene modules of cell signaling and cell-cell interactions.[94]

Various algorithms have been developed to generate computational models informed by single-cell and spatial omics data. These computational models may be generated using different methodologies and formalisms including ordinary differential equations (ODE's), Boolean networks, neural networks, etc. based on the desired functionality of the model. For example, SCODE was developed to ease the integration of scRNA-seq data with dynamic models of cellular function to infer regulatory networks and describe expression dynamics with ODE's.[74] However, one limitation of SCODE is that it does not consider the zero-inflation problem with scRNA-seq data, which may affect the parameter tuning of the model, leading to false modeling predictions. In an additional modeling algorithm, scBONITA infers Boolean regulatory rules and logical gates for developing executable network models from single-cell omics data.[90] The methodology is limited by the requirement of a prior knowledge network. In Table 5, we highlight these computational algorithms as well as several others that can be utilized for building and simulating models that elucidate functional cell-cell interactions and communication within heterogeneous tissues.

SCALING TO MULTI-ORGAN, PHYSIOLOGY-BASED MODELS INFORMED BY SINGLE-CELL OMICS

Single-cell and spatial omics data may also provide a powerful means for developing and parameterizing multi-organ, physiology-based models. Publicly available omics databases, such as HuBMAP,[10] Tabula Sapiens,[11] The Human Cell Atlas,[9] and SPARC,[103] have been developed for a wide range of tissue and organ types and can be utilized for developing multi-organ network models with organ-specific cellular and molecular interactions (Figure 6A). Classically, our understanding of the physiology of organs and tissues has come from "Not So Big" data. This knowledge can be supplemented with information from publicly available cell atlases which provide single-cell transcriptomics, proteomics, and metabolomics data, as well as other databases that provide physiological data on human tissues from a diverse background of patients. Once the data have been collated from their individual sources, organ-specific cell types can be further investigated. For illustrative purposes, we highlight a few cell types that may be extracted from the clusters of brain (neurons, microglia, and astrocytes) and heart (neurons and cardiomyocytes) cells. Once cell types have been identified for the organs of interest, the functionality of each cell type/state can be deciphered based on the average across patient samples from the cell atlases. Then, generic cellular and molecular interaction network models can be developed and integrated to build a higher-level multi-organ model with cellular interactions (Figure 6A). Within the brain module of the multi-organ model, cellular interactions exist between astrocytes, neurons, and microglia and within the neuronal cell submodule, there exist molecular interactions between various cytokines, i.e., TNFa, TGFb, and IL-10. Organ models interact with each other in one of two ways: via the blood supply or neural input. While the brain can interact with all other organs through neural innervation, the heart can interact with all organs through the blood supply. A similar methodology is proposed in Anderson and Vadigepalli (2016) for modeling cytokine regulatory network dynamics driving neuroinflammation in the central nervous system.[18] Illustrative studies that systematically use "Big" and "Not So Big" data from end to end for structuring, parameterizing and tuning high-fidelity multiscale, multiorgan models of pathophysiology are hard to find in the existing literature. Here, we discuss a framework for building such models by harnessing the data resources of multiple scales.

Scaling further toward multiple organs, Thiele et al. (2020) developed a whole-body metabolic model, which captured the metabolism of 26 organs and 6 blood cell types.[106] The model was parameterized using physiological, dietary, and metabolomic data and could recapitulate known inter-organ metabolic cycles and energy use. Similarly, Zhang et al. (2020) developed a method that utilizes multi-organ single-cell data from mice to systematically simulate cellular metabolism through constraint-based, context-specific, genome-scale metabolic modeling (GEM).[107] The Tabula Muris scRNA-seq dataset, which includes data from nearly 100,000 cells from 20 organs and tissues in

Table 5. Highlighted algorithms for computational modeling informed by single-cell and spatial omics data

Algorithm Name	Modeling Type	Input Data Type/s	Algorithm Features
SCODE[74]	ODE	scRNA-seq	To infer regulatory networks and describe expression dynamics
GRISLI[95]	ODE	scRNA-seq	To infer regulatory networks and describe expression dynamics
scGNN[96]	Neural Network	scRNA-seq	To provide a hypothesis-free deep learning framework for modeling heterogeneous gene expression patterns and cell-cell relationships
scFEA[97]	Graph Neural Network	scRNA-seq	To estimate cell-wise metabolic flux, metabolic stress, and the effect of metabolic gene perturbances
DeepVelo[98]	ODE/Neural Network	scRNA-seq	To formulate transcriptome dynamics on different time scales, measure the instability of cell states, and identify developmental driver genes via perturbation analysis
scBONITA[99]	Boolean	scRNA-seq	To infer executable dynamic pathway models and perform perturbation analyses to identify high impact genes
SCNS[73]	Boolean	Time course scRNA-seq/qPCR	To reconstruct and analyze executable models that drive cell transitions and make predictions about the effect of gene perturbations
CellNOptR[72]	ODE/Boolean	Single-cell proteomics	To build predictive logic models of signaling networks
SOTIP[100]	Graph-based	Spatial transcriptomics/ proteomics/metabolomics	To generate a unified graph of tissue microenvironments and their interactions to quantify spatial heterogeneity, identify the spatial domain, and perform differential microenvironment analysis
SPACE-GM[101]	Neural Network	Spatial proteomics	To model tumor microenvironments as local subgraphs to capture the distinctive cellular interactions associated with different clinical outcomes
MISTy[102]	ODE/Random Forest Machine Learning Model	Spatial transcriptomics/ proteomics/metabolomics	To extract relationships about cell-cell interactions and communication in a tissue microenvironment at various levels of resolution (within the local cellular niche and at the level of tissue structure)

mice,[108] was utilized for context-specific modeling of each tissue type and cell ontology class and subsequent metabolic flux prediction. RNA-seq profiles of a selected group of cells are then converted into gene expression confidence scores and assigned to each gene to represent the likelihood that the gene is expressed in that group of cells. The authors found that simulation of NAD+ biosynthesis activity in 7 different mouse tissues, namely the heart, liver, brain, kidney, lung, skeletal muscle, and spleen, showed a significant linear correlation with experimental measurements.

A recent study also highlights the power of multi-organ, multiscale modeling informed by single-cell omics and physiological data[105] (Figure 6B). Briefly, single-cell qPCR data from the nucleus tractus solitarius (NTS), a region of the brainstem responsible for integrating sensory information, was collected and used to define cellular subtypes in a model of the NTS.[109,110] At the physiological level, neurons project from the NTS to the dorsal motor nucleus of the vagus (DMV) and nucleus ambiguous (NA), both of which then project to the intrinsic cardiac nervous system (ICN) at the heart. The ICN then integrates sensory information and regulates beat-to-beat contractions of the heart (Figure 6B). Recently, our group proposed the concept of cardiovascular health controlled by two differing sources: DMV input mediated by muscarinic receptors that activate over 10's of seconds ("slow" lane) and NA input mediated by nicotinic receptors that activate in milliseconds ("fast" lane). Subsequently, a multi-organ cellular and physiologically informed network model of the fast and slow lanes of cardiovascular control was developed to predict the influence of the two lanes on heart health.[104,105] The closed-loop model comprises multiple organs, including the brain, heart, lungs, and vasculature, thus, the organs, their subregions of interest, and the organ-specific cell types can

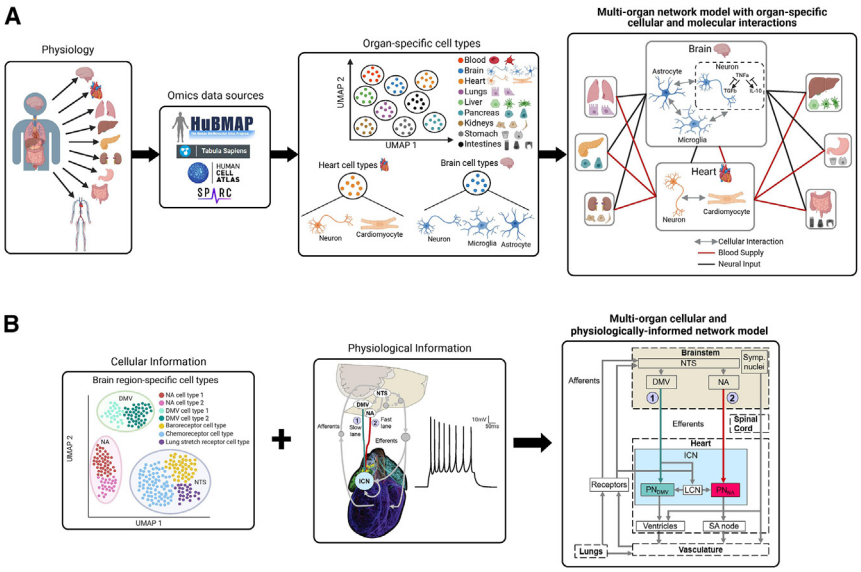

Figure 6. Computational models informed by physiology

(A) A multi-organ model can be developed by first gathering physiological information from the organs of interest. Publicly available databases can be utilized to extract single-cell information for each organ. Processing the single-cell data for each organ permits identification of organ-specific cell types of interest. Molecular interactions from the organ-specific cell types can be inferred from the data sources. A multi-organ network model with organ-specific cellular and molecular interactions can then be generated.

(B) Cellular and physiological information can be integrated to generate a multi-organ cellular and physiologically informed network model of the "fast" and "slow" lanes of cardiovascular control.[104] Single-cell data can be obtained for brain regions of interest and electrophysiology experiments on neuronal responses to perturbations can be performed. A model of neural control of the heart can then be generated with information on multi-organ and cellular interactions derived from the performed experiments. NTS = nucleus tractus solitarius, DMV = dorsal motor nucleus of the vagus, NA = nucleus ambiguus, Symp. = sympathetic, LCN = local circuit neurons, PN = principal neurons, SA = sinoatrial. Figure 6B adapted from Gee et al., (2023)[104] and Gee et al., (2023).[105] Created using biorender.com.

communicate with one another via neural input as well as the blood supply.

A study from Nandi et al. (2022) utilized single-cell transcriptomics data to validate a library of single cortical neuron electrophysiology models developed using morphology and electrophysiology data.[111] Individual cell types were identified using an approach that integrated transcriptomic, morphological, and electrophysiological data. Subsequently, single-cell models of voltage-gated ion channels were developed and simulated. The authors found that the differences in model conductance explained the electrophysiological discrepancies observed between the cortical cell subclasses. In an additional study, a computational model of Purkinje fiber single-cell electrophysiology was developed utilizing collected experimental data from human voltage-gated channels.[112] Model simulations were able to recapitulate the unique electrical properties of the Purkinje fibers, and simulating selective ion channel blockades reproduced responses to pharmacological challenges characteristic of isolated Purkinje fibers *in vitro*.[112] While these models with physiological detail do not explicitly incorporate multi-organ interactions, they can be assembled with higher level generic multi-organ models to develop multi-organ models with neural specificity.

PATIENT-SPECIFIC MODELING, DIGITAL TWIN MODELS, AND IN-SILICO CLINICAL TRIALS

Computational models are utilized to study biological systems in health and disease contexts and elucidate the heterogeneity and variability across patients within a given disease state by incorporating patient-specific omics data. For instance, Verma et al. (2019) shows that there exists a wide range of variability in the liver's ability to regenerate following partial resection.[113] This suggests the need for personalized modeling and simulation to predict the regulatory and functional behavior of tissues in health and disease in a patient-specific manner.

Genome-scale metabolic models (GEMs) provide a unique and necessary modeling framework for developing context and patient-specific models informed by single-cell multiomics data.[114,115] The metabolic functionalities of the models can then be assessed to further understand how metabolic pathways may be dysregulated with disease. Several software suites, including MATLAB's cobra and RAVEN toolboxes, provide all functions necessary for modeling, analyzing, and predicting GEM phenotypes.[116,117] A generalized GEM, namely Human1,[114] Recon2,[115] etc., may be integrated with both liver-specific transcriptomics and proteomics data from a patient such that a GEM specific to the individual patient's liver can be generated (Figure 7A). For this purpose, Gustafsson et al. (2022) developed fast tINIT (ftINIT), a method for generating context-specific GEMs from pools or clusters of scRNA-seq profiles.[12] Using this methodology, the Human1 metabolic network is minimized by solving for the gene-protein reaction (GPR) rules using the supplied scRNA-seq data. The network is optimized using mixed integer linear programming (MILP), and the gaps in the network are filled by providing a list of essential metabolic tasks. Following the generation of a context-specific GEM, metabolic fluxes can be predicted by flux balance analysis (FBA). When solving for the metabolic fluxes, FBA assumes that there is no change in metabolites over time (i.e., the system is at steady state), despite the dynamic fluctuations of metabolites seen in nature. A user-set optimization function is then optimized, and the fluxes can be further bounded by experimentally observed constraints such as collected liver metabolomics data (Figure 7A). Given the strict steady state assumption in FBA, Chandrasekaran et al. (2017) developed an in-house systems approach to relax the assumption, resulting in metabolic fluxes that can be solved dynamically.[118] The algorithm assumes that

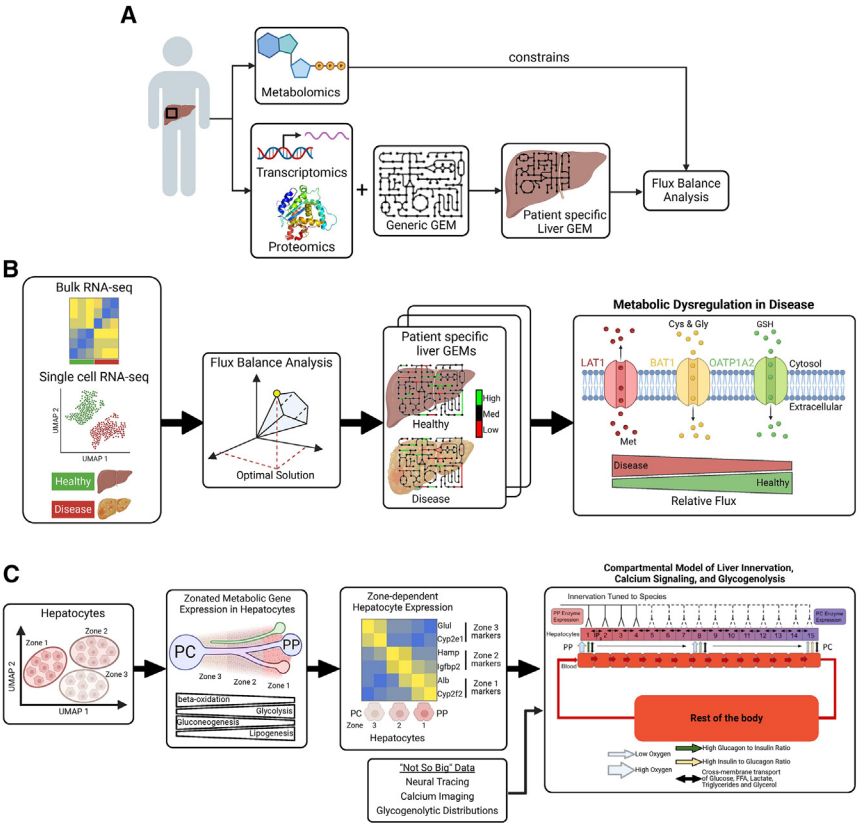

Figure 7. Patient-specific models informed by omics data

(A) Metabolomics, transcriptomics, and proteomics data can be collected from a patient's liver sample. A patient-specific genome scale metabolic model (GEM) of the liver can then be generated by integrating the transcriptomics and proteomics data with a generic GEM (i.e., Human1[9] or Recon2[115]). Metabolic fluxes are constrained using the metabolomics data and predicted by flux balance analysis.

(B) Bulk and single-cell RNA-seq data can be utilized to generate context-specific metabolic models in health and disease (i.e., liver disease). Metabolic fluxes can be predicted by flux balance analysis and significantly perturbed metabolic pathways/subsystems can be identified in health vs. disease. For example, our analysis of liver transcriptomics data from alcoholic liver disease identified significant metabolic dysregulation in the glutathione (GSH) metabolic pathway. Specifically, the metabolic flux activity of specific solute transporters (LAT1, BAT1, OATP1A2) within the GSH pathway decreased with liver disease, while healthy livers showed an increase in flux along the pathway.

(C) Zone-specific hepatocyte populations can be elucidated from single-cell omics data sources. The metabolic expression for genes in the B-oxidation and gluconeogenesis pathways decreases from zone 3 to zone 1, while it increases from zone 3 to zone 1 for genes in the glycolysis and lipogenesis pathways. Marker expression for each of the zonated hepatocyte populations within the "Big" data can be utilized in conjunction with "Not So Big" experimental data (i.e., neural tracings, calcium imaging, and glycogenolytic distribution analyses) to parameterize and structure a computational model of liver innervation, calcium signaling, and glycogenolysis. Additionally, the extent of innervation to the liver can be tuned in the model to the species of interest based on physiological evidence from the literature. Figure 7B adapted from Manchel et al., (2022).[119] Figure 7C adapted from Verma, Manchel et al., (2021).[88] Created using biorender.com.

the rate of change of each metabolite in the FBA equation is proportional to the measured rate of change of the metabolites from time course metabolomics data.[118]

A semi-automated tool that utilizes a similar methodology to that of Figure 7A is XomicsToModel.[120] This tool integrates transcriptomic, proteomic, and metabolomic data with a generic GEM to extract a context-specific model that is stoichiometrically, thermodynamically, and flux consistent.[120] The enzyme activity within the metabolic model is computed using the proteomics and transcriptomics data, while the metabolite constraints are obtained from the metabolomics data. In Ambikan et al. (2022), a similar approach was also used to develop patient-specific COVID-19 GEMs.[121] The GEMs were generated by integrating whole blood transcriptomics data with a generic GEM, and subsequent FBA was performed by constraining the model's exchange reactions with plasma metabolomics data.

Our group utilized a variation of ftINIT, namely tINIT,[122] which integrates bulk omics data with the generic Human1 metabolic model to generate context-specific metabolic models. We developed liver disease-state-specific metabolic models using publicly available bulk RNA-seq data from the livers of patients with liver disease.[119] Subsequently, the SPOT algorithm was utilized to calculate a suitable objective function by maximizing the Pearson correlation between a flux vector and its corresponding gene expression data.[123] Next, E-Flux2 was used to perform standard FBA and minimize the Euclidean norm of the flux vector such that a unique solution is found.[123] This methodology can be applied to single-cell transcriptomics data to generate context-specific metabolic models of specific cell types/states and the plausible flux solution space is constrained according to the metabolic gene expression in these cell types/states (Figure 7B). Changes in predicted metabolic fluxes can be compared across cell types/states to identify significantly perturbed metabolic pathways/subsystems in health vs. disease. For example, our analysis of liver transcriptomics data from alcoholic liver disease identified significant metabolic dysregulation in the glutathione (GSH) metabolic pathway. Specifically, the metabolic flux activity of specific solute transporters (LAT1, BAT1, OATP1A2) within the GSH pathway decreased with liver disease, while healthy livers showed an increase in flux along the pathway (Figure 7B).

Previously developed models can be informed by "Big" and "Not So Big" data resulting in models with patient-specificity. For example, a computational model of liver innervation, calcium signaling, and glycogenolysis, as described in Verma, Manchel et al. (2021),[88] can be further parameterized and structured by

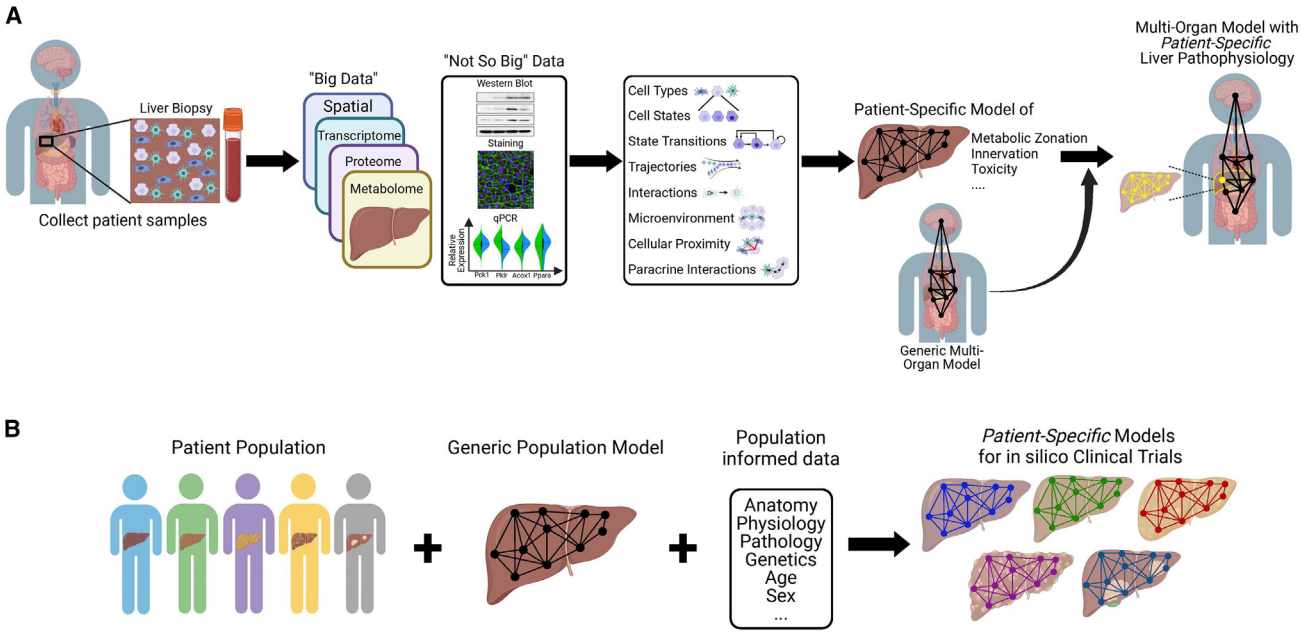

Figure 8. Patient-specific modeling for the clinic and in silico clinical trials

(A) Patient-specific blood and liver tissue samples can be collected and various "Big" and "Not So Big" data can be sampled such that cell types, cell states, state transitions, trajectories, and interactions can be elucidated from the non-spatial data and information about the microenvironment, cellular proximities, and paracrine interactions can be elucidated from the spatial data. Patient-specific liver models can then be developed for a multitude of different functions such as metabolic zonation, innervation, or toxicity. The patient specific liver model can then be integrated with a generic multi-organ model such that a patient-specific, multi-organ model with liver specificity is generated.

(B) Data can be collected from the livers of various patients within a population. These data can then be integrated with a generic population model of the liver and population-informed data such as organ anatomy, physiology, pathology, genetics, age, and sex such that the patient-specific models can be utilized for in silico clinical trials. Created using biorender.com.

patient-specific omics data (Figure 7C). Liver biopsies can be taken from a range of individuals and subsequent single-cell RNA-seq seq can be performed. Dimensionality reduction of the data will reveal inter-patient variability, but the hepatocytes will continue to cluster together based on their location within the liver lobule. Cells nearest the periportal region are termed zone 1, cells in the mid-lobular region are zone 2, and cells nearest the pericentral region are zone 3. Further analysis of the differentially zonated single-cells will show decreasing expression of beta-oxidation and gluconeogenesis-related genes from zone 3 to zone 1, consistent with experimental data. Oppositely, there will be increasing expression of glycolysis and lipogenesis-related genes from zone 3 to zone 1. Specific genes markers for the three zones can be determined and these gene expression values may be utilized for tuning the patient-specific parameters in the model. For example, Glul and Cyp2e1 peak in zone 3, Hamp and Igfbp2 peak in zone 2, and Alb and Cyp2f2 peak in zone 1, however, there exists variability across cells. The computational model can then be generated by combining the information gained from single-cell omics ("Big") data with pre-existing "Not So Big" data resulting from neural tracing, calcium imaging, and glycolytic distribution experiments. While the "Not So Big" data are important for structuring the model, the "Big" data are useful for model parameterization, which can be adjusted after collecting and analyzing new patient-specific data.

While it would be advantageous to collect or have access to multiple omics data sources for a single patient to bring deeper specificity to an individualized patient model, this is usually infeasible and impractical in the clinical setting. Therefore, we propose a way forward to generate organ-specific models that are embedded in a broader, more generally applicable manner that can be tuned to individual patients. Multi-organ models with patient-specific organ specificity can also be generated utilizing both "Big" and "Not So Big" data, as in Figure 8A. For instance, if one were to develop a patient-specific multi-organ model with liver specificity, one could utilize the vast range of publicly available patient-specific multiomics data (transcriptomics, proteomics, metabolomics, spatial). In addition to collating the "Big" data, "Not So Big" data (i.e., western blots, immunostaining, and qPCR experiments) can be extracted from relevant literature to enhance our understanding of the liver-specific cell types, cell states, state transitions, state trajectories, and state interactions. Additional spatial information about the tissue microenvironment, cellular proximities, and paracrine interactions can be extracted from spatial transcriptomics datasets to generate a model of liver functionality specific to metabolic zonation, innervation, toxicity, etc., depending on the purpose of the study. This model can then be brought to the patient-specific level when additional measurements that are commonly collected in the clinic are utilized to tune (restructure or reparameterize) the model. A generic multi-organ model (proposed in Figure 6) can

then be integrated with the patient-specific, liver-specific model (proposed in Figure 7), resulting in a multi-organ model with a patient-specific liver (Figure 8A).

Patient-specific models may also be developed for the purpose of in silico clinical trials (Figure 8B). Extensive diversity and variability exist between the livers of patients within a population. The variability can be accounted for in large scale modeling studies by generating patient-specific liver models of integrated individualized omics data with a generic model of the liver. The model can be further refined by population-informed data including but not limited to liver anatomy, physiology, pathology, genetics, age, and sex. This results in a patient-specific liver model with population-informed data, which can then be utilized for in-silico clinical trials of various medications (Figure 8B). While in silico clinical trials are in the very early stages of development and have still yet to be utilized in the clinic, one group showed the potential of such modeling efforts in accurately recapitulating the relapse rates among various multiple sclerosis patient cohorts.[124]

The concept of digital twin models, or surrogate models that can generate data for a target patient, has recently sparked interest in the field of biology as there are direct translatable opportunities in bringing computational models with patient-specificity to the clinic. Such computational models may begin to bridge the gap between the clinician and the biologist as they can provide the clinician with patient-specific information that can be used to aid medical diagnoses and guide medication management. Digital twin models can also be used as a means for implementing closed-loop control for reprogramming an existing faulty physiological system such as the pancreas in diabetic patients[125] or the immune system in sepsis patients.[126] In a recent study, Vodovotz (2023) discusses the possibilities and challenges of integrating single-cell omics data analysis and mathematical modeling concepts to describe multi-organ disease pathophysiology toward the development of digital twin models.[127] Importantly, Vodovotz (2023) addresses the significant gap between disparate single-cell data-driven models, which assess and visualize inter-relationships among variables of "Big" data, and mechanistic models, which are typically equation-based and allow for mechanism-based reproduction of biological variables from "Not So Big" data. In this review, we propose that data driven models informed by both single-cell and spatial "Big" data in combination with "Not So Big" data can yield information about cell state transitions, interactions, and trajectories at the tissue microenvironment level. These features can then be extracted from the data-driven models and serve to intermediate the parameterization and structuring of multiscale, multiorgan, mechanistic models.

Several efforts, including the Virtual Physiological Human (VPH) and EDITH European Virtual Human Twin (VHT) initiatives, have developed in silico patient-specific physiological and pathological models in the form of digital twins.[128,129] The VPH and VHT models can begin to incorporate patient-specific variability by utilizing data from the Human Pangenome Project,[130] which was developed to represent the global genomic variation across individuals. These patient-specific models can be further detailed by incorporating patient-specific single-cell and spatial omics data and then utilized for the diagnosis, prognosis, prevention, and treatment of diseases in a patient-specific manner. Digital twin modeling and simulation may therefore allow clinicians to predict a patient's transition from a healthy to a diseased state at various levels of specificity, including the single-cell, microenvironment, and tissue levels.

POTENTIAL GAPS AND EMERGING OPPORTUNITIES

While significant strides have been made in the field of biology toward computational modeling and simulation of various biological systems, there are still necessary obstacles that must be overcome to generate systems pathophysiological models informed by "Big" and "Not So Big" data before bringing these efforts to the clinic. Firstly, the single-cell omics experimental design is crucial for properly quantifying cell populations within the tissue of interest. For instance, the single-cell sampling and subsequent experimentation may not result in cell state clusters that are quantitatively representative of the tissue microenvironment as different cell types have different susceptibilities for survival through the experimental process. Emerging methods based on single nuclear RNA sequencing (snRNA-seq) allow for dissociation of tissues that may otherwise not be readily dissociated into single-cell suspensions. Additionally, snRNA-seq methods have high sensitivity and perform well for cell type classification but do not contain spatial information.[131] Therefore, it is essential that the single-cell data are used properly to develop models that are accurately built, parameterized, and simulated.

Single-cell omics data are also sparse with a high dropout rate, which introduces a challenge in downstream data analysis as observed zeros in the data may represent true biological zeros or technical noise. However, methods of imputing data have shown a vast amount of variability in their accuracy.[132] Another option for addressing the issue of data sparsity is by using statistical models that explicitly account for the data sparsity, sampling variation, and noise.[133] When integrating single-cell omics data across samples or patients, batch effects may be introduced as an additional challenge. Various batch correction algorithms have been developed to correctly align the different datasets, while preserving the key biological variations.[134] Therefore, the patient-specific variation can be preserved for parameterization and structuring of computational models.

Following single-cell data analysis and visualization in the form of a UMAP or tSNE plot, groups of cells can be clustered to identify the cellular communities within the tissue. One gap in this step is the parameterization of the clustering algorithm that detects the number of cell type communities or clusters in the data. The resolution parameter within the clustering function can be tuned and adjusted such that there are more or less clusters depending on the user's input. For instance, if one is clustering macrophages, a low-resolution input value may only identify the M0, M1, and M2 macrophage states, while a high-resolution input value may identify the subsets of M2 macrophage states, namely the M2a, M2b, and M2c states. While varying the resolution input values is more intuitive for identifying macrophage states, cluster and cell type/state identification may not be as simple for other datasets. However, there is power in course graining and tuning the resolution parameter as the

computational model can then be pruned to the desired resolution.

Patient-specific computational models of cellular and molecular networks with spatial resolution can be additionally informed by imaging data at various length scales. For example, radiological imaging modalities including magnetic resonance imaging (MRI) and computerized tomography (CT) can be utilized to inform computational models at the whole-body physiological scale, while microscopic and molecular imaging modalities such as bright field microscopy, light sheet microscopy, and single molecule fluorescence *in situ* hybridization (smFISH) can be utilized to inform computational models at the tissue, cellular, and subcellular level.[135] Imaging data can be acquired from individual patients and then integrated with a single-cell omics-informed computational model to build executable surrogate models that can predict the static and dynamic behavior of various organs during disease progression, treatment, and post-surgical assessment.

Furthermore, there are significant challenges in sampling data when scaling computational models across different spatial and temporal scales. The data necessary for building and parameterizing models at distinct scales may be vastly different, necessitating strict and specific guidelines for experimental sampling at various time scales and resolution. Furthermore, proper simulation practices must be adhered to in terms of model development and credibility such as those presented in Viceconti and Emili (2024).[136] However, one may utilize the single-cell reference datasets (i.e., HuBMAP,[10] Human Cell Atlas,[9] The Tabula Sapiens[11]) as a starting point for modeling at various spatial resolutions from patient-specific, inter-organ cell type interactions (low resolution) to patient-specific, tissue-specific cell state interactions (high resolution). One modeling approach is to utilize low resolution data to first develop a higher-level organ model for which the model can be taken into various biological contexts once further biological complexities are incorporated (as in Figure 8). Oppositely, one may include the biological intricacies of a specific biological process and then scale it to the organ level.

The omics revolution led to the advancement of many high-resolution single-cell and spatial omics technologies that can be integrated to inform computational models of various biological systems at different scales. At the whole-body, physiological scale, multi-organ models can be generated by assembling various organ-specific models together such that a single-cell omics informed, patient-specific, multi-organ model is developed. For example, we propose that organ-specific computational models can be built individually but can interact with one another to generate a multi-organ model that can inform higher-level physiology (Figure 6). Furthermore, we propose that an individualized patient-specific model of the liver that is parameterized and structured by spatial, multi-omics data can then be utilized to inform models for in silico clinical trials (Figures 7 and 8).

Single-cell data provides a window into the functionality of individual cell types and cellular communication, which is pertinent to our understanding of tissues as a whole. Furthermore, the functionality of individual cells can be traced back to the signaling molecules which regulate them. Utilizing this information and other features extracted from "Big" and "Not So Big" data, we propose that systems pathophysiological computational models can be built at various scales including the tissue, microenvironment, cellular, and subcellular level. Model simulation will then facilitate our deeper understanding of the regulatory and functional behavior as well as heterogeneity within and across cell types in the tissue. In this review, we propose that single-cell omics data with spatial resolution can be utilized for informing, structuring, and parameterizing various types of computational models including those with patient-specificity. Furthermore, computational models of individual organs can be assembled to generate multi-organ models that can predict overall physiology. Additionally, we discuss the implications of digital twin and personalized models in the clinical setting and how they can be used for in silico clinical trials. Altogether, the future of single-cell and spatial multi-omics informed computational modeling and simulation is vast and has the potential to shape the future of healthcare and aid clinicians in pertinent decision making.

ACKNOWLEDGMENTS

The authors would like to acknowledge financial support for this study from the National Institute on Alcohol Abuse and Alcoholism R01 AA018873, T32 AA007463 (PI: R.V.), National Heart, Lung, and Blood Institute R01 HL161696, National Institutes of Health Common Fund Program SPARC OT2 OD030534 (PI: R.V.), National Institute on Alcohol Abuse and Alcoholism F31 AA030214 (PI: A.M.), National Science Foundation Graduate Research Fellowship 1940700 (PI: M.G.). The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

AUTHOR CONTRIBUTIONS

Conceptualization, R.V.; writing—original draft, A.M.; writing—review and editing, A.M., M.G., and R.V.; visualization, A.M. and R.V.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Hodgkin, A.L., and Huxley, A.F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. *117*, 500–544. https://doi.org/10.1113/jphysiol.1952.sp004764.
2. Michaelis, L., and Menten, M.L. (1913). Die kinetik der invertinwirkung. Biochem. Z. *49*, 333–369.
3. Tyson, J.J. (1991). Modeling the cell division cycle: cdc2 and cyclin interactions. Proc. Natl. Acad. Sci. USA *88*, 7328–7332. https://doi.org/10.1073/pnas.88.16.7328.
4. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210. https://doi.org/10.1093/nar/30.1.207.
5. Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., et al. (2007). ArrayExpress–a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. *35*, D747–D750. https://doi.org/10.1093/nar/gkl995.
6. Haug, K., Cochrane, K., Nainala, V.C., Williams, M., Chang, J., Jayaseelan, K.V., and O'Donovan, C. (2020). MetaboLights: a resource evolving in response to the needs of its scientific community. Nucleic Acids Res. *48*, D440–D444. https://doi.org/10.1093/nar/gkz1019.

7. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/nar/gkab1038.
8. The Cancer Genome Atlas Program - NCI https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.
9. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The human cell atlas. Elife *6*, e27041. https://doi.org/10.7554/eLife.27041.
10. The Human BioMolecular Atlas Program - HuBMAP | NIH Common Fund https://commonfund.nih.gov/HuBMAP.
11. Tabula Sapiens Consortium*; Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. https://doi.org/10.1126/science.abl4896.
12. Gustafsson, J., Anton, M., Roshanzamir, F., Jörnsten, R., Kerkhoven, E.J., Robinson, J.L., and Nielsen, J. (2023). Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data. Proc. Natl. Acad. Sci. USA *120*, e2217868120. https://doi.org/10.1073/pnas.2217868120.
13. Park, J., Zhu, H., O'Sullivan, S., Ogunnaike, B.A., Weaver, D.R., Schwaber, J.S., and Vadigepalli, R. (2016). Single-Cell Transcriptional Analysis Reveals Novel Neuronal Phenotypes and Interaction Networks Involved in the Central Circadian Clock. Front. Neurosci. *10*, 481. https://doi.org/10.3389/fnins.2016.00481.
14. Moss, A., Robbins, S., Achanta, S., Kuttippurathu, L., Turick, S., Nieves, S., Hanna, P., Smith, E.H., Hoover, D.B., Chen, J., et al. (2021). A single cell transcriptomics map of paracrine networks in the intrinsic cardiac nervous system. iScience *24*, 102713. https://doi.org/10.1016/j.isci.2021.102713.
15. Liver Cell Atlas https://livercellatlas.org/.
16. Emont, M.P., Jacobs, C., Essene, A.L., Pant, D., Tenen, D., Colleluori, G., Di Vincenzo, A., Jørgensen, A.M., Dashti, H., Stefek, A., et al. (2022). A single-cell atlas of human and mouse white adipose tissue. Nature *603*, 926–933. https://doi.org/10.1038/s41586-022-04518-2.
17. Single Cell Portal https://singlecell.broadinstitute.org/single_cell.
18. Anderson, W.D., and Vadigepalli, R. (2016). Modeling cytokine regulatory network dynamics driving neuroinflammation in central nervous system disorders. Drug Discov. Today Dis. Model. *19*, 59–67. https://doi.org/10.1016/j.ddmod.2017.01.003.
19. Simmune Project | NIH: National Institute of Allergy and Infectious Diseases https://www.niaid.nih.gov/research/simmune-project.
20. Hoehme, S., and Drasdo, D. (2010). A cell-based simulation software for multi-cellular systems. Bioinformatics *26*, 2641–2642. https://doi.org/10.1093/bioinformatics/btq437.
21. Marino, S., Hult, C., Wolberg, P., Linderman, J.J., and Kirschner, D.E. (2018). The role of dimensionality in understanding granuloma formation. Computation *6*, 58. https://doi.org/10.3390/computation6040058.
22. Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., and Koltay, P. (2015). Technologies for Single-Cell Isolation. Int. J. Mol. Sci. *16*, 16897–16919. https://doi.org/10.3390/ijms160816897.
23. Welzel, G., Seitz, D., and Schuster, S. (2015). Magnetic-activated cell sorting (MACS) can be used as a large-scale method for establishing zebrafish neuronal cell cultures. Sci. Rep. *5*, 7959. https://doi.org/10.1038/srep07959.
24. Datta, S., Malhotra, L., Dickerson, R., Chaffee, S., Sen, C.K., and Roy, S. (2015). Laser capture microdissection: Big data from small samples. Histol. Histopathol. *30*, 1255–1269. https://doi.org/10.14670/HH-11-622.
25. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214. https://doi.org/10.1016/j.cell.2015.05.002.
26. Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201. https://doi.org/10.1016/j.cell.2015.04.044.
27. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049. https://doi.org/10.1038/ncomms14049.
28. Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463–1467. https://doi.org/10.1126/science.aaw1219.
29. Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P., Macosko, E.Z., and Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat. Biotechnol. *39*, 313–319. https://doi.org/10.1038/s41587-020-0739-1.
30. Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020). High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell *183*, 1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026.
31. Xia, K., Sun, H.-X., Li, J., Li, J., Zhao, Y., Chen, L., Qin, C., Chen, R., Chen, Z., Liu, G., et al. (2022). The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in Arabidopsis leaves. Dev. Cell *57*, 1299–1310.e4. https://doi.org/10.1016/j.devcel.2022.04.011.
32. He, S., Bhatt, R., Brown, C., Brown, E.A., Buhr, D.L., Chantranuvatana, K., Danaher, P., Dunaway, D., Garrison, R.G., Geiss, G., et al. (2022). High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. Nat. Biotechnol. *40*, 1794–1806. https://doi.org/10.1038/s41587-022-01483-z.
33. Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Ferrante, T.C., Terry, R., Turczyk, B.M., Yang, J.L., Lee, H.S., Aach, J., et al. (2015). Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat. Protoc. *10*, 442–458. https://doi.org/10.1038/nprot.2014.191.
34. Zenobi, R. (2013). Single-cell metabolomics: analytical and biological perspectives. Science *342*, 1243259. https://doi.org/10.1126/science.1243259.
35. Marx, V. (2019). A dream of single-cell proteomics. Nat. Methods *16*, 809–812. https://doi.org/10.1038/s41592-019-0540-6.
36. Kelly, R.T. (2020). Single-cell Proteomics: Progress and Prospects. Mol. Cell. Proteomics *19*, 1739–1748. https://doi.org/10.1074/mcp.R120.002234.
37. Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. *19*, 161. https://doi.org/10.1186/s13059-018-1547-5.
38. Specht, H., Emmott, E., Petelski, A.A., Huffman, R.G., Perlman, D.H., Serra, M., Kharchenko, P., Koller, A., and Slavov, N. (2021). Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome Biol. *22*, 50. https://doi.org/10.1186/s13059-021-02267-5.
39. Zhu, Y., Piehowski, P.D., Zhao, R., Chen, J., Shen, Y., Moore, R.J., Shukla, A.K., Petyuk, V.A., Campbell-Thompson, M., Mathews, C.E., et al. (2018). Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. Nat. Commun. *9*, 882. https://doi.org/10.1038/s41467-018-03367-w.
40. Li, Z.-Y., Huang, M., Wang, X.-K., Zhu, Y., Li, J.-S., Wong, C.C.L., and Fang, Q. (2018). Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell

Proteomic Analysis. Anal. Chem. *90*, 5430–5438. https://doi.org/10.1021/acs.analchem.8b00661.

41. Shao, X., Wang, X., Guan, S., Lin, H., Yan, G., Gao, M., Deng, C., and Zhang, X. (2018). Integrated Proteome Analysis Device for Fast Single-Cell Protein Profiling. Anal. Chem. *90*, 14003–14010. https://doi.org/10.1021/acs.analchem.8b03692.
42. Guevremont, R. (2004). High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. J. Chromatogr. A *1058*, 3–19. https://doi.org/10.1016/j.chroma.2004.08.119.
43. Cong, Y., Motamedchaboki, K., Misal, S.A., Liang, Y., Guise, A.J., Truong, T., Huguet, R., Plowey, E.D., Zhu, Y., Lopez-Ferrer, D., and Kelly, R.T. (2020). Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. Chem. Sci. *12*, 1001–1006. https://doi.org/10.1039/d0sc03636f.
44. Tracey, L.J., An, Y., and Justice, M.J. (2021). CyTOF: An Emerging Technology for Single-Cell Proteomics in the Mouse. Curr. Protoc. *1*, e118. https://doi.org/10.1002/cpz1.118.
45. Liu, R., and Yang, Z. (2021). Single cell metabolomics using mass spectrometry: Techniques and data analysis. Anal. Chim. Acta *1143*, 124–134. https://doi.org/10.1016/j.aca.2020.11.020.
46. Taylor, M.J., Lukowski, J.K., and Anderton, C.R. (2021). Spatially resolved mass spectrometry at the single cell: recent innovations in proteomics and metabolomics. J. Am. Soc. Mass Spectrom. *32*, 872–894. https://doi.org/10.1021/jasms.0c00439.
47. Passarelli, M.K., Newman, C.F., Marshall, P.S., West, A., Gilmore, I.S., Bunch, J., Alexander, M.R., and Dollery, C.T. (2015). Single-Cell Analysis: Visualizing Pharmaceutical and Metabolite Uptake in Cells with Label-Free 3D Mass Spectrometry Imaging. Anal. Chem. *87*, 6696–6702. https://doi.org/10.1021/acs.analchem.5b00842.
48. Agüi-Gonzalez, P., Jähne, S., and Phan, N.T.N. (2019). SIMS imaging in neurobiology and cell biology. J. Anal. At. Spectrom. *34*, 1355–1368. https://doi.org/10.1039/C9JA00118B.
49. Amantonico, A., Urban, P.L., Fagerer, S.R., Balabin, R.M., and Zenobi, R. (2010). Single-cell MALDI-MS as an analytical tool for studying intrapopulation metabolic heterogeneity of unicellular organisms. Anal. Chem. *82*, 7394–7400. https://doi.org/10.1021/ac1015326.
50. Stopka, S.A., Khattar, R., Agtuca, B.J., Anderton, C.R., Paša-Tolić, L., Stacey, G., and Vertes, A. (2018). Metabolic Noise and Distinct Subpopulations Observed by Single Cell LAESI Mass Spectrometry of Plant Cells in situ. Front. Plant Sci. *9*, 1646. https://doi.org/10.3389/fpls.2018.01646.
51. Stopka, S.A., Rong, C., Korte, A.R., Yadavilli, S., Nazarian, J., Razunguzwa, T.T., Morris, N.J., and Vertes, A. (2016). Molecular imaging of biological samples on nanophotonic laser desorption ionization platforms. Angew. Chem. *128*, 4558–4562. https://doi.org/10.1002/ange.201511691.
52. Bergman, H.-M., and Lanekoff, I. (2017). Profiling and quantifying endogenous molecules in single cells using nano-DESI MS. Analyst *142*, 3639–3647. https://doi.org/10.1039/c7an00885f.
53. Pan, N., Rao, W., Kothapalli, N.R., Liu, R., Burgett, A.W.G., and Yang, Z. (2014). The single-probe: a miniaturized multifunctional device for single cell mass spectrometry analysis. Anal. Chem. *86*, 9376–9380. https://doi.org/10.1021/ac5029038.
54. Liu, R., Pan, N., Zhu, Y., and Yang, Z. (2018). T-Probe: An Integrated Microscale Device for Online In Situ Single Cell Analysis and Metabolic Profiling Using Mass Spectrometry. Anal. Chem. *90*, 11078–11085. https://doi.org/10.1021/acs.analchem.8b02927.
55. Genshaft, A.S., Li, S., Gallant, C.J., Darmanis, S., Prakadan, S.M., Ziegler, C.G.K., Lundberg, M., Fredriksson, S., Hong, J., Regev, A., et al. (2016). Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. Genome Biol. *17*, 188. https://doi.org/10.1186/s13059-016-1045-6.
56. Frei, A.P., Bava, F.-A., Zunder, E.R., Hsieh, E.W.Y., Chen, S.-Y., Nolan, G.P., and Gherardini, P.F. (2016). Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat. Methods *13*, 269–275. https://doi.org/10.1038/nmeth.3742.
57. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868. https://doi.org/10.1038/nmeth.4380.
58. Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. Nat. Biotechnol. *35*, 936–939. https://doi.org/10.1038/nbt.3973.
59. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. *24*, 695–713. https://doi.org/10.1038/s41580-023-00615-w.
60. Miao, Z., Humphreys, B.D., McMahon, A.P., and Kim, J. (2021). Multi-omics integration in the age of million single-cell data. Nat. Rev. Nephrol. *17*, 710–724. https://doi.org/10.1038/s41581-021-00463-x.
61. Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat. Methods *21*, 1470–1480. https://doi.org/10.1038/s41592-024-02201-0.
62. Theodoris, C.V., Xiao, L., Chopra, A., Chaffin, M.D., Al Sayed, Z.R., Hill, M.C., Mantineo, H., Brydon, E.M., Zeng, Z., Liu, X.S., and Ellinor, P.T. (2023). Transfer learning enables predictions in network biology. Nature *618*, 616–624. https://doi.org/10.1038/s41586-023-06139-9.
63. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
64. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36. https://doi.org/10.1186/gb-2013-14-4-r36.
65. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.
66. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162. https://doi.org/10.1038/s41592-019-0667-5.
67. Zhou, P., Wang, S., Li, T., and Nie, Q. (2021). Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. Nat. Commun. *12*, 5609. https://doi.org/10.1038/s41467-021-25548-w.
68. Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genom. *19*, 477. https://doi.org/10.1186/s12864-018-4772-0.
69. Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. Science *295*, 1669–1678. https://doi.org/10.1126/science.1069883.
70. Novak, B., and Tyson, J.J. (1993). Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. J. Cell Sci. *106*, 1153–1168. https://doi.org/10.1242/jcs.106.4.1153.
71. Forger, D.B., and Peskin, C.S. (2003). A detailed predictive model of the mammalian circadian clock. Proc. Natl. Acad. Sci. USA *100*, 14806–14811. https://doi.org/10.1073/pnas.2036281100.
72. Terfve, C., Cokelaer, T., Henriques, D., MacNamara, A., Goncalves, E., Morris, M.K., van Iersel, M., Lauffenburger, D.A., and Saez-Rodriguez, J. (2012). CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. BMC Syst. Biol. *6*, 133. https://doi.org/10.1186/1752-0509-6-133.

73. Woodhouse, S., Piterman, N., Wintersteiger, C.M., Göttgens, B., and Fisher, J. (2018). SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. BMC Syst. Biol. *12*, 59. https://doi.org/10.1186/s12918-018-0581-y.

74. Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N., Hayashi, T., and Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics *33*, 2314–2321. https://doi.org/10.1093/bioinformatics/btx194.

75. Bower, J.M., and Bolouri, H. (2001). Computational modeling of genetic and biochemical networks (The MIT Press). https://doi.org/10.7551/mitpress/2018.001.0001.

76. Edelstein-Keshet, L. (2005). Mathematical Models in Biology (Society for Industrial and Applied Mathematics). https://doi.org/10.1137/1.9780898719147.

77. Shmulevich, I., and Dougherty, E.R. (2010). Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks (Society for Industrial and Applied Mathematics). https://doi.org/10.1137/1.9780898717631.

78. Livingstone, D.J. (2008). Artificial neural networks: Methods and applications. Methods Mol. Biol. *458*.

79. Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., and Collins, J.J. (2018). Next-Generation Machine Learning for Biological Networks. Cell *173*, 1581–1592. https://doi.org/10.1016/j.cell.2018.05.015.

80. Nazari, F., Pearson, A.T., Nör, J.E., and Jackson, T.L. (2018). A mathematical model for IL-6-mediated, stem cell driven tumor growth and targeted treatment. PLoS Comput. Biol. *14*, e1005920. https://doi.org/10.1371/journal.pcbi.1005920.

81. Cook, D., Achanta, S., Hoek, J.B., Ogunnaike, B.A., and Vadigepalli, R. (2018). Cellular network modeling and single cell gene expression analysis reveals novel hepatic stellate cell phenotypes controlling liver regeneration dynamics. BMC Syst. Biol. *12*, 86. https://doi.org/10.1186/s12918-018-0605-7.

82. Sha, Y., Wang, S., Zhou, P., and Nie, Q. (2020). Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data. Nucleic Acids Res. *48*, 9505–9520. https://doi.org/10.1093/nar/gkaa725.

83. Walesky, C.M., Kolb, K.E., Winston, C.L., Henderson, J., Kruft, B., Fleming, I., Ko, S., Monga, S.P., Mueller, F., Apte, U., et al. (2020). Functional compensation precedes recovery of tissue mass following acute liver injury. Nat. Commun. *11*, 5785. https://doi.org/10.1038/s41467-020-19558-3.

84. Waclaw, B., Bozic, I., Pittman, M.E., Hruban, R.H., Vogelstein, B., and Nowak, M.A. (2015). A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature *525*, 261–264. https://doi.org/10.1038/nature14971.

85. Jagiella, N., Müller, B., Müller, M., Vignon-Clementel, I.E., and Drasdo, D. (2016). Inferring Growth Control Mechanisms in Growing Multi-cellular Spheroids of NSCLC Cells from Spatial-Temporal Image Data. PLoS Comput. Biol. *12*, e1004412. https://doi.org/10.1371/journal.pcbi.1004412.

86. Verma, A., Antony, A.N., Ogunnaike, B.A., Hoek, J.B., and Vadigepalli, R. (2018). Causality analysis and cell network modeling of spatial calcium signaling patterns in liver lobules. Front. Physiol. *9*, 1377. https://doi.org/10.3389/fphys.2018.01377.

87. Dichamp, J., Cellière, G., Ghallab, A., Hassan, R., Boissier, N., Hofmann, U., Reinders, J., Sezgin, S., Zühlke, S., Hengstler, J.G., et al. (2023). In vitro to in vivo acetaminophen hepatotoxicity extrapolation using classical schemes, pharmacodynamic models and a multiscale spatial-temporal liver twin. Front. Bioeng. Biotechnol. *11*, 1049564. https://doi.org/10.3389/fbioe.2023.1049564.

88. Verma, A., Manchel, A., Narayanan, R., Hoek, J.B., Ogunnaike, B.A., and Vadigepalli, R. (2021). A spatial model of hepatic calcium signaling and glucose metabolism under autonomic control reveals functional consequences of varying liver innervation patterns across species. Front. Physiol. *12*, 748962. https://doi.org/10.3389/fphys.2021.748962.

89. Höhme, S., Hengstler, J.G., Brulport, M., Schäfer, M., Bauer, A., Gebhardt, R., and Drasdo, D. (2007). Mathematical modelling of liver regeneration after intoxication with CCl(4). Chem. Biol. Interact. *168*, 74–93. https://doi.org/10.1016/j.cbi.2007.01.010.

90. Hoehme, S., Hammad, S., Boettger, J., Begher-Tibbe, B., Bucur, P., Vibert, E., Gebhardt, R., Hengstler, J.G., and Drasdo, D. (2023). Digital twin demonstrates significance of biomechanical growth control in liver regeneration after partial hepatectomy. iScience *26*, 105714. https://doi.org/10.1016/j.isci.2022.105714.

91. Duriez, M., Jacquet, A., Hoet, L., Roche, S., Bock, M.-D., Rocher, C., Haussy, G., Vigé, X., Bocskei, Z., Slavnic, T., et al. (2020). A 3D human liver model of nonalcoholic steatohepatitis. J. Clin. Transl. Hepatol. *8*, 359–370. https://doi.org/10.14218/JCTH.2020.00015.

92. Friedman, A., and Hao, W. (2017). Mathematical modeling of liver fibrosis. Math. Biosci. Eng. *14*, 143–164. https://doi.org/10.3934/mbe.2017010.

93. Cang, Z., Wang, Y., Wang, Q., Cho, K.W.Y., Holmes, W., and Nie, Q. (2021). A multiscale model via single-cell transcriptomics reveals robust patterning mechanisms during early mammalian embryo development. PLoS Comput. Biol. *17*, e1008571. https://doi.org/10.1371/journal.pcbi.1008571.

94. Singh-Varma, A., Shah, A.M., Liu, S., Zamora, R., Monga, S.P., and Vodovotz, Y. (2023). Defining spatiotemporal gene modules in liver regeneration using Analytical Dynamic Visual Spatial Omics Representation (ADViSOR). Hepatol. Commun. *7*, e0289. https://doi.org/10.1097/HC9.0000000000000289.

95. Aubin-Frankowski, P.-C., and Vert, J.-P. (2020). Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference. Bioinformatics *36*, 4774–4780. https://doi.org/10.1093/bioinformatics/btaa576.

96. Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., Wang, C., Fu, H., Ma, Q., and Xu, D. (2021). scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. Nat. Commun. *12*, 1882. https://doi.org/10.1038/s41467-021-22197-x.

97. Alghamdi, N., Chang, W., Dang, P., Lu, X., Wan, C., Gampala, S., Huang, Z., Wang, J., Ma, Q., Zang, Y., et al. (2021). A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. Genome Res. *31*, 1867–1884. https://doi.org/10.1101/gr.271205.120.

98. Chen, Z., King, W.C., Hwang, A., Gerstein, M., and Zhang, J. (2022). DeepVelo: Single-cell transcriptomic deep velocity field learning with neural ordinary differential equations. Sci. Adv. *8*, eabq3745. https://doi.org/10.1126/sciadv.abq3745.

99. Palshikar, M.G., Palli, R., Tyrell, A., Maggirwar, S., Schifitto, G., Singh, M.V., and Thakar, J. (2022). Executable models of immune signaling pathways in HIV-associated atherosclerosis. NPJ Syst. Biol. Appl. *8*, 35. https://doi.org/10.1038/s41540-022-00246-5.

100. Yuan, Z., Li, Y., Shi, M., Yang, F., Gao, J., Yao, J., and Zhang, M.Q. (2022). SOTIP is a versatile method for microenvironment modeling with spatial omics data. Nat. Commun. *13*, 7330. https://doi.org/10.1038/s41467-022-34867-5.

101. Wu, Z., Trevino, A.E., Wu, E., Swanson, K., Kim, H.J., D'Angio, H.B., Preska, R., Charville, G.W., Dalerba, P.D., Egloff, A.M., et al. (2022). Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. Nat. Biomed. Eng. *6*, 1435–1448. https://doi.org/10.1038/s41551-022-00951-w.

102. Tanevski, J., Flores, R.O.R., Gabor, A., Schapiro, D., and Saez-Rodriguez, J. (2022). Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. Genome Biol. *23*, 97. https://doi.org/10.1186/s13059-022-02663-5.

103. SPARC Portal https://sparc.science/.

104. Gee, M.M., Hornung, E., Gupta, S., Newton, A.J.H., Cheng, Z.J., Lytton, W.W., Lenhoff, A.M., Schwaber, J.S., and Vadigepalli, R. (2023). Unpacking the multimodal, multi-scale data of the fast and slow lanes of the cardiac vagus through computational modelling. Exp. Physiol. https://doi.org/10.1113/EP090865.

105. Gee, M.M., Lenhoff, A.M., Schwaber, J.S., Ogunnaike, B.A., and Vadigepalli, R. (2023). Closed-loop modeling of central and intrinsic cardiac nervous system circuits underlying cardiovascular control. AIChE J. *69*, e18033. https://doi.org/10.1002/aic.18033.

106. Thiele, I., Sahoo, S., Heinken, A., Hertel, J., Heirendt, L., Aurich, M.K., and Fleming, R.M. (2020). Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. Mol. Syst. Biol. *16*, e8982. https://doi.org/10.15252/msb.20198982.

107. Zhang, Y., Kim, M.S., Nguyen, E., and Taylor, D.M. (2020). Modeling metabolic variation with single-cell expression data. Preprint at bioRxiv. https://doi.org/10.1101/2020.01.28.923680.

108. Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators (2018). Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, and Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367–372. https://doi.org/10.1038/s41586-018-0590-4.

109. Park, J., Brureau, A., Kernan, K., Starks, A., Gulati, S., Ogunnaike, B., Schwaber, J., and Vadigepalli, R. (2014). Inputs drive cell phenotype variability. Genome Res. *24*, 930–941. https://doi.org/10.1101/gr.161802.113.

110. Park, J.H., Gorky, J., Ogunnaike, B., Vadigepalli, R., and Schwaber, J.S. (2020). Investigating the effects of brainstem neuronal adaptation on cardiovascular homeostasis. Front. Neurosci. *14*, 470. https://doi.org/10.3389/fnins.2020.00470.

111. Nandi, A., Chartrand, T., Van Geit, W., Buchin, A., Yao, Z., Lee, S.Y., Wei, Y., Kalmbach, B., Lee, B., Lein, E., et al. (2022). Single-neuron models linking electrophysiology, morphology, and transcriptomics across cortical cell types. Cell Rep. *40*, 111176. https://doi.org/10.1016/j.celrep.2022.111176.

112. Sampson, K.J., Iyer, V., Marks, A.R., and Kass, R.S. (2010). A computational model of Purkinje fibre single cell electrophysiology: implications for the long QT syndrome. J. Physiol. *588*, 2643–2655. https://doi.org/10.1113/jphysiol.2010.187328.

113. Verma, B.K., Subramaniam, P., and Vadigepalli, R. (2019). Model-based virtual patient analysis of human liver regeneration predicts critical perioperative factors controlling the dynamic mode of response to resection. BMC Syst. Biol. *13*, 9. https://doi.org/10.1186/s12918-019-0678-y.

114. Robinson, J.L., Kocabaş, P., Wang, H., Cholley, P.-E., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V., et al. (2020). An atlas of human metabolism. Sci. Signal. *13*, eaaz1482. https://doi.org/10.1126/scisignal.aaz1482.

115. Thiele, I., Swainston, N., Fleming, R.M.T., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013). A community-driven global reconstruction of human metabolism. Nat. Biotechnol. *31*, 419–425. https://doi.org/10.1038/nbt.2488.

116. Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc. *14*, 639–702. https://doi.org/10.1038/s41596-018-0098-2.

117. Wang, H., Marcišauskas, S., Sánchez, B.J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., and Kerkhoven, E.J. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. PLoS Comput. Biol. *14*, e1006541. https://doi.org/10.1371/journal.pcbi.1006541.

118. Chandrasekaran, S., Zhang, J., Sun, Z., Zhang, L., Ross, C.A., Huang, Y.-C., Asara, J.M., Li, H., Daley, G.Q., and Collins, J.J. (2017). Comprehensive Mapping of Pluripotent Stem Cell Metabolism Using Dynamic Genome-Scale Network Modeling. Cell Rep. *21*, 2965–2977. https://doi.org/10.1016/j.celrep.2017.07.048.

119. Manchel, A., Mahadevan, R., Bataller, R., Hoek, J.B., and Vadigepalli, R. (2022). Genome-Scale Metabolic Modeling Reveals Sequential Dysregulation of Glutathione Metabolism in Livers from Patients with Alcoholic Hepatitis. Metabolites *12*, 1157. https://doi.org/10.3390/metabo12121157.

120. Preciat, G., Wegrzyn, A.B., Thiele, I., Hankemeier, T., and Fleming, R.M. (2021). XomicsToModel: Omics data integration and generation of thermodynamically consistent metabolic models. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.08.467803.

121. Ambikan, A.T., Yang, H., Krishnan, S., Svensson Akusjärvi, S., Gupta, S., Lourda, M., Sperk, M., Arif, M., Zhang, C., Nordqvist, H., et al. (2022). Multi-omics personalized network analyses highlight progressive disruption of central metabolism associated with COVID-19 severity. Cell Syst. *13*, 665–681.e4. https://doi.org/10.1016/j.cels.2022.06.006.

122. Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M., and Nielsen, J. (2014). Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol. Syst. Biol. *10*, 721. https://doi.org/10.1002/msb.145122.

123. Kim, M.K., Lane, A., Kelley, J.J., and Lun, D.S. (2016). E-Flux2 and SPOT: Validated Methods for Inferring Intracellular Metabolic Flux Distributions from Transcriptomic Data. PLoS One *11*, e0157101. https://doi.org/10.1371/journal.pone.0157101.

124. Sips, F.L.P., Pappalardo, F., Russo, G., and Bursi, R. (2022). In silico clinical trials for relapsing-remitting multiple sclerosis with MS TreatSim. BMC Med. Inf. Decis. Making *22*, 294. https://doi.org/10.1186/s12911-022-02034-x.

125. Masison, J., Beezley, J., Mei, Y., Ribeiro, H., Knapp, A.C., Sordo Vieira, L., Adhikari, B., Scindia, Y., Grauer, M., Helba, B., et al. (2021). A modular computational framework for medical digital twins. Proc. Natl. Acad. Sci. USA *118*, e2024287118. https://doi.org/10.1073/pnas.2024287118.

126. Namas, R.A., Mikheev, M., Yin, J., Barclay, D., Jefferson, B., Mi, Q., Billiar, T.R., Zamora, R., Gerlach, J., and Vodovotz, Y. (2023). An adaptive, negative feedback circuit in a biohybrid device reprograms dynamic networks of systemic inflammation in vivo. Front. Syst. Biol. *2*, 926618. https://doi.org/10.3389/fsysb.2022.926618.

127. Vodovotz, Y. (2023). Towards systems immunology of critical illness at scale: from single cell 'omics to digital twins. Trends Immunol. *44*, 345–355. https://doi.org/10.1016/j.it.2023.03.004.

128. VPH Institute | Virtual Physiological Human - International non-profit organisation https://www.vph-institute.org/.

129. European Virtual Human Twin https://www.edith-csa.eu/.

130. Wang, T., Antonacci-Fulton, L., Howe, K., Lawson, H.A., Lucas, J.K., Phillippy, A.M., Popejoy, A.B., Asri, M., Carson, C., Chaisson, M.J.P., et al. (2022). The Human Pangenome Project: a global resource to map genomic diversity. Nature *604*, 437–446. https://doi.org/10.1038/s41586-022-04601-8.

131. Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. *38*, 737–746. https://doi.org/10.1038/s41587-020-0465-8.

132. Hou, W., Ji, Z., Ji, H., and Hicks, S.C. (2020). A systematic evaluation of single-cell RNA-sequencing imputation methods. Genome Biol. *21*, 218. https://doi.org/10.1186/s13059-020-02132-x.

133. Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N.,

Mahfouz, A., et al. (2020). Eleven grand challenges in single-cell data science. Genome Biol. *21*, 31. https://doi.org/10.1186/s13059-020-1926-6.

134. Tran, H.T.N., Ang, K.S., Chevrier, M., Zhang, X., Lee, N.Y.S., Goh, M., and Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol. *21*, 12. https://doi.org/10.1186/s13059-019-1850-9.

135. Verma, A., Manchel, A., Melunis, J., Hengstler, J.G., and Vadigepalli, R. (2022). From Seeing to Simulating: A Survey of Imaging Techniques and Spatially-Resolved Data for Developing Multiscale Computational Models of Liver Regeneration. Front. Syst. Biol. *2*, 917191. https://doi.org/10.3389/fsysb.2022.917191.

136. Viceconti, M., and Emili, L. (2024). Toward good simulation practice: best practices for the use of computational modelling and simulation in the regulatory process of biomedical products (Springer Nature Switzerland). https://doi.org/10.1007/978-3-031-48284-7.