Title

COMPASS: A Web-Based COMPosite Activity Scoring System to Navigate Health and Disease Through Deterministic Digital Biomarkers

Authors

Saptarshi Sinha[1, 3]* and Pradipta Ghosh[1-3]*

Departments of [1]Cellular and Molecular Medicine and [2]Medicine, and [3]Institute for Network Medicine, University of California, San Diego, CA, 92093, USA.

*Correspondence to sasinha@health.ucsd.edu (SS) or prghosh@ucsd.edu (P.G).

Keywords (two to six)

Gene expression analysis · Data-driven thresholding · Composite activity scoring · Pathway quantification · Precision medicine · Cohort anchoring · New Approach Methods (NAMs)

Abstract

Quantifying pathway activation in absolute, reproducible terms is central to systems biology and precision medicine. COMPASS (COMPosite Activity Scoring System) provides a deterministic, ontology-free framework that extracts gene-specific activation thresholds from expression data, standardizes deviation from these boundaries, and aggregates direction-encoded genes into per-sample composite activity scores. Implemented as an intuitive web application, COMPASS enables non-coding users to upload expression matrices, define custom gene sets, and instantly generate activity plots and ROC–AUC statistics for biological or clinical comparisons. Across diverse datasets, COMPASS yields stable, interpretable, and transferable digital biomarkers that benchmark the "humanness" of model systems, quantify differentiation or immune states, and track therapeutic response trajectories. By directly linking expression, threshold, deviation, and directionality, COMPASS replaces permutation-based enrichment with closed-form logic, delivering a transparent, mechanistic, and reproducible quantification system for pathway activity.

Impact statement

COMPASS provides reproducible composite pathway activity scores from transcriptomic data, enabling standardized benchmarking across cohorts and model systems and accelerating translational biomarker evaluation.

Introduction

Understanding how genes act collectively to define cellular states remains a central challenge in systems biology. Yet, enrichment-based frameworks, such as GSEA[1], GSVA[2], ssGSEA[3], PLAGE[4], AUCell[5], and VAM[6] measure *relative* enrichment, not *absolute* activity. These approaches rely on evolving ontologies and permutation statistics, which introduce blind spots and reproducibility failures (**Fig. 1a**-*left*). As pathway definitions shift, scores fluctuate across datasets, platforms, and time, raising persistent concerns about the reliability of ontology-dependent analyses in translational research[7-9].

To overcome these limitations, we developed **COMPASS (COMPosite Activity Scoring System)**, a deterministic, transparent, and web-based framework that quantifies gene-set activity directly from expression data (**Fig. 1a**-*middle*). COMPASS enables researchers, including those without coding expertise, to convert gene-expression matrices into standardized, reproducible indices of pathway activation without random permutations or curated hierarchies. It transforms biological variability into measurable logic, yielding stable, interpretable, and transferable digital biomarkers of biological activity.

COMPASS operationalizes the foundational Boolean principle that biological systems compute through thresholds and feedback, capturing the continuum of cellular decisions from health to disease [expanded here[10]]. Modeled for long using the *Bo*olean *N*etwork *E*xplorer (BoNE[11]) and named recently as **AXIOM-Bio**[10] (*Abstract eXpression Inference and Ontology-free Modeling for Biological Logic and Disease*), this framework for creating foundational models of disease continuum states has demonstrated reproducibility across species and diseases[10,12]. Despite the proven track record of enabling objective benchmarking of digital biomarkers, assessing the "humanness" of animal models and NAMs, and tracking disease trajectories or therapeutic reversals[10], BoNE's computational complexity has limited accessibility. By embedding this logic in an intuitive web platform (https://compass.precsn.com/), **COMPASS democratizes Boolean logic**, enabling users to convert raw expression data into deterministic digital biomarkers with a few clicks, bridging foundational biological **reasoning**, computational **rigor**, regulatory **reproducibility**, and clinical and/or experimental **relevance** (**Fig. 1a**-*right*).

Mathematical transparency, the defining feature of COMPASS, is achieved through a deterministic workflow of three sequential steps: *thresholding, standardization,* and *aggregation* (**Fig. 1b**; see ***Online Methods***). Each step corresponds to a concrete, interpretable biological operation rather than an abstract statistical transformation. In the thresholding step, COMPASS identifies for every gene its intrinsic expression boundary—the inflection where its state transitions from "off" to "on." This data-driven threshold, determined automatically by adaptive segmentation algorithms, pinpoints where biological probability collapses into commitment. The standardization step aligns genes of varying dynamic range onto a unified scale, transforming noisy expression profiles into interpretable distance measures that indicate how far each gene lies from its activation boundary. Finally, the aggregation step integrates these standardized deviations across

genes within a defined set, applying direction-specific weights to capture the balance of opposing influences. The resulting composite score represents the net activation of a pathway in each sample.

In practice, these mathematically transparent operations occur entirely in the background (**Fig. 1c**). Users simply upload a tab-delimited expression matrix (txt, tsv, csv), define custom gene signatures and sample groups, and initiate analysis through an intuitive point-and-click interface (**Fig S1**). No pathway databases, metadata, or programming are required. The backend executes all computations automatically, generating downloadable boxplots and ROC–AUC curves that benchmark pathway activation against clinical or biological groups. Every run from identical input yields identical output, ensuring the reproducibility required for regulatory-grade science.

The most important output is the framework's ability to transform qualitative gene signatures into quantitative digital biomarkers. By linking each computational step to a measurable biological principle, i.e., expression, threshold, deviation, or directionality, COMPASS replaces randomization with determinism. This direct mapping of data to decision logic defines its mathematical transparency: users can trace each result to its underlying threshold and deviation, eliminating both statistical opacity and annotation bias. Because thresholds are derived directly from the data rather than from evolving ontologies, COMPASS is platform-agnostic and transferable across tissues, cohorts, and species. It thus aligns biological logic independent of normalization or batch effects, converting pathway analysis from a statistical comparison into a mechanistic measurement system. It harmonizes results across studies and captures subtle transitions such as health-to-disease progression, differentiation dynamics, or treatment-induced reversibility with reproducible precision.

Results

COMPASS Workflow and Implementation

The COMPASS workflow transforms a raw gene-expression matrix into deterministic, quantitative measures of pathway activity through a three-step process: (i) thresholding, (ii) standardization, and (iii) aggregation (**Fig. 1b**). Unlike enrichment-based methods that rely on predefined ontologies or random permutations, COMPASS derives gene-specific thresholds directly from expression data, converts deviations from these thresholds into standardized metrics, and integrates them to compute composite activity scores. All computations are closed form and deterministic: identical input produces identical output, ensuring reproducibility, auditability, and mathematical transparency.

1. Threshold determination (biological decision boundaries): At the core of COMPASS lies the *StepMiner*[13] algorithm, an adaptive algorithm originally developed to identify sharp transcriptional transitions from noisy expression data. For each gene *g*, the expression values across samples are sorted, and a step function is fit to this ordered distribution. The transition index (*Tg*) represents the inflection point that minimizes within-group variance—corresponding to the biological boundary between low ("off") and high ("on") expression states.

For each gene g and sample i, COMPASS computes a normalized deviation from the threshold as:

$$E_{g,i} = \begin{cases} \mu_{g,0}, & i < T_g \\ \mu_{g,1}, & i \geq T_g \end{cases} \quad (1)$$

Where T_g denotes the transition index minimizing within-group variance. This threshold defines the binary states “low (0)” and “high (1)” expression. This adaptive segmentation identifies the **intrinsic activation threshold** for each gene directly from data, without requiring any external annotation or training. Conceptually, this threshold identifies where biological probability collapses into commitment (“off” → “on”)[10]. In doing so, thresholding reveals the cellular decision boundary.

To minimize stochastic fluctuations around this boundary, COMPASS applies a small confidence offset (σ) to the threshold, defining the final decision boundary of each gene g as:

$$T^*{}_g = T_g + \sigma \quad (2)$$

where, $\sigma = 0.5\ \log_2 Units$. This conservative shift in threshold ensures that only values confidently above or below the step contribute to the score.

2. Standardization (normalization across genes): Once thresholds are established, each gene’s deviation from its intrinsic boundary is standardized relative to its variance, yielding a dimensionless measure of distance from the decision point: For every gene g and sample i, the deviation from this intrinsic threshold is standardized as:

$$Z_{g,i} = \frac{E_{g,i} - T_g}{\sigma_g} \quad (3)$$

where $E_{g,i}$ is the observed expression and σ_g the gene-specific standard deviation. $Z_{g,i}$ thus measures the distance from the decision boundary[11], producing a hybrid metric: discrete when categorical precision is needed (i.e., digital) and continuous when population gradients matter (i.e., analog). The standardized deviation used in COMPASS is then calculated as:

$$Z_{g,i} = \frac{E_{g,i} - T^*{}_g}{3\sigma_g} \quad (4)$$

Dividing by $3\sigma_g$ (instead of 1×σ_g) compresses extreme outliers and brings all genes to a comparable dynamic scale. This preserves biological directionality while reducing threshold-adjacent noise.

This standardization step converts heterogeneous expression profiles into a common coordinate system in which each $Z_{(g,i)}$ representing how strongly a gene deviates from its activation threshold.

3. Composite aggregation (pathway-level integration): For any gene-set, $G = \{g_1, g_2, \dots, g_m\}$, COMPASS computes the **composite activity score** for each sample *i* as the weighted mean of standardized deviations:

$$C_i = \frac{1}{m} \sum_{j=1}^{m} w_j \, Z_{g_j,i} \quad (5)$$

where weights w_j encode gene directionality (+1 = activation, −1 = repression). The resulting composite score (C_i) reflects the net balance of activation and repression, providing an interpretable, quantitative index of pathway activity. A higher positive value indicates dominance of activation, while a negative value indicates repression. Published use cases include pro- versus anti-inflammatory macrophage states[14], differentiation versus stemness[15], or metabolic activation versus inhibition, within a single continuous framework.

This closed-form calculation eliminates the randomness inherent to permutation-based enrichment methods. It also ensures that each component—expression, threshold, deviation, and directionality—corresponds to an observable biological quantity rather than a statistical abstraction.

4. Benchmarking and performance metrics: To assess biological and clinical relevance, COMPASS automatically benchmarks the resulting composite activity scores using ROC–AUC analysis, comparing the separability of defined sample groups (for example, disease vs control or responder vs non-responder).

The Area Under the Curve (AUC) is computed as:

$$AUC = P\,(C_i^A > C_j^B) \quad (6)$$

where C_i^A and C_j^B denote the composite scores of state A (e.g., treated or disease) and state B (e.g., control or healthy) samples, respectively. This provides a universal, model-agnostic measure of discriminatory power that reflects the internal biological coherence of each gene-set's activity pattern and serves as a statistical summary the same.

All calculations are performed automatically within the web interface. Users can download publication-ready boxplots and ROC curves along with associated t, F, and p statistics. The absence of stochastic resampling ensures identical results across runs, datasets, and computational environments.

5. Implementation and accessibility: COMPASS is implemented as a web-based analytical engine with an intuitive graphical interface (**Fig. S1**). The user interface is available through the Institute for Network Medicine, UC San Diego (https://compass.precsn.com/). All analyses are performed locally within the web interface to preserve data privacy. Users upload a tab- or comma-delimited matrix (txt, tsv, csv), where genes occupy rows and samples occupy columns, along with optional metadata to define comparison groups. Gene-set definitions are fully customizable, with flexibility to assign directionality weights (+/−). Once data are

uploaded, all backend computations—including thresholding, normalization, and statistical validation—are executed automatically.

Outputs include:

- **Boxplots** summarizing pathway activity across user-defined groups, accompanied by pairwise *t*-test statistics.
- **ROC–AUC plots** displaying discrimination between states with associated *F* and *p* values.
- **Composite activity tables** containing standardized *Z*-scores and composite scores for each pathway and sample.

All results are downloadable in CSV or PDF format. The deterministic nature of COMPASS ensures that the same input always yields identical output, enabling full traceability and reproducibility for regulatory submissions or multi-site validation.

In terms of accessibility, unlike conventional computational frameworks, COMPASS is designed to be intuitive and accessible to both computational and non-computational users:

- **Minimal Input Requirement**: Only a gene expression matrix is required; no need for pathway databases, sample identifier, or clinical annotations.
- **Flexible Group Comparisons**: Users can define and compare any number of sample groups directly from their uploaded data (e.g., WT vs KO, responders vs non-responders, multiple treatment conditions).
- **No Coding or Statistical Expertise Required**: All statistical analyses, such as, composite score calculation, group-wise comparisons, ROC–AUC, p-values are performed automatically.
- **Interactive and User-Friendly**: The interface enables point-and-click selection of genes, pathways, and groups, generating publication-ready figures without manual scripting (**Fig. S1a-b**).
- **Scalable and Reproducible**: Every step is deterministic, ensuring identical results across runs which is critical for regulatory submissions and clinical validation.

Reproducibility is guaranteed by the deterministic nature of the algorithm: identical inputs produce identical numerical and graphical outputs, ensuring that results can be independently verified across laboratories or regulatory reviews. Through these three core operations, i.e., thresholding to define biological boundaries, standardization to normalize deviations, and aggregation to integrate weighted gene effects, COMPASS converts raw gene-expression matrices into quantitative, deterministic indices of pathway activity. By linking mathematical transparency to biological interpretability, COMPASS enables objective, reproducible quantification of cellular logic across model systems and clinical cohorts.

Applications and Benchmarking

COMPASS supports diverse use cases (see **Fig. 1d-f**): benchmarking organoids, animal models, or New Approach Methodologies (**NAMs**) against human cohorts to quantify *humanness* and biological *relevance*[15-21]; computing composite activity scores predictive of outcomes[15]; tracking drug response or toxicity[19]; and mapping disease-driver cell states such as alveolar cytopathy, immune cell processes (reactivity vs. tolerance) or differentiation[11,14,15,22].

To demonstrate cross-cohort, cross-model, and cross-species applicability, we applied COMPASS to a range of previously published gene signatures spanning oncology, infectious disease, inflammatory disorders, and cellular state mapping (**Fig. 2a-o**). Using deterministic composite activity scores, COMPASS consistently recapitulated biological signal across human cohorts, animal models, and organoid systems—validating its capacity for ontology-free translation and cohort-level anchoring. For example, in colorectal cancer, a 50-gene stemness signature[15] quantified the efficacy of a differentiation-inducing therapeutic across cell lines, xenografts, and patient-derived organoids, revealing concordant pathway suppression in all models (**Fig. 2a–c**). In respiratory viral injury[23], COMPASS detected induction of a lung damage signature across infected human biopsies, hamster lungs, and adult stem cell–derived lung organoids (**Fig. 2d–f**), highlighting conserved molecular pathology.

Diagnostic and prognostic signatures for sepsis[24], breast cancer[25], and idiopathic pulmonary fibrosis[26] generated robust discriminatory scores aligned with clinical classifications (**Fig. 2g–i**). In two independent IBD clinical trials, a bioenergetic barrier signature[19] stratified responders from non-responders across therapeutic classes and tracked remission longitudinally (**Fig. 2j–m**). Finally, COMPASS quantified cell-state transitions—capturing induction of DATPs in fibrosis models[27] and defining graded polarization from macrophage reactivity to tolerance[14] (**Fig. 2n–o**).

Through these capabilities, COMPASS bridges computation, bench biology and translational medicine, offering a scalable and user-friendly platform for digital biomarker discovery and systems-level disease modeling[10]. Its accessibility allows any biologist to navigate data using deterministic metrics rather than probabilistic inference, bringing the mathematical logic of life to practical use in translational and regulatory settings.

Comparison with Existing Technologies and Approaches: **Table 1** highlights the distinctive features of COMPASS relative to enrichment-based and latent-variable methods. Unlike GSEA, which compares only two discrete conditions, COMPASS models graded, multi-class trajectories. Datasets spanning disease evolution, therapy response, or circadian cycles can be analyzed along continuous activation axes, allowing longitudinal tracking of biological programs such as epithelial–mesenchymal transition or immune exhaustion. By computing absolute rather than relative activity, COMPASS provides quantitative grounding for digital biomarkers that trace biological and therapeutic trajectories. But perhaps the most distinctive feature of

COMPASS is its conceptual grounding in cellular decision-making, which in turn improves *reasoning* and *explainability* of the conclusions drawn. COMPASS extends Boolean logic into analog quantification. Whereas Boolean frameworks such as BoNE[11] identifies digital "*if–then*" rules that govern cellular decisions, COMPASS quantifies the transitions as populations traverse those rules, thus integrating discrete and continuous representations of biology. Together, these approaches operationalize the **Five Rules of Biological Logic**[10]—that life computes through thresholds, maintains fidelity through feedback, and progresses along reversible, measurable continua, thereby providing a quantitative bridge between binary logic and analog biology (see *Logic of Life*[10]). Unlike enrichment-based methods (GSEA[1], GSVA[2], ssGSEA[3], PLAGE[4], AUCell[5], and VAM[6]) that depend on ranked gene lists and random permutations, COMPASS deterministically measures biological activity without reference to predefined ontologies. Each variable—expression, threshold, deviation, or directionality—corresponds to an observable biological quantity, not a statistical abstraction. In contrast to matrix-factorization or deep-learning models such as PCA, ICA, NMF, MOFA^{+}, or variational autoencoders[28,29]), COMPASS avoids latent dimensions, retaining interpretability while eliminating stochastic variation.

Because thresholds are derived directly from data, COMPASS is **platform-agnostic** and transferable across tissues, cohorts, and species, aligning biological logic independent of normalization or batch effects. This transparency turns COMPASS from a statistical model into a **mechanistic measurement system**, that produces deterministic, reproducible metrics that can serve as **digital biomarkers** for disease state, drug response, or biological fidelity. It also enables **cross-study harmonization**, generating quantitative measures that capture subtle biological transitions with reproducible precision.

Unlike GSEA, which can only compare *two* discrete conditions, COMPASS models **graded, multi-class trajectories**. If datasets include samples representing the evolution from health to disease, normal to carcinoma, or response to resistance, or even circadian rhythms, COMPASS can capture these transitions, enabling **longitudinal tracking of phenotypes**. This makes COMPASS uniquely capable of quantifying therapeutic trajectories (e.g., during biologic therapy) or disease progression (e.g., rising stemness along the adenoma–carcinoma continuum). Thus, COMPASS **transforms pathway activity into a universal digital biomarker,** enabling objective, reproducible alignment of human, organoid, and animal models within a single logical framework for systems medicine.

Discussion

COMPASS has a few limitations. Although deterministic and interpretable, COMPASS depends on well-normalized transcriptomic data and assumes unimodal gene distributions, potentially limiting performance in sparse or noisy single-cell datasets. Its current weighting scheme treats all genes equally, without accounting for causal topology or network hierarchy. Future versions will integrate context-dependent weighting based on Boolean network connectivity to strengthen causal inference. At present, COMPASS models transcriptome-level activity; extending its logic to proteomic, metabolomic, and epigenomic data will test its generalizability. Integration with longitudinal and spatial datasets may enable trajectory mapping of therapeutic responses and disease progression, while embedding COMPASS in interactive dashboards could facilitate real-time digital biomarker tracking for translational and regulatory use.

Data Availability: Example expression matrices and gene-signature files used to generate Figure 2 and related analyses are available at our GitHub repository: https://github.com/sinha7290/COMPASS. The COMPASS web application implementing the composite activity scoring framework is freely accessible at https://compass.precsn.com/. Further information and requests for resources should be directed to and will be fulfilled by, Saptarshi Sinha, sasinha@health.ucsd.edu.

Acknowledgments: The authors wish to apologize for not citing many important publications on this topic due to the limit of references allowed. P.G. is supported by NIH grants R01-AI141630 and R01-AI55696. PG was also supported by the Leona M. and Harry B. Helmsley Charitable Trust and the Propel a Cure Foundation. S.S. was supported through The American Association of Immunologists (AAI) Intersect Fellowship Program for Computational Scientists and Immunologists.

Declaration of interests: The authors declare no conflicts of interest.

Materials and methods

Gene expression datasets and preprocessing: Previously published bulk gene-expression datasets used for benchmarking COMPASS are described in the main text. For each study, normalized expression matrices were obtained from the original publication, GEO/ArrayExpress, or were reprocessed from raw counts using standard RNA-seq pipelines (alignment, quantification, and normalization). When raw data were available, counts were converted to transcripts per million (TPM) or counts-per-million and $\log_2$-transformed after adding a small offset (typically $\log_2(x + 1)$). When only processed data were available (e.g., $\log_2$ intensity values from microarrays), those matrices were used directly after confirming that sample-wise distributions were unimodal and approximately comparable across groups. Genes with zero or near-zero variance across samples were excluded because they do not contribute meaningfully to composite scores. All COMPASS analyses were performed on gene-by-sample matrices in which rows correspond to genes and columns correspond to samples, with samples annotated by user-defined biological or clinical group labels.

Gene signatures: For all analyses in this study, gene lists and direction assignments were curated from prior experimental work, literature-derived pathways, or Boolean implication networks, as detailed in the main text.

COMPASS composite activity scoring: Composite activity scores were calculated using COMPASS as defined in Equations (1–5) in the main text. Briefly, $\log_2$-normalized gene expression matrices (e.g., TPM or counts-per-million) were used as input. For each gene, COMPASS estimated a data-driven activation threshold by fitting a step function to the ordered expression values and identifying the position that minimized the within-group variance, thereby classifying samples as “low” or “high” expressors. Expression deviations from this threshold were then standardized by the across-sample standard deviation to generate direction-aware z-scores. For each predefined gene set or signature, per-sample composite activity scores were obtained as the weighted average of these z-scores, with +1 weights assigned to genes expected to increase and −1 weights to genes expected to decrease with pathway activation. The resulting scores are deterministic (no permutations) and were used to generate boxplots, group-wise comparisons, and receiver operating characteristic (ROC) curves for user-specified biological or clinical contrasts. All analyses were performed using the COMPASS web application (https://compass.precsn.com).

Quantification and Statistical Analysis: For each gene module or signature, COMPASS computed a single continuous composite activity score per sample as described in the main text. These scores served as the primary quantitative readout for all downstream analyses. For group-wise comparisons, samples were assigned to user-defined biological or clinical groups, and composite scores were compared using two-sided Welch’s t-tests (unequal variances). Raw p-values were assembled into a group × group matrix and adjusted for multiple testing using both Bonferroni and Benjamini–Hochberg false discovery rate (FDR) procedures; both raw and adjusted values were exported for interpretation. For classification performance, a designated

reference group was treated as the negative class and each other group as the positive class. Receiver operating characteristic (ROC) curves and corresponding areas under the curve (AUCs) were computed from composite scores using standard empirical estimators, and AUC p-values were obtained from asymptotic normal approximations. All tests were two-sided and based on complete-case data. No permutations, resampling, or stochastic optimization were used, so given the same expression matrix, signature file, and group labels, COMPASS returns identical scores and statistical outputs. Analyses were implemented in Python within the COMPASS web application using standard scientific libraries (numpy, pandas, scipy, scikit-learn), with visualization modules generating boxplots and ROC curves for export.

Tables and Figures

Table 1 | Comparative features of COMPASS versus existing gene-set activity frameworks. Acronyms: GSEA, Gene Set Enrichment Analysis; GSVA, Gene Set Variation Analysis; ssGSEA, Single-Sample Gene Set Enrichment Analysis; PLAGE, Pathway-Level Analysis of Gene Expression; AUCell, Area-Under-the-Curve Cell Scoring; VAM, Variance-Adjusted Mahalanobis.

Feature	COMPASS	GSEA	GSVA	ssGSEA	PLAGE	AUCell / VAM
Full Name	COMPosite Activity Scoring System	Gene Set Enrichment Analysis	Gene Set Variation Analysis for microarray & RNA-Seq	Single-Sample Gene Set Enrichment Analysis	Pathway Level Analysis of Gene Expression	AUC-based Cell Scoring & Variance-Adjusted Mahalanobis
Key Feature	Deterministic, ontology-free, web-based digital biomarker quantification	Permutation-based enrichment; ontology-dependent	Unsupervised variation of gene-set enrichment across samples	Per-sample enrichment scoring; extension of GSEA	SVD-based modeling of pathway activity	Single-cell activity scoring via AUC and variance-adjusted methods
Lead Institution	Institute for Network Medicine, UC San Diego, USA	Broad Institute, MIT-Harvard	Universitat Pompeu Fabra & Sage Bionetworks	Broad Institute, MIT-Harvard	Columbia University & KRIBB (Korea)	Center for Genomic Regulation (Spain) / University of Alabama
Publication (Year)	Sinha & Ghosh, 2025 (This Study)	Subramanian et al., PNAS 2005[1]	Hänzelmann et al., BMC Bioinformatics 2013[2]	Barbie et al., Nature 2009[3]	Kim & Volsky, BMC Bioinformatics 2005[4]	Aibar et al., Nature Methods 2017[5]; Frost et al., Bioinformatics 2021[6]
Data-derived thresholding	Yes (adaptive per gene)	No	No	No	No	No
Deterministic output (run-to-run identical)	Yes	No (permutation-based)	Yes	Yes	Yes	Yes
Coding required to run	No (web app)	No (web app / desktop client)	Yes (R)	Yes (R)	Yes (R)	Yes (R / Python)
Assigned gene weights (+ / − per gene)	Yes (activation vs repression encoded)	No	No	No	No	No
Absolute per-sample activity score (digital biomarker)	Yes (composite activity index)	No (relative enrichment only)	Yes	Yes	Yes	Yes
Ontology-independent (custom gene sets allowed)	Yes (fully user-defined)	Yes	Yes	Yes	Yes	Yes
Single-sample scoring	Yes	No (requires two-class contrast)	Yes	Yes	Yes	Yes
Handles >2 groups / graded trajectories	Yes (multi-class and continua)	No (two-group only)	Yes	Yes	Yes	Yes
Robust with very small n (<5 samples / group)	Yes	No (needs group sizes for permutations)	Moderate	Moderate	Moderate	Yes
Noise handling via intrinsic, data-driven thresholds	Yes (per-gene decision boundary)	No	No	No	No	No
Speed	Very fast (closed form; no permutations)	Slow (permutation testing)	Fast	Fast	Fast	Fast
Interpretability (reasoning and explainability)	High (Grounded in Foundational Model of Biological Logic[10])	Moderate, knowledge based (enrichment probability)	Moderate, knowledge based	Moderate, knowledge based	Moderate, knowledge based	Moderate, knowledge based
Transparency of computation	High (no latent dimensions / no sampling)	Low (randomization, ranked lists)	Medium	Medium	Medium	Medium
Cross-dataset reproducibility	High (deterministic, platform-agnostic)	Low–Medium	Medium	Medium	Medium	Medium
Availability / licensing	Open source	Open source	Open source (R package)	Open source (R package)	Open source (R package)	Open source
Web-based interface	Yes (no coding required)	Yes (Broad GSEA Web / desktop app)	No	No	No	Partial (via SCENIC / portals)

Table 1. Comparative features of COMPASS versus existing gene-set activity frameworks.

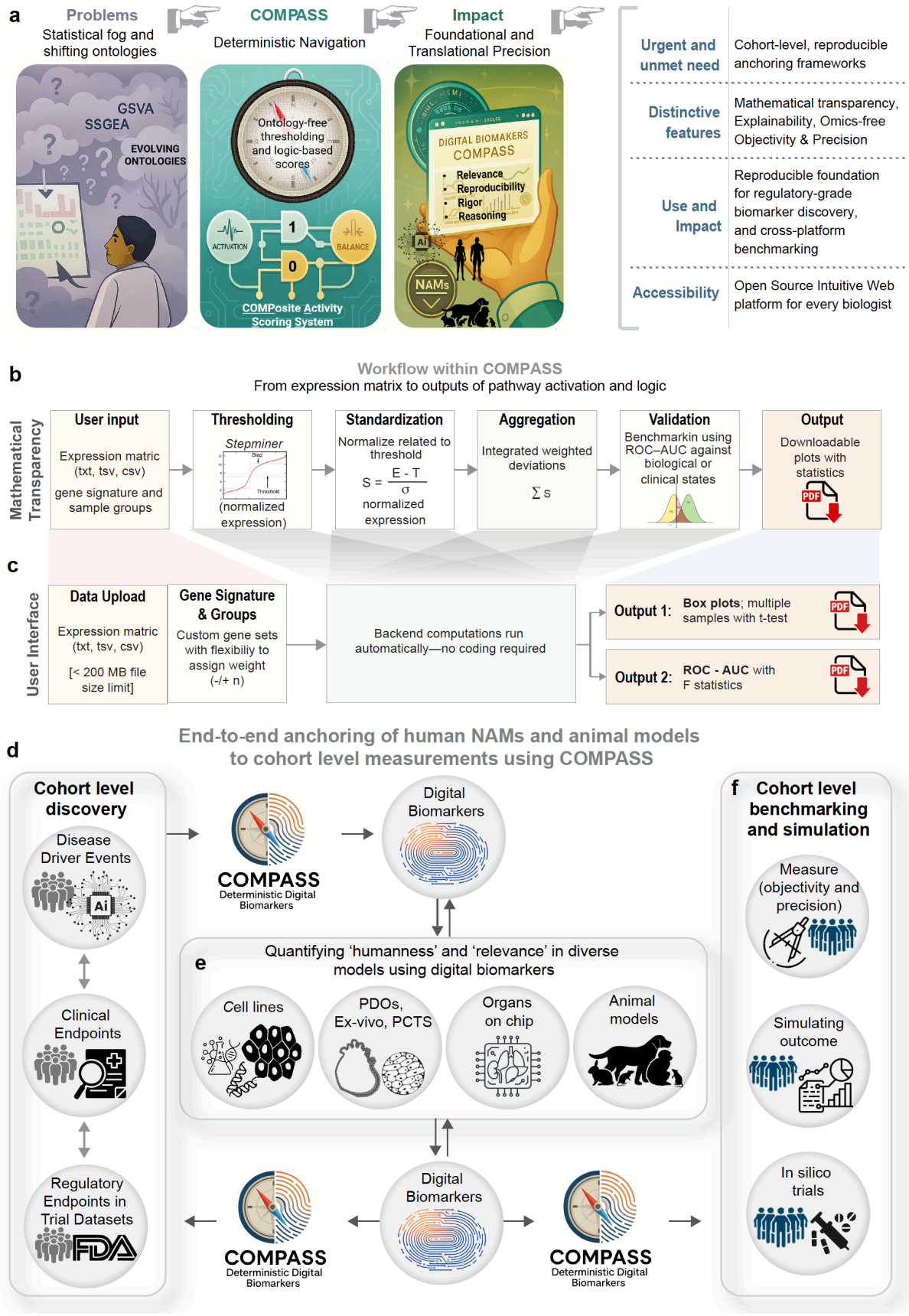

Figure 1. Overview of the COMPASS Framework for translational Integration of Digital Biomarkers Across Models and Cohorts Using COMPASS.

The conceptual foundation (a), computational workflow (b), and workflow for translational integration of digital biomarkers for analyzing gene expression data to derive logic-based pathway activation scores and digital biomarkers.

(a) Conceptual Foundation: COMPASS addresses limitations of traditional logic modeling by offering a deterministic, reproducible, and accessible framework for translational precision medicine.

(b) Computational Workflow: Workflow schematic showing user inputs (expression matrix, gene signature, logic), automated backend computations, and outputs including box plots and ROC-AUC analyses.

(c) User Interface: Example of an intuitive interface and downloadable outputs demonstrating COMPASS's ability to stratify groups and assess pathway activation with statistical significance, while remaining accessible to all biologists.

(d-e) Translational Workflow: Workflow schematic illustrates a three-part framework for aligning diverse biological models with cohort-level measurements through digital biomarkers and the COMPASS system. (**c**) **Cohort-Level Discovery**: Disease driver events, clinical endpoints, and regulatory endpoints are identified from large-n human trial datasets. These serve as anchors for downstream modeling and validation. (**d**) **Model-Level Quantification**: Diverse experimental platforms—including cell lines, patient-derived organoids (PDOs), ex vivo tissues (e.g., precision-cut tissue slices, live or fixed), organs-on-chip, and animal models—are evaluated using digital biomarkers. COMPASS objectively quantifies the humanness and relevance of each model's response, enabling cross-comparison and translational alignment. (**e**) **Cohort-Level Benchmarking and Simulation**: Validated digital biomarkers are used to simulate cohort-level responses with high objectivity and precision. In silico trials further extend predictive capacity, supporting regulatory decision-making and personalized medicine. Central to this framework, COMPASS acts as a translational bridge linking small-n mechanistic studies with large-n population insights and enabling iterative refinement of digital biomarkers across biological scales.

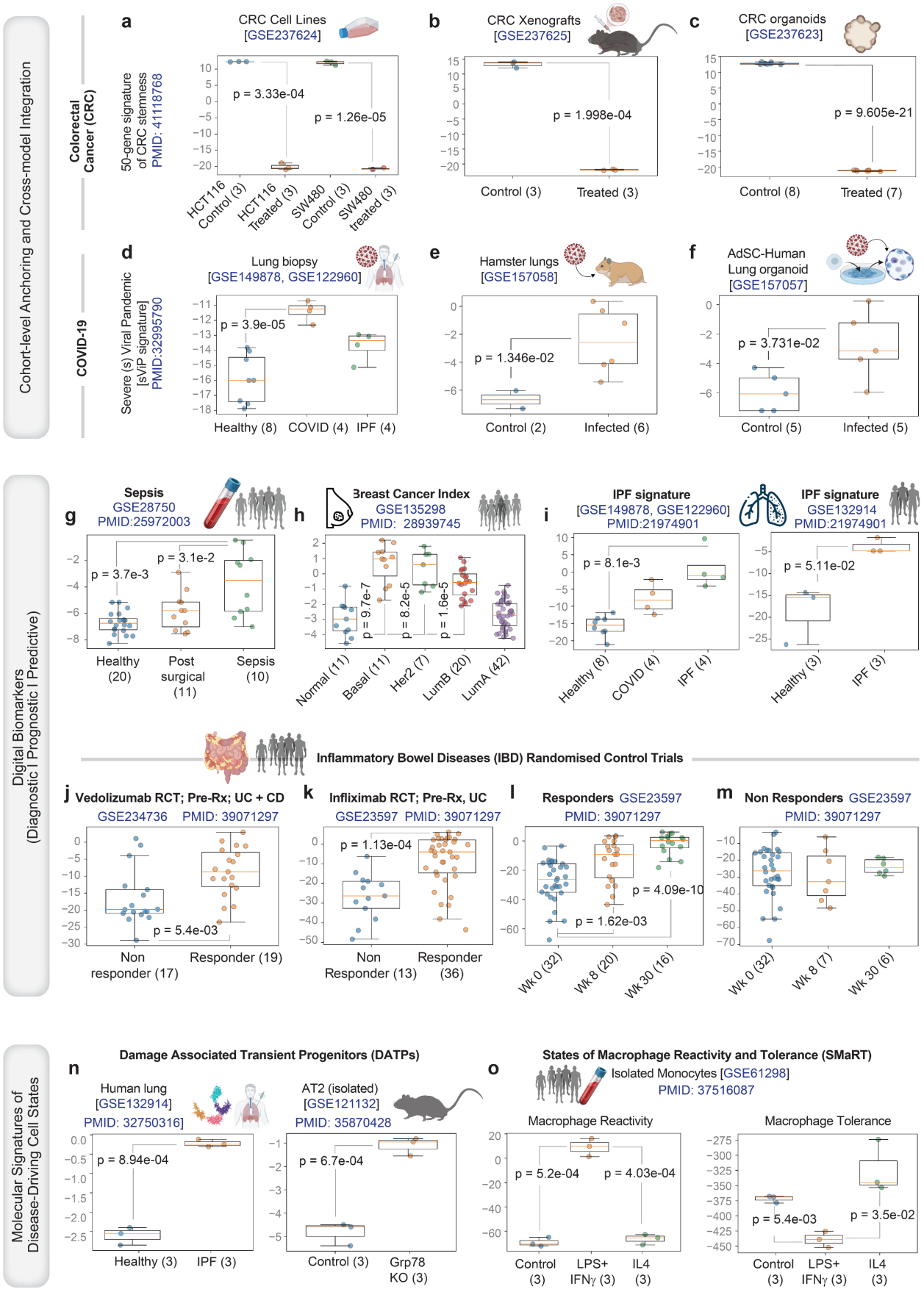

Figure 2. Cohort-level anchoring and cross-model benchmarking using COMPASS.

COMPASS enables deterministic quantification of pathway activity across cohorts, tissues, species, and disease models, allowing unified navigation of biological logic and clinical relevance. Published examples illustrating such cross-species and cross-model ontology-free navigation between human, organoid, and animal systems are shown.

(a–c) *Colorectal cancer (CRC)*: A 50-gene CRC stemness signature[15] was used to score the efficacy of a first-in-class differentiation inducing therapeutic agent across cell lines (a), xenografts (b), and patient-derived organoids (c), demonstrating consistent efficacy across experimental models[15].

(d–f) *COVID-19 and lung injury*: A 20-gene signature[23] indicative of severe alveolar damage in the setting of respiratory viral pandemics is induced in CoV-infected human lungs[30] (d), hamster lungs[31] (e), and human adult stem cell-derived lung organoids[21] (f), recapitulating conserved pathway activation across species and model systems (datasets and PMIDs annotated).

(g–i) *Diagnostic and prognostic digital biomarkers*: Distinct gene sets corresponding to sepsis (g; 11-gene diagnostic gene set, identified using a comprehensive time-course-based multicohort analysis of sepsis[24]), breast cancer subtypes (h; 7-gene **BCI**™ signature, the only genomic test that is both prognostic and predictive and is recognized by the NCCN and the ASCO® Clinical Practice Guideline to identify patients who might benefit from adjuvant anti-estrogen therapy beyond 5 years), and idiopathic pulmonary fibrosis (I; a 153-gene IPF diagnostic signature[26], comprised of 76 up and 77 down genes) generated robust discriminatory activity scores, matching known clinical groupings.

(j–m) *Digital biomarkers for therapeutic response prediction and tracking*: A 34-gene signature[19], indicative of mucosal barrier bioenergetic state, predicts therapeutic response in Inflammatory bowel disease (IBD) clinical trial datasets, distinguishing responders (high-34) from non-responders (low-34), regardless of class of drug (j-k) and tracks remission during therapy among responders in longitudinal datasets. Wk, week; UC, Ulcerative colitis; CD, Crohn's disease.

(n–o) *Molecular signatures of disease-driving cell states*: A 6-gene signature of damage-associated transient progenitors (DATP[27]) is induced in human IPF and murine models that recapitulate fundamental processes in IPF (i.e., impaired ER stress response[31]; n). Gene signatures of macrophage[14] reactivity (48 genes) and tolerance (232 genes) accurately captured macrophage polarization continua (reactivity vs tolerance; o) across cytokine-stimulated monocytes, revealing deterministic transitions that define physiologic and pathologic spectra.

REFERENCES:

1. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550. 10.1073/pnas.0506580102.

2. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7. 10.1186/1471-2105-14-7.

3. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108-112. 10.1038/nature08460.

4. Tomfohr, J., Lu, J., and Kepler, T.B. (2005). Pathway level analysis of gene expression using singular value decomposition. BMC Bioinformatics *6*, 225. 10.1186/1471-2105-6-225.

5. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat Methods *14*, 1083-1086. 10.1038/nmeth.4463.

6. Frost, H.R. (2020). Variance-adjusted Mahalanobis (VAM): a fast and accurate method for cell-specific gene set scoring. Nucleic Acids Res *48*, e94. 10.1093/nar/gkaa582.

7. Timmons, J.A., Szkop, K.J., and Gallagher, I.J. (2015). Multiple sources of bias confound functional enrichment analysis of global -omics data. Genome Biol *16*, 186. 10.1186/s13059-015-0761-7.

8. Wijesooriya, K., Jadaan, S.A., Perera, K.L., Kaur, T., and Ziemann, M. (2022). Urgent need for consistent standards in functional enrichment analysis. PLoS Comput Biol *18*, e1009935. 10.1371/journal.pcbi.1009935.

9. Haynes, W.A., Tomczak, A., and Khatri, P. (2018). Gene annotation bias impedes biomedical research. Sci Rep *8*, 1362. 10.1038/s41598-018-19333-x.

10. Sinha, S., and Ghosh, P. (2025). AXIOM-Bio: A Foundational Model for Biological Logic and Disease. Zenodo. https://doi.org/10.5281/zenodo.17459815.

11. Sahoo, D., Swanson, L., Sayed, I.M., Katkar, G.D., Ibeawuchi, S.R., Mittal, Y., Pranadinata, R.F., Tindle, C., Fuller, M., Stec, D.L., et al. (2021). Artificial intelligence guided discovery of a barrier-protective therapy in inflammatory bowel disease. Nat Commun *12*, 4246. 10.1038/s41467-021-24470-5.

12. Sahoo, D. (2012). The power of boolean implication networks. Front Physiol *3*, 276. 10.3389/fphys.2012.00276.

13. Sahoo, D., Dill, D.L., Tibshirani, R., and Plevritis, S.K. (2007). Extracting binary signals from microarray time-course data. Nucleic Acids Res *35*, 3705-3712. 10.1093/nar/gkm284.

14. Ghosh, P., Sinha, S., Katkar, G.D., Vo, D., Taheri, S., Dang, D., Das, S., and Sahoo, D. (2023). Machine learning identifies signatures of macrophage reactivity and tolerance that predict disease outcomes. EBioMedicine *94*, 104719. 10.1016/j.ebiom.2023.104719.

15. Sinha, S., Alcantara, J., Perry, K., Castillo, V., Ondersma, A.K., Banerjee, S., McLaren, E., Espinoza, C.R., Taheri, S., Vidales, E., et al. (2025). CANDiT: A machine learning framework for differentiation therapy in colorectal cancer. Cell Rep Med, 102421. 10.1016/j.xcrm.2025.102421.

16. Tindle, C., Fonseca, A.G., Taheri, S., Katkar, G.D., Lee, J., Maity, P., Sayed, I.M., Ibeawuchi, S.R., Vidales, E., Pranadinata, R.F., et al. (2024). A living organoid biobank of patients with Crohn's disease reveals molecular subtypes for personalized therapeutics. Cell Rep Med *5*, 101748. 10.1016/j.xcrm.2024.101748.

17. Sinha, S., Castillo, V., Espinoza, C.R., Tindle, C., Fonseca, A.G., Dan, J.M., Katkar, G.D., Das, S., Sahoo, D., and Ghosh, P. (2022). COVID-19 lung disease shares driver AT2 cytopathic features with Idiopathic pulmonary fibrosis. EBioMedicine *82*, 104185. 10.1016/j.ebiom.2022.104185.

18. Vo, D.T., Fuller, M.R., Tindle, C., Anandachar, M.S., Das, S., Sahoo, D., and Ghosh, P. (2021). SPT6 loss permits the transdifferentiation of keratinocytes into an intestinal fate that resembles Barrett's metaplasia. iScience *24*, 103121. 10.1016/j.isci.2021.103121.

19. Sinha, S., McLaren, E., Mullick, M., Penrose, H.M., Tindle, C., Singh, S., Boland, B.S., and Ghosh, P. (2025). F.O.R.W.A.R.D: A Data-Driven Framework for Network-Based Target Prioritization in Drug Discovery. bioRxiv. 10.1101/2024.07.16.602603.

20. Penrose, H.M., Sinha, S., Tindle, C., Zablan, K., Le, H.N., Neill, J., Ghosh, P., and Boland, B.S. (2025). A Living Organoid Biobank of Crohn's Disease Patients Reveals Distinct Clinical Correlates of Molecular Subtypes of Disease. medRxiv. 10.1101/2025.04.01.25325058.

21. Tindle, C., Fuller, M., Fonseca, A., Taheri, S., Ibeawuchi, S.R., Beutler, N., Katkar, G.D., Claire, A., Castillo, V., Hernandez, M., et al. (2021). Adult stem cell-derived complete lung organoid models emulate lung disease in COVID-19. Elife *10*. 10.7554/eLife.66417.

22. Sinha, S., Farfel, A., Luker, K.E., Parker, B.A., Yeung, K.T., Luker, G.D., and Ghosh, P. (2024). Growth signaling autonomy in circulating tumor cells aids metastatic seeding. PNAS Nexus *3*, pgae014. 10.1093/pnasnexus/pgae014.

23. Sahoo, D., Katkar, G.D., Khandelwal, S., Behroozikhah, M., Claire, A., Castillo, V., Tindle, C., Fuller, M., Taheri, S., Rogers, T.F., et al. (2021). AI-guided discovery of the invariant host response to viral pandemics. EBioMedicine *68*, 103390. 10.1016/j.ebiom.2021.103390.

24. Sweeney, T.E., Shidham, A., Wong, H.R., and Khatri, P. (2015). A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci Transl Med *7*, 287ra271. 10.1126/scitranslmed.aaa5993.

25. Zhang, Y., Schroeder, B.E., Jerevall, P.L., Ly, A., Nolan, H., Schnabel, C.A., and Sgroi, D.C. (2017). A Novel Breast Cancer Index for Prediction of Distant Recurrence in HR. Clin Cancer Res *23*, 7217-7224. 10.1158/1078-0432.CCR-17-1688.

26. Meltzer, E.B., Barry, W.T., D'Amico, T.A., Davis, R.D., Lin, S.S., Onaitis, M.W., Morrison, L.D., Sporn, T.A., Steele, M.P., and Noble, P.W. (2011). Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. BMC Med Genomics *4*, 70. 10.1186/1755-8794-4-70.

27. Choi, J., Park, J.E., Tsagkogeorga, G., Yanagita, M., Koo, B.K., Han, N., and Lee, J.H. (2020). Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. Cell Stem Cell *27*, 366-382.e367. 10.1016/j.stem.2020.06.020.

28. Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J.C., and Stegle, O. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol *21*, 111. 10.1186/s13059-020-02015-1.

29. Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner, F., Huber, W., and Stegle, O. (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol *14*, e8124. 10.15252/msb.20178124.

30. Ackermann, M., Verleden, S.E., Kuehnel, M., Haverich, A., Welte, T., Laenger, F., Vanstapel, A., Werlein, C., Stark, H., Tzankov, A., et al. (2020). Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. N Engl J Med *383*, 120-128. 10.1056/NEJMoa2015432.

31. Borok, Z., Horie, M., Flodby, P., Wang, H., Liu, Y., Ganesh, S., Firth, A.L., Minoo, P., Li, C., Beers, M.F., et al. (2020). Grp78 Loss in Epithelial Progenitors Reveals an Age-linked Role for Endoplasmic Reticulum Stress in Pulmonary Fibrosis. Am J Respir Crit Care Med *201*, 198-211. 10.1164/rccm.201902-0451OC.