Supplementary Methods and Extended Results

Commutator-Based Prioritization of Metabolite Measurements under Partial Observability

Anas Enoch, MD
Mohammed VI University of Health Sciences (UM6SS)

Note. This supplement contains material moved from the main manuscript to meet journal length requirements, together with extended derivations, ablation analyses, and per-cohort benchmark results. Section labels follow the main manuscript where applicable; cross-references to main-text sections are noted inline. All statistical claims and formal results in this supplement correspond to methods and references cited in the main manuscript.

Contents

Supplementary Methods

0.1 Limitations of enumerative and imputation-based strategies

A direct strategy to assess ambiguity would enumerate possible completions of missing metabolites and evaluate pathway consistency across all admissible configurations. However, the number of latent metabolite assignments grows combinatorially with pathway size, rendering exhaustive enumeration computationally infeasible. Moreover, enumeration obscures structural sources of ambiguity and does not directly inform measurement prioritization.

Imputation-based methods avoid combinatorial explosion but collapse uncertainty into point estimates. As a consequence, they may produce pathway interpretations that appear stable despite being poorly constrained by the underlying measurements.

These limitations motivate a framework that (i) preserves structural uncertainty explicitly, (ii) quantifies pathway-level ambiguity without enumerating latent states, and (iii) links ambiguity quantification to experimental design.

0.2 Pathway underdetermination under partial metabolite observability

Terminology note. We use *identifiability* loosely to denote the stability of pathway interpretation under partial observation, rather than classical parameter identifiability in the dynamical-systems sense (e.g., Raue et al. [1], which concerns unique parameter recovery from time-series data). Where precision is required, we use *underdetermination* to denote the degree to which a pathway state is ambiguous given partial metabolite coverage.

In targeted and semi-targeted metabolomics, pathway activity is inferred from incomplete metabolite panels. Under such partial observability, distinct biochemical configurations may produce indistinguishable measured profiles. This creates structural ambiguity that cannot be resolved without additional measurements.

We formalize this phenomenon through a pathway-level underdetermination functional U_k, which quantifies how strongly pathway k remains compatible with multiple biochemical configurations given the observed metabolite panel.

The functional integrates three components:

- Transport entropy, capturing dispersion of feasible pathway alignments under partial observation,
- Alignment instability across conditions, reflecting structural sensitivity to masking,
- A structural growth index (SGI), quantifying expansion of reachable biochemical configurations when metabolites are unobserved.

High U_k values indicate that pathway interpretation is structurally unstable and should not be considered reliable without additional targeted measurements.

Measurement prioritization. Measurement selection is formulated as an optimization problem over the sensitivity of $\mathcal{U}_k$ with respect to candidate metabolite observations:

$$m^\star = \arg \max_{m \in \mathcal{M}_{\text{unobs}}} \widehat{\Delta}_m \mathcal{U}_k. \tag{1}$$

Together, these operators define a unified computational pipeline for pathway comparison, ambiguity quantification, and measurement-driven disambiguation.

0.3 Scope and applicability

Although motivated by metabolomics, the operators apply to partially observed pathway graphs. We focus this study on metabolomics settings and leave other modalities to future work.

Future work will evaluate prospective panel design in experimental settings where the recommended metabolite is actually measured and pathway ambiguity reassessed. Integration with isotope tracing data would provide orthogonal benchmark evaluation of the ambiguity-reduction claim independent of the masking protocol.

Supplementary Methods S1. Operator stability benchmark

S1.1 Benchmark design

Six preprocessing operators were evaluated under three perturbation families (transition-kernel noise at $\rho \in \{0.05, 0.10, 0.20, 0.30\}$; metabolite dropout; bootstrap resampling) across three primary cohorts. For each combination, the coefficient of variation of FGW distance ($\mathrm{CV}_{\mathrm{FGW}}$), transport-plan drift, top-3 Jaccard stability, and cross-pathway Kendall τ were recorded.

S1.2 Operator selection rule

The stable operator is selected by lowest mean $\mathrm{CV}_{\mathrm{FGW}}$ across perturbation regimes. In the low-node regime tested here ($|\mathcal{M}_k| \leq 25$), non-projection operators achieve $\mathrm{CV}_{\mathrm{FGW}} < 0.03$ and $\tau > 0.96$, while projection-based operators (`jl`, `randproj`) achieve $\mathrm{CV}_{\mathrm{FGW}} \approx 0.31$–$0.34$ and $\tau \approx 0.33$–0.34.

Table 1: *

Supplementary Table S1. Operator stability summary (primary cohorts, all perturbation regimes combined). The selected operator (**none**) has the lowest $\mathrm{CV}_{\mathrm{FGW}}$ and highest cross-pathway τ. Projection operators are excluded from the primary benchmark.

Operator	Mean $\mathrm{CV}_{\mathrm{FGW}}$	Transport drift	Top-3 Jaccard	Kendall τ
none	**0.023**	**0.019**	**0.94**	**0.97**
l2	0.026	0.022	0.93	0.96
pca_fixed	0.028	0.031	0.91	0.95
pca_var95	0.027	0.028	0.92	0.94
jl	0.331	0.931	0.41	0.33
randproj	0.328	0.929	0.42	0.34

S1.3 Why projection operators are unstable in this regime

Projecting from $d \approx 22$ features to $d_{\mathrm{proj}} = 8$ introduces projection variance that dominates perturbation-induced variance in pathways with $|\mathcal{M}_k| \leq 8$. Under fixed-scale normalization, compressed features produce cost matrices whose entries are small relative to the Sinkhorn regularization parameter $\varepsilon = 0.5$, pushing transport plans toward uniformity. Under max-normalization this effect is masked; the normalization fix (Figure 4 in the main manuscript) is therefore a prerequisite for valid operator comparison rather than merely a preprocessing choice.

Supplementary Methods S2. Regime-dependent findings

S2.1 Operating-regime characterisation

The commutator advantage is pathway-size dependent. In the primary cohorts, where pathways have $|\mathcal{M}_k| \leq 25$ and curated edge topology is available, the advantage is large ($\tau = 0.599$ vs 0.086). In the external COVID-19 cohorts, the advantage is preserved for $|\mathcal{M}_k| \leq 40$ nodes but collapses to $\delta\tau \approx -0.006$ for large lipid pathways ($|\mathcal{M}_k| > 40$).

The collapse mechanism: when all metabolites are highly correlated, the correlation-distance structure matrix $C_{ij} = \sqrt{2(1 - r_{ij})}$ becomes approximately flat. The commutator's operator-geometry signal saturates because operator incompatibility $\|[L_{\text{mask}}, E_m]\|_F$ converges to the same value for every candidate. Variance-based baselines, which ignore structure entirely, become competitive in this regime.

S2.2 Supplementary Table S2: per-pathway external cohort results

Table 2: *

Supplementary Table S2. Commutator vs variance: per pathway-size-class results on external cohorts (hard subset, $n_{\text{hidden}} \geq 2$). $\delta\tau = \tau_{\text{comm}} - \tau_{\text{var}}$. The operating-regime boundary at $|\mathcal{M}_k| = 40$ is visible in both cohorts.

Cohort	Size class	Comm. τ	Var. τ	$\delta\tau$	n pathways
ST001849 (LC/MS)	$\|\mathcal{M}_k\| \leq 6$	0.400	0.190	+0.210	5
ST001849	$\|\mathcal{M}_k\| = 7$–15	0.292	0.191	+0.101	7
ST001849	$\|\mathcal{M}_k\| = 16$–40	0.175	0.072	+0.103	4
ST001849	$\|\mathcal{M}_k\| > 40$	0.061	0.067	−0.006	3
ST002829 (Metabolon)	$\|\mathcal{M}_k\| \leq 6$	0.374	0.154	+0.220	6
ST002829	$\|\mathcal{M}_k\| = 7$–15	0.301	0.183	+0.118	8
ST002829	$\|\mathcal{M}_k\| = 16$–40	0.183	0.086	+0.097	4
ST002829	$\|\mathcal{M}_k\| > 40$	0.064	0.070	−0.006	3

The $\delta\tau \approx -0.006$ collapse at $|\mathcal{M}_k| > 40$ is reproduced quantitatively across both cohorts and both platforms (LC/MS and Metabolon), confirming it as a regime property rather than a dataset artefact. Full per-pathway results are available in the project repository (`ST001849_benchmark_results.csv`, `ST002829_benchmark_results.csv`).

Supplementary Methods S3. Information-theoretic and Bayesian baselines

The following three baselines were added in the final revision to provide stronger comparison points. All are evaluated under the same masking protocol and reveal-defined oracle as the other predictors.

S3.1 Mutual information baseline (`mutual_info`). For each candidate hidden metabolite $m \in \mathcal{H}_k$, the mutual information between the metabolite values and the binary condition label is estimated using the Kraskov–Stögbauer–Grassberger (KSG) k-nearest neighbour estimator [2] on the masked pathway feature matrix. The candidate with the highest estimated mutual information is ranked first. This baseline captures purely marginal statistical dependence on the condition label, without accounting for pathway structural relationships.

S3.2 Bayesian optimal experimental design heuristic (`bayes_oed`). Following the framework of [3], a Bayesian OED heuristic scores each candidate by the expected information gain (EIG) it would provide to the posterior over pathway state. The posterior is approximated as a Gaussian over the FGW alignment parameters, and the EIG is approximated by the marginal increase in the trace of the Fisher information matrix when candidate m is added to the observed set:

$$s_{\mathrm{OED}}(m) = \mathrm{tr}[\mathcal{I}(\theta \mid \mathcal{O}_k \cup \{m\}) - \mathcal{I}(\theta \mid \mathcal{O}_k)]\,,$$

where $\mathcal{I}(\theta \mid \cdot)$ denotes the Fisher information of the FGW alignment parameters given the observed metabolites. The Fisher matrix contribution is approximated via a first-order Taylor expansion of the FGW objective around the current transport plan.

S3.3 Active feature acquisition (`active_acq`). A greedy single-step active feature acquisition rule following the myopic approximation of [4]. For each candidate m, a logistic regression classifier is retrained on the observed feature set augmented with the candidate's values (imputed from training data via k-NN for the evaluation step only). The candidate is scored by the predicted reduction in cross-entropy loss on a held-out validation set:

$$s_{\mathrm{AFA}}(m) = \mathcal{L}(\hat{y} \mid \mathcal{O}_k) - \mathcal{L}(\hat{y} \mid \mathcal{O}_k \cup \{m\}).$$

This captures the single-step classification benefit of acquiring m, without considering pathway structural relationships.

Computed results on primary cohorts. All three baselines were evaluated on the same masking protocol as the primary benchmark (ST000356, ST003390, ST003506):

Predictor	Regret	Top-1	Top-3	τ
`bayes_oed`	0.165	0.568	0.697	0.100
`active_acq`	0.159	0.516	0.702	0.017
`mutual_info`	0.257	0.372	0.696	−0.517

Interpretation. A three-tier τ hierarchy emerges across all eleven predictors: operator-geometric ($\tau = 0.40$–0.60) > connectivity-weighted differential ($\tau = 0.10$–0.20) > label-marginal ($\tau = -0.52$ to 0.09). The `mutual_info` result ($\tau = -0.517$) is the strongest negative finding: mutual information against the condition label is actively anti-correlated with the structural oracle in 52.7% of trials. This occurs because the oracle is defined by pathway-alignment geometry, not classification signal. The `bayes_oed` and `active_acq` baselines show positive but weak τ (0.100 and 0.017), indicating that information-theoretic reasoning about the condition label captures some structural signal but substantially less than the operator-commutator.

Supplementary Methods S4. Targeted acquisition simulation protocol

The targeted acquisition simulation used the following protocol, applied to ST000356 (breast cancer, $n = 149$ cases, $n = 82$ controls).

1. For each of 30 retained pathways, randomly mask 20% of metabolites as the "hidden" set (50 independent masking trials per pathway).

2. Compute the underdetermination functional $\mathcal{U}_k$ on the observed panel.

3. Apply each acquisition strategy to select the next metabolite to reveal.

4. Reveal the selected metabolite, recompute $\mathcal{U}_k$, record $\Delta\mathcal{U}_k$.

5. For the classification evaluation (Figure 10 in the main manuscript): train logistic regression on the partial panel using 70% of samples, evaluate AUC on the remaining 30%. Repeat for 1–10 sequential additions.

All imputation in step 3 (for `active_acq` only) uses k-NN with $k = 5$ on the training split; imputed values are used only for acquisition scoring and are not incorporated into the final panel. The simulation is a computational evaluation under controlled masking and does not involve new experimental measurements.

Supplementary Methods S5. Perturbation-oriented validation (ST001865)

S5.1 Dataset description

ST001865 is an untargeted metabolomics dataset (MetaboLights) comparing hypoxia ($n = 8$) with normoxia ($n = 8$) in a matched cell-line design. Raw data were converted from transposed matrix format using a dedicated parsing script that extracted condition labels from the metadata row, transposed the sample-metabolite matrix, and normalised metabolite identifiers. The processed dataset contained 110 metabolite features across 16 samples. Condition labels were mapped via explicit override (`CASE_CTRL_OVERRIDE = {"ST001865": ("Hypoxia","Normoxia")}`) to ensure correct case/control assignment.

S5.2 Pathway coverage

Intersection of the 110 measured metabolites with the core pathway map yielded four pathways meeting the minimum coverage threshold ($|\mathcal{M}_k| \geq 3$):

- Arginine and Proline Metabolism (6 metabolites): Arginine, Citrulline, Creatine, Creatinine, Ornithine, Proline.
- Alanine–Aspartate–Glutamate Metabolism (3 metabolites)
- Glycerophospholipid Metabolism (3 metabolites)
- Valine–Leucine–Isoleucine Degradation (3 metabolites)

The three 3-metabolite pathways produce $n_{\text{hidden}} = 1$ under any masking rate (forced selection); only Arginine and Proline Metabolism generates a non-trivial acquisition problem ($n_{\text{hidden}} = 2$–3).

S5.3 Benchmark protocol

Masking rates $\rho \in \{0.40, 0.50, 0.60\}$, 20 trials per rate (60 total), seven competing strategies, single seeded run (`GLOBAL_SEED=42`, `np.random.seed(42)`). All FGW computations used `SINK_MAX_ITER=300`, `SINK_REG=0.5`, `ALPHA_FGW=0.5`.

S5.4 Oracle metabolite frequency

Over 60 trials on Arginine and Proline Metabolism, the reveal-defined oracle selected: Citrulline 45% (27/60), Creatine 30% (18/60), Arginine 13% (8/60), Proline 12% (7/60). Citrulline's high oracle frequency reflects its structural position at the arginine/urea cycle interface: high network degree and strong cross-condition alignment perturbation.

S5.5 Bootstrap confidence intervals (Citrulline recovery)

Bootstrap resampling ($B = 10{,}000$) of the $n = 27$ Citrulline-oracle trials:

Method	Recovery rate	95% CI
`gnc_commutator`	92.6%	[81.5%, 100%]
`random`	66.7%	[48.1%, 85.2%]
`degree`	51.9%	[33.3%, 70.4%]
`mb2d_transport`	33.3%	[14.8%, 51.9%]
`variance`	3.7%	[0%, 11.1%]
`diffabundance`	3.7%	[0%, 11.1%]
`surrogate`	3.7%	[0%, 11.1%]

Fisher exact odds ratio for commutator vs variance/diffabundance/surrogate: OR= 325, $p < 0.001$. The 95% CI for the commutator does not overlap with the CI for any abundance-based heuristic.

References

[1] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[2] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.

[3] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

[4] Weihao Gong, Sebastian Tschiatschek, Sebastian Nowozin, Richard E. Turner, José Miguel Hernández-Lobato, and Cheng Zhang. Icebreaker: Element-wise efficient information acquisition with a Bayesian deep latent Gaussian model. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.