Pillar I — Multi-Omics Convergence

Proteomics · Metabolomics · Integration across biological scales | Pre-session reading brief · ~10 min read

Purpose of this brief. This document prepares you for the upcoming brainstorm session on extending CytoRNA beyond transcriptomics. It sets out the scientific problem, the landscape of available approaches, and open questions we would like the group to explore. It does not prescribe solutions — your expertise and your experimental experience are what the session is designed to draw on.

The Biological Problem

CytoRNA characterises single-cell transcriptional states with high resolution. The question driving this pillar is: how much biology are we missing by measuring RNA alone?

mRNA – protein correlation. Genome-wide Pearson correlation between mRNA abundance and protein abundance is approximately r ≈ 0.40–0.60 across mammalian cell types (Schwanhaäusser et al., *Nature* 2011; Wilhelm et al., *Nature* 2014). The gap arises from post-transcriptional regulation: miRNA-mediated repression, RBP-controlled mRNA stability, ribosome occupancy differences, and ubiquitin-proteasome turnover. High mRNA does not guarantee high protein. Low protein does not indicate low mRNA. These relationships are biologically informative in their own right and are invisible to RNA-seq.

The metabolome. ~8,000 intracellular small molecules represent the integrated biochemical flux state of a cell — the net outcome of all enzymatic activity. Metabolite concentrations reflect cellular physiology more directly than gene expression. For CardioVIVE specifically, the foetal–adult cardiomyocyte maturation transition is fundamentally a metabolic switch: from glycolysis (HIF1α-driven, Warburg-like) to β-oxidation of long-chain fatty acids (PPARα/PGC-1α axis). This transition is detectable only indirectly from scRNA-seq (via FAO enzyme gene expression proxies). Direct LC-MS measurement of acetyl-CoA, lactate/pyruvate ratio, and β-hydroxybutyrate provides biochemical ground truth.

What we cannot see from RNA alone

Sarcomeric protein stoichiometry (MYH7:MYH6, TNNI3:TNNI1 mass ratios) · Titin phosphorylation at S3991 and S12884 regulating passive stiffness · SERCA2a / PLN / RyR2 protein abundance governing Ca^{2+} cycling · Fatty acid β-oxidation flux rates · Metabolic coupling between VCM and PEDC populations

The Technology Landscape

Three complementary measurement layers are now routine in well-equipped proteomics and metabolomics facilities:

Quantitative Proteomics

DIA-MS (data-independent acquisition) — library-free quantification of ≥5,000 proteins per sample at ≤1% FDR. TMT/iTRAQ multiplexing for high-throughput cohorts. Phosphoproteomics, ubiquitinomics, and acetylomics for PTM-level resolution.

Untargeted Metabolomics

LC-MS/MS reverse-phase (lipids, acylcarnitines) and HILIC (amino acids, nucleotides, organic acids). GC-MS for volatile and derivatisable metabolites. NMR for absolute quantification of ~50 key metabolites in physiological units.

13C Isotope Tracing

13C-labelled substrates (palmitate, glucose, glutamine) allow direct measurement of metabolic flux — which pathway is actually running, not just which enzymes are expressed. MFA (metabolic flux analysis) gives quantitative flux maps.

Integration across layers is the scientifically non-trivial step. MOFA+ (Multi-Omics Factor Analysis) performs joint dimensionality reduction across paired omics matrices, decomposing shared and dataset-specific sources of variance into interpretable latent factors. PALS (Pathway Activity Level Scoring) ranks KEGG pathways by combining evidence from all three omics layers simultaneously.

The AIDO principle

The AIDO (AI-Driven Digital Organism) architecture treats omics layers as nodes in a multi-scale biological graph: genome → epigenome → transcriptome → proteome → metabolome → phenome. The power is in finding signals consistent across layers (high-confidence biology) versus signals present in only one layer (artifact or transient perturbation). A CytoRNA 2.0 multi-omics object would be AIDO-export compatible — making CytoHub a contributor to next-generation biological foundation model training.

Where CytoHub Sits Today

What CytoRNA already delivers	What is currently missing
• 14-step scRNA-seq pipeline (QC → UMAP → 22 figures) • 9 cardiac cell types resolved from Brett dataset (~15k cells) • DGE, marker genes, cell-cell communication inference • QC-governed, audit-trailed, SOP-compliant workflow • GEO source registry; reproducibility package per run	• Protein abundance, stoichiometry, and PTM state • Metabolic flux and substrate utilisation rates • Post-transcriptional regulatory layer (miRNA, RBP effects) • mRNA–protein discordance map (the ‘hidden regulation’ layer) • Cross-layer pathway scoring (which pathways are truly active?)

Proposed Extension — Four Modules

The following modules represent one possible architecture for multi-omics integration into CytoRNA. We are bringing these to the brainstorm for critical review, not as fixed decisions.

1

Proteomics ingestion & QC

Accept DIA-MS (.mzML/.raw) processed via DIA-NN v1.8+; TMT/iTRAQ exports from MaxQuant or Proteome Discoverer; SILAC ratios for protein synthesis/degradation studies. QC dashboard: peptide/protein FDR, replicate CV, missing value heatmap, sample correlation matrix — mirroring CytoRNA's existing scRNA-seq QC panels.

2

Metabolomics ingestion, identification & statistical analysis

Accept LC-MS peak tables (MZmine 3, XCMS, MS-DIAL) and NMR outputs (Chenomx). Spectral library matching against HMDB 5.0 and METLIN at MSI confidence levels 1–4. Univariate (Welch t-test, FDR, volcano), multivariate (PLS-DA, OPLS-DA with permutation testing), and Metabolite Set Enrichment Analysis (MSEA/QEA via KEGG). 13C isotopologue tracing with IsoCor correction and MFA flux estimation.

3

Transcriptome–proteome integration

Pearson/Spearman mRNA–protein correlation genome-wide; discordance map highlighting post-transcriptionally regulated genes. MOFA+ joint dimensionality reduction on paired scRNA-seq + proteomics matrices; latent factor decomposition with per-omic variance explained and GSEA-based factor annotation. CITE-seq ADT incorporation where surface protein panels are available.

4

Tri-omics joint pathway analysis & CardioVIVE atlas

PALS pathway activity scoring across all three layers simultaneously; metabolite–gene regulatory inference connecting metabolic reprogramming to upstream TF activity. CardioVIVE 9-cell-type metabolic reference atlas integrating Brett scRNA-seq cluster assignments with published cardiac metabolite and protein profiles. AIDO-compatible JSON-LD export for multi-omics data.

Open Questions for the Brainstorm

We do not have answers to the following. Your experimental and clinical experience is the input we need:

- Which biological questions in your current work require protein- or metabolite-level data that you are currently unable to obtain or interpret efficiently?
- For CardioVIVE validation: what is the single most important protein or metabolic readout that would strengthen our maturation, disease-modelling, or drug-screening claims?
- What are the practical barriers in your lab to generating matched proteomics or metabolomics data — sample quantity, instrument access, turnaround time, cost? How do these affect what is feasible?
- Are there existing datasets (internal or published) where proteomics and/or metabolomics data already exist alongside transcriptomics, that we should be integrating now?
- What does 'good' look like for a multi-omics analysis report? What would make you trust and act on the integrated output?
- Where are the biggest risks of over-interpreting cross-layer correlations? What controls or validation experiments would you require before accepting a multi-omics finding as real?

Indicative 12-Month Delivery Sketch

This is a proposed direction, not a commitment. The brainstorm will refine or redirect it.

Q1 Months 1–3	Q2 Months 4–6	Q3 Months 7–9	Q4 Months 10–12
DIA-MS ingestion · proteomics QC dashboard · limma differential abundance · volcano plots	Protein-set enrichment · STRING PPI network · phosphoproteomics · LC-MS metabolomics ingestion · MSEA	MOFA+ integration · mRNA–protein discordance · PLS-DA / OPLS-DA · 13C flux analysis · IsoCor	Tri-omics PALS · CardioVIVE metabolic atlas preprint · AIDO export API

What we are asking from you in the session

Challenge the framing. Add the biological questions we have not thought of. Tell us where the gaps are in your own work that multi-omics could address — and where you think it is unlikely to help. The roadmap should be shaped by scientific need, not by what is technically easiest to build.