A Scalable AI Framework for Multi-Study Single-Cell Analysis of GLP-1 Biology

Alissa Cait[1], Jessica Corrado[1], Dan Ryder[12]

[1]Bridge Informatics, Inc., 160 Alewife Brook Pkwy, Cambridge, MA 02138 [2]Corresponding author and Chief Executive Officer

THE PROBLEM

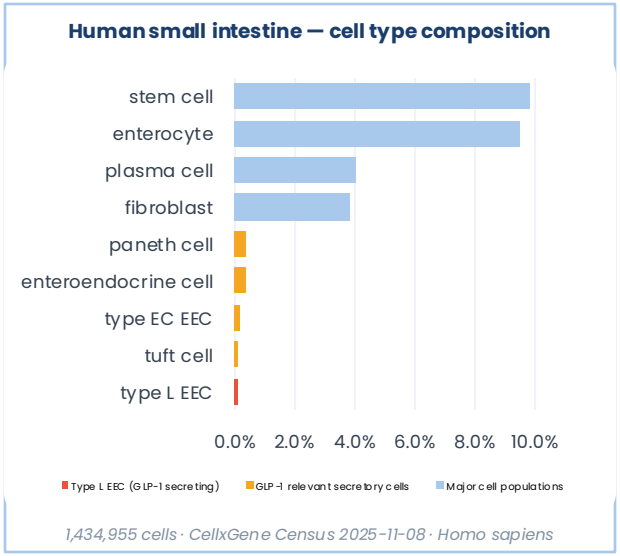
Why standard workflows fall short

GLP-1R is expressed across pancreatic beta cells, intestinal L-cells, hypothalamic neurons, vagal afferents, and cardiomyocytes. The most pharmacologically relevant cell states are often rare, like specific enteroendocrine cell (EEC) subtypes and dedifferentiated beta cells.

The problem is: standard CPU-based workflows require downsampling to remain tractable. Downsampling is precisely where rare populations disappear. A single study rarely provides enough cells to robustly characterize populations at 0.08% abundance.

The solution is to combine multiple large human gut datasets, directly addressing this challenge. Aggregating studies gives you a statistically robust population of rare cells of interest. But naive merging introduces batch effects across donors, labs, and sequencing technologies. Solving both problems simultaneously, scale and harmonization, requires a purpose-built computational framework.

Figure 1: cell type composition of the human small intestine, cells of interest are a small fraction of total cells.



THE SOLUTION

DATA INPUTS — HUMAN GUT scRNA-seq STUDIES

Study A	Study B	Study C
Human gut	Human gut	Human gut
scRNA-seq multi-donor	scRNA-seq different lab	scRNA-seq different tech

RAPIDS single-cell — No downsampling
GPU preprocessing · normalization · PCA · neighbor graph · no downsampling

scVI — Joint latent space
Variational integration · cross-donor, cross-lab batch correction

scGPT — Perturbation-ready
Foundation model embedding · cell state representation · perturbation-ready

Rare cell detection	Perturbation prediction	Cross-study integration

Let's Connect!

www.bridgeinformatics.com
/company/bridge-informatics
@Bridge_Info_USA
dan.ryder@bridgeinformatics.com

WHAT THIS FRAMEWORK ENABLES

By combining GPU-native preprocessing, variational integration, and foundation model embeddings, this framework makes previously intractable analyses less challenging and more feasible.

01 RAPIDS single-cell

Full-scale analysis with no compromise

Standard pipelines subsample to thousands of cells to stay computationally tractable. RAPIDS single-cell runs GPU-native preprocessing on millions of cells without downsampling, preserving the rare EEC populations (0.08% of gut cells) that define the most interesting GLP-1 biology. Analysis that previously took hours completes in minutes.

50×+ faster than CPU-based scanpy

02 scVI

Cross-study integration, harmonized biology

Human gut datasets from different labs, donors, and sequencing technologies carry strong batch effects that corrupt naïve merges. scVI learns a joint latent representation that separates technical variation from biological signal, enabling comparison across studies without distorting the underlying cell type structure. The result: 9 EEC subtypes resolved from a single integrated embedding.

9 EEC resolved subtypes

03 scGPT

Perturbation modeling of in silico drug response

scGPT is pretrained on tens of millions of single cells. Fine-tuned on GLP-1R-expressing populations, it predicts how gene programs shift in response to semaglutide or tirzepatide at single-cell resolution, stratified by EEC subtype. This enables hypothesis generation about off-target effects and subtype-specific drug responses without exhaustive in vivo experiments.

<1% of gut cells are type L EECs

[1] Dicks S, et al. GPU-accelerated single-cell analysis at scale with rapids-singlecell. arXiv 2026. doi:10.48550/arXiv.2603.02402 [2] Lopez R, et al. Deep generative modeling for single-cell transcriptomics. Nature Methods 2018;15:1053–1058. [3] Cui H, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods 2024;21:1470–1480. [4] CellxGene Census 2025-11-08. Chan Zuckerberg Initiative.