Supplementary Material for Conditional Similarity Triplets Enable Covariate-Informed Representations of Single-Cell Data

1. ROC CURVES IN CYTOF DATASETS

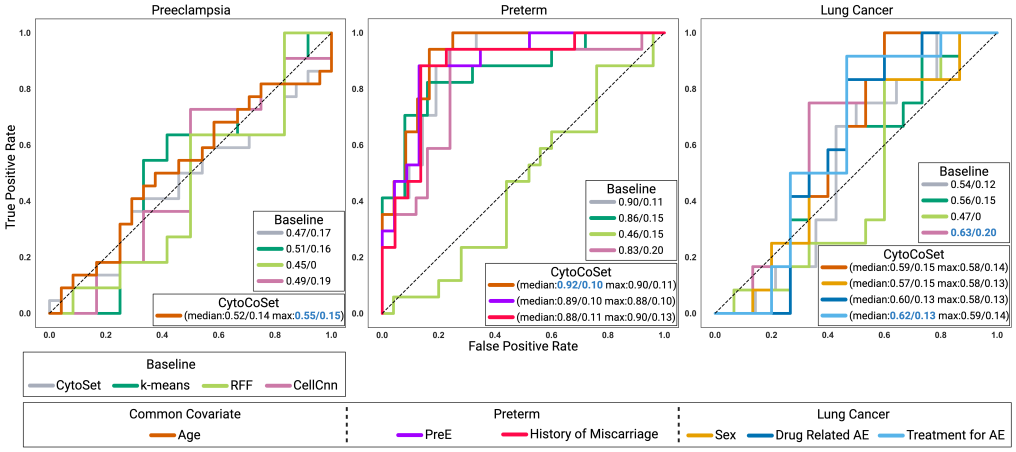

Fig. S1. Classification Accuracy in CyTOF Datasets. CytoCoSet and baseline methods were evaluated for their capacity to produce per-sample encodings that could accurately predict a binary clinical outcome in three CyTOF (Preeclampsia, Preterm, Lung Cancer) datasets. All classification experiments were done using 30 different train/test splits, and ROC curves corresponding to the particular trial that achieved the mean AUC. The legend for each ROC curve shows the (mean/standard deviation) for various combinations of method and covariate. AUC results for median and max pooling strategies employed to compute Random Fourier Features for selecting triplets in CytoCoSet are also shown.

2. SCALABILITY OF CYTOCOSET TRAINING PROCESS

We empirically evaluated the scalability of CytoCoSet by profiling the model training time as a function of the number of downsampled cells per profiled sample. The number of downsampled (e.g. randomly selected cells) per sample was varied between 200 and 1,000. This is in contrast to the default setting of sampling 1,024 cells per sample. Fig. S2 shows the accuracies (a-b) and run-time (c). Broadly, the training process and accuracies are not strongly influenced by the number of downsampled cells. Notably, as the preeclampsia dataset is large, we see that larger downsamples do not necessarily increase accuracy, but do modestly increase the training time.

3. PERFORMANCE METRICS ACROSS METHODS AND DATASETS

Table S2 presents performance comparisons across various baseline models, including CytoSet, k-means + frequency feature engineering, CellCnn, and RFF, for predicting binary clinical outcomes across three datasets: Preeclampsia, Preterm, and Lung Cancer. The metrics used are Accuracy S1.1, Precision S1.2, Recall S1.3, and F1 Score S1.4. We evaluated CytoCoSet under different choices of pooling operation in the RFF step. In the Preeclampsia dataset, k-means + feature engineering achieved the highest accuracy (0.608) and precision (0.6), while CytoCoSet-Age

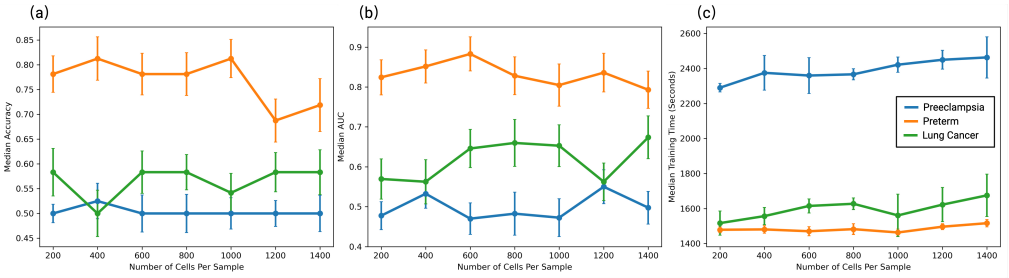

Fig. S2. Scalability of CytoCoSet as a function of cells downsampled per sample. The line plot illustrates the (a) median accuracy, (b) median AUC, and (c) median training time as a function of the number of downsampled cells per-sample in the three CyTOF datasets (Preeclampsia, Preterm, and Lung Cancer). Each classification experiment was conducted with 10 randomized train/test splits. Error bars represent the 95% confidence intervals around the median values.

Accuracy	$\frac{TP+TN}{TP+FN+TN+FP}$	(S1.1)
Precision	$\frac{TP}{TP+FP}$	(S1.2)
Recall	$\frac{TP}{TP+FN}$	(S1.3)
F1 Score	$\frac{2\times Precision\times Recall}{Precision+Recall}$	(S1.4)

Table S1. Accuracy Measures Metrics Equations. The equations rely on the assumption that a test sample falls into one of four categories: False Positive (FP), where the system labels a negative as positive; False Negative (FN), where it labels a positive as negative; True Positive (TP), and True Negative (TN), where the system correctly predicts the label. Here, we use TP, TN, FP, and FN to represent the counts of true positives, true negatives, false positives, and false negatives, respectively.

achieved the best recall (0.818). However, CellCnn had the highest F1 Score (0.615), indicating that it balances precision and recall effectively in this dataset. In the Preterm dataset, CytoCoSet (notation CytoCoSet-X used to denote the accuracy obtained using CytoCoSet with covariate X) outperformed other models with the highest values across all metrics: accuracy (0.853 in CytoCoSet-Age through median pooling and CytoCoSet-Miscarriage History through max pooling), precision (0.812 in CytoCoSet-PreE through max pooling)), recall (0.941 in CytoCoSet-Age), and F1 Score (0.842 in CytoCoSet-Age), making it the most effective model for this dataset. Last in the Lung Cancer dataset, CytoCoSet-TreatmentforAE(both median and max) and CytoCoSet-Sex (max) achieved the highest accuracy (0.666), recall (0.833), and F1 Score (0.689), indicating superior performance for this dataset. However, CellCnn had a slightly higher precision (0.666). Overall, CytoCoSet (particularly CytoCoSet-Age and CytoCoSet-TreatmentforAE) demonstrating superior performance in accurately predicting clinical outcomes compared to baseline models across different datasets. In the Preterm and Lung Cancer datasets, CytoCoSet models achieved the highest scores, highlighting CytoCoSet as the most effective method for clinical outcome prediction in these dataset.

Table S2. Accuracy Metrics Across Datasets. Accuracy, precision, recall, and F1 score for Preeclampsia, Preterm, and Lung Cancer CyTOF datasets. Notation for CytoCoSet variates as CytoCoSet-X (Y) indicates CytoCoSet using covariate X and pooling method Y.

Preeclampsia				
	Accuracy	precision	Recall	F1 Score
CytoSet	0.456	0.448	0.59	0.509
k-means	**0.608**	**0.6**	0.545	0.571
CellCnn	0.565	0.533	0.727	**0.615**
RFF	0.52	0.5	0.091	0.153
CytoCoSet-Age (Median)	0.456	0.461	**0.818**	0.59
CytoCoSet-Age (Max)	0.413	0.439	**0.818**	0.571
Preterm				
	Accuracy	precision	Recall	F1 Score
CytoSet	0.815	0.75	0.882	0.81
k-means	0.809	0.736	0.823	0.777
CellCnn	0.738	0.666	0.705	0.685
RFF	0.5	0.416	0.588	0.487
CytoCoSet-Age (Median)	**0.853**	0.761	**0.941**	**0.842**
CytoCoSet-Age (Max)	0.846	0.761	**0.941**	**0.842**
CytoCoSet-PreE (Median)	0.8	.8	0.705	0.75
CytoCoSet-PreE (Max)	0.825	**0.812**	0.764	0.787
CytoCoSet-Miscarriage History (Median)	0.794	.8	0.705	0.75
CytoCoSet-Miscarriage History (Max)	**0.853**	0.789	0.88	0.833
Lung Cancer				
	Accuracy	precision	Recall	F1 Score
CytoSet	0.615	0.571	0.666	0.615
k-means	0.555	0.5	0.5	0.5
CellCnn	**0.666**	**0.6**	0.75	0.666
RFF	0.444	0.444	1.0	0.615
CytoCoSet-Age (Median)	0.555	0.5	0.666	0.571
CytoCoSet-Age (Max)	0.555	0.5	0.666	0.571
CytoCoSet-Sex (Median)	0.629	0.562	0.75	0.642
CytoCoSet-Sex (Max)	**0.666**	0.588	**0.833**	**0.689**
CytoCoSet-DrugRelatedAE (Median)	0.592	0.538	0.583	0.559
CytoCoSet-DrugRelatedAE (Max)	0.629	0.562	0.75	0.642
CytoCoSet-TreatmentforAE (Median)	**0.666**	0.588	**0.833**	**0.689**
CytoCoSet-TreatmentforAE (Max)	**0.666**	0.588	**0.833**	**0.689**