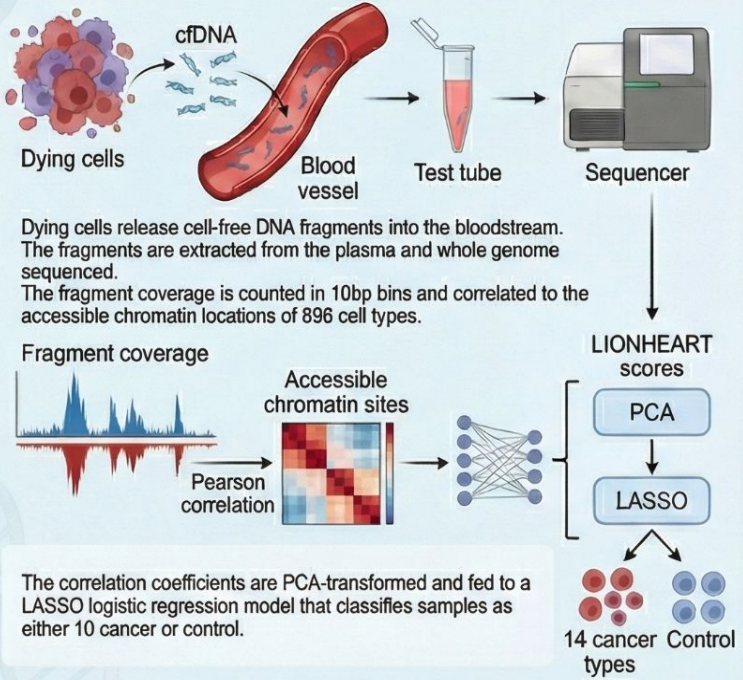
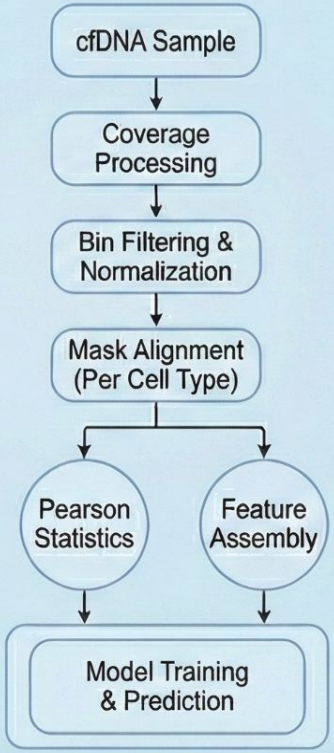
Reproducing LIONHEART: A Novel cfDNA-Based Cancer Detection Method

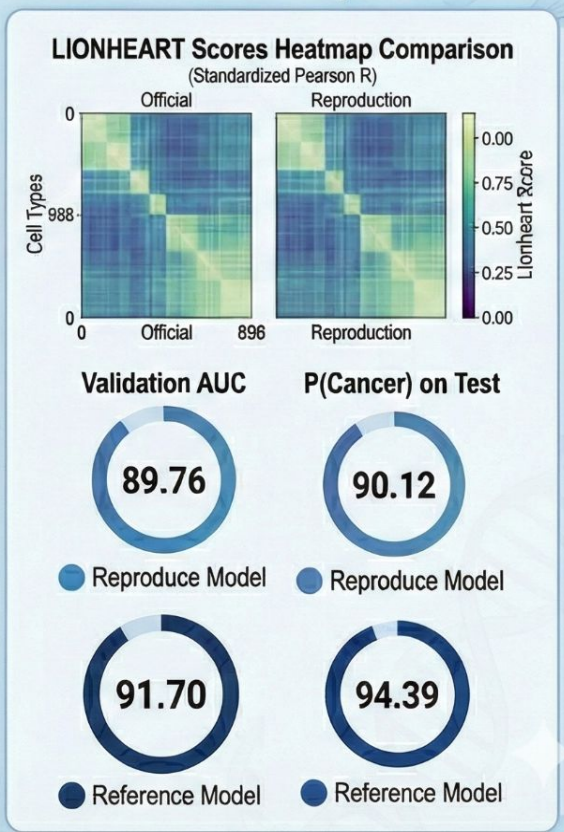
Method



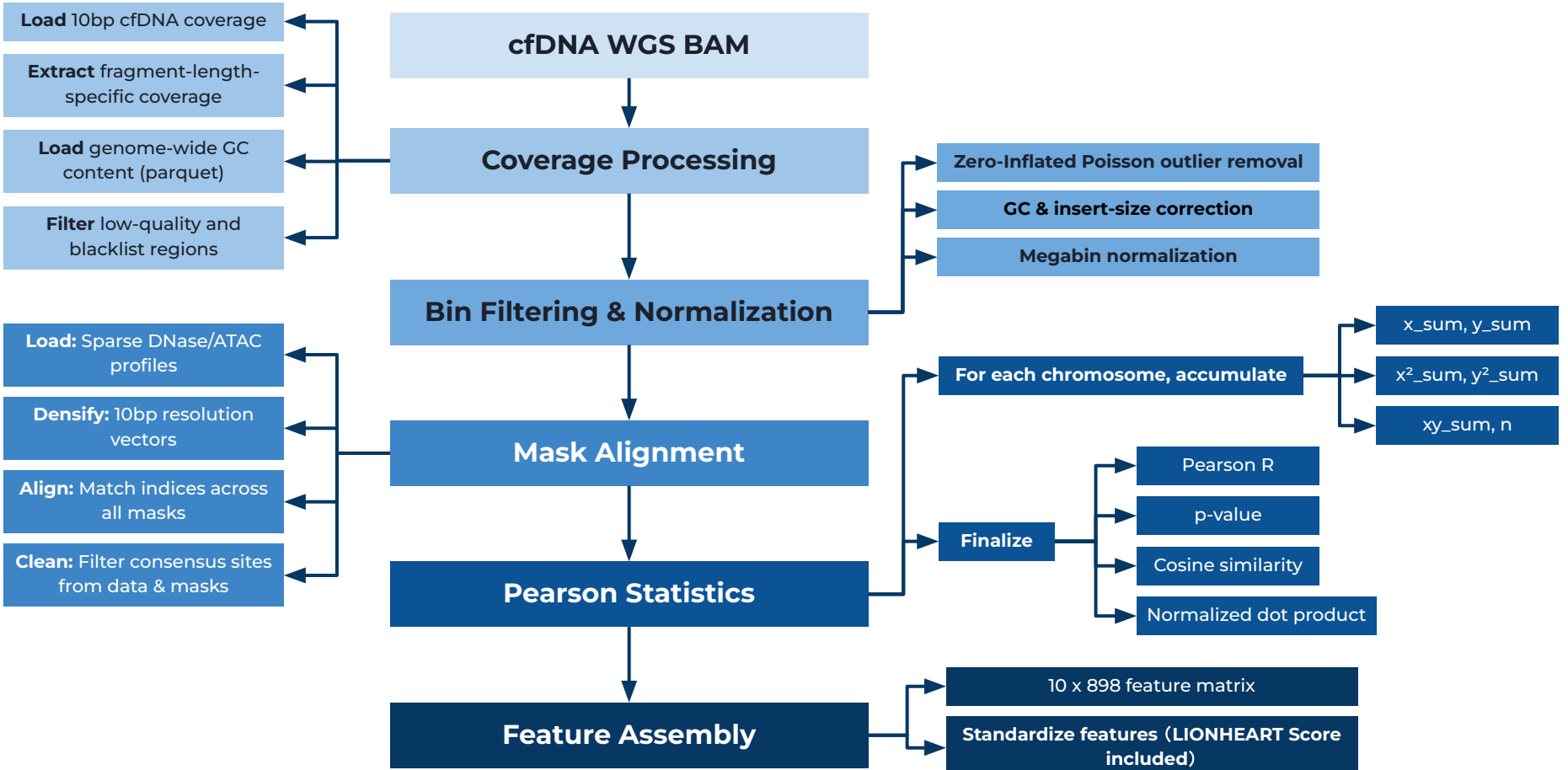
Reproduction Pipeline

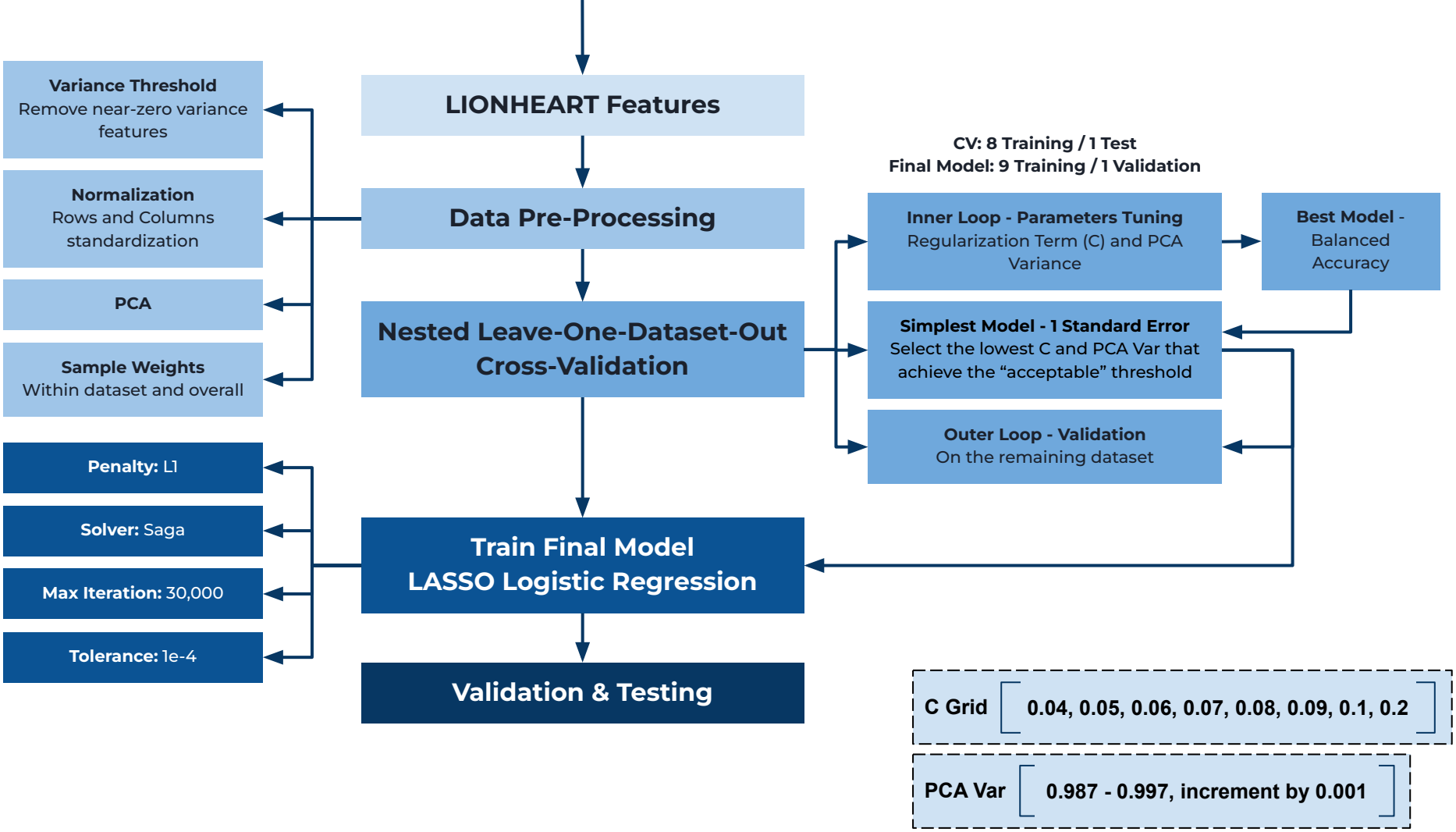


Results Comparison

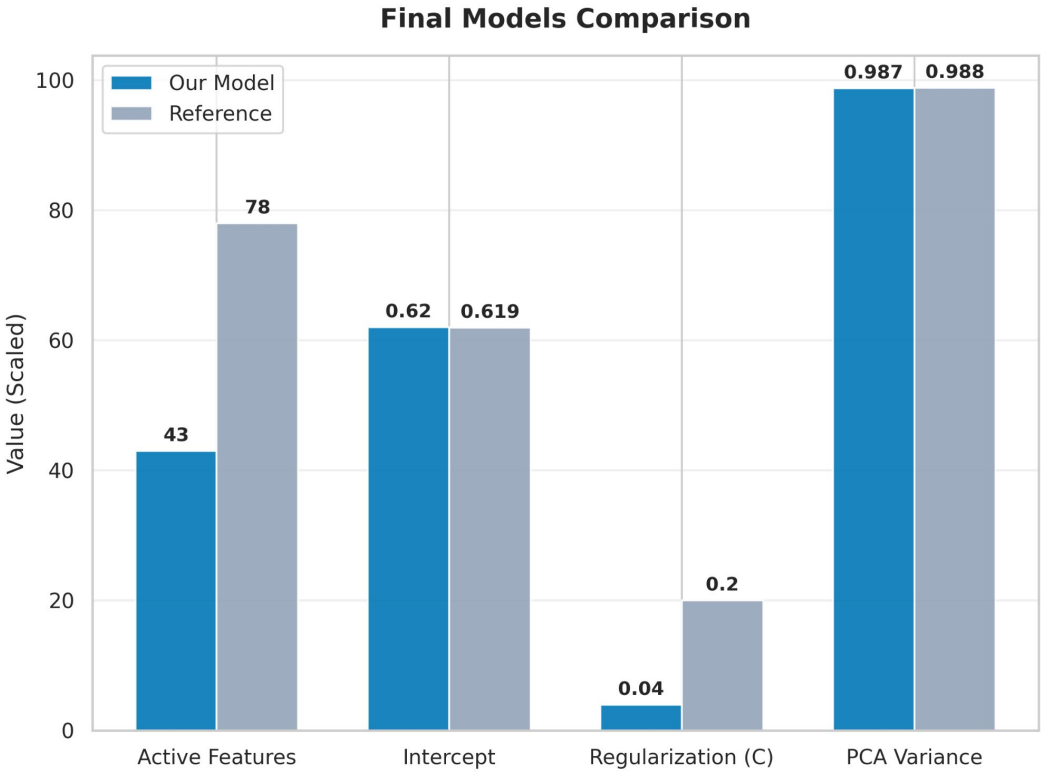


REPRODUCTION PIPELINE




LIONHEART Features
Data Pre-Processing
Variance Threshold
Remove near-zero variance features
Normalization
Rows and Columns standardization
PCA
Sample Weights
Within dataset and overall
Nested Leave-One-Dataset-Out Cross-Validation
CV: 8 Training / 1 Test
Final Model: 9 Training / 1 Validation
Inner Loop - Parameters Tuning
Regularization Term (C) and PCA Variance
Best Model - Balanced Accuracy
Simplest Model - 1 Standard Error
Select the lowest C and PCA Var that achieve the "acceptable" threshold
Outer Loop - Validation
On the remaining dataset
Penalty: L1
Solver: Saga
Max Iteration: 30,000
Tolerance: 1e-4
Train Final Model
LASSO Logistic Regression
Validation & Testing
C Grid
0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2
PCA Var
0.987 - 0.997, increment by 0.001

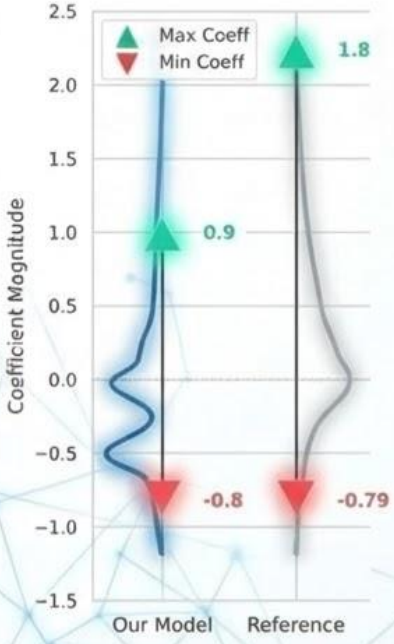
MODELING RESULTS

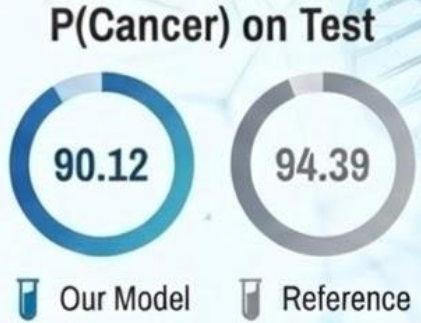


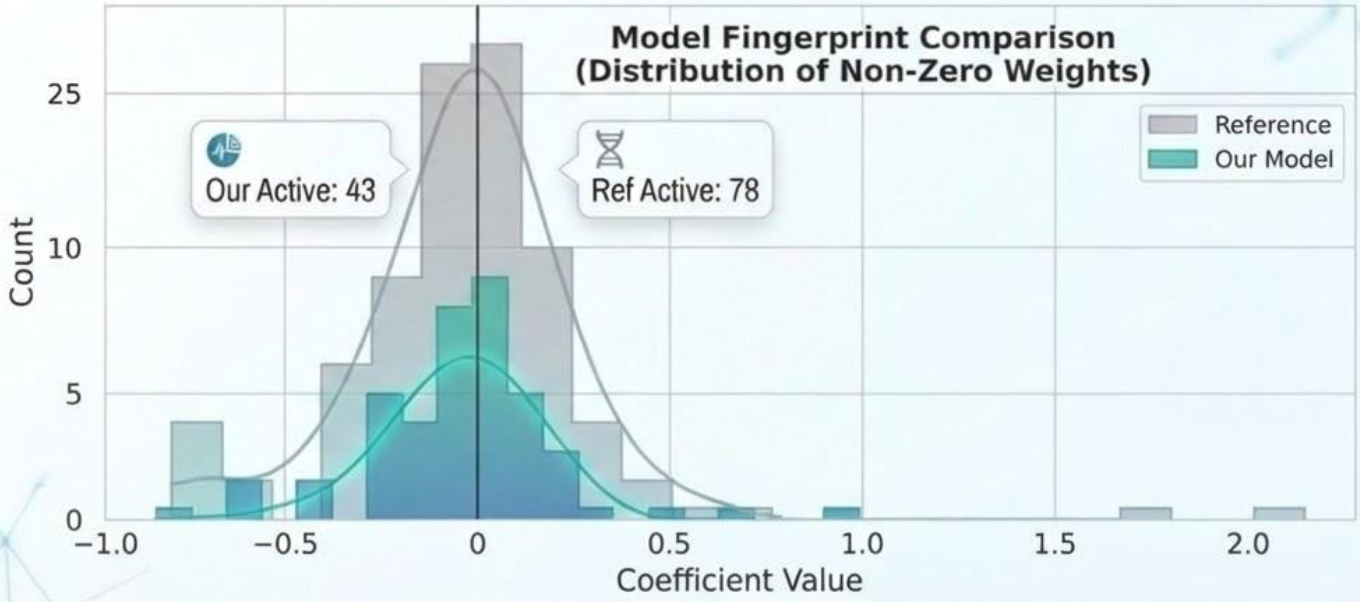
- **Stronger penalty** → Much more conservative model
- **45% fewer features** to achieve a comparable AUC
- Identified the **“core”** biological signals

MODELING RESULTS

Final Model Coefficients

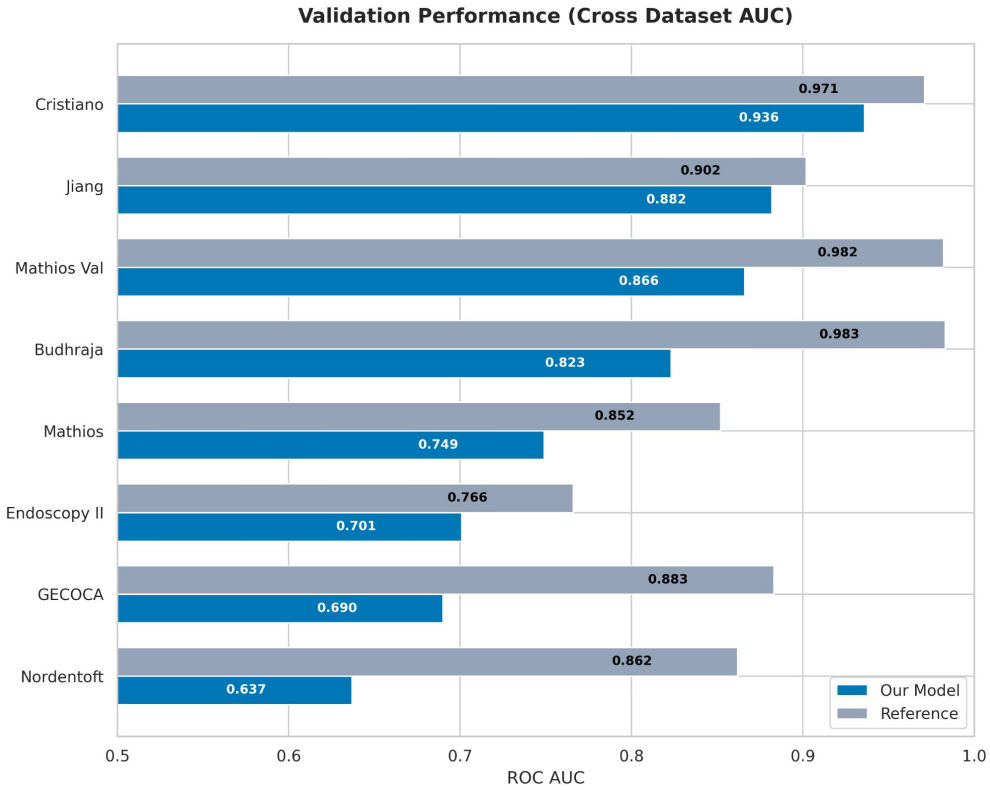






MODELING RESULTS

Cross Validation Performance



Reference Stats

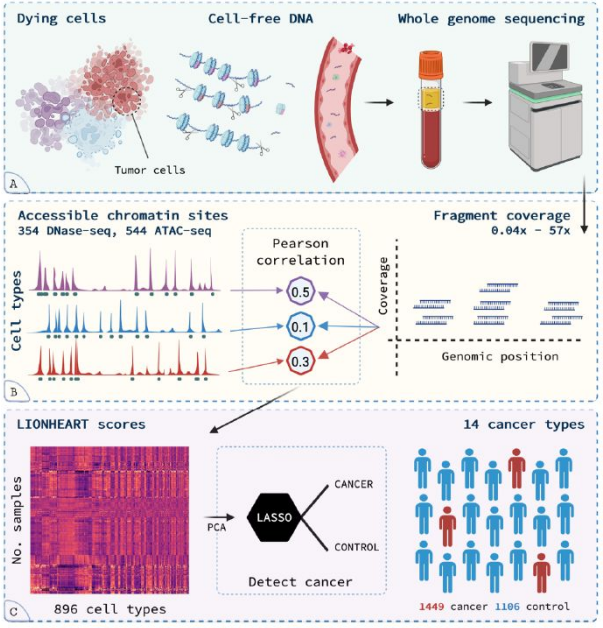
0.90	0.983	0.766
Weighted Mean	Max (Budhraja)	Min (Endoscopy)

Reproduced Stats

0.802	0.936	0.637
Weighted Mean	Max (Cristiano)	Min (Nordentoft)

Reproducing LIONHEART: A Novel cfDNA-Based Cancer Detection Method

Method

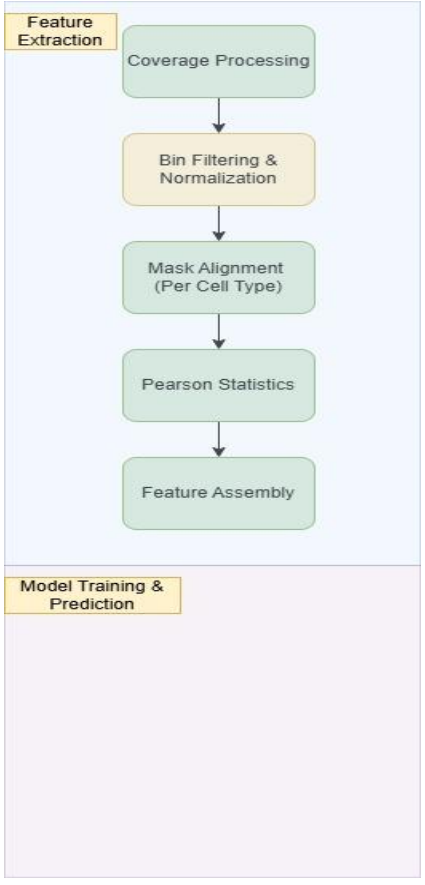
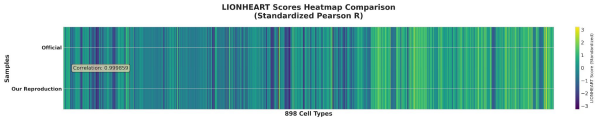
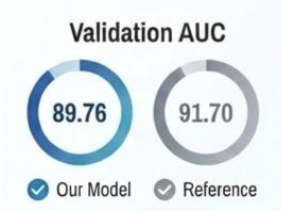
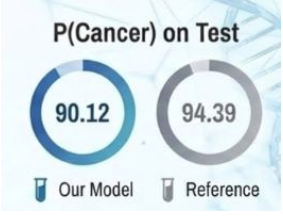


Dying cells release cell-free DNA fragments into the bloodstream. The fragments are extracted from the plasma and whole genome sequenced.

The fragment coverage is counted in 10bp bins and correlated to the accessible chromatin locations of 898 cell types.

The correlation coefficients are PCA-transformed and fed to a LASSO logistic regression model that classifies samples as either 10 cancer or control.

Reproduction Pipeline



Results Comparison