Supplementary Table S1.

Optimization results of meth-SemiCancer and the machine-learning-based classifiers with the different combination of the parameters. Grid search was adopted for the model tuning, and the hyperparameters showing the best accuracy were selected.

meth-SemiCancer

# of hidden nodes	Pre-training		Fine-tuning		Dropout	Alpha	Accuracy
	Learning rate	Training epoch	Learning rate	Training epoch			
1000-500	1e-5	1000	1e-3	2000	0.7	0.001	0.8018
1000-500	1e-5	1000	1e-3	2000	0.7	0.005	0.8113
1000-500	1e-5	1000	1e-3	2000	0.7	0.01	0.8113
1000-500	1e-5	1000	1e-3	2000	0.7	0.05	0.8018
1000-500	1e-5	1000	1e-3	2000	0.7	0.1	0.8018
1000-500	1e-5	1000	1e-3	2000	0.7	0.5	0.8018
1000-500	1e-3	1000	1e-3	2000	0.7	0.001	0.8113
1000-500	1e-3	1000	1e-3	2000	0.7	0.005	0.8113
1000-500	1e-3	1000	1e-3	2000	0.7	0.01	0.8207
1000-500	1e-3	1000	1e-3	2000	0.7	0.05	0.8207
1000-500	1e-3	1000	1e-3	2000	0.7	0.1	0.8113
1000-500	1e-3	1000	1e-3	2000	0.7	0.5	0.8113
1000-500	1e-5	1500	1e-3	3000	0.7	0.001	0.8301
1000-500	1e-5	1500	1e-3	3000	0.7	0.005	0.8301
1000-500	1e-5	1500	1e-3	3000	0.7	0.01	0.8301
1000-500	**1e-5**	**1500**	**1e-3**	**3000**	0.7	**0.05**	**0.8490**
1000-500	1e-5	1500	1e-3	3000	0.7	0.1	0.8301
1000-500	1e-5	1500	1e-3	3000	0.7	0.5	0.8207
1000-500	1e-3	1500	1e-3	3000	0.7	0.001	0.8207
1000-500	1e-3	1500	1e-3	3000	0.7	0.005	0.8207
1000-500	1e-3	1500	1e-3	3000	0.7	0.01	0.8301
1000-500	1e-3	1500	1e-3	3000	0.7	0.05	0.8301
1000-500	1e-3	1500	1e-3	3000	0.7	0.1	0.8301
1000-500	1e-3	1500	1e-3	3000	0.7	0.5	0.8207

Support vector machine (SVM)

Kernel	Penalty parameter (C)	RBF kernel coeff (Gamma)	Accruacy
Linear	2^{-5}	-	0.8066
Linear	$\mathbf{2^{-3}}$	**-**	**0.8113**
Linear	2^{-1}	-	0.7925
Linear	2^{1}	-	0.7925
Linear	2^{3}	-	0.7925
Linear	2^{5}	-	0.7925
RBF	2^{-5}	2^{-15}	0.5755
RBF	2^{-5}	2^{-13}	0.5755
RBF	2^{-5}	2^{-11}	0.5755
RBF	2^{-5}	2^{-9}	0.5755
RBF	2^{-5}	2^{-7}	0.5755
RBF	2^{-5}	2^{-5}	0.5755
RBF	2^{-5}	2^{-3}	0.5755
RBF	2^{-5}	2^{-1}	0.5755
RBF	2^{-5}	2^{1}	0.5755
RBF	2^{-5}	2^{3}	0.5755
RBF	2^{-3}	2^{-15}	0.5755
RBF	2^{-3}	2^{-13}	0.5755
RBF	2^{-3}	2^{-11}	0.5755
RBF	2^{-3}	2^{-9}	0.7547
RBF	2^{-3}	2^{-7}	0.7547
RBF	2^{-3}	2^{-5}	0.5991
RBF	2^{-3}	2^{-3}	0.5755
RBF	2^{-3}	2^{-1}	0.5755
RBF	2^{-3}	2^{1}	0.5755
RBF	2^{-3}	2^{3}	0.5755
RBF	2^{-1}	2^{-15}	0.5755
RBF	2^{-1}	2^{-13}	0.5755
RBF	2^{-1}	2^{-11}	0.7547
RBF	2^{-1}	2^{-9}	0.7500
RBF	2^{-1}	2^{-7}	0.7830
RBF	2^{-1}	2^{-5}	0.7170
RBF	2^{-1}	2^{-3}	0.5755
RBF	2^{-1}	2^{-1}	0.5755
RBF	2^{-1}	2^{1}	0.5755
RBF	2^{-1}	2^{3}	0.5755
RBF	2^{1}	2^{-15}	0.5755

RBF	2^{1}	2^{-13}	0.7547
RBF	2^{1}	2^{-11}	0.7594
RBF	2^{1}	2^{-9}	0.8019
RBF	2^{1}	2^{-7}	0.7972
RBF	2^{1}	2^{-5}	0.7311
RBF	2^{1}	2^{-3}	0.5708
RBF	2^{1}	2^{-1}	0.5755
RBF	2^{1}	2^{1}	0.5755
RBF	2^{1}	2^{3}	0.5755
RBF	2^{3}	2^{-15}	0.7547
RBF	2^{3}	2^{-13}	0.7642
RBF	2^{3}	2^{-11}	0.7972
RBF	2^{3}	2^{-9}	0.8019
RBF	2^{3}	2^{-7}	0.8066
RBF	2^{3}	2^{-5}	0.7264
RBF	2^{3}	2^{-3}	0.5708
RBF	2^{3}	2^{-1}	0.5755
RBF	2^{3}	2^{1}	0.5755
RBF	2^{3}	2^{3}	0.5755
RBF	2^{5}	2^{-15}	0.7642
RBF	2^{5}	2^{-13}	0.7972
RBF	2^{5}	2^{-11}	0.7972
RBF	2^{5}	2^{-9}	0.8066
RBF	2^{5}	2^{-7}	0.8066
RBF	2^{5}	2^{-5}	0.7264
RBF	2^{5}	2^{-3}	0.5708
RBF	2^{5}	2^{-1}	0.5755
RBF	2^{5}	2^{1}	0.5755
RBF	2^{5}	2^{3}	0.5755

Random Forest (RF)

Split criteria (criterion)	# of trees (estimators)	The minimum # of samples in a leaf node (min_samples_leaf)	Accruacy
Gini impurity	100	1	0.7689
Gini impurity	100	2	0.7736
Gini impurity	**100**	**3**	**0.8019**
Gini impurity	100	4	0.7547
Gini impurity	100	5	0.7830
Gini impurity	300	1	0.7500
Gini impurity	300	2	0.7783
Gini impurity	300	3	0.7736
Gini impurity	300	4	0.7736
Gini impurity	300	5	0.7594
Gini impurity	500	1	0.7783
Gini impurity	500	2	0.7736
Gini impurity	500	3	0.7547
Gini impurity	500	4	0.7547
Gini impurity	500	5	0.7736
Gini impurity	700	1	0.7642
Gini impurity	700	2	0.7547
Gini impurity	700	3	0.7736
Gini impurity	700	4	0.7689
Gini impurity	700	5	0.7736
Gini impurity	900	1	0.7736
Gini impurity	900	2	0.7689
Gini impurity	900	3	0.7689
Gini impurity	900	4	0.7783
Gini impurity	900	5	0.7689
Information gain	100	1	0.7830
Information gain	100	2	0.7689
Information gain	100	3	0.7830
Information gain	100	4	0.7783
Information gain	100	5	0.7689
Information gain	300	1	0.7783
Information gain	300	2	0.7736
Information gain	300	3	0.7642
Information gain	300	4	0.7736
Information gain	300	5	0.7830

Information gain	500	1	0.7830
Information gain	500	2	0.7689
Information gain	500	3	0.7972
Information gain	500	4	0.8019
Information gain	500	5	0.7736
Information gain	700	1	0.7877
Information gain	700	2	0.7783
Information gain	700	3	0.7783
Information gain	700	4	0.7736
Information gain	700	5	0.7689
Information gain	900	1	0.7783
Information gain	900	2	0.7736
Information gain	900	3	0.7830
Information gain	900	4	0.7783
Information gain	900	5	0.7736

Logistic Regression (LR)

max_iter (Maximum number of iterations to converge)	Penalty parameter (C)	Accuracy
100	2^{-5}	0.7925
100	2^{-3}	0.8113
100	$\mathbf{2^{-1}}$	**0.8302**
100	2^{1}	0.8255
100	2^{3}	0.8302
100	2^{5}	0.8255
200	2^{-5}	0.7925
200	2^{-3}	0.8113
200	2^{-1}	0.8302
200	2^{1}	0.8255
200	2^{3}	0.8302
200	2^{5}	0.8255
300	2^{-5}	0.7925
300	2^{-3}	0.8113
300	2^{-1}	0.8302
300	2^{1}	0.8255
300	2^{3}	0.8302
300	2^{5}	0.8255
400	2^{-5}	0.7925
400	2^{-3}	0.8113
400	2^{-1}	0.8302
400	2^{1}	0.8255
400	2^{3}	0.8302
400	2^{5}	0.8255
500	2^{-5}	0.7925
500	2^{-3}	0.8113
500	2^{-1}	0.8302
500	2^{1}	0.8255
500	2^{3}	0.8302
500	2^{5}	0.8255

Supplementary Table S2.

Average breast cancer subtype classification performance results of BCtypeFinder with other methods for 87 samples of GSE72245 dataset.

Metric	BCtypeFinder	meth-SemiCancer	SVM	RF	LR
Accuracy	**0.816**	0.736	0.655	0.545	0.667
F1-score	**0.822**	0.748	0.668	0.505	0.673
MCC	**0.752**	0.664	0.603	0.522	0.623
AUC	**0.883**	0.848	0.811	0.747	0.825

Supplementary Table S3.

Average classification performance results of BCtypeFinder with other comparison methods based on the 10-fold cross validation on TCGA-BRCA dataset.

Metric	BCtypeFinder	meth-SemiCancer	SVM	RF	LR
Accuracy	**0.849**	0.814	0.811	0.764	0.821
F1-score	**0.843**	0.803	0.807	0.734	0.814
MCC	**0.763**	0.707	0.703	0.621	0.716
AUC	**0.845**	0.819	0.83	0.745	0.825

Supplementary Table S4.
Performance improvement results of BCtypeFinder testing GSE72245 dataset by measuring the prediction performance on each training phase.

Metric	Pre-training	SSL	SSL + Subtype alignment
Accuracy	0.720	0.730	**0.816**
F1-score	0.723	0.731	**0.822**
MCC	0.673	0.678	**0.752**
AUC	0.856	0.857	**0.883**