Supplementary Data for

Ying-Ying Xu, Fan Yang, Yang Zhang, and Hong-Bin Shen, An image-based multi-label human protein subcellular localization predictor (*i*Locator) reveals protein mislocalizations in cancer tissues.

Supplementary text

Evaluating metrics. In this study, 5 multi-label metrics were used to evaluate the performance of the classifier model. Suppose there are L classes. Let $\hat{Y}_{t_j} = [\hat{y}_1^j, \hat{y}_2^j, \cdots, \hat{y}_L^j]$ denotes the predicted label vector of the jth test sample t_j, while $Y_{t_j} = [y_1^j, y_2^j, \cdots, y_L^j]$ is the corresponding real vector. The 5 metrics are defined below:

1) Subset accuracy

$$Subset\_accuracy = \frac{1}{q}\sum_{j=1}^{q} \Phi\left[\!\left[\hat{Y}_{t_j} = Y_{t_j} \right]\!\right] \tag{S1}$$

where $\Phi[\![\bullet]\!] = \begin{cases} 1, & \bullet \text{ is true} \\ 0, & otherwise \end{cases}$.

Subset accuracy is the fraction of samples whose predicted label set is the same as the true label set. This metric is severe and ignores the much difficulty against single-label learning. Yet it is direct-viewing, and can reflect the performance of the classification.

2) Accuracy

$$Accuracy = \frac{1}{q}\sum_{j=1}^{q} point(\hat{Y}_{t_j}) \tag{S2}$$

Each test sample prediction can be scored by:

$$point(\hat{Y}_{t_j}) = \frac{\sum_{l=1}^{L} \Phi\left[\!\left[y_l^j = 1, \hat{y}_l^j = 1 \right]\!\right]}{\sum_{l=1}^{L} \Phi\left[\!\left[y_l^j = 1, or\ \hat{y}_l^j = 1 \right]\!\right]} \tag{S3}$$

Accuracy is more lenient to errors than subset accuracy because if not all the predicted labels of a sample are correct, then subset accuracy gives 0, but accuracy gives a value between 0 and 1, reflecting the degree of partial correctness.

3) Recall

For a class l,

$$Recall(l) = \frac{1}{\sum_{j=1}^{q} \Phi\left[\!\left[y_l^j = 1 \right]\!\right]} \sum_{t_j \in \{t_j | y_l^j = 1\}} point(\hat{Y}_{t_j}) \tag{S4}$$

Then the uniform recall of the total testing samples is computed as:

$$Recall = \frac{1}{L}\sum_{l=1}^{L} Recall(l) \tag{S5}$$

4) Precision

We can obtain precision in a similar way:

$$Precision(l) = \frac{1}{\sum_{j=1}^{q} \Phi\left[\!\left[\hat{y}_l^j = 1 \right]\!\right]} \sum_{t_j \in \{t_j | \hat{y}_l^j = 1\}} point(\hat{Y}_{t_j}) \tag{S6}$$

$$Precision = \frac{1}{L}\sum_{l=1}^{L} Precision(l) \tag{S7}$$

The above two metrics are extensions of the classic definitions to measure recall and precision of each class in traditional single-label learning. Recall is the fraction of true labels that are correctly predicted, while precision is the fraction of predicted labels that are correctly predicted.

5) Label accuracy

For a class l,

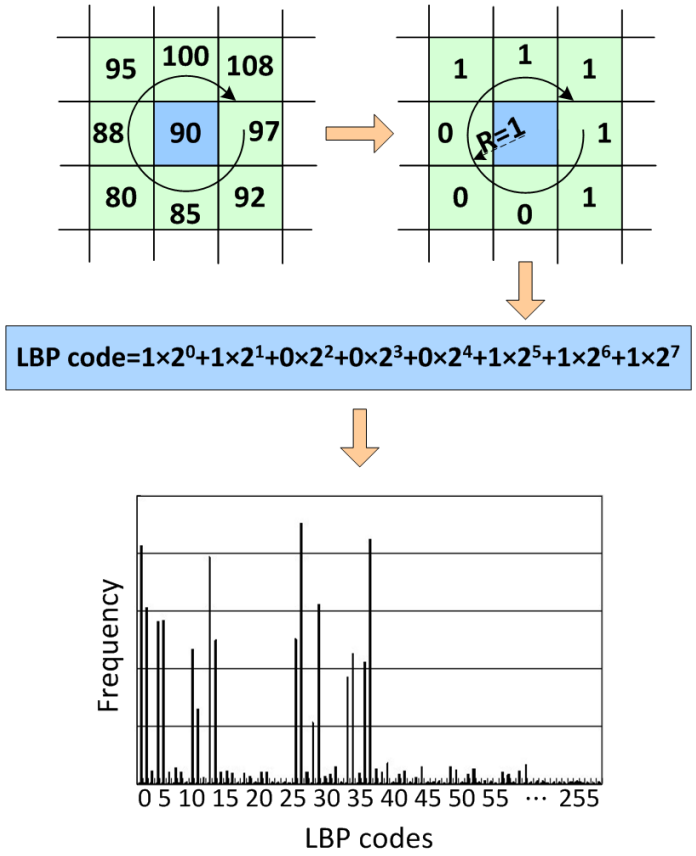
$$Label\_accuracy(l) = \frac{1}{q}\sum_{j=1}^{q}\left[\!\left[\hat{y}_l^j = y_l^j \right]\!\right] \tag{S8}$$

$$Average\_label\_accuracy = \frac{1}{L}\sum_{l=1}^{L} Label\_accuracy(l) \tag{S9}$$

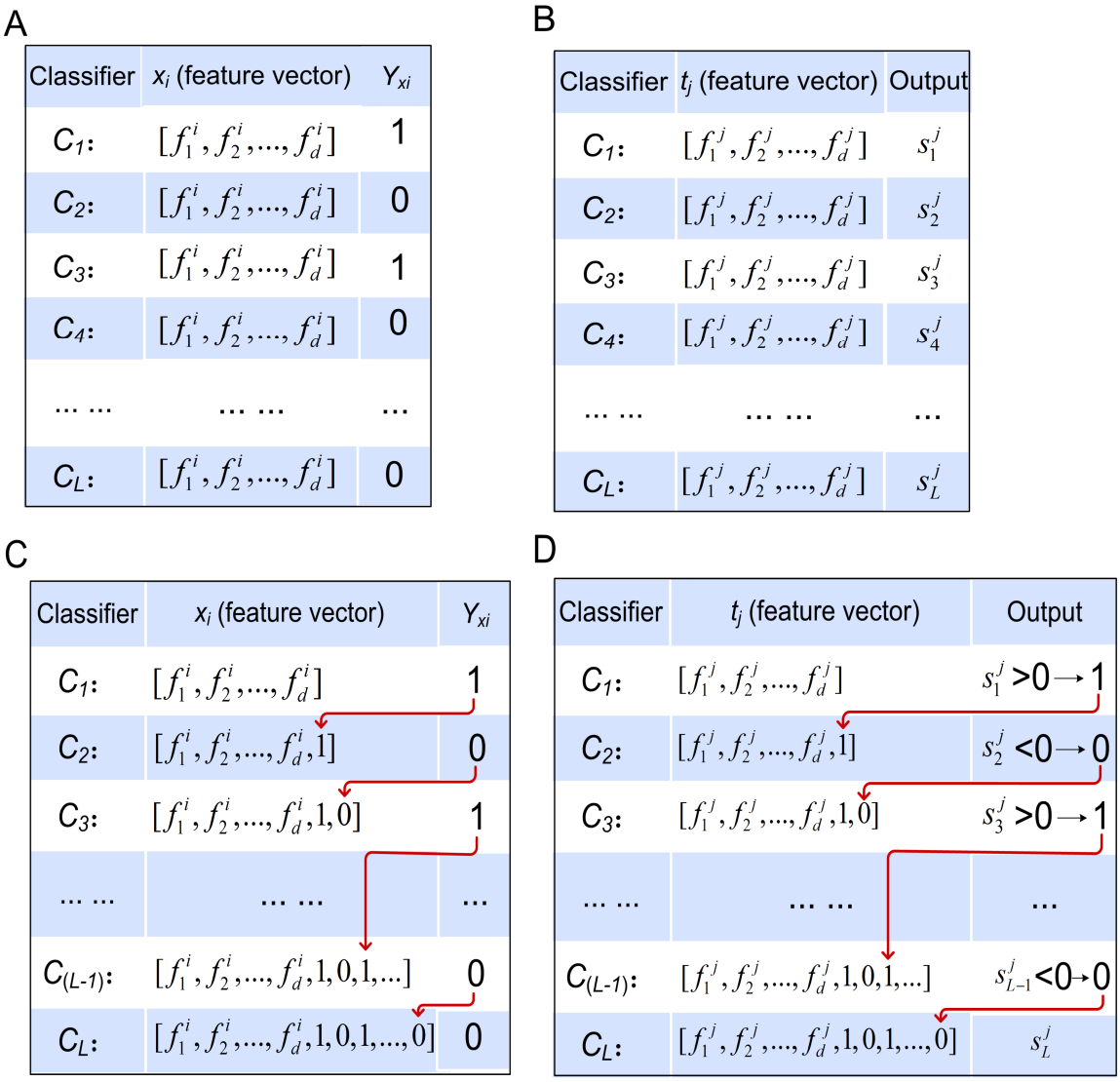
Label accuracy evaluates the prediction accuracy for each label, from which we can identify which subcellular locations are easier to recognize. The average label accuracy computes the average of L accuracies of labels, and can reflect the total performance.

Supplementary Table 1 The dataset is composed of the normal image dataset and the cancer image dataset, where the numbers of the former is shown in column 4, and the latter in column 5.

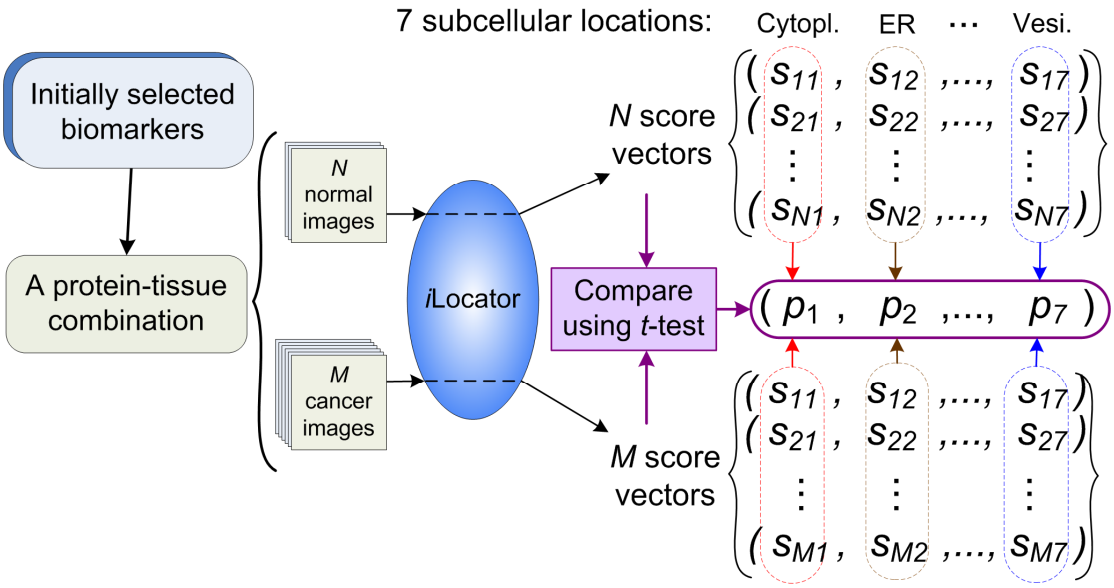
Antibody ID	Protein name	Subcellular locations in normal tissues	Number of images	
			normal	cancer
2384	Cysteine--tRNA ligase	Cytopl.	123	140
17097	Methionine--tRNA ligase	Cytopl.	113	130
29804	Aspartate--tRNA ligase	Cytopl.	117	127
2321	Major vault protein	Cytopl.	107	126
5853	Trafficking kinesin-binding protein 1	Cytopl.	111	130
3901	Endoplasmin	ER	95	146
18884	Protein disulfide-isomerase	ER	125	135
5480	E3 ubiquitin-protein ligase synoviolin	ER	125	133
992	Golgin subfamily A member 5	Gol.	123	123
10638	Golgi membrane protein 1	Gol.	126	136
37770	Arylsulfatase B	Lyso.	100	124
41788	Ceroid-lipofuscinosis neuronal protein 5	Lyso.	130	137
1523	60 kDa heat shock protein	Mito.	115	129
4479	AFG3-like protein 2	Mito.	125	142
28202	Carnitine O-palmitoyltransferase 2	Mito.	113	137
20637	Alpha-aminoadipic semialdehyde synthase	Mito.	114	131
6669	Bcl-2-associated transcription factor 1	Nucl.	116	140
3890	Interferon-inducible protein 4	Nucl.	114	132
6429	Transcription initiation factor TFIID subunit 7	Nucl.	90	127
13606	Huntingtin-interacting Protein 1	Vesi.	118	137
30372	Synaptotagmin-2	Vesi.	113	122
5922	Four and a half LIM domains Protein 2	Cytopl.+ Nucl.	103	125
1873	Alpha-actinin-4	Cytopl.+ Nucl.	116	126
571	Setrol-4-alpha-carboxylate 3-dehydrogenase	ER+ Vesi.	123	127
26485	BCKAD-E2	Cytopl.+ Mito.	121	136
6964	Ras-related protein Rab-7a	Lyso.+ Vesi.	124	140
29722	Aryl hydrocarbon receptor	Cytopl. + Nucl.	130	130
3734	Ras-related GTP-binding protein A	Gol.+ Lyso.+ Vesi.	110	128
	Total		3240	3696



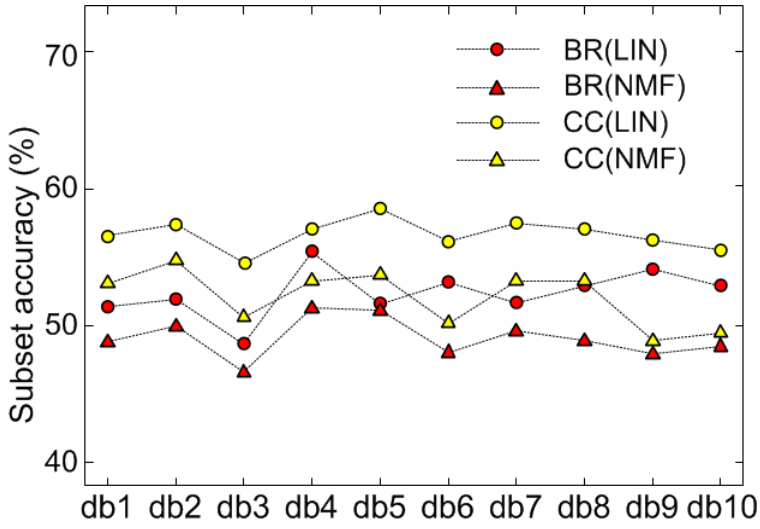
Supplementary Fig. 1 An example of calculating LBP code. For a pixel of the gray image, we use g_c to represent its gray value, while (g_0, g_1, ⋯, g_{U-1}) correspond to the gray values of the U neighbor pixels around g_c with the radius R. The LBP codes can be generated as $LBP_{U,R} = \sum_{k=0}^{U-1} s(g_k)2^k$, where $s(g_k) = \begin{cases} 1, & g_k \geq g_c \\ 0, & g_k < g_c \end{cases}$. A histogram of the magnitudes of these codes was plotted according to their magnitudes. And then, LBP features are extracted from this histogram. We set U=8, R=1. These LBP codes range from 0 to 255, so the calculated dimension of LBP features is 256.



Supplementary Fig. 2 The training and testing process of BR and CC. L is the total number of classes. Both of BR and CC have L classifiers, i.e. $C_1, C_2, \cdots, C_L$. (**A**) and (**C**) show the different training procedures of BR and CC for a sample $x_i = [f_1^i, f_2^i, \cdots, f_d^i]$, whose label vector is $Y = [1,0,1,0,\cdots,0]$. (**B**) and (**D**) illustrate the different testing procedures of BR and CC for a sample $t_j = [f_1^j, f_2^j, \cdots, f_d^j]$, and its predicted score vector $[s_1^j, s_2^j, \cdots, s_L^j]$ in the chain, where $s_1^j > 0, s_2^j < 0, s_3^j > 0, \cdots, s_{(L-1)}^j < 0$.



Supplementary Fig. 3 The process of using *t*-test to compare normal and cancer images of one protein-tissue combination.



Supplementary Fig. 4 The subset accuracies of db1~db10 classifiers using NMF and LIN separation approaches on both BR and CC modes.