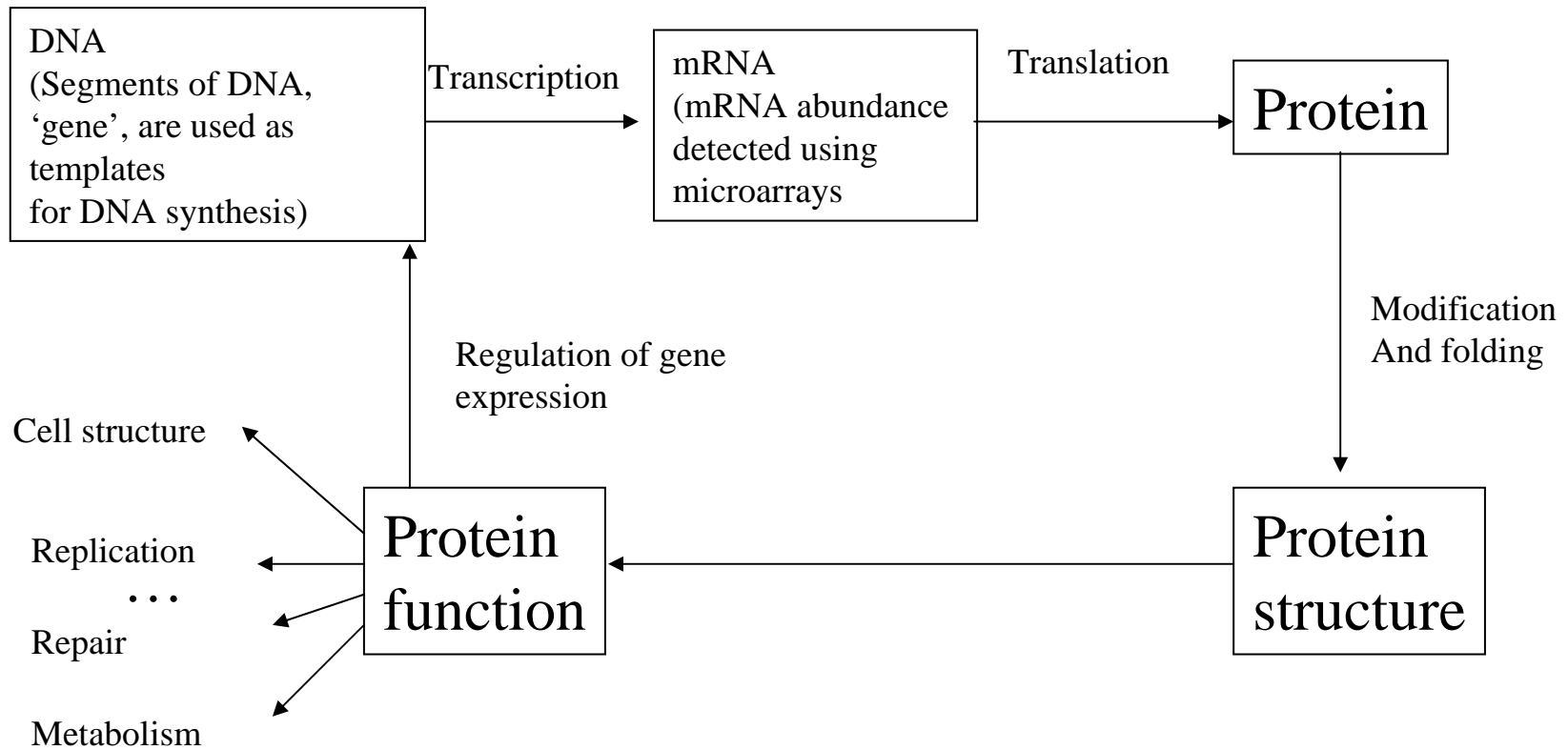


Cancer classification based on microarray gene expression

Bio-signal analysis

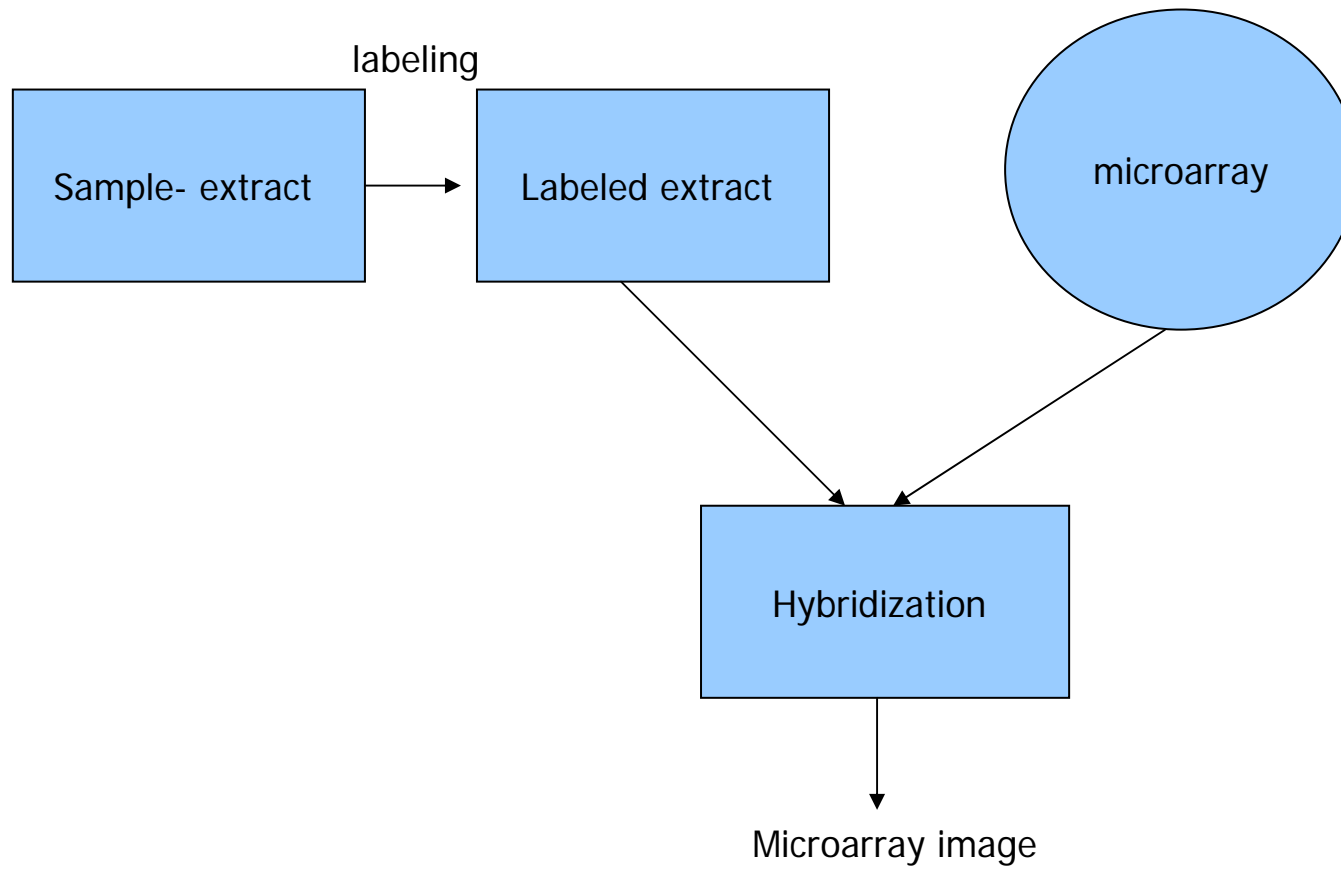
Molecular biology



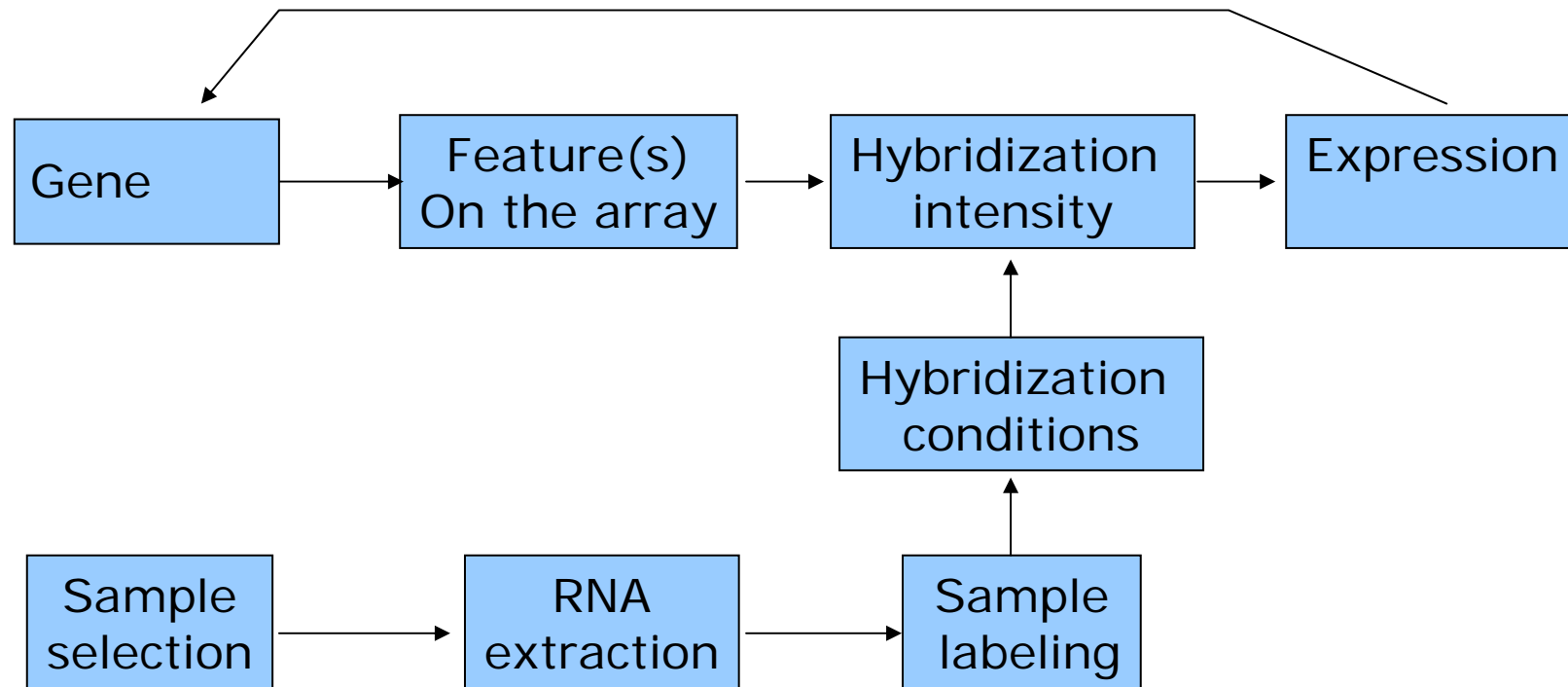
SRBCT Database

- Small round blue cell tumor
- Four categories
 - NB, neuroblastoma
 - RMS, rhabdomyosarcoma
 - NHL, non-Hodgkin lymphoma
 - EWS, Ewing family of tumors

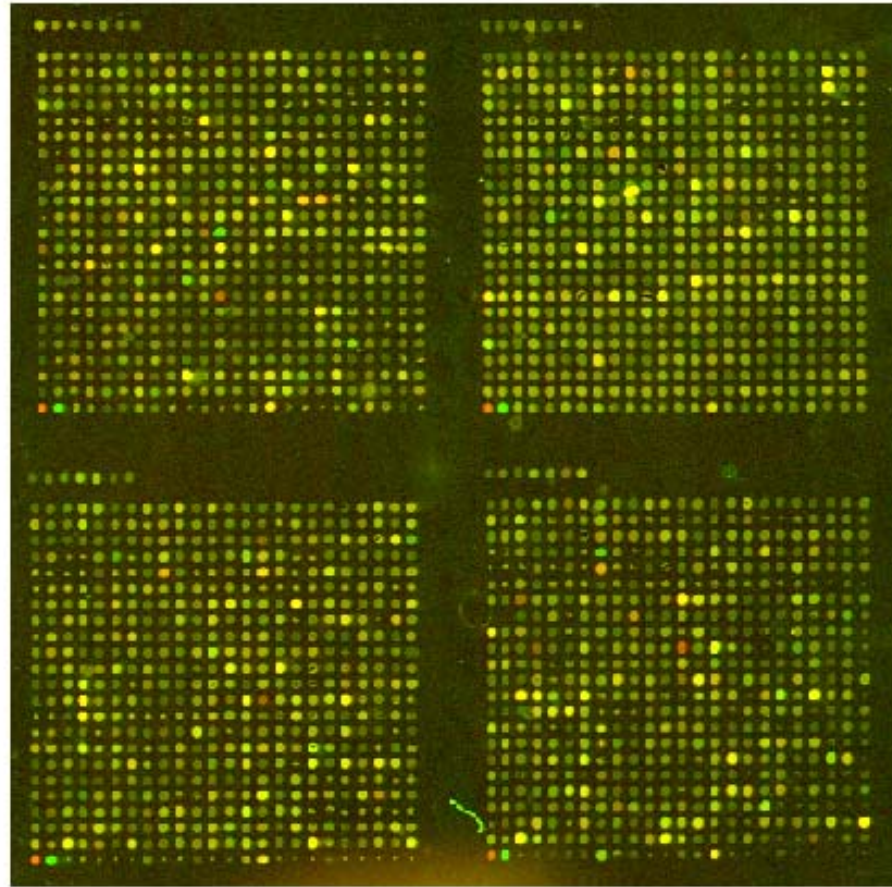
Microarray gene expression



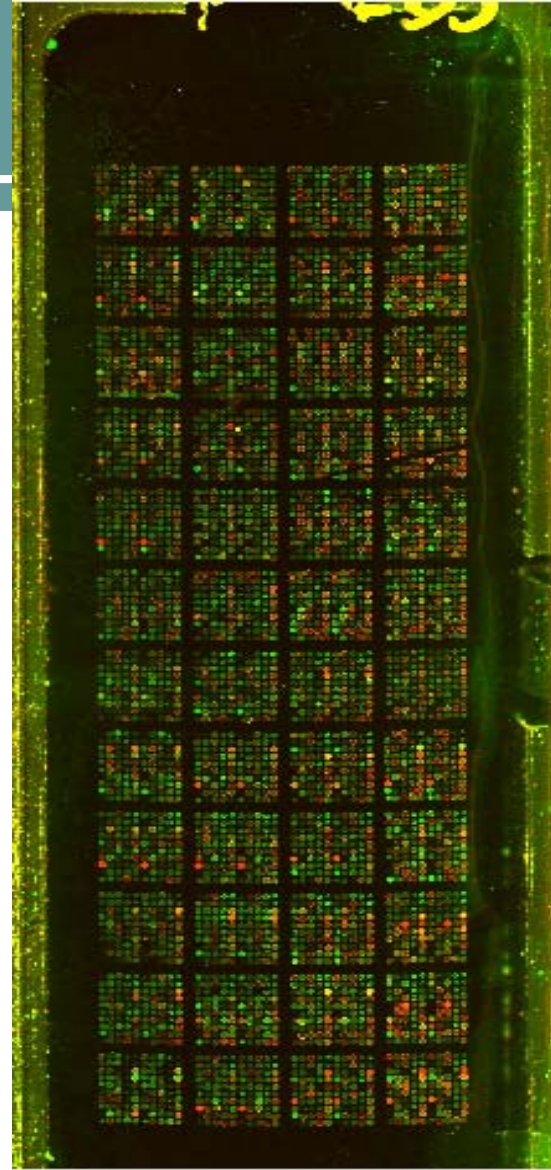
Microarray experiment



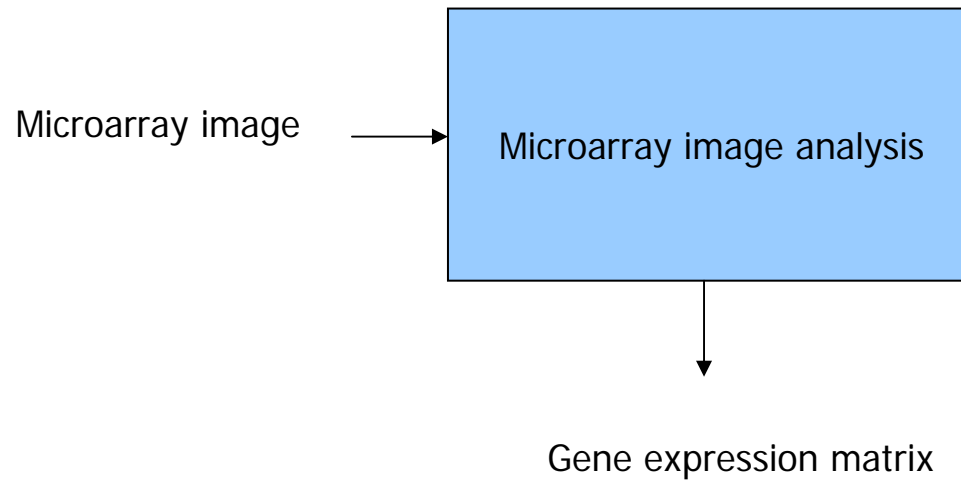
Microarray images



Microarray image



Microarray image analysis



Cancer categories

- Load cancer categories

```
Y=load('data\SRBCTY.dat')
```

	Category
Sample 1	
Sample 2	
⋮	
⋮	
⋮	
Sample 83	

1 NB, neuroblastoma

2 RMS, rhabdomyosarcoma

3 NHL, non-Hodgkin lymphoma

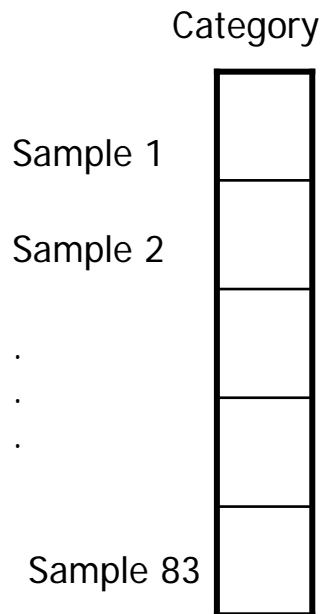
4 EWS, Ewing family of tumors

Training and testing data

- The first sixty samples are used to form training data. Remains form testing data.

Classification

- Discriminate one type of cancer from the others.



0: 1 NB, neuroblastoma

2 RMS, rhabdomyosarcoma

3 NHL, non-Hodgkin lymphoma

1: 4 EWS, Ewing family of tumors

Gene screening

- Every d , e.g. 5, consecutive genes are grouped.
- Specify a gene segment

```
group_no=201; d=5;
```

```
gene_seg = (group_no-1)*d: group_no*d
```

```
x=X(1:60, gene_seg)';
```

- Find the testing error of using the i th group of genes for cancer classification

Data flow

Training data
 $x: dxN, d=5, N=60$
 $y: 1xN$

```
[E,NetDef,W1,W2]=noisecan(x',y',M);  
[Y_sim,E,PI] = nneval(NetDef,W1,W2, x, y,1);
```

Testing data
 $x_test: 5x23$
 $y_test: 1x23$

```
[Y_sim,E,PI] = nneval(NetDef,W1,W2, x_test, y_test,1);
```

Testing errors

Exercise

SRBCT cancer classification by learning MLP

- Load gene matrix
- Load sample category
- Select one category to be discriminated from the others
- Select one gene group
- Form training data: x and y
- Form testing data: x_{test} , y_{test}
- Train an MLP network
- Test
- Output the testing error

Gene self-organization

- Yeast
- Gene matrix
- [yeastdata822.zip](#)