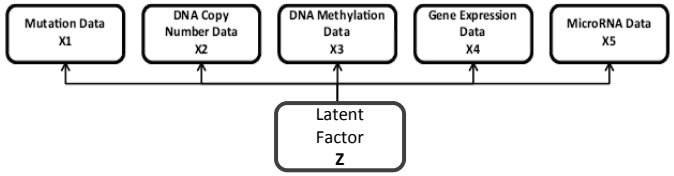
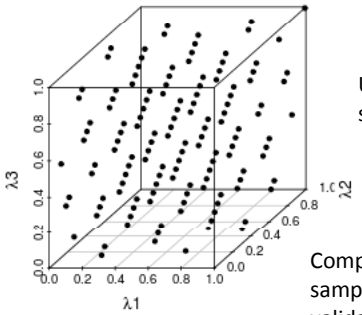
Supplementary Figure 1

A

iCluster model

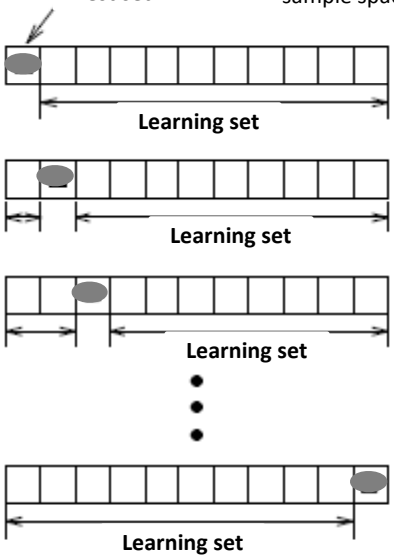




Uniform sampling of sparsity parameter sets λ

Compute reproducibility index (RI) given a sampled parameter set in 5-fold cross-validation. Choose λ over the entire sample space with maximum RI

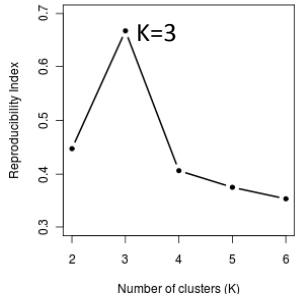
B



U\V	V_1	V_2	...	V_C	Sums
U_1	n_{11}	n_{12}	...	n_{1C}	a_1
U_2	n_{21}	n_{22}	...	n_{2C}	a_2
:	:	:	...	:	:
U_R	n_{R1}	n_{R2}		n_{RC}	a_R
Sums	b_1	b_2	...	b_C	

A reproducible index (RI) based on the adjusted Rand Index between predicted (U) and fitted (V) test set cluster membership

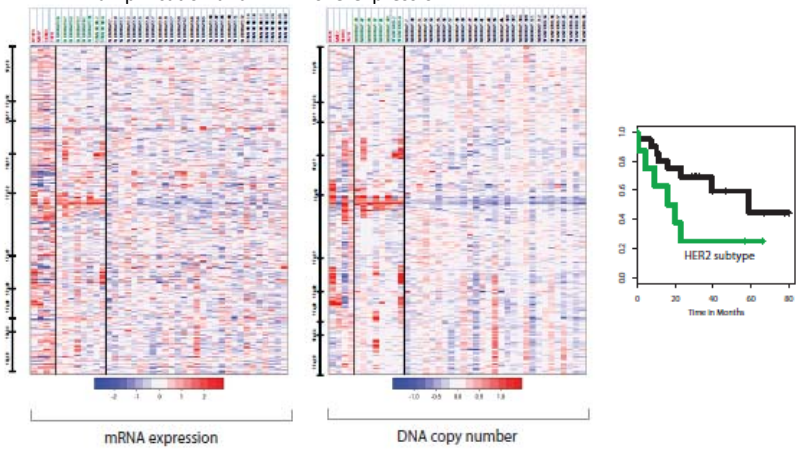
C



- Repeat **A** and **B** for different K.
- Choose the number of clusters (K) that has the highest reproducibility index

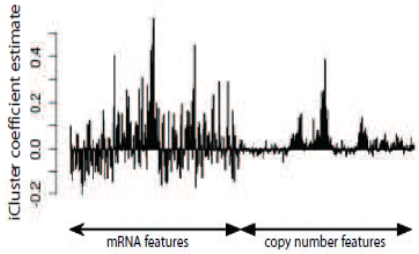
D

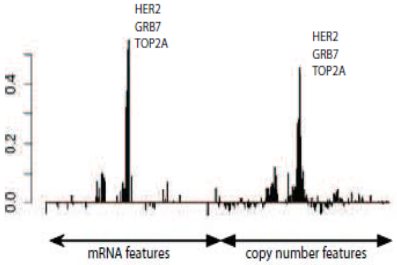
HER2 subtype (green labels) jointly identified by concordant DNA amplification and mRNA overexpression



E

Standard feature space



F

Sparse feature space