Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response

Stefano Monti
Broad Institute of MIT & Harvard
smonti@broad.mit.edu
http://www.broad.mit.edu/cancer

Diffuse Large B-cell Lymphoma
(DLBCL)

- Most common lymphoid malignancy (~40%)
- Significant clinical & genetic heterogeneity
- Hypothesis:
  - gene expression profiling will reveal disease heterogeneity

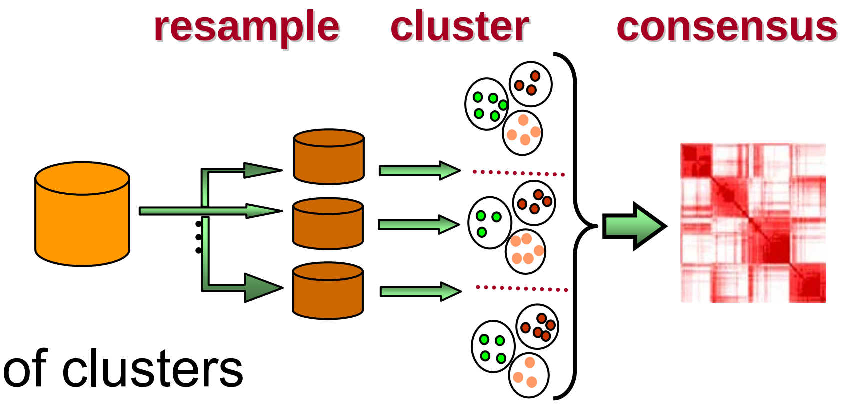
DLBCL expression profiling

- 176 DLBCL samples.
- Nodal biopsies from untreated patients.
- Affymetrix U133A/B chips (~42K probes)
- Goal: unsupervised analysis to discover novel substructure

Analysis of DLBCL Substructure

by unsupervised analysis

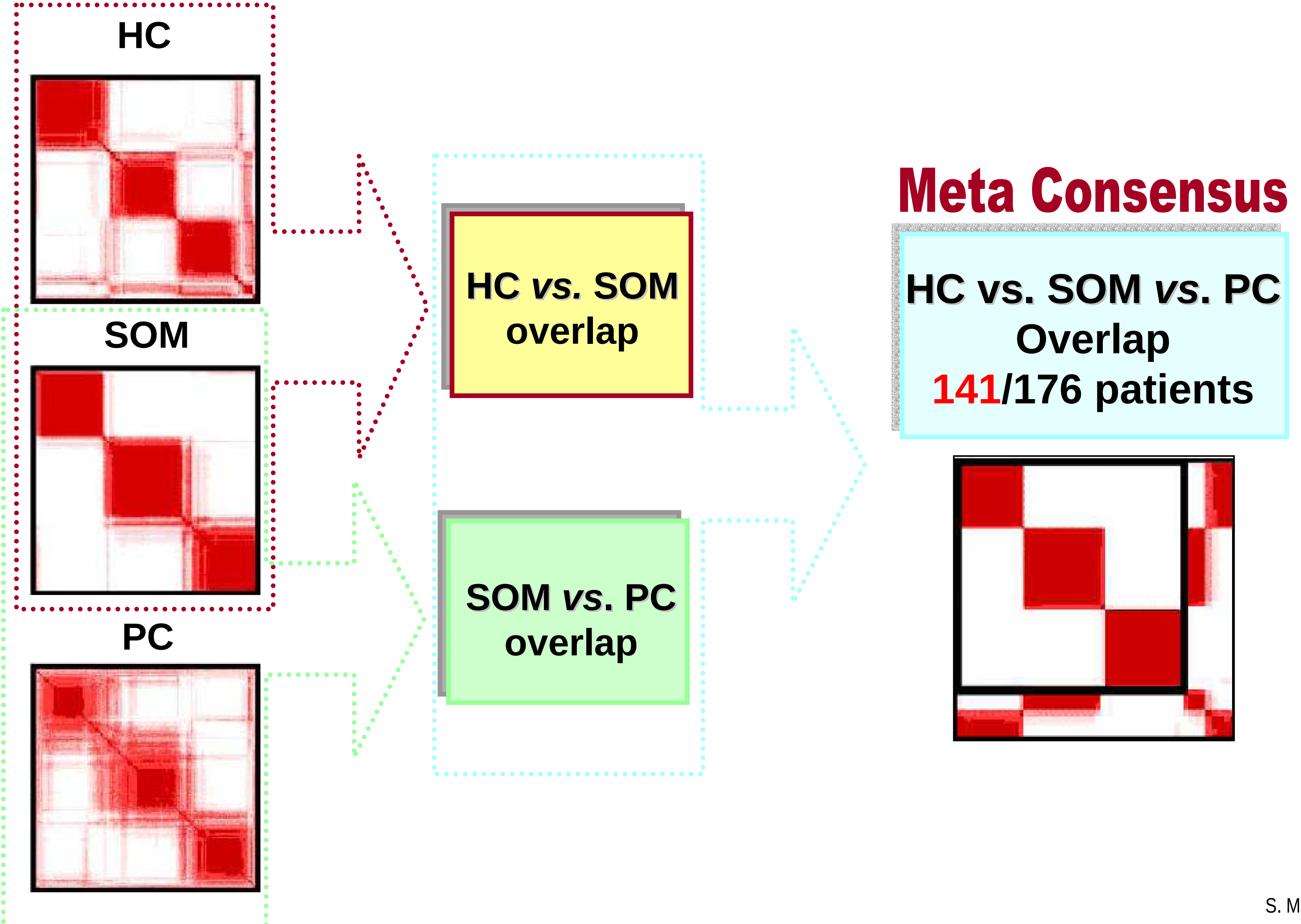
- Consensus clustering
  - identifies robust clusters
  - resampling-based method
  - automatically selects the number of clusters



resample cluster consensus

- Used with 3 different clustering algorithms
  - Hierarchical clustering (HC)
  - Self-organizing Map (SOM)
  - Probabilistic Clustering (PC)

Consensus Clustering of DLBCL



HC

SOM

PC

HC *vs.* SOM overlap

SOM *vs.* PC overlap

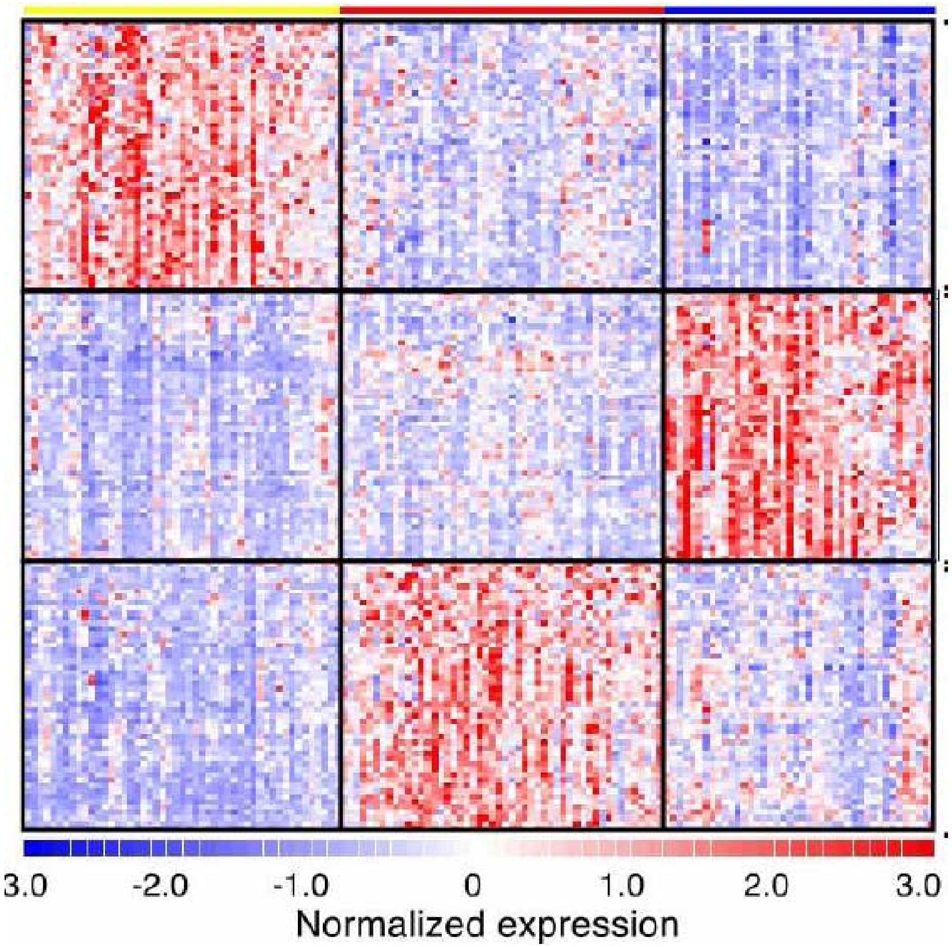
Meta Consensus

HC vs. SOM *vs.* PC Overlap

141/176 patients

DLBCL Consensus Clusters

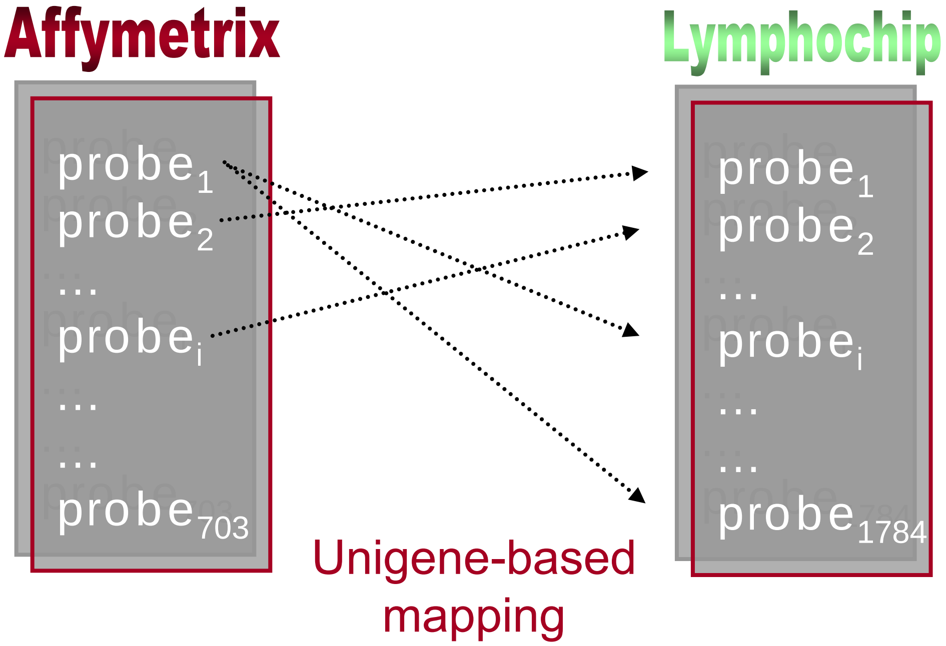
OxPhos **BCR/Proliferation** **Host Response**



Validation of Consensus Clusters

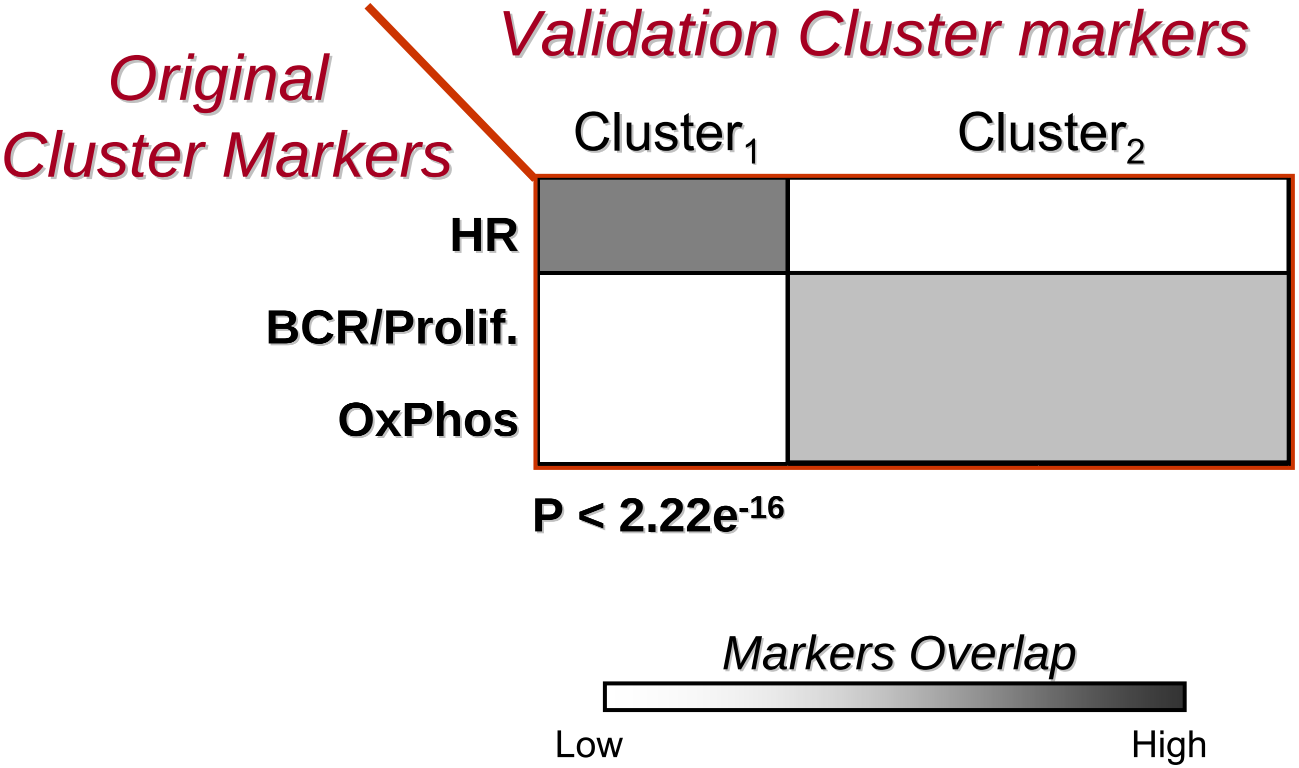
on Independent Database

- 221 DLBCL samples on Lymphochip [Rosenwald, et al., NEJM 2003]
- Cross-platform mapping



Validation of Consensus Clusters

on Independent Database



Validation Cluster markers

Original Cluster Markers

	Cluster$_1$	Cluster$_2$
HR	▓	☐
BCR/Prolif.	☐	░
OxPhos		

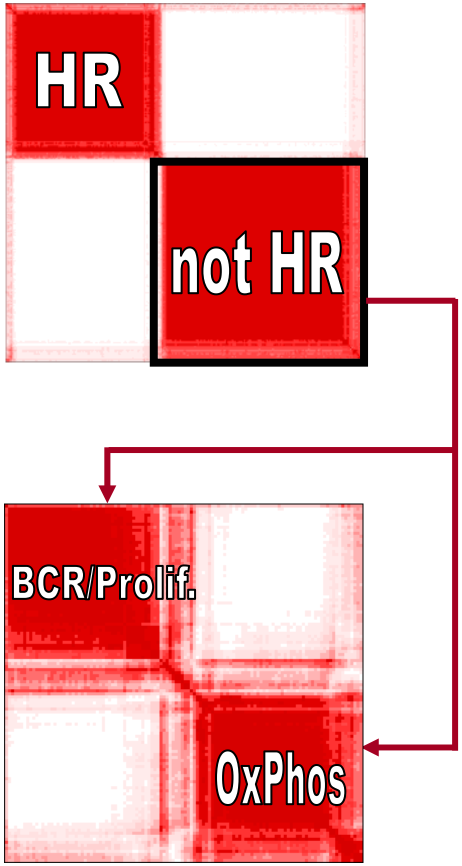
P < 2.22e^{-16}

Markers Overlap

Low High

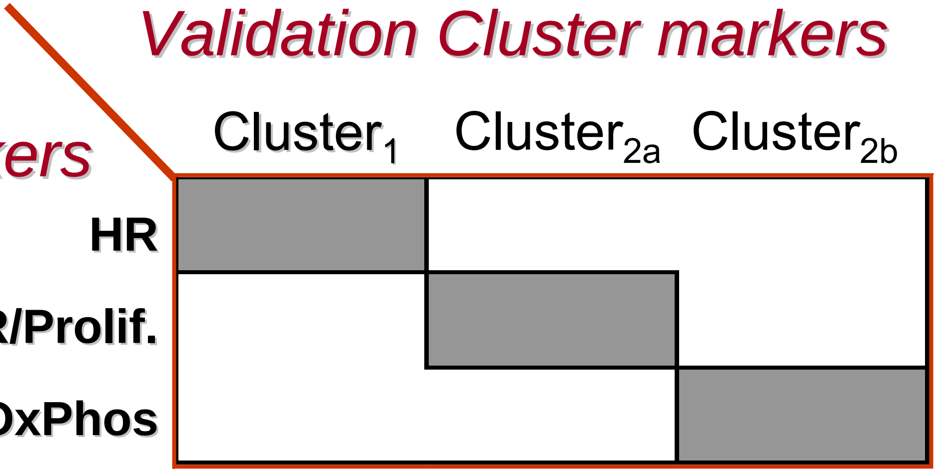
Validation of Consensus Clusters

on Independent Database



Original Cluster Markers

Validation Cluster markers



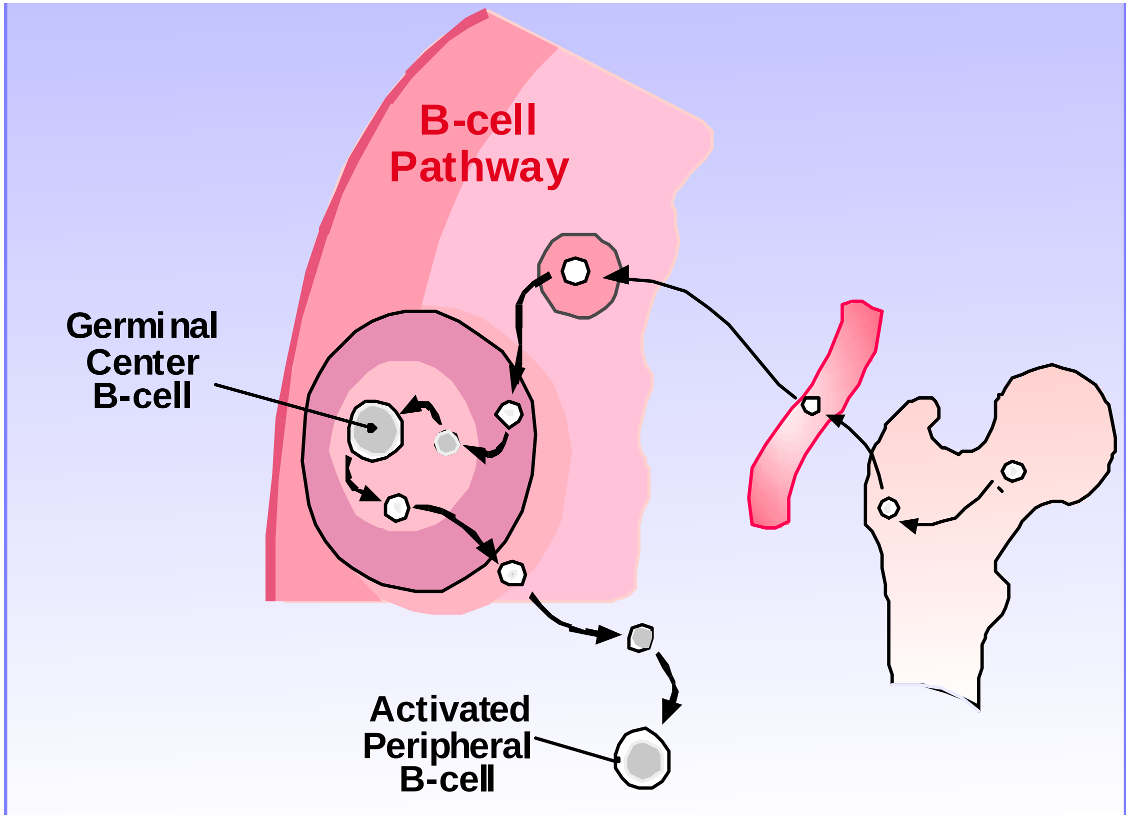
	Cluster$_1$	Cluster$_{2a}$	Cluster$_{2b}$
HR	▨		
BCR/Prolif.		▨	
OxPhos			▨

P < 2.22e^{-16} **P < 0.009**

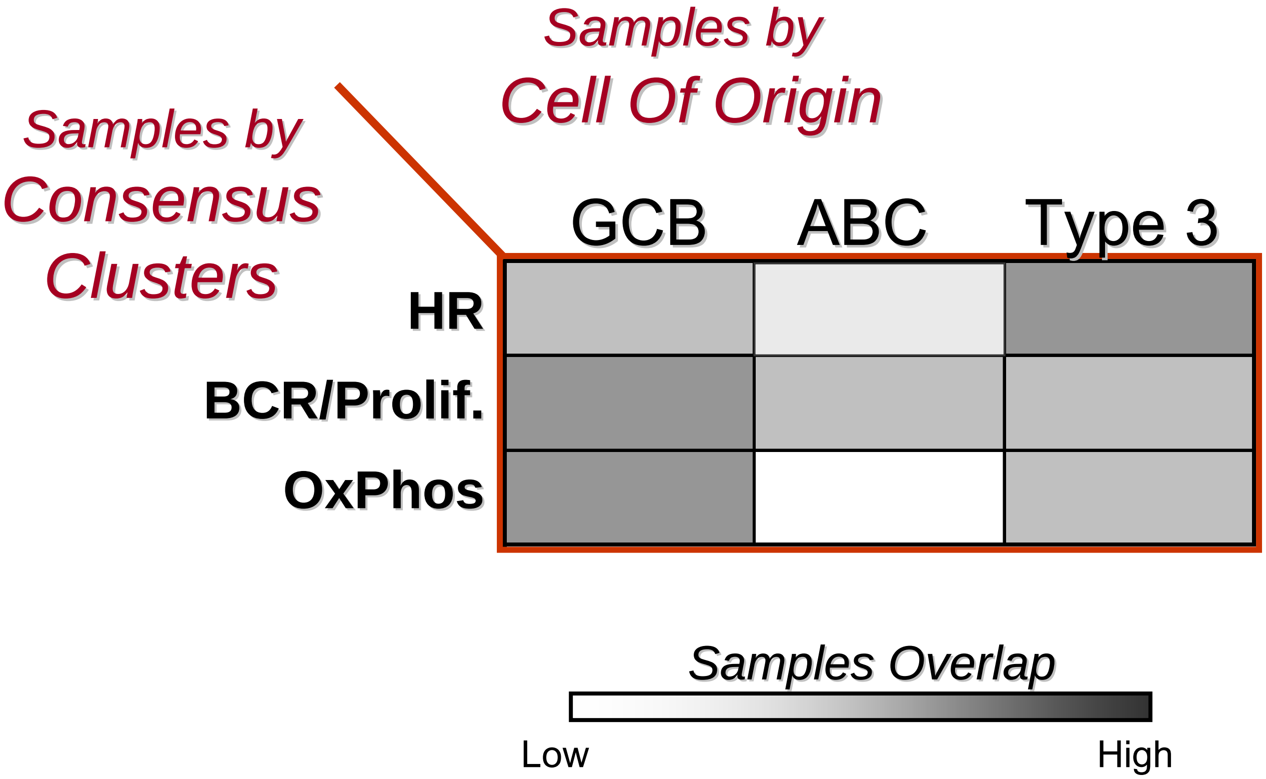
Markers Overlap

Low High

Consensus Clusters *and* Cell Of Origin



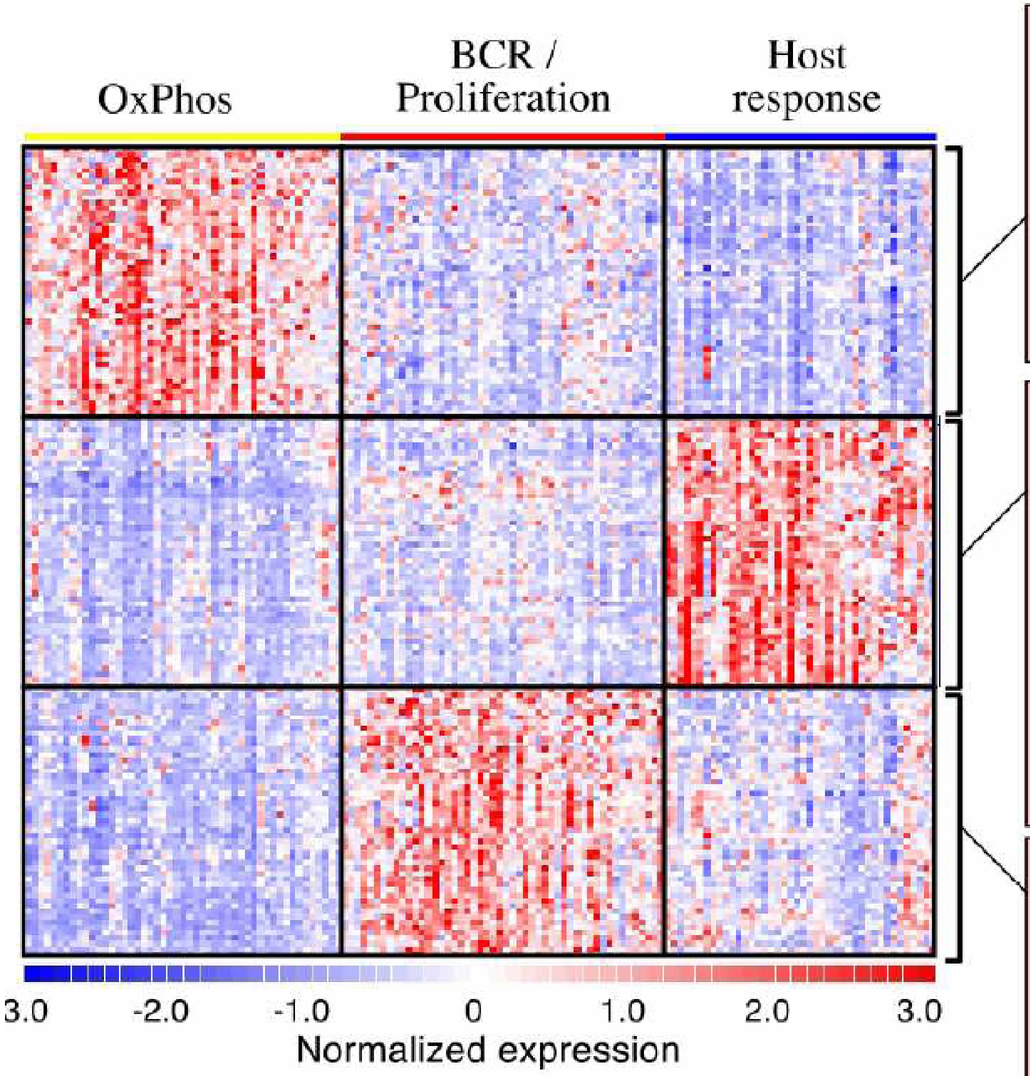
Consensus Clusters *and* Cell Of Origin



Non-overlapping information

COO assignment based on methodology described in [Wright, et al., PNAS 2003].

DLBCL Consensus Clusters



NADH dehydrogenase 1 α/β subcomplex 1
Cytochrome c oxidase (COX) 7A2L
ATP binding protein
Proteosome α 5
ATP synthase, mitochon. FO complex, subunit c iso 3
ATP synthase, mitochon. F1 complex, γ polypeptide 1
Proteosome α 2
Proteosome α 6
Mitochondrial ribosomal protein L3
Translocase of inner mitochondrial membrane 8B
NADH dehydrogenase 1 β subcomplex 1

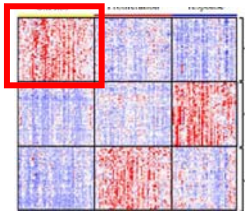
Stromal cell-derived factor 1
TNF-related death ligand 1β (APRIL)
IFN-induced transmembrane protein 2
TNFRSF1β
LAMP1
GATA3
cMAF
CD3ε
Linker for activation of T-cells
CD2
T-cell immune regulator 1
TNFRSF1α
Integrin β2
IFN regulatory factor 1

CD79A
Phospholipase C γ 2
MAP4K1
CD22
CD37
Postmeiotic segreg. increased-2-like 9, 8, 2, 11, 3
Proliferation-associated protein 100
Ki67
Inositol polyphosphate-5-phosphatase
CHL1-related helicase

H '04

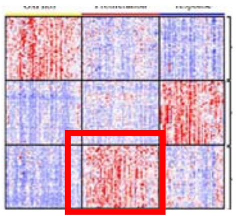
OxPhos cluster

- Genes involved in oxidative phosphorylation (p≤.002) and mitochondrial function (p≤.003)
  - BFL-1/A1 (anti-apoptotic BCL-2 family member)
  - members of NADH dehydrogenase complex
  - members of the COX complex
  - ATP synthase components



BCR/Proliferation cluster

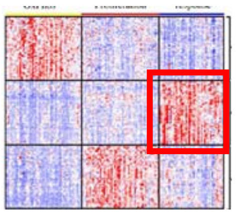
- BCR signaling components
  - CD19, IG, CD79a, BLK, SYK, PLCgamma2, MAP4K
- B-cell transcription factors
  - PAX5, OBF-1, E2A, BCL6, STAT6, MYC
- Cell cycle regulatory genes
  - CDK2, MCM
- DNA repair genes
  - PMS2, H2AX, PTIP, p53



Host Response (HR) cluster

signature largely defined by the associated host response rather than the tumor itself

- Components of TCR (TCRα/β, CD3), CD2, T/NK cell activation, and complement cascade.
- Co-regulated inflammatory mediators
- More abundant monocyte/macrophage & dendritic cell transcripts
- Interferon-induced genes, TNF ligands/receptors, cytokine receptors.



Tumor Infiltrating Lymphocytes (TILs)* (morphology)

Consensus clusters	> 20 TILs/HPF
HR	65%
BCR/proliferation	14%
OxPhos	11%

p < .0001

* Morphologically normal (CD2+) lymphocytes with round/oval nuclei and delicately dispersed chromatin.

TILs and Dentritic Cells in HR tumors

(immunostaining)

- Increased # of CD2$^+$/CD3$^+$ T-cells (p≤.005)
- Increased # of GILT$^+$ dendritic cells (DC) (p=.06)
  - Interdigitating DCs (S100$^+$, CD1a$^-$ CDC123$^-$), (p<.009)
  - correlated with TILs (p<.0001)

HR tumors and T-cell/histiocyte-rich LBCLs

- HR tumors appear
    - in younger patients (p=0.04);
    - with higher incidence of splenic (p=0.02) and BM involvement (p=0.03).

- 8/10 patients with T-cell/Histiocyte-rich LBCLs fall in the HR cluster

Genetic abnormalities

in DLBCL consensus clusters

Nearly absent in the HR cluster

Genetic abnormality	OxPhos (n=27)	BCR/ Prolif. (n=50)	HR (n=29)	Total (n=116)
t(14;18)	8 (22%)	5 (10%)	1 (3%)	14 (12%)
t(3; ...)	2 (5%)	8 (16%)	1 (3%)	11 (9%)
None	27 (73%)	37 (74%)	27 (93%)	91 (78%)

*

p = .059

* Columns may not add to 100% because of rounding.

Summary

- Identified three robust clusters by transcriptional profiling and confirmed in independent series.

- HR cluster characterized by:
    - inflammatory/immune cell infiltrate
    - fewer known genetic lesions
    - distinct clinical features

- Different mechanisms of transformation?

Participants

Christine Ladd
David Peck
Todd Golub

Kim Last
Andrew Lister

Paul Kurtin
Tom Habermann

Francoise Berger
Gilles Salles

Nancy Lee Harris
Laurence de Leval

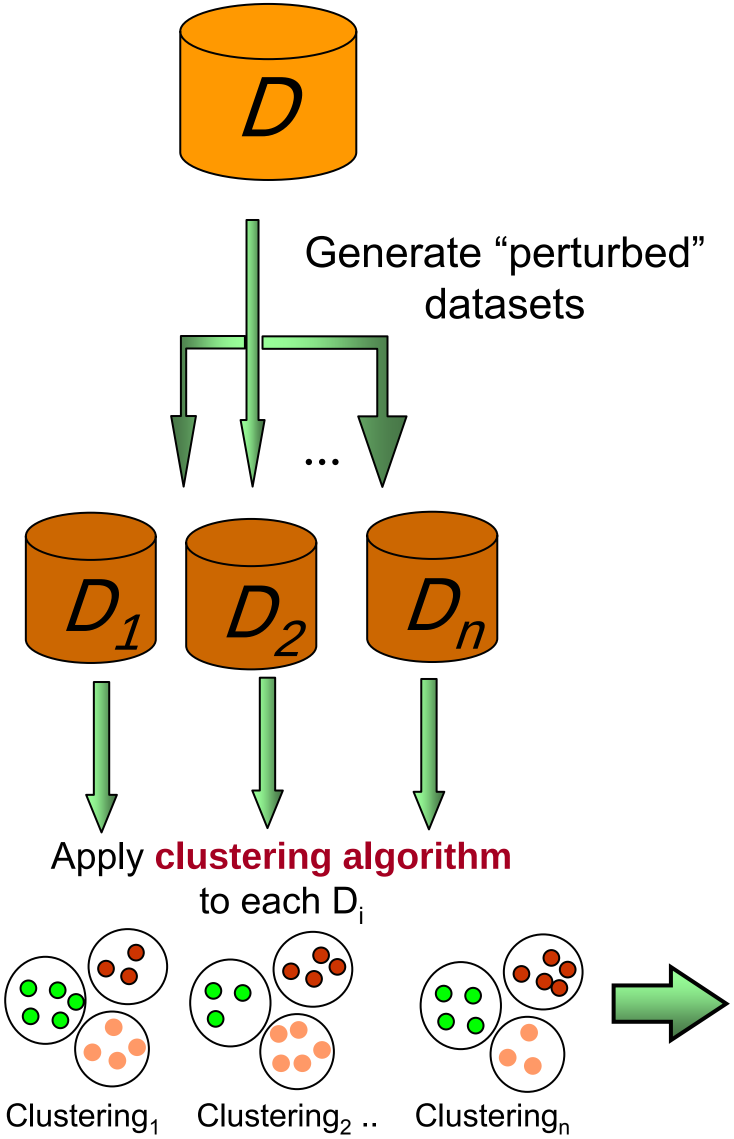
Giorgio Cattoretti
Riccardo Dalla Favera

Andrew Weng
Jeffery Kutok
Paola Dal Cin
Geraldine Pinkus
Jon Aster

Kerry Savage
Friedrich Feuerhake
Ricardo Aguiar
Peter Smith
Erxi Wu
Donna Neuberg
Margaret Shipp

Appendix

Consensus Clustering



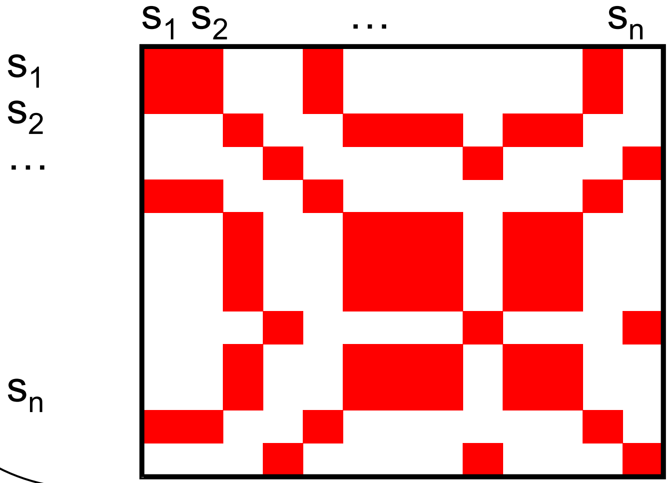
Generate “perturbed” datasets

Apply **clustering algorithm** to each D_i

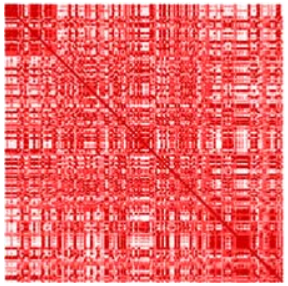
Clustering_1 Clustering_2 .. Clustering_n

Consensus matrix: counts proportion of times two samples are clustered together.

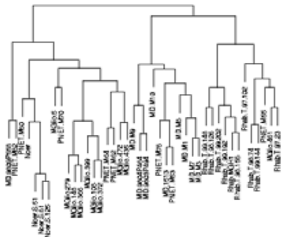
- **RED** (**1**) two samples **always** cluster together
- **WHITE** (**0**) two samples **never** cluster together



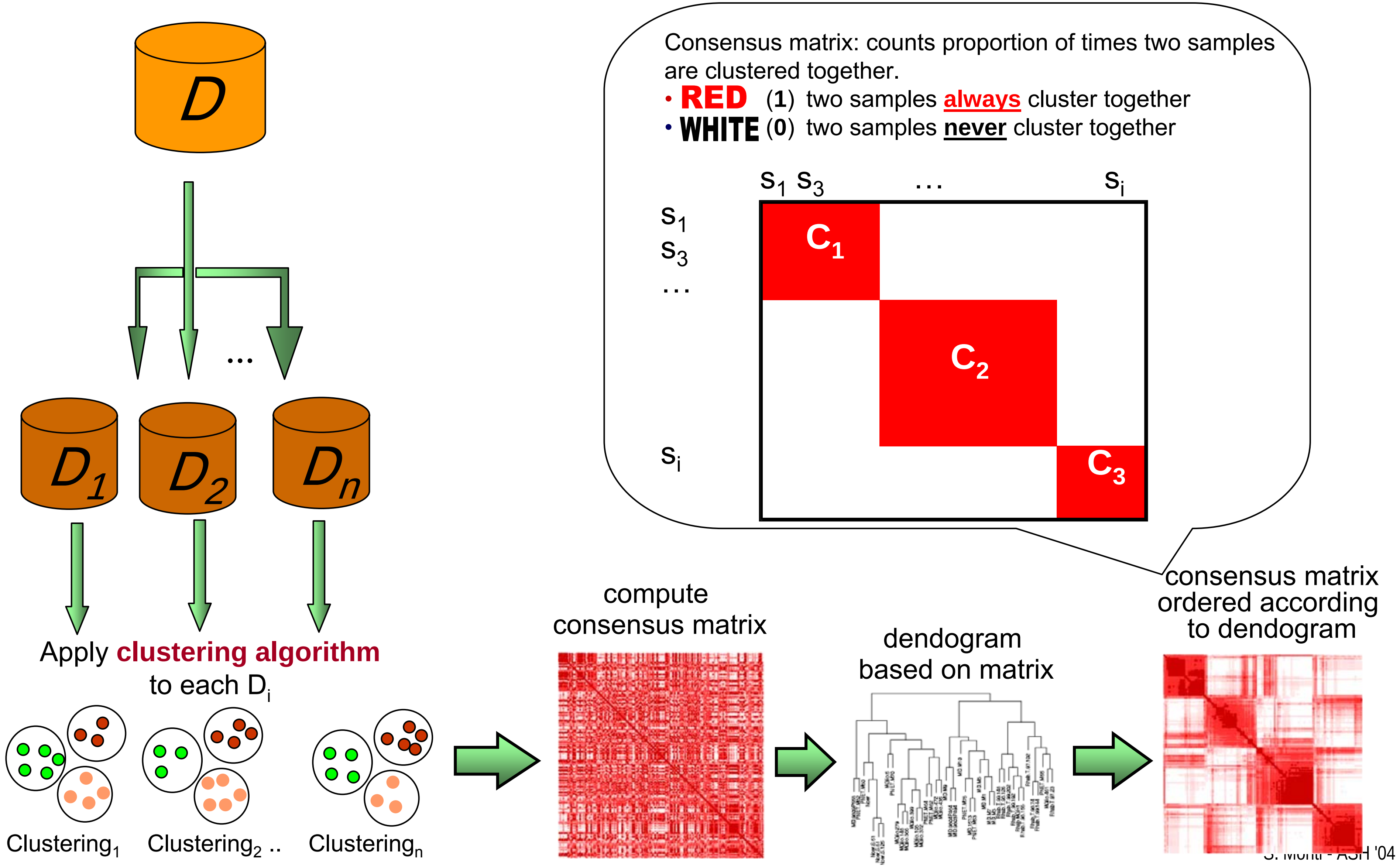
compute consensus matrix



dendogram based on matrix



Consensus Clustering



Consensus matrix: counts proportion of times two samples are clustered together.

- **RED** (**1**) two samples **always** cluster together
- **WHITE** (**0**) two samples **never** cluster together

Apply **clustering algorithm** to each D_i

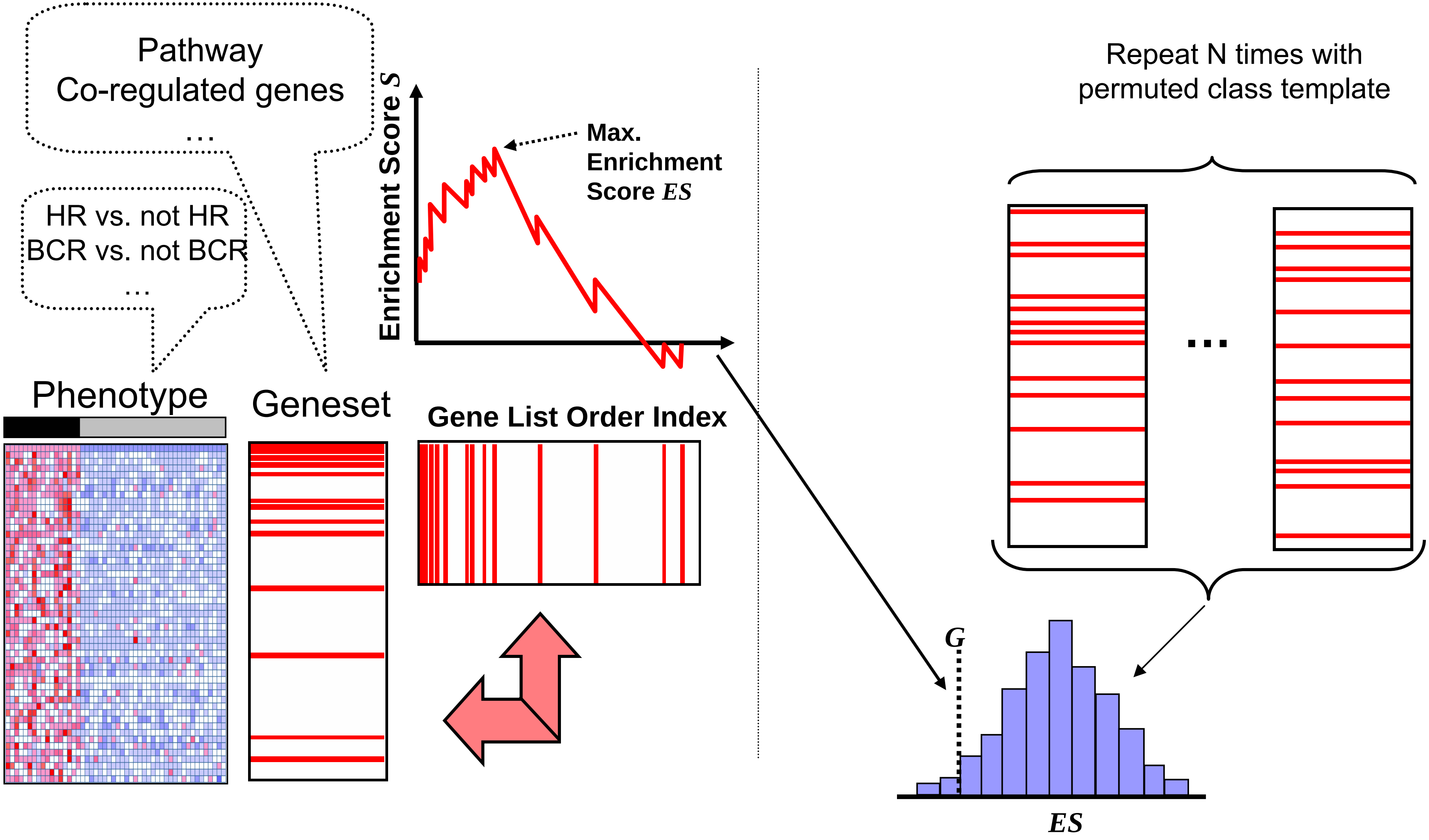
Clustering_1 Clustering_2 .. Clustering_n

compute consensus matrix

dendogram based on matrix

consensus matrix ordered according to dendogram

Gene Set Enrichment Analysis



Pathway
Co-regulated genes
…

HR vs. not HR
BCR vs. not BCR
…

Phenotype

Geneset

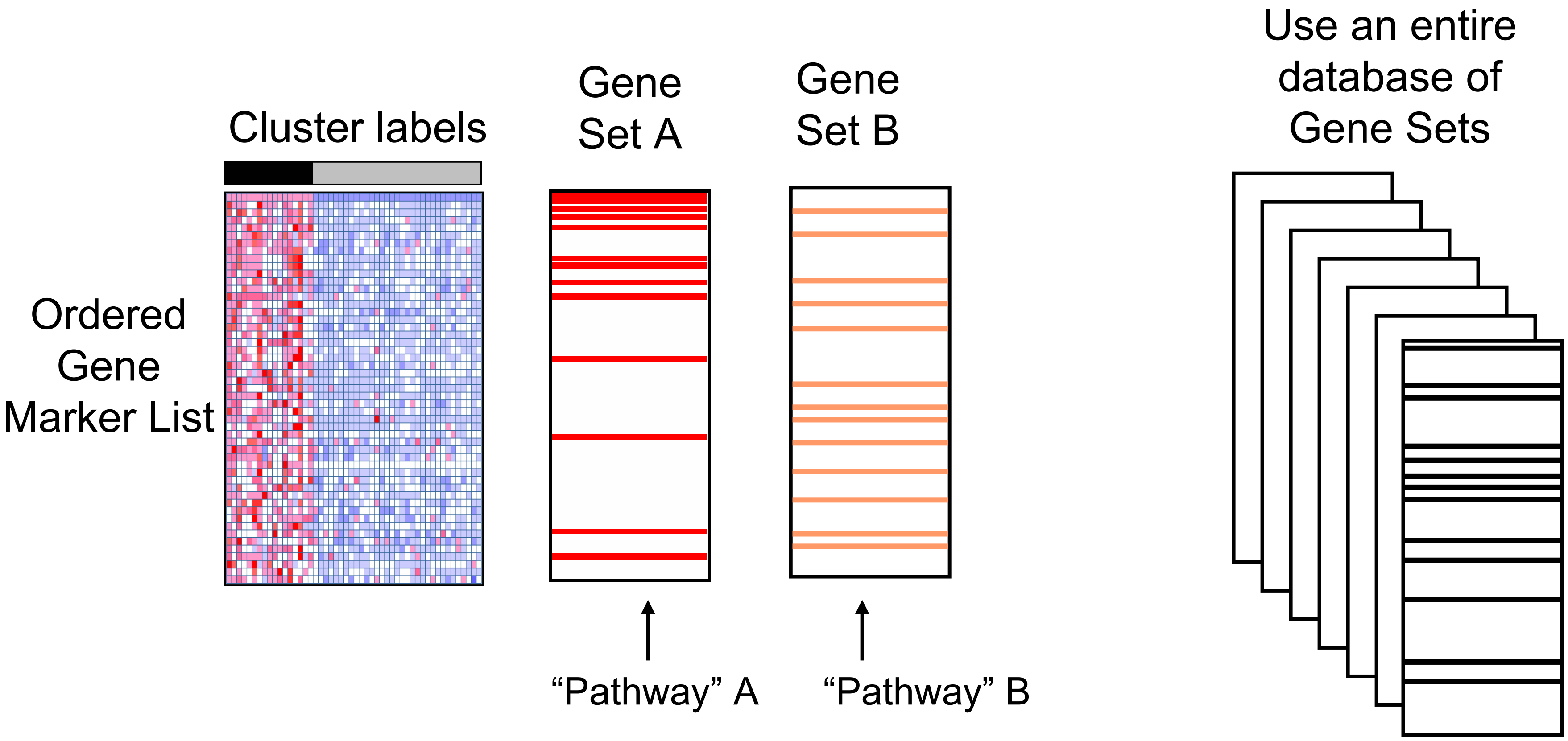
Gene List Order Index

Max. Enrichment Score *ES*

Repeat N times with permuted class template

Clusters' annotation

by GSEA



Use an entire database of Gene Sets

Cluster labels

Gene Set A

Gene Set B

Ordered Gene Marker List

"Pathway" A

"Pathway" B