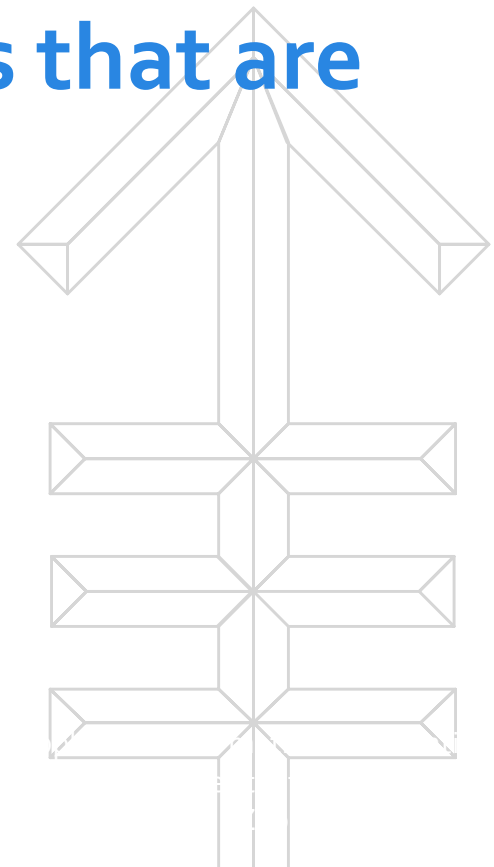Memorial Sloan Kettering Cancer Center™
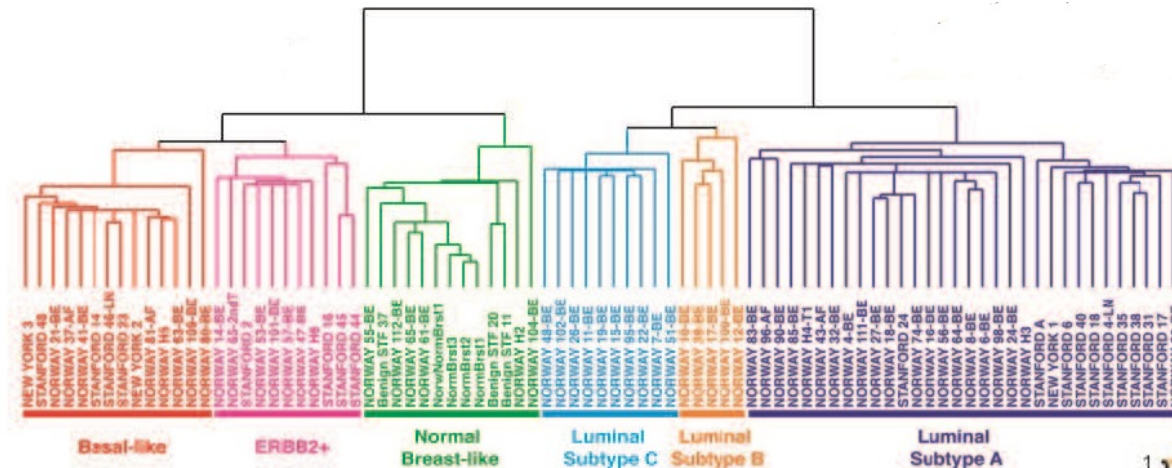
COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH
BIOSTATISTICS

Application of a method for identifying disease subtypes that are etiologically heterogeneous

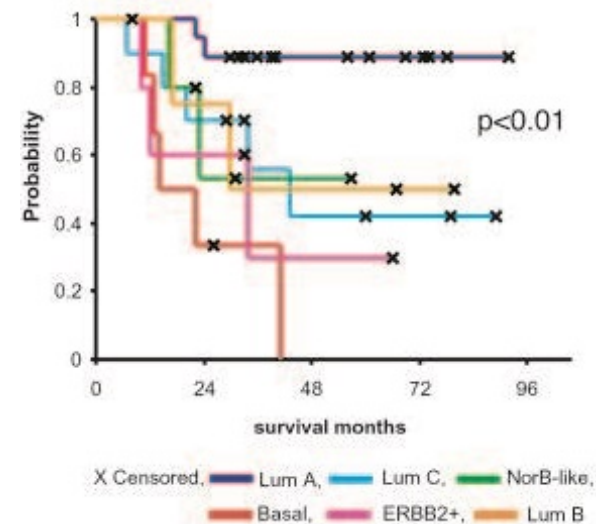Emily C. Zabor, Shuang Wang, Colin B. Begg

Breast cancer is biologically diverse and subtypes of disease risk and prognosis are recognized
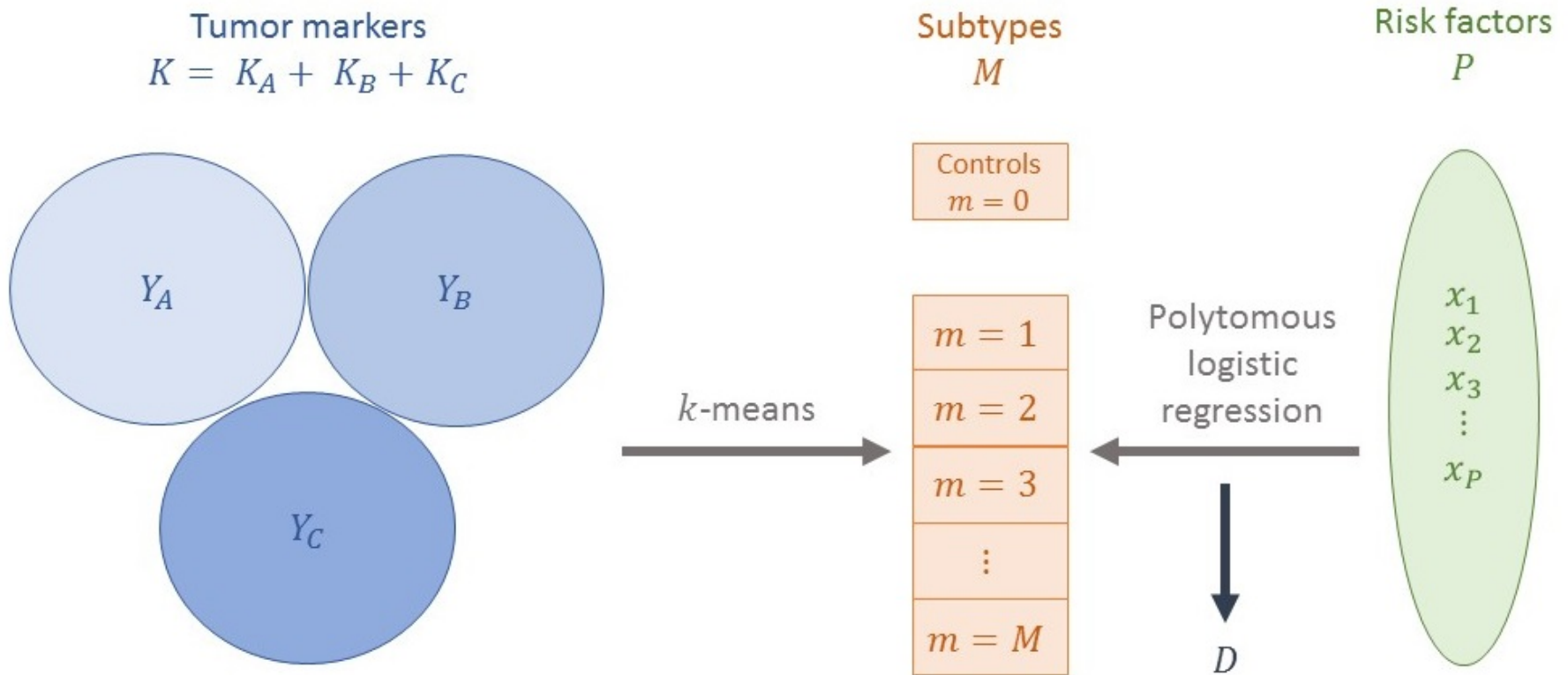

Primary molecular subtypes of breast cancer can be approximated with 4 classes based on 3 IHC markers:

- Estrogen receptor (ER)
- Progesterone receptor (PR)
- Human epidermal growth receptor (Her2)
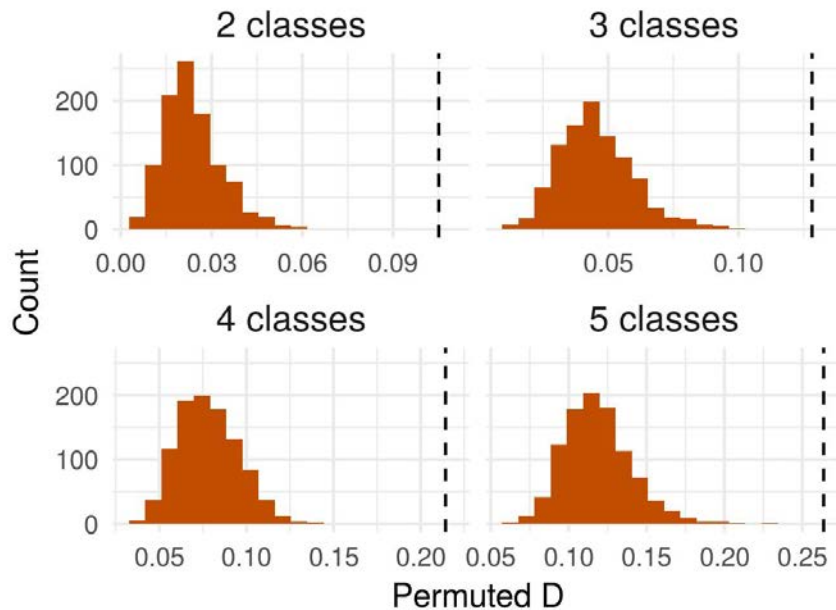
Sorlie et al (2001). Gene Expression Patterns of Breast Carcinomas Distinguish Tumor Subclasses with Clinical Implications. PNAS (19): 10869-74.

Memorial Sloan Kettering Cancer Center

Application of a method for identifying disease subtypes that are etiologically heterogeneous - Zabor EC, Wang S, Begg CB

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH
BIOSTATISTICS

Seek to cluster tumor marker data and optimize a scalar measure of etiologic heterogeneity, D

Memorial Sloan Kettering Cancer Center.

Application of a method for identifying disease subtypes that are etiologically heterogeneous - Zabor EC, Wang S, Begg CB

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH
BIOSTATISTICS

Discovery analysis selects 112 genes and identifies 4 classes

Class size	D difference	P-value
3 vs 2	0.023	0.454
4 vs 2	0.109	0.002
5 vs 2	0.159	0.012
4 vs 3	0.086	0.007
5 vs 3	0.135	0.015
5 vs 4	0.049	0.358

- 532 breast cancer cases included in clustering
- Top-ranked genes of 406 total selected for inclusion based on individual gene D

Memorial Sloan Kettering Cancer Center.

Application of a method for identifying disease subtypes that are etiologically heterogeneous - Zabor EC, Wang S, Begg CB

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH
BIOSTATISTICS

Age at diagnosis, menopausal status, and race are driving the heterogeneity across subtypes
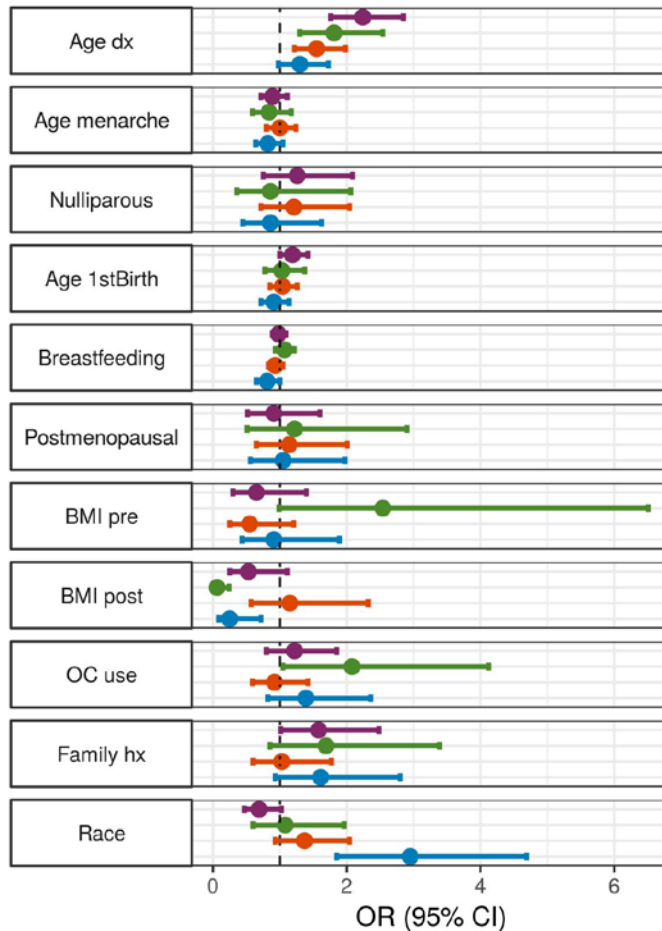
Class size	D
Discovery	0.214
Validation	0.245
Traditional IHC 4-class	0.148

- Validation produces reasonably similar gene rankings
- Optimal 4-class D exceeds that in traditional IHC 4-class
- ER is a key gene in distinguishing subtypes

Application of a method for identifying disease subtypes that are etiologically heterogeneous - Zabor EC, Wang S, Begg CB