Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer

Motivation

Observation: breast-cancer patients at same stage have different outcomes

Problem: existing outcome predictors are poor
- lymph nodes
- histological grade

Goals:
- identify categories of breast cancer
- predict outcome based on gene expression
- decide therapy accordingly

Treatment

"Chemotherapy or hormonal therapy reduces risk of distant metastases by approximately one-third; however 70-80% of patients receiving this treatment would have survived without it."

A main contribution of this method is lower false positives.

Result Overview

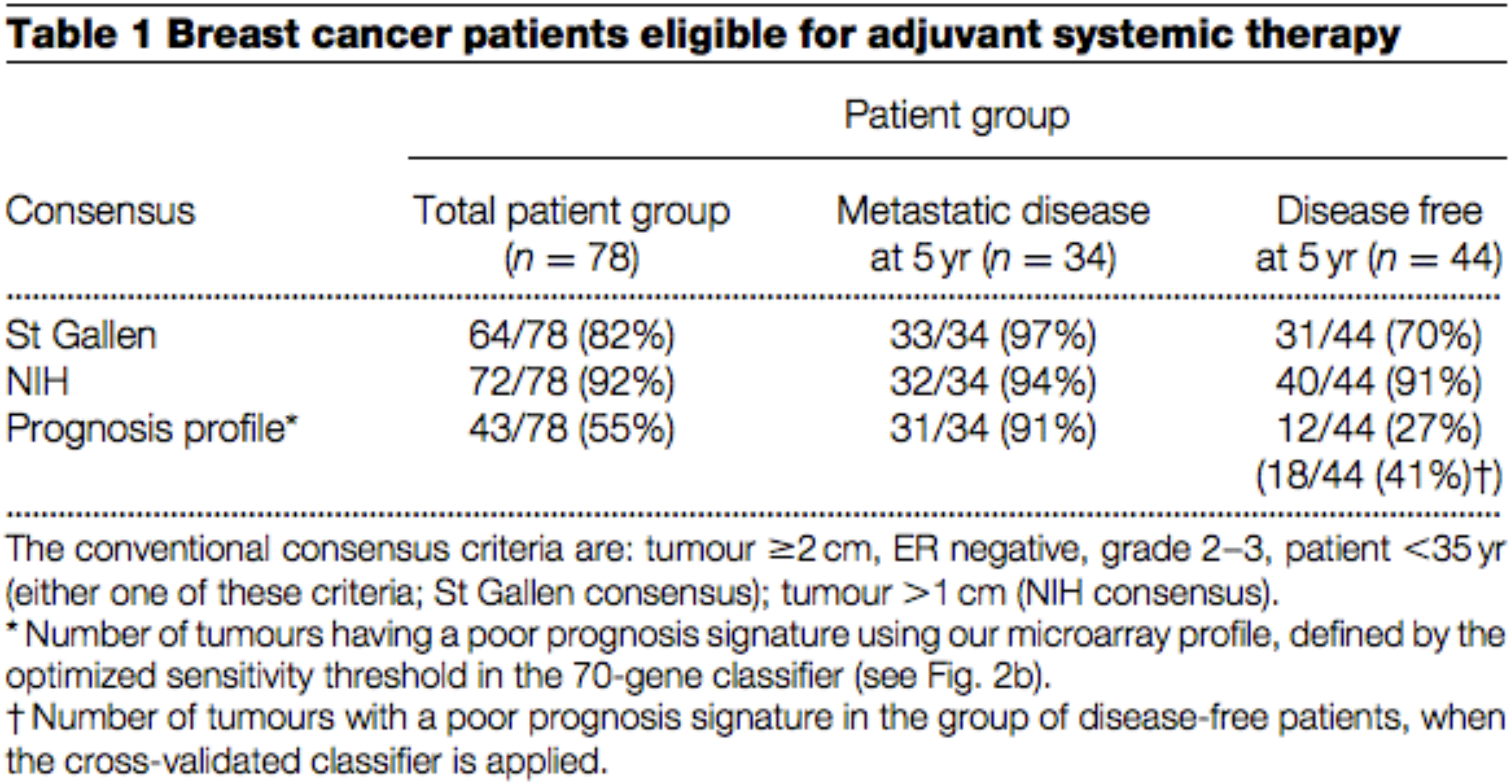


Table 1 Breast cancer patients eligible for adjuvant systemic therapy

	Patient group		
Consensus	Total patient group (*n* = 78)	Metastatic disease at 5 yr (*n* = 34)	Disease free at 5 yr (*n* = 44)
St Gallen	64/78 (82%)	33/34 (97%)	31/44 (70%)
NIH	72/78 (92%)	32/34 (94%)	40/44 (91%)
Prognosis profile*	43/78 (55%)	31/34 (91%)	12/44 (27%) (18/44 (41%)†)

The conventional consensus criteria are: tumour ≥2 cm, ER negative, grade 2–3, patient <35 yr (either one of these criteria; St Gallen consensus); tumour >1 cm (NIH consensus).

* Number of tumours having a poor prognosis signature using our microarray profile, defined by the optimized sensitivity threshold in the 70-gene classifier (see Fig. 2b).

† Number of tumours with a poor prognosis signature in the group of disease-free patients, when the cross-validated classifier is applied.

Result Overview

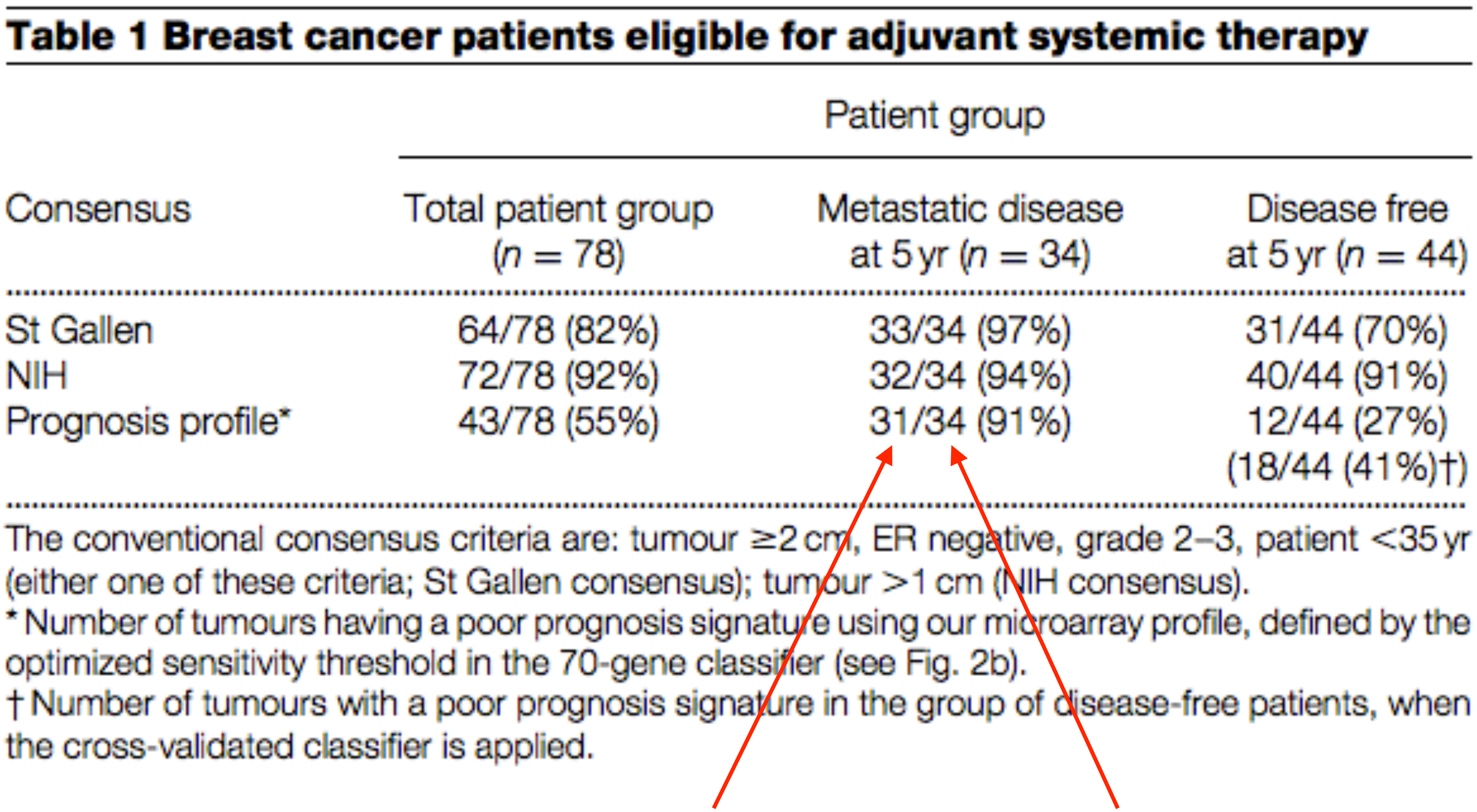


Table 1 Breast cancer patients eligible for adjuvant systemic therapy

Consensus	Patient group Total patient group ($n = 78$)	 Metastatic disease at 5 yr ($n = 34$)	 Disease free at 5 yr ($n = 44$)
St Gallen	64/78 (82%)	33/34 (97%)	31/44 (70%)
NIH	72/78 (92%)	32/34 (94%)	40/44 (91%)
Prognosis profile*	43/78 (55%)	31/34 (91%)	12/44 (27%) (18/44 (41%)†)

The conventional consensus criteria are: tumour ≥2 cm, ER negative, grade 2–3, patient <35 yr (either one of these criteria; St Gallen consensus); tumour >1 cm (NIH consensus).

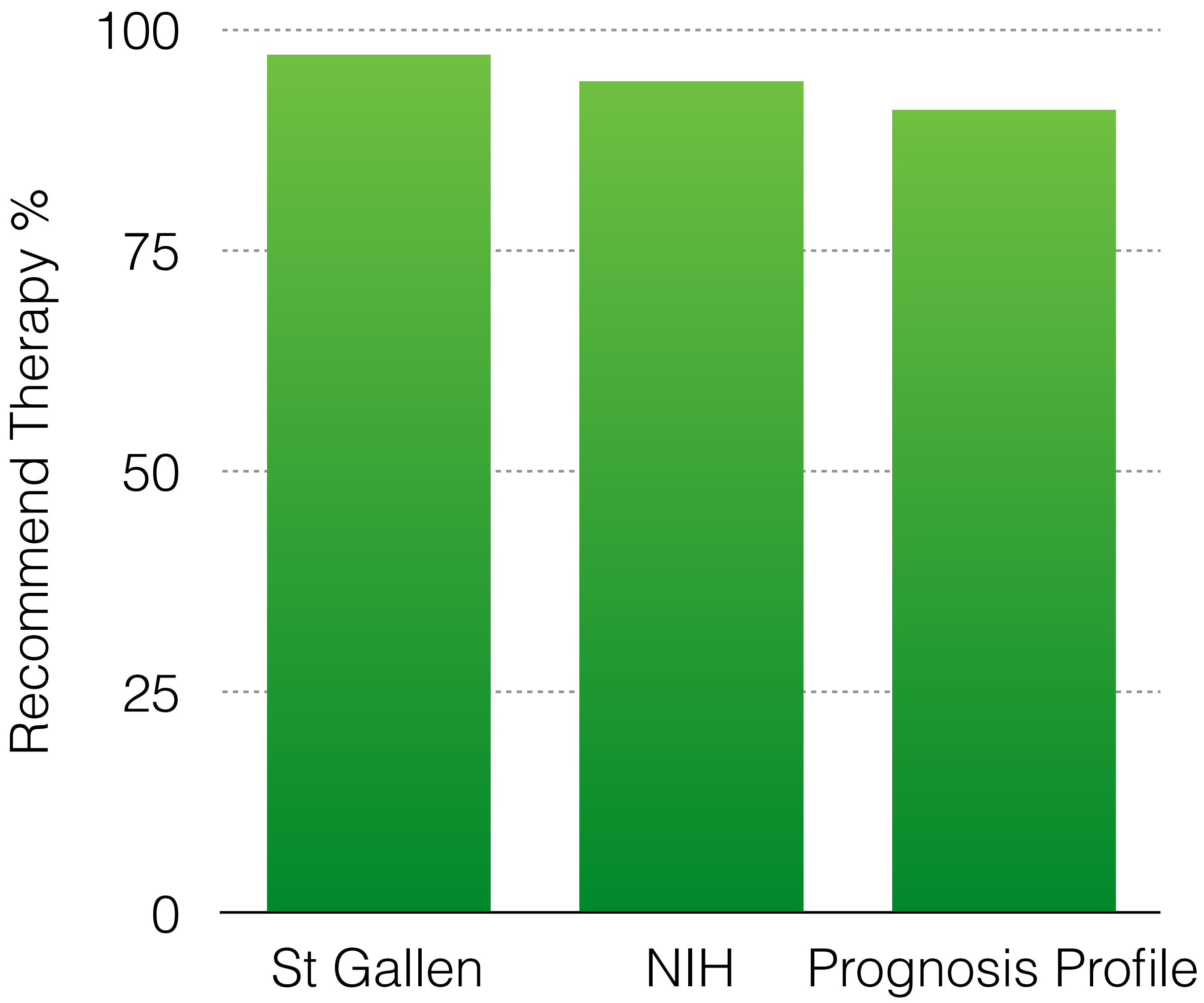
* Number of tumours having a poor prognosis signature using our microarray profile, defined by the optimized sensitivity threshold in the 70-gene classifier (see Fig. 2b).

† Number of tumours with a poor prognosis signature in the group of disease-free patients, when the cross-validated classifier is applied.

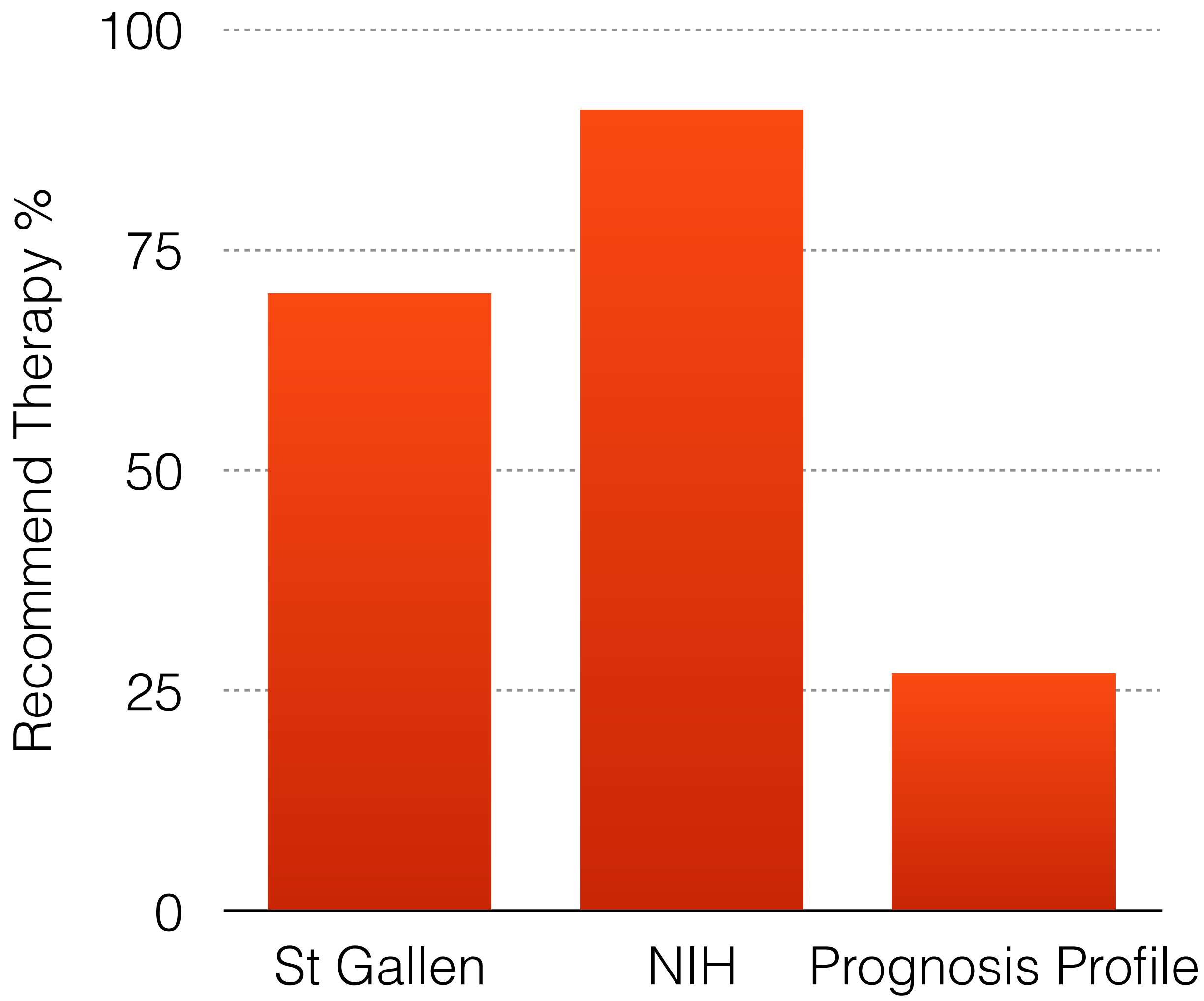
gets therapy

needs therapy

Result Overview



needs therapy



does not need therapy

Study

1. unsupervised clustering, look for tumor categories
2. supervised learning, find prognosis reporter genes

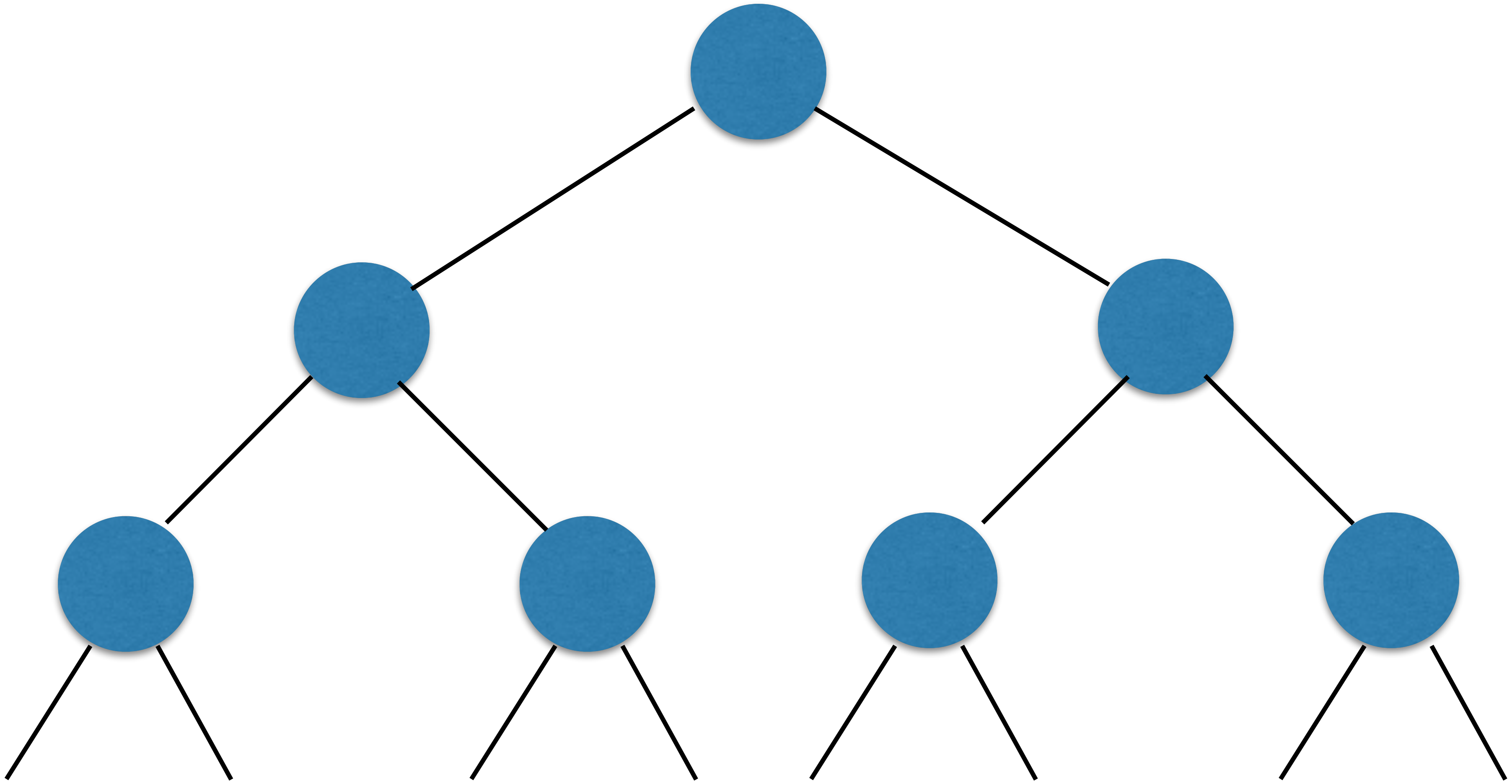
Study

1. unsupervised clustering, look for tumor categories

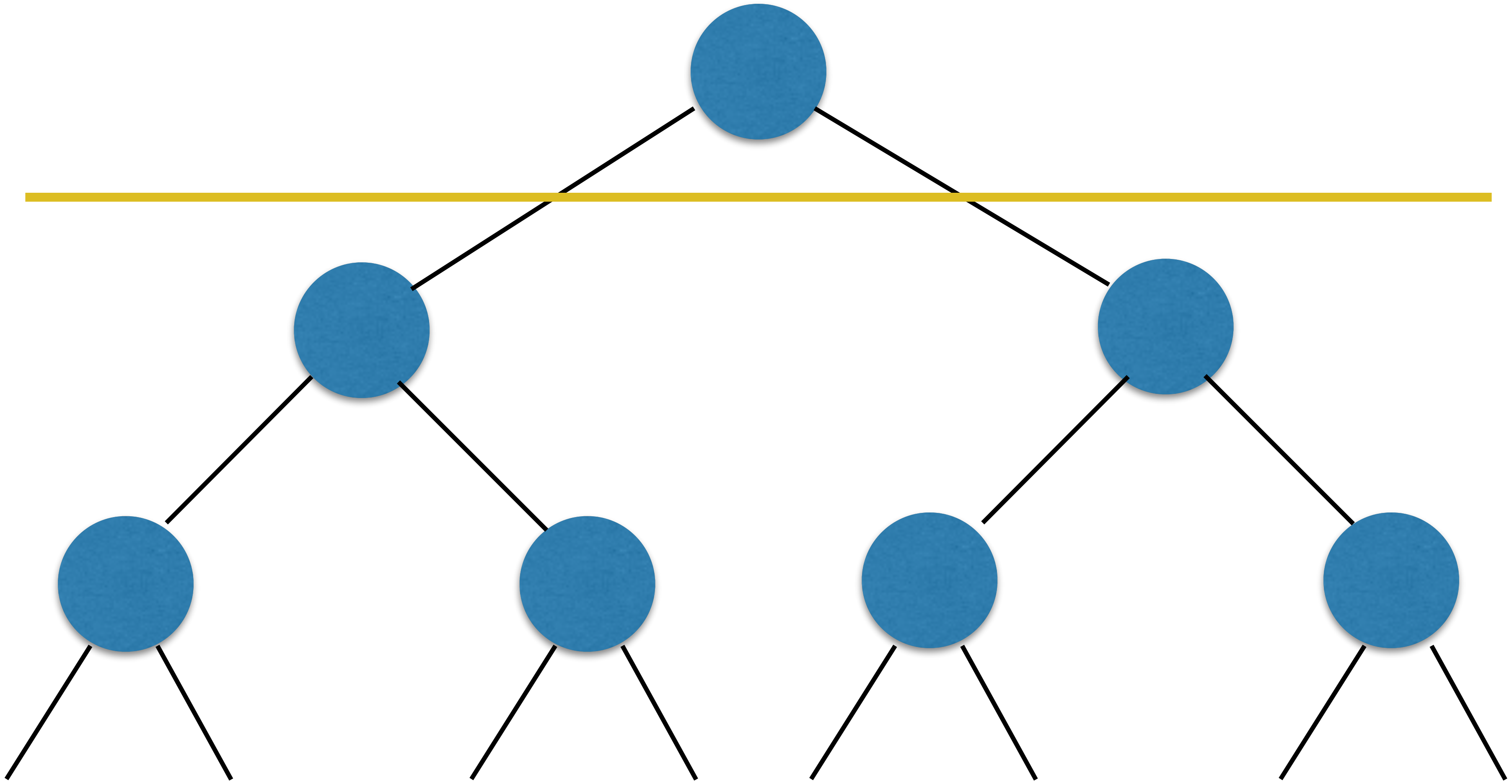
2. supervised learning, find prognosis reporter genes

Unsupervised Hierarchical Clustering: Dendogram

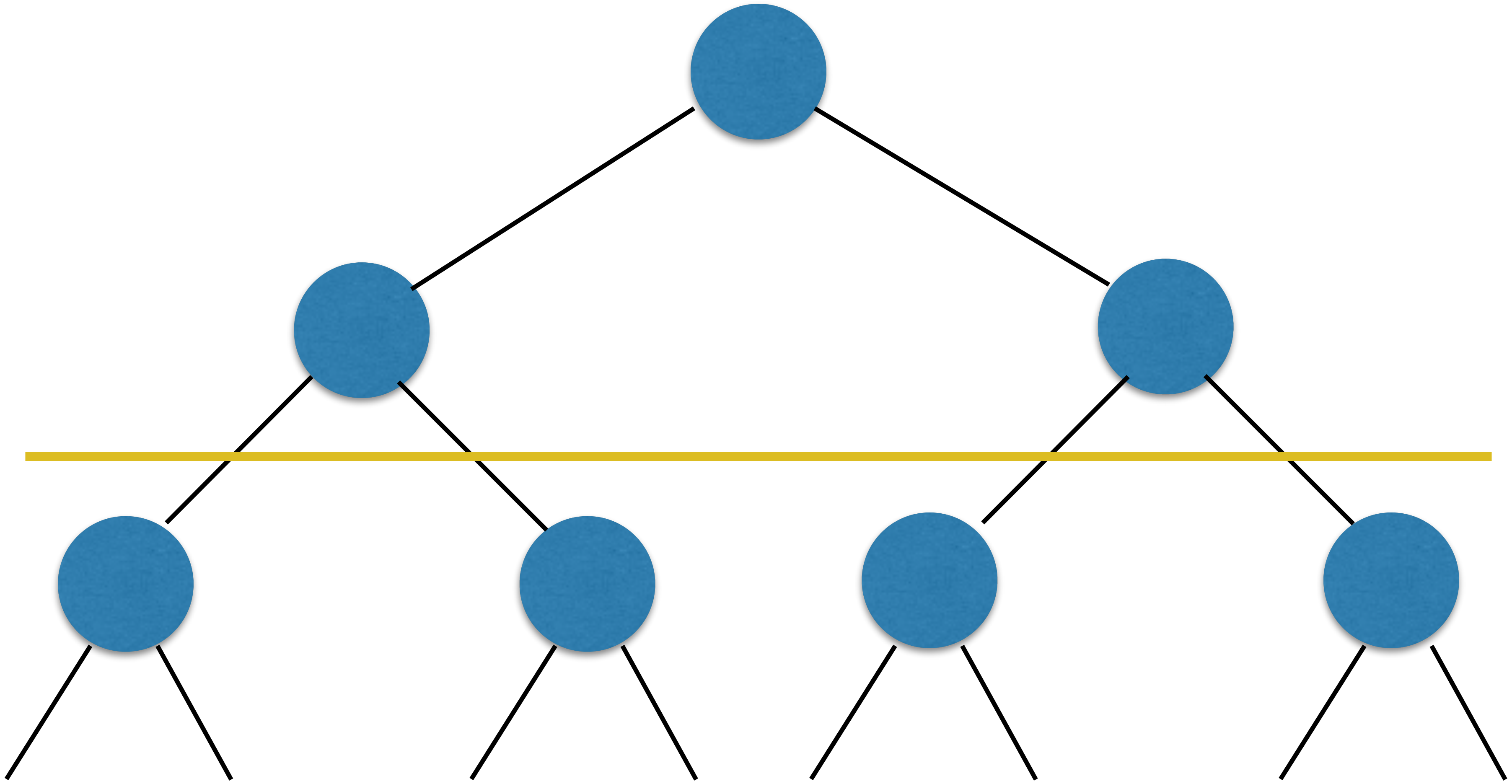
http://youtu.be/XJ3194AmH40?t=5m



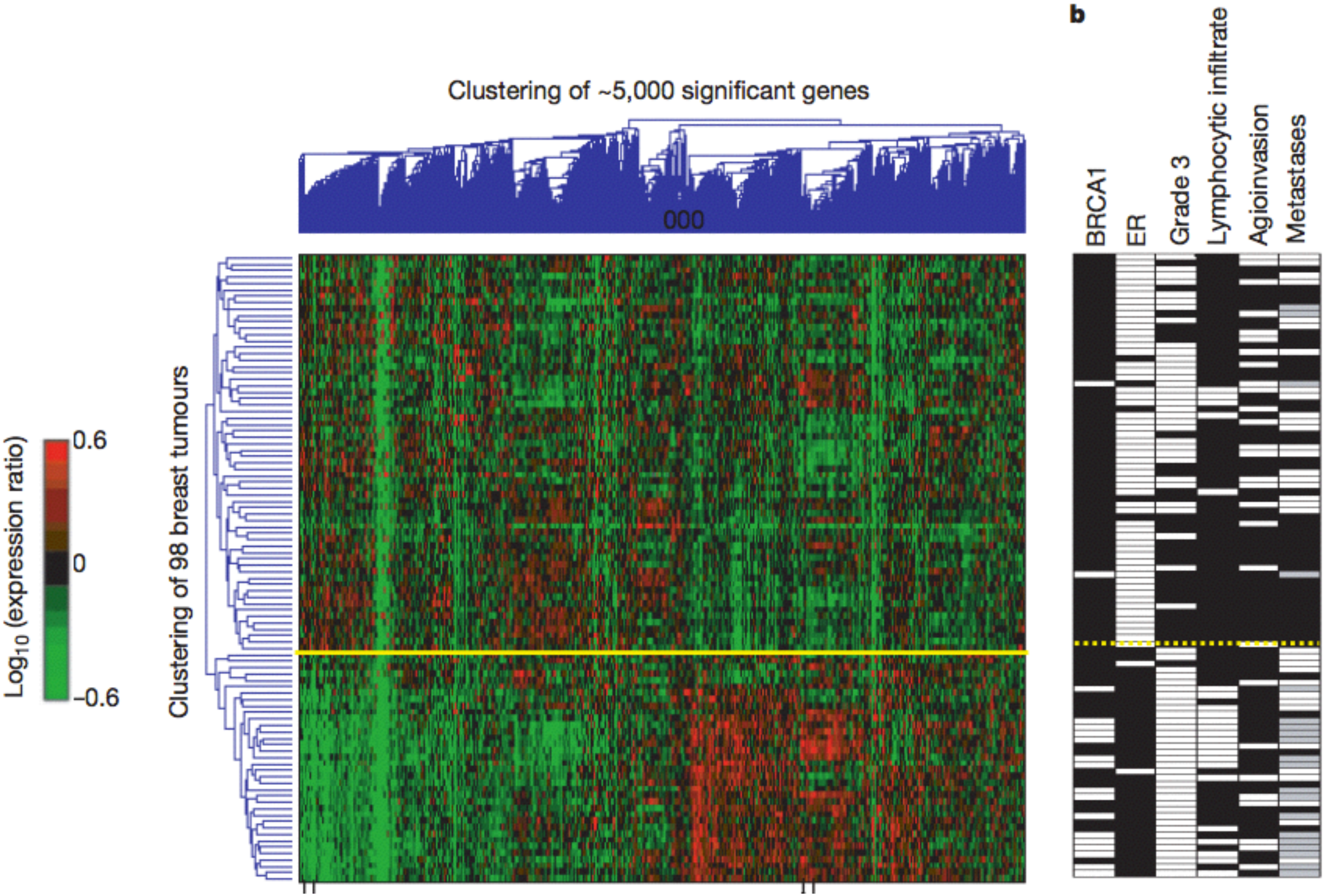
2 categories



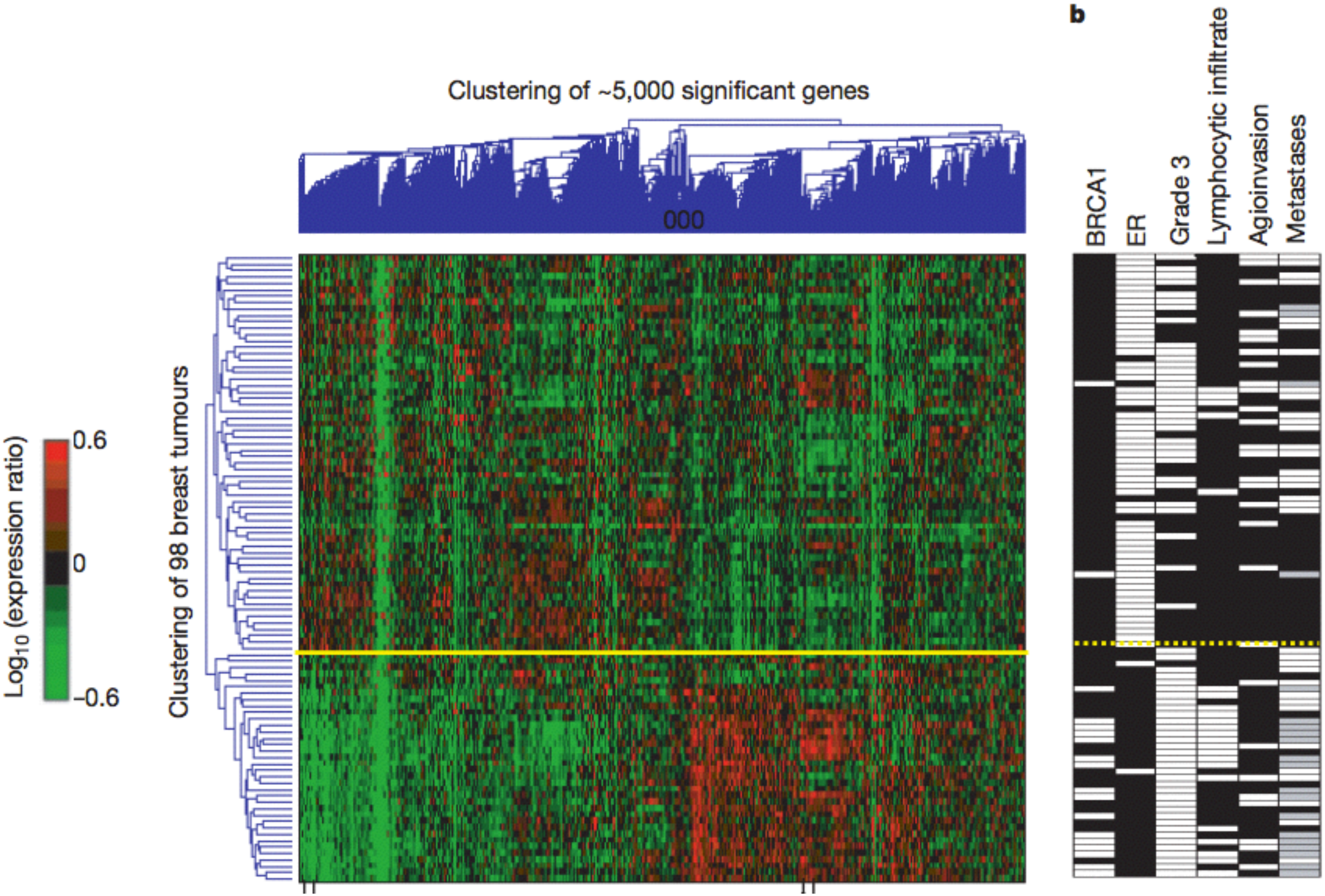
4 categories



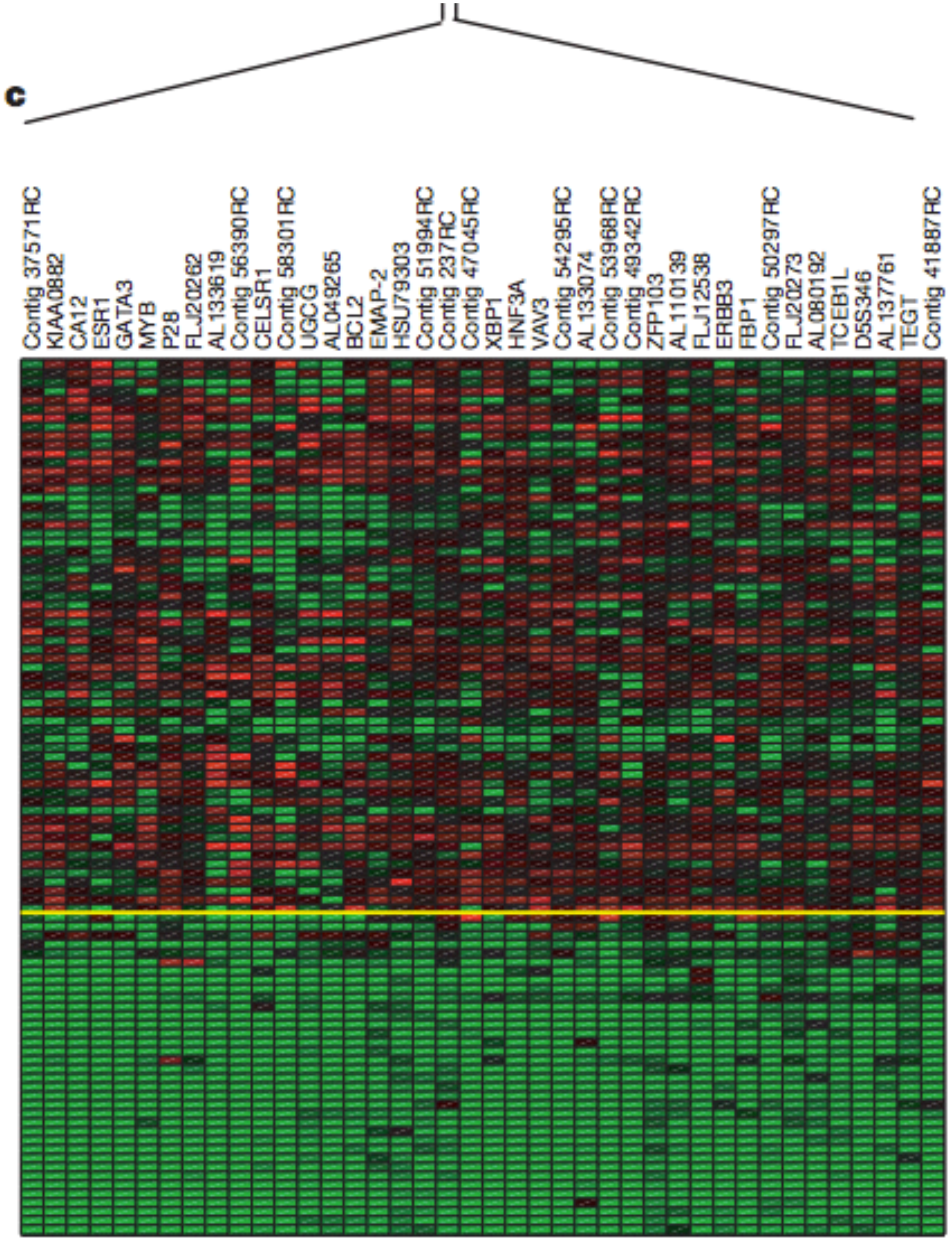
page 531


Clustering of ~5,000 significant genes
000
Log$_{10}$ (expression ratio)
0.6
0
–0.6
Clustering of 98 breast tumours
b
BRCA1
ER
Grade 3
Lymphocytic infiltrate
Agioinvasion
Metastases

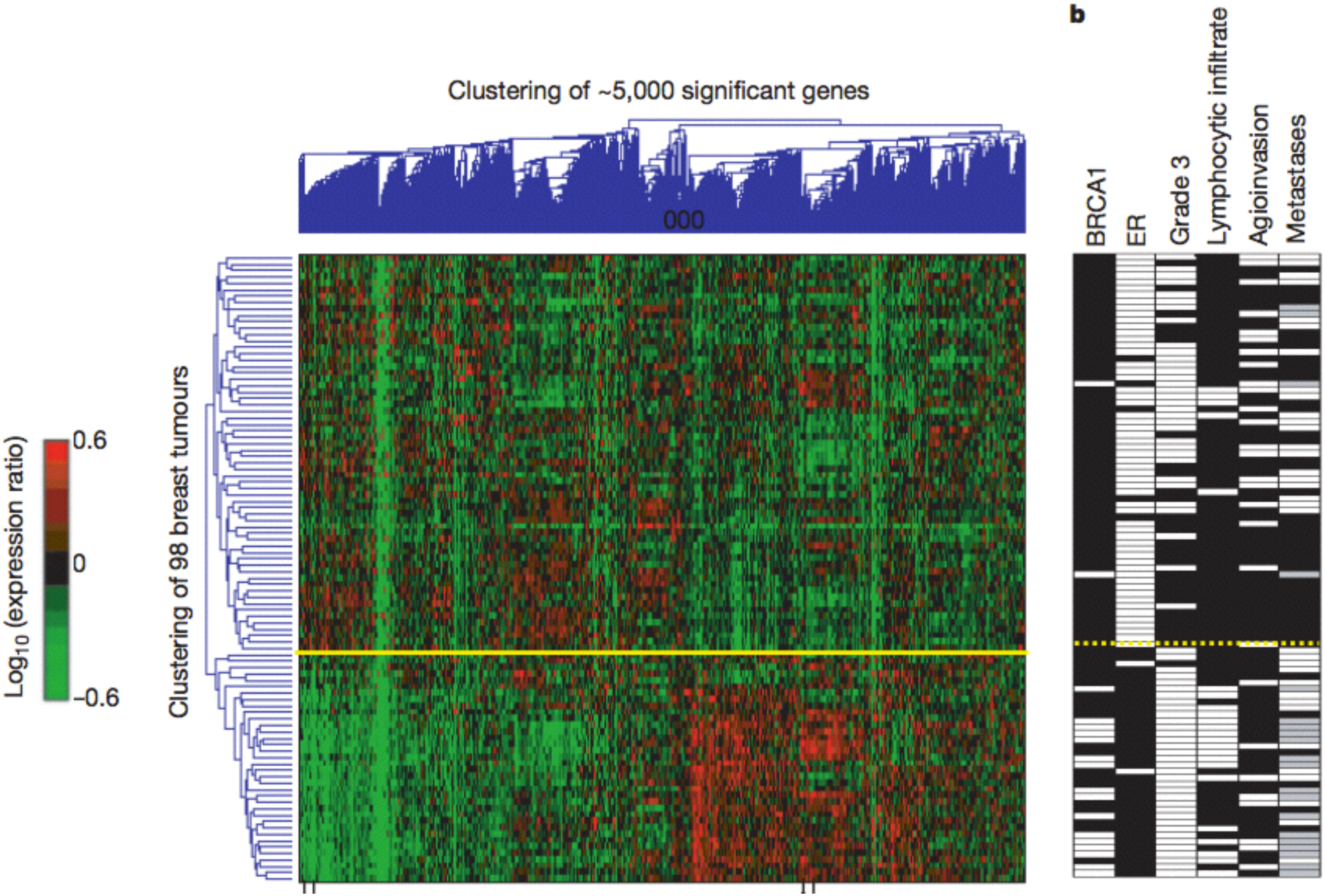
page 531



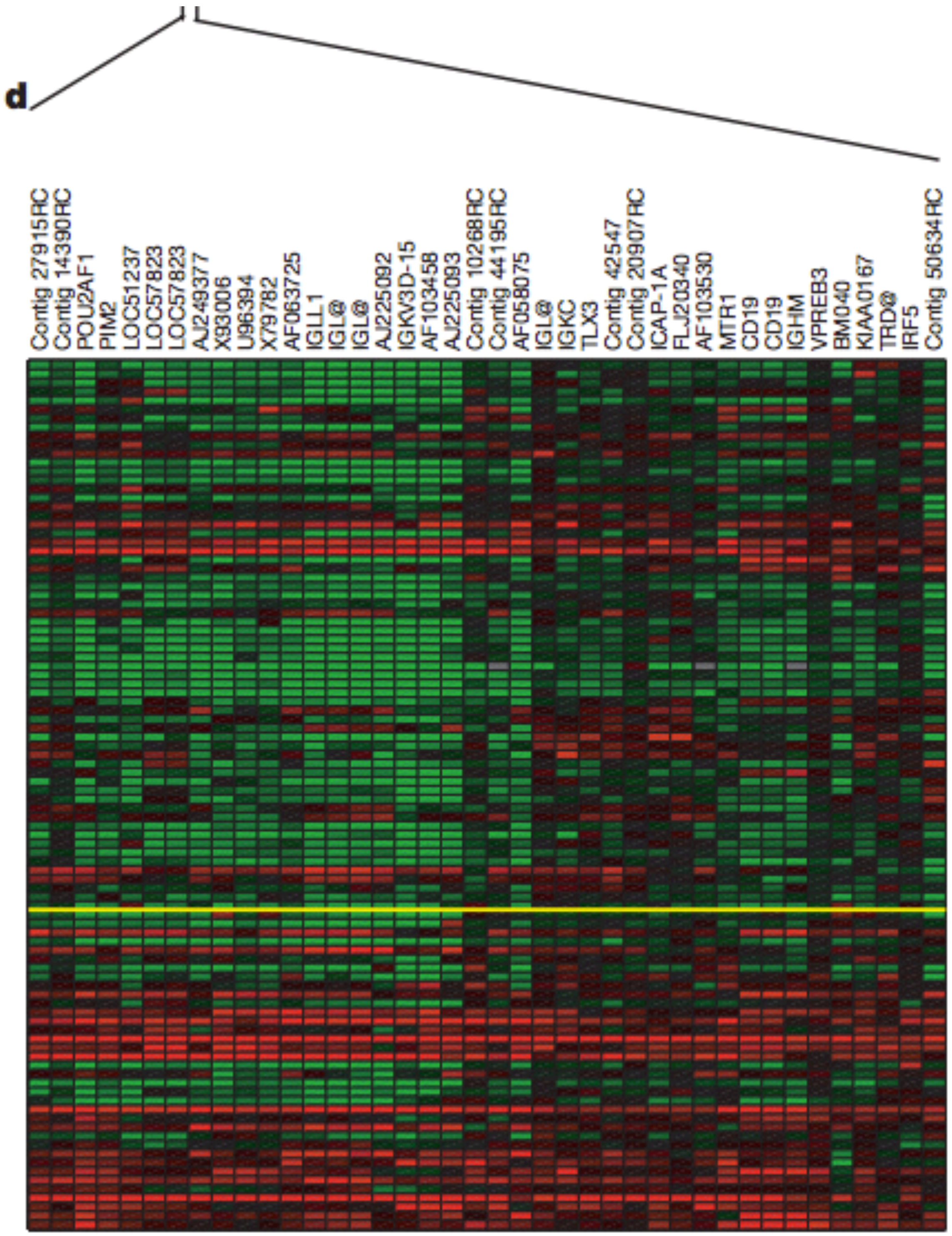
co-regulates with ER-α


c
Contig 37571RC
KIAA0882
CA12
ESR1
GATA3
MYB
P28
FLJ20262
AL133619
Contig 56390RC
CELSR1
Contig 58301RC
UGCG
AL049265
BCL2
EMAP-2
HSU79303
Contig 51994RC
Contig 237RC
Contig 47045RC
XBP1
HNF3A
VAV3
Contig 54295RC
AL133074
Contig 53968RC
Contig 49342RC
ZFP103
AL110139
FLJ12538
ERBB3
FBP1
Contig 50297RC
FLJ20273
AL080192
TCEB1L
D5S346
AL137761
TEGT
Contig 41887RC

page 531



co-regulates with
lymphocytic infiltrate


d
Contig 27915RC
Contig 14390RC
POU2AF1
PIM2
LOC51237
LOC57823
LOC57823
AJ249377
X93006
U96394
X79782
AF063725
IGLL1
IGL@
IGL@
AJ225092
IGKV3D-15
AF103458
AJ225093
Contig 10268RC
Contig 44195RC
AF058075
IGL@
IGKC
TLX3
Contig 42547
Contig 20907RC
ICAP-1A
FLJ20340
AF103530
MTR1
CD19
CD19
IGHM
VPREB3
BM040
KIAA0167
TRD@
IRF5
Contig 50634RC

Study

1. unsupervised clustering, look for tumor categories
2. supervised learning, find prognosis reporter genes

Study

1. unsupervised clustering, look for tumor categories
2. <u>supervised learning, find prognosis reporter genes</u>

Method

Start with **25K** genes (some double counting)

From these, identify **5K** that are significantly regulated

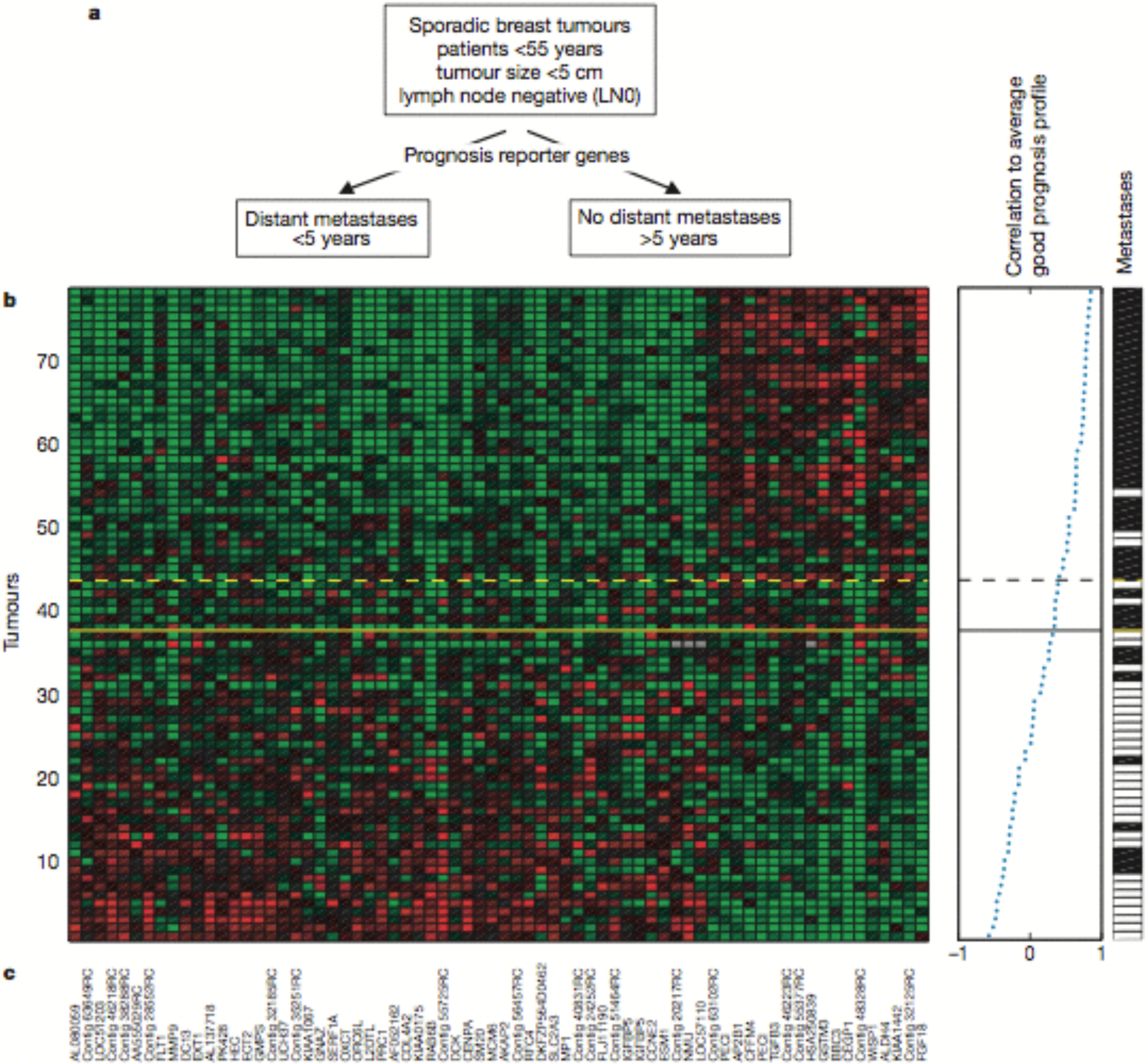
From these, identify **231** significantly associated with disease outcome

From these, identify **70** as classification features

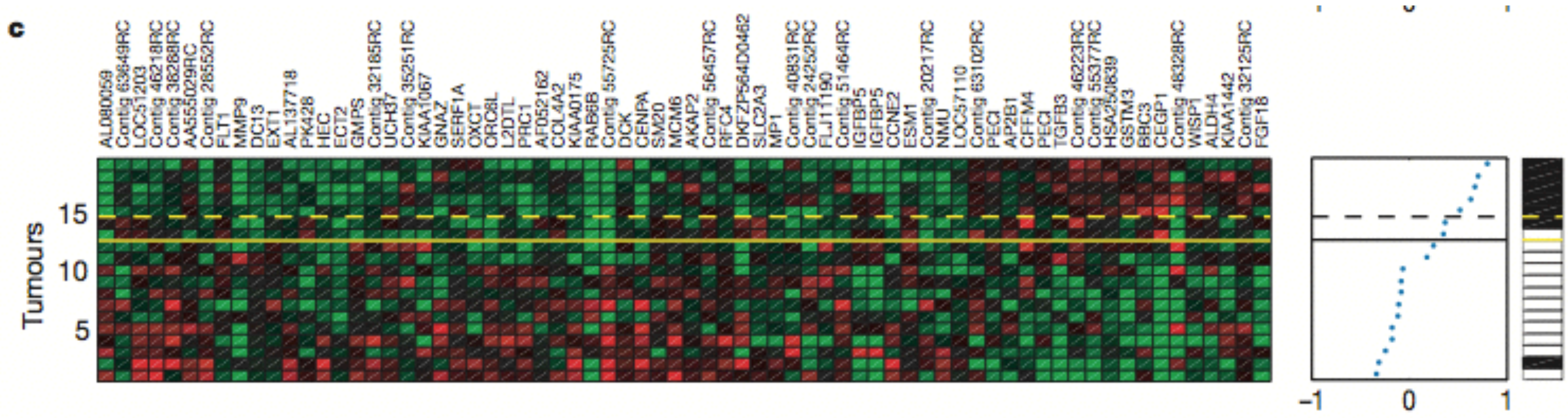
Supervised Learning on Prognosis Signatures

page 532

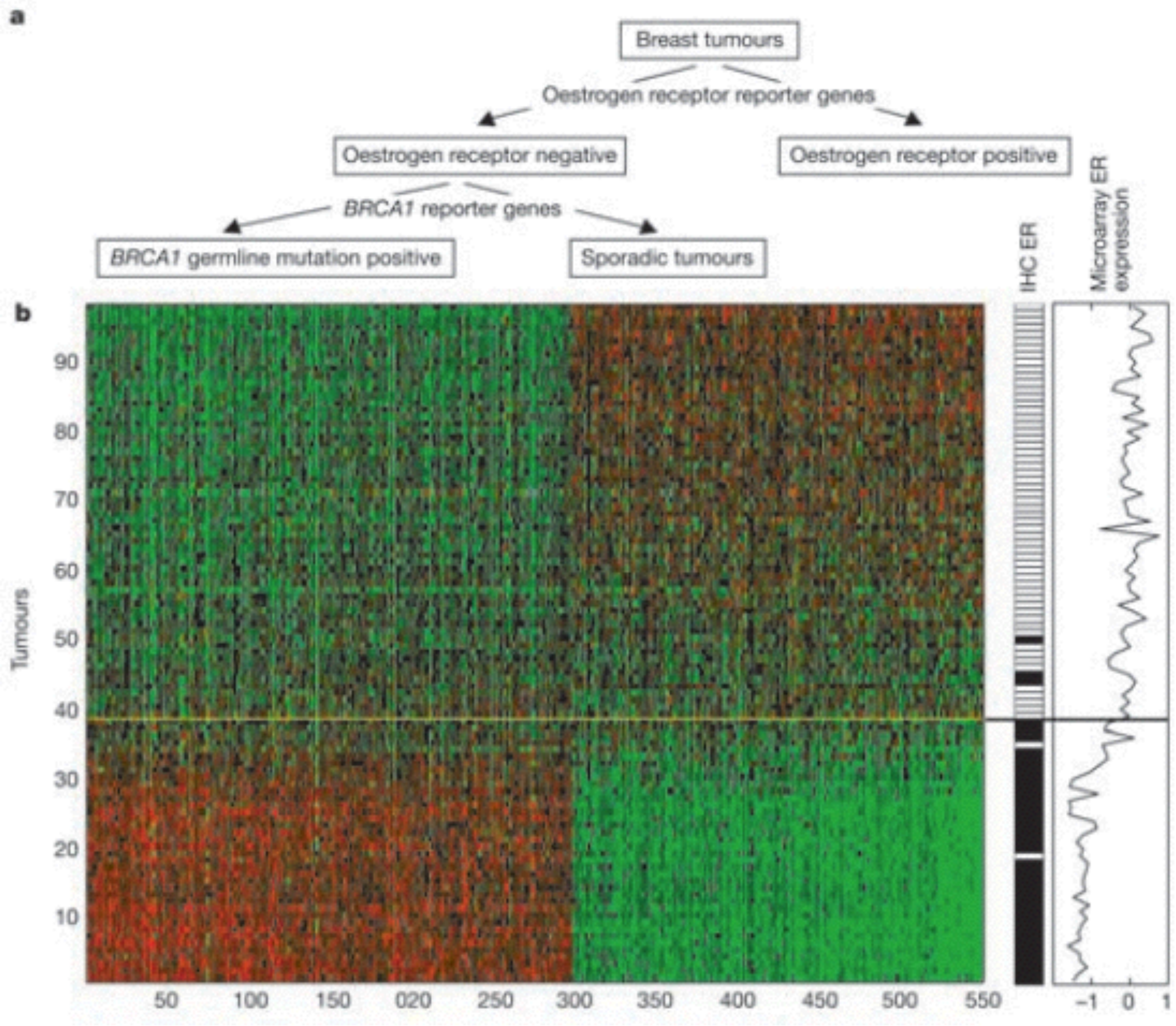
Training Data (78 tumors)



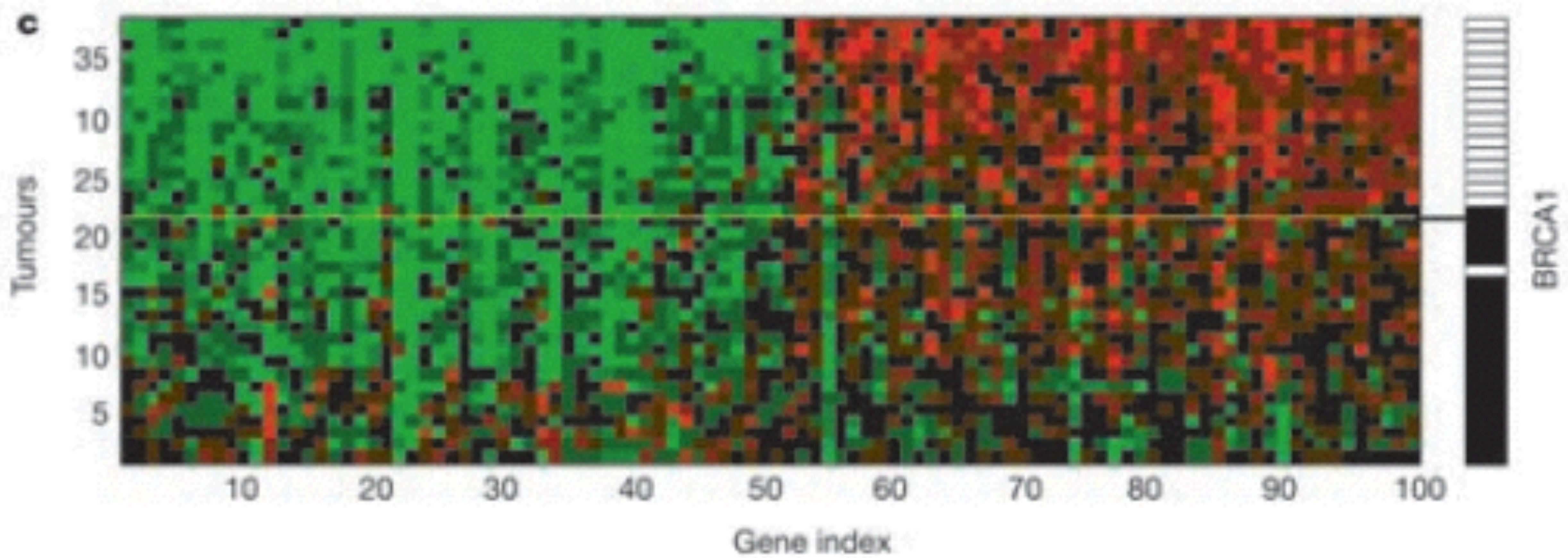
Test Data (19 tumors)



Supervised Learning on ER and BRCA1 Signatures


a
Breast tumours
Oestrogen receptor reporter genes
Oestrogen receptor negative
Oestrogen receptor positive
BRCA1 reporter genes
BRCA1 germline mutation positive
Sporadic tumours
b
Tumours
90
80
70
60
50
40
30
20
10
50
100
150
020
250
300
350
400
450
500
550
IHC ER
Microarray ER expression
−1
0
1

page 533


c
35
10
25
20
15
10
5
Tumours
10
20
30
40
50
60
70
80
90
100
Gene index
BRCA1

Summary

New ideas (for 826):

- new label: clinical outcome
- use of unsupervised learning
- accuracy vs. sensitivity tradeoffs