Comprehensive analysis integrating both clinico-pathological and gene expression data in more than 1500 samples: Proliferation captured by gene expression grade index appears to be the strongest prognostic factor in breast cancer

Christos Sotiriou, Pratyaksha Wirapati, Sherene Loi, Benjamin Haibe-Kains, Christine Desmedt, Andrew M. Tutt, Paul Ellis, Marc Buyse, Gianluca Bontempi, Mauro Delorenzi and Martine Piccart

Background

•Prognosis of breast cancer has been extensively studied using several clinico-pathological parameters and tumor markers reflecting different stages of disease and breast tumor biology

•Although the development of high-throughput gene expression technologies has allowed the identification of several "molecular signatures" predicting clinical outcome, no attempt has been made yet to perform a comprehensive analysis integrating both clinico-pathological and gene expression data

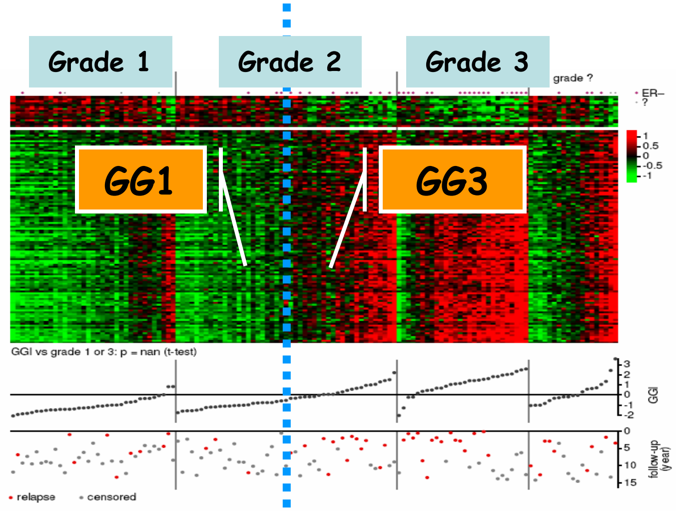
Defining Genomic Grade in Breast Carcinoma

128 probe sets of "genomic grade" (97 genes) FDC>0

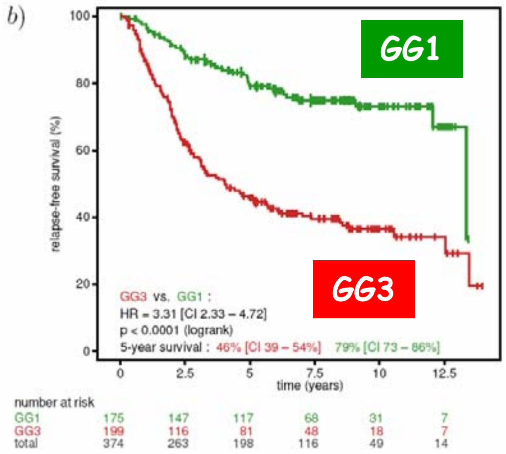
UBE2C RACGAP1 C10orf3 KPNA2
PTTG1 KIF4A TPX2 FOXM1 KIF20A
STK6 STK6 DLG7 DDX39 MELK CCNA2
MYBL2 KIAA0186 BIRC5 NUDT1 KPNA2
KIF2C KIFC1 SPAG5 ASPM CDC20 FEN1
TIMELESS ESPL1 CENPA MCM2
DONSON CDC2 CCNB1 CDCA8 KIF11
DKFZp762E1312 MCM10 CDKN3 MARS
CENPA CCNB2 TRIP13 LMNB1 CDC2
TROAP AURKB FLJ20641 BUB1B
CENPE CCNE2 CDC2 FSHPRH1 BRRN1
HMMR POLQ PMSCL1 MKI67 GTSE1
ZWINT GMPS TMPO RRM2 KLIP1 FEN1
MKI67 KIF2C PLK1 BLM BUB1
LOC146909 OIP5 K-ALPHA-1 SHMT2
DC13 H2AFZ MCM4 UBE2S TUBA6 TTK
FLJ10156 C20orf24 MARS RRM2 MKI67
CENPF PRC1 BM039 K-ALPHA-1
CDC25A NUSAP1 KNTC2 EXO1 MCM4
BIRC5 MAD2L1 UBE2N MGC5528 CDK2
ESPL1 HCAP-G CCT5 SLC7A5 CDCA3
ORMDL2 KIF14 PTDSR K-ALPHA-1
BIRC5 RNASEH2A HIST1H4B HMGB3
NEK2 KNSL7 SNRPC MKI67 EZH2
DNAJC9 DC12 TPRT COX7B MRPS17
SIL FBXO5 HCAP-G HN1 POLR2K
NUTF2 MCM6 MCM4 VRK1 PKMYT1
RAD51 …

$$GGI = scale\left[\sum_{j\in G_3} x_j - \sum_{j\in G_1} x_j - cutoff\right]$$

Define GGI score (Gene-expression Grade Index)



Clinical Outcome



(J Natl Cancer Inst., 2006)

Aims

- To confirm our results in larger datasets

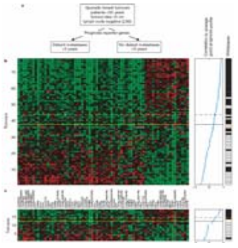
- To define additional gene expression indices related to breast cancer biology

- To elucidate the relationship between traditional clinico-pathological markers, gene expression patterns and their interaction with prognosis

Methods : Defining Molecular Indices

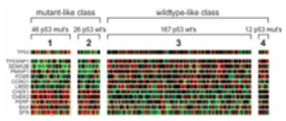
2 Published datasets

Van de Vijver *et al.*
NEJM, 2002



Agilent
N=295

Miller *et al.*
PNAS, 2005



Affymetrix
N=251

gene X_1
.
.
.
.
gene X_{40000}

Prototype genes

ESR1 = ER related
+
ERBB2 = Her2-neu
+
STK6 = proliferation/ GGI
+
PLAU = Stroma/invasion
+
STAT1 = immune response
+
VEGF = angiogenesis

A multivariate linear regression model is fitted to estimate the contribution of each prototype for the prediction of the expression of each gene on the arrays

Gene Expression Molecular Indices

$$INDEX = \sum_{j \in P} x_j - \sum_{j \in N} x_j$$

Each gene expression index is defined by the difference of the sums of the positively and negatively correlated genes for the chosen prototype

⇩

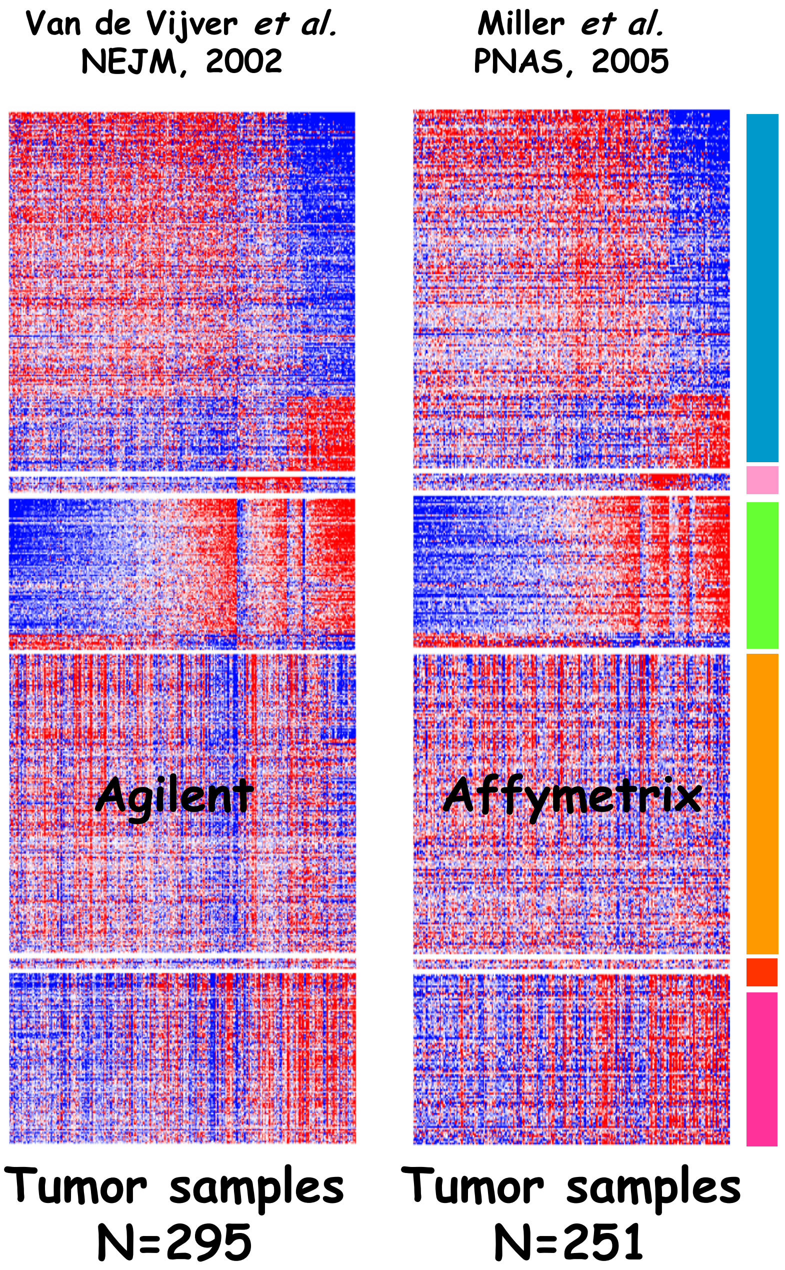
Mapped and computed on SEVERAL published microarray datasets
(Sorlie et al. PNAS 2003; Sotiriou et al PNAS 2003; van De Vijver NEJM 2002; Wang et al The Lancet 2005; Foekens et al JCO 2006; Miller et al PNAS 2005; Sotiriou et al JNCI 2006; Buyse et al JNCI in press; N=8)

N ~ 1500 tumor samples

Three main characteristics of the molecular indices

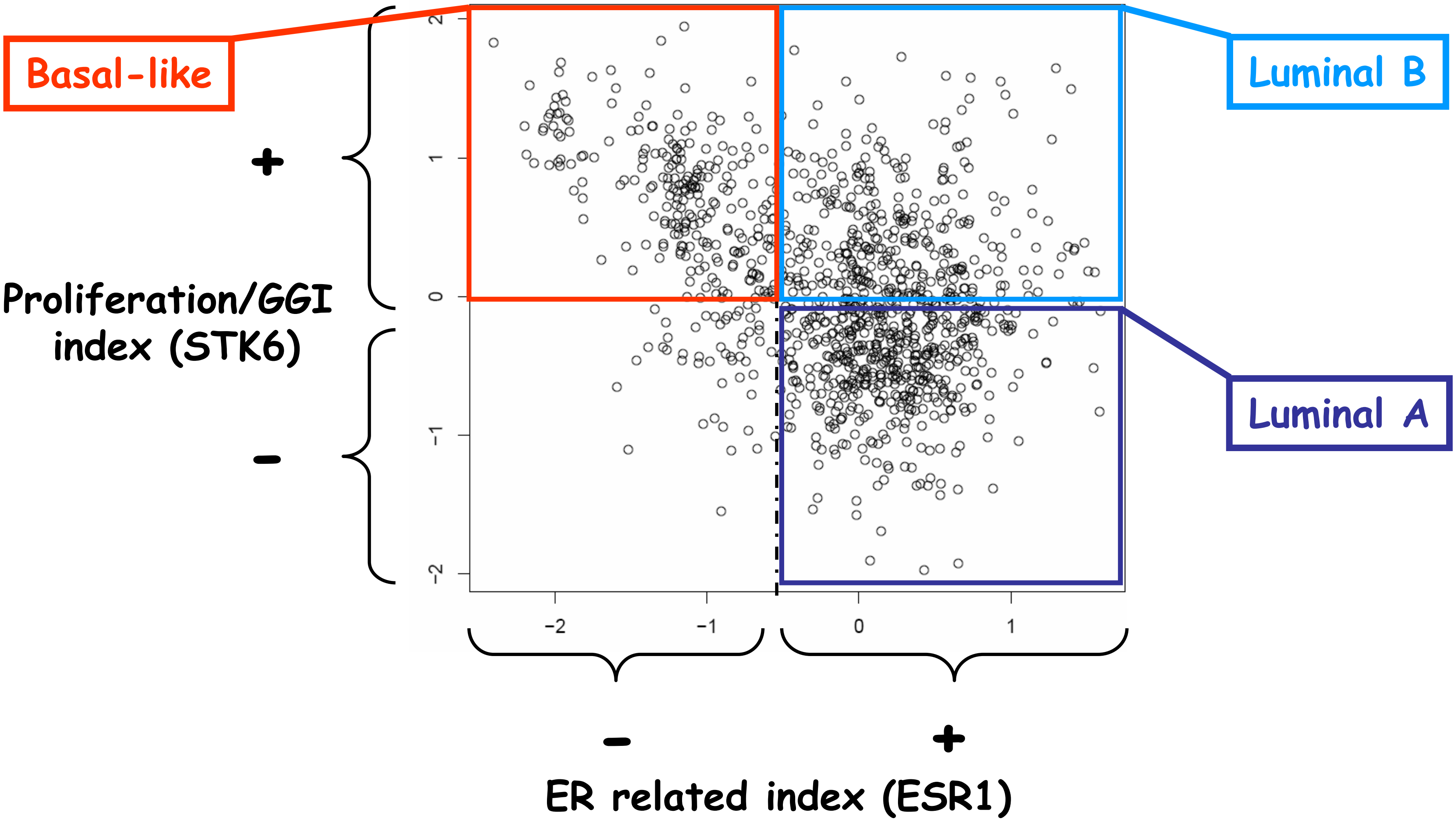
- They are identified as genes that are co-expressed consistently with the prototypes in Agilent and Affymetrix data
- They are identified without looking at clinical variables
- They are identified without looking at gene annotation (except for the selected prototypes)

Results



	No genes
ER related (ESR1) Luminal: GATA3, XBP1, TFF1, TFF3, MYB, PR, BCL2... Basal: KRT5, KRT6B, KRT16, ATF4, FYN, S100A9 …	N=874
ERBB2: GRB7,TCAP,PERLD1... (17q11-22 amplicon)	N=68
Proliferation/GGI (STK6): TOP2A, MYBL2, CCNA2, CCNB1, CCNE2, CDC20…	N=664
Stroma/invasion (PLAU): PLAUR, SERPINE, COL5A2, MMP14, MMP12, DCN, ADAM12…	N=495
Angiogenesis (VEGF): CA9,PDX …	N=17
Immune response (STAT1): GZMA, HLA-A, IFIT1, CXCL11, CXCL9, MX1…	N=288

Relationship between the ESR1 and the proliferation index values



Basal-like

Luminal B

Luminal A

+

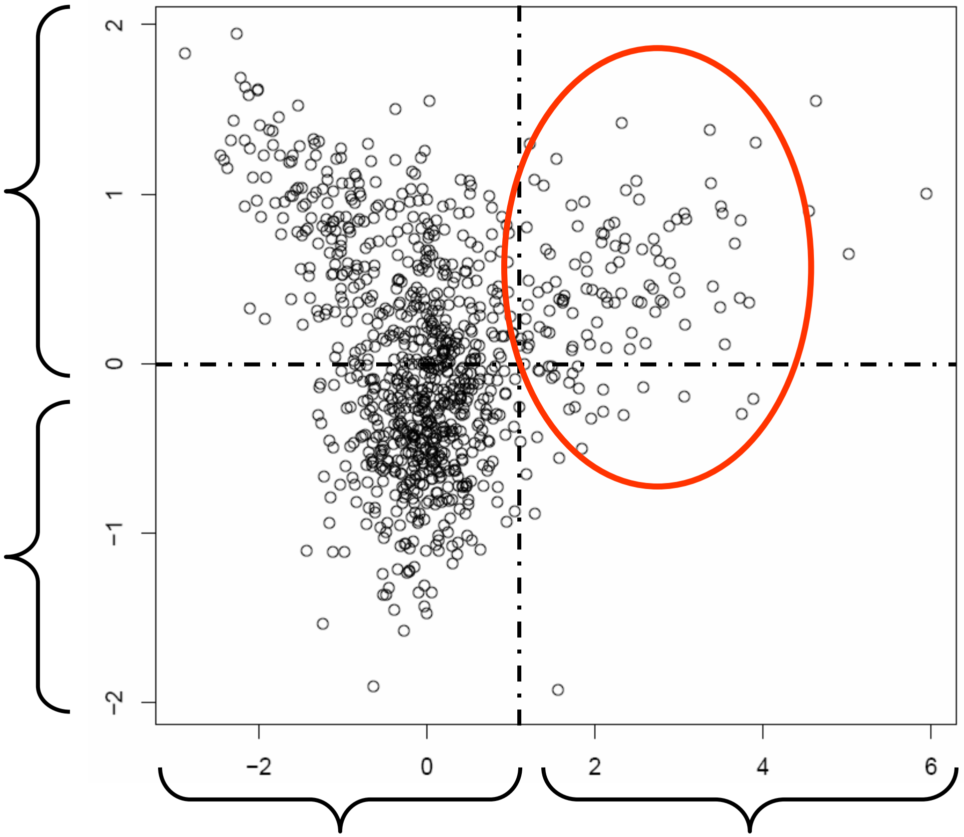
Proliferation/GGI index (STK6)

-

- +

ER related index (ESR1)

Relationship between the HER-2 and the proliferation index values



+

Proliferation/GGI Index (STK6)

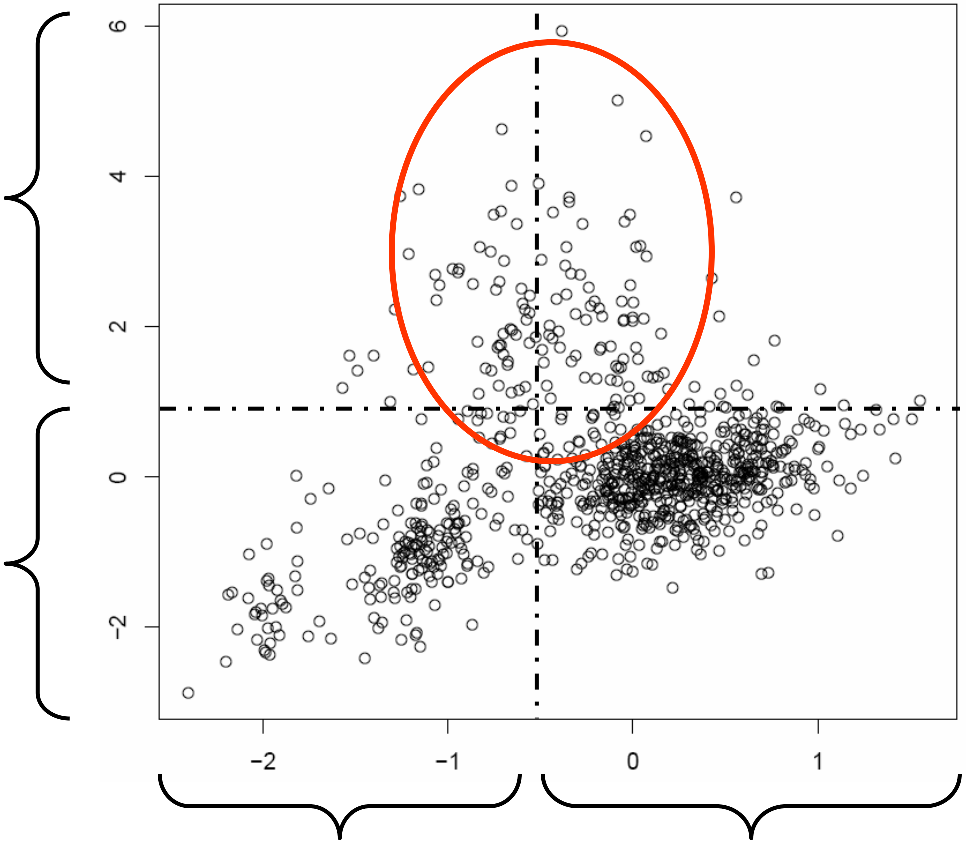
–

–

+

HER-2 index (ERBB2)

Relationship between expression levels of HER-2 and ESR1 indices



+

HER-2 index
(ERBB2)

−

− +

ER related index (ESR1)

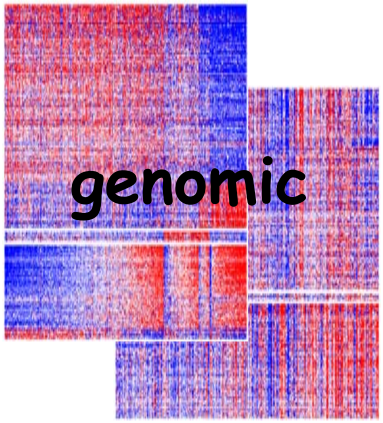
Association between clinical parameters and gene expression indices

	ER related (ESR1)	HER-2 (ERBB2)	Proliferation /GGI (STK6)	Stroma (PLAU)	Angiogenesis (VEGF)	Immune response (STAT1)
Age (<50 ys)	↓	-	-	-	-	-
Size (>2 cm)	-	-	↑	-	↑	-
Nodal Status (+)	-	-	↑	-	↑	-
Histological Grade (High)	↓	↑	↑	-	↑	↑
ER Status (-)	↓	↑	↑	-	↑	↑

Clinical Relevance?

Clinico-pathological

&



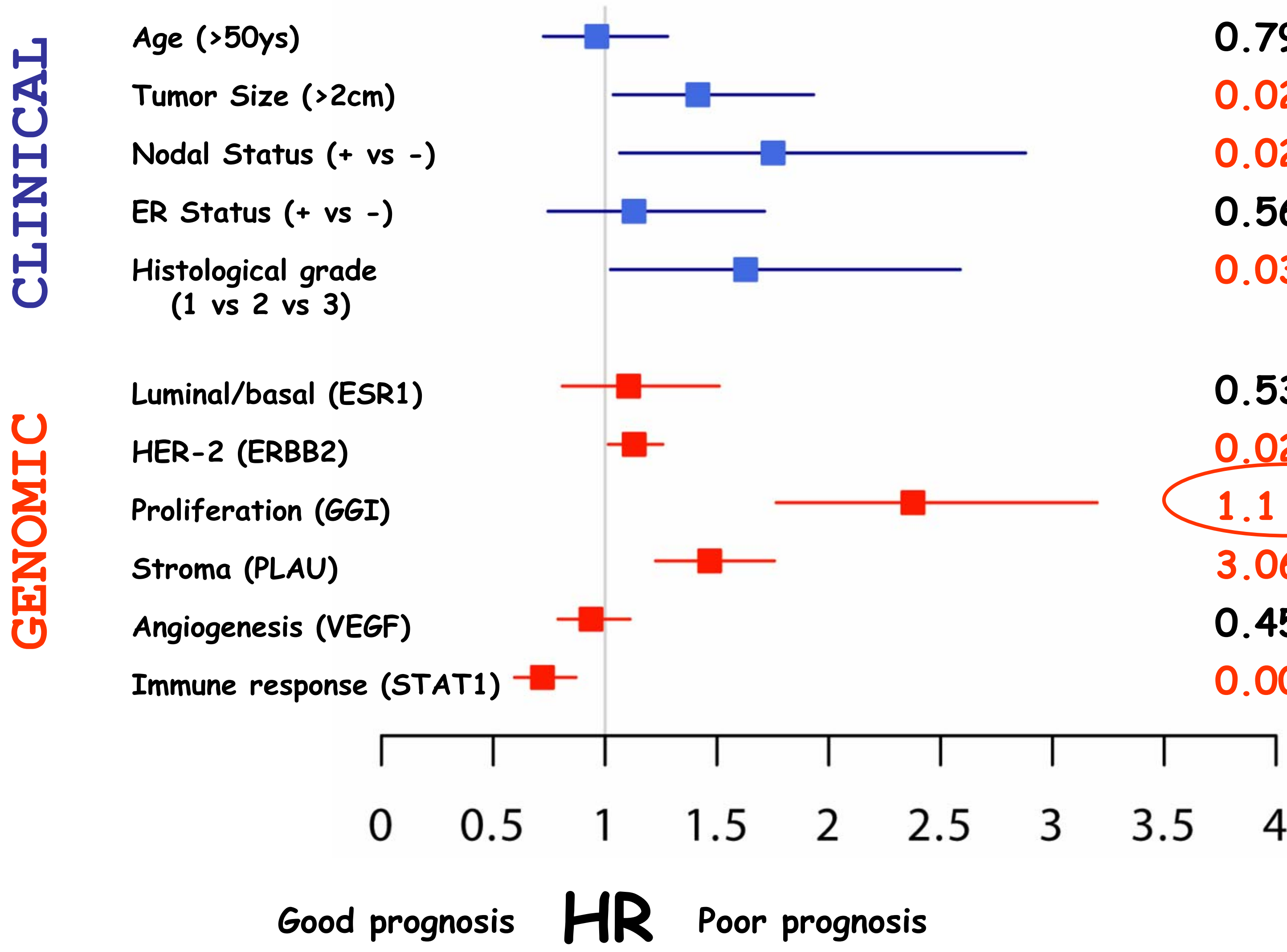
genomic

Datasets

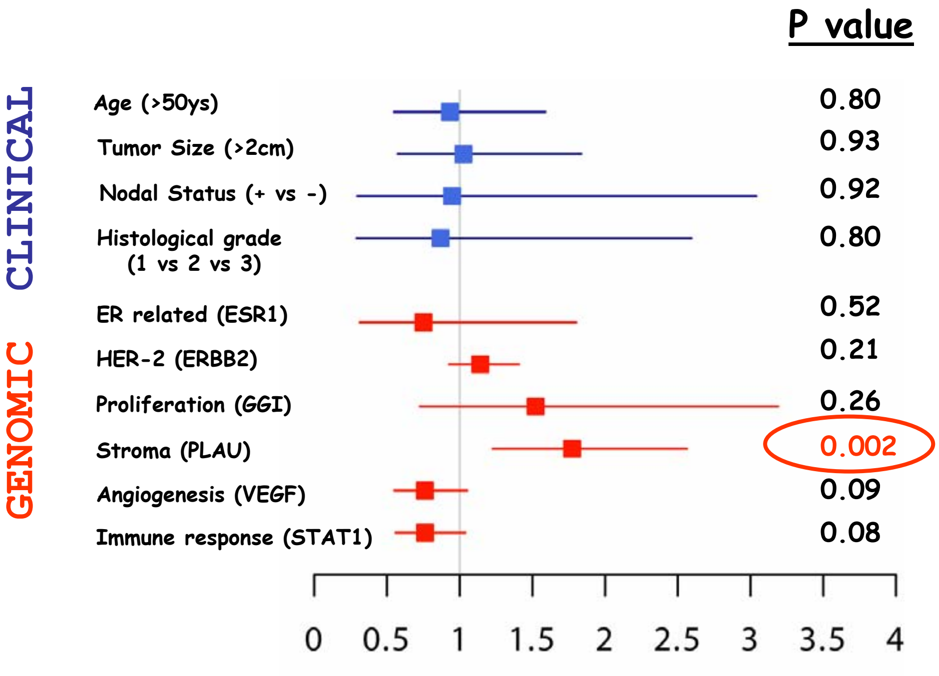
UNTREATED
POPULATION

N~1100

Multivariate analysis
RFS (10 years)

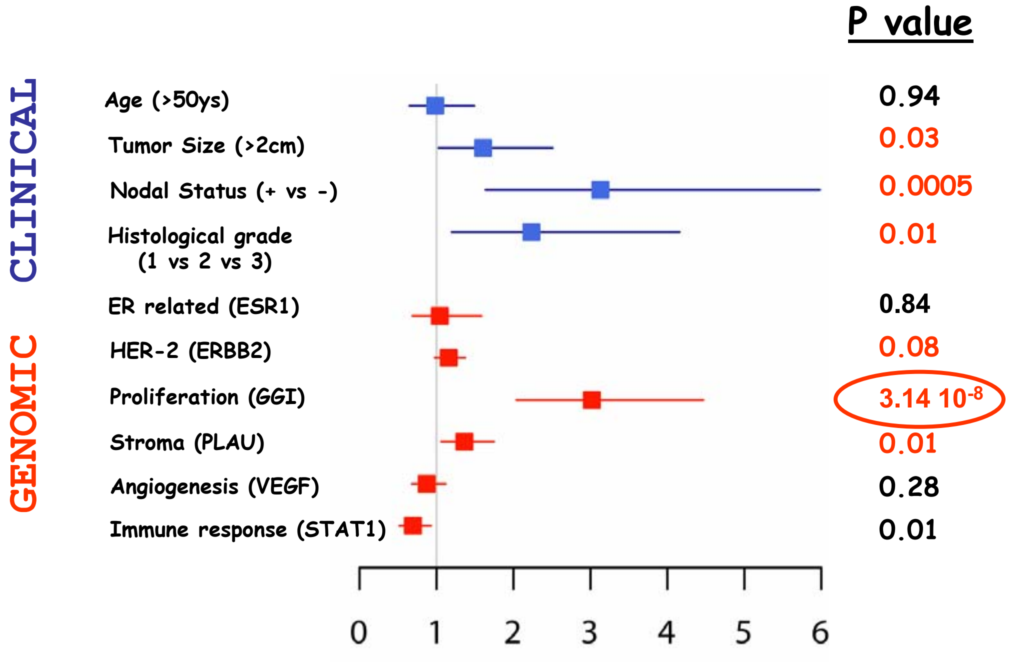


	P value
Age (>50ys)	0.79
Tumor Size (>2cm)	0.02
Nodal Status (+ vs -)	0.02
ER Status (+ vs -)	0.56
Histological grade (1 vs 2 vs 3)	0.03
Luminal/basal (ESR1)	0.53
HER-2 (ERBB2)	0.02
Proliferation (GGI)	$1.1\ 10^{-8}$
Stroma (PLAU)	$3.06\ 10^{-5}$
Angiogenesis (VEGF)	0.45
Immune response (STAT1)	0.0007



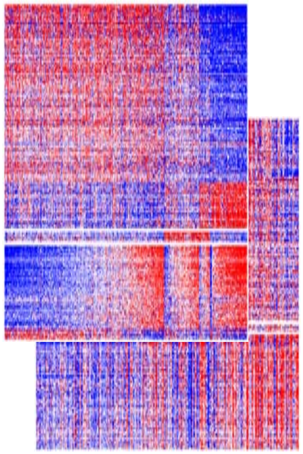
Multivariate analysis in ER-NEGATIVE tumours RFS (10 years)

Multivariate analysis in ER-POSITIVE tumours RFS (10 years)



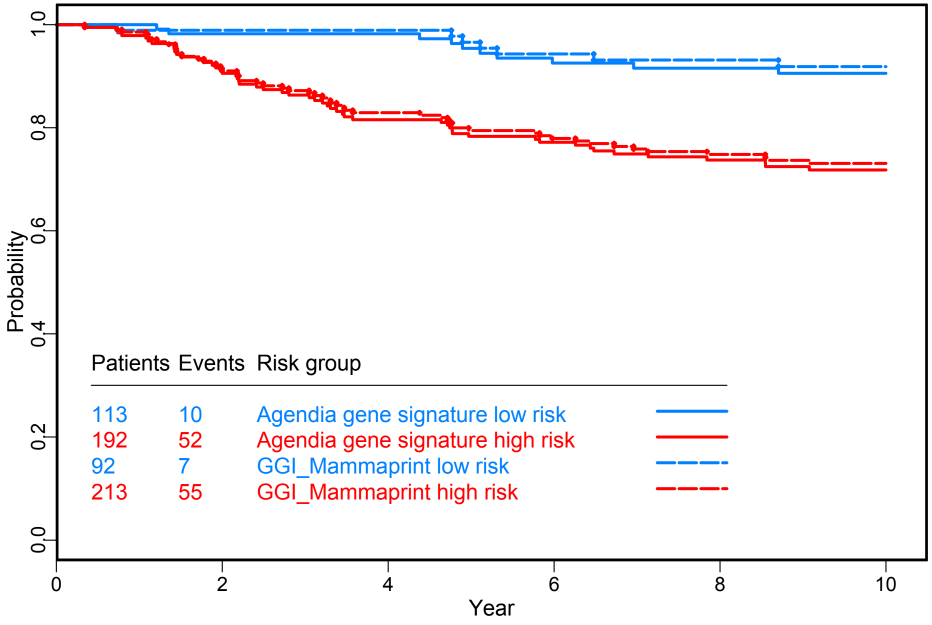
How important are proliferation genes in <u>prognostic</u> gene signatures?

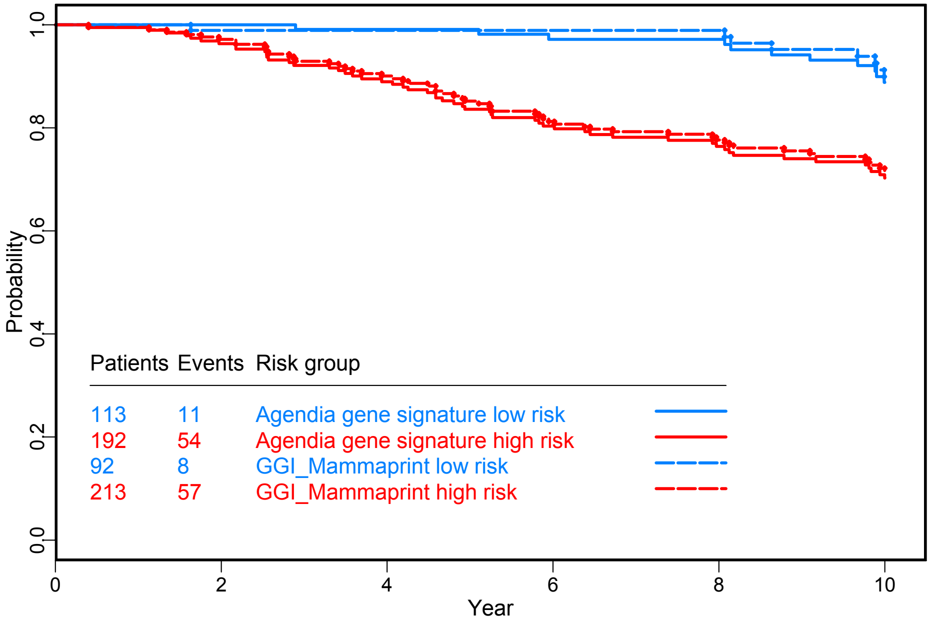
Dissecting Gene Expression Signatures…



Microarray Indices	Amsterdam No (%) 70 *genes* (Van de Vijver *et al.* NEJM, 2002)	Rotterdam No (%) 76 *genes* (Wang *et al.* The Lancet, 2005)
ESR1 = luminal/basal	35 (50)	17 (18)
ERBB2 = Her2-neu	6 (9)	2 (3)
STK6 = proliferation/GGI	34 (49)	30 (39)
PLAU = stroma/invasion	10 (14)	1 (1)
STAT1 = immune response	4 (6)	6 (8)
VEGF = angiogenesis	7 (10)	4 (5)
NA = undetermined	9 (9)	30 (39)

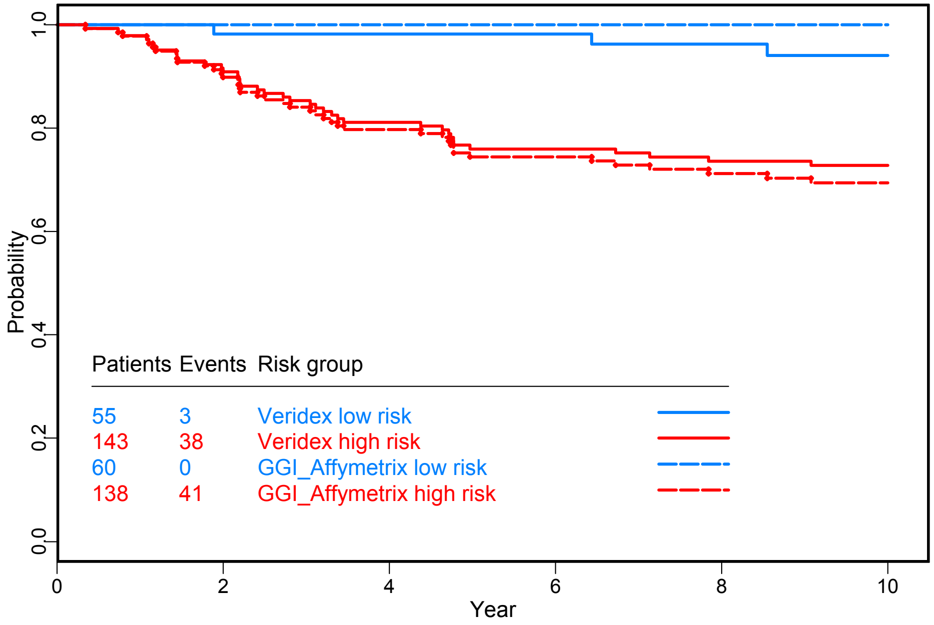
TRANSBIG VALIDATION SERIES
Using MammaPrint® (AGENDIA)
70-gene (AMSTERDAM) versus Proliferation (GGI)


TDM

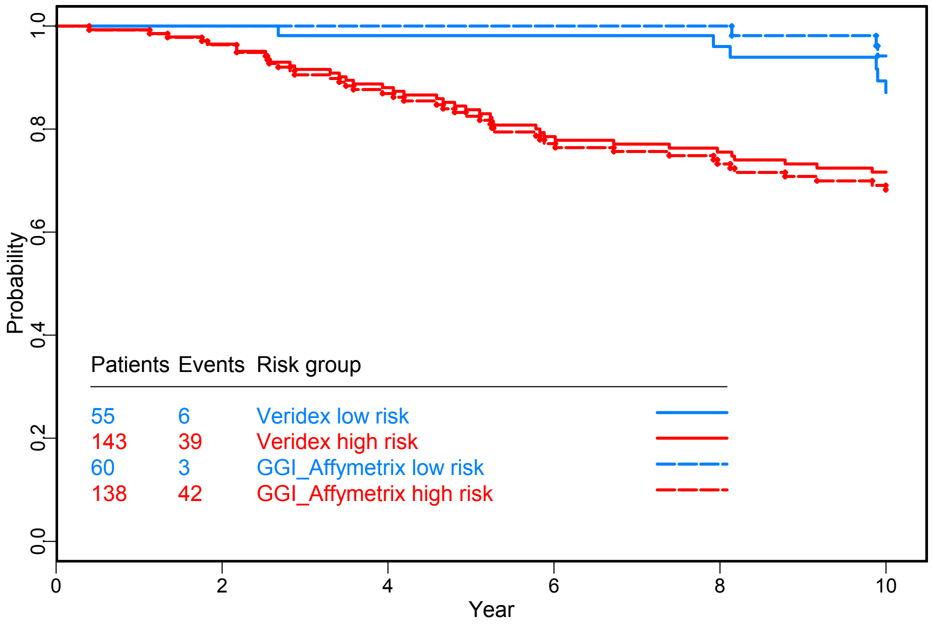

Overall Survival

TRANSBIG VALIDATION SERIES
Using Affymetrix® (VERIDEX)
76-gene (ROTTERDAM) versus Proliferation (GGI)

TDM



Overall Survival



Conclusions (1)

1. ESR1 and HER-2 indices seem to be the most prominent discriminators dichotomizing breast tumors into two main subsets in agreement with previously proposed breast cancer subtypes

2. Almost all ER-negative and HER-2-positive tumors were associated with high proliferation index scores

3. In contrast, ER-positive tumors showed a whole range of proliferation index values

Conclusions (2)

4. Proliferation captured by the gene expression grade index (GGI) appears to be a key biological factor in breast cancer and one of the most significant indicators predicting clinical outcome far beyond ER. Its weight seems to be far more important in ER-positive disease

5. Proliferation genes seem to be an important component of many existing prognostic gene signatures

6. Tumor size and nodal status retain prognostic value in addition to proliferation

7. Current molecular indices need to be refined

Acknowledgements

- Translational Research Unit
  Jules Bordet Institute
  Brussels Belgium

  Benjamin Haibe-Kains
  Christine Desmedt
  Francoise Lallemand
  Sherene Loi
  Gianluca Bontempi (ULB)

- Swiss Institute of Bioinformatics
  Asa Wirapati and Mauro Delorenzi

- IDDI
  Statistics: Marc Buyse

  Mentor: Martine Piccart

Collaborators

- TRANSBIG consortium

- VERIDEX
  Yixin Wang, Yi Zhang

- Erasmus MC (Rotterdam)
  John Foekens, Jan Klijn

Funding Support: J.C. Heuson Breast Cancer Foundation; Fondation Belge Contre le Cancer, Fonds National de Recherche Scientifique (FNRS), MEDIC Foundation