ONLINE-ONLY SUPPLEMENTARY MATERIAL

Index

Supplementary Table 1. Comparative analysis of epidemiological and clinicopathological characteristics of GBC vs. non-GBC

Characteristics		GBC (N = 70)	Non-GBC		
			GEICAM/9906 (N = 293)	Málaga Cohort (N = 96)	El ALAMO III (N = 1473)
Age at diagnosis, median		35	38	38	38
		*	p<0.001	p=0.002	p<0.001
HR-negative (n/total N)		42.9% (30/70)	15.7% (33/210)	28.4% (27/95)	24.1% (330/1367)
		**	OR = 0.25 95%CI = (0.13, 0.48) p<0.001	OR = 0.53 95%CI = (0.26, 1.07) p=0.069	OR = 0.42 95%CI = (0.25, 0.72) p<0.001
Histological grade 3, % (n/total N)		63.3% (38/60)	44.3% (121/273)	46.9% (38/81)	40.5% (479/1184)
		**	OR = 2.16 95%CI = 1.18, 4.06) P=0.009	OR = 1.95 95%CI = 0.94, 4.11) P=0.062	OR = 2.54 95%CI = 1.44, 4.57) p<0.001
Ki67 High (≥20%) % (n/ total N)		89.6% (43/48)	21.9% (46/210)	64.5% (60/93)	61.3% (209/341)
		**	OR = 0.03 95%CI = (0.01, 0.1) p<0.001	OR = 0.22 95%CI = (0.05, 0.71) P = 0.005	OR = 0.19 95%CI = (0.05, 0.56) p<0.001
First- and second-degree family history of breast and/or ovarian cancer, % (n/ total N)		52.3% (34/65)	-	-	27.3% (321/1178)
		**	-	-	OR = 2.69 95%CI = (1.59, 4.55) p<0.001
Mean age at first partum, median		32	-	-	26[1]
		*	-	-	p<0.001
First therapy[2], %	CT[3]	Neoadj.: 48.6% Adj.: 41.4% 1st line ABC: 7.1%	Adj.: 100%	Neoadj.: 100%	Neoadj.: 13.7% Adj.: 75.1% 1st line ABC:3.7%
	Only HT	1%	0	0	4%
	No systemic therapy	1%	0	0	3%

[1]343 (27%) Nulliparous patients in El ALAMO III

[2]6 *de novo* metastatic patients in the GBC population (8.6%) and 58 (3.9%) in El ALAMO III

[3]CT combined or not with HT or targeted therapy.

* Wilcoxon´s rank sum test with continuity correction.

**p-value of Fisher´s exact test.

CT: chemotherapy; HR: hormonal receptors; HT: hormonotherapy; n: number of patients with the indicated characteristic; total N: number of patients with available data; Neoadj.: neoadjuvant treatment; Adj.: adjuvant treatment; ABC: advanced breast cancer; OR: odd ratio; CI: confidence interval.

Supplementary Table 2. Intrinsic subtype classification of the GBC and control population tumors

Intrinsic subtype	GBC (n = 50) n (%)	Non-GBC * (n = 287) n (%)	p-value (Fishers' test)
Luminal A	7 (14.0)	81 (28.2)	**0.04**
Luminal B	10 (20.0)	96 (33.4)	0.86
HER2-enriched	11 (22.0)	70 (24.4)	0.07
Basal-like	22 (44.0)	40 (13.9)	**<0.01**

* Patients diagnosed at age ≤42 years from GEICAM/9906 and Malaga Cohort

Supplementary Table 3. Comparative analysis of GBC and non-GBC population by intrinsic subtypes

Age at diagnosis (years) (n = 50)	**Luminal A** (t test pairwise)*	**Luminal B** (t test pairwise)*	**HER2-enriched** (t test pairwise)*	**p-value (analysis of variance model)**
Luminal A (Median = 39.1)	-	-	-	0.016
Luminal B (Median = 33.5)	0.06	-	-	
HER2-enriched (Median = 36.2)	0.86	1.0		
Basal-like (Median = 35.2)	0.02	1.0	0.6	

* t test pairwise comparisons between group levels with corrections for multiple testing

Supplementary Table 4. Lymphocyte infiltration in GBC tumors in all samples (n = 55) and stratified by intrinsic subtypes and by basal-like vs. non-basal-like subtype.

Tumors	**Low TIL content (≤ 10%)**	**Intermediate TIL content (11-59%)**	**High TIL content (≥ 60%)**	
Total (n = 55)	31 (56.36%)	14 (25.45%)	10 (18.18%)	
By intrinsic subtype (n = 50)	**Low TIL content (≤ 10%) (n = 28)**	**Intermediate TIL content (11-59%) (n = 13)**	**High TIL content (≥ 60%) (n = 9)**	**p-value (Fishers' test)**
Luminal A (n = 7)	5 (71.43%)	1 (14.29%)	1 (14.29%)	0.340
Luminal B (n = 10)	5 (50%)	4 (40%)	1 (10%)	
HER2-enriched (n = 11)	8 (72.73%)	3 (27.27%)	0 (0%)	
Basal-like (n = 22)	10 (45.45%)	5 (22.73%)	7 (31.82%)	
By basal subtype (n = 50)	**Low TIL content (≤ 10%) (n = 28)**	**Intermediate TIL content (11-59%) (n = 13)**	**High TIL content (≥ 60%) (n = 9)**	**p-value (Fishers' test)**
Basal-like (n = 22)	10 (45.45%)	5 (22.73%)	7 (31.82%)	0.089
Non-basal (n = 28)	18 (64.29%)	8 (28.57%)	2 (7.14%)	

Supplementary Table 5. Up and down-regulated genes in GBC vs. No-GBC patients by intrinsic subtype.

Genes with FDR < 0.05 are listed

Up-regulated genes in GBC	Down-regulated genes in GBC
Luminal A	
MPP1	*ERBB3*
PNP	*ESR1*
MYBL2	*UBE2T*
GAL	*LEPRE1*
F11R	*VAMP8*
	FOXC1
	FOXA1
	PLOD1
	ACTR3B
	SPINT2
	RRAGD
	CXXC5
	ANLN
	KRT8
	DDR1
	CAV1
	EXO1
	MAPT
Luminal B	
PNP	*LEPRE1*
RRM2	*UBE2T*
MYBL2	*CAV1*
MKI67	*FOXC1*
F11R	*ERBB3*
TYMS	*PLOD1*
MPP1	*FBN1*
NDC80	*VAMP8*
PIK3CA	*ACTR3B*
PTTG1	*VIM*
CCNB1	*EGFR*
NUF2	*LHFP*
MIA	*RAD17*
KIF2C	*RRAGD*
GRHL2	*MMP11*
UBE2C	*EMP3*
SLC39A6	
GAL	
MELK	
RAB25	
FGFR4	
PHGDH	
GPR160	
CLDN3	
HER2-enriched	
PNP	*VAMP8*
MPP1	*ERBB3*
SLC39A6	*UBE2T*

PTEN	*DDR1*
MYBL2	*CD24*
	EXO1
	RAD17
	CXXC5
	LHFP
Basal-like	
MYBL2	*BAG1*
GAL	*MDM2*
UBE2C	*MKI67*
NUF2	*ANLN*
GATA3	*LEPRE1*
KRT19	*FOXC1*
TYMS	*ACTR3B*
KIF2C	*ERBB3*
PIK3CA	*MLPH*
PNP	*CEP55*
CENPF	*NDC80*
PTTG1	*CXXC5*
MPP1	*SLC39A6*
PHGDH	*FGFR4*
ZEB1	*CD24*
PLOD1	*TMEM45B*
CCNB1	*MELK*
CAV1	*RAD17*
SLC16A3	*SPINT1*
SH2B3	*MMP11*
CDH3	*EXO1*
RRAGD	*MAPT*
CDC20	*AXL*
NT5E	*BLVRA*
JUP	*ERBB2*
PGR	*CCNE1*
CLDN4	*VIM*
EMP3	*CDC6*
ESRP1	
FBN1	
ESR1	
F11R	
CLDN3	
RAB25	
GRB7	
FLVCR2	
FOXA1	
PVRL3	
RRM2	
FABP5	
UBE2T	
NDRG1	
MYC	
PTEN	
DDIT4	
BCL2	
EPCAM	

GRHL2	

The non-GBC control populations are: GEICAM/2006-11 (LEA) clinical trial [22] for Luminal A and B subtypes, GEICAM/2009-03 (ConvertHER) study [23] for HER2-enriched subtype and GEICAM/2003-11-CIBOMA/2004-01 clinical trial [24] for basal-like subtype.

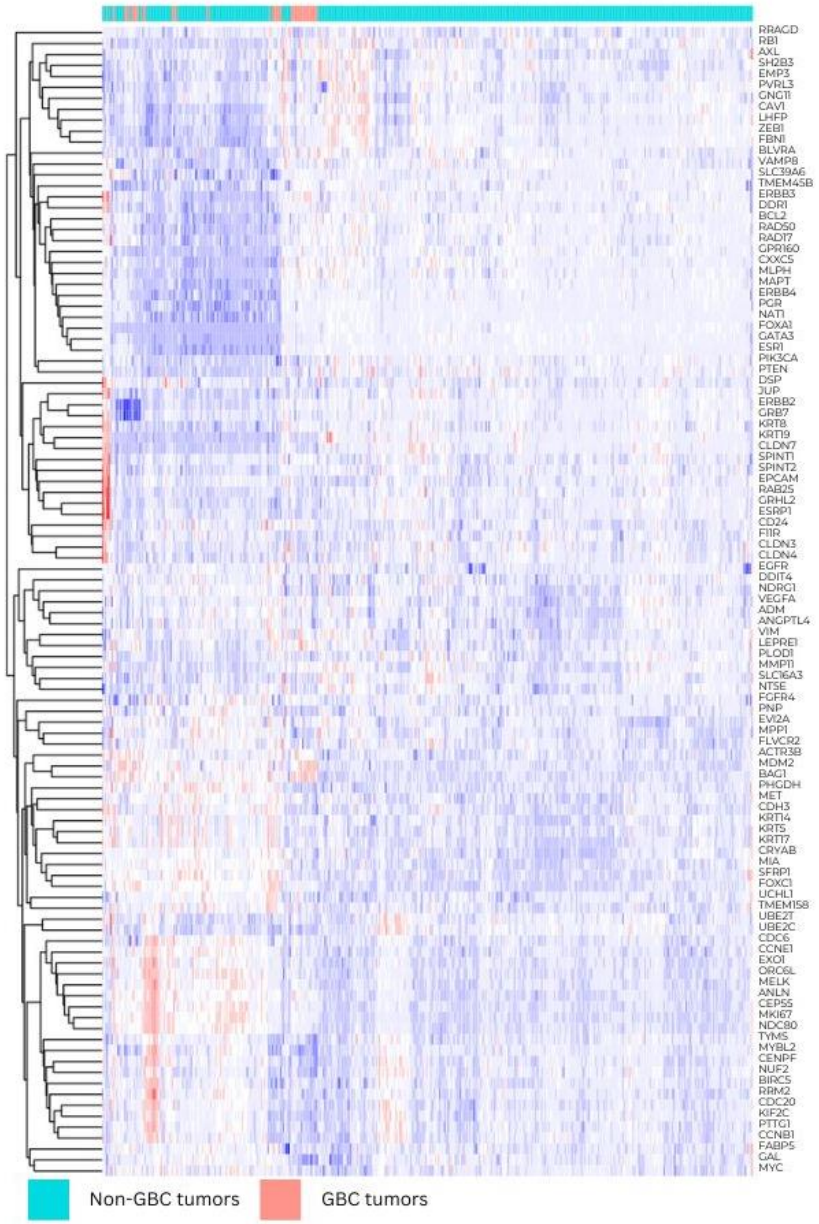
Supplementary Table 6. List of participating hospitals and central laboratories, indicating PI or person responsible for the service

PIs	Hospital	# patients recruited
Anna Lluch	Hospital Clínico Universitario de Valencia	29 (41,4%)
Amparo Ruiz / Ángel Guerrero	Instituto Valenciano de Oncología	24 (34,5%)
Juan de la Haba	Hospital Universitario Reina Sofía de Córdoba	17 (24,1%)
PIs	**Laboratory**	**Analysis / Aim**
Federico Rojo	Pathology unit. Hospital Universitario Fundación Jiménez Díaz	Pathology review and tumor sample management
Aleix Prat	Grupo de Genómica Traslacional. VHIO.	Gene expression analysis
Regina Peña Enriquez	Pathology unit. Hospital Universitario Reina Sofía de Córdoba	Tumor-infiltrating lymphocyte (TIL) analysis
Pablo Mínguez	Bioinformatic unit. Hospital Universitario Fundación Jiménez Díaz	Bioinformatic analysis

PI: principal investigator;

Supplementary Figure 1. Heatmap of gene expression values (normalized) for GBC and non-GBC tumors.

Unweighted Pair-Group Method with Arithmetic mean (UPGMA) clustering shows samples aggregation of same type cancers.



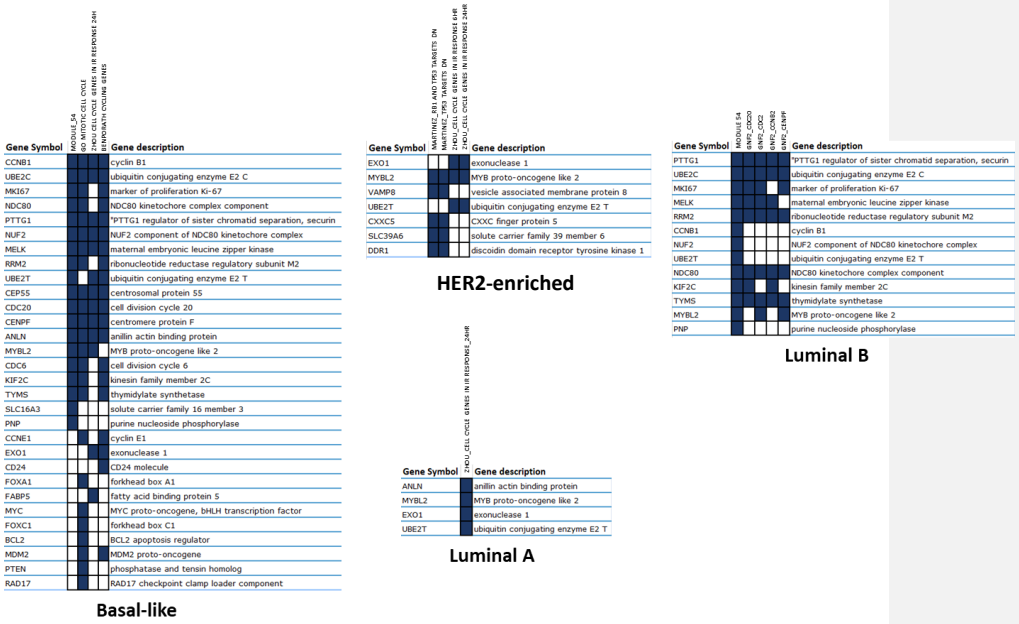
Comentado [DF1]: Reviewer 1 comment: supp figure 1 is illegible.

Comentado [DF2R1]: Corregido

Supplementary Figure 2. Gene Set Enrichment Analysis.

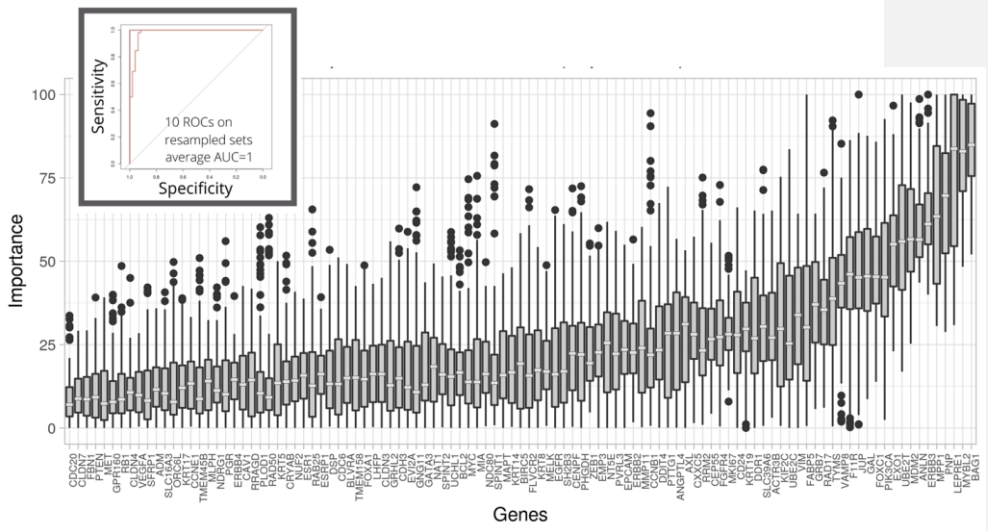
Cell cycle-related gene/gene set overlap matrix in GBC/gene set overlap matrix in the different breast cancer intrinsic subtypes.

Dark blue color shows the genes implicated in each geneset.



Supplementary Figure 3. Gene importance contribution in neural network (NN) models including GBC and non-GBC samples.

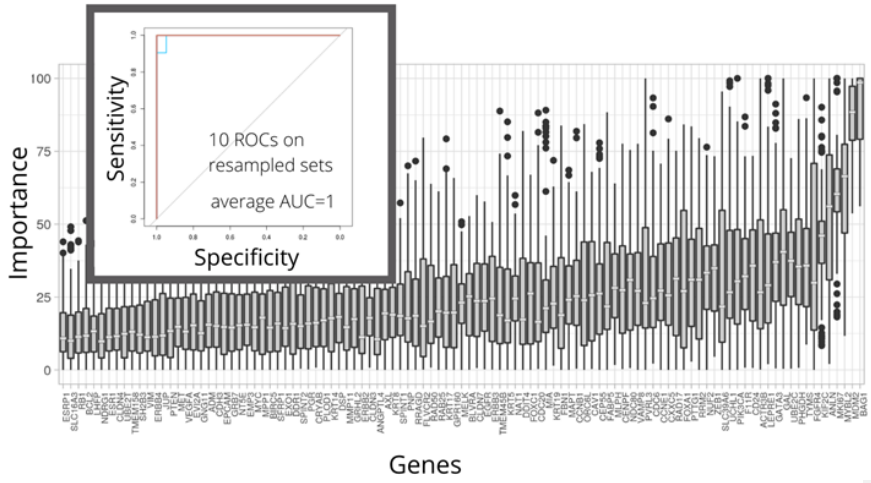
The upper panel shows the NN ROC of all GBC and non-GBC samples. A ROC (total = 10) was drawn for each NN model performed over resampled non-GBC samples.



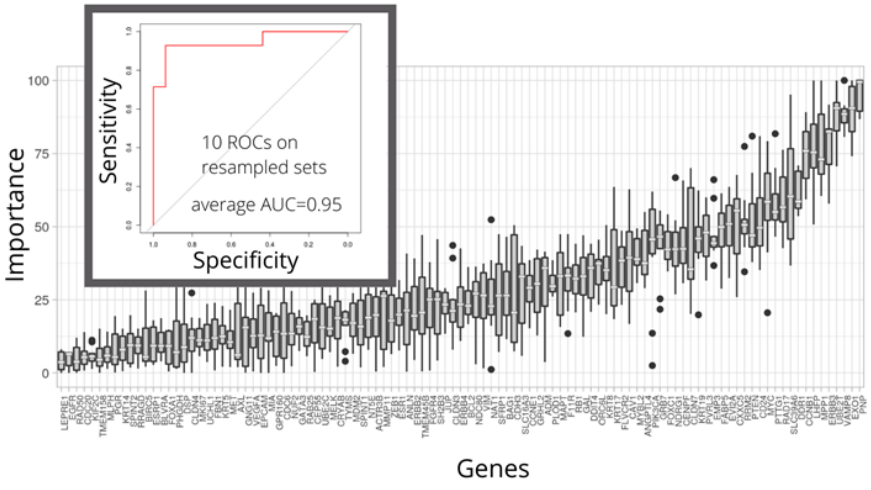
Supplementary Figure 4. Gene importance contribution in neural network (NN) models including all GBC and non-GBC samples.

The upper panel for each intrinsic subtype shows the NN ROC of the gestational and non-gestational breast cancer samples. A ROC (total = 10) is drawn for each NN model performed over resampled non-pregnancy samples. In every model, genes are sorted by their contribution in the capacity of sample classification, the so-called importance.

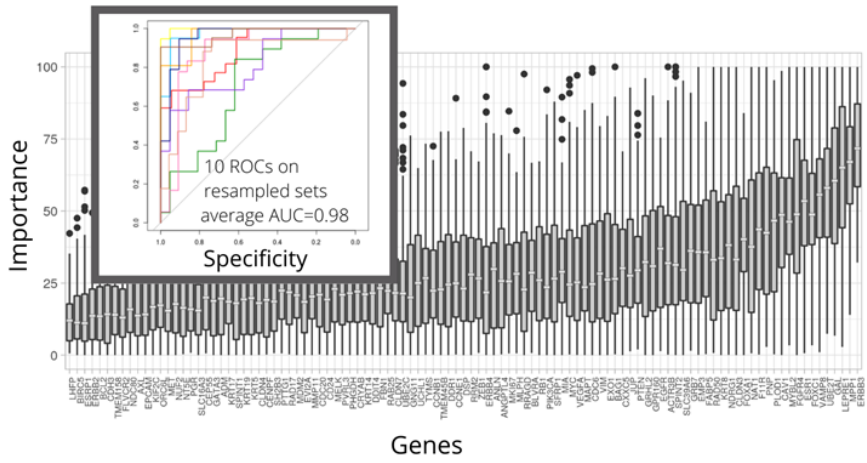
A) Basal-like



B) HER2-enriched



C) Luminal A



D) Luminal B