Sample parameters affecting the clinical relevance of RNA biomarkers in translational breast cancer research

Vassiliki Kotoula, Konstantine T. Kalogeras, George Kouvatseas, Despoina Televantou, Ralf Kronenwett, Ralph M. Wirtz, George Fountzilas

ESM_1:

CLINICAL DATA, TUMOR CHARACTERISTICS & SUPPLEMENTAL METHODS

ESM_1_1: Description of standard patient, clinical and tumour phenotypic characteristics in the three matched sample series. (A) MD and NMD primary infiltrating breast carcinomas; (B) MD and NMD metastatic lymph nodes; (C) primary infiltrating breast carcinomas with paired metastatic lymph nodes. The same parameters are presented for the entire HE10/97 patient cohort.

	A (N=98)		**B** (N=72)		**C** (N=93)		**entire cohort** (N=367)	
	median	range	**median**	range	**median**	range	**median**	range
Age (years)	50.4	24.4 - 78.0	52.3	27.6 - 74.5	52.3	23.8 - 78.0	50.4	22.5 - 78
	N cases	**%**	**N cases**	**%**	**N cases**	**%**	**N cases**	**%**
Menopausal status	**98**		**72**		**93**		**358**	
pre	55	56.1	35	48.6	47	50.5	196	53.4
post	43	43.9	37	51.4	46	49.5	171	46.6
Tumor size	**98**		**72**		**93**		**366**	
<2 cm	41	41.8	32	44.4	37	39.8	118	32.2
2-5 cm	46	46.9	29	40.3	43	46.2	185	50.5
>5 cm	11	11.2	11	15.3	13	14.0	63	17.2
Nodal involvement	**98**		**72**		**93**		**366**	
0-3 nodes	27	27.5	21	29.2	21	22.6	94	25.7
≥4 nodes	71	72.5	51	70.8	72	77.4	272	74.3
Grade	**98**		**72**		**93**		**366**	
I-II	56	57.1	32	44.4	43	46.2	184	50.3
III-Undifferentiated	42	42.9	40	55.6	50	53.8	182	49.7
ER IHC	**97**		**71**		**92**		**362**	
negative	19	19.6	16	22.5	20	21.7	110	30.4
positive	78	80.4	52	77.5	72	78.3	252	69.6
PgR IHC	**98**		**70**		**90**		**358**	
negative	34	34.7	19	27.1	29	32.2	133	37.2
positive	64	65.3	51	72.9	61	67.8	225	62.8
HER2 status*	**91**		**63**		**87**		**330**	
negative	68	74.7	43	68.3	58	66.7	231	70
positive	23	25.3	20	31.7	29	33.3	99	30
TREATMENT	**98**		**72**		**93**		**367**	
radiotherapy	78	79,6	45	62,5	69	74,2	293	80,5
hormonal treatment	93	94,9	64	88,9	83	89,2	330	90,7
E-T-CMF	49	50,0	30	41,7	43	46,2	165	45,0
E-CMF	49	50,0	42	58,3	50	53,8	202	55,0
DFS (months)								
Mean	76,1		71,8		71,1		72.5	
Median	88,3		84,3		84,5		86.0	
Std. Deviation	34,4		32,5		32,8		34.1	
Min - Max	9.4 - 121.5		6.1 - 115.7		4.0 - 115.7		3.6 - 126.3	
OS (months)								
Mean	86,4		83,4		84,5		83.6	
Median	94,2		86,7		89,9		90.2	
Std. Deviation	25,8		23,3		22,7		25.6	
Min - Max	19.0 - 122.8		27.8 - 115.7		20.9 - 115.7		7 - 126.3	

*: HER2 IHC 3+ and/or FISH amplified; IHC: immunohistochemistry; DFS: disease free survival; OS: overall survival

Tissue processing

H&E sections from each tissue block were evaluated for: tumour presence and histologic grade; TCC% in whole sections (TCC_NMD); TCC% in the areas that were marked for dissection (TCC_MD); and, for the presence of in situ carcinoma (ISC), hyperplastic lesions and normal breast epithelial components. All histologic components were recorded as continuous variables, as shown in Table **ESM_1_2.**

ESM_1_2: Histologic and phenotypic characteristics of primary infiltrative breast carcinomas in sample cohort A (N=98). Numbers represent % of each parameter in whole sections.

	normal	hyperplasia	in situ carcinoma
presence, N cases	**58**	**39**	**33**
Mean	14,4	7,3	11,5
Median	7,5	7,5	7,5
Std. Deviation	14,5	5,3	12,3
Minimum	0,5	3,0	3,0
Maximum	65,0	25,0	65,0

Serial unstained thick sections (10μm) were used for RNA extraction and were processed (a) as NMD sections (whole sections), and (b) for MD, whereby carcinoma tissue areas were dissected manually with thin scalpels from sections placed on slides. The number of sections used in each case (1 – 3) varied according to the corresponding available tissue areas in NMD sections and in MD fragments. As previously evaluated for the RNA extraction method employed in this study, for tissues containing ~70% sectioned cells, 1 section for tissue areas at 5cm^2 yielded the same RNA content with 3 sections of tissue areas at 1.25cm^2 (data not shown). NMD sections and MD tissue fragments were placed into separate Eppendorf tubes and processed for RNA extraction.

RNA extraction and mRNA expression investigations

The magnetic bead-based method employed (VERSANT Tissue Preparation System, Siemens Healthcare Diagnostics) involves an extra step of DNase I digestion for ensuring the presence of pure RNA and the absence of DNA in the sample. The method was previously validated for yielding molecular templates of adequate quality from FFPE sections for the assessment of gene expression with quantitative real time PCR.

RNA samples were assessed by 1-step reverse-transcription quantitative real-time PCR (RT-qPCR) for the relative expression of the following genes: *ESR1*, *ERBB2*, *MAPT*, *MMP7*, and *RACGAP1*. A transcript of *RPL37A*, (ribosomal protein L37a) was assessed for sample normalization and for the evaluation of RNA content. Details of the FAM-TAMRA labelled Taqman assays for all transcripts, including PCR efficiencies that were calculated as $E=1^{(10\text{-slope})}$, are shown in Table **ESM_1_3**. SuperScript® III Platinum® One-Step Quantitative RT-PCR System with ROX (Life Technologies) according to the manufacturer's instructions was used. Experiments were performed on an ABI PRISM® 7900HT (Applied Biosystems) with 30 min at 50 °C and 2 min at 95 °C followed by 40 cycles of 15s at 95 °C, and 30s at 60 °C. Paired sampled were included in the same run. All samples were run in triplicates and were considered eligible for analysis when *RPL37A* CT (cycle threshold) values were <32 (triplicate mean values). Relative expression levels (relative quantification, RQ) of the target mRNAs were calculated as $(40 - [\text{mean } CT_{target} - \text{mean } CT_{RPL37A}])$ to yield positively correlated numbers and to facilitate comparisons.

RPL37A was selected as the endogenous reference because it was initially identified as one among the least variably expressed genes in 399 fresh tissue data sets from breast cancer samples, normal control tissues and non breast cancer tissue samples (e.g. colorectal cancer) by GeneData software packages. Subsequently, RPL37A was verified as a stably expressed reference gene in breast cancer by a model-based variance estimation approach [1] using Affymetrix U133A gene expression data sets from 379 breast cancer tumors.

Table ESM_1_3

Taqman assays employed in this study

gene	Oligo ID	Oligo sequences 5' - 3'	Amplicon length (nts)	Assay location	PCR Efficacy in %
RPL37A#	forward	TGTGGTTCCTGCATGAAGACA	65	c. ex 3 - 4	86
	reverse	GTGACAGCGGAAGTGGTATTGTAC		NM_000998.4	
	probe	TGGCTGGCGGTGCCTGGA			
ESR1	forward	GCCAAATTGTGTTTGATGGATTAA	73	3'-UTR	96
	reverse	GACAAAACCGAGTCACATCAGTAATAG		NM_000125.3 and all ESR1	
	probe	ATGCCCTTTTGCCGATGCA		alternative mRNAs	
ERBB2	forward	CCAGCCTTCGACAACCTCTATT	87	c. ex 27 - 27	96
	reverse	TGCCGTAGGTGTCCCTTTG		NM_004448.2 and all ERBB2	
	probe	ACCAGGACCCACCAGAGCGGG		variants	
MAPT	forward	GTGGCTCAAAGGATAATATCAAACAC	82	c. ex 11-12	94
	reverse	ACCTTGCTCAGGTCAACTGGTT		NM_001123066.3 and all MAPT	
	probe	AGACTATTTGCACACTGCCGCCT*		mRNAs	
MMP7	forward	CGGGAGGCATGAGTGAGCTA	86	c. ex 1 - 2	90
	reverse	GGCATTTTTTGTTTCTGAGTCATAGA		NM_002423.3	
	probe	AGTGGGAACAGGCTCAGGACTATCTCAAGAG			
RACGAP1	forward	GAATGTGCGGAATCTGTTTGAG	86	c. ex 1- 2	98
	reverse	TCGCCAACTGGATAAATTGGA		NM_013277.3, NM_001126103.1,	
	probe	ACTGAGAATCTCCACCCGGCGCA*		NM_001126104.1	

= reference gene (endogenous reference); * = antisense probes; c. = coding

ADDITIONAL REFERENCE, ESM_1

1. Andersen CL, Jensen JL, Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets Cancer research 64:5245-5250