EUROPEAN JOURNAL OF CANCER 48 (2012) 12–23

available at www.sciencedirect.com

journal homepage: www.ejconline.com

Melanoma antigen family A identified by the bimodality index defines a subset of triple negative breast cancers as candidates for immune response augmentation

Thomas Karn [a,*], *Lajos Pusztai* [b], *Eugen Ruckhäberle* [a], *Cornelia Liedtke* [c], *Volkmar Müller* [d], *Marcus Schmidt* [e], *Dirk Metzler* [f], *Jing Wang* [g], *Kevin R. Coombes* [g], *Regine Gätje* [a], *Lars Hanker* [a], *Christine Solbach* [a], *Andre Ahr* [a], *Uwe Holtrich* [a], *Achim Rody* [a], *Manfred Kaufmann* [a]

[a] *Department of Obstetrics and Gynecology, J.W. Goethe-University, Frankfurt, Germany*
[b] *Department of Breast Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA*
[c] *Department of Obstetrics and Gynecology, University of Muenster, Muenster, Germany*
[d] *Department of Obstetrics and Gynecology, University Hospital Hamburg-Eppendorf, Hamburg, Germany*
[e] *Department of Obstetrics and Gynecology, Gutenberg-University, Mainz, Germany*
[f] *Department of Biology II, Ludwig-Maximilians-University, Munich, Germany*
[g] *Department of Bioinformatics and Computational Biology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA*

ARTICLE INFO

Article history:
Available online 7 July 2011

Keywords:
Bimodal expression
Immune therapy
CT-X antigens
Prognostic markers
Subtypes of breast cancer
Microarray analysis

ABSTRACT

Background: Molecular markers displaying bimodal expression distribution can reveal distinct disease subsets and may serve as prognostic or predictive markers or represent therapeutic targets. Oestrogen (ER) and human epidermal growth factor receptor 2 (HER2) receptors are strongly bimodally expressed genes in breast cancer.

Material and methods: We applied a novel method to identify bimodally expressed genes in 394 triple negative breast cancers (TNBC). We identified 133 bimodally expressed probe sets (128 unique genes), 69 of these correlated to previously reported metagenes that define molecular subtypes within TNBC including basal-like, molecular-apocrine, claudin-low and immune cell rich subgroups but 64 probe sets showed no correlation with these features.

Results: The single most prominent functional group among these uncorrelated genes was the X chromosome derived Cancer/Testis Antigens (CT-X) including melanoma antigen family A (MAGE-A) and Cancer/Testis Antigens (CTAG). High expression of CT-X genes correlated with worse survival in multivariate analysis (HR 2.02, 95% CI 1.27–3.20; $P = 0.003$). The only other significant variable was lymph node status. The poor prognosis of patients with high MAGE-A expression was ameliorated by the concomitant high expression of immune cell metagenes (HR 1.87, 95% CI 0.96–3.64; $P = 0.060$), whereas the same immune metagene had lesser prognostic value in TNBC with low MAGE-A expression.

Conclusions: MAGE-A antigen defines a very aggressive subgroup of TNBC; particularly in the absence of immune infiltration in the tumour microenvironment. These observations suggest a therapeutic hypothesis; TNBC with MAGE-A expression may benefit the most from further augmentation of the immune response. Novel immune stimulatory drugs

* *Corresponding author:* Address: Department of Obstetrics and Gynecology, J.W. Goethe-University, Theodor-Stern-Kai 7, D-60590 Frankfurt, Germany. Tel.: +49 69 6301 4120; fax: +49 69 6301 7025.
E-mail address: t.karn@em.uni-frankfurt.de (T. Karn).

0959-8049/$ - see front matter
doi:10.1016/j.ejca.2011.06.025

such as (anti-cytotoxic T-lymphocyte antigen-4 CTLA-4) directed therapies provide a realistic opportunity to directly test this hypothesis in the clinic.

1. Introduction

Bimodal gene expression indicates that two distinct subpopulations exist in the data that may correspond to clinically important subtypes of a disease. Several analytical methods were proposed to identify bimodally or multi-modally expressed genes in high dimensional data sets.[1–4] We recently developed a new computational tool that assigns a continuous '*bimodality index*' (BI) score to each probe set and allows ranking of genes by bimodality in a given data set.[5] The larger the separation between the two modes of distribution and the more balanced the size of the groups is, the larger the BI index. A large number of genes show strong bimodal expression in breast cancer but many of these are also highly co-expressed with oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2).[6–8] Separate analysis of hormone receptor-based subtypes of breast cancer is important to avoid rediscovery of molecular markers of known subtypes. Subtype restricted analysis is more likely to identify novel prognostic and predictive markers that could add to existing classification.[9] We hypothesised that bimodally expressed genes within ER, PgR and HER2-negative breast cancers (triple negative, TNBC) may reveal natural and clinically relevant subsets of this cancer type. To test this hypothesis we applied our method to expression data of TNBC. Our TNBC-specific analysis revealed over 130 bimodally expressed genes, many corresponding to known metagenes that were previously suggested as stratification tools for TNBC. However, among the previously unknown markers we identified several Cancer/Testis Antigens (CT-X). These antigens are predominantly expressed in human germ line cells, and are absent in adult somatic tissues.[10] However, they have been shown to be re-expressed in various cancers. Because of their highly specific expression in cancer cells and their ability to induce immune responses mediated by cytotoxic T cells, these antigens are in the focus of efforts to develop vaccines against them as potential cancer treatment.[11,12]

2. Methods

This analysis was performed following the REMARK recommendations for tumour marker studies.[13] Fig. 1 illustrates the analytical strategy and the flow of samples through the study, including the number of cases used in each stage of the analysis. All analyses were performed using the R software environment (http://www.r-project.org/) and SPSS version 17.0 (SPSS Inc., Chicago, Illinois). Chi square test was used to assess associations between categorical parameters. All reported P values are two sided and $P \leqslant 0.05$ was considered significant.

2.1. Assembly of a combined Affymetrix dataset from triple negative breast cancers

To generate a homogeneous dataset for the analysis of bimodally expressed genes we used (i) only one array platform (Affymetrix U133 gene chips) and (ii) included only samples defined as triple negative based on the mRNA expression levels of ER, PR, and HER2 as previously described.[6] To obtain the largest possible sample size for analysis we started with gene expression data corresponding to $n = 3488$ primary breast cancers representing 28 different datasets (Supplementary Table S1). Only primary invasive breast cancers and no metastases were included in this starting cohort. In addition, to avoid any effects of treatment on gene expression, only samples from patients untreated at the time of sample acquisition were included. In the case of neoadjuvant treatment only pretherapeutic biopsies were used in the analyses. All expression data were processed with the MAS5.0 algorithm[14] of the *affy* package[15] of the Bioconductor software project.[16] Subsequently data from each array were $\log_2$-transformed, median-centered, and the expression values of all the probe sets from the U133A arrays were multiplied by a scale factor S so that the magnitude (sum of the squares of the values) equals one. Within this data set, we identified 579 triple negative breast cancers (TNBC) based on the expression of ER, PgR, and HER2 as detected by the microarray.[6] Next, we calculated a comparability metric C for each primary dataset to identify the most comparable samples. This metric C is derived from the sum of the squared differences of the mean (μ) within a specific dataset and among all datasets, respectively, normalised by the standard deviation (σ) calculated for all genes (g) on the array:

$$C_{\text{dataset}_i} = \sum_{g=1}^{n} \left(\frac{\mu_g \text{dataset}_i - \mu_g \text{total}}{\sigma_g \text{total}} \right)^2 \quad (1)$$

All datasets were sorted according to this metric and the top 15 datasets with the lowest values (norm. $C \leqslant 0.03$) including 394 samples combined were selected as the analysis cohort (Supplementary Fig. S5). The rational of this metric is based on the assumption that the mean of a genes expression within a dataset should be similar between different dataset. Therefore this mean is compared to the global mean and the difference normalised by the respective variance. When summing up these differences for all genes on the array the resulting comparability metric C gives a measurement to which extent the arrays in the specific dataset differ from the complete cohort of all arrays. Importantly this method does not intend to define which datasets are 'right' of 'wrong' per se. Moreover, larger datasets will dominate over smaller datasets because of their higher impact on the global mean. As demonstrated in Supplementary Fig. S5 the method allows to identify a combined cohort of rather homogeneous arrays.

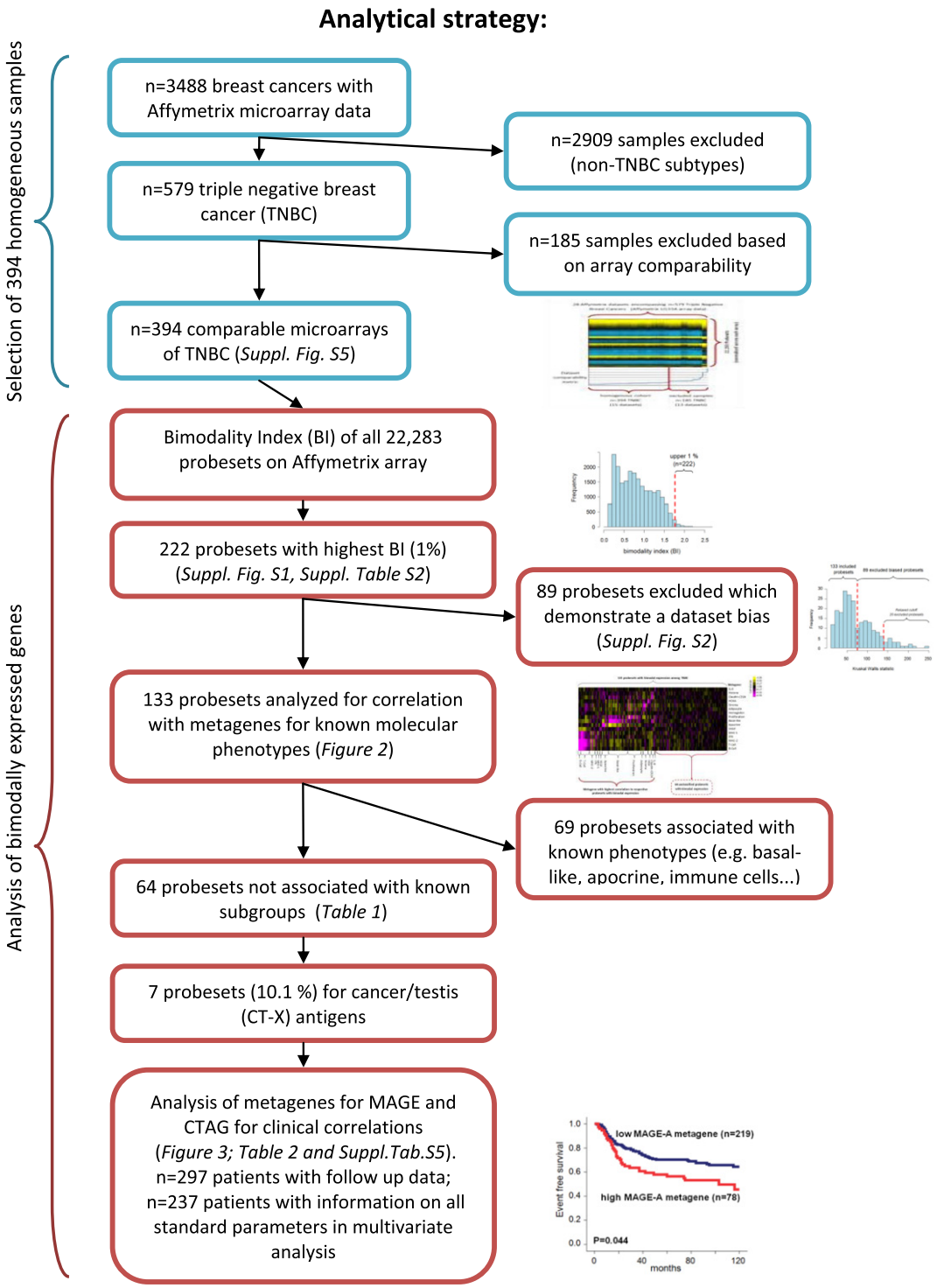

Fig. 1 – Analytical strategy. The outline of the analysis strategy is schematically shown. The upper part shows the selection of the homogenous sample cohort of 394 triple negative breast cancers (TNBC). The lower part displays the analysis of bimodally expressed genes in this cohort according to the *bimodality index* (BI) and the subsequent selection of those genes providing information independent to previous known molecular factors.

2.2. *Identification of genes with a bimodal expression*

We applied the R function *bimodalIndex*[3] from the package *ClassDiscovery* (http://bioinformatics.mdanderson.org/Software/OOMPA/) to the expression data of 394 TNBC samples. All 22,283 probe sets available on the Affymetrix U133A array were included in the analysis. Probe sets were ranked according to the bimodality index score (BI) and based on the distribution of Bi values the upper 1% ($n = 222$ probe sets) were selected for futher analyses (Supplementary Fig. S1). Next,

Table 1 – 133 Bimodally expressed probe sets.

Assigned to metagene	Affymetrix probe set ID[a]	Gene-Symbol[a]	Maximum correlation to metagene	Gene title	KEGG pathway	OMIM disease information
Unclassified	204965_at	GC	–0.083	Group-specific component (vitamin D binding protein)		Susceptibility to Graves disease
Unclassified	217129_at	–	–0.085	Clone RP1-20N2		
Unclassified	216710_x_at	ZNF287	–0.086	Zinc finger protein 287		
Unclassified	216140_at	–	0.092	DKFZp434K1126 Zp434K1126		
Unclassified	217450_at	IRS3L	–0.095	Similar to insulin receptor substrate like protein		
Unclassified	216498_at	–	0.096	Clone RP3-336K20		
Unclassified	216303_s_at	MTMR1	0.098	Myotubularin related protein 1	Fructose and mannose metabolism	
Unclassified	209942_x_at	MAGEA3	–0.100	Melanoma antigen family A, 3		
Unclassified	216592_at	MAGEC3	–0.103	Melanoma antigen family C, 3		
Unclassified	AFFX-LysX-M_at		–0.105		Lysine biosynthesis	
Unclassified	221674_s_at	CHRD	0.108	Chordin	TGF-β signalling pathway	
Unclassified	217210_at	1422_g_at	0.110	Clone 8B22		
Unclassified	222277_at		–0.111	Hypothetical protein LOC100133696		
Unclassified	203815_at	GSTT1	0.111	Glutathione S-transferase theta 1	Glutathione metabolism	
Unclassified	217246_s_at	DIAPH2	–0.115	Diaphanous homologue 2 (Drosophila)	Regulation of actin cytoskeleton	Premature ovarian failure
Unclassified	207706_at	USH2A	–0.116	Usher syndrome 2A (autosomal recessive, mild)		Retinitis pigmentosa, AR, without hearing loss, Usher syndrome, type 2A
Unclassified	219865_at	HSPC157	–0.120	Hypothetical LOC29092		
Unclassified	220567_at	ZNFN1A2	0.121	IKAROS family zinc finger 2 (Helios)		
Unclassified	219831_at	CDKL3	–0.121	Cyclin-dependent kinase-like 3		
Unclassified	216951_at	FCGR1A	–0.123	Fc fragment of IgG, high affinity Ia, receptor (CD64)	Hematopoietic cell lineage	Familial deficiency of phagocytic IgG receptor I
Unclassified	215555_at	LOC148936	–0.124			
Unclassified	202885_s_at	PPP2R1B	–0.124	Protein phosphatase 2 regulatory subunit A, β	Oocyte meiosis	Lung cancer
Unclassified	216626_at		–0.126			Susceptibility to pancreatic cancer, whole genome association study

(*continued on next page*)

Table 1 (*continued*)

Assigned to metagene	Affymetrix probe set ID[a]	Gene-Symbol[a]	Maximum correlation to metagene	Gene title	KEGG pathway	OMIM disease information
Unclassified	215315_at	ZNF549	0.129	Zinc finger protein 549		
Unclassified	210289_at	NAT8	−0.130	N-acetyltransferase 8		
Unclassified	214612_x_at	MAGEA6	−0.131	Melanoma antigen family A, 6		
Unclassified	34471_at	MYH8	0.132	Myosin, heavy chain 8	Tight junction	Carney complex variant, Trismus-pseudocamptodactyly syndrome
Unclassified	210834_s_at	PTGER3	−0.132	Prostaglandin E receptor EP3	Calcium signalling pathway	
Unclassified	215312_at		−0.132	RAD52 pseudogene		
Unclassified	214579_at	DJ462O23.2	0.133	NIPA-like domain containing 3		
Unclassified	221040_at	CAPN10	0.136	Calpain 10		Diabetes mellitus, non-insulin-dependent, Diabetes mellitus, non-insulin-dependent 1
Unclassified	221279_at	GDAP1	0.137	Ganglioside-induced differentiation-associated protein 1		Genome-wide association study, Charcot-Marie-Tooth disease, axonal, type 2 K
Unclassified	217228_s_at	ASB4	−0.137	Ankyrin repeat and SOCS box-containing 4		
Unclassified	220675_s_at	ADPN	0.138	Patatin-like phospholipase domain containing 3	Tyrosine metabolism	Genome-wide association studies of plasma levels of liver enzymes
Unclassified	215356_at	FLJ13072	0.140	Tudor domain containing 12		
Unclassified	220997_s_at	DIAPH3	−0.142	Diaphanous homologue 3 (Drosophila)	Regulation of actin cytoskeleton	
Unclassified	204545_at	PEX6	0.144	Peroxisomal biogenesis factor 6		Peroxisomal biogenesis disorder
Unclassified	210709_at		−0.144	Lipopolysaccharide specific response-67 protein, mRNA sequence		
Unclassified	216463_at		0.144	DKFZp434P2435 Zp434P2435		
Unclassified	202707_at	UMPS	0.146	Uridine monophosphate synthetase	Pyrimidine metabolism	Oroticaciduria
Unclassified	214521_at	HES2	0.152	Hairy and enhancer of split 2 (Drosophila)		
Unclassified	215846_at		−0.154	DKFZp564O1172 Zp564O1172		
Unclassified	203902_at	HEPH	−0.154	Hephaestin		
Unclassified	207470_at	DKFZp566H0824	−0.155	Hypothetical LOC54744		
Unclassified	207703_at	NLGN4Y	−0.156	Neuroligin 4, Y-linked		
Unclassified	207021_at	ZPBP	0.159	Zona pellucida binding protein		

Unclassified	206294_at	HSD3B2	−0.161	Hydroxy-delta-5-steroid dehydrogenase	Steroid hormone biosynthesis	3-β-hydroxysteroid dehydrogenase, type II, deficiency
Unclassified	219589_s_at	FLJ10922	−0.161	Transmembrane protein 143		
Unclassified	216134_at	KIAA1013	0.165	FERM domain containing 4B		
Unclassified	219247_s_at	ZDHHC14	−0.170	Zinc finger, DHHC-type containing 14		
Unclassified	212831_at	EGFL5	0.171	Multiple EGF-like-domains 9		
Unclassified	215349_at		−0.173	HCG1730474)		
Unclassified	221092_at	ZNFN1A3	0.179	IKAROS family zinc finger 3 (Aiolos)		
Unclassified	203290_at	HLA-DQA1	0.183	Major histocompatibility complex, class II, DQ alpha 1	Cell adhesion molecules (CAMs)	Genome-wide association study for coeliac disease
Unclassified	220055_at	ZNF287	−0.183	Zinc finger protein 287		
Unclassified	202067_s_at	LDLR	0.188	Low density lipoprotein receptor	Endocytosis	Polygenic dyslipidemia, Genome-wide association analysis
Unclassified	222189_at	FLJ20294	0.189	FLJ20294		
Unclassified	1320_at	PTPN21	0.190	Protein tyrosine phosphatase, non-receptor type 21		
Unclassified	217339_x_at	CTAG1B	0.192	Cancer testis antigen NY-ESO-1		
Unclassified	215915_at	CED-6	0.192	GULP		
Unclassified	210546_x_at	CTAG1B	0.193	Cancer testis antigen NY-ESO-1		
Unclassified	201372_s_at	CUL3	−0.197	Cullin 3	Ubiquitin mediated proteolysis	
Unclassified	214008_at	PTK9	−0.198	CDNA FLJ52233		
Unclassified	215733_x_at	CTAG2	0.199	Cancer testis antigen NY-ESO-1		
B-Cell	2 Probe sets	See Supplementary Table S3	⩾0.2			
T-Cell	8 Probe sets	See Supplementary Table S3	⩾0.2			
MHC-2	5 Probe sets	See Supplementary Table S4	⩾0.2			
IFN	2 Probe sets	See Supplementary Table S4	⩾0.2			
MHC-1	2 Probe sets	See Supplementary Table S4	⩾0.2			
VEGF	2 Probe sets	See Supplementary Table S4	⩾0.2			
Apocrine	6 Probe sets	See Supplementary Table S4	⩾0.2			
Basal-like	17 Probe sets	See Supplementary Table S4	⩾0.2			
Proliferation	12 Probe sets	See Supplementary Table S4	⩾0.2			
Haemoglobin	1 Probe set	See Supplementary Table S4	⩾0.2			
Adipocyte	1 Probe set	See Supplementary Table S4	⩾0.2			
Stroma	4 Probe sets	See Supplementary Table S4	⩾0.2			
HOXA	3 Probe sets	See Supplementary Table S4	⩾0.2			
Claudin-CD24	3 Probe sets	See Supplementary Table S4	⩾0.2			
IL-8	1 Probe set	See Supplementary Table S4	⩾0.2			

[a] Full information is presented only for the 64 '*unclassified*' probe sets with a correlation <0.2 to any metagene. For those 69 probe sets which were assigned to a metagene based on a correlation ⩾0.2 only summary data are given. The complete information of all 133 probe sets is given in Supplementary Table S4.

we removed probe sets which displayed a bias related to the datasets. To assess bias, we used Kruskal Wallis statistic comparing the expression of each probe sets with the primary dataset vector among the 394 TNBC. A cutoff for exclusion of probe sets due to strong association with a data set (i.e. laboratory-bias or sampling) was derived from the distribution of the Kruskal Wallis statistic (Supplementary Fig. S2).

2.3. Correlation with molecular phenotypes in TNBC

Several investigators described molecular subgroups within TNBC defined by groups of highly co-expressed genes (i.e. metagenes) that vary in expression among TNBC. To determine if our bimodal genes correspond to or serve as surrogates for these previously described metagenes, we calculated the correlation between each of our (bimodally expressed) genes and 16 metagenes known to represent different cell populations and different molecular variants of TNBC. These metagenes included the intrinsic genes of the basal-molecular class,[17] an apocrine/androgen receptor signalling signature[18,19] five signatures related to different types of immune cells,[20–23] a stromal signature,[24] the claudin-CD24 signature,[25,26] markers of blood[27] and adipocytes,[17] as well as an angiogenesis signature[28] and an inflammatory signature.[29] The probe sets representing these 16 metagenes have been assembled by unsupervised methods described in an independent publication[30] and are listed in Supplementary Table S3. Metagene values were defined as mean expression of the individual probe sets that define the metagene. Next, the expression of each of the bimodal genes (i.e. probe sets) was correlated with the expression values of the 16 metagenes. Genes that showed high correlation with one or more of the 16 metagenes were considered as surrogate for that metagene and therefore less interesting for classification purposes. Probe sets whose correlation to any metagene did not reach a pre-specified cutoff (see Section 3) were assigned as '*unclassified*' and subsequently inspected in more detail.

2.4. Survival analyses

Follow-up data were available for 297 of the 394 TNBC samples used for identification of bimodal genes (Supplementary Table S1). All survival intervals were measured from the time of surgery to the survival endpoint that was available for that datasets. In 11 datasets ($n = 160$), the end point was relapse free survival (RFS) and in six other dataset ($n = 137$) it was distant metastasis free survival (DMFS). RFS includes local recurrences as events whereas DMFS does not. In order to plot Kaplan–Meier survival curves and perform survival analysis of the pooled data, we combined both types of endpoints into a single event free survival (EFS) endpoint that includes either RFS or DMFS whichever is available for the particular case. We have previously shown that the effect of using these different endpoints was rather small in the overall dataset.[6] All results from the pooled survival analyses were also verified by examining the effect of the different endpoints in stratified analyses. Follow-up data for those women in whom the survival end point was not reached were censored at the last follow-up or at 120 months. Subjects with missing values were excluded. We constructed Kaplan–Meier curves and used the log-rank test to determine the univariate significance of the variables. In order to plot Kaplan–Meier curves, patients were dichotomised into low or high expression groups using thresholds derived from the bimodal distribution of the CT-X metagenes and in the case of the continuous distribution of the B-cell metagene based on its prognostic value among TNBC (see Supplementary Material and Supplementary Fig. S8). Cox regression analysis was applied to analyse the univariate hazard ratio of individual metagenes as continuous factors. A Cox proportional-hazards model was used to simultaneously examine the effects of multiple covariates on survival. The effect of each individual variable was assessed with the use of the Wald test and described by the hazard ratio and 95% confidence intervals (95%CI).

3. Results

We hypothesised that bimodally expressed genes within TNBC may reveal clinically important subsets of this cancer type and could also draw attention to potential novel therapeutic targets. To test this hypothesis we applied our method to pooled Affymetrix gene expression data of 394 TNBC (Supplementary Table S1, Fig. 1). To minimise batch and inter-laboratory variation we analysed only highly comparable arrays and data set-biased genes were also filtered (see Section 2). Supplementary Fig. S1 displays the distribution of the BI scores for all 22,283 probe sets represented on the Affymetrix U133 chips. The top 1% of the probe sets ($n = 222$) with the highest BI values were selected for further analysis, this corresponds to a BI score threshold of >1.768 (Supplementary Table S2). As reference, the BI scores of ER, PgR, and HER2 (probe sets 205225_at, 208305_at, and 216836_s_at) when different breast cancer subtypes are analysed together are 1.96, 1.73, and 1.63, respectively.[5] Probe sets that showed highly variable median expression values across individual data sets using Kruskal Wallis statistic (i.e. data set bias) were removed from further analysis. A stringent Kruskal Wallis metric <75 flagged 89 probe sets as biased (Supplementary Fig. S2) resulting in a final list of 133 probe sets that were strongly bimodally expressed independent of data sets (Table 1).

Next, we examined if these probe sets correspond to known co-expressed gene clusters (i.e. metagenes). We calculated the Pearson correlation between the 133 bimodally expressed probe sets and each of 16 distinct metagenes (metagene probe set IDs are presented in Supplementary Table S3) previously described in TNBC data.[17–27,29] Supplementary Table S4 lists the highest correlation coefficient and the corresponding metagenes for each of the 133 probe sets. Fig. 2 shows a heat map of the correlation matrix for the 133 probe sets and 16 metagenes in the 394 TNBC samples. Sixty-four probe sets (48%) showed correlation <0.2 to any metagene and these were designated as '*bimodally expressed unclassified*' that could represent potential new classification features (Table 1). Among these 64 probe sets, 7 (11%) targeted genes that belong to the cancer/testis (CT-X) antigen family, including MAGE (melanoma antigen family) A3, A6 and C3, and CTAG (cancer testis antigen) 1B and 2. There were no other large functional groups among the remaining 57

unclassified probe sets. There was a high correlation between MAGE-A3 and -A6 expression (but not C3) (Supplementary Fig. S3A) and we also observed a strong correlation between the three probe sets that targeted CTAG 1B and the single probe set that targeted CTAG2 (Supplementary Fig. S3B). Because of these high correlations, we used the combined average expressions of MAGE-A3 and 6 and CTAG1B and CTAG2 as measures of MAGE-A and CTAG expression, respectively. Cutoffs to define low and high expression groups for these two metagenes and for MAGE-C3 expression were established from the bimodal expression distributions of these genes (Supplementary Fig. S4). Twenty-six percent of the 394 TNBC samples ($n = 102$) were assigned to high expression of the MAGE-A metagene, 28% to high expression of MAGE-C3 ($n = 110$), and 22% to high expression of the CTAG metagene ($n = 86$). Only 12% displayed expression of both MAGE-A and CTAG metagenes and only 6% expression of all three metagenes.

We next examined the prognostic value of these antigens and plotted Kaplan Meier survival curves for low and high CT-X expression groups (Fig. 3A–C). TNBC with high expression of either MAGE-A or CTAG had poor prognosis compared to low expression, the 5-year event-free survivals (EFS) were 59% (SE ± 5.6%) versus 70% (±3.2%), $P = 0.044$ (Fig. 3A); and 60% (±6.3%) versus 69% (±3.1%), $P = 0.029$ (Fig. 3C); respectively. MAGE-C3 expression had no prognostic value (Fig. 3B). We also compared the clinical characteristics of patients with tumours displaying high or low expression of MAGE-A and CTAG metagenes. Supplementary Table S5 shows that age, tumour size, and nodal status were all equally distributed across the groups with only a somewhat higher percentage of histological grade 3 tumours in the group with high CTAG expression (88.3% versus 70.7%; $P = 0.037$). We also performed a multivariate Cox regression analysis including 237 samples for which information on lymph node status, age, tumour size, histological grade and follow-up were available. Only lymph node status (HR 2.17, 95% CI 1.00–3.69; $P = 0.050$) and high expression of the MAGE-A metagene (HR 2.02, 95% CI 1.27–3.20; $P = 0.003$) were significant independent prognostic factors (Table 2). Similar results were obtained for the CTAG metagene (Supplementary Table S6).

We and others have previously shown that lymphocyte infiltration of TNBC is associated with improved prognosis.[20–23] The high expression of CT-X antigens among these tumours might represent a *bona fide* target for an immune response and could identify a group of patients where local immune response may alter otherwise poor prognosis. We examined whether the presence or absence of CT-X expression alters the prognostic value of a B cell signature. Cancers with high expression of MAGE-A and also high expression of B-cell metagene showed a strong trend for better survival compared to high MAGE-A cancers with low B-cell signature (5 yr-EFS 67 ± 6% versus 42 ± 10%; $P = 0.06$; Fig. 3E). The prognostic value of the B-cell metagene was lower in cancers

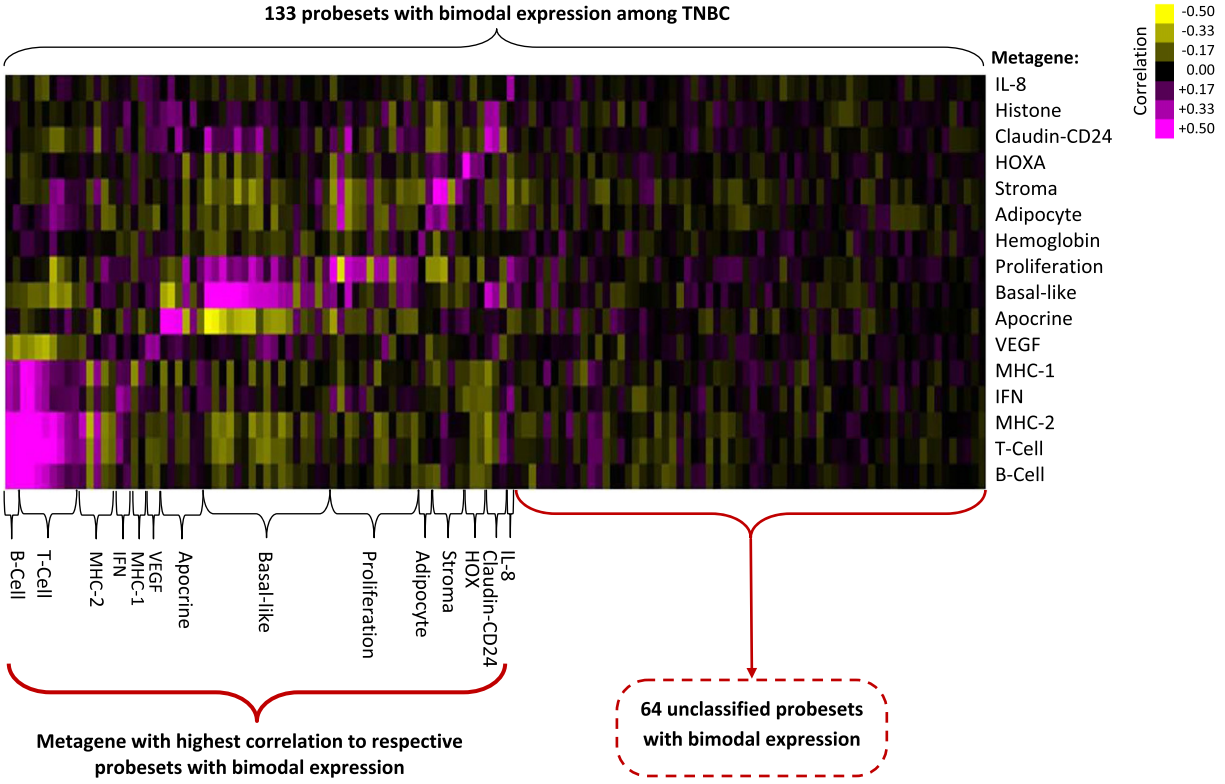

Fig. 2 – Colour representation of the correlation matrix of 133 probe sets with bimodal expression and 16 metagenes in TNBC. Shown is a colour representation of the correlation matrix of the 133 probe sets with bimodal expression from Table 1 and the 16 metagenes. Probe sets are grouped according to the assigned metagene and sorted according to their correlation from left to right in decreasing order. Positive correlation values are represented by magenta and negative correlation values by yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

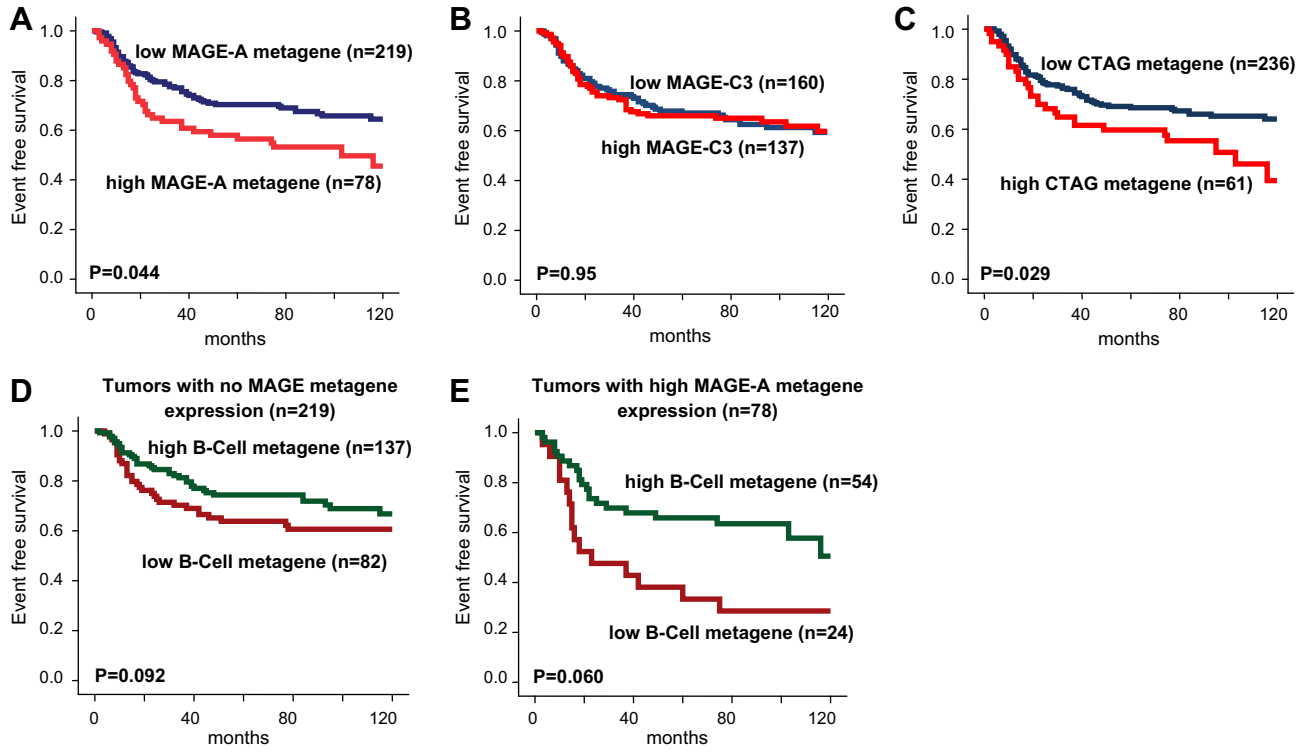

Fig. 3 – Prognosis of TNBC according to MAGE and Cancer/Testis Antigen (CTAG) metagene expression. Kaplan–Meier analysis of event free survival of the 297 TNBC patients with follow up information. Samples were stratified either according to the expression of the MAGE-A metagene in (A), MAGE-C3 in (B), or the CTAG metagene in (C) using the cutoff values derived from their distribution. The prognostic value of a B-cell metagene as a surrogate marker for lymphocyte infiltration was analysed in (D) and (E) for TNBC patients either without (D) or with high MAGE-A metagene expression (E), respectively.

Table 2 – Multivariate Cox analysis of event free survival according to clinical variables and expression of MAGE-A genes.

Variable		No. of patients[a]	Hazard ratio	95% CI	P-value[b]
Lymph node status	LNP versus LNN	27 versus 210	2.17	1.00–3.69	**0.050**
Age	>50 versus ⩽50	113 versus 124	0.69	0.44–1.10	0.116
Tumour size	⩽2 cm versus >2 cm	71 versus 166	0.80	0.48–1.32	0.38
Histological grading	G3 versus G1&2	166 versus 71	1.11	0.67–1.82	0.69
MAGE-A expression	High versus Low	59 versus 178	2.02	1.27–3.20	**0.003**

[a] Information on all parameters was available for 237 of the 297 TNBC samples with follow up data.
[b] Significant *P*-values are given in **bold**.

with low MAGE-A expression (5 yr-EFS 74 ± 4% versus 62 ± 5%; $P = 0.09$; Fig. 3D). When interaction between CTAG expression and the B-cell metagene was examined, no similar relationship was seen (Supplementary Fig. S9). These observations suggest a therapeutic hypothesis; TNBC with MAGE-A expression may benefit the most from further augmentation of the immune response.[11,12]

4. Discussion

We have applied a recently developed bioinformatic method the *bimodality index* (BI) to a cohort of triple negative breast cancers (TNBC). Thus this cohort was rather homogenous with respect to the well known molecular subtypes of breast cancer.[7] Nevertheless, the persisting heterogeneity of this single subtype was demonstrated by the identification of several bimodally expressed genes. Many of those genes correlated well to different previously described molecular features as the basal-like,[17] the apocrine,[18,19] the claudin-low[25,26] subtypes as well as infiltration of immune cells.[20–23] But still bimodally expressed markers were identified which are not associated with those characteristics. Most of them did not show correlations to each other and future studies are needed on their functional and prognostic role. However the most prominent group of genes was X chromosome derived Cancer/Testis Antigens (CT-X). CT-X antigens are markers which are predominantly expressed in human germ line cells, but not in somatic tissues,[10] but become frequently activated in different cancer types. CT-X gene products are also targets of immune responses mediated by cytotoxic T cells

in some cancers, and strategies in developing vaccines that induce these responses has gained much interest.[11,12] In breast cancer previous studies based on PCR detection suggested prevalences of 13-60%[31–34] for CTAG1B and 6–19% for MAGE genes.[35] However the immunohistochemical detection of the protein was far lower with only 1-2%.[34,36] In our study we found rather high frequencies of expression of MAGE-A, MAGE-C, and CTAG1B mRNAs in TNBC of 26.1%, 27.9% and 21.8%, respectively. The distribution in different subtypes of breast cancer is shown in Supplementary Fig. S7 revealed that these markers are mainly confined to TNBC. This explains their high frequency we observed. A very recent publication by Grigoriadis et al. analysed the expression of CT-X antigens in unselected breast cancers using massively parallel signature sequencing (MPSS) and microarray analysis.[37] The findings of this study were similar to our results with a high expression of CTAG and MAGE families in up to 26% of ER negative breast cancers. Moreover, the authors of this study also confirmed the correlation with a negative ER status by immunohistochemical analysis.[37]

We have observed a poor prognosis for the subset of TNBC which displayed high expression of CTAG1B and MAGE-A metagenes ($P = 0.002$). Similar results of a worse prognosis for CT-X expression have also been described for other types of cancer.[38] In our cohort of TNBC the expression of MAGE-A and CTAG metagenes was an independent prognostic factor in multivariate regression analysis (HR 2.02, 95% CI 1.27–3.20; $P = 0.003$, and HR 2.32, 95% CI 1.41–3.82; $P = 0.001$, respectively). In contrast, with the exception of lymph node status other known prognostic factors in breast cancer such as age, tumour size, and histological grading were not significant in this analysis. Most TNBC are usually of highly proliferating and grading is not as important for prognosis in this subtype as it is in ER positive disease. TNBCs are also often associated with younger age but the impact of age and tumour size for prognosis within this subtype is not yet fully clear. Still it cannot be excluded that a bias in our cohort is the reason for the lack of significance of these factors. Since we used a mixed cohort from different datasets e.g. considerable heterogeneity regarding treatment of patients can exist.

Until now both the prognosis and the therapeutic options in TNBC are rather limited.[39,40] In this respect an immunotherapy targeting CT-X antigens as highly specific markers for cancer cells might be an interesting option. This is especially the case since the very subset of tumours expressing these antigens was characterised by an even worse prognosis in our analyses. Today immunotherapy is mostly placed as a possible addendum after initial treatment for patients with minimal residual disease. Several clinical trials of vaccines against members of the MAGE-A and CTAG1B families are already in progress in melanoma, lung cancer, and ovarian cancer.[41–46] Thus it seems a reasonable approach to test these vaccines also in those patients with TNBC showing high expression of the corresponding antigens. Moreover a possible clue of the immune system for these tumours was highlighted by the pronounced prognostic effect of the B-cell metagene in tumours with high MAGE-A expression (Fig. 3D and E). The expression of this B-cell metagene corresponds to the presence of tumour infiltrating lymphocytes,[20] However, as we and others have shown lymphocyte infiltration in breast cancer generally represents a mixture of both B- and T-cells. This is demonstrated by immunohistochemistry and the strong correlation of B-cell and T-cell metagene expression[20] which bear nearly identical information. B-cell and T-cell metagenes can be used as a surrogate marker for infiltration of both types of lymphocytes. Therefore TNBC with MAGE-A expression may benefit the most from further augmentation of the immune response.[11,12] Novel immune stimulatory drugs such as anti-CTLA-4 directed therapies have shown highly promising results especially in melanoma[47] but have also been applied in others cancers[48,49] including breast cancer.[50] These drugs may provide a realistic opportunity to directly test the hypothesis of immune augmentation in TNBC with MAGE-A expression in the clinic.

Conflict of interest statement

None declared.

Acknowledgements

We thank Katherina Kourtis and Samira Adel for expert technical assistance. This work was supported by grants from the the H.W. & J. Hector-Stiftung, Mannheim; the Margarete Bonifer-Stiftung, Bad Soden; the BANSS-Stiftung, Biedenkopf; and the Dr. Robert Pfleger-Stiftung, Bamberg. These foundations had no role in planning of the study and writing of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2011.06.025.

REFERENCES

1. Ertel A, Tozeren A. Switch-like genes populate cell communication pathways and are enriched for extracellular proteins. *BMC Genomics* 2008;**9**:3.
2. Teschendorff AE, Naderi A, Barbosa-Morais NL, Caldas C. PACK: Profile Analysis using Clustering and Kurtosis to find molecular classifiers in cancer. *Bioinformatics* 2006;**22**(18):2269–75.
3. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol* 2007;**8**(8):R157.
4. Ertel A. Bimodal gene expression and biomarker discovery. *Cancer Inform* 2010;**9**:11–4.
5. Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR. The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Inform* 2009;**7**:199–216.
6. Karn T, Metzler D, Ruckhäberle E, et al. Data driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer. *Breast Cancer Res Treat* 2009;**45**:5.

7. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med* 2009;**360**(8):790–800.
8. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 2008;**10**(4):R65.
9. Tabchy A, Valero V, Vidaurre T, et al. Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. *Clin Cancer Res* 2010;**16**(21):5351–61.
10. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* 2005;**5**(8):615–25.
11. Scanlan MJ, Gure AO, Jungbluth AA, Old LJ, Chen YT. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol Rev* 2002;**188**:22–32 [Review].
12. Van Baren N, Bonnet MC, Dréno B, et al. Tumoral and immunologic response after vaccination of melanoma patients with an ALVAC virus encoding MAGE antigens recognized by T cells. *J Clin Oncol* 2005;**23**(35):9008–21.
13. McShane LM, Altman DG, Sauerbrei W, et al. Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. Reporting recommendations for tumor marker prognostic studies. *J Clin Oncol* 2005;**23**(36):9067–72.
14. Affymetrix (2001) Statistical algorithms reference guide, Technical report, Affymetrix.
15. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy – analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004;**20**(3):307–15.
16. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;**5**(10):R80.
17. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;**406**(6797):747–52.
18. Farmer P, Bonnefoi H, Becette V, et al. Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 2005;**24**(29):4660–71.
19. Doane AS, Danso M, Lal P, et al. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* 2006;**25**(28):3994–4008.
20. Rody A, Holtrich U, Pusztai L, et al. T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. *Breast Cancer Res* 2009;**11**(2):R15.
21. Desmedt C, Haibe-Kains B, Wirapati P, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* 2008;**14**(16):5158–65.
22. Schmidt M, Böhm D, von Törne C, et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 2008;**68**(13):5405–13.
23. Bianchini G, Qi Y, Alvarez RH, et al. Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor-positive and -negative cancers. *J Clin Oncol* 2010;**28**(28):4316–23.
24. Farmer P, Bonnefoi H, Anderle P, et al. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med* 2009;**15**(1):68–74.
25. Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, et al. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res* 2009;**69**(10):4116–24.
26. Creighton CJ, Li X, Landis M, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci USA* 2009;**106**(33):13820–5.
27. Whitney AR, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA* 2003;**100**(4):1896–901.
28. Hu Z, Fan C, Livasy C, et al. A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med* 2009;**6**:7–9.
29. Bièche I, Chavey C, Andrieu C, et al. CXC chemokines located in the 4q21 region are up-regulated in breast cancer. *Endocr Relat Cancer* 2007;**14**(4):1039–52.
30. Rody A, Karn T, Liedtke C, et al. Identification of a clinically relevant gene signature in triple negative and basal-like breast cancer. *Cancer Res* 2010;**70**(Suppl. 24) [Abstract No. S5–5].
31. Mischo A, Kubuschok B, Ertan K, et al. Prospective study on the expression of cancer testis genes and antibody responses in 100 consecutive patients with primary breast cancer. *Int J Cancer* 2006;**118**(3):696–703.
32. Bandić D, Juretić A, Sarcević B, et al. Expression and possible prognostic role of MAGE-A4, NY-ESO-1 and HER-2 antigens in women with relapsing invasive ductal breast cancer: retrospective immunohistochemical study. *Croat Med J* 2006;**47**(1):32–41.
33. Taylor M, Bolton LM, Johnson P, Elliott T, Murray N. Breast cancer is a promising target for vaccination using cancer-testis antigens known to elicit immune responses. *Breast Cancer Res* 2007;**9**(4):R46.
34. Sugita Y, Wada H, Fujita S, et al. NY-ESO-1 expression and immunogenicity in malignant and benign breast tumors. *Cancer Res* 2004;**64**(6):2199–204.
35. Otte M, Zafrakas M, Riethdorf L, et al. MAGE-A gene expression pattern in primary breast cancer. *Cancer Res* 2001;**61**(18):6682–7.
36. Theurillat JP, Ingold F, Frei C, et al. NY-ESO-1 protein expression in primary breast carcinoma and metastases: correlation with CD8+ T-cell and CD79a+ plasmacytic/B-cell infiltration. *Int J Cancer* 2007;**120**(11):2411–7.
37. Grigoriadis A, Caballero OL, Hoek KS, et al. CT-X antigen expression in human breast cancer. *Proc Natl Acad Sci USA* 2009;**106**(32):13493–8.
38. Gure AO, Chua R, Williamson B, et al. Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer. *Clin Cancer Res* 2005;**11**(22):8055–62.
39. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 2007 Apr;**13**(8):2329–34.
40. Irvin Jr WJ, Carey LA. What is triple-negative breast cancer? *Eur J Cancer* 2008;**44**(18):2799–805.
41. Bender A, Karbach J, Neumann A, et al. LUD 00–009 phase 1 study of intensive course immunization with NY-ESO-1 peptides in HLA-A2 positive patients with NY-ESO-1-expressing cancer. *Cancer Immun* 2007;**7**:16.
42. Atanackovic D, Altorki NK, Cao Y, et al. Booster vaccination of cancer patients with MAGE-A3 protein reveals long-term immunological memory or tolerance depending on priming. *Proc Natl Acad Sci USA* 2008;**105**(5):1650–5.
43. Jager E, Karbach J, Gnjatic S, et al. Recombinant vaccinia/fowlpox NY-ESO-1 vaccines induce both humoral and cellular NY-ESO-1-specific immune responses in cancer patients. *Proc Natl Acad Sci USA* 2006;**103**(39):14453–8.
44. Valmori D, Souleimanian NE, Tosello V, et al. Vaccination with NY-ESO-1 protein and CpG in Montanide induces integrated antibody/Th1 responses and CD8 T cells

through cross-priming. *Proc Natl Acad Sci USA* 2007;**104**(21):8947–52.
45. Odunsi K, Qian F, Matsuzaki J, et al. Vaccination with an NY-ESO-1 peptide of HLA class I/II specificities induces integrated humoral and T cell responses in ovarian cancer. *Proc Natl Acad Sci USA* 2007;**104**(31):12837–42.
46. Davis ID, Chen W, Jackson H, et al. Recombinant NY-ESO-1 protein with ISCOMATRIX adjuvant induces broad integrated antibody and CD4(+) and CD8(+) T cell responses in humans. *Proc Natl Acad Sci USA* 2004;**101**(29):10697–702.
47. Eggermont AM, Testori A, Maio M, Robert C. Anti-CTLA-4 antibody adjuvant therapy in melanoma. *Semin Oncol* 2010;**37**(5):455–9. Review.
48. Calabro L, Danielli R, Sigalotti L, Maio M. Clinical studies with anti-CTLA-4 antibodies in non-melanoma indications. *Semin Oncol* 2010;**37**(5):460–7.
49. Carthon BC, Wolchok JD, Yuan J, et al. Preoperative CTLA-4 blockade: tolerability and immune monitoring in the setting of a presurgical clinical trial. *Clin Cancer Res* 2010;**16**(10):2861–71.
50. Vonderheide RH, LoRusso PM, Khalil M, et al. Tremelimumab in combination with exemestane in patients with advanced breast cancer and treatment-associated modulation of inducible costimulator expression on patient T cells. *Clin Cancer Res* 2010;**16**(13):3485–94.