Research Article

Evaluation of auto-antibody serum biomarkers for breast cancer screening and *in silico* analysis of sero-reactive proteins

Parvez Syed[1], Klemens Vierlinger[1], Albert Kriegner[1], Khulan Sergelen[1], Johana Luna-Coronell[1], Christine Rappaport-Fürhauser[2], Christa Nöhammer[1], Christian F Singer[2] and Andreas Weinhäusel[1]

[1]Austrian Institute of Technology – AIT, Health & Environment, Molecular Medicine, Vienna, Austria
[2]Department of Obstetrics and Gynaecology, Medical University of Vienna, Vienna, Austria

Received on April 22, 2012; Accepted on June 2, 2012; Published on June 16, 2012

Correspondence should be addressed to Andreas Weinhäusel; Phone: +43 50550445, Fax: +43 505504450, E-mail: andreas.weinhaeusel@ait.ac.at

Abstract

Aberrantly expressed proteins in tumours evoke an immunological response. These immunogenic proteins can serve as potential biomarkers for the early diagnosis of cancers. In this study, we performed a candidate marker screen on macroarrays containing 38,016 human proteins, derived from a human fetal-brain expression library, with the pools of sera from breast cancer patients (1 pool of benign samples, 3 pools of ductal carcinoma and 2 pools of lobular carcinoma) and 1 pool of sera from healthy women. A panel of 642 sero-reactive clones were deduced from these macroarray experiments which include 284 in-frame clones. Over-representation analyses of the sero-reactive in-frame clones enabled the identification of the sets of genes over-expressed in various pathways of the functional categories (KEGG, Transpath, Pfam and GO). Protein microarrays, generated using the His-tag proteins derived from the macroarray experiments, were used to evaluate the sera from breast cancer patients (24 malignant, 16 benign) and 20 control individuals. Using the PAM algorithm we elucidated a panel of 50 clones which enabled the correct classification prediction of 93% of the breast-nodule positive group (benign & malignant) sera from healthy individuals' sera with 100% sensitivity and 85% specificity. This was followed by over-representation analysis of the significant clones derived from the class prediction.

Introduction

Within the European countries in 2008 there were an estimated 3.2 million new cases of cancer and 1.7 million cancer related deaths. Out of the 1.7 million cancer cases, 129,000 (7.5% of all forms of cancer) were cases of breast cancer (Ferlay *et al.* 2010). As a result, there is a great anticipation to identify novel biomarkers for diagnosing breast cancer.

An immunological response can be evoked by a mutated or an aberrantly expressed protein resulting in the production of auto-antibodies. In the context of cancer, these immunogenic proteins are known as tumour-associated antigens (TAA). The corresponding tumour auto-antibodies could be used as biomarkers for early diagnosis and prognosis of cancer (Anderson & LaBaer 2005, Casiano *et al.* 2006, Sanchez-Carbayo 2006). Proteins like ANXA11, p53, HIP1 and ECPKA are known to serve as TAA biomarkers for various cancers (Bradley *et al.* 2005, Fernandez-Madrid *et al.* 2004, Nesterova *et al.* 2006, Soussi 2000). Tomaino *et al*. (2007) used Western blot analysis to identify auto-antibodies against pancreatic ductal adenocarcinoma (PDAC) associated antigens from the PDAC sera. In addition, various studies have elucidated a range of TAAs in breast cancer, such as MUC1, HSP90, HER2/neu, c-myc, NY-ESO-1/LAGE-1 and Lipophilin B (Carter *et al.* 2003, Chapman *et al.* 2007, Conroy *et al.* 1995, Disis *et al.* 1994). However, it has been shown that measurement of a single TAA is neither sensitive nor specific enough to be used as a diagnostic biomarker. Assessment of auto-antibodies to a tailor-made panel of TAAs may have a promising diagnostic potential (Piura & Piura 2011). Various studies have re-

ported panels of TAAs which differentiated the breast cancer patients from healthy controls with higher specificity but low sensitivity (Table 1).

For TAA profiling both macro- and microarrays are used. Macroarrays, blotted onto polyvinylidene fluoride (PVFD) membranes, are spotted with *E. coli* clones expressing recombinant proteins. Using macroarrays (with the hEx1 library), Ludwig *et al*. (2009) could differentiate glioma sera from healthy controls with a specificity and sensitivity of 90.3% and 87.3%, respectively. On the other hand, microarrays are spotted with purified recombinant proteins. Babel *et al.* (Babel *et al.* 2009), used protein microarrays, containing 8000 human GST-tagged proteins, to differentiate sera from 20 colorectal cancer (CRC) patients and healthy individuals. They reported that antibodies against PIM1, MAPKAPK3, STK4, SRC, and FGFR4 were found in high abundance in cancer samples and antibodies against ACVR2B were present in abundance in healthy controls (Babel *et al.* 2009).

In this article we describe the identification of a panel of 642 sero-reactive clones from a collection of 38,016 recombinant protein expressing clones (hEx1 library (Büssow *et al.* 2000)) using macroarrays and sera from breast cancer patients and healthy controls. After identification of the panel of sero-reactive clones we used the "GeneTrail" gene set analysis toolkit to find the genes which are significantly over-represented and are accumulated into certain functional categories (Transpath, Pfam and GO). GeneTrail is an efficient software tool which enables a statistical evaluation of high-throughput genomic or proteomic data sets with regards to the enrichment of functional categories. Furthermore, the genes expressed by the 642 sero-reactive clones were compared to the SEREX (serological expression of cDNA expression libraries) database and their role in cancer is discussed. Using the recombinant proteins derived from the 642 sero-reactive clones, protein microarrays were generated which enabled dis-

Table 1. TAA panels identified in breast cancer patients reported in various studies.

TAA/panel of TAA	Sensitivity (%)	Specificity (%)	Study size	Ages (Mean average in years)		Method used	Ref.
ASB-9 SERAC1 RELT	80	100	87 patients & 87 controls	n.a		cDNA T7 phage library protein screening with ELISA	(Zhong *et al.* 2008)
p16 p53 c-myc	43.9	97.6	41 patients & 82 controls	n.a		ELISA	(Looi *et al.* 2006, Zhong *et al.* 2008)
PPIA PRDX2 FKBP52 MUC1 HSP60	73	85	60 primary breast cancer patients, 82 carcinoma in situ patients & 93 controls	55 (Patients)		ELISA	(Desmetz *et al.* 2009)
p53 c-myc HER2 NY-ESO-1 BRCA2 MUC1	64	85	97 patients & 94 controls	59 (Patients)	54 (Controls)	ELISA	(Chapman *et al.* 2007)
IMP1 p62 Koc p53 c-MYC cyclin B1 survivin	70	95	64 Chinese patients, 82 healthy Chinese controls & 264 healthy USA controls	n.a		ELISA	(Koziol *et al.* 2003, Zhang *et al.* 2003)

Table 2. Clinical and pathological characteristics of the sera used in macro- and microarray screenings. Pools 1-7 were used for the macroarray experiments. Pools 1 and 2 consist of sera from patients with benign fibroadenoma and healthy controls, respectively. Pools 3-5 comprise sera from patients with ductal carcinoma while pools 6 and 7 contain sera from patients with lobular carcinoma. The data enlisted in the columns, Control, Benign and Malignant, are the samples used for microarray experiments.

	Pool 1	**Pool 2**	**Pool 3**	**Pool 4**	**Pool 5**	**Pool 6**	**Pool 7**	**Control**	**Benign**	**Malignant**
Number of samples	10	10	10	10	10	10	10	20	16	24
Median age (years)	43	73	71	57.5	65.5	54	63	77	45[a]	60
Grading[b]										
G1			5			3				6
G2			5	10		7	9			11
G3					10		1			5
P53 Positive			1	1	9	2	1			
Hormone receptor positive										
Her2/neu				2	3		2			8
Estrogen			10	10	1	10	10			18
Progesterone			10	8		10	9			
pT stage[c](%)										
Tx, Tis, T1; T1a, T1b, T1c, T1mic, T2, T3; T4b			0; 0; 0; 20; 30; 40; 10; 0; 0; 0.	0; 10; 0; 0; 0; 10; 0; 60; 0; 10.	0; 0; 0; 10; 0; 30; 10; 40; 0; 0.	0; 0; 0; 0; 20; 50; 0; 10; 10; 0.	0; 0; 0; 0; 0; 20; 0; 60; 20; 0.			4.17; 4.17; 16.67; 29.17; 4.17; 12.50; 0; 4.17
pN stage[d](%)										
Nx; N0; N1; N1a, N1biv; N1mi; N2a; N3			0; 90; 0; 0; 0; 0; 0; 0.	10; 0; 10; 20; 10; 0; 30; 10	0; 60; 0; 0; 0; 0; 20; 10	0; 90; 0; 0; 0; 0; 0; 0.	0; 0; 10; 50; 0; 10; 0; 0.			20; 50; 10; 10; 10
Menopause status[e]										
Pre-menopause	5		3	2	2	3	1		5	3
Peri-menopause						1				
Post-menopause	1		7	7	7	5	9			18

[a]Data available for 14 patients. [b]Data available for 22 malignant patients used in microarray experiments. G1 (low-grade), G2 (intermediate grade) and G3 (high-grade). Low-grade tumours are usually slow growing and are less likely to spread. High-grade tumours are likely to grow more quickly and are more likely to spread. [c]Data available for 24 malignant patients used in microarray experiments. [d]Data available for all patients (40 samples, Pools 3-6) and 9 samples from Pool 7; used in macroarray experiments and data available for 20 Malignant patients. [e]Data available for 47 patient (Pools 3-7) and 6 benign samples (Pool 1); used in macroarray experiments and data available for 26 patients; used in microarray experiments.

tinguishing serum samples from breast-nodule positive patients (benign and malignant) and healthy controls.

Materials and Methods

Serum Samples

Serum samples were obtained after approval from patients and healthy women and were stored at -80°C. The study was approved by the Ethics Committee of the Medical University of Vienna and the General Hospital of Vienna (study number: 143/2007). For macroarray experiments, an aliquot (80 µL) of each serum sample was used for the generation of 7 serum pools. For microarray experiments, 60 serum samples (malignant *n*=24; benign *n*=16; healthy *n*=20) were used. The pathological and clinical cohort characteristics of the breast cancer samples can be found in Table 2.

Candidate marker screening

Protein macroarrays, containing duplicates of 38,016 clones (hEx1 library) were purchased from RZPD (now Source Bioscience), Germany. The protein features were generated by expression of spotted *E. coli* clones, which harbour an expression vector, pQE30NST. The expressed recombinant proteins are His-Tagged. Duplicate clones are present on a set of 2 macroarrays and the macroarrays were processed according the detailed protocol for membrane processing which can be found on the Source Bioscience homepage (http://www.lifesciences.sourcebioscience.com/media/290406/sbs_ig_manual_proteinarray_v1.pdf).

In a pre-test, the reliability of auto-antibody screening on PVFD membranes containing 38,016 fetal brain proteins was evaluated using the native serum samples and the IgG-purified serum fraction isolated by affinity purification of immunoglobulins. The purification of IgG from the serum was done using Melon™ Gel IgG Purification Kit (Thermo Scientific) and the procedure was followed as per the manufacturer's instructions. In this pre-test, an individual serum sample was tested against a pool of 10 healthy control serum samples (also including the single individual sample) with and without the Melon™ Gel IgG Purification in order to decide whether to apply serum or the affinity enriched Ig-fraction onto the macroarrays.

Based on the results derived from the pre-test we decided to use the pools of native serum samples to perform a candidate marker screen on PVFD membranes containing 38,016 human proteins derived from hEx1, a human fetal-brain expression library. In order

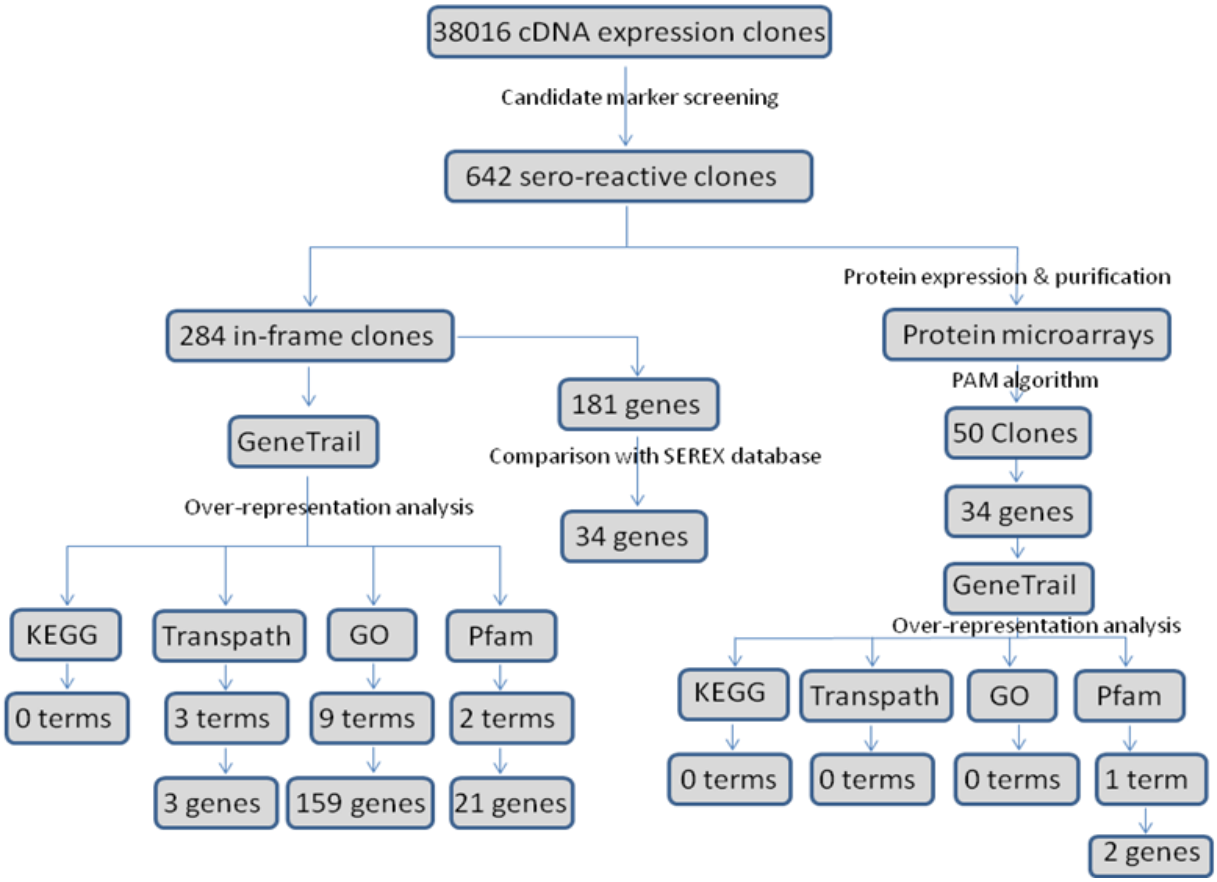

Figure 1. An overview of the number of clones and genes identified in this study.

to have a measure of the reproducibility of the macroarrays, all the membranes were hybridized with a male-serum sample (with no personal or familial breast cancer history). The membranes were then stripped and blinded duplicates of each pool of patient (Pool 3-7) and non-malignant (Pool 1 & 2) sera were applied onto the macroarrays. Thereafter, data was generated upon signal detection according to the protocol from RZPD, Germany.

The selection of the clones was done on the basis of sero-reactivity in all experiments. A total of 642 sero-reactive clones (after excluding duplicates) from different screening experiments were considered for the production of microarrays.

GeneTrail analysis

GeneTrail analysis was done for 284 in-frame clones among the panel of 642 sero-reactive clones. A statistical approach of Over-Representation Analysis (ORA) was followed for the comparison of the test set with the reference set ("Heidelberg human fetal brain"), provided by the gene set analysis tool (Backes *et al.* 2007, Keller *et al.* 2007). The analyses were performed with the following parameters: Multiple testing adjustment method: false discovery rate (FDR), significance level threshold (α-level): 0.05.

Protein microarray production and processing

E. coli clones were cultured using the autoinduction protocol according to Stempfer *et al.* (2010). Recombinant protein expression was induced by cultivation of *E. coli* clones in autoinduction medium (SB medium) and purified using Ni-NTA agarose (Qiagen). Elution of His-Tag proteins was performed using elution buffer (50 mM KH_2PO_4 and 50 mM K_2HPO_4, pH 8.0, 500 mM imidazole, 0.01% SDS and 0.01% NaN3). Purified His-Tag proteins were then spotted on AR-Chip Epoxy slides (Preininger *et al.* 2004). Each microarray consisted of 4 sub-arrays with protein antigens printed in duplicates. Clarified *E. coli* lysate with a concentration of 0.5 mg/mL was used as a positive control and plain buffer spots as a negative control. Processing of the protein microarrays was performed as described previously (Stempfer *et al.* 2010). The processed microarray images were captured using an Axon Genepix 4000A microarray scanner (Molecular Devices, Union City, CA). Median fluorescence intensities after subtraction of local background were calculated from the scanned array images and used for the data analysis.

Statistical analysis

The statistical analysis of the data from the scanned images of macroarrays was performed using R version 2.10.0 (R Development Core Team 2005). For microarray data analyses in addition to R, BRB-ArrayTools Version: 3.6.0 - Stable Release (Simon & Lam 2009) were also used.

For class prediction, we used the Prediction Analysis for Microarrays (PAM) algorithm. The PAM algorithm uses the "nearest shrunken centroid" method which identifies a subset of significant genes/clones for the best classification of the samples (Tibshirani *et al.* 2002). Cross-validation of the predicted class and the true class was performed.

Results

In brief, from the collection of 38,016 cDNA expression clones 642 clones were selected based on their sero-reactivity. Over-representation analysis was performed using 284 in-frame clones. Protein microarrays were generated using the purified proteins from the 642 sero-reactive clones. Using these protein microarrays breast-nodule positive samples could be differentiated from healthy controls. A schematic overview of the results obtained during the course of the study is shown in Figure 1.

Evaluation of purified IgG versus serum for membrane screening

Clones on the membranes which were reactive to native serum samples (pooled serum samples and single serum sample) and purified IgG (same as above) were compared. Signals of duplicate spots were counted as positive signals within the colour-range of 0-4 based on the staining intensity of the spots (Figure 1S; see

Table 3. Number of clones with overlapping reactivity within different samples analysed.

	Purified IgG-Single (1)	Purified IgG-Pool (2)	Native sera-Pool (3)	Native serum-Single (4)
Purified IgG- Single (1)	22	7	11	19
Purified IgG- Pool (2)	7	32	21	11
Native sera- Pool (3)	11	21	67	31
Native serum- Single (4)	19	11	31	125

The numbers (1-4) in the brackets correspond to the lanes in the Figure 1S (see supplementary data).

Table 4. Comparison with the SEREX database. The genes encoded by the in-frame clones were compared to genes enlisted in the SEREX database and TAA related published literature.

Gene symbol	Cancer study in SEREX db	TAA study-Cancer	Gene symbol	Cancer study in SEREX db	TAA study-Cancer
ACTG1	Colon, Fibrosarcoma		***MARK3***	Prostate	
*ALDOA**	Breast, lung	Melanoma (Suzuki *et al.* 2010)	***MAZ****	Squamous cell carcinoma, colon adenomacarcinoma,	Hodgkin's disease (Bataller *et al.* 2003)
ANKHD1	Renal cell carcinoma, glioma, prostate		***MRPS24***	Prostate,	
ATP5B	Malignant fibrous histiocytoma		***PDAP1***	Fibrosarcoma	
BAG5	Melanoma		***PRDX1****	Melanoma	Oesophageal squamous cell carcinoma (Zhang *et al.* 2011)
CD320	Prostate cancer		***PRKRA***	Testis	
CDC42BPB	Renal carcinoma RCC, thyroid		***RBM5***	Renal cancer	
*CENPB**	Melanoma	Breast cancer (Atalay *et al.* 2005, Atalay *et al.* 2010), small cell lung cancer (Briasoulis *et al.* 2008)	***RPL5***	Colon cancer	
CKB	Colon adenocarcinoma		***RPS12***	Renal cell carcinoma	
*EEF2**	Head neck cancer	Melanoma, hepatocellular carcinoma (Li *et al.* 2008)	***RPS13***	Testis	
FDFT1	Fibrosarcoma		***RUFY1***	Prostate cancer, stomach cancer	
FKBP3	Stomach cancer, melanoma		***SMARCA4***	Melanoma, prostate	
*GAPDH**	Breast cancer	Melanoma	***STUB1***	Colorectal adenocarcinoma, breast carcinoma, prostate cancer, ovarian cancer, glioma	
HIST1H1C	Testis		***TP53****	Colorectal adenocarcinoma, breast cancer, colon cancer	Hepatocellular carcinoma (Liu *et al.* 2011a), ovarian cancer (Lu *et al.* 2011), lymphocytic leukemia (Messmer *et al.* 2011), breast cancer (Dalifard *et al.* 1999)
HSPH1	Colorectal adenocarcinoma, melanoma, glioma, lung, pancreas adenocarcinoma,		***TRIM21***	Breast cancer	
IDH2	Breast		***TTC3***	Stomach cancer, glioma, prostate cancer	
IK	Testis		***ZNF232***	Breast carcinoma	

*antigens against which auto-antibodies have been reported through various cancer studies

supplementary data). A total of 170 sero-reactive clones were identified during this experiment. 32 and 67 clones reacted positively to pooled purified IgG and native pooled serum samples, respectively, whereas 22 and 125 clones were observed reacting positively to purified IgG and native serum sample, respectively (Table 3). Based on the number of clones showing a positive reaction, we decided to use native sera for membrane screening.

Antigen Identification on Macroarrays

Macroarrays were hybridized with pooled samples (pools 1-7) after being processed with a single serum-control (reference) and then stripped. Hierarchical clustering results of the reference serum sample on different membranes used for sample analysis are shown in Figure 2S (right part; see aupplementary data) and the number of the sero-reactive clones from each membrane can be found in Table 1S (see supplementary data). The correlation coefficient values derived from the processed membranes with the same reference serum range from 0.68 to 0.98. Analysis of signal intensities derived from the membranes, processed with blinded duplicates (Pools 1-7) was performed and sero-reactive clones were identified (left part of Figure 2S; see supplementary data). The correlation coefficients of the two runs of each serum pool (Pool 1-7) on macroarrays were found to be ranging from 0.12 to 0.89.

A total of 1691 sero-reactive clones were found, including the clones identified from the "IgG versus serum" pre-test. Of all these, 642 were identified as unique clones showing sero-reactivity in all the macroarray experiments. 284 out of 642 clones were confirmed (based on their DNA sequence) to be cloned in-frame. Out of the 284 in-frame clones, 71 reacted positively to the serum samples from benign breast cancer patients, while 41 and 133 showed a positive reaction to the serum samples from health control and malignant breast cancer patients, respectively.

We decided to use all the 642 clones found positive within all the experiments for protein expression and thereby use the subsequent proteins for the production of protein microarrays.

In silico analysis of sero-reactive clones

Out of 284 in-frame clones, 181 code for unique proteins. Upon comparison of the 181 genes with 1545 genes from the SEREX database (http://www.licr.org/D_programs/d4a1i_SEREX.php), we found 34 genes over-lapping between the lists. These 34 genes have been reported in the SEREX database from a variety of cancer studies. Among them, 7 genes (ALDOA, CENBP, EEF2, GAPDH, MAZ, PRDX1 and TP53) are reported in various cancer studies as TAAs (Table 4).

Using GeneTrail, *in silico* analysis of the 284 in-frame clone protein sequences (test set) was performed to retrieve information about their functional categories (KEGG, Transpath, Pfam and GO) as well as their sub-categories, protein families, domains and pathways (Table 5). The number of genes annotated in the test set to the selected functional categories was found to be 168, out of 284 sequences, while the number of genes annotated in the "Heidelberg human fetal brain" reference set were 3527, out of 3553 sequences. It was found that the observed number of genes involved in cellular processes and various pathways was higher than expected. For example, the expected number of genes involved in the sub-category "cellular process" was 121 while the observed number was found to be 139 when compared to the reference set, with a p-value of 0.03. This indicates the over-representation of genes involved in the respective functional categories in breast cancer. Some of the sub-categories which were enriched in the test set when compared to the reference set are cellular process (GO), wnt pathway (Transpath) and R3H domain (Pfam). The sum of the genes found over-represented in all the enriched subcategories of Transpath, Pfam and GO were found to be 3, 21 and 159, respectively. No sub-category pertaining to KEGG was found enriched in the test set compared to the reference set. A detailed list of sub-categories, the genes encoded by the sero-reactive clones and the number of expected and observed genes are shown in the Tables 2S, 3S and

Table 5. Prediction of classes (Benign, Malignant and Control) using the classifier from PAM algorithm. A cross-tabulation of the classes in rows (true) versus columns (predicted) and the corresponding sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) is shown in the table.

Class	Benign	Malignant	Normal	Sensitivity	Specificity	PPV	NPV
Benign	2	13	1	0.125	0.932	0.4	0.745
Malignant	2	15	7	0.625	0.556	0.484	0.69
Normal	1	4	15	0.75	0.8	0.652	0.865

Table 6. Classifier clones derived from PAM which correctly classified breast-nodule samples from healthy controls.

Clone	Gene	E value	Control-Intensities	Breast-nodule intensities	Ratio of intensities	Clone	Gene	E value	Control-Intensities	Breast-nodule intensities	Ratio of intensities
MPMGp800O19569	*YBX1*	0	1793	853	2.1	MPMGp800E07573	*NUBP2*	0	1772	1321	1.3
MPMGp800A04578	*HIP1R*	0	1594	865	1.8	MPMGp800K01579	*YBX1*	0	1626	2150	0.8
MPMGp800I19548	*YBX1*	0	1747	2831	0.6	MPMGp800M08528#	*PODXL2*	0	983	1353	0.7
MPMGp800F12540	*PRPF19*	0	975	1405	0.7	MPMGp800F17571	*RDBP*	0	1500	1940	0.8
MPMGp800H22523	*RBM10*	0	684	1043	0.7	MPMGp800G17568	*RNF187*	5E-119	818	1040	0.8
MPMGp800P06511	*YBX1*	0	1303	2006	0.6	MPMGp800K23566	*H2AFY*	0	1716	1326	1.3
MPMGp800I18557	*CPLX2*	0	898	1524	0.6	MPMGp800L15517*			680	837	0.8
MPMGp800C05534*			4670	3077	1.5	MPMGp800B14528#	*LPHN1*	0	6491	4815	1.3
MPMGp800N23548	*YBX1*	0	1684	2591	0.6	MPMGp800D08553	*EIF3C*	0	7569	5638	1.3
MPMGp800P01595*			514	744	0.7	MPMGp800K10577#	*JUP*	0	1280	1624	0.8
MPMGp800J06581	*YBX1*	0	2020	2966	0.7	MPMGp800F05518#	*SPAG7*	0	3939	5041	0.8
MPMGp800H22512#	*CENPB*	0	2616	3923	0.7	MPMGp800C17586	*EEF2*	0	4071	3146	1.3
MPMGp800K07565	*YBX1*	0	1993	2936	0.7	MPMGp800K22574#	*SNX5*	0	979	1175	0.8
MPMGp800H05540	*OSBPL7*	0	6652	9692	0.7	MPMGp800H22541#	*HIST1H1C*	0	1910	2449	0.8
MPMGp800I15594	*ARPP21*	0	1873	2618	0.7	MPMGp800M24582	*PRDX1*	0	1323	1004	1.3
MPMGp800I07520#	*SRRT*	0	1743	2575	0.7	MPMGp800O02506#	*SPAG7*	0	5536	6937	0.8
MPMGp800P13536	*H2AFY*	0	4062	6036	0.7	MPMGp800O13595#	*PKM2*	0	725	608	1.2
MPMGp800P08541#	*TANK*	0	846	1187	0.7	MPMGp800C06602	*CKB*	0	12854	9937	1.3
MPMGp800N08514	*MAZ*	0	3892	2507	1.6	MPMGp800H07541	*MAZ*	0	6275	4741	1.3
MPMGp800L16562#	*SPAG7*	0	6575	9168	0.7	MPMGp800M05558	*PIM3*	0	4082	3033	1.3
MPMGp800G16536*			968	1283	0.8	MPMGp800J24571	*CBLL1*	0	1175	1354	0.9
MPMGp800M18568#	*MAZ*	0	7833	4627	1.7	MPMGp800N11538	*C16orf13*	1.63E-135	1043	1211	0.9
MPMGp800N14581#	*RPS3A*	0	576	774	0.7	MPMGp800E06542#	*MAZ*	0	16454	13877	1.2
MPMGp800J12588#	*SMAR-CA1*		5879	3965	1.5	MPMGp800M08567	*EPB41L3*	0	1042	920	1.1
MPMGp800G05508	*AKR7A2*	0	1315	1736	0.8	MPMGp800K16540*			7281	6380	1.1

#In-frame clones. *Clones whose sequences were not available.

4S (see supplementary data).

Protein microarray analysis

We used the BRB array tools to analyze the data derived from the microarrays processed with patient and healthy control sera. Using the PAM algorithm, we identified 45 significant clones enabling the classification of benign, malignant and control samples (Table 5S; see supplementary data). Out of 16 benign breast cancer samples 13 were predicted as malignant and 1 as control. Out of 24 malignant samples, 15 were correctly identified as malignant and out of 20 control samples, 15 were identified as healthy controls (Table 5).

Since the majority of benign samples were identified as malignant, we decided to compare the breast-nodule positive samples with the healthy controls. We identified 50 significant clones which enabled the classification of breast-nodule positive samples and healthy controls (Table 6). These clones gave 93% correct classification prediction of breast-nodule positive sera from normal sera with 100% sensitivity and 85% specificity. 4 out of 16 control samples were predicted as breast-nodule positive, while all of the 40 breast-nodule positive samples were correctly predicted (Table 7).

Concerning the lists derived using the PAM algorithm, 12 clones were found significant in both. The lists of significant clones were compared to the list of positively reacting clones to breast-nodule positive sera and healthy control sera. 40 clones were found to react positively to the breast-nodule positive sera and 9 reacted positively to the healthy control sera, exclusively. 14 clones reacted positively to sera from both patients and controls.

To find the set of genes among the 34 genes encoded by the 50 significant clones (that gave 93% correct classification prediction) which are over-represented in the functional categories like KEGG, Transpath, Pfam and GO we used GeneTrail with "Heidelberg human fetal brain" as the reference set. The parameters for the analysis were identical to the ones previously used for the analysis of the 284 in-frame clones. The number of genes found annotated within the test set of 43 genes for KEGG, Transpath, GO and Pfam were found to be 7, 1, 27 and 26, respectively. However, no genes related to any of the KEGG, Transpath and GO were found to be over-represented in the test set when compared to the reference set. 2 genes, ARPP21 and SPAG7 were found to be over-represented in the R3H domain sub-category of Pfam (p-value of 0.001). The expected number of genes was 0.05 while the observed number of genes was 2.

Discussion

Over the years, macroarrays spotted with cDNA expression clones have been used for TAA profiling. Macroarrays spotted with hEx1 cDNA expression library clones have been used for the identification of auto-antibodies from patients with glioma, chronic obstructive pulmonary disease (COPD) and Wilm's tumour (Leidinger *et al.* 2009, Schmitt *et al.* 2011). Auto-antibodies are known to be present in the serum prior to the onset of breast, lung and prostate cancer (Abendstein *et al.* 2000, Lubin *et al.* 1995, Trivers *et al.* 1996). This opens up the possibility of using these antibodies as serological tools for the early diagnosis and management of cancer.

We used these macroarrays for identifying a panel of 642 sero-reactive clones from a collection of 38,016 cDNA expression clones. An initial experiment was conducted to check the performance of the macroarrays when hybridized with purified IgG and native serum. We observed that the number of positive clones was higher when using native sera, compared to purified IgG. In this regard, we decided to use native serum samples for TAA profiling.

To test for reproducibility, a reference serum was hybridized on the macroarrays which were then stripped and hybridized with blinded duplicates of serum pools from breast cancer patients and healthy controls (Pools 1-7). Blinded duplicates of the serum pools were used to avoid experimental bias. Signal intensities derived from the sero-reactive clones were used for hierarchical clustering. Although the results from the single control serum analysed on every single membrane did cluster in a distinct tree, the sum of the positive clones detected from each pool in both of the repeated analyses did not cluster with respect to the

Table 7. Prediction of classes (Breast-nodule positive and Control) using the classifier from the PAM algorithm. A cross-tabulation of the classes in rows (true) versus columns (predicted) and the corresponding sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) is shown.

Class	Control	Breast-nodule positive	Sensitivity	Specificity	PPV	NPV
Control	16	4	0.8	1	1	0.909
Breast-nodule positive	0	40	1	0.85	0.93	1

sample groups "normal", "benign" and "5 different pools of ductal and lobular breast tumour" (Pools 1-7) (Figure 2). A total of 642 clones were found positive within all the macroarray experiments (including positive clones detected along the pre-test).

Out of the panel of 642 sero-reactive clones identified from the macroarray experiments, 284 clones are cloned in-frame. 181 proteins were found to be encoded by the 284 clones, out of which 34 protein encoding genes were found to be enlisted in the SEREX database. These genes were reported in the database from various cancer studies. Through literature search we found 7 (*ALDOA, CENPB, EEF2, GAPDH, MAZ, PRDX1* and *TP53*) out of 34 genes to be reported as TAAs against a variety of cancers (Table 5). In a study conducted by Suzuki *et al*. (2010) on the identification of melanoma antigens by a serological proteome approach, 5 genes (ALDOA, EEF2, GAPDH, ENO1 and HNRNP) showed high reactivity in patient sera incubated with G361 cell line protein spots, as compared to melanocytes. In another study antibodies against ALDOA were identified in the sera of patients with Hepatocellular carcinoma (Looi *et al.* 2008). The CENPB gene has been reported to be significantly expressed in autoimmune diseases (Nakano *et al.* 2000) and several studies have shown CENPB along with TP53 to be markedly associated with breast cancer survival and prognosis (Kulic *et al.* 2010). Over -expression of the genes CENBP, MAZ and PRDX1 was postulated to be linked to regulation of tumour progression, proliferation and metastasis (Liang *et al.* 2004, Zaytseva *et al.* 2008). PRDX1 was found to be overexpressed in human oesophagus squamous cell carcinoma and MAZ protein isolated from a cerebellar expression library showed significant reactivity against sera from patients with Hodgkin's disease (Bataller *et al.* 2003).

Information on the molecular mechanisms is important in understanding cellular behaviour and in predicting the reasons for dysregulation, which may lead to cancer (Krull *et al.* 2006). *In silico* analysis was performed with the aim of identifying any set of genes, among the genes expressed by the sero-reactive clones, which cluster together in accordance with certain functional categories like Transpath, Pfam and GO and are over-represented in breast cancer. Transpath is a database which provides information on signalling molecules, their reactions and the pathways these molecules are involved in (Schacherer *et al.* 2001). KEGG is a collection of databases related to genomes, enzymatic pathways and biological chemicals in the cells (Kanehisa *et al.* 2004). Pfam is a database of protein families based on multiple sequence alignments and profile hidden Markov models (Bateman *et al.* 2004, Liu *et al.* 2011b). GO is an initiative which helps to standardize the representation of genes and gene product attributes across species and databases (The Gene Ontology Consortium 2000). GO provides structured ontologies which classify the gene products with regards to biological processes, cellular components and molecular functions irrespective of species (Lee *et al.* 2007). In a meta-analysis conducted by Chopra, global cancer maps for KEGG, GO and Pfam were created based on 23 breast cancer microarray expression data sets. These maps revealed "hotspots" of activation and de-activation of breast cancer (Chopra 2009).

In order to have a better understanding of the genes/proteins encoded by the sero-reactive clones and their overexpression in various pathways, we employed a web based toolkit called GeneTrail. We compared the 284 in-frame clones (test set) with a reference set ("Heidelberg human fetal brain"). No genes were found to be over-represented in any of the KEGG pathways in the test. A significant over-representation of the genes involved in various enriched subcategories of Pfam, Transpath and GO was observed. A detailed list of over-represented genes pertaining to the pathways and protein families are shown in the supplementary data.

The duplicates of the macroarrays processed with the same serum samples identified a varying number of positive clones (Table 1S; see supplementary data) showing limited reproducibility. The membranes were purchased and were produced so that the spotted cDNA expression clones are grown on the membranes. Recombinant protein expression is induced directly on these membranes and protein immobilization is performed upon lysis of the bacterial cell. It may be presumed that macroarrays, despite being derived from the same batch, present higher variability compared to arrays printed with formerly purified proteins. Although the reproducibility of the macroarrays was not good enough to draw conclusions, we could identify a sizable panel of clones which we used for recombinant protein expression and purification. Protein microarrays serve as a very good alternative to protein macroarrays and have certain advantages. One of them is that the signals derived from macroarrays are not as dynamic as compared to the 16 bit ($0\text{-}2^{16}$) dynamic range of standard microarrays. Only a few microliters (approximately 10 μL) of serum sample are enough for the validation of auto-antibody signatures. In our previous experiment, we observed that the signal patterns obtained by microarray analysis of brain and lung tumour patients' sera were highly reproducible (R=0.92-0.96) (Stempfer *et al.* 2010).

The panel of 642 sero-reactive clones obtained from the macroarray screenings were used for the ex-

pression of His-tag proteins. These recombinant proteins were used for the production of targeted protein microarrays for TAA profiling using serum samples from breast cancer patients (n=24), females with benign fibroadenomas (n=16) and control individuals (n=20). Upon statistical evaluation of the signal intensities derived from the processed microarrays using the PAM algorithm, we could differentiate serum samples obtained from breast-nodule positive patients with 100% sensitivity 85% specificity. When we tried to differentiate all three classes (benign, malignant and healthy controls), we had only 53% correct classification prediction. Furthermore, GeneTrail analysis of the genes expressed by the classifier clones showed enrichment of the R3H domain.

Conclusion

We used macroarrays for a broad screening and could deduce a panel of 642 sero-reactive clones from an expression library consisting of 38,016 recombinant protein expressing clones. *In silico* analysis of the in-frame clones revealed enrichment of functional categories like Transpath, Pfam and GO in breast cancer. Using the recombinant proteins derived from 642 sero-reactive clones we generated a targeted array for TAA profiling using patient sera and controls. With these protein microarrays, breast-nodule positive (benign and malignant) sera could be differentiated from healthy control sera using 50 clones derived from the PAM algorithm.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We want to thank Michael Stierschneider, Agnes Burg, Rudolf Pichler, Silvia Schönthaler and Ronald Kulovics (AIT) for their help with the *E. coli* cultivation, protein isolation and printing microarrays. Above all, we thank the Jubiläumsfonds der Österreichischen National bank for funding this study (project no. 12551).

References

Abendstein B, Marth C, Muller-Holzner E, Widschwendter M, Daxenbichler G & Zeimet AG 2000 Clinical significance of serum and ascitic p53 autoantibodies in epithelial ovarian carcinoma. *Cancer* **88** 1432-1437.

Anderson KS & LaBaer J 2005 The sentinel within: exploiting the immune system for cancer biomarkers. *J.Proteome.Res.* **4** 1123-1133.

Atalay C, Atalay G, Yilmaz KB & Altinok M 2005 The role of anti-CENP-B and anti-SS-B antibodies in breast cancer. *Neoplasma* **52** 32-35.

Atalay C, Dogan L & Atalay G 2010 Anti-CENP-B antibodies are associated with prolonged survival in breast cancer. *Future Oncol* **6** 471-477.

Babel I, Barderas R, az-Uriarte R, Martinez-Torrecuadrada JL, Sanchez-Carbayo M & Casal JI 2009 Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. *Mol Cell Proteomics* **8** 2382-2395.

Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, Muller R, Meese E & Lenhof HP 2007 GeneTrail--advanced gene set enrichment analysis. *Nucleic Acids Res* **35** W186-W192.

Bataller L, Wade DF, Graus F, Rosenfeld MR & Dalmau J 2003 The MAZ protein is an autoantigen of Hodgkin's disease and paraneoplastic cerebellar dysfunction. *Ann Neurol* **53** 123-127.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C & Eddy SR 2004 The Pfam protein families database. *Nucleic Acids Res* **32** D138-D141.

Bradley SV, Oravecz-Wilson KI, Bougeard G, Mizukami I, Li L, Munaco AJ, Sreekumar A, Corradetti MN, Chinnaiyan AM, Sanda MG & Ross TS 2005 Serum antibodies to huntingtin interacting protein-1: a new blood test for prostate cancer. *Cancer Res* **65** 4126-4133.

Briasoulis E, Kamposioras K, Tzovaras V, Pafitanis G, Kostoula A, Mavridis A & Pavlidis N 2008 CENP-B specific anti-centromere autoantibodies heralding small-cell lung cancer. A case study and review of the literature. *Lung Cancer* **60** 302-306.

Büssow K, Nordhoff E, Lubbert C, Lehrach H & Walter G 2000 A human cDNA library for high-throughput protein expression screening. *Genomics* **65** 1-8.

Carter D, Dillon DC, Reynolds LD, Retter MW, Fanger G, Molesh DA, Sleath PR, McNeill PD, Vedvick TS, Reed SG, Persing DH & Houghton RL 2003 Serum antibodies to lipophilin B detected in late stage breast cancer patients. *Clin Cancer Res* **9** 749-754.

Casiano CA, Mediavilla-Varela M & Tan EM 2006 Tumor-associated antigen arrays for the serological diagnosis of cancer. *Mol Cell Proteomics* **5** 1745-1759.

Chapman C, Murray A, Chakrabarti J, Thorpe A, Woolston C, Sahin U, Barnes A & Robertson J 2007 Autoantibodies in breast cancer: their use as an aid to early diagnosis. *Ann Oncol* **18** 868-873.

Chopra P: Data Mining Techniques to Enable Large-scale Exploratory Analysis of Heterogeneous Scientific Data. PhD Thesis. North Carolina State University, Department of Computer Science; 2009.

Conroy SE, Gibson SL, Brunstrom G, Isenberg D, Luqmani Y & Latchman DS 1995 Autoantibodies to 90 kD heat-shock protein in sera of breast cancer patients. *Lancet* **345** 126.

Dalifard I, Daver A & Larra F 1999 Cytosolic p53 protein and serum p53 autoantibody evaluation in breast cancer.

Comparison with prognostic factors. *Anticancer Res* **19** 5015-5022.
Desmetz C, Bascoul-Mollevi C, Rochaix P, Lamy PJ, Kramar A, Rouanet P, Maudelonde T, Mange A & Solassol J 2009 Identification of a new panel of serum autoantibodies associated with the presence of in situ carcinoma of the breast in younger women. *Clin Cancer Res* **15** 4733-4741.
Disis ML, Calenoff E, McLaughlin G, Murphy AE, Chen W, Groner B, Jeschke M, Lydon N, McGlynn E, Livingston RB & . 1994 Existent T-cell and antibody immunity to HER -2/neu protein in patients with breast cancer. *Cancer Res* **54** 16-20.
Ferlay J, Parkin DM & Steliarova-Foucher E 2010 Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer* **46** 765-781.
Fernandez-Madrid F, Tang N, Alansari H, Granda JL, Tait L, Amirikia KC, Moroianu M, Wang X & Karvonen RL 2004 Autoantibodies to Annexin XI-A and Other Autoantigens in the Diagnosis of Breast Cancer. *Cancer Res* **64** 5089 -5096.
Kanehisa M, Goto S, Kawashima S, Okuno Y & Hattori M 2004 The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32** D277-D280.
Keller A, Backes C & Lenhof HP 2007 Computation of significance scores of unweighted Gene Set Enrichment Analyses. *BMC Bioinformatics* **8** 290.
Koziol JA, Zhang JY, Casiano CA, Peng XX, Shi FD, Feng AC, Chan EK & Tan EM 2003 Recursive partitioning as an approach to selection of immune markers for tumor diagnosis. *Clin Cancer Res* **9** 5120-5126.
Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O & Wingender E 2006 TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res* **34** D546-D551.
Kulic A, Sirotkovic-Skerlev M, Jelisavac-Cosic S, Herceg D, Kovac Z & Vrbanec D 2010 Anti-p53 antibodies in serum: relationship to tumor biology and prognosis of breast cancer patients. *Med Oncol* **27** 887-893.
Lee D, Redfern O & Orengo C 2007 Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol* **8** 995-1005.
Leidinger P, Keller A, Heisel S, Ludwig N, Rheinheimer S, Klein V, Andres C, Hamacher J, Huwer H, Stephan B, Stehle I, Lenhof HP & Meese E 2009 Novel autoantigens immunogenic in COPD patients. *Respir Res* **10** 20.
Li L, Chen SH, Yu CH, Li YM & Wang SQ 2008 Identification of hepatocellular-carcinoma-associated antigens and autoantibodies by serological proteome analysis combined with protein microarray. *J Proteome Res* **7** 611-620.
Liang QJ, Lu XF, Cheng XL, Luo S, He DC & Wang YC 2004 [The active expression of CenpB, a constitutive protein in the centromeres of chromosomes, in breast cancer tissues]. *Yi.Chuan Xue Bao.* **31** 236-240.
Liu H, Zhang J, Wang S, Pang Z, Wang Z, Zhou W & Wu M 2011a Screening of autoantibodies as potential biomarkers for hepatocellular carcinoma by using T7 phase display system. *Cancer Epidemiol* doi:10.1016/ j.canep.2011.04.001.
Liu X, Lv B & Guo W 2011b The size distribution of protein families within different types of folds. *Biochem Biophys Res Commun* **406** 218-222.
Looi K, Megliorino R, Shi FD, Peng XX, Chen Y & Zhang JY 2006 Humoral immune response to p16, a cyclin-dependent kinase inhibitor in human malignancies. *Oncol-Rep* **16** 1105-1110.
Looi KS, Nakayasu ES, Diaz RA, Tan EM, Almeida IC & Zhang JY 2008 Using proteomic approach to identify tumor -associated antigens as markers in hepatocellular carcinoma. *J Proteome Res* **7** 4004-4012.
Lu D, Kuhn E, Bristow RE, Giuntoli RL, Kjaer SK, Shih I & Roden RB 2011 Comparison of candidate serologic markers for type I and type II ovarian cancer. *Gynecol Oncol* **122** 560-566.
Lubin R, Zalcman G, Bouchet L, Tredanel J, Legros Y, Cazals D, Hirsch A & Soussi T 1995 Serum p53 antibodies as early markers of lung cancer. *Nat Med* **1** 701-702.
Ludwig N, Keller A, Heisel S, Leidinger P, Klein V, Rheinheimer S, Andres CU, Stephan B, Steudel WI, Graf NM, Burgeth B, Weickert J, Lenhof HP & Meese E 2009 Improving seroreactivity-based detection of glioma. *Neoplasia* **11** 1383-1389.
Messmer BT, Nour-Omid TS, Ghia E, Sanchez AB & Kipps TJ 2011 Autoantibodies against p53 are associated with chromosome 17p deletions in chronic lymphocytic leukemia. *Leuk Res* **35** 965-967.
Nakano M, Ohuchi Y, Hasegawa H, Kuroda T, Ito S & Gejyo F 2000 Clinical significance of anticentromere antibodies in patients with systemic lupus erythematosus. *J Rheumatol* **27** 1403-1407.
Nesterova MV, Johnson N, Cheadle C, Bates SE, Mani S, Stratakis CA, Khan IU, Gupta RK & Cho-Chung YS 2006 Autoantibody cancer biomarker: extracellular protein kinase A. *Cancer Res* **66** 8971-8974.
Piura E & Piura B 2011 Autoantibodies to tailor-made panels of tumor-associated antigens in breast carcinoma. *J Oncol* **2011** 982425.
Preininger C, Bodrossy L, Sauer U, Pichler R & Weilharter A 2004 ARChip epoxy and ARChip UV for covalent on-chip immobilization of pmoA gene-specific oligonucleotides. *Anal Biochem* **330** 29-36.
R Development Core Team 2005 R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.
Sanchez-Carbayo M 2006 Antibody arrays: technical considerations and clinical applications in cancer. *Clin Chem* **52** 1651-1659.
Schacherer F, Choi C, Gotze U, Krull M, Pistor S & Wingender E 2001 The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics* **17** 1053-1057.
Schmitt J, Heisel S, Keller A, Leidinger P, Ludwig N, Habel N, Furtwangler R, Nourkami-Tutdibi N, Wegert J, Grundy P, Gessler M, Graf N, Lenhof HP & Meese E 2011 Multicenter study identified molecular blood-born protein signatures for Wilms Tumor. *Int J Cancer* doi:10.1002/ijc.26419.
Simon R & Lam A 2009 *BRB-ArrayTools User Guide, ver-*

sion 3.2. Biometric Research Branch, National Cancer Institute. http://linus.nci.nih.gov/brb. Biometric Research Branch, National Cancer Institute. http://linus.nci.nih.gov/brb.

Soussi T 2000 p53 Antibodies in the sera of patients with various types of cancer: a review. *Cancer Res* **60** 1777-1788.

Stempfer R, Syed P, Vierlinger K, Pichler R, Meese E, Leidinger P, Ludwig N, Kriegner A, Nohammer C & Weinhausel A 2010 Tumour auto-antibody screening: performance of protein microarrays using SEREX derived antigens. *BMC Cancer* **10** 627.

Suzuki A, Iizuka A, Komiyama M, Takikawa M, Kume A, Tai S, Ohshita C, Kurusu A, Nakamura Y, Yamamoto A, Yamazaki N, Yoshikawa S, Kiyohara Y & Akiyama Y 2010 Identification of melanoma antigens using a Serological Proteome Approach (SERPA). *Cancer Genomics Proteomics* **7** 17-23.

The Gene Ontology Consortium 2000 Gene ontology: tool for the unification of biology. *Nat Genet* **25** 25-29

Tibshirani R, Hastie T, Narasimhan B & Chu G 2002 Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U.S.A* **99** 6567-6572.

Tomaino B, Cappello P, Capello M, Fredolini C, Ponzetto A, Novarino A, Ciuffreda L, Bertetto O, De AC, Gaia E, Salacone P, Milella M, Nistico P, Alessio M, Chiarle R, Giuffrida MG, Giovarelli M & Novelli F 2007 Autoantibody signature in human ductal pancreatic adenocarcinoma. *J Proteome Res* **6** 4025-4031.

Trivers GE, De B, V, Cawley HL, Caron G, Harrington AM, Bennett WP, Jett JR, Colby TV, Tazelaar H, Pairolero P, Miller RD & Harris CC 1996 Anti-p53 antibodies in sera from patients with chronic obstructive pulmonary disease can predate a diagnosis of cancer. *Clin Cancer Res* **2** 1767-1775.

Zaytseva YY, Wang X, Southard RC, Wallis NK & Kilgore MW 2008 Down-regulation of PPARgamma1 suppresses cell growth and induces apoptosis in MCF-7 breast cancer cells. *Mol Cancer* **7** 90.

Zhang J, Wang K, Zhang J, Liu SS, Dai L & Zhang JY 2011 Using proteomic approach to identify tumor-associated proteins as biomarkers in human esophageal squamous cell carcinoma. *J Proteome Res* **10** 2863-2872.

Zhang JY, Casiano CA, Peng XX, Koziol JA, Chan EK & Tan EM 2003 Enhancement of antibody detection in cancer using panel of recombinant tumor-associated antigens. *Cancer Epidemiol Biomarkers Prev* **12** 136-143.

Zhong L, Ge K, Zu JC, Zhao LH, Shen WK, Wang JF, Zhang XG, Gao X, Hu W, Yen Y & Kernstine KH 2008 Autoantibodies as potential biomarkers for breast cancer. *Breast Cancer Res* **10** R40.