High-throughput detection of fusion genes in cancer using the Sequenom MassARRAY platform

Maryou BK Lambros^{1,*}, Paul M Wilkerson^{1,*}, Rachael Natrajan¹, Neill Patani¹, Vidya Pawar¹, Radost Vatcheva¹, Marthe Mansour¹, Mirja Laschet², Beatrice Oelze², Nicholas Orr³, Susanne Muller² and Jorge S Reis-Filho¹

Fusion genes have pivotal roles in the development and progression of human cancer and offer potential for rational drug design. Massively parallel sequencing has identified a panoply of in-frame expressed fusion genes, but early reports suggest that the majority of these are present at very low prevalence or are private events. Conventional methods for the identification of recurrent expressed fusion genes in large cohorts of cancers (eg fluorescence *in situ* hybridization (FISH) and reverse transcriptase PCR (RT–PCR)) are time consuming and prone to artifacts. Here, we describe a novel high-throughput strategy for the detection of recurrent fusion genes in cancer based on the Sequenom MassARRAY platform. Fusion genes were initially identified by massively parallel sequencing of breast cancer cell lines. For each fusion gene, two Sequenom probes were designed. Primary human breast cancers and cancer cell lines were interrogated for 10 fusion genes. Sensitivity, specificity, and predictive values of the MassARRAY method were then determined using FISH and qRT–PCR as the 'gold standard.' By combining two probes per fusion gene, the negative and positive predictive values were 100 and 71.4%, respectively. All fusion genes identified by massively parallel sequencing were accurately detected. No recurrent fusion genes were found. The MassARRAY-based approach described here may, therefore, be employed as a high-throughput screening tool for known fusion genes in human cancer. In keeping with other highly sensitive assays, further refinement of this technique is necessary to reduce the number of false-positive results. *Laboratory Investigation* (2011) **91**, 1491–1501; doi:10.1038/labinvest.2011.110; published online 1 August 2011

KEYWORDS: breast cancer; fusion gene; genetics; high-throughput; MassARRAY; Sequenom

The identification of oncogenic fusion genes in cancer provides an opportunity for rational drug design, and is exemplified by chronic myeloid leukemia, where neoplastic cells expressing the BCR-ABL fusion gene are specifically sensitive to imatinib.¹ With the advent of massively parallel sequencing, it has become apparent that chromosomal rearrangements are not limited to hematological malignancies and sarcomas, but are in fact common events in carcinomas.^{2–10} For example, fusion genes involving ETS-family members, such as ERG (SLC45A3-ERG), ETV1 (TMPRSS2-ETV1), ETV4 (TMPRSS2-ETV4), and ETV5 (SLC45A3-ETV5), have been shown to be common in prostate carcinomas.^{2,6,11-16} Some special types of breast cancer have recently been shown to be characterized by the presence of recurrent fusion genes resulting from chromosomal translocations, including ETV6-NTRK3 in secretory carcinomas^{5,17–21} and MYB-NFIB in adenoid cystic carcinomas.4,22-24 The

majority of fusion genes identified by massively parallel sequencing, however, appear to exist at very low prevalence or represent private events (ie identified only in the index case).^{7,9,10,14,15} This observation and the fact that chromosomal rearrangements are more frequently identified within amplified regions of the genome have led some to conclude that the majority of these rearrangements are likely to constitute passenger events.^{7,8}

The probability of a given expressed fusion gene to constitute an oncogenic driver is higher if it is recurrently found in cancers.⁷ Traditionally, testing the recurrence of known expressed fusion genes was based on a combination of fluorescence *in situ* hybridization (FISH) and reverse transcription PCR (RT–PCR). In comparison to contemporary analytical studies, these techniques are labor intensive, time consuming (eg FISH on tissue microarrays, including probe

Received 13 May 2011; revised 25 May 2011; accepted 26 May 2011

¹Molecular Pathology Team, The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK; ²Sequenom GmbH, Hamburg, Germany; ³Genetic Epidemiology Group, The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK

Correspondence: Professor JS Reis-Filho, MD, PhD, FRCPath, The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.

E-mail: jorge.reis-filho@icr.ac.uk

^{*}These authors contributed equally to this work

design and validation, and scoring) and/or prone to false-positive results (eg mis-priming and amplification of intronic sequences with RT–PCR, particularly when using DNA extracted from formalin-fixed paraffin-embedded (FFPE) material).

MassARRAY technology (iPLEX®) constitutes a highly sensitive, high-throughput tool for the genotyping of singlenucleotide polymorphisms²⁵⁻³⁰ and for mutation screening.^{31–33} To detect nucleic acid changes, target regions are initially amplified using multiplex PCR and subsequently hybridized to custom-designed primers, then subjected to a single base extension reaction using single mass-modified nucleotides. Spotting of the products onto a matrix chip and subsequent ionization then enables real-time detection by the MassARRAY mass spectrometer (Sequenom, San Diego, CA, USA). This process is based on the principle of matrixassisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Briefly, laser energy is absorbed by the matrix, resulting in partial vaporization of the illuminated substrate, with minimal damage or fragmentation. Ionized samples are then transferred electrostatically into the mass spectrometer, which enables detection by mass-to-charge ratio. This platform has some advantages over conventional detection systems. Product mass is directly determined with high analytical accuracy (0.1-0.01%), without reliance upon the indirect analysis of fluorescent or radioactive reporters. The high-throughput nature of the technology is due to a combination of the ability to multiplex PCR reactions (up to 36-plex, with as little as 10 ng per multiplex), and rapid analysis. Furthermore, small PCR amplicons (70-120 bp) are suitable for the assessment of FFPE tissue, facilitating the identification of recurrent fusion genes in large annotated cohorts.

Given the versatility of the MassARRAY technology, we have successfully developed a protocol for the detection of known, expressed fusion genes. The aims of this proof-ofprinciple study were (i) to develop a robust strategy for the high-throughput detection of known, expressed fusion genes, (ii) to test this protocol in a cohort of 10 cancer cell lines known to harbor fusion genes previously detected by massively parallel sequencing, and (iii) to apply these techniques to detect fusion genes in a cohort of FFPE breast cancers.

MATERIALS AND METHODS

Cancer Cell Lines and Human Breast Cancer Tissues

Nine breast cancer cell lines (HCC1143, HCC1187, HCC1395, HCC1599, HCC1937, HCC1954, HCC2157, HCC2218, and HCC38) and HeLa cells, which were used as a negative control for the known fusion genes, were obtained from ATCC (see Stephens *et al*⁷ for phenotypic characteristics and growth conditions). DNA samples from these cell lines were previously subjected to massively parallel paired-end DNA sequencing.⁷ Ten fusion genes were identified among five breast cancer cell lines (HCC1187, HCC1395, HCC1599, HCC1937, and HCC38) and validated using RT–PCR and

FISH (Table 1). Cell pellets of two cell lines (HCC38 and HCC1937) were formalin fixed overnight and embedded in paraffin following routine protocols to simulate the same sample preparation that FFPE human breast cancer tissues undergo.

To evaluate this novel assay in primary human breast cancers, 125 FFPE invasive breast cancer samples were randomly selected from a cohort of 245 cases that have been previously described.³⁴ These samples were predominantly estrogen and progesterone receptor positive (86 and 78%, respectively), did not express HER2 (89%), and were of high histological grade (61% grade III; Supplementary Table 1). Two of the 10 fusion genes examined (NFIA-EHF and SLC26A6-PRKAR2A) were previously investigated in the 125 samples assessed by MassARRAY in this study by FISH and RT-PCR.⁷ They were, therefore, included in the assessment of the performance of this assay along with the detection of all 10 fusion genes in the panel of 10 cancer cell lines (nine breast cancer and HeLa cells). The study design, including the cell lines used for each aspect of the assessment of probe performance, is illustrated in Figure 1. This study is compliant with the required sections of the REMARK guidelines.35

RNA Extraction and RT

Representative 8 μ m thick FFPE sections of the breast tumor samples were microdissected to ensure >70% tumor cell content as previously described.³⁶ Total RNA was extracted from breast cancer tissues using the RNeasy FFPE RNA Isolation Kit (Qiagen) followed by an additional DNase treatment as previously described.³⁷ RNA quantification was performed using the Ribogreen Quant-iT reagent (Invitrogen, UK) according to the manufacturer's instructions. Total RNA was extracted from breast cancer cell lines using the RNAeasy kit (Invitrogen) according to the manufacturer's instructions. RT was performed using Superscript III

Table 1 Summary of cancer cell lines used as positive controlsand the fusion genes expressed

Fusion gene	Positive control	Reference
NFIA-EHF	HCC1937	7
PLXND1-TMCC1	HCC1187	7
CYP39A1-EIF3K	HCC1395	7
EFTUD2-KIF18B	HCC1395	7
KCNQ5-RIMS1	HCC1395	7
PLA2R1-RBMS1	HCC1395	7
ERO1L-FERMT2	HCC1395	7
CYTH1-PRPSAP1	HCC1599	7
MBOAT2-PRKCE	HCC38	7
SLC26A6-PRKAR2A	HCC38	7



Figure 1 Diagrammatic representation of the design of the study. A panel of breast cancer cell lines were previously subjected to massively parallel sequencing. Five cell lines were found to harbor 10 novel fusion genes (red boxes). Four breast cancer cell lines devoid of these 10 fusion genes were used as negative controls in this study, with HeLa cells as a further control (gray box). From these 10 cell lines, cDNA samples from two were pooled in a spiking experiment to simulate a non-modal clone harboring a fusion gene (orange boxes). Cell pellets from two further cell lines were FFPE to assess the performance of the Sequenom fusion probes in tissues prepared in this way (green boxes). Of the 10 fusion genes assessed in this study, two were investigated by means of FISH and RT–PCR in a cohort of 125 breast cancer samples (blue box) and reported elsewhere.⁷ Sequenom MassARRAY fusion assays were designed for all 10 fusion genes, as illustrated in Figure 2. All 10 cell lines were interrogated with these 10 fusion gene assays, while the 125 breast cancer samples were interrogated with RT–PCR and FISH. The resulting data were used to determine the performance of each probe. The 125 breast cancer samples were then subjected to Sequenom MassARRAY analysis using the fusion gene assays for the remaining eight fusion genes as screening tool in this cohort. Positive results using these assays were then validated using RT–PCR.

(Invitrogen) converting 500 ng of total RNA, with triplicate reactions undertaken for each sample as previously described.³⁷ In addition, in all cases of primary breast cancer where MassARRAY analysis yielded positive results, validation of the presence of the fusion gene was performed using not only the RNA sample subjected to MassARRAY analysis, but also with RNA obtained from a re-microdissected sample. To assess the performance of this assay in a biologically relevant scenario, where the fusion gene is harbored by a non-modal tumor cell population within a heterogeneous tumor, cDNA from HCC1937 (known to harbor the *NFIA*-*EHF* fusion gene) was pooled with HCC1954 (devoid of this fusion gene) in serial dilutions of 0, 10, 25, 50, 75, 90, and 100%.

PCR Primer and Extension Probe Design

Five PCR multiplexes were designed using Sequenom's Assay Designer software (available from https://www.mysequenom. com/Home), two multiplexes for probe 1 and three for probe 2. Sequences for initial PCR primers and extension primers for each fusion gene assessed in this study are shown in Table 2, and details of multiplexes, unextended and extended primer masses are detailed in Supplementary Table 2. Amplification primer pairs were designed so the forward and reverse primers hybridize to each partner of the fusion gene (Figure 2). Therefore, in the absence of a fusion gene, no PCR product is available for the single base extension step. Two single base extension probes were designed to detect the presence of each of 10 previously characterized fusion genes. In each case, probe 1 was designed to specifically amplify the region of the breakpoint with an extension primer upstream and downstream, extending a single base beyond the breakpoint to the partner gene (Figure 2). Probe 2 (consisting of a single primer) was designed to hybridize to the breakpoint region, spanning 5-6 bases on either side of the fused genes, and extend to the 3'-end of the fusion gene (Figure 2).

Sequenom MassARRAY Protocol

Initial amplification PCR multiplex reactions were performed in 384-well plates using 2.5 ng of cDNA, 100 nM PCR primers, 2.75 mM MgCl₂, 200 µM dNTP, and 0.1 U of hot-start enzyme (PCR enzyme Kit, Sequenom) per reaction, under the following conditions: 94°C for 5 min, followed by 45 cycles of 94°C (30 s), 56°C (30 s), 72°C (60 s), and a final extension step at 72°C for 3 min. Samples were kept at 4°C until further analysis. Shrimp alkaline phosphatase (SAP) dephosphorylates unincorporated dNTPs by cleaving phosphate groups from the 5'-termini, removing all excess dNTPs. A 2µl mixture, comprising $0.3 \,\mu$ l SAP (1 U/µl), 0.17 µl SAP buffer (10 ×) and 1.53 µl H₂O, was added to the PCR products, which were then subjected to the following conditions: 37°C for 40 min, followed by 85°C for 10 min. Samples were kept at 4°C until further analysis.

Samples were then subjected to the iPLEX extension reaction, during which the primers hybridize to their target regions, which are extended by a single mass-modified nucleotide. Extension primers and multiplexes are detailed in Supplementary Table 2. The final extension reaction contains 0.222 μ l iPLEX buffer (10 ×), 0.2 μ l iPLEX termination mix, 0.041 μ l iPLEX enzyme, 0.4926 μ l H₂O, and 1.0444 μ l of the extension primers at their optimized concentrations (Sequenom). The extension reaction conditions were as follows: 94°C for 30 s, followed by 40 cycles of 94°C for 5 s, followed by 5 cycles of 52°C for 5 s, and 80°C for 5 s and a final step of 72°C for 3 min. The iPLEX reaction products were then treated with a cationic exchange resin for 30 min to remove salts, such as Na⁺, K⁺, and Mg⁺ ions, to minimize background noise. In all, 15 nl of the resulting

	iuit oi uu custoin acsignea i cu pi				
Fusion gene	Forward primer	Reverse primer	Probe 1 forward unextended primer	Probe 1 reverse unextended primer	Probe 2 unextended primer
CYP39A1-EIF3K	ACGTTGGATGGCAGTGCTAAAAAGACATTC	ACGTTGGATGACATTGACCGCTGGCTGCTG	ATTCTCGGGGGATCTGTCG	AAGACATTCTTTGGAATTGATG	GGATCTGTCGGCATCAA
CYTH1-PRPSAP1	ACGTTGGATGTCTTGGCCACGAACAGATTC	ACGTTGGATGACCATGGAGGAGGACGACA	GAGGACGACAGGCTAC	TTCTTTATTTCAACTCTTGTTT	GTACGACAGGCTACGAAACAA
EFTUD2-KIF18B	ACGTTGGATGGCAACGCACAGCACCTTTAG	ACGTTGGATGGGCTGGTCTGCAAAAGTTC	TTTGGCGAGTTCACA	ACCTTTAGGGCCAAC	AGTTCACAGGTTGGC
ERO1L-FERMT2	ACGTTGGATGACTATACAGTCCTGGGCTTG	ACGTTGGATGTGGACCAAGCATGATGATTC	ACTTCTGTGAAGCTGAT	GGCCTGGGCTTGAATATATACTTC	CGTGAAGCTGATGGAAGTA
KCNQ5-RIMS1	ACGTTGGATGAGCTTTCAAACTGTTGATGC	ACGTTGGATGCTGGGCGTTCATCTACCAC	CCTCTACCACGCTTTCG	CAAACTGTTGATGCAATCTC	CCACAACGCTTTCGTGAGATT
MBOAT2-PRKCE	ACGTTGGATGTCCAGAAGGATGTTGTCCAG	ACGTTGGATGCGGAAGGATGAAGAACTGAC	GGACCAGAGGGATTTAGCTGTAA	TTGTCCAGTTTCAAATCC	AGTGATTTAGCTGTAAGGGATTT
NFIA-EHF	ACGTTGGATGTCTCGGTCTTTCCAGGTGTTC	ACGTTGGATGGGAACTCGATTTATATTTGGC	GGCTACTTTGTGCATGCAGCA	GGTGTTCATGATGGAAGGCT	ATGCAGCAGAGCCTT
PLAZR1-RBMS1	ACGTTGGATGAGGTTGGTAGGATCTTGTTC	ACGTTGGATGGAGCAAGAATTCATCCACAG	TCCACAGCAAAATAAAGC	CTITITGTAGGATCTTGTTCCTGTTG	CAAAATAAAGCGCAACAG
PLXND1-TMCC1	ACGTTGGATGCAAGGCTGCTTACTTCCAAC	ACGTTGGATGCCACTGCCTGACAGATACAC	ACACGGTGTCGGAGG	CCTTCCTTCCAACCGTTCGAT	TGTCGGAGGAATCGAA
SLC26A6-PRKAR2A	ACGITGGATGTCCTGAAGTCTGCATCTCTG	ACGTTGGATGAGCAACAACGTTGAGGACTG	GAGGACTGCAAGATGAT	CCCTCAGTTTTAGGATGAATCAC	CTGCAAGATGATGGTGATT

Table 2 Summary of all custom-designed PCR primers and extension primers used



Figure 2 Illustration of high-throughput method for detection of recurrent fusion genes. PCR amplification of a small amplicon from the breakpoint region (a) was followed by several steps of sample preparation (see Materials and Methods) and hybridization with custom-designed extension primers (b). Only in the presence of a fusion gene is a PCR product generated for the extension reaction to take place. Samples were spotted onto a SpectroChip after cation removal (c). Samples were then analyzed by the MALDI-TOF MS system (d). In the final analysis, if the predicted alleles are detected by the probe combination, a fusion gene is present (red arrow), whereas if peaks are seen representing unextended primers (black arrow), with no alleles detected, no fusion gene is present (e).

products were spotted onto the MassARRAY SpectroCHIP II with a nano-dispenser (Sequenom), followed by insertion into the mass spectrometer of the MassARRAY System (Sequenom).

Data Analysis

MassARRAY results were analyzed with the SpectroTYPER 4.0. software. Design information was used to calculate the expected position of the correct analyte in the spectra. In the presence of a targeted fusion gene, PCR products are generated and single base extension will result in an allelespecific peak (Figure 2). The accurate detection of a fusion gene by MassARRAY using this assay requires both the forward and reverse primers of probe 1, and a single primer for probe 2 to detect the expected allele (according to primer design), whereas a negative MassARRAY result is recorded if either primer of probe 1 and/or the single primer of probe 2 fail to detect an allele. True positive results were defined as samples classified as positive by MassARRAY that were previously determined to harbor the same fusion genes by massively parallel sequencing, FISH or RT-PCR. The results of the MassARRAY analysis were interpreted by two observers (MBK and SM) blinded to the results of FISH and RT-PCR.

Sensitivity was defined as the proportion of FISH/RT–PCR positive tests, which were identified as positive by Sequenom MassARRAY. Specificity was defined as the proportion of FISH/RT–PCR negative tests, which were identified as negative by Sequenom MassARRAY. Negative predictive value was defined as the proportion of negative tests by Sequenom MassARRAY that were negative by FISH/RT–PCR. Positive predictive value was defined as the proportion of positive tests by Sequenom MassARRAY that were positive by FISH/RT–PCR.

The sensitivity, specificity, and negative and positive predictive values of this assay were determined by combining the results from MassARRAY analysis of all 10 fusion genes in the 10 cancer cell lines (nine breast cancer cell lines and HeLa cells) and of two fusion genes (ie *NFIA-EHF* and *SLC26A6-PRKAR2A*) in 125 breast cancer samples. These results were compared with the previously determined fusion gene status of these samples by massively parallel sequencing and RT–PCR in the cell lines, or RT–PCR and FISH in the series of primary breast cancers (Figure 1).

RESULTS

MassARRAY analysis identified all known fusion genes correctly (ie 100% sensitivity and negative predictive value) with either probe 1 or probe 2 (Table 3). Importantly, when samples were evaluated with either probe independently, the performance of this assay was suboptimal, given the high number of false-positive results and the limited specificity (95.0 and 94.4% for probe 1 and probe 2, respectively) and positive predictive value (37 and 34.5% for probe 1 and probe 2, respectively; Table 3 and Supplementary Table 3).

Table 3 Performance of the probes in human breast cancer cell lines and tissues.

	Probe 1	Probe 2	Combined probes
True positive	10	10	10
False positive	17	19	4
False negative	0	0	0
True negative	323	321	336
Sensitivity	100.0%	100.0%	100.0%
Specificity	95.0%	94.4%	98.8%
Positive predictive value	37.0%	34.5%	71.4%
Negative predictive value	100.0%	100.0%	100.0%
Accuracy	95.1%	94.6%	98.9%

When the results of the two probes were combined, the specificity (98.8%) and positive predictive value (71.4%) were both increased, without a reduction in sensitivity (100%) or negative predictive value (100%; Table 3).

Given the high sensitivity and negative predictive value of the MassARRAY assay to detect expressed fusion genes, we sought to determine if any of the 10 fusion genes identified in the cell lines by massively parallel sequencing would be present in any of the breast cancers included in this study. First, to ascertain the efficacy of the assay to detect expressed fusion genes in FFPE samples, cDNA derived from FFPE pellets of two breast cancer cell lines harboring three fusion genes (HCC38 harboring MBOAT2-PRKCE and SLC26A6-PRKCE fusion genes, and HCC1937 harboring the NFIA-EHF fusion gene) were used as positive controls. Among the 10 fusion genes assessed, three were present in one of the cell lines, one of which was a false positive (ie HCC1937 does not harbor the PLXND1-TMCC1 fusion gene; Table 4 and Supplementary Table 4). We next investigated the presence of the remaining eight fusion genes in the series of 125 primary breast cancers. Of these eight fusion genes, seven were not detected in any of the 125 breast tumor samples. The remaining fusion gene (CYP39A1-EIF3K) was positive in four cases using this MassARRAY fusion assay. Independent validation with RT-PCR using custom-designed primers, the same RNA subjected to Sequenom MassARRAY assay, and RNA extracted from independent samples of the same cancers failed to validate any of the results as true positives, demonstrating that no primary breast cancer tested harbored any of the fusion genes evaluated (Figure 3; Supplementary Table 5). These observations lend further support to the suggestion that the majority of expressed fusion genes in usual types of breast cancer are private.

To determine if the assay described here could detect a fusion gene expressed by a non-modal cancer cell population, we performed spiking experiments by mixing cDNA from HCC1937 cells, which express the *NFIA-EHF* fusion gene, with cDNA from HCC1954 cells, which do not harbor this

FFPE-treated cell lines $(n = 2)$				Primary breast cancers ($n = 125$)						
Fusion ID	Probe ID	True negative	False negative	True positive	False positive	Fusion ID	Probe ID	Negative	True positive	False positive
SLC26A6-PRKAR2A	Probe 1	1	0	1	0					
	Probe 2	1	0	1	0					
	Combined	1	0	1	0					
NFIA-EHF	Probe 1	1	0	1	0					
	Probe 2	1	0	1	0					
	Combined	1	0	1	0					
MBOAT2-PRKCE	Probe 1	1	0	1	0	MBOAT2-PRKCE	Probe 1	125	0	0
	Probe 2	1	0	1	0		Probe 2	125	0	0
	Combined	1	0	1	0		Combined	125	0	0
CYP39A1-EIF3K	Probe 1	2	0	0	0	CYP39A1-EIF3K	Probe 1	125	0	0
	Probe 2	2	0	0	0		Probe 2	121	0	4
	Combined	2	0	0	0		Combined	125	0	0
PLA2R1-RBMS1	Probe 1	2	0	0	0	PLA2R1-RBMS1	Probe 1	125	0	0
	Probe 2	2	0	0	0		Probe 2	125	0	0
	Combined	2	0	0	0		Combined	125	0	0
KCNQ5-RIMS1	Probe 1	2	0	0	0	KCNQ5-RIMS1	Probe 1	125	0	0
	Probe 2	2	0	0	0		Probe 2	125	0	0
	Combined	2	0	0	0		Combined	125	0	0
ERO1L-FERMT2	Probe 1	2	0	0	0	ERO1L-FERMT2	Probe 1	125	0	0
	Probe 2	2	0	0	0		Probe 2	125	0	0
	Combined	2	0	0	0		Combined	125	0	0
EFTUD2-KIF18B	Probe 1	2	0	0	0	EFTUD2-KIF18B	Probe 1	125	0	0
	Probe 2	2	0	0	0		Probe 2	125	0	0
	Combined	2	0	0	0		Combined	125	0	0
CYTH1-PRPSAP1	Probe 1	2	0	0	0	CYTH1-PRPSAP1	Probe 1	125	0	0
	Probe 2	2	0	0	0		Probe 2	125	0	0
	Combined	2	0	0	0		Combined	125	0	0
PLXND1-TMCC1	Probe 1	0	0	0	2	PLXND1-TMCC1	Probe 1	125	0	0
PLXND1-TMCC1	Probe 2	1	0	0	1		Probe 2	125	0	0
	Combined	1	0	0	1		Combined	125	0	0

Table 4 Screening of eight fusion genes identified by massively parallel sequencing in breast cancer cell lines in 125 FFPE primary breast cancers, and in a FFPE simulation using two cell lines

fusion gene. The Sequenom MassARRAY assay was able to detect a fusion gene expressed by the equivalent of as little as 10% of the cancer cells within a tumor population, in agreement with the results obtained by RT–PCR analysis of the same mixtures of cDNA samples (Figure 4).

This novel fusion gene detection assay has favorable sensitivity and negative predictive value in both cell lines and FFPE breast cancer samples and is able to detect fusion genes in non-modal clonal populations of cancers. Its specificity and positive predictive value, however, are suboptimal.

DISCUSSION

In this proof-of-principle study, we describe a novel application of the Sequenom MassARRAY platform as a robust high-throughput screening tool for the detection of known expressed fusion genes. Probes were designed to specifically target the breakpoint of the fusion transcript. Unlike the DNA breakpoint of recurrent fusion genes, which is specific to each sample, the breakpoint at the transcript level is identical in recurrent expressed fusion genes (eg *ETV6-NTRK3* in secretory carcinomas of the breast³⁸). The use of



Figure 3 Validation of Sequenom MassARRAY results by RT–PCR. All positive Sequenom results were tested by an independent RT–PCRs in this study. When the 125 FFPE breast cancer samples were assessed using Sequenom fusion probes for all 10 fusion genes (ie two genes used to define the sensitivity, specificity, and predictive values, and eight additional fusion genes), 14 cases were recorded as positive using the criteria described in Materials and methods. Independent RT–PCRs demonstrated objectively that these were false-positive results. Of all Sequenom MassARRAY fusion gene assays designed, the probes for *SLC26A6-PRKAR2A* gave the highest number of false-positive results. RT–PCR results from cases considered positive by the Sequenom MassARRAY assay: (a) *SLC26A6-PRKAR2A*, (b) *CYP39A1-EIF3K*, (c) *NFIA-EHF*, (d) *KCNQ5-RIMS1*, (e) *ERO1L-FERMT2*, (f) *CYTH1-PRPSAP1*, and (g) *PLA2R1-RBM1*. Positive controls (ie RNA extracted from the cell line know to express a given fusion gene as defined by massively parallel sequencing) and negative controls (no template control – NTC), including a RT–PCR negative control (RT neg), were included in each experiment. The results of the Sequenom MassARRAY assay for each probe are shown below the gel images for each case.

two probes per fusion gene enabled the sensitivity and negative predictive value to be maximized (both 100%), with acceptable specificity (98.8%), but suboptimal positive predictive value (71.4%). We have also used the assay we developed to investigate the presence of eight fusion genes detected in breast cancer cell lines in a cohort of 125 FFPE primary breast cancers. Although four cases were considered to harbor a fusion gene using either of the two MassARRAY probes described here, they were proven to be false-positive results using an orthogonal method (ie RT–PCR) in two biological replicates of the samples.

One source of artifacts when these assays are applied to RNA extracted from FFPE tissues is the degraded nature of RNA extracted. In fact, one could expect a higher proportion of false-positive results if degraded RNA rather than RNA extracted from cell lines or fresh frozen tissues was used. This hypothesis was not confirmed. Out of 200 assays (with both probes) performed on RNA extracted from all cell lines, 26 were false-positive results (13%), whereas out of the 2500 assays performed on RNA extracted from FFPE samples, only 14 were false-positive results (0.6%), less than would have been expected given the source of the RNA. To confirm that the performance of this assay was not adversely affected by tissue fixation and sample preparation, two breast cancer cell lines known to harbor some of the fusion genes evaluated were subjected to formalin fixation and paraffin embedding. RNA was extracted from these cell lines and subjected to the MassARRAY assay. Both fusion genes were accurately detected. These observations suggest that the assay described here is unlikely to be compromised by tissue fixation and sample preparation.

Although the sensitivity and negative predictive value of the MassARRAY assay developed were optimal (ie no falsenegative results), false-positive results were observed. The source of the false-positive results is yet to be fully elucidated. It is perhaps surprising that a greater number of false-positive tests were found when RNA extracted from cell lines was assessed. Possible explanations for this observation include (i) incomplete removal of DNA from the RNA samples, (ii) the iPLEX reactions may have been saturated with cDNA from the cell lines, and (iii) formation of primer dimers or hairpin formation, particularly in the absence of the target fusion gene. Importantly, though, despite the suboptimal specificity and positive predictive value, the high sensitivity and negative predictive value render the MassARRAY assay described here as a useful screening tool for the identification of expressed fusion genes. Hence, this assay allows large collections of cancer samples to be rapidly and efficiently



Figure 4 Spiking experiments to test the Sequenom MassARRAY assay for the detection of expressed fusion genes in non-modal populations of cancer cells. Pooled cDNA samples from HCC1937 and HCC1954 were mixed in a spiking experiment to simulate the expression of a fusion gene in non-modal cell populations (please see Materials and Methods). The Sequenom MassARRAY assay detected the *NFIA-EHF* fusion gene in samples containing as little as 10% cDNA from HCC1937. Cluster plots including all HCC1937:HCC1954 cDNA mixtures, depicting the results of probe 1 forward (probe 1F) primer, probe 1 reverse (probe 1R) primer, and probe 2 for the *NFIA-EHF* fusion gene (**a**). Mass spectra charts for the detection of the *NFIA-EHF* fusion gene in mixtures of 100% HCC1937 cDNA: 0% HCC1937 cDNA: 0% HCC1937 cDNA: 0% HCC1937 cDNA: 0% HCC1937 cDNA: 10% HCC1954 cDNA. Please note that in HCC1937 100, and 10%, the *NFIA-EHF* peak was detected (red arrow) and no unextended primer peak was observed (black arrow), whereas in HCC1937 on, only the unextended primer peak was confirmed by RT–PCR (**b**).

screened for the presence of multiple expressed fusion genes in a single experiment. Cases without the target fusion gene can be confidently excluded, therefore, leading to a substantial reduction in the workload for the subsequent validation steps, should the entire screening process be undertaken by RT–PCR or FISH.

The limitations of the present study reflect those inherent to the MassARRAY platform, and, for that matter, any platform that does not allow for *de novo* sequencing for the detection of fusion transcripts. First, the approach described is not suitable for the identification of novel fusion genes and is of limited utility for discovery studies. Second, using the probes described, MassARRAY can only be employed to detect expressed fusion genes that lead to the formation of chimeric transcripts with recurrent breakpoints at the RNA level; fusion genes resulting from promoter swapping events cannot be identified. Third, in common with other highly sensitive assays, false positives can occur with appreciable frequency. In fact, even by combining the results of probes 1 and 2, the positive predictive value was still 71.4%. Therefore, in the presence of a positive result by the MassARRAY method described here, validation with orthogonal methods (ie RT–PCR and FISH) remains necessary. Finally, we did not detect any of the 10 fusion genes identified in the cell line panel in any of the 125 primary breast cancers screened using the MassARRAY fusion assay. This may be due to the fact that the expressed fusion genes identified in cancer cell lines may have arisen as artifacts of *in vitro* culturing, and do not readily recapitulate the changes seen in primary tumors. It is conceivable that if fusion genes were detected by massively parallel sequencing of primary breast cancers, and their recurrence interrogated by the MassARRAY methodology described here in a large cohort of phenotypically matched primary breast cancers, recurrent events would have been identified.

In conclusion, we have extended the repertoire of existing applications of the MassARRAY platform by designing an assay that has optimal sensitivity and negative predictive value, and would constitute a useful screening method for expressed fusion genes. The increasing numbers of genomes subjected to massively parallel sequencing will undoubtedly require scalable platforms for the validation of fusion genes identified; the method described here may provide a solution for this need. Importantly, the limited specificity and positive predictive values of the Sequenom MassARRAY assay described here do not allow for its use as a stand-alone diagnostic test for the detection of expressed fusion genes. Confirmation of the results with established 'gold standards,' such as FISH and RT–PCR, remains necessary.

Supplementary Information accompanies the paper on the Laboratory Investigation website (http://www.laboratoryinvestigation.org)

ACKNOWLEDGEMENTS

This project was funded in part by Breakthrough Breast Cancer. PW is funded by a Wellcome Trust Clinical Fellowship grant. NP is supported in part by a Royal College of Surgeons of England Clinical Research Fellowship. VP and JSRF are funded in part by European Community's Seventh Framework Programme (FP7/2010-2014) under the Grant agreement number 242006. JSR-F is a recipient of the 2010 CRUK Future Leaders Prize. We acknowledge NHS funding to the NIHR Biomedical Research Centre.

DISCLOSURE/CONFLICT OF INTEREST

The authors declare no conflict of interest.

- Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 2001;344:1031–1037.
- 2. Brenner JC, Chinnaiyan AM. Translocations in epithelial cancers. Biochim Biophys Acta 2009;1796:201–215.
- 3. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 2007;7:233–245.
- 4. Persson M, Andren Y, Mark J, *et al.* Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc Natl Acad Sci USA 2009;106:18740–18744.
- 5. Tognon C, Knezevich SR, Huntsman D, *et al.* Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. Cancer Cell 2002;2:367–376.
- 6. Tomlins SA, Rhodes DR, Perner S, *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;310:644–648.

- Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 2009;462: 1005–1010.
- Campbell PJ, Stephens PJ, Pleasance ED, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 2008;40:722–729.
- 9. Berger MF, Levin JZ, Vijayendran K, *et al.* Integrative analysis of the melanoma transcriptome. Genome Res 2010;20:413–427.
- 10. Palanisamy N, Ateeq B, Kalyana-Sundaram S, *et al.* Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med 2010;16:793–798.
- 11. Attard G, Clark J, Ambroisine L, *et al.* Heterogeneity and clinical significance of ETV1 translocations in human prostate cancer. Br J Cancer 2008;99:314–320.
- 12. Esgueva R, Perner S, C JL, *et al.* Prevalence of TMPRSS2-ERG and SLC45A3-ERG gene fusions in a large prostatectomy cohort. Mod Pathol 2010;23:539–546.
- 13. Helgeson BE, Tomlins SA, Shah N, *et al.* Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. Cancer Res 2008;68:73–80.
- 14. Maher CA, Kumar-Sinha C, Cao X, *et al.* Transcriptome sequencing to detect gene fusions in cancer. Nature 2009;458:97–101.
- 15. Maher CA, Palanisamy N, Brenner JC, *et al.* Chimeric transcript discovery by paired-end transcriptome sequencing. Proc Natl Acad Sci U S A 2009;106:12353–12358.
- 16. Svensson MA, LaFargue CJ, MacDonald TY, *et al.* Testing mutual exclusivity of ETS rearranged prostate cancer. Lab Invest 2011;91: 404–412.
- 17. Euhus DM, Timmons CF, Tomlinson GE. ETV6-NTRK3–Trk-ing the primary event in human secretory breast cancer. Cancer Cell 2002;2:347–348.
- Lambros MB, Tan DS, Jones RL, et al. Genomic profile of a secretory breast cancer with an ETV6-NTRK3 duplication. J Clin Pathol 2009;62:604–612.
- Reis-Filho JS, Natrajan R, Vatcheva R, *et al.* Is acinic cell carcinoma a variant of secretory carcinoma? A FISH study using ETV6'split apart' probes. Histopathology 2008;52:840–846.
- 20. Makretsov N, He M, Hayes M, *et al.* A fluorescence *in situ* hybridization study of ETV6-NTRK3 fusion gene in secretory breast carcinoma. Genes Chromosomes Cancer 2004;40:152–157.
- 21. Letessier A, Ginestier C, Charafe-Jauffret E, *et al.* ETV6 gene rearrangements in invasive breast carcinoma. Genes Chromosomes Cancer 2005;44:103–108.
- 22. Marchio C, Weigelt B, Reis-Filho JS. Adenoid cystic carcinomas of the breast and salivary glands (or 'The strange case of Dr Jekyll and Mr Hyde' of exocrine gland carcinomas). J Clin Pathol 2010;63:220–228.
- Mitani Y, Li J, Rao PH, *et al.* Comprehensive analysis of the MYB-NFIB gene fusion in salivary adenoid cystic carcinoma: incidence, variability, and clinicopathologic significance. Clin Cancer Res 2010;16:4722–4731.
- 24. West RB, Kong C, Clarke N, *et al.* MYB expression and translocation in adenoid cystic carcinomas and other salivary gland tumors with clinicopathologic correlation. Am J Surg Pathol 2011;35:92–99.
- Deininger MW, Goldman JM, Lydon N, et al. The tyrosine kinase inhibitor CGP57148B selectively inhibits the growth of BCR-ABLpositive cells. Blood 1997;90:3691–3698.
- Falzoi M, Mossa A, Congeddu E, et al. Multiplex genotyping of CYP3A4, CYP3A5, CYP2C9 and CYP2C19 SNPs using MALDI-TOF mass spectrometry. Pharmacogenomics 2010;11:559–571.
- Humeny A, Rodel F, Rodel C, et al. MDR1 single nucleotide polymorphism C3435T in normal colorectal tissue and colorectal carcinomas detected by MALDI-TOF mass spectrometry. Anticancer Res 2003;23:2735–2740.
- James MR, Hayward NK, Dumenil T, et al. Epidermal growth factor gene (EGF) polymorphism and risk of melanocytic neoplasia. J Invest Dermatol 2004;123:760–762.
- 29. Werner M, Sych M, Herbon N, *et al.* Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. Hum Mutat 2002;20:57–64.
- 30. Yu JT, Mao CX, Zhang HW, *et al.* Genetic association of rs11610206 SNP on chromosome 12q13 with late-onset Alzheimer's disease in a Han Chinese population. Clin Chim Acta 2011;412:148–151.

- 31. Thomas RK, Nickerson E, Simons JF, *et al.* Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat Med 2006;12:852–855.
- 32. Fumagalli D, Gavin PG, Taniyama Y, *et al*. A rapid, sensitive, reproducible and cost-effective method for mutation profiling of colon cancer and metastatic lymph nodes. BMC Cancer 2010;10:101.
- MacConaill LE, Campbell CD, Kehoe SM, et al. Profiling critical cancer gene mutations in clinical tumor samples. PLoS One 2009; 4:e7887.
- 34. Dedes KJ, Natrajan R, Lambros MB, *et al.* Down-regulation of the miRNA master regulators Drosha and Dicer is associated with specific subgroups of breast cancer. Eur J Cancer 2011;47:138–150.
- McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 2005;93:387–391.
- 36. Lacroix-Triki M, Suarez PH, MacKay A, *et al.* Mucinous carcinoma of the breast is genomically distinct from invasive ductal carcinomas of no special type. J Pathol 2010;222:282–298.
- Geyer FC, Weigelt B, Natrajan R, et al. Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. J Pathol 2010;220:562–573.
- Li Z, Tognon CE, Godinho FJ, et al. ETV6-NTRK3 fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex. Cancer Cell 2007;12:542–558.