

# Evaluating the repair of DNA derived from formalin-fixed paraffin-embedded tissues prior to genomic profiling by SNP–CGH analysis

Abdel Nasser Hosein<sup>1,\*</sup>, Sarah Song<sup>2,3,\*</sup>, Amy E McCart Reed<sup>3</sup>, Janani Jayanthan<sup>3</sup>, Lynne E Reid<sup>3</sup>, Jamie R Kutasovic<sup>3</sup>, Margaret C Cummings<sup>3,4</sup>, Nic Waddell<sup>2</sup>, Sunil R Lakhani<sup>3,4,5</sup>, Georgia Chenevix-Trench<sup>1</sup> and Peter T Simpson<sup>3</sup>

Pathology archives contain vast resources of clinical material in the form of formalin-fixed paraffin-embedded (FFPE) tissue samples. Owing to the methods of tissue fixation and storage, the integrity of DNA and RNA available from FFPE tissue is compromised, which means obtaining informative data regarding epigenetic, genomic, and expression alterations can be challenging. Here, we have investigated the utility of repairing damaged DNA derived from FFPE tumors prior to single-nucleotide polymorphism (SNP) arrays for whole-genome DNA copy number analysis. DNA was extracted from FFPE samples spanning five decades, involving tumor material obtained from surgical specimens and postmortems. Various aspects of the protocol were assessed, including the method of DNA extraction, the role of Quality Control quantitative PCR (qPCR) in predicting sample success, and the effect of DNA restoration on assay performance, data quality, and the prediction of copy number aberrations (CNAs). DNA that had undergone the repair process yielded higher SNP call rates, reduced log R ratio variance, and improved calling of CNAs compared with matched FFPE DNA not subjected to repair. Reproducible mapping of genomic break points and detection of focal CNAs representing high-level gains and homozygous deletions (HD) were possible, even on autopsy material obtained in 1974. For example, DNA amplifications at the *ERBB2* and *EGFR* gene loci and a HD mapping to 13q14.2 were validated using immunohistochemistry, *in situ* hybridization, and qPCR. The power of SNP arrays lies in the detection of allele-specific aberrations; however, this aspect of the analysis remains challenging, particularly in the distinction between loss of heterozygosity (LOH) and copy neutral LOH. In summary, attempting to repair DNA that is damaged during fixation and storage may be a useful pretreatment step for genomic studies of large archival FFPE cohorts with long-term follow-up or for understanding rare cancer types, where fresh frozen material is scarce.

*Laboratory Investigation* (2013) 93, 701–710; doi:10.1038/labinvest.2013.54; published online 8 April 2013

**KEYWORDS:** array CGH; breast cancer; DNA restoration; DNA copy number alterations; formalin fixed, paraffin-embedded tissue; SNP array

Archived formalin-fixed paraffin-embedded (FFPE) tumors represent a rich reservoir of tissue samples for cancer research. Being able to access this resource to obtain high-quality molecular data is important for investigating the biology underlying rare tumor types, complex malignant processes, such as metastatic progression, and long-term clinical outcome studies. Molecular analyses of FFPE samples, however, is extremely challenging, as the DNA and RNA isolated from such samples is often degraded and chemically

modified during formalin fixation, and as a result of the long storage times. Nevertheless, several molecular assays have been developed that can tolerate low-quality nucleic acids to generate genomic or expression data that is informative to tumor biology.

Chromosomal genomic hybridization (CGH), first described over two decades ago,<sup>1</sup> is a cytogenetic technique that interrogates chromosomal imbalances across the genome.<sup>2</sup> It has been instrumental in defining genome-wide DNA copy

<sup>1</sup>Cancer Genetics Laboratory, Queensland Institute of Medical Research, Herston QLD, Australia; <sup>2</sup>Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia; <sup>3</sup>The University of Queensland, UQ Centre for Clinical Research, Royal Brisbane and Women's Hospital, Herston, Brisbane, QLD, Australia; <sup>4</sup>Pathology Queensland, Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia and <sup>5</sup>The University of Queensland, School of Medicine, Brisbane, QLD, Australia

Correspondence: Dr PT Simpson, PhD, UQ Centre for Clinical Research, The University of Queensland, Building 71/918, Royal Brisbane and Women's Hospital, Herston, QLD 4029, Australia.

E-mail: p.simpson@uq.edu.au

\*These authors contributed equally to this work.

Received 10 January 2013; revised 25 February 2013; accepted 28 February 2013

number changes that occur in human cancers and can readily tolerate fragmented DNA derived from FFPE tissue. Early array-based CGH (aCGH) platforms were dependent on bacterial artificial chromosomes containing large fragments of the human genome<sup>3–6</sup> but have been largely supplanted by commercially available platforms utilizing long oligonucleotide-based or single-nucleotide polymorphism (SNP) markers as probes.<sup>7–10</sup>

Initially designed for use in genetic association studies, SNP arrays have also furthered our understanding of structural variation in cancer. SNP arrays (eg, from vendors such as Affymetrix or Illumina) are the platform of choice by large consortia such as the International Cancer Genome Consortium (ICGC), The Cancer Genome Atlas (TCGA), and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)<sup>11–13</sup> for studying somatic DNA copy number alterations in fresh frozen (FF) tumor samples, as, unlike aCGH platforms, they are also capable of detecting copy neutral loss of heterozygosity (LOH) events. There are, however, limited data demonstrating the tolerance of this technology to compromised DNA quality derived from large series of FFPE tumors. Comparative analyses of matched pairs of FF and FFPE tumors demonstrate that SNP array technology can yield reasonable DNA copy number data.<sup>14–19</sup> However, reduced data quality obtained from FFPE samples on SNP arrays can lead to the false detection of copy number aberrations (CNAs) that are not identified using oligonucleotide aCGH platforms or using matched FF tumors on SNP-based microarrays.<sup>20</sup> Oligonucleotide aCGH platforms developed by Agilent or Nimblegen are thus considered as the more robust at tolerating degraded DNA from FFPE tissue samples than SNP arrays. They provide both high resolution and precision for detecting high-level amplifications, single-copy alterations, and resolving chromosomal break points,<sup>20,21</sup> but lack the capacity for detecting copy neutral LOH.

Several technical developments have been introduced that may help optimize performance of FFPE-derived DNA in SNP array analysis. For example, the introduction of a prequalifier PCR step provides a means of predicting which FFPE DNA samples might yield sufficient quality SNP array data.<sup>15,20,22</sup> The recently developed Oncoscan FFPE platform (Affymetrix), a high-resolution SNP array based on molecular inversion probes, appears to perform well in comparison with aCGH platforms and so might prove to be a successful method for studying FFPE samples for CNAs.<sup>23,24</sup> Finally, there are now several commercial methods available designed to repair DNA that is damaged by tissue processing or sample storage. In these, overhangs and gaps in degraded DNA and adducts in chemically modified DNA sequences are enzymatically repaired, and short DNA fragments are ligated to produce longer fragments. In the current study we have assessed the effectiveness of DNA 'restoration' as a pretreatment step prior to SNP-CGH analysis of archival FFPE tissues, including tumor blocks processed up to 50 years ago.

## MATERIALS AND METHODS

### Clinical Cohort

Ethical approval for the research was obtained from the Human Research Ethics Committees of the Royal Brisbane and Women's Hospital (RBWH) and The University of Queensland. The biospecimens used in this study were obtained from the RBWH pathology department or from the Brisbane Breast Bank, and involved archival FFPE tissue samples, FF tumor tissue, and DNA from blood. In total, tissue specimens from 25 cases were used, encompassing (i) surgical tissue samples (tumor and matched normal DNA from blood) from two cases diagnosed in 2010, (ii) surgical tissue samples (tumor and matched normal) from five cases diagnosed between 1987 and 1990, and (iii) tissue samples taken during autopsies performed on 18 patients who died from metastatic breast cancer between 1959 and 2001. An overview of the specimens and the analyses performed are outlined in Table 1, with further experimental details given in Supplementary Table 1. Notably, case Q590 had tumor DNA from both an FFPE block and a matched FF piece of tissue; for cases Q590 and 007, the tumor DNA was analyzed by SNP-CGH with and without undergoing the restoration process; and for five cases (007, 261, 276, 318, 540), the DNA was extracted using two different DNA extraction methods to see if this affected data quality.

### DNA Extraction

For all FFPE samples, tumor-rich areas were identified from a freshly cut hematoxylin and eosin-stained section, and cores of tumor tissue were punched using a 1-mm diameter tissue microarray needle. Cores were dewaxed and rehydrated according to the standard protocols. Two methods of DNA extraction were used: the DNeasy Blood and Tissue Kit (Qiagen Pty, Chadstone, VIC, Australia) and the High Pure DNA Template Preparation Kit (Roche Australia Pty, Castle Hill, NSW, Australia), which was recommended by Illumina. Both were performed according to the manufacturer's instructions with the following exceptions for both techniques: (i) some tissue samples (Supplementary Table 1) were pretreated in 1 M sodium thiocyanate overnight at 37 °C to remove crosslinks and (ii) for all cases an extended tissue digestion step was performed over three nights with supplementary Proteinase K (Invitrogen) added every 24 h. Eluted DNA was assessed for purity using the Nanodrop-2000, and double-stranded DNA was quantified using the Qubit fluorometer (Invitrogen) as per manufacturer's instructions.

### DNA Restoration and Infinium Homozygous Deletions Assay

The FFPE Quality Control qPCR (QC-qPCR) assay was performed as per manufacturer's instructions (Illumina, San Diego, CA, USA). Triplicate real-time PCR reactions were performed on the Roche LightCycler 480 using 2 ng of DNA for each FFPE sample, and the assay-supplied (QCT) DNA

**Table 1 Overview of samples analyzed for DNA restoration and SNP-CGH assay**

Sample ID	Sample type	Year	Tissue type	Normal/tumor	Tumor type	DNA extraction	DNA restoration
74-T	Autopsy	1974	FFPE	N + T	NA	R/Q	Y
80-T	Autopsy	1980	FFPE	N + T	NA	Q	Y
A3E26	Autopsy	1967	FFPE	N	NA	Q	Y
A4K26	Autopsy	1960	FFPE	N	NA	Q	Y
A14L10	Autopsy	1979	FFPE	N	NA	Q	Y
276T	Surgical	1988	FFPE	N + T	IDC, grade 3, ER – PgR – HER2 +	R/Q	Y
318T	Surgical	1989	FFPE	N + T	IDC, grade 2, ER + PgR + HER2 –	R/Q	Y
540T	Surgical	1990	FFPE	N + T	IDC, grade 3, ER + PgR + HER2 +	R/Q	Y
0007T	Surgical	1987	FFPE	N + T	IDC, grade 3, ER – PgR – HER2 –	R/Q	Y + N
Q590T	Surgical	2010	FFPE/FF	N + T	IDC, grade 3, ER + PgR + HER2 +	R	Y + N
Q607T	Surgical	2010	FFPE	N + T	IDC, grade 3, ER + PgR – HER2 –	R	Y

Abbreviations: NA, data not applicable; FFPE, formalin-fixed paraffin embedded; FF, fresh frozen; N, normal; T, tumor; ER, estrogen receptor; PgR, progesterone receptor; +, positive; –, negative; IDC, invasive carcinoma of no special type; R, Roche DNA extraction method; Q, Qiagen DNA extraction method; Y, DNA sample restored; N, DNA sample not restored.

control. The cycle threshold (Ct) value of the QCT template control was subtracted from the Ct value for each FFPE sample value to calculate the final quality control Ct value for each FFPE DNA sample. FFPE DNA samples with Ct values  $\leq 5$  are reported to be of sufficient quality to proceed with in the DNA restoration and Infinium assay, whereas samples with values  $> 5$  are considered less likely to yield reliable SNP array data. The Infinium HD FFPE DNA Restore protocol was followed as directed, beginning with 100 ng of DNA of each sample. Following DNA restoration, the Infinium HD FFPE Assay was performed according to standard protocol using the Human CytoSNP FFPE-12v2.1 arrays ( $> 262,000$  SNPs, Illumina). The chips were scanned using the iScan (Illumina), and overall SNP call rates, B-allele frequency (BAF), and log R ratio (LRR) values for each SNP were extracted, calculated, and exported using GenomeStudio version 2010.3 (Illumina). The data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE43406. The BAF, LRR, and copy number calls were visualized/calculated using the genocN method,<sup>25</sup> implemented in R version 2.15.0. To measure the extent of noise for each SNP array, the central 50 percent of the LRR was extracted, which represents the total relative copy number values that are not changed across the whole genome. The variance of these values was then calculated; a larger variance score indicated SNP array data with increased noise, and hence a decreased capacity to call DNA copy number changes.

### Immunohistochemistry

Immunohistochemistry (IHC) was performed under the following conditions: ER, 1:100 dilution, clone 6F11 (Novocastra, Leica Microsystems Pty, North Ryde, NSW,

Australia); PgR, 1:100 dilution, clone 16 (Novocastra); HER2, 1:200 dilution, code A0485 (Dako Australia Pty., Campbellfield, VIC, Australia); and EGFR, 1:100 dilution, clone 31G7 (Invitrogen, Life Technologies Australia Pty Mulgrave, VIC, Australia). All antibodies required citrate antigen retrieval except EGFR, which required chymotrypsin-based retrieval. The MACH 1 Universal HRP-Polymer kit (Biocare Medical, Concord, CA, USA) was used for detection.

### TaqMan Copy Number Variation Verification

TaqMan Copy Number Variation assays (Applied Biosystems, Life Technologies Australia Pty Mulgrave, VIC, Australia) were performed as per manufacturer's instructions. Specifically, 20 ng of DNA was analyzed per well using the TaqMan Universal Master Mix, the human RNase P reference probe (#4403326), and either the CAB39L (13q14.2c or Chr.13:49900240; Hs05292107\_cn) or 13q14.3b (Chr.13:52514984 location; Hs05302865\_cn) target probes. Verification was performed twice, initially using the same DNA that was applied to restoration and Infinium assay, and secondly, using DNA extracted from tumor cells that were enriched by laser-capture microdissection (LCM).<sup>26</sup> The latter was also performed on duplicate LCM samples; using 7  $\mu$ l of neat (unpurified) DNA in the assay. Plates were run on the OneStepPlus system and data analyzed using CopyCaller Software (Applied Biosystems).

### Meta-Analysis Of 13q14 Deletion Using TCGA Genomic Data

Data from 527 breast cancers was downloaded from TCGA (downloaded December 2011)<sup>11</sup> to assess the frequency of the 13q14.2 deletion in breast cancer and the effect this has on the expression of genes from this genomic region. The data set contained gene expression data (log<sub>2</sub> lowess normalized;

$n = 527$ ) and DNA copy number data (segmental values defining tumor DNA genomic events;  $n = 500$ ). The expression levels of genes were examined in tumors, stratified according to breast cancer intrinsic molecular subtypes (basal-like, HER2, luminal A, luminal B, and normal-like;<sup>27</sup>). This analysis was performed in R version 2.15.0.

## RESULTS

The effectiveness of the DNA restoration assay for defining DNA CNAs with the Infinium SNP–CGH assay was evaluated in a collection of archival pathology blocks. All DNA samples underwent a quality assessment using the QC–qPCR assay before a subset of these samples was selected for DNA restoration and SNP–CGH analysis (Table 1 and Supplementary Table 1). Several parameters of the process were assessed, including (i) whether the FFPE QC–qPCR assay could reliably predict success in downstream applications, (ii) if DNA restoration improves the quality of the data obtained from the Infinium assay, (iii) if the method of DNA extraction affects data quality, and (iv) if the method can identify recurrent DNA copy number alterations.

### The QC–qPCR Reaction as a Predictor of Array Quality

DNA was extracted from FFPE blocks comprising 13 surgical tissue samples from six patients and 30 autopsy tissue samples from 18 patients. The QC–qPCR was performed on all samples to identify FFPE DNA samples of sufficient quality to analyze by DNA restoration and Infinium assay. QC values were variable for both surgical tissue samples (range:  $-1.0$ – $11.3$ ) and for autopsy samples (range:  $2.1$ – $13.7$ ), with acceptable values considered to be  $<5$ . We included DNA samples with poor QC values in DNA restoration and Infinium for experimental verification of the assay, and from this data we found overall that lower QC values predicted for higher SNP call rates (Supplementary Table 1, Supplementary Figure 1). We calculated the variance or degree of noise in the LRR data for each array (Supplementary Table 1) to measure the observation that samples exhibiting low QC values had less noisy data. We saw that lower QC values correlated with lower LRR variance or noise (Supplementary Figure 1). This variance also correlated with QC–qPCR values and SNP call rates (Supplementary Figure 1). For each of these assessments, there were exceptions, in which samples with low QC–qPCR values yielded low SNP call rates and/or high variance, but generally the data suggested that the QC–qPCR assay provided a reasonable prediction of array quality.

### Effectiveness of DNA Restoration on Infinium Assay Data Quality

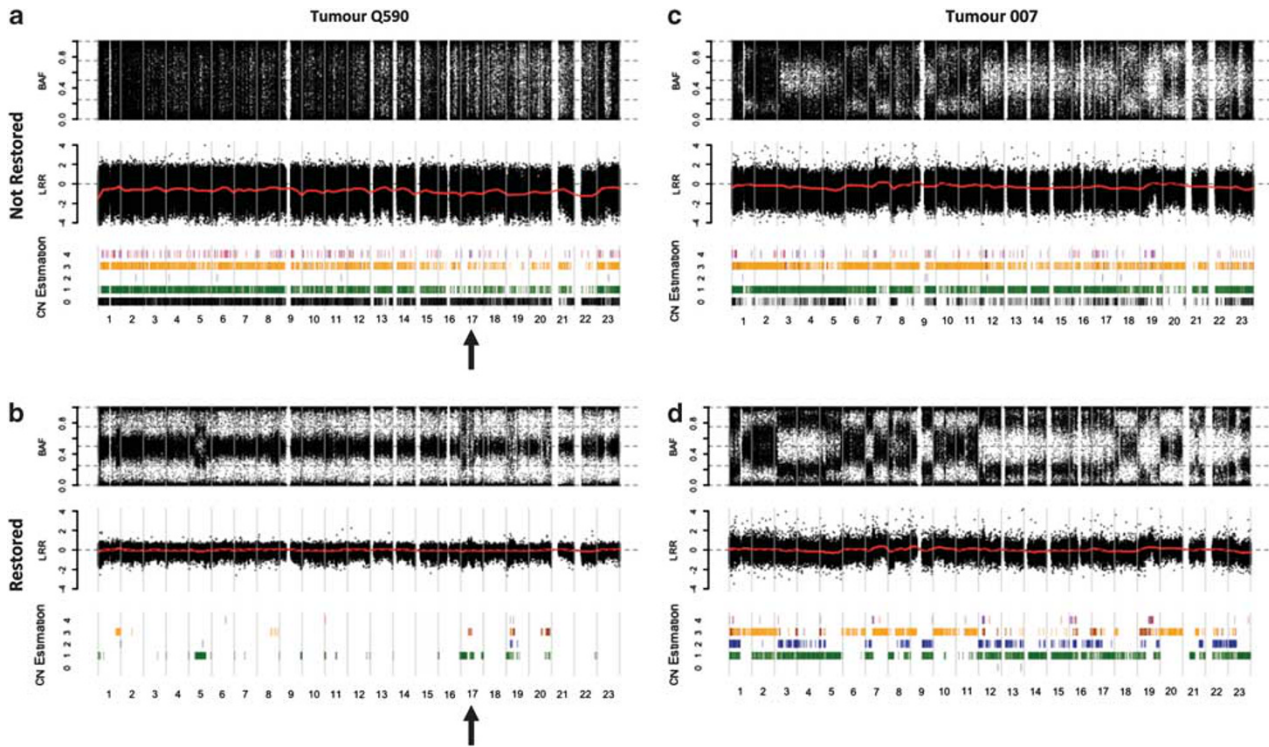
The role DNA restoration has in improving DNA sample performance in the Infinium assay was investigated directly using two tumor samples (tumor Q590 diagnosed in 2010 and tumor 007 diagnosed in 1987) with and without restoration treatment (Figure 1, Supplementary Figures 2 and 3), and between matched FF and FFPE samples of tumor

Q590 (Supplementary Figure 2). We assessed SNP call rates, BAF and LRR plots, LRR variance, and the ability to call copy number changes as measures of performance. Improvement was observed in all aspects of this assessment for both tumors following DNA restoration (Supplementary Table 1). SNP call rates for tumor Q590 improved markedly from  $0.699$ – $0.964$ , to be comparable to the  $0.977$  SNP call rate obtained for the matched FF tumor. The LRR variance improved from  $0.26270$  (without restoration) to  $0.01090$  (with restoration). Tumor 007 had ‘with and without restoration’ data available from both Roche and Qiagen-extracted DNA. For Roche-extracted DNA there was an improvement in SNP call rates (from  $0.796$ – $0.839$ ) and variance (from  $0.14394$ – $0.02481$ ); for Qiagen-extracted DNA, SNP call rates did not improve with restoration treatment ( $0.831$ – $0.841$ ) but LRR variance did (from  $0.04993$ – $0.0134$ ; Supplementary Table 1). These improvements were indicative of enhanced data quality following restoration and improved detection for calling CNAs in these FFPE samples (see below; Figure 1, Supplementary Figures 2 and 3).

We also found good consistency in the SNP genotype calls between restored DNA and nonrestored DNA for these two tumors (Supplementary Table 2). Firstly, we identified the SNPs called in both samples and then identified the concordance in genotype called. For tumor Q590, there were three replicates of restored DNA tested against a nonrestored sample and there was an  $84.5$ – $86.11\%$  concordance observed (from  $188\,214$  to  $189\,904$  SNPs). For tumor 007, there was data for both Qiagen (two replicates of restored *versus* a nonrestored sample) and Roche-extracted DNA (one restored *versus* nonrestored comparison), and from these data there was a  $97.97$ – $99.75\%$  concordance in the SNP genotypes that were called (from  $235\,951$  to  $244\,460$  SNPs).

The matched FF *versus* FFPE analysis for tumor Q590 was performed to determine if the repaired DNA resulted in accurate DNA copy number predictions in the FFPE samples. BAF plots and genomic break points called by GenoCN were replicated between the matching FF and FFPE samples (Supplementary Figure 2). However, as has been reported previously,<sup>28</sup> the magnitude of the copy number change was reduced in the FF sample relative to the FFPE sample. For instance, deletions detected on chromosomes 5 and 17 in the FFPE samples were classified as copy neutral LOH in the FF sample (Supplementary Figure 2). This is probably due to differences in the proportion of contaminating normal cells in the tissue preparation prior to DNA extraction, as the FFPE tissue was macrodissected by taking cores of tissue to enrich for tumor cells, and hence had a higher tumor cellularity relative to the FF prepared DNA (estimated at  $70\%$  *versus*  $50\%$ ).<sup>28,29</sup>

To measure the reproducibility of the DNA restoration assay, eight FFPE DNA samples were assayed as technical replicates (seven as duplicates and one as a triplicate; Supplementary Table 1). SNP call rates of all replicate pairs were within  $0.038$  of each other. Replicates 74-T, 007T, and



**Figure 1** The effectiveness of DNA Restoration in SNP-CGH analysis of FFPE DNA samples. Visual inspection of the BAF and log R ratio (LRR) plots highlights the improvement in quality of the data obtained, following restoration. This was demonstrated by comparing the SNP array data obtained from the same FFPE DNA samples processed both without the DNA restoration step (a and c) and with the DNA restoration step (b and d), prior to the Infinium assay. This was done for two FFPE tumor samples: tumour Q590, archived in 2010 (a and b) and tumor 007, archived 1987 (c and d). The DNA copy number changes identified by GenoCN is shown for each specimen (DNA copy number states: 2 = normal, 1 and 0 = deletion, 3 and 4 = gains). Genome position indicated along X axis for all plots, from chromosome 1–23 (X). Arrow indicates position of *ERBB2* amplification (also shown in Figure 2).

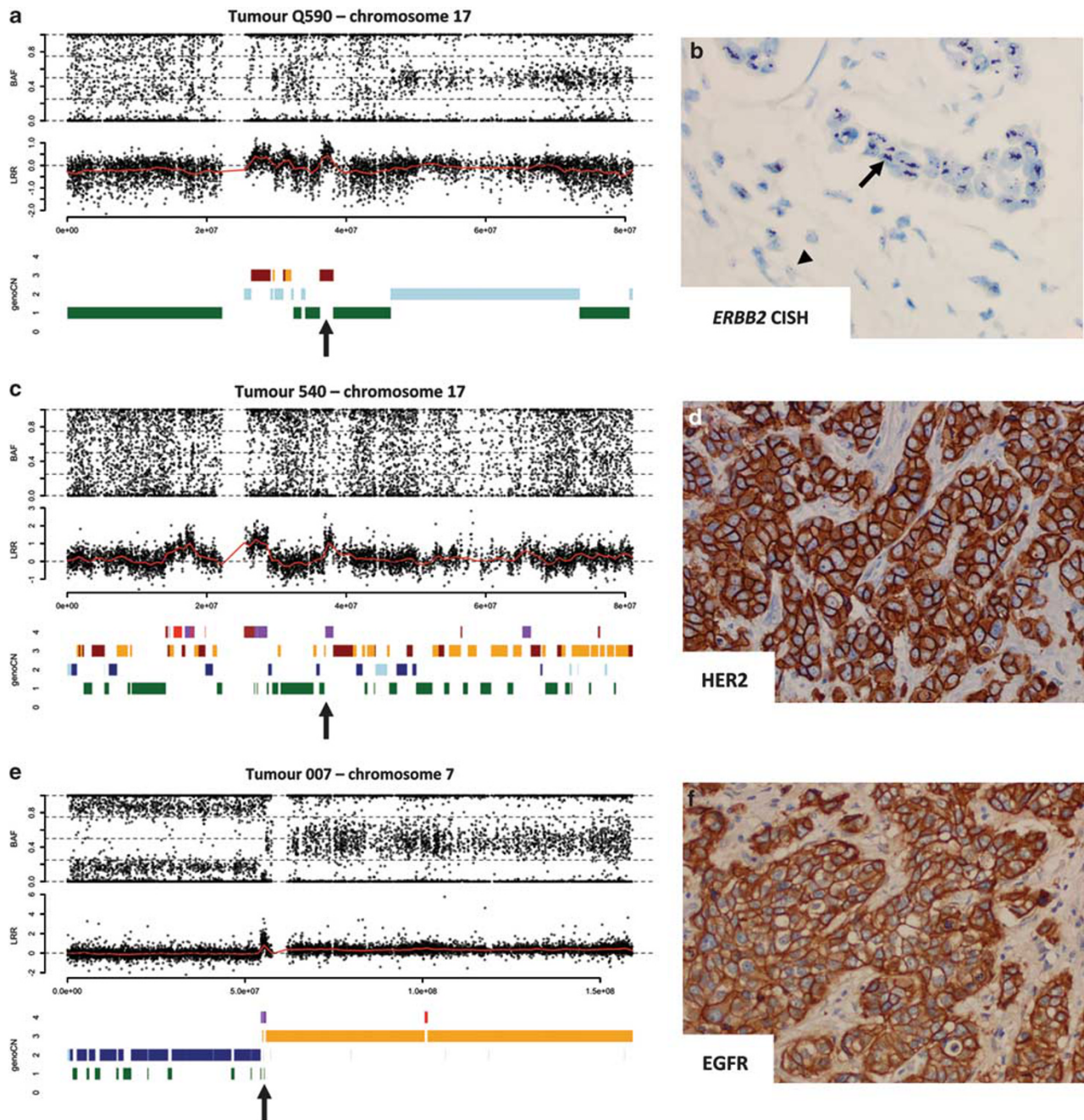
Q590-T were derived from tumor DNA and so could serve to assess reproducibility in identifying somatic CNAs. Defining DNA break points resulting in changes in copy number states as defined by GenoCN, and identifying focal high-level gains was quite robust between each set of replicates (supplementary Figures 2–4). For example, in tumor 74, five out of six regions of high-level gain (copy number state of 4; 2q11, 8q23-q24, 12q13, 15q26, 17q12, 17q25) were detected in both replicates (arrows in Supplementary Figure 4). Furthermore, two focal HD (mapping to chr12: 8,008,179-8,123,306 and chr12: 27476545-27498107) were also detected in both replicates of T007, which were not detected in the nonrestored DNA sample from the same case (Supplementary Figure 3). In each case, however, the GenoCN software found defining single-copy alterations was difficult and so there was more variability observed between replicates, particularly in differentiating between LOH (green copy number states in GenoCN plots) and copy neutral LOH (dark blue copy number states) (Supplementary Figures 2–4).

### Comparison of Different DNA Extraction Techniques and Their Impact on Array Data Quality

The DNA restoration and Infinium assay recommended the High Pure DNA Template Preparation Kit (Roche) for

DNA preparation. The DNeasy Blood and Tissue DNA kit (Qiagen) is also widely used, and so we evaluated both methods. DNA from 22 samples was extracted using both methods and neither method consistently outperformed the other in terms of QC-qPCR values (Supplementary Figure 5). Ten of the samples were analyzed by Infinium assay, and in terms of SNP call rates and LRR variance; the Qiagen method either matched ( $n=7$ ) or outperformed ( $n=3$ ) the Roche method (Supplementary Table 1, Supplementary Figure 5). The improvement observed using Qiagen-extracted DNA was notable in cases that were considered borderline or poor DNA samples according to the QC-qPCR assay. For example, DNA sample T74 and tumor 540 both yielded 10% higher SNP call rates with the Qiagen extraction method. This translated into improvements in the quality of BAF and LRR plots, and the ability to call CNAs that were impossible to discern using Roche-extracted DNA, such as high-level gains on chromosomes 8 and 17 (Supplementary Figure 6 and 7).

We tested whether the pretreatment of FFPE tissues in sodium thiocyanate to remove crosslinks<sup>22</sup> helped or hindered the DNA restoration process and subsequent Infinium assay. This comparison was performed on only tumors Q590 and Q607, where it had little impact; and hence we cannot infer whether this step would improve data quality



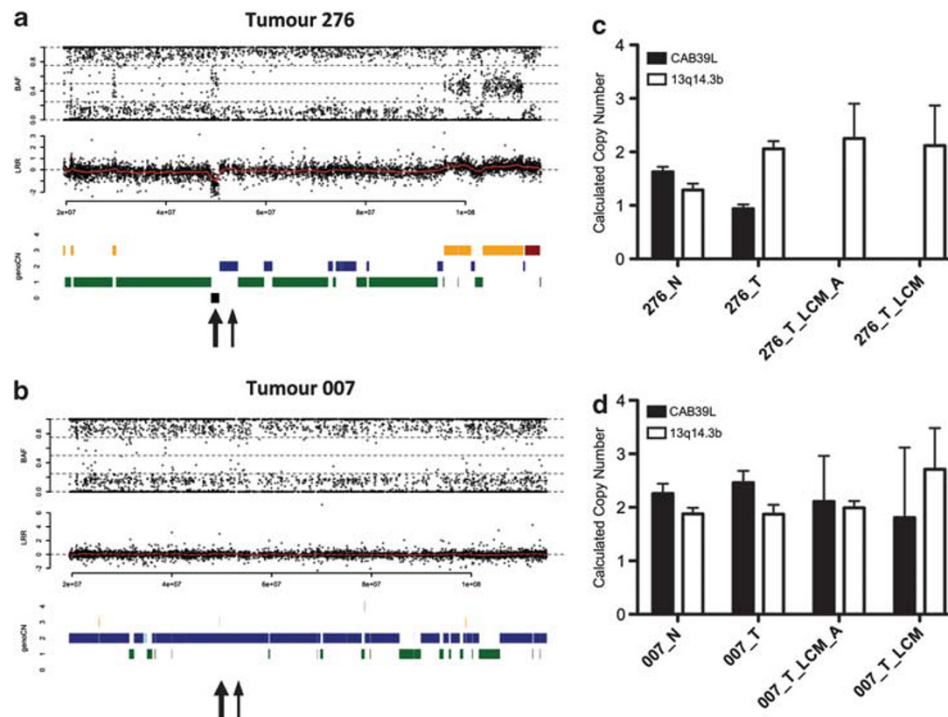
**Figure 2** Validation of DNA copy number amplifications detected in restored FFPE samples. DNA aberrations were validated by immunohistochemistry (IHC) and *in situ* hybridization. Tumor samples Q590 (a) and 540 (c) both exhibited amplification of 17q12-q21, encompassing the *ERBB2* gene (arrow). These were validated by chromogenic *in situ* hybridization (b) or IHC (d). Tumor 007 (e) harbored an amplification at 7p11.2 encompassing the *EGFR* gene; overexpression of EGFR protein was validated by IHC (f; scored as 3+ in 100% of cells).

on the older and more challenging samples, all of which were extracted using this pretreatment step.

### Validation Of Specific DNA Copy Number Changes Detected Using Restored FFPE Samples

The Illumina restoration protocol and CytoSNP Infinium assay detected both gain and deletion DNA copy number changes important in tumor development, some of which

were validated. Tumor Q590 had amplification at the *ERBB2* locus 17q12 following DNA restoration, but there was no evidence of this beyond the background variability seen across the entire chromosome in the nonrestored FFPE sample (Figure 1). This tumor was classified as HER2 3+ (positive) by IHC and harbored 15 copies of the *ERBB2* gene as quantified by chromogenic *in situ* hybridization (CISH) (Figure 2a and b, Table 1). Tumor 540 (from 1990)



**Figure 3** Validation of a predicted homozygous deletion (HD) on chromosome 13q14. Tumor 276 harbored a predicted HD in chromosome 13 (chr13: 49,059,577-50,624,140; thick arrow in **a**), which was not evident in other tumors, including tumor 007 (**b**). To validate the HD, quantitative PCR (qPCR) was performed on DNA from both tumors using assays for *CAB39L* in the deleted region (chr13: 49,900,240 at 13q14.2c; thick arrow in **a** and **b**) and a control region 13q14.3b of normal copy number (chr13: 52,514,984; thin arrow in **a** and **b**). qPCR was performed on the same DNA as used in the Infinium assay DNA from both tumor (T) and normal (N) and on two independent LCM enrichments of the same tumor samples. qPCR data for tumor 276 (**c**) shows reduced *CAB39L* copy number in tumor versus normal, which is exemplified following laser-capture microdissection (LCM) enrichment, whereas qPCR data for tumor 007 (**d**) shows copy number of  $\sim 2$  in N and T.

performed poorly in QC-qPCR and Infinium assays, yet LRR plots and GenoCN were still useful for defining copy number alterations, including amplification of *ERBB2/HER2*, which was validated by IHC (Figure 2c and d, Table 1). Tumor 007 (from 1987) had an amplification mapping to 7p11.2 and encompassing the *EGFR* gene (7p11.2) (Figure 2e and f and Supplementary Figure 3). This was validated by IHC, confirming that the triple negative classified tumor was EGFR positive in 100% of tumor cells. It should be noted, however, that both nonrestored and restored DNA samples were able to detect this amplification, indicating that restoration is not always required to detect high-level copy number increases (Supplementary Figure 3).

Tumors 007 and 276 harbored putative homozygous deletion (HD) on chromosome 12 (chr12: 8,008,179-8,123,306; Supplementary Figure 3) and chromosome 13 (chr13: 49,047,417-50,624,140; Figure 3), respectively. To confirm the chromosome 13 HD, real-time qPCR was carried out targeting a gene within the deleted region (*CAB39L*) and one to a control region (13q14.3C) in an adjacent chromosomal region with normal copy number (Figure 3). qPCR was performed on tumor and normal DNA from case 276, and on tumor and normal DNA from a second case (007) that was copy number normal in this genomic region. As found in the array data, qPCR showed that the deleted region had half the copy

number of the adjacent region in the tumor DNA, whereas the two regions were approximately equal in normal DNA and in the control tumor 007. The qPCR assay was repeated on LCM-enriched tumor epithelium, to eliminate 'normal' DNA from contaminating stromal cells; there was no evidence of *CAB39L* DNA in tumor 276, whereas the copy number was normal in tumor 007 (Figure 3). The copy number of the adjacent genomic region 13q14.3 was normal in all samples, suggesting 13q14.2 was indeed homozygously deleted in tumor cells of case 276 (Figure 3).

The major candidate tumor suppressor gene in this region is likely to be *RBI*; however, there are limited data regarding the role and frequency of this alteration in breast cancer, and the genes particularly targeted. We have previously reported a HD mapping to a similar region in a breast tumor from a patient with an inherited *BRCA2* mutation (chr13: 47,869,937-49,878,203), and through integrated analysis of gene expression profiling data from the same case we identified nine genes (*CAB39L*, *SETDB2*, *PHF11*, *RCBTB1*, *EBPL*, *KPNA3*, *TRIM13*, *KCNRG*, *DLEU1*) that were also down-regulated.<sup>30</sup> In tumor 276, the deletion spanned 17 genes (*RCBTB2*, *CYSLTR2*, *FNDC3A*, *MLNR*, *CDADC1*, *CAB39L*, *SETDB2*, *PHF11*, *RCBTB1*, *ARL11*, *EBPL*, *KPNA3*, *CTAGE10P*, *SPRYD7*, *DLEU2*, *TRIM13*, *KCNRG*). The resolution of the array used here is not sufficient to

accurately define the break points of this deletion, but it appeared to involve the 3' end of the *RB1* gene. Without having gene expression data available for this case and to appreciate the impact of this deletion in breast cancer, we analyzed SNP-CGH-defined DNA copy number and gene expression levels of all genes within the deleted region in tumors from the TCGA resource. Deletion in this genomic region was identified in 34/500 (6.8%) tumors (18 were of Luminal B subtype, nine were Luminal A, five were Basal and two were Her2). Particularly low levels of expression (relative to the median of all tumors) were noted for most genes in this region in up to 5/527 tumors, although expression of *RB1* and *RCBTB2* was more frequently downregulated (Supplementary Figure 8). Several tumors (eg, TCGA-BH-A0EE, TCGA-AN-A04D, TCGA-A8-A09V) showed consistent loss of expression of multiple genes from this region, suggesting that the expression was driven by loss of DNA (Supplementary Figure 8). This was confirmed by integrating gene expression data and DNA copy number from two of these tumors for which both types of data were available (TCGA-BH-A0EE and TCGA-AN-A04D) (Supplementary Figure 8). The deletion in tumor TCGA-AN-A04D mapped to a very similar region to that identified in tumor 276, and involved *RB1*. The deletion in tumor TCGA-BH-A0EE was more complex, involving two separate regions centered on *RB1* and *EBPL*.

## DISCUSSION

Here we report an evaluation of Illumina's FFPE DNA restoration solution as a pretreatment step for FFPE-derived DNA prior to whole-genome CNA analysis using SNP arrays. The DNA restoration process uses a combination of enzyme incubation steps to repair damaged DNA and to ligate fragments of double-stranded DNA to create a more appropriate template DNA for the whole-genome amplification step of the Infinium assay. The performance of this process has been tested in a particularly challenging series of FFPE samples, including surgical material from 1987 to 1990 (up to 25 years old) and a series of FFPE blocks taken during postmortem for metastatic breast cancer between 1960 and 1980 (up to 52 years old).

From a technical point of view, we found that the inclusion of the DNA restoration step significantly improved the performance of FFPE DNA in the Infinium assay, generating higher SNP call rates and reduced LRR variance ('noise') compared with matched FFPE DNA samples not pretreated with the restoration step. As with previous studies,<sup>15,20,22</sup> we show that the FFPE quality control PCR assay is a reasonable predictor of sample success in which DNA samples with low QC values (<5) had higher SNP call rates and lower LRR variance relative to samples with higher QC values (>5). It is important to note that a low QC value was indicative that a DNA sample would respond well to the restoration protocol in preparation for the Infinium assay, rather than being of sufficient quality to be directly assayed on the array. This was

clearly demonstrated by FFPE tumor Q590 DNA, which yielded a low QC value (1.63), yet failed the Infinium assay when the DNA was not restored. Finally, DNA extracted using the Qiagen kit provided notable improvements in data quality and mapping of CNAs compared with the Roche kit in these challenging FFPE samples, and this was particularly evident with tumors 540 and 74 that had poor QC-qPCR scores. The reasons for this are unclear, given the general similarities between the two column-based extraction and purification methods, and it may be a sample-specific phenomenon.

The accuracy and reproducibility of identifying DNA CNAs were tested with technical replicates and a comparison between DNA derived from a FF tumor and a matched FFPE block from the same case. These highlighted the reliability in which the assay could detect high-level gains, which can be major drivers of tumor behavior. For example, focal high-level amplifications at 17q12 (harboring the *ERBB2* gene) and 7p11.2 (containing the *EGFR* gene) were detected and validated by other means. A tumor obtained from postmortem in 1974 yielded quite noisy SNP array data, yet despite this several high-level CNAs (8q23-q24,12q13,15q26,17q12,17q25) were reproducibly detected in both replicates. These are common events in breast cancer and known drivers of phenotype. Depending on the quality of the DNA obtained from the FFPE samples, high-level CNAs can be identified in nonrestored DNA. For instance, tumor 007 performed quite well in the Infinium assay without undergoing DNA restoration (Supplementary Figure 3) with several amplifications, including at 7p11.2, 12p12.1, 16p12.1, 16p13.3 and 16p13.2, being mapped using both restored and nonrestored DNA. Others have reported reasonable success of FFPE samples in SNP array experiments,<sup>14-19</sup> so this is not unexpected. However, the clarity and magnitude of the amplification may be enhanced in LRR plots of restored DNA (for instance for the high-level CNAs on 16p). Furthermore, two HDs, as defined by GenoCN, were detected in this tumor at 12p13.31 (involving the genes *SLC2A3/GLUT3*) and 12p11.23 (involving *STK38L* and *ARNTL2*). These very focal deletions were detected in both technical replicates of Qiagen-extracted, restored DNA, but not in nonrestored DNA. These data suggest that the restoration process provides some subtle but important improvements in data quality and CNA detection.

HD are also considered important CNA events in driving tumorigenesis and tend to occur in genomic regions housing tumor suppressor genes. A second HD was identified in case 267 at 13q14.2, was verified by qPCR, and has been previously reported in breast cancer.<sup>30,31</sup> The size and complexity of this deletion is variable in breast tumors, and sometimes involves *RB1* as the most likely tumor suppressor gene,<sup>31</sup> as illustrated in tumors TCGA-AN-A04D and TCGA-BH-A0EE (Supplementary Figure 8). A meta-analysis of data from the TCGA suggested that the deletion also drives the loss of expression of multiple genes (eg, *CAB39L*, *EBPL*, *SETDB2*, *TRIM13*) that may also have a role in the behavior



of about 1% of breast tumors. Curtis *et al.*<sup>12</sup> also performed an integrated analysis of genomic and gene expression data from ~2000 breast cancers and reported, in supplementary data, that the 13q14 deletion drives the loss of expression of several genes within this region (eg, *SUCLA2*, *MED4*, *RBI*, *P2RY5*, *PHF11*, *TRIM13*) and occurs in a similar small proportion (<1%) of tumors, suggesting that this HD is a recurrent but rare event in breast cancer.

The mapping of genomic break points was also quite reproducible between replicates and in the FF *versus* FFPE comparison. The most obvious issue, however, in defining CNAs using FFPE DNA in this assay is in differentiating between loss (green, GenoCN state = 1) and copy neutral LOH (dark blue, GenoCN state = 2). There was frequent variability in this assessment between technical replicates and in the FF *versus* the matched FFPE comparison: for example, in chromosomes 5p and 17 of case Q590; chromosomes 7p, 12, 16q, and 21q of tumor 007; and chromosomes 1 and 4 of tumor 74. We consider tumor cellularity differences between the FF (~50%) and FFPE (~70%) samples to be a contributor to this problem in this comparison, but for the technical replicates, it is most likely due to inherent noise generated by FFPE DNA samples and the subsequent difficulty in differentiating between loss of one allele only and copy neutral LOH (the loss of one allele and duplication of the other allele). A different analysis algorithm might handle this better, and increasing the tumor cellularity of samples analyzed might also help. Many studies apply a threshold of 70–80% tumor cellularity for FF samples to generate sufficient sensitivity in somatic mutation calling,<sup>11–13</sup> and so for these more challenging samples it may be important to enrich further to 90–100% for FFPE samples to reduce the normal cell DNA content to an absolute minimum.

The method of choice for analyzing genome-wide CNAs in FFPE samples is a difficult one, taking into account the cost, accuracy in detecting CNAs, platform resolution, input DNA amount, and accessibility to the different technologies.<sup>19</sup> The DNA restoration and Infinium assay has performed reasonably well here, although the value of applying these FFPE samples to SNP arrays rather than oligo-based aCGH platforms was in this detection of allele-specific aberrations, yet this proved somewhat problematic. As we have not directly compared with the oligo-based aCGH platforms, such as those provided by Agilent and Nimblegen or the Affymetrix OncoScan FFPE platform, it remains unclear which is the most accurate and robust. The tolerance of these assays for low input DNA amounts is an important consideration, as valuable clinical material may be limited by FFPE block thickness, tumor size, or the requirement of microdissection to reduce the contribution of contaminating normal tissue. The restoration protocol required 100 ng of double-stranded DNA, which is less than the recommended 250 ng or more required for the Agilent or Nimblegen assays. Reduced DNA input amounts have been tested using Agilent, Nimblegen, and Affymetrix OncoScan FFPE platforms,<sup>24</sup> and

these assays do show tolerance for lower input amounts (50–100 ng), but there is also some increase in data noise.

There has been a considerable drive in recent years to move tumor genome studies to whole-genome or exome-sequencing platforms to benefit from major improvements in sensitivity and detection of somatic CNAs and small nucleotide variants (eg, substitutions, insertions, and deletions). While the power of these approaches is being realized for FF tumors,<sup>11–13</sup> there are currently few studies to have realized this potential for FFPE samples, owing to the inherent problems of working with chemically modified and damaged DNA, the requirements for larger amounts of genomic DNA and/or high depth of sequencing coverage to minimize ambiguous read mapping and the subsequent added bioinformatics challenges that this brings.<sup>28,32–35</sup> Large consortia such as the ICGC and TCGA also continue to analyze tumor genomes by SNP arrays as an important support tool for CNA detection and assessment of tumor cellularity,<sup>11–13</sup> and so for these reasons it remains important to optimize CGH platforms for whole-genome CNA detection of FFPE tumor samples.

In summary, we have assessed a method for repairing damaged DNA derived from challenging FFPE tissue samples prior to application to genome-wide SNP-CGH arrays. The repaired DNA performed better on SNP-CGH arrays than nonrepaired DNA, and provided an improved means to map genomic break points and detect CNAs that are biologically relevant to driving tumorigenesis, such as focal high-level amplifications and HD. However, one of the benefits of utilizing SNP arrays over other aCGH platforms is in the capacity to reproducibly define single-copy CNAs and differentiate between LOH and copy neutral LOH, and we found this problematic with these FFPE samples.

Further work is required to determine whether this or other available methods for repairing damaged DNA is beneficial for other genomic applications requiring FFPE resources, such as oligonucleotide aCGH, whole-genome methylation analysis on SNP arrays, or sequencing-based applications in which current protocols for archival materials are extremely challenging.

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

#### ACKNOWLEDGEMENTS

We would like to express our thanks to Sibylle Cocciardi (Queensland Institute for Medical Research) and Dr Brian Fritz (Illumina) for technical assistance. We would also like to acknowledge the contribution of tissue donors and research groups to the generation of the Brisbane Breast Bank and TCGA data resource. Peter Simpson is a recipient of a fellowship from the National Breast Cancer Foundation, Australia. This work was funded by National Health and Medical Research, Australia—Program Grant #1017028.

#### Disclaimer

Illumina had no role in study design, data interpretation or manuscript preparation.

**DISCLOSURE/CONFLICT OF INTEREST**

The authors received support from Illumina to evaluate this assay.

- Kallioniemi A, Kallioniemi OP, Sudar D, *et al*. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;258:818–821.
- Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* 2005;37(Suppl):S11–S17.
- Devries S, Nyante S, Korkola J, *et al*. Array-based comparative genomic hybridization from formalin-fixed, paraffin-embedded breast tumors. *J Mol Diagn* 2005;7:65–71.
- Loo LW, Grove DI, Williams EM, *et al*. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Res* 2004;64:8541–8549.
- Simpson PT, Reis-Filho JS, Lambros MB, *et al*. Molecular profiling pleomorphic lobular carcinomas of the breast: evidence for a common molecular genetic pathway with classic lobular carcinomas. *J Pathol* 2008;215:231–244.
- Natrajan R, Lambros MB, Rodriguez-Pinilla SM, *et al*. Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin Cancer Res* 2009;15:2711–2722.
- Wicker N, Carles A, Mills IG, *et al*. A new look towards BAC-based array CGH through a comprehensive comparison with oligo-based array CGH. *BMC Genomics* 2007;8:84.
- Curtis C, Lynch AG, Dunning MJ, *et al*. The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* 2009;10:588.
- Coe BP, Ylstra B, Carvalho B, *et al*. Resolving the resolution of array CGH. *Genomics* 2007;89:647–653.
- Halper-Stromberg E, Frelin L, Ruczinski I, *et al*. Performance assessment of copy number microarray platforms using a spike-in experiment. *Bioinformatics* 2011;27:1052–1060.
- TCGA. Comprehensive molecular portraits of human breast tumors. *Nature* 2012;490:61–70.
- Curtis C, Shah SP, Chin SF, *et al*. The genomic and transcriptomic architecture of 2,000 breast tumors reveals novel subgroups. *Nature* 2012;486:346–352.
- Stephens PJ, Tarpey PS, Davies H, *et al*. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012;486:400–404.
- Thompson ER, Herbert SC, Forrest SM, *et al*. Whole genome SNP arrays using DNA derived from formalin-fixed, paraffin-embedded ovarian tumor tissue. *Hum Mutat* 2005;26:384–389.
- Jacobs S, Thompson ER, Nannya Y, *et al*. Genome-wide, high-resolution detection of copy number, loss of heterozygosity, and genotypes from formalin-fixed, paraffin-embedded tumor tissue using microarrays. *Cancer Res* 2007;67:2544–2551.
- Tuefferd M, De Bondt A, Van Den Wyngaert I, *et al*. Genome-wide copy number alterations detection in fresh frozen and matched FFPE samples using SNP 6.0 arrays. *Genes Chromosomes Cancer* 2008;47:957–964.
- Lips EH, Dierssen JW, van Eijk R, *et al*. Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. *Cancer Res* 2005;65:10188–10191.
- Oosting J, Lips EH, van Eijk R, *et al*. High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays. *Genome Res* 2007;17:368–376.
- Alvarez K, Kash SF, Lyons-Weiler MA, *et al*. Reproducibility and performance of virtual karyotyping with SNP microarrays for the detection of chromosomal imbalances in formalin-fixed paraffin-embedded tissues. *Diagn Mol Pathol* 2010;19:127–134.
- Nasri S, Anjomshoa A, Song S, *et al*. Oligonucleotide array outperforms SNP array on formalin-fixed paraffin-embedded clinical samples. *Cancer Genet Cytogenet* 2010;198:1–6.
- Hostetter G, Kim SY, Savage S, *et al*. Random DNA fragmentation allows detection of single-copy, single-exon alterations of copy number by oligonucleotide array CGH in clinical FFPE samples. *Nucleic Acids Res* 2010;38:e9.
- van Beers EH, Joosse SA, Ligtenberg MJ, *et al*. A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer* 2006;94:333–337.
- Johnson CE, Goringe KL, Thompson ER, *et al*. Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma. *Breast Cancer Res Treat* 2012;133:889–898.
- Krijgsman O, Israeli D, Haan JC, *et al*. CGH arrays compared for DNA isolated from formalin-fixed, paraffin-embedded material. *Genes Chromosomes Cancer* 2012;51:344–352.
- Sun W, Wright FA, Tang Z, *et al*. Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res* 2009;37:5365–5377.
- Simpson PT, Gale T, Reis-Filho JS, *et al*. Columnar cell lesions of the breast: the missing link in breast cancer progression? A morphological and molecular analysis. *Am J Surg Pathol* 2005;29:734–746.
- Hu Z, Fan C, Oh DS, *et al*. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;7:96.
- Wood HM, Belvedere O, Conway C, *et al*. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic Acids Res* 2010;38:e151.
- Song S, Nones K, Miller D, *et al*. qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE* 2012;7:e45835.
- Waddell N, Arnold J, Cocciardi S, *et al*. Subtypes of familial breast tumors revealed by expression and copy number profiling. *Breast Cancer Res Treat* 2010;123:661–677.
- Jonsson G, Staaf J, Vallon-Christersson J, *et al*. The retinoblastoma gene undergoes rearrangements in BRCA1-deficient basal-like breast cancer. *Cancer Res* 2012;72:4028–4036.
- Schweiger MR, Kerick M, Timmermann B, *et al*. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. *PLoS ONE* 2009;4:e5548.
- Yost SE, Smith EN, Schwab RB, *et al*. Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. *Nucleic Acids Res* 2012;40:e107.
- Kerick M, Isau M, Timmermann B, *et al*. Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med Genomics* 2011;4:68.
- Lipson D, Capelletti M, Yelensky R, *et al*. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat Med* 2012;18:382–384.