

Evaluation of Phi29-based whole-genome amplification for microarray-based comparative genomic hybridisation

Edurne Arriola^{1,2}, Maryou BK Lambros¹, Chris Jones³, Tim Dexter¹, Alan Mackay¹, David SP Tan¹, Narinder Tamber¹, Kerry Fenwick¹, Alan Ashworth¹, Mitch Dowsett^{1,2} and Jorge S Reis-Filho¹

¹The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, UK; ²Department of Academic Biochemistry, Institute of Cancer Research, London, UK and ³Section of Paediatric Oncology, Institute of Cancer Research, Sutton, UK

For the optimal performance of high throughput genomic technologies sufficient yields of high-quality DNA are crucial. Following microdissection, most samples fail to produce sufficient quantities of DNA for genome-wide experiments. Various PCR-based amplification methods have been used, but these usually produce nonuniform representations of the genome. Bacteriophage Phi29 DNA polymerase random-primed DNA amplification is based on isothermal multiple displacement amplification. We sought to define the genome representation of this method in a bacterial artificial chromosome microarray comparative genomic hybridisation (aCGH) platform. Test genomic female DNA was amplified using Phi29 amplification at four different starting concentrations (0.5, 5, 10 and 50 ng). These products were combined with unamplified and amplified genomic female DNA as reference. In addition, 50 ng of DNA from five microdissected breast cancer frozen samples, were amplified using the same method. Three combinations were performed: unamplified test with unamplified reference, amplified test with unamplified reference and both amplified tumour and reference DNA. aCGH was performed with an in-house 16 K BAC platform (a resolution of ~100 Kb). Pearson's correlation tests and hierarchical clustering were performed to compare the profiles obtained. aCGH profiles obtained with amplified test and unamplified reference female genomic DNA showed copy number biases throughout the genome. These biases were more conspicuous with smaller amounts of starting material and mapped to regions of known copy number polymorphisms. When similar concentrations of test and reference DNA were amplified, the biases were significantly reduced, rendering accurate profiles. For the tumours, representative profiles were obtained when both test and reference DNA were amplified. Phi29 amplification induces copy number biases and unamplified material remains the gold standard for copy number analysis. For accurate results using Phi29 amplification, samples subjected to aCGH analysis should be combined with reference DNA amplified with the same method, using similar amounts of starting template.

Laboratory Investigation (2007) 87, 75–83. doi:10.1038/labinvest.3700495

Keywords: Phi29; bias; DNA; microarray; CGH

Genome-wide techniques have played a major role in unravelling the molecular genetic pathways associated with tumorigenesis and tumour progression.^{1,2} Comparative genomic hybridisation (CGH) was the first molecular method to allow a systematic analysis of chromosomal copy number changes throughout the entire genome.³ This method has been instrumental in the characterisation of the

molecular genetic profiles of several types of human cancer.^{1,4–7} However, owing to the limited resolution of CGH, few tumour-suppressor genes or oncogenes have been identified using this technique. In recent years, microarray-based comparative genomic hybridisation (aCGH) has been devised to overcome the resolution limitation of CGH.^{1,5–7}

The potential of aCGH as a method to define the genomic profile of human tumours and to be used in the context of clinical outcome and treatment response prediction is enormous.^{1,5–7} This is particularly important as targeted treatments for cancer patients based on specific genomic alterations (eg, *HER2* amplification) are becoming clinically

Correspondence: Dr JS Reis-Filho, MD, PhD, The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK.
E-mail: Jorge.Reis-Filho@icr.ac.uk
Received 31 July 2006; revised and accepted 5 October 2006

available. Unlike expression profiling analysis, where variable proportions of tumour and stromal components can be subjected to microarray experiments, aCGH requires a purity of >75% of neoplastic cells.⁸ With the advent of laser capture microdissection,⁹ it has become possible to study the genetic features of limited number of cells or small lesions. The limiting step for coupling laser capture microdissection or other microdissection methods with microarray CGH has been the amount of DNA retrieved with these methods. For the study of breast cancer this has become of paramount importance, given that most diagnosis are currently made on core needle biopsies, leaving very limited material for molecular analysis.

Whole-genome amplification (WGA) methods have been developed in order to obtain adequate DNA yields with the highest possible fidelity to the original profile.^{10–14} PCR-based amplification methods, including degenerate oligonucleotide primer polymerase chain reaction (DOP-PCR)¹⁰ and single-cell comparative genomic hybridisation (SCOMP),¹¹ have been shown to provide a DNA yield sufficient for CGH analysis, however, some regions are reported to be preferentially amplified.^{12,13} Awareness of these biases and improvements in CGH technology, including the use of standard reference intervals for CGH analysis,^{15,16} has allowed the genetic study of amplified DNA. However, to reliably translate the increased resolution of microarray-based CGH into the identification of gene-specific copy number changes, unbiased amplification methods are required.

In this study, we sought to define the genome representation of Bacteriophage Phi29 DNA polymerase amplification in a high-density bacterial aCGH platform and to determine whether this method can be used to study the genome-wide profiles of microdissected samples.

Materials and methods

Samples and DNA Extraction

For the analysis of the accuracy of Phi29 in amplifying varying amounts of high-quality DNA, we used DNA extracted from peripheral blood lymphocytes of healthy female and male volunteers.

From 30 available breast cancer core biopsies, we selected five cases, where DNA extraction yielded >4 μ g, for the comparison between the genetic profiles obtained with unamplified DNA and Phi29 amplified products. Fresh frozen sections of the tumours were stained with nuclear fast red and microdissected with a sterile needle (Terumo corporation[®], Japan) under a stereomicroscope (Olympus SZ61[®], Japan) to obtain a percentage of tumour cells in the remaining tissue greater than 75%.⁵ This study has been approved by The Royal Marsden Hospital Research Ethics Committee.

Genomic DNA was extracted using a standard proteinase K digestion followed by phenol/chloroform extraction and resuspended in TE buffer pH:7.5.⁵ The concentration of the samples was measured by spectrophotometry (Ultrospec[®] 3100 pro, Amersham Bioscience, Amersham, UK).

Whole-Genome Amplification

GenomiPhi DNA amplification kit[®] was used following the manufacturer's instructions. Briefly, 1 μ l of template was mixed with 9 μ l of sample buffer. The mixture was denatured at 95°C for 3 min and immediately cooled at 4°C for a minimum of 10 min. A combination of 9 μ l of reaction buffer with 1 μ l of the enzyme was added to the cooled sample and then incubated at 30°C for 16 h overnight. Inactivation of the enzyme was performed by heating the sample for 10 min at 65°C which was then cooled to 4°C. Postamplification cleanup was achieved by ethanol precipitation using the sodium acetate/EDTA (1.5 M sodium acetate pH>8 and 250 mM EDTA) buffer. Samples were resuspended in 22 μ l of TE (pH: 7.5).

Quality Control PCR for Phi29 Amplification

Given that negative controls of Phi29 amplified reactions also show a smear in agarose gels, we determined whether the amplified products contained actual genomic sequences by performing a PCR,⁷ using primers specific for GAPDH and ER alpha genes: GAPDH (F: acagtccatgccatcactgc; R: gctgacaaagtggctcgtg; 400 bp) and ER alpha gene (F: gggagaatgttgaaacacaag; R: ccagttgatcatgtgaaccag; 300 bp). A total reaction volume of 30 μ l comprised 100 ng of genomic DNA, 1 \times Taq buffer, 1.5 mM MgCl₂, 0.2 mM dNTPs, and 0.133 μ M of each forward and reverse primers and 1 U of Taq polymerase (Invitrogen, Paisley, UK). PCR reactions were performed in a Peltier thermal cycler (PTC-225) starting with 4 min at 94°C and 34 cycles (1 min at 94°C 1 min at 55°C and 3 min at 72°C, followed by 7 min at 72°C, ending at 15°C). Products were run in a 2% agarose gel electrophoresis with ethidium bromide staining.

Experimental Design

To determine whether Phi29 produces unbiased genome-wide amplification with different amounts of starting DNA, we used as reference either unamplified female genomic DNA or Phi29 amplified female genomic DNA at a concentration of 50 ng/ μ l. We then tested Phi29 amplified female genomic samples at concentrations of 0.5, 5, 10 and 50 ng/ μ l. We also performed experiments combining male genomic and female genomic DNA, in order to define the log₂ ratios in the presence of a known 1:2 copy number ratio (ie, chromosome X in male vs

Table 1 List of experiments performed with female and male genomic DNA at different starting concentrations of original DNA

Sample	Test DNA	Reference DNA
F1F1, F2F2	Unamplified female	Unamplified female
G0.5GN	Amplified female 0.5 ng/ μ l	Unamplified female
G0.5GG	Amplified female 0.5 ng/ μ l	Amplified female 50 ng/ μ l
G0.5G0.5	Amplified female 0.5 ng/ μ l	Amplified female 0.5 ng/ μ l
G5GN	Amplified female 5 ng/ μ l	Unamplified female
G5GG	Amplified female 5 ng/ μ l	Amplified female 50 ng/ μ l
G5G5	Amplified female 5 ng/ μ l	Amplified female 5 ng/ μ l
G10GN	Amplified female 10 ng/ μ l	Unamplified female
G10GG	Amplified female 10 ng/ μ l	Amplified female 50 ng/ μ l
G10G10	Amplified female 10 ng/ μ l	Amplified female 10 ng/ μ l
G50GN	Amplified female 50 ng/ μ l	Unamplified female
G50GG	Amplified female 50 ng/ μ l	Amplified female 50 ng/ μ l
MF	Unamplified male	Unamplified female
GMGN	Amplified male 50 ng/ μ l	Unamplified female
GMGG	Amplified male 50 ng/ μ l	Amplified female 50 ng/ μ l

female samples). For this purpose, we combined unamplified male genomic DNA with unamplified female genomic DNA, amplified male genomic DNA with unamplified female genomic DNA and both amplified DNA samples at the same starting concentration (50 ng/ μ l). Table 1 summarises the experiments we performed with male and female genomic DNA.

Following the findings from the experiments above, we combined female genomic DNA, both as test and reference, amplified using the same starting amount of DNA (0.5, 5, 10 ng).

We next tested whether the tumour DNA results obtained with Phi29 were accurate and representative, using the profiles obtained with unamplified DNA as the 'gold standard'. Each tumour was subjected to three separate analyses (Table 2): unamplified tumour DNA with unamplified reference (pooled healthy female genomic) DNA, Phi29 amplified tumour DNA with unamplified reference DNA, and Phi29 amplified tumour and reference DNA.

To study the genomic regions (comprising 3 or more consecutive BACs) consistently harbouring genomic gains or losses due to amplification bias, we selected the experiments where amplified healthy female genomic DNA was combined with unamplified healthy female genomic DNA as reference. Those regions that were consistently gained or lost in >2 experiments were considered as recurrent biases. These regions were subsequently mapped to define their GC content and we investigated whether they preferentially mapped to regions of known copy number polymorphisms. A frequency plot was generated to illustrate each BAC showing Phi29 amplification-induced bias in >2 experiments (Figure 5).

Microarray CGH

The aCGH platform used for this study was constructed in the Breakthrough Breast Cancer

Table 2 List of experiments performed with breast cancer samples (all amplified products were at a starting concentration of 50 ng/ μ l)

Sample	Test	Reference
M67ANN	67A unamplified	Unamplified female DNA
M67AGN	67A amplified	Unamplified female DNA
M67AGG	67A amplified	Amplified female DNA
M148BNN	148B unamplified	Unamplified female DNA
M148BGN	148B amplified	Unamplified female DNA
M148BGG	148B amplified	Amplified female DNA
M152BNN	152B unamplified	Unamplified female DNA
M152BGN	152B amplified	Unamplified female DNA
M152BGG	152B amplified	Amplified female DNA
M218ANN	218A unamplified	Unamplified female DNA
M218AGN	218A amplified	Unamplified female DNA
M218AGG	218A amplified	Amplified female DNA
M229ANN	229A unamplified	Unamplified female DNA
M229AGN	229A amplified	Unamplified female DNA
M229AGG	229A amplified	Amplified female DNA

Research Centre and comprises ~16 000 clones, spaced at approximately 100 kb throughout the genome and spotted onto Corning GAPSII-coated glass slides (Corning, NY, USA). Labelling, hybridisation and washes were carried out essentially as previously described.^{5,17,18} Briefly, 400 ng of test and reference genomic DNA were labelled with Cy3-dCTP or Cy5-dCTP (Amersham Bioscience, Amersham, UK) using a Bioprime labelling kit (Invitrogen, Paisley, UK) according to the manufacturers protocol modified to incorporate 0.6 mM dCTP, and 1.2 mM dATP, dGTP and dTTP. The unincorporated nucleotides were removed with MinElute purification columns (Qiagen Ltd, West Sussex, UK). All experiments were performed in duplicate ('dye swaps') to minimise dye-biases.

Image Acquisition and Data Analysis

Slides were scanned using an Axon 4000B scanner (Axon Instruments, Burlingame, CA, USA) and images were processed using Genepix Pro 3.0 image analysis software (Axon Instruments). The log₂ ratios were normalised for spatial and intensity-dependent biases using a two-dimensional loess regression and then averaged across the 'dye-swaps'. This left a final data set of 13 263 clones with unambiguous mapping information according to the March 2006 build of the human genome (hg17). Data were smoothed using a local polynomial adaptive weights smoothing (aws) procedure for regression problems with additive errors.^{5,17} Thresholds for defining genomic gains and losses were obtained using data from unamplified female vs female and female vs male genomic DNA, as previously described.^{5,18} A categorical analysis was applied to each clone on the array after classification as gain, loss or no-change according to their smoothed log₂ ratio values. Smoothed log₂ ratio values < -0.182 were categorized as losses, those > 0.182 as gains, and those in between as unchanged. Pearson's

correlation test was used to compare the profiles obtained with amplified and unamplified experiments performed for the same tumour. Identical hybridisations of amplified material (test and reference) from two of the tumour samples were subjected to array CGH on different days on different batches of array CGH slides and used as a control for an accurate interpretation of Pearson's correlations results. Unsupervised clustering was performed on smoothed and 'thresholded' data using Ward's method based on Euclidean distance. Data processing and analysis was carried out in R 2.0.1 (<http://www.r-project.org/>) and BioConductor 1.5 (<http://www.bioconductor.org/>), making extensive use of modified versions of the packages aCGH, marray and aws in particular.

Results

DNA Extraction, Amplification and PCR Quality Control

All tumour samples had greater than 75% invasive component after microdissection. The DNA extraction yielded total products of around 1.4–9 μg (median = 3.25 μg), with an OD260/280 ratio >1.8.

Amplification of genomic female DNA with 0.5, 5, 10 and 50 ng yielded equivalent amounts of final product for all amplifications ranging from 4.3 to 6.4 μg (median = 5.35 μg ; 100- to 10 000-fold amplification), suggesting that fold amplification was inversely correlated with the starting amount of DNA.

Amplification of the tumour samples was carried out from a starting concentration of 50 ng. Genomi-Phi[®] provided a 90- to 108-fold amplification, with the amount of DNA obtained ranging from 4.4 to 5.4 μg (median = 4.85 μg). Gene-specific PCR for amplified genomic DNA and tumour samples yielded products of the expected sizes (400 bp length), whereas products obtained in the negative controls failed to show specific bands, rather a smear was observed (Supplementary Figure 1).

Assessment of Fidelity of Phi29 WGA

The profiles obtained with Phi29 amplified female genomic DNA samples varied according to the starting DNA concentration and the use of amplified

reference and test DNA. Analysis of the molecular genetic profiles obtained with starting amounts of 0.5 and 5 ng DNA revealed significant biases induced by Phi29 amplification, when combined with unamplified reference (Table 3, Figure 1a, d, g, j and m). When amplified test samples were hybridised with a standard amplified reference (50 ng), a significant stepwise decrease in the percentage of BACs showing either gains or losses was observed with increasing amounts of starting material (Table 3, Figure 1b, e, h, k and n). The bias was completely cancelled out when combining both test and reference amplified from 50 ng (Figure 1k). In all experiments, these biases were more conspicuous in the comparisons carried out using non-amplified reference DNA compared to those with amplified reference from 50 ng of DNA.

Given that the Phi29 amplification bias was completely cancelled out in the experiment using independent reactions with 50 ng of test and reference starting material (Table 3, Figure 1k), we next tested how the bias would be affected when matched starting amounts of DNA were used for both reference and test. We observed that <0.5% of the BACs showed Phi29 amplification-induced copy number biases when products from separate reactions with the same starting amount were combined (Table 3, Figure 1c, f, i, k and n). Optimal results when using amplified material were obtained in the experiments where both test and reference were amplified using the same amount of starting DNA (Figure 1c, f, i, k and n).

We then performed hierarchical cluster analysis for all combinations of genomic DNA experiments and found that unamplified test and reference samples segregated together with those carried out with both amplified test and reference DNA using the same starting amount of DNA (Figure 2).

Breast Tumours

Based upon the results obtained with female genomic DNA, the starting concentration for all tumour samples before amplification was 50 ng/ μl . In all cases, the combination of both amplified test and reference had a higher Pearson's correlation value (r^2) than the combination of amplified with

Table 3 Number of clones and percentage of the genome showing Phi 29-induced apparent copy number changes

Reference	Bias n clones (% genome)		
	Unamplified DNA	Amplified (50 ng general ref)	Matched amount of DNA ref
<i>Test</i>			
Phi amp 0.5 ng	3500 (26.39)	6193 (46.69)	62 (0.47)
Phi amp 5 ng	817 (6.16)	1669 (12.58)	5 (0.04)
Phi amp 10 ng	479 (3.61)	198 (1.49)	60 (0.45)
Phi amp 50 ng	354 (2.67)	0 (0)	0 (0)

Gains and losses were defined according to the thresholds of 0.182 and -0.182, respectively.

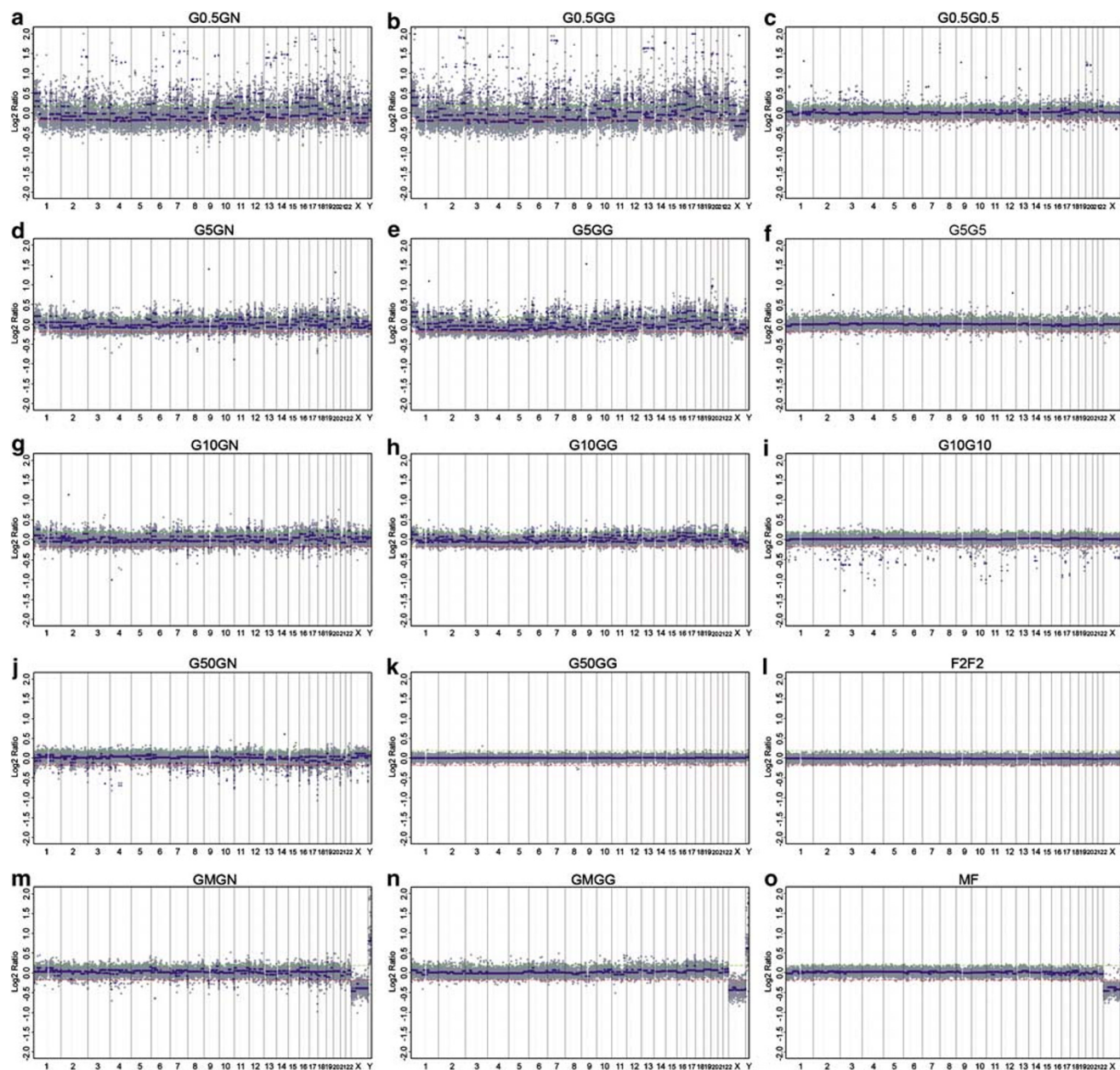


Figure 1 Evaluation of amplification bias using BAC aCGH. Representative genome plots of each experiment are illustrated. \log_2 ratios and aws-ratios are plotted in grey and blue, respectively, on the y axis against each clone according to genomic location on the x axis. The centromere is represented by a vertical dotted line. Horizontal dashed lines correspond to ratios of 0.182 and -0.182 . Experiments were performed using female genomic DNA amplified from four different starting amounts of material (rows 1–4: 0.5, 5, 10, 50 ng) and three test-reference combinations (columns 1–3): amplified test with unamplified reference (**a**: G0.5GN, **d**: G5GN, **g**: G10GN, **j**: G50GN); amplified test with amplified ‘standard’ reference with a starting amount of 50 ng of DNA (**b**: G0.5GG, **e**: G5GG, **h**: G10GG, **k**: G50GG); and both test and reference amplified from the same starting amount of template (**c**: G0.5G0.5, **f**: G5G5, **i**: G10G10). Row 5 shows the male DNA (test)-female DNA (reference) combinations: GMGN (**m**): amplified test with unamplified reference; GMGG (**n**): both test and reference amplified from 50 ng of starting template; MF (**o**): both test and reference un-amplified. F2F2 (**l**): unamplified female DNA as test and reference. Note that the genome plots show biased profiles when combining amplified with unamplified reference (**a**, **d**, **g**, **j**, **m**), particularly in the experiments using lower amounts of DNA as template (**a**, **d**, **g**). These biases are cancelled out when combining both test and reference DNA amplified from the same starting amount of DNA (**c**, **f**, **i**, **k**, **n**).

unamplified material. When profiles obtained with unamplified DNA were compared with those obtained with amplified test and reference, we found a very good correlation in two tumours ($r^2 > 0.8$) (Figure 3), good correlation in two cases ($r^2 > 0.7$)

and moderate correlation for the last case ($r^2 = 0.5$) (M67A). The correlation for the control experiments performed in different days was in both cases $r^2 \geq 0.9$. We then performed hierarchical clustering with all experiments. As expected, all experiments

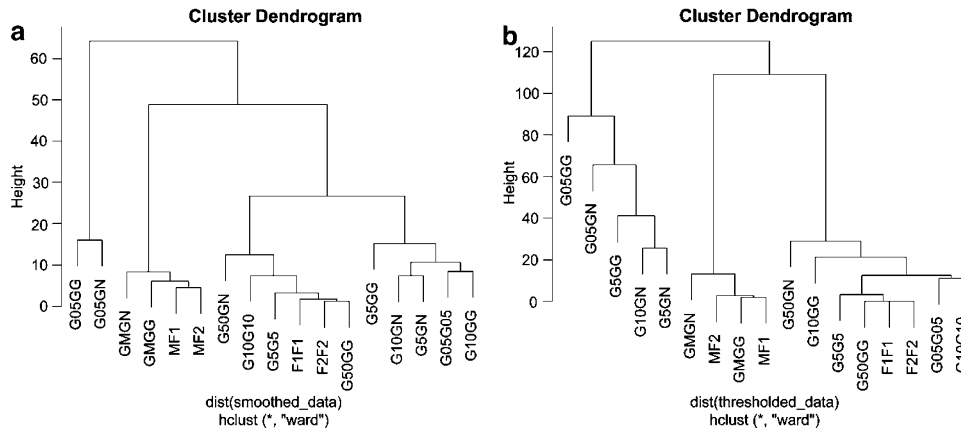


Figure 2 Unsupervised hierarchical clustering of aws-smoothed (a) and ‘thresholded’ (b) data from the genomic female and male DNA experiments using Ward’s method based on Euclidean distance. The dendrograms show that the unamplified female combinations (F1F1, F2F2) considered gold-standard segregates together with those experiments where amplified test was combined with amplified reference using the same starting amount of DNA. G0.5GN, G5GN, G10GN, G50GN: amplified test at the described starting amount of DNA with unamplified reference. G0.5GG, G5GG, G10GG, G50GG: amplified test at the described starting amount of DNA with amplified reference from 50 ng of starting template. G0.5G0.5, G5G5, G10G10: both test and reference amplified from the same starting amount of template (as described).

from those tumours with good-to-very good correlation scores segregated together (Figure 4).

Assessment of the Phi29-Induced Genomic Bias

To define the regions with recurrent copy number biases, experiments combining amplified female genomic DNA and unamplified female genomic reference DNA were used. Applying the thresholds of 0.182 (‘gains’) and -0.182 (‘losses’), recurrent increases in copy number due to amplification bias (‘gains’) were more frequent than decreases in copy number (‘losses’). We found 62 regions of consistent gain and 10 regions of loss (Figure 5 and Supplementary Table 1).

We then investigated whether regions of amplification bias mapped to genomic regions with known copy number variations (CNVs) (<http://projects.tcag.ca/variation/>). Thirty-eight out of the 62 regions of ‘gains’ (60%) mapped to known CNV regions. Eighty per cent of ‘losses’ consistently mapped to previously described CNV. These regions of misrepresentation were found to be GC-rich, with a mean GC of 45.6% (s.d. = 22.8, median = 45.46%).

Discussion

Bacteriophage Phi29 DNA polymerase amplification is based on an isothermal reaction that uses random hexamer primers to anneal to the genomic template at multiple sites to initiate replication.^{19,20} This isothermal technique is reported to have several advantages when compared to PCR-based methods, including the linear amplification of the DNA and the proofreading property of the enzyme, which leads to an increased accuracy and therefore, guarantees higher fidelity of the process.²¹ As synthesis proceeds, strand displace-

ment of complementary DNA generates new single-stranded DNA available to be primed by additional primers. Although the Phi29 WGA clearly results in recurrent genomic misrepresentations, this technique is reported to have a threefold error rate for over/under representation, which is significantly lower than the 1000-fold reported for PCR-based WGA methods.²⁰ Moreover, Phi29 WGA has been successfully applied to SNP and sequencing studies.^{22,23} On the other hand, a previous microarray CGH study demonstrated that Phi29 induces representational distortion, likely to be resultant from variability in priming density and processing of repetitive and polymorphic sequences.²⁴

In our hands, Phi29 amplification provided 90 (for 50 ng)—10 000 (for 0.5 ng) fold amplification, depending on the amount of starting template. Given the higher fold amplification with lower amounts of starting material, the highest amplification levels and proportionally highest yields were obtained with the lowest starting amount of material (ie, 0.5 ng). However, our study demonstrates that Phi29 amplification induces misrepresentation of several regions, showing artefactual ‘gains’ and ‘losses’ on the array CGH profiles and that these amplification biases are more frequent when low amounts (0.5–10 ng) of starting template are used, suggesting that the bias is more conspicuous in reactions with higher fold amplification.

Our findings provide evidence to suggest that Phi29 copy number change biases are not random (ie, map to specific genomic regions) and depend on the starting amount of template DNA, given that these are corrected (ie, <0.5% of genomic biases) when both test and reference DNA are amplified using the same amount of starting material. Mutation analysis and single-nucleotide polymorphism studies have also demonstrated that Phi29-induced

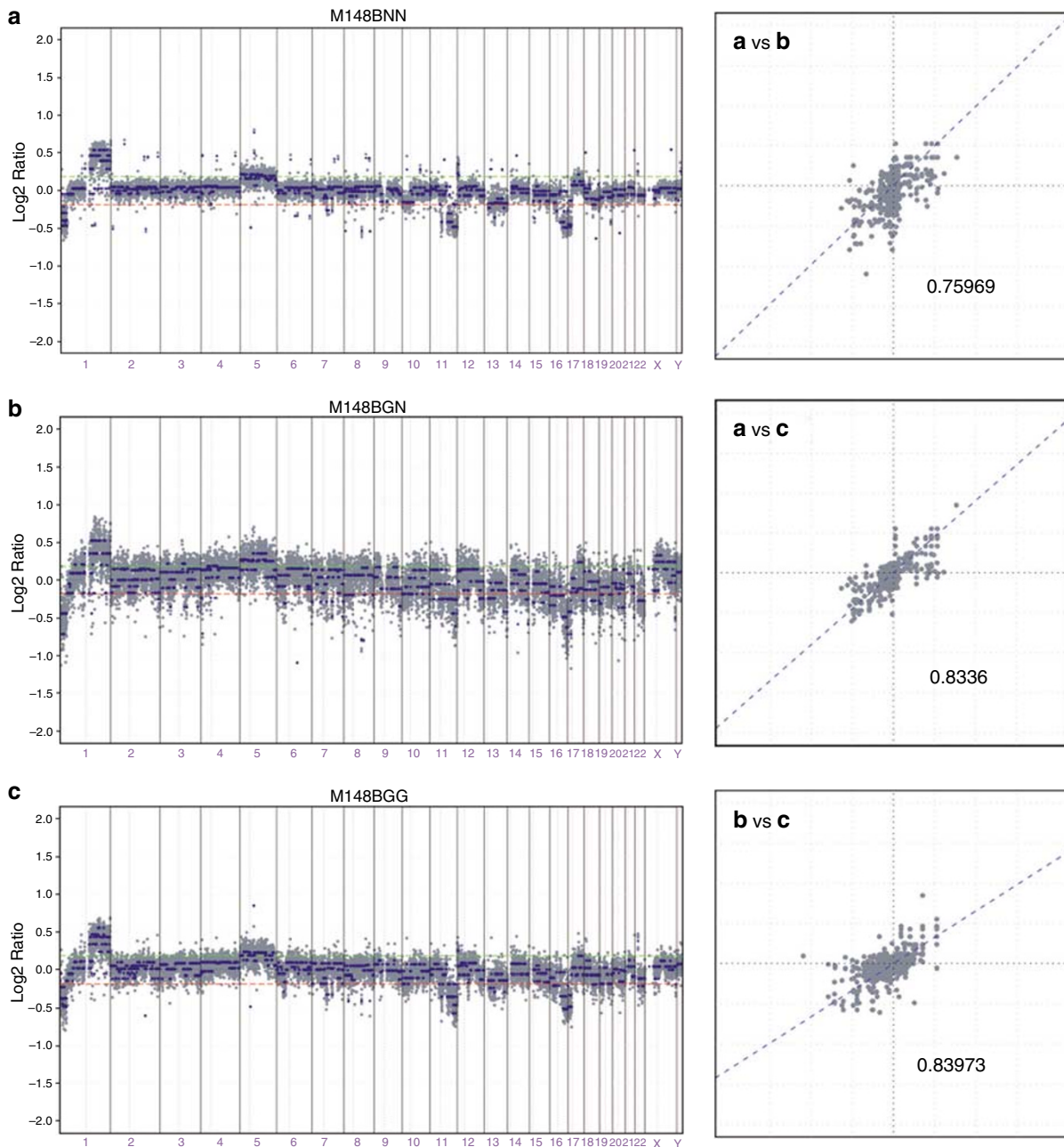


Figure 3 Evaluation of the correlation between the different combinations of tumour DNA with female reference. Left: Genome plots of sample M148B (**a**: unamplified test with unamplified reference; **b**: amplified test with unamplified reference; **c**: amplified test with amplified reference) log₂ ratios and aws-ratios are plotted in grey and blue, respectively, on the y axis against each clone according to genomic location on the x axis. The centromere is represented by a vertical dotted line. Horizontal dashed lines correspond to ratios of 0.182 and -0.182. Note that the experiments performed with both amplified test and reference DNA show a more similar profile to the experiment performed with unamplified test and reference. Right: correlation plot of log₂ ratios for each BAC clone between experiments, with linear regression and Pearson's correlations coefficients given. The correlation with the experiment using both unamplified test and reference is higher for the experiment using both amplified test and reference ($r^2 = 0.83$) compared to the one using unamplified reference ($r^2 = 0.76$).

copy number changes are nonrandom and can be effectively corrected using amplified reference DNA.^{22,25,26} Given that the bias is dependent on the starting amount of DNA and does not affect all loci in every sample, the statistical methods proposed by Cardoso *et al*²⁴ to circumvent this problem, may be effective in some but not all scenarios. However, our

results support the use of amplified reference^{22,25,26} to compensate for these areas of regional misrepresentation in order to obtain accurate results.

The high-resolution of the aCGH platform used in the present study allowed for an accurate mapping of the regions with Phi29 copy number biases. Using DNA from the same pooled healthy female source

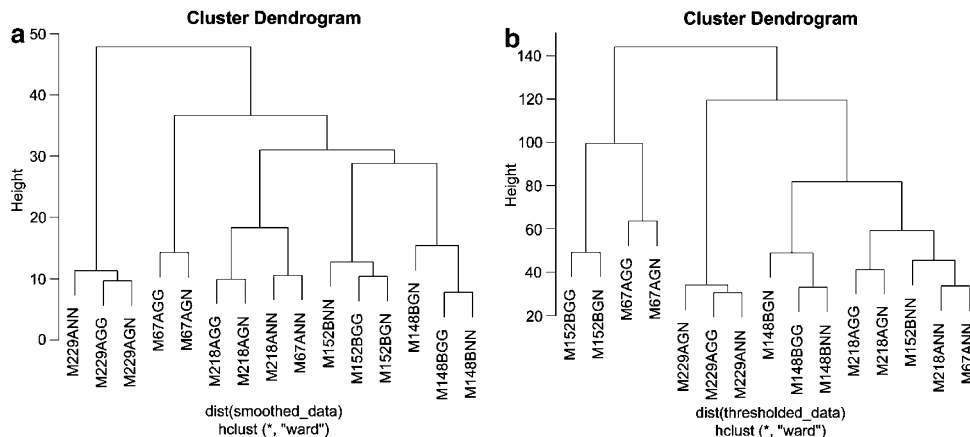


Figure 4 Unsupervised hierarchical clustering of smoothed (a) and 'thresholded' (b) data from microdissected breast cancer samples using Ward's method based on Euclidean distance. The experiments of the different combinations of the tumours segregate together, except for M67A, and also M152B when 'thresholded' data were used.

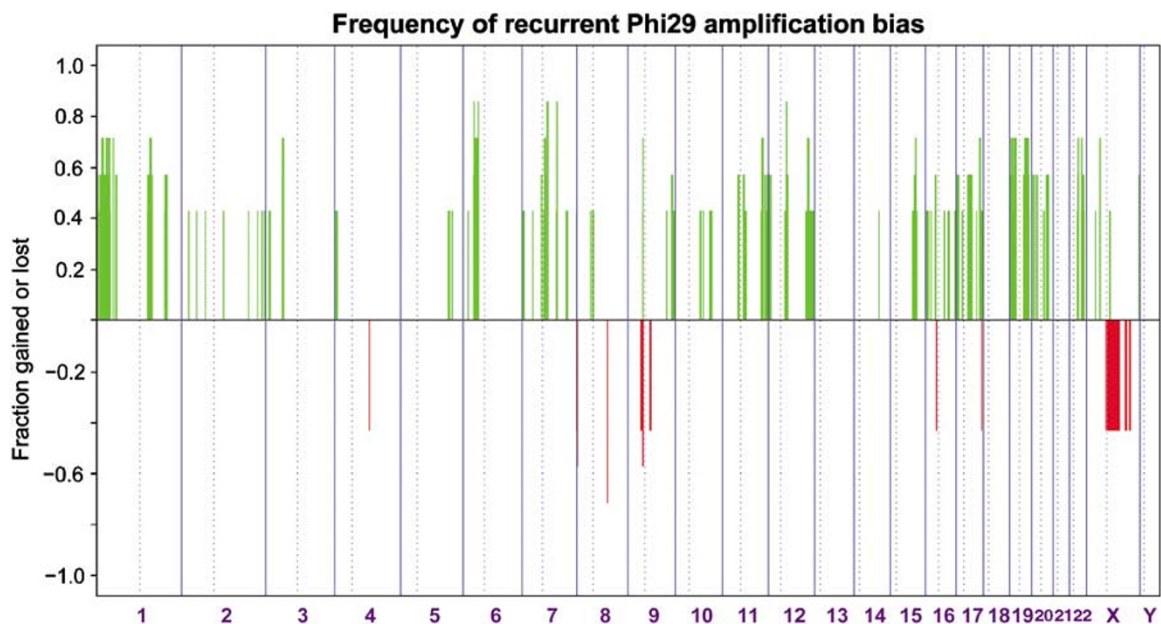


Figure 5 Frequency and genomic mapping of BACs showing recurrent (present in > 2 experiments: > 29%) Phi29 amplification-induced copy number biases. This figure was generated using the experiments where amplified test was combined with unamplified reference or when reference was amplified from 50 ng (G0.5GN, G0.5GG, G5GN, G5GG, G10GN, G10GG and G50GN). NB: G50GG was not included, given that reference and test were amplified from matched amounts of starting material. Individual BAC clones are plotted according to genomic location along the x axis. The proportion of tumours in which each clone is 'gained' (green bars) or 'lost' (red bars) is plotted along the y axis. Vertical dotted lines represent chromosome centromeres.

amplified in separate reactions, we have observed that a high percentage of these false 'gains' and 'losses' map to known CNVs.^{27–29} In addition, several loci of bias that did not correspond to known CNVs, mapped either to flanking regions of CNVs or telomeric (eg, 5q35, 17q25) and pericentromeric areas (eg, 11q13.1, 16p11) (Figure 5). In previous reports, regions of recurrent amplification bias have been located in areas of highly repetitive elements,^{22,25,26} however, this is the first time that these biases have been shown to map to known CNVs in the human genome. Although the underlying mechanism of this preferential amplification

of specific areas by Phi29 polymerase remains unclear, it appears to be related with the fold amplification rates and to the amount of starting material.

Bredel *et al*²⁵ have observed a correlation between the GC content and the biases (underrepresentation) induced by Phi29 amplification. This enrichment in the GC content had been shown to affect the efficiency of polymerase reaction amplification.³⁰ In our study, the regions of misrepresentation also showed enrichment for GC repeats (average of 45.6%), which is slightly higher than that of whole-genome (41%) GC content.

In conclusion, although Phi29 amplification provides efficient WGA, this method still induces copy number biases on BAC aCGH analysis, which can be reduced with increased amounts of starting material and almost completely cancelled out by using amplified test and reference DNA from the similar starting amounts of template. However, unamplified material rendered the best results and should still be considered the 'gold standard' for molecular genetic profiling with microarray-based comparative genomic hybridisation.

References

- Albertson DG, Collins C, McCormick F, *et al*. Chromosome aberrations in solid tumors. *Nat Genet* 2003;34:369–376.
- Weber BL. Cancer genomics. *Cancer Cell* 2002;1:37–47.
- Kallioniemi A, Kallioniemi OP, Sudar D, *et al*. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;258:818–821.
- Tirkkonen M, Tanner M, Karhu R, *et al*. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* 1998;21:177–184.
- Reis-Filho JS, Simpson PT, Jones C, *et al*. Pleomorphic lobular carcinoma of the breast: role of comprehensive molecular pathology in characterization of an entity. *J Pathol* 2005;207:1–13.
- Albertson DG, Pinkel D. Genomic microarrays in human genetic disease and cancer. *Hum Mol Genet* 2003;12(Spec no 2):R145–R152.
- 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.
- Weiss MM, Hermsen MA, Meijer GA, *et al*. Comparative genomic hybridisation. *Mol Pathol* 1999;52:243–251.
- Bonner RF, Emmert-Buck M, Cole K, *et al*. Laser capture microdissection: molecular analysis of tissue. *Science* 1997;278:1481, 3.
- Little SE, Vuononvirta R, Reis-Filho JS, *et al*. Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA. *Genomics* 2006;87:298–306.
- Stoecklein NH, Erbersdobler A, Schmidt-Kittler O, *et al*. SCOMP is superior to degenerated oligonucleotide primed-polymerase chain reaction for global amplification of minute amounts of DNA from microdissected archival tissue samples. *Am J Pathol* 2002;161:43–51.
- Cheung VG, Nelson SF. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc Natl Acad Sci USA* 1996;93:14676–14679.
- Telenius H, Carter NP, Bebb CE, *et al*. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 1992;13:718–725.
- Obermann EC, Junker K, Stoehr R, *et al*. Frequent genetic alterations in flat urothelial hyperplasias and concomitant papillary bladder cancer as detected by CGH, LOH, and FISH analyses. *J Pathol* 2003;199:50–57.
- Kirchhoff M, Gerdes T, Rose H, *et al*. Detection of chromosomal gains and losses in comparative genomic hybridization analysis based on standard reference intervals. *Cytometry* 1998;31:163–173.
- Kirchhoff M, Gerdes T, Maahr J, *et al*. Automatic correction of the interfering effect of unsuppressed interspersed repetitive sequences in comparative genomic hybridization analysis. *Cytometry* 1997;28:130–134.
- Natrajan R, Williams RD, Hing SN, *et al*. Array CGH profiling of favourable histology Wilms tumours reveals novel gains and losses associated with relapse. *J Pathol* 2006;210:49–58.
- Reis-Filho J, Pinheiro C, Lambros M, *et al*. EGFR amplification and lack of activating mutations in metastatic breast carcinomas. *J Pathol* 2006;209:445–453.
- Blanco L, Bernad A, Lazaro JM, *et al*. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem* 1989;264:8935–8940.
- Dean FB, Hosono S, Fang L, *et al*. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 2002;99:5261–5266.
- Esteban JA, Salas M, Blanco L. Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J Biol Chem* 1993;268:2719–2726.
- Paez JG, Lin M, Beroukhim R, *et al*. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* 2004;32:e71.
- Rook MS, Delach SM, Deyneko G, *et al*. Whole genome amplification of DNA from laser capture-microdissected tissue for high-throughput single nucleotide polymorphism and short tandem repeat genotyping. *Am J Pathol* 2004;164:23–33.
- Cardoso J, Molenaar L, de Menezes RX, *et al*. Genomic profiling by DNA amplification of laser capture microdissected tissues and array CGH. *Nucleic Acids Res* 2004;32:e146.
- Bredel M, Bredel C, Juric D, *et al*. Amplification of whole tumor genomes and gene-by-gene mapping of genomic aberrations from limited sources of fresh-frozen and paraffin-embedded DNA. *J Mol Diagn* 2005;7:171–182.
- Lage JM, Leamon JH, Pejovic T, *et al*. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res* 2003;13:294–307.
- Freeman JL, Perry GH, Feuk L, *et al*. Copy number variation: new insights in genome diversity. *Genome Res* 2006;16:949–961.
- Iafate AJ, Feuk L, Rivera MN, *et al*. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949–951.
- Sharp AJ, Locke DP, McGrath SD, *et al*. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005;77:78–88.
- Arezi B, Xing W, Sorge JA, *et al*. Amplification efficiency of thermostable DNA polymerases. *Anal Biochem* 2003;321:226–235.

Supplementary Information accompanies the paper on the Laboratory Investigation website (<http://www.nature.com/labinvest>)