Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples

Gang Wang, Cameron Brennan[1], Martha Rook[2], Jia Liu Wolfe[2], Christopher Leo[3], Lynda Chin[1], Hongjie Pan, Wei-Hua Liu, Brendan Price and G. Mike Makrigiorgos*

Department of Radiation Oncology, [1]Department of Medical Oncology and [3]Arthur and Rochelle Belfer Cancer Genomics Center, Dana Farber Cancer Institute, Harvard Medical School, Boston, MA, USA and [2]Variagenics Inc. Cambridge, MA, USA

ABSTRACT

Analysis of genomic DNA derived from cells and fresh or fixed tissues often requires whole genome amplification prior to microarray screening. Technical hurdles to this process are the introduction of amplification bias and/or the inhibitory effects of formalin fixation on DNA amplification. Here we demonstrate a balanced-PCR procedure that allows unbiased amplification of genomic DNA from fresh or modestly degraded paraffin-embedded DNA samples. Following digestion and ligation of a target and a control genome with distinct linkers, the two are mixed and amplified in a single PCR, thereby avoiding biases associated with PCR saturation and impurities. We demonstrate genome-wide retention of allelic differences following balanced-PCR amplification of DNA from breast cancer and normal human cells and genomic profiling by array-CGH (cDNA arrays, 100 kb resolution) and by real-time PCR (single gene resolution). Comparison of balanced-PCR with multiple displacement amplification (MDA) demonstrates equivalent performance between the two when intact genomic DNA is used. When DNA from paraffin-embedded samples is used, balanced PCR overcomes problems associated with modest DNA degradation and produces unbiased amplification whereas MDA does not. Balanced-PCR allows amplification and recovery of modestly degraded genomic DNA for subsequent retrospective analysis of human tumors with known outcomes.

INTRODUCTION

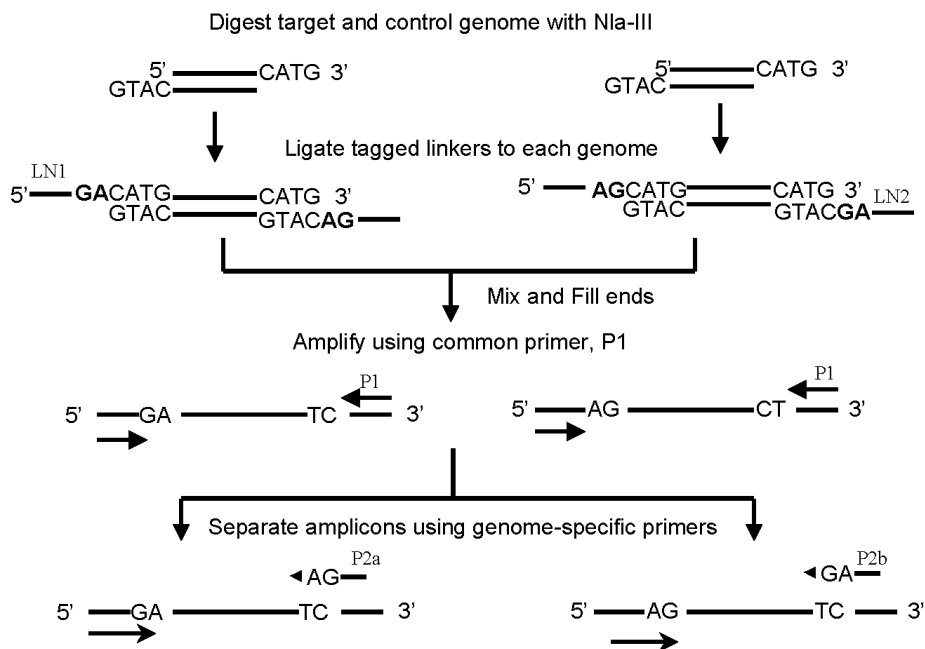Genetic profiling-based diagnosis promises to refine (1) and potentially revolutionize (2) the existing cancer staging system and the management of early disease. Array-based comparative genomic hybridization (array-CGH) offers global views of cancer genomes by detecting amplification or deletion of cancer genes (3–10), whereas techniques like real-time PCR (11) can be used for validation and quantification of the identified genomic changes.

However, such multiplexed analysis of genetic changes in tumors requires 'micrograms' of pure tumor DNA (12,13). Routine tumor biopsies often consist of heterogeneous mixtures of stromal cells plus tumor cells with a wide range of genetic profiles (14). Techniques such as fine needle aspiration and laser capture microdissection (LCM), allow for removal of minute amounts of fresh or archived tumor tissue (14), thereby isolating homogeneous populations of normal or tumor cells (15–17). DNA extracted from such a small number of cells has to be amplified to provide sufficient material for microarray screening. Whole genome amplification may be carried out via conventional PCR. In fact, PCR may amplify whole genomic DNA from as little as a single cell (13,18). However, the exponential mode of DNA amplification, the concentration-dependent PCR saturation and the lack of reproducibility due to stray impurities are notorious for the introduction of bias (11). Consequently, different quantitative relationships between two genes are usually observed before and after PCR amplification. Whole genome amplification methods other than PCR have been described [reviewed in (19)], including the promising multiple displacement amplification (MDA) (20). MDA operates on long DNA templates and produces linearly amplified genomic DNA when starting from intact genomes obtained from cell cultures or fresh tissue. However, the amplification efficiency of MDA is diminished as the molecular weight of the starting material decreases, which is problematic for amplification of formalin-fixed archival DNA or low molecular weight DNA from deteriorated forensic samples (21).

Here we describe a PCR-based approach to amplify genomic DNA of two different origins, one from cancer cells and another from normal cells. This method does not require intact, long genomic DNA as starting material and

---

*To whom correspondence should be addressed at Dana Farber—Brigham and Women's Cancer Center, Brigham and Women's Hospital, Level L2, Radiation Therapy, 75 Francis Street, Boston, MA 02115, USA. Tel: +1 617 525 7122l; Fax: +1 617 587 6037; Email: mmakrigiorgos@partners.org
Present address:
Jia Liu Wolfe, Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA

Digest target and control genome with Nla-III

5' ═══CATG 3'   5' ═══CATG 3'
GTAC═══   GTAC═══

Ligate tagged linkers to each genome

LN1
5'══**GA**CATG═══CATG 3'   5'══**AG**CATG═══CATG 3' LN2
GTAC═══GTAC**AG**══   GTAC═══GTAC**GA**══

Mix and Fill ends

Amplify using common primer, P1

                                    P1   P1
5' ══GA ═══TC ══ 3'   5' ══AG ═══CT ══ 3'

Separate amplicons using genome-specific primers

                ◄AG═ P2a   ◄GA═ P2b
5' ══GA ═══TC ══ 3'   5' ══AG ═══TC ══ 3'

Figure 1. Protocol used for the unbiased amplification of two genomic DNAs via balanced-PCR.

allows removal of amplification bias caused by PCR saturation and impurities down to the single gene level. Genomic DNA is first digested with a 4 bp cutting restriction nuclease. Following ligation of composite linkers to the two DNAs, the samples are mixed and PCR amplified in a single tube (Fig. 1). The single tube amplification of the mixed samples is aimed at eliminating PCR biases related to PCR saturation and impurities, since the polymerase cannot distinguish among alleles originated from normal or cancer genomes. A nested, genome-specific primer is subsequently used in a low-yield, second PCR to re-separate DNA fragments from the two original genomes on the basis of nucleotide 'tags' incorporated in the composite linkers. We previously demonstrated the utility of this balanced-PCR approach for the unbiased amplification of cDNA prior to gene expression microarray screening (22). The increased complexity of genomic DNA relative to cDNA required modification of our original approach. We describe an improved single tube procedure that allows application of balanced-PCR to genomic DNA obtained from about 1000 cells, and we demonstrate its use for array-CGH and real-time PCR quantification of gene copy numbers from normal and breast cancer cells and for modestly degraded DNA obtained from paraffin-embedded tissue.

MATERIALS AND METHODS

Cell lines and genomic DNA

Breast cancer cells BT-474 and human mammary epithelial cells (HMEC) were obtained from the American Type Culture Collection (Manassas, VA) and from Cambrex (Rockland, ME), respectively, and were cultured as per the companies' recommendations. Total genomic DNA was then isolated from cultured cells using the QIAamp™ DNA Mini Kit (Qiagen, Valencia, CA). Genomic DNA from paraffin-embedded tissue was extracted using the Qiagen EZ1™ paraffin kit.

Single tube procedure for balanced-PCR

The linkers and primers used for the balanced-PCR protocol in Figure 1 were synthesized by Oligos Etc. Inc. (Wilsonville, OR) and are depicted in Table 1. A single tube procedure was used for digestion and ligation of BT474 ('target') and HMEC ('control') genomic DNA with genome-specific linkers. Genomic DNA (5 ng) was digested in a 5 µl total reaction volume using restriction enzyme NlaIII (10 units/µl stock, 37°C, 2 h; New England Biolabs, Beverly, MA) in 1× buffer (50 mM Tris–HCl, pH 7.5, 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP, 25 µg/ml BSA). NlaIII was subsequently inactivated by incubation at 70°C for 1 h. Composite linkers LN1 and LN2 (0.3 µl from a 2.8 µg/µl stock in a 10 µl reaction volume) were then ligated to DNA from BT474 (target) and HMEC (control) cells, respectively, using T4 DNA ligase (New England Biolabs) at room temperature for 1 h. After inactivation of ligase at 65°C for 40 min, the linker-ligated target and control DNAs were mixed.

The DNA mixture was PCR-amplified using the common oligonucleotide P1 in a Tech-Gene™ PCR thermocycler (Techne, Princeton, NJ) with Advantage 2 DNA polymerase (BD Biosciences, Palo Alto, CA). Thermocycling conditions were: 8 min at 72°C; 1 min at 95°C; 20 × (30 s at 95°C and 60 s at 72°C); 5 min at 72°C. Following thorough DNA purification with a QIAquick™ PCR Purification Kit to remove unincorporated primer P1, PCR products were quantified using a PicoGreen assay (Molecular Probes, Eugene, OR). To re-separate PCR products originating from target and control genomes, a low-yield PCR was carried out using primers P2a (BT474 target genome) or P2b (HMEC control

Table 1. Linkers, probes and primers for PCR

	Linkers and primers for balanced PCR
LN1	AACTGTGCTATCCGAGGGAAAGGACATG
LN2	AACTGTGCTATCCGAGGGAAAGAGCATG
P1	AGGCAACTGTGCTATCCGAGGGAA
P2a	AACTGTGCTATCCGAGGGAAAGGA
P2b	AACTGTGCTATCCGAGGGAAAGAG

Name, GI no.	Real-time PCR primers and probes	
HB-EGF, 29735304	Forward	CCCCAGTTGCCGTCTAGGA
	Reverse	CGGACATACTCTGTTTGGCACTT
	Probe	CCCATAATTGCTTTGCCAAAATACCAGAGC
HER2, 29739994	Forward	GGATGTGCGGCTCGTACAC
	Reverse	TGACATGGTTGGGACTCTTGAC
	Probe	ACTTGGCCGCTCGGAACGTGC
IL9R, 29746178	Forward	CCTTGTTGCTGTGTCCATCTTTC
	Reverse	CCTGGGCGACAGCTTGAA
	Probe	CCTGCTGACTGGCCCGACCTACC
E2F1, 17458490	Forward	TGGCTGGGCGTGTAGGA
	Reverse	CGCTCCATTAAAGCTTCAATCA
	Probe	AGAGCACTTCTGTCTTAAAGGTTTTTT
TBP, 27484631	Forward	GGGCATTATTTGTGCACTGAGA
	Reverse	AGCAGCACGGTATGAGCAACTGTCAGA
	Probe	CACCGCGCAGCGTGACTGT
RAN, 34194620	Forward	TGGAGCCCAGCGTCAGA
	Reverse	CGCTGCACCGCTGACAT
	Probe	TCTAGTTTTTATAGGCAGCTGTCCTGT
TOP1, 17484369	Forward	GACAGCCCCGGATGAGAAC
	Reverse	AAGAATTGCAACAGCTCGATTG
	Probe	TCCCAGCGAAGATCCTTTCTTATAACCGTG
TFR, 29728873	Forward	GCCAATGAGGTCTGAAATGGA
	Reverse	GGCCTTATTCCTGCAATCAACA
	Probe	CTTCTGCTGGATAAAATGAGGTTCAA
CYC, 29745697	Forward	GCCATGGAGCGCTTTGG
	Reverse	TCCACAGTCAGCAATGGTGATC
	Probe	TCCAGGAATGGCAAGACCAGCAAGA
GAPDH, 29744218	Forward	CGTCCTTGACTCCCTAGTGTC
	Reverse	CCGTAAAACCGCTAGTAGCC
	Probe	ATGGGAGGTGATCGGTGCTGGTT
HoxB5, 29738788	Forward	CCGAGAAGGAGTTTACAAAGT
	Reverse	CGCATACATAGCAAAACGAA
	Probe	CTTGATTTGTGGATCGTGGTCGTTA
PCK1, 17484369	Forward	CGAGAGAGAGATCCTTGCCTT
	Reverse	TTCAGATCTGCTCACGGTGT
	Probe	CAGTAGGAGCAAGAGAGGGCAAGTGTT
RAE1, 17484369	Forward	TATTTCCTATGTTTGGGGTG
	Reverse	CAAGACCCTTCTAAACCACT
	Probe	TGTACGAGTTGGTCTTAGCGGTATTG

genome) which contain two-nucleotide 'tags' at their ends that distinguish the two genomes. In each reaction, 1–2 ng from the first PCR product was amplified using the Titanium PCR kit (BD Biosciences) with the following thermocycling conditions: 1 min at 95°C; 10 × (30 s at 95°C and 60 s at 72°C); 5 min at 72°C. Alternatively, instead of BT474 DNA, the target DNA used for balanced-PCR amplification was DNA (10 ng) extracted from paraffin-embedded tissue.

The efficiency of NlaIII was routinely monitored during balanced-PCR, as previously described (22), and we have found that restriction digestion is >95% complete. The ligation efficiency was also monitored; however, this is somewhat less critical, since every sample is normalized to internal housekeeping genes (GAPDH) and therefore a reduced ligation

efficiency should affect both the housekeeping gene amplification and the particular gene tested.

Multiple displacement amplification (MDA)

MDA was performed for target (BT474) and control (HMEC) genomic DNAs using the Repli-g™ whole genome amplification kit (Molecular Staging, New Haven, CT) according to kit instructions. Briefly, 5 ng of either BT474 or HMEC genomic DNA was brought to a final volume of 2.5 μl with sterile, distilled water. A reaction master mix was prepared by adding 12.5 μl of 4× mix, 0.5 μl of DNA polymerase mix and 34.5 μl of sterile, distilled water. The reaction master mix was added to the DNA, and samples were incubated at 30°C for 16 h, following which the enzyme was

heat-denatured at 65°C for 3 min. The concentration of amplified samples was determined using a PicoGreen DNA quantification assay (Molecular Probes). Alternatively, the target DNA used for MDA amplification was DNA (50 ng) extracted from paraffin-embedded tissue.

Quantitation using real-time (TaqMan) PCR

Real-time PCR, TaqMan (23) assays, were performed to determine the relative copy number of specific genes in target DNA (BT474 or DNA from paraffin-embedded tissue) relative to control DNA (HMEC) for unamplified genomic DNA, balanced-PCR amplified DNA and MDA-amplified DNA. TaqMan assays were performed using AmpliTaq Gold™ (Applied Biosystems, Foster City, CA) in an ABI Prism 7900HT detection system. Some experiments were also performed using Platinum *Taq* DNA Polymerase (Invitrogen, Carlsbad, CA) in a Smart-Cycler™ (Cepheid, Sunnyvale, CA). Primers and probes for exonic regions of 13 genes (Table 1) were designed using Oligo software (v. 6.65, Molecular Biology Insights Inc., West Cascade, CO) and PrimerExpress software (Applied Biosciences, ABI, Foster City, CA) and were obtained from Bioresearch Technologies (Novato, CA). Three independent triplicates of quantitative PCR experiments were performed for each gene to generate an average relative copy number and standard deviation. For each triplicate, 3 ng of DNA was added to a final volume of 70 µl with a final concentration of $1\times$ ABI TaqMan master mix™, 4 µM each primer and 2 µM probe. This reaction mix was split into three different 20 µl PCRs and thermo-cycled. The cycling program was one cycle at 50°C for 2 min, one cycle at 95°C for 10 min, and 40 cycles at 95°C for 15 s and 60°C for 1 min. The relative genomic copy number was calculated using the comparative threshold (Ct) method (11). Briefly, the threshold cycle (C_T) for each gene was determined using the thermocycler software and the average of three independent Cts/DNA was calculated. The copy number of the target gene normalized to an endogenous reference and relative to calibrator is given by the formula $2^{-\Delta\Delta C_T}$. GAPDH was used as an endogenous reference, and ΔC_T was calculated by subtracting the average GAPDH C_T from the average C_T of the gene of interest. A variety of calibrator DNAs were used to calculate $\Delta\Delta C_T$ ($\Delta C_{T \text{ DNA of interest}} - \Delta C_{T \text{ calibrator DNA}}$). For BT474 or paraffin samples amplified via balanced-PCR, co-amplified HMEC DNA was used as a calibrator. For unamplified BT-474 or unamplified paraffin DNA, unamplified HMEC was used as calibrator. For MDA-amplified BT474 or paraffin DNA, MDA-amplified HMEC was used as a calibrator.

Array-CGH using cDNA microarrays

Array-based comparative genomic hybridization (Array-CGH) was performed on Agilent Human 1 cDNA microarrays using NlaIII digested DNA from unamplified BT474 and HMEC genomic DNA, balanced-PCR-amplified DNA, and MDA-amplified DNA. Alternatively, BT474 DNA was replaced with paraffin-extracted DNA. For each labeling reaction, 2 µg of digested DNA (amplified or unamplified) was used. Each sample pair was dye-swap labeled for hybridization. Briefly, DNA samples (2 µg) were denatured in the presence of Random Primer and Reaction Buffer (Invitrogen BioPrime Labeling Kit) at 98°C for 5 min, and then cooled to

2°C for 5 min. The denatured sample was incubated with Klenow fragment, dNTP mix (2.0 mM dATP dGTP dTTP, 1.0 mM dCTP in 10 mM Tris pH 8.0, 1 mM EDTA) and Cy3 or Cy5 dCTP nucleotides (1 mM; Perkin Elmer) for 2 h at 37°C. Reactions were terminated using EDTA (0.5 M, pH 8.0) Cy3 and Cy5 reaction pairs (labeled pair = Cy5-sample:Cy3-reference; reversed labeled pair = Cy3-sample:Cy5-reference) were pooled, precipitated and resuspended in 18.5 µl of 0.514% SDS. Samples were mixed with blocking solution concentrated from 50 µl of human Cot-1 DNA (1 mg/ml; Gibco), 20 µl of yeast tRNA (5 mg/ml; Gibco) and 4 µl (dA)–poly(dT) (5 mg/ml; Sigma). SSC was added to a final concentration of $3.4\times$ and 2.5 µl of Deposition Control Target (Operon) was added to a final volume of 30 µl. For hybridization, samples are denatured at 98°C for 2 min, then cooled at 37°C for 30 min under light-protection with foil. Labeled reactions in a volume of 27.5 µl were pipetted onto Agilent Human 1 cDNA arrays. Hybridization was carried out for 18–20 h in a 65°C water bath. After hybridization was complete, arrays were washed in $2\times$ SSC–SDS [100 ml of $20\times$ SSC, 0.03% SDS (10%) (v/v)] at 65°C for 5 min, followed by additional 5 min wash steps in $1\times$ SSC, then $0.2\times$ SSC, each at room temperature. After drying, hybridized arrays were scanned on an Axon scanner and spot finding and flagging were accomplished using GenePix software. Custom tools developed at the Belfer Center for Cancer Genomics (C. Brennan and L. Chin, manuscript in preparation) including cDNA-to-chromosome mapping, exclusion of non-reporters, ratio calculation, normalization and visualization were used to compile the CGH profiles from these array data points.

RESULTS

Single tube balanced-PCR protocol

We explored the application of balanced-PCR to the amplification of whole genomic DNA and the detection of changes in gene copy number via array-CGH and real-time PCR. The complex nature of genomic DNA required modification of the originally reported protocol developed for gene expression profiling (22), and a single tube approach was employed for DNA digestion and linker ligation. The single tube approach results to higher reproducibility when working with small amounts of DNA, since it avoids an intermediate purification step and is convenient to perform. NlaIII endonuclease is used to digest DNA (Fig. 1) to generate fragments that contain recessed 5′ ends and 3′ overhangs, which can be linker ligated without addition of an adaptor. This design feature allows the use of a single tube process without purification, because PCR artifacts are known to occur in the presence of excessive adaptors. The linker length has been reduced to 28 bp from the original 44 bp, since shorter linkers avoid PCR suppression effects by reducing hairpin formation (24). Distinction between the genome-specific primers P2a and P2b is based on two nucleotide 'tags' on their 3′ end (5′-AG-3′ versus 5′-GA-3′; Fig. 1). The two base mismatch at the 3′ end of the primers P2a and P2b prevents P2a from amplifying sequences from the LN1-ligated (target) genome and vice versa, while it retains similarity in the remaining part of the primer sequence. The lack of cross-talk between the genome-specific primers is
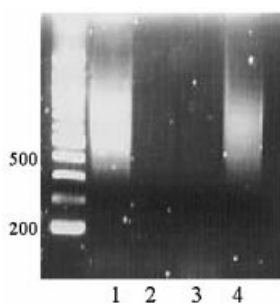
Figure 2. Evaluation of the specificity of primers P2a and P2b for amplifying target and control genomes ligated to LN1 and LN2, respectively. The protocol of Figure 1 was applied to co-amplify and, subsequently, to re-separate the two genomes. Lane 1, P2a-amplified genome; lane 4, P2b-amplified genome; lanes 2 and 3, the products depicted in lanes 1 and 4 were further amplified for 10 cycles using P2b and P2a primers, respectively.

demonstrated in Figure 2, where target and control genomic DNA were amplified as per the protocol in Figure 1 and then separated using primers P2a and P2b (lanes 1 and 4, respectively). The products of lanes 1 and 4 were subsequently amplified for 10 additional cycles using primers P2a and P2b (lanes 2 and 3, respectively), i.e. the 'wrong' primers. The lack of product in lanes 2 and 3 demonstrates the specificity of the two primers, P2a and P2b, for their respective genomes.

Reproducibility of array-CGH profiling

To evaluate the reproducibility of the overall procedure-balanced-PCR amplification plus array-CGH screening, the experiment was repeated two independent times starting with 5 ng each of HMEC and BT474 DNA. The results from replicate experiments were compared to derive an estimate of the combined errors due to random variations in the efficiency of digestion, ligation and balanced-PCR amplification, and signal differences/defects of individual cDNA microarrays. A generally good agreement was demonstrated between replicate experiments as depicted for chromosomes 17 and 20 in Figure 3. Concordance between the two sets of data was $R^2 = 0.51$, which increased substantially if nearest neighbor averaging was applied to the data ($R^2 = 0.71$, 0.79 and 0.87 for averaging signals by two, five and 12 nearest neighbors along each chromosome). Whether signals from neighbor chromosomal sites were averaged or not, genomic loci with relatively high gene-dosage alterations could still be detected with high reproducibility among different experiments (vide infra). These results indicate that the array signals tend to fluctuate randomly and signal variability is similar to the previously reported levels for replicate array-CGH experiments (21). To balance the need of improving signal reproducibility and preserving the highest resolution that microarrays can offer, a two-nearest neighbor averaging was applied in array-CGH data analysis. By following this approach, it was estimated that the average distance between successive chromosomal regions in the resulting data sets is ~300 kb.

Genomic copy number screening (array-CGH) of breast cancer cells

Gene copy number ratios in BT474 (target) and HMEC (control) genomic DNAs were compared to each other prior to

and after balanced-PCR amplification. First, 5 μg (~1 000 000 cells) of unamplified BT474 and HMEC genomic DNA was directly labeled and hybridized to cDNA microarrays and the resulting array-CGH profiles of copy number ratios are shown in Figure 4. The reported differences between the well studied BT474 breast cancer cell line and normal human female (HMEC) were reproduced in this comparison, including the multiple amplification regions in chromosomes 17q and 20q, the amplifications in chromosomes 9, 11 and 14 and the deletions in chromosome 10 previously observed by conventional CGH (4,25) and array-CGH (5,26). Next, 5 ng (~1000 cells) of genomic DNA from BT474 and HMEC cells was amplified using balanced-PCR and analyzed for comparative gene dosage via array-CGH (Fig. 5). The results demonstrate an overall pattern of gene amplifications and deletions resembling that of unamplified DNA (shaded areas in Fig. 5). The comparison was also performed using MDA-amplified material and the concordance among balanced-PCR amplified, MDA-amplified and unamplified samples was further analyzed for chromosomes 17 and 20 where marked gene dosage changes were observed. Figure 6 depicts two-nearest neighbor-smoothed gene dosage data for target (BT474) versus control female (HMEC) DNA for chromosomes 17 and 20 using these two amplification methods. It is evident that both balanced-PCR and MDA are capable of reproducing the major genetic changes occurring in the genome of the cancerous BT474 cells. For chromosome 17, array-CGH data demonstrated a correlation coefficient $R^2 = 0.67$ (two-nearest neighbor averaging) and $R^2 = 0.90$ (12-nearest neighbor averaging) when comparing fold change using balanced-PCR-amplified DNA with unamplified DNA. The same analysis conducted using MDA-amplified DNA (Fig. 6) generated $R^2 = 0.77$ (two-nearest neighbor averaging) and $R^2 = 0.88$ (12-nearest neighbor averaging). Comparable levels of concordance were also derived by analysis on chromosome 20. The concordance levels for balanced-PCR and MDA are similar to the concordance observed in the replicate-reproducibility studies depicted in Figure 3. Since replicate balanced-PCR experiments generated similar levels of concordance to that observed when amplified and unamplified samples are compared, it was concluded that the two amplification methods, balanced-PCR and MDA, did not introduce substantial bias during DNA amplification (i.e. amplification bias < array-CGH bias). Many of the genes included in the amplified regions of chromosomes 17 and 20 have a well established association with cancer. For example, RAE 1, PCK, HOX and HER2 are highly amplified in BT474 cells and are a prognostic marker for breast tumors (25,27,28). Amplification in these genes was clearly depicted among all replicate experiments in the array-CGH data for both of the amplification methodologies tested.

Real-time PCR measurement of gene copy number in target versus control cells

For many research and diagnostic applications, the array-CGH-identified gene copy number changes need to be further verified via real-time PCR. To evaluate the two amplification methodologies, balanced-PCR and MDA, on a gene-by-gene level, we chose genes that are located in chromosomal regions where gene amplification was observed in array-CGH profiling: HER2, PCK, RAE and HOX. Genes were also selected
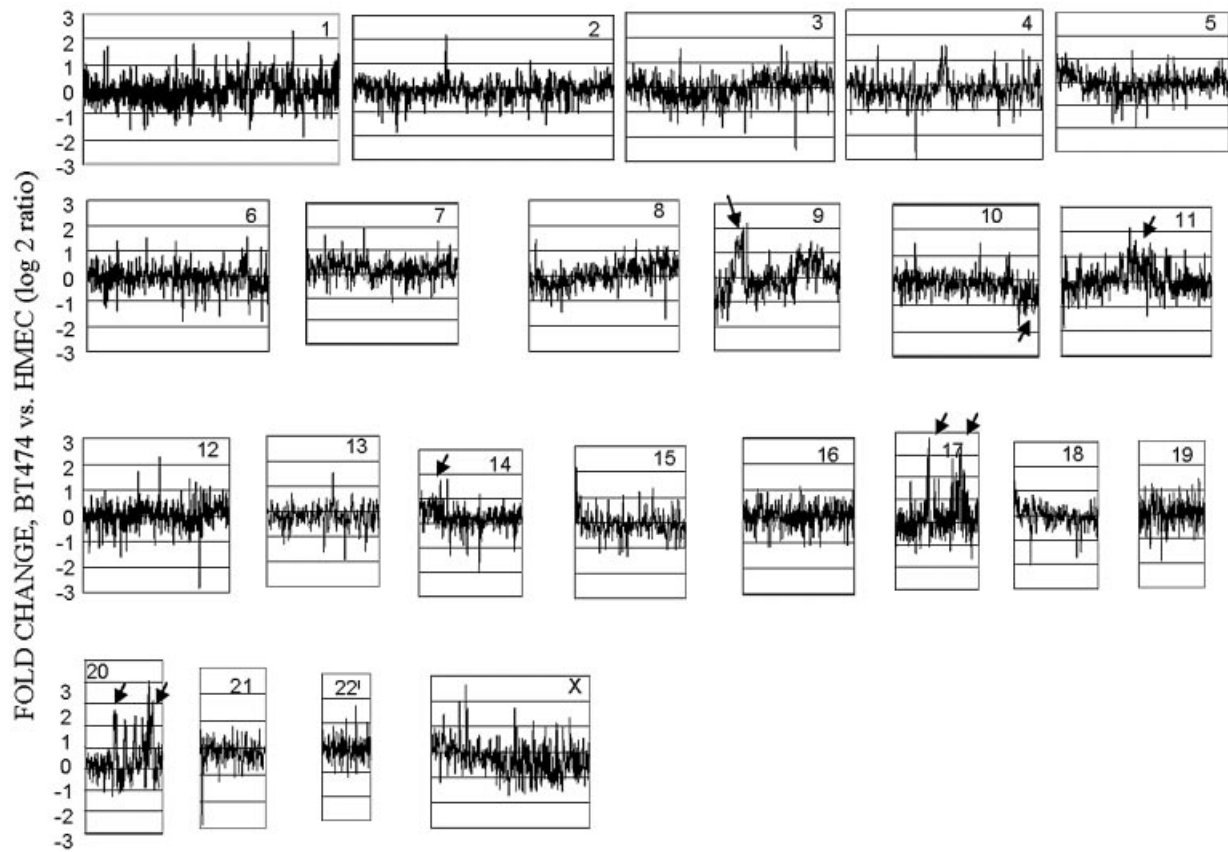
CHROMOSOME 17

CHROMOSOME 20
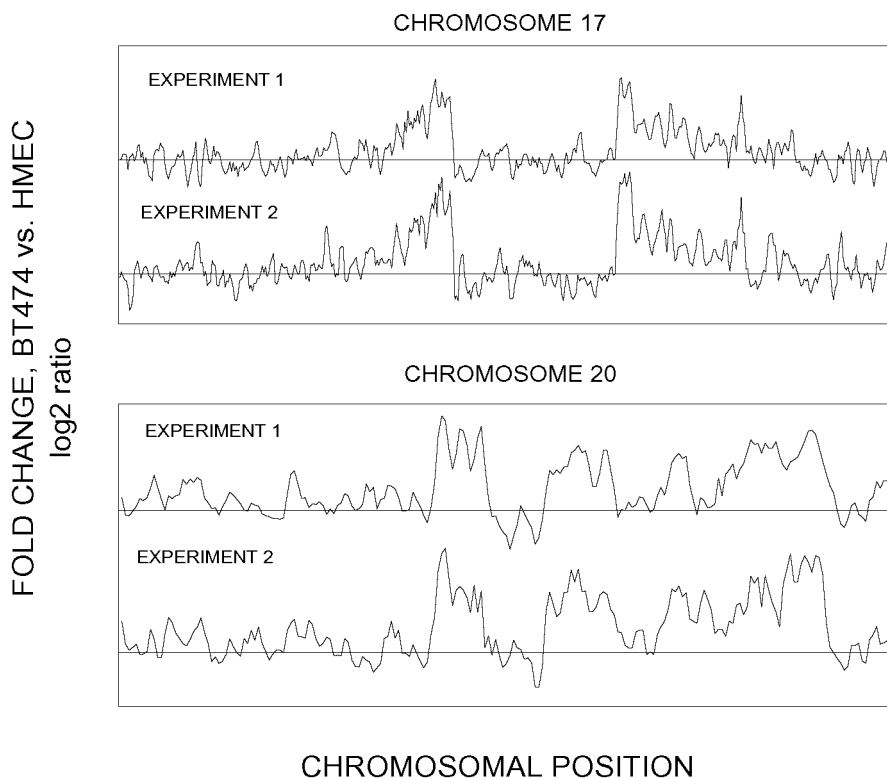
Figure 3. Reproducibility of array-CGH screening of samples amplified via balanced-PCR. In two independent experiments, genomic DNA from BT474 and HMEC cells was amplified via balanced-PCR and then screened on different human cDNA microarrays. Fold change versus chromosomal position for chromosomes 17 (481 genes) and 20 (218 genes) are depicted.

from regions that do not indicate amplification: E2F, TOP1, RAN, Tfr, HBEGF, IL9R, TBP and CYC. TaqMan assay-derived copy number ratios ('fold change' between BT474 and HMEC DNA) were then compared for amplified versus unamplified samples (Fig. 7). Genetic amplification, or lack of amplification, was correctly indicated for both, unamplified and balanced PCR-amplified DNA, for 11 of the 12 genes examined. One gene (HOX) was classified as a false negative since no amplification would have been demonstrated following a blind screen of balanced-PCR amplified samples. It is noteworthy that the array-CGH data for the HOX gene demonstrated good agreement between balanced-PCR and unamplified samples (fold change of 6.1 and 8, respectively). These data seem to suggest that the reason for the false negative in HOX may lie with the specific use of balanced-PCR amplified DNA in TaqMan assays. For example, since DNA amplified via balanced-PCR is NlaIII digested, potential NlaIII polymorphisms could affect TaqMan primer/probe binding sites in the target or the control DNA.

In a real-time PCR screen similar to that conducted for balanced-PCR, MDA amplification also indicated generally good agreement of genetic differences observed for unamplified DNA for 11 of the 12 genes examined (Fig. 7). One gene (TOP1) was classified as a false positive, since a blind screen would have demonstrated significant (6-fold) gene amplification for MDA-amplified samples, but not for unamplified or balanced-PCR amplified DNA.

Screening of DNA from formalin-fixed, paraffin-embedded tissue

DNA obtained from paraffin-embedded tissue (glioblastoma, <5 years years since formalin fixation) was either used directly (unamplified) for array-CGH or real-time PCR screening, or was first amplified via balanced-PCR or MDA and subsequently screened using HMEC DNA as the co-amplified control. DNA obtained from formalin-fixed tissue was modestly degraded (gel electrophoresis profile depicted in Fig. 8A). Following amplification via balanced-PCR the sample was screened via array-CGH and real-time PCR. The array-CGH profiling successfully revealed the main features obtained from direct screening of unamplified samples (Fig. 8B and C). In Frame B, array-CGH profiles from all 23 chromosomes are depicted and regions of amplification in chromosome 4 are indicated. In Figure 8C the chromosomal region from chromosome 4 flanking the amplified region of interest (~7 Mb long) is shown. To examine reproducibility, the experiment was conducted in duplicate and both array-CGH profiles demonstrated the same chromosome 4 feature (Fig. 8C). Similarly, when examined via Taqman real-time PCR, samples amplified via balanced-PCR demonstrated concordance with unamplified DNA for eight out of nine genes examined (Fig. 8D). In contrast, MDA universally generated low or insignificant amplification of formalin-fixed DNA and array-CGH/real-time PCR screening failed to produce substantial signals. These data indicate that, for

Figure 4. Array-CGH screening of genomic DNA from human female BT474 and HMEC cells, using unamplified DNA. Chromosomes 1–23 are depicted and arrows indicate regions of known amplifications and deletions for the BT474 cell line.

formalin-fixed samples of modest degradation, such as the one depicted in Figure 8A, balanced-PCR can be successfully used for array-CGH and real-time PCR evaluation.

DISCUSSION

The ability of balanced-PCR to overcome problems associated with amplification of modestly degraded DNA may be associated with the initial digestion of DNA followed by adaptor ligation, which generates a substantial number of DNA fragments lacking formalin-associated DNA damage, and which can then be amplified. Evidence exists that amplification performed in this manner is not substantially inhibited by formalin-induced DNA damage. Klein and colleagues described SCOMP (13,29), which utilizes DNA digestion and adaptor ligation to perform whole genome PCR amplification and comparative genomic hybridization when starting from a single cell. Because SCOMP utilizes digested, low molecular weight DNA as starting material, it was capable of efficient amplification of DNA from formalin-fixed samples and was found to be superior to DOP-PCR (29). However, the issue of amplification bias using SCOMP was not adequately addressed since the method was not validated at high resolution, i.e. via array-CGH or on a gene-by-gene basis. Due to the aforementioned PCR shortcomings, SCOMP is expected to cause substantial amplification bias. In our hands,

SCOMP produced skewed results on a gene-by-gene basis (data not shown).

Therefore, in this work we adapted balanced-PCR, which removes biases associated with PCR saturation and impurities (22), to the amplification of genomic DNA followed by array-CGH or real-time PCR quantification of gene copy number. We utilized 5 ng of genomic DNA, an equivalent to ~1000 cells, which is similar to the amount of DNA usually obtained from LCM microdissection (~5–20 ng). Upon high-resolution examination of gene copy numbers using array-CGH, balanced-PCR demonstrated an unbiased representation of the true allelic differences between the breast cancer cell line BT474 and normal mammary epithelial cells, indicating that the method can be applied for the genome-wide examination of genetic differences among cell lines or minute tumor biopsies and normal tissues. A parallel examination using real-time PCR demonstrated that the resulting gene copy differences between tumor and normal breast genomes are generally larger than array-CGH data, both for amplified and unamplified samples. This 'dynamic range compression' is commonly observed with array-CGH (21) and indicates the importance of performing TaqMan-based verification of array-detected gene-dosage changes. To further evaluate the performance of balanced-PCR we compared it with MDA. MDA is currently considered the method of choice for certain genomics applications due to the low incidence of non-specific amplification artifacts or bias among alleles and for enabling
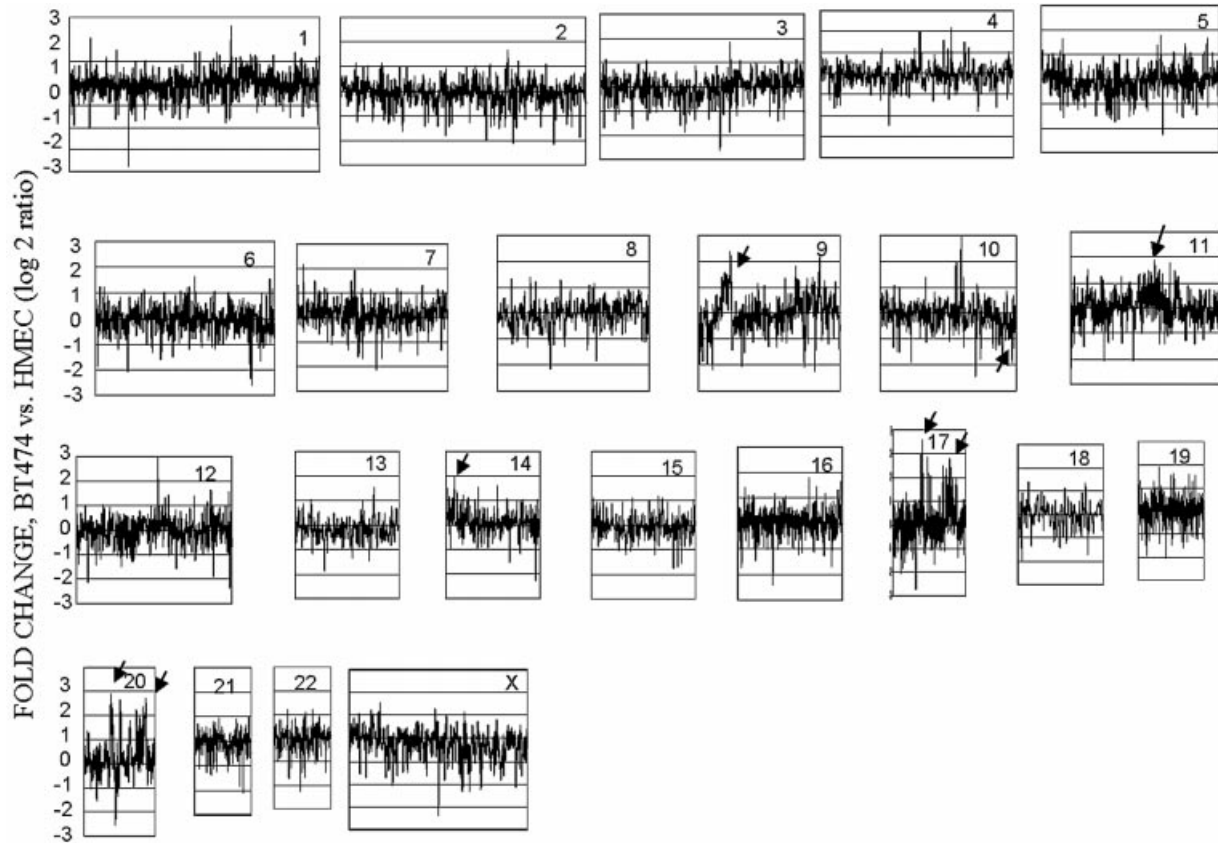
Figure 5. Array-CGH screening of genomic DNA from human female BT474 and HMEC cells, using balanced-PCR amplified DNA. Chromosomes 1–23 are depicted and arrows indicate highlighted regions of known amplifications and deletions for the BT474 cell line.
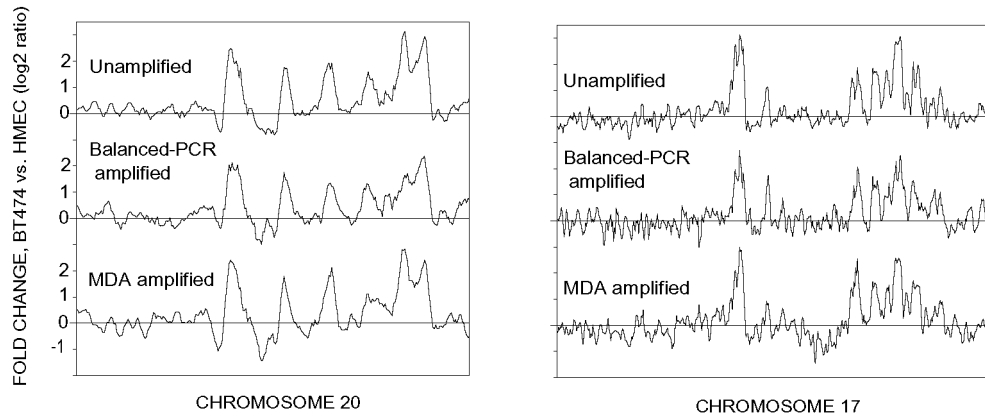

Figure 6. Array-CGH screening of chromosomes 17 and 20 from human female BT474 and HMEC cells: comparison of results using unamplified DNA (top curve), balanced-PCR amplified DNA (middle curve) and MDA amplified DNA (bottom curve).

genome-wide genotyping of small samples (30–32). In a direct comparison of balanced-PCR with MDA, when using fresh DNA samples, both methods demonstrated an approximately equivalent performance and resulted in a satisfactory amplification of previously described, tumor-related differences among the two cell lines. MDA amplification results in amplified DNA of higher molecular weight, thus it may be more appropriate for situations where a representation of most genomic regions is required, or where undigested DNA is required for subsequent analysis. Since balanced-PCR cannot

effectively amplify large (>2 kb) fragments which may potentially exist due to the location of successive NlaIII sites in a genome, the method is expected to amplify a small fraction [a 'representation' (12)] of the genome rather than the entire genome. When DNA from fresh samples is used, it may be advisable to perform both balanced-PCR and MDA amplifications whenever possible, since an agreement with regards to gene amplification and deletion by the two methods may provide higher detection accuracy. Based on our quantitation results, the gene copy number variation for 12
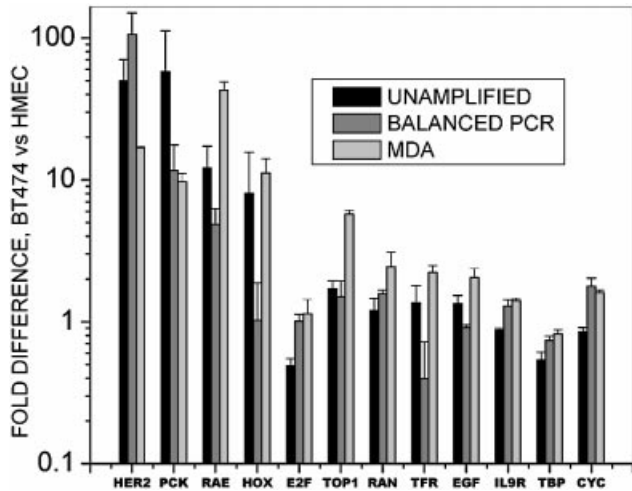
Figure 7. Real-time PCR screening (TaqMan assay) of relative gene copy numbers for breast cancer cells (BT747, 'target') versus HMEC cells ('control'). First column (black), amplification directly from unamplified genomic DNA. Second column (dark gray), amplification from balanced-PCR amplified genomic DNA. Third column (light gray), amplification from MDA amplified genomic DNA.

out of 12 genes would have been called accurately if only the consensus results were considered.

On the other hand, MDA demonstrated an almost complete failure to amplify material from formalin-fixed sample of modestly degraded DNA, which balanced-PCR was capable of amplifying. Several well preserved formalin-fixed tissue samples fall in this category and therefore may be amplified successfuly via balanced-PCR. The nucleotide 'tags' incorporated in the primers P2a and P2b during balanced-PCR can potentially be varied to include many distinct nucleotide combinations, each amplifying a different linker LN_1, LN_2, LN_3, …, LN_N. Consequently, it should be feasible to mix N genomes simultaneously and amplify them in a PCR. Thereby, large sets of archived samples could be amplified in a single, unbiased PCR amplification to provide an essentially unlimited resource of amplified materials. This resource may not only enable investigators who utilize different microarray platforms to perform inter-comparison studies, but also facilitate the establishment of tissue banks for clinicopathological studies in the future.

In summary, we have developed a balanced-PCR whole-genome amplification methodology and shown its effective-
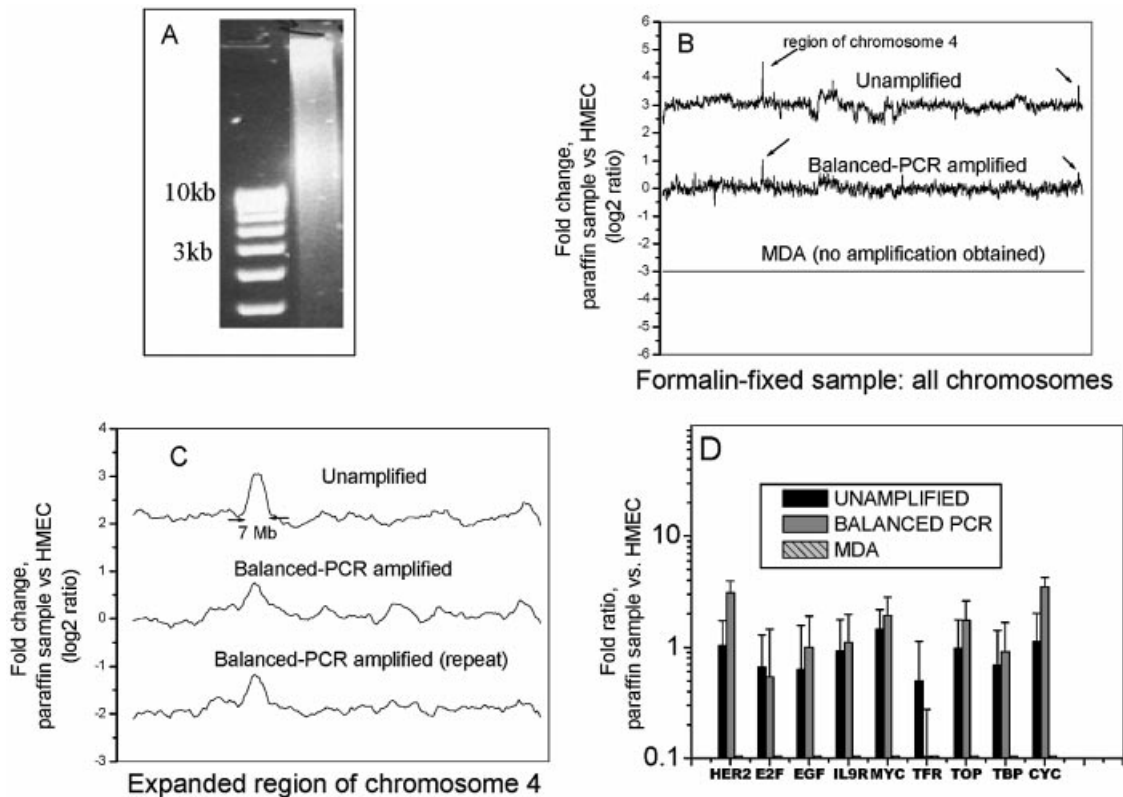
Figure 8. Screening of DNA from paraffin-embedded DNA. (**A**) Gel electrophoresis profile from a formalin-fixed, paraffin-embedded sample indicating DNA degradation. (**B**) Array-CGH screening of all 23 chromosomes using unamplified DNA (top curve), balanced-PCR-amplified DNA (middle curve) and MDA amplified DNA (bottom curve). (**C**) Chromosome 4 area of interest, indicating a 7 Mb amplification region in the unamplified and the balanced-PCR amplified sample. Duplicate experiments on two different arrays are depicted. (**D**) Evaluation of single genes using unamplified DNA, balanced-PCR amplified DNA and MDA-amplified DNA using Taqman real-time PCR.

ness in measuring gene amplifications and deletions at high resolution via array-CGH and real-time PCR. This method should allow effective amplification of DNA from archives containing modestly degraded paraffin-embedded DNA and the study of cancers whose tissue is limited, e.g. head/neck CA and pancreatic CA. Further applications in pre-implantation diagnosis, biotechnology and forensics can be envisioned.

ACKNOWLEDGEMENTS

REFERENCES

1. Lakhani,S.R. and Ashworth,A. (2001) Microarray and histopathological analysis of tumours: the future and the past? *Nature Rev. Cancer*, **1**, 151–157.
2. Burki,N.G., Caduff,R., Walt,H., Moll,C., Pejovic,T., Haller,U. and Ward,D.C. (2000) Comparative genomic hybridization of fine needle aspirates from breast carcinomas. *Int. J. Cancer*, **88**, 607–613.
3. Kallioniemi,A., Kallioniemi,O.P., Sudar,D., Rutovitz,D., Gray,J.W., Waldman,F. and Pinkel,D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, **258**, 818–821.
4. Kallioniemi,A., Kallioniemi,O.P., Piper,J., Tanner,M., Stokke,T., Chen,L., Smith,H.S., Pinkel,D., Gray,J.W. and Waldman,F.M. (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl Acad. Sci. USA*, **91**, 2156–2160.
5. Pollack,J.R., Perou,C.M., Alizadeh,A.A., Eisen,M.B., Pergamenschikov,A., Williams,C.F., Jeffrey,S.S., Botstein,D. and Brown,P.O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.*, **23**, 41–46.
6. Pinkel,D., Segraves,R., Sudar,D., Clark,S., Poole,I., Kowbel,D., Collins,C., Kuo,W.L., Chen,C., Zhai,Y. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet.*, **20**, 207–211.
7. Albertson,D.G., Ylstra,B., Segraves,R., Collins,C., Dairkee,S.H., Kowbel,D., Kuo,W.L., Gray,J.W. and Pinkel,D. (2000) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nature Genet.*, **25**, 144–146.
8. Visakorpi,T., Kallioniemi,A.H., Syvanen,A.C., Hyytinen,E.R., Karhu,R., Tammela,T., Isola,J.J. and Kallioniemi,O.P. (1995) Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res.*, **55**, 342–347.
9. Saramaki,O., Willi,N., Bratt,O., Gasser,T.C., Koivisto,P., Nupponen,N.N., Bubendorf,L. and Visakorpi,T. (2001) Amplification of EIF3S3 gene is associated with advanced stage in prostate cancer. *Am. J. Pathol.*, **159**, 2089–2094.
10. El Gedaily,A., Bubendorf,L., Willi,N., Fu,W., Richter,J., Moch,H., Mihatsch,M.J., Sauter,G. and Gasser,T.C. (2001) Discovery of new DNA amplification loci in prostate cancer by comparative genomic hybridization. *Prostate*, **46**, 184–190.
11. Heid,C., Stevens,J., Livak,K. and Williams,P. (1996) *Real Time Quantitative PCR. Genome Methods: Genome Research.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
12. Lucito,R., Nakimura,M., West,J.A., Han,Y., Chin,K., Jensen,K., McCombie,R., Gray,J.W. and Wigler,M. (1998) Genetic analysis using genomic representations. *Proc. Natl Acad. Sci. USA*, **95**, 4487–4492.
13. Klein,C.A., Schmidt-Kittler,O., Schardt,J.A., Pantel,K., Speicher,M.R. and Riethmuller,G. (1999) Comparative genomic hybridization, loss of heterozygosity and DNA sequence analysis of single cells. *Proc. Natl Acad. Sci. USA*, **96**, 4494–4499.
14. Rubin,M.A. (2002) Understanding disease cell by cell. *Science*, **296**, 1329–1330.
15. Emmert-Buck,M.R., Bonner,R.F., Smith,P.D., Chuaqui,R.F., Zhuang,Z., Goldstein,S.R., Weiss,R.A. and Liotta,L.A. (1996) Laser capture microdissection. *Science*, **274**, 998–1001.
16. Becker,I., Becker,K.F., Rohrl,M.H. and Hofler,H. (1997) Laser-assisted preparation of single cells from stained histological slides for gene analysis. *Histochem. Cell Biol.*, **108**, 447–451.
17. Assersohn,L., Gangi,L., Zhao,Y., Dowsett,M., Simon,R., Powles,T.J. and Liu,E.T. (2002) The feasibility of using fine needle aspiration from primary breast cancers for cDNA microarray analyses. *Clin. Cancer Res.*, **8**, 794–801.
18. Nelson,D.L., Ledbetter,S.A., Corbo,L., Victoria,M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter,D.H. and Caskey,C.T. (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl Acad. Sci. USA*, **86**, 6686–6690.
19. Hawkins,T.L., Detter,J.C. and Richardson,P.M. (2002) Whole genome amplification–applications and advances. *Curr. Opin. Biotechnol.*, **13**, 65–67.
20. Dean,F.B., Hosono,S., Fang,L., Wu,X., Faruqi,A.F., Bray-Ward,P., Sun,Z., Zong,Q., Du,Y., Du,J. *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl Acad. Sci. USA*, **99**, 5261–5266.
21. Lage,J.M., Leamon,J.H., Pejovic,T., Hamann,S., Lacey,M., Dillon,D., Segraves,R., Vossbrinck,B., Gonzalez,A., Pinkel,D. *et al.* (2003) Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res.*, **13**, 294–307.
22. Makrigiorgos,G.M., Chakrabarti,S., Zhang,Y., Kaur,M. and Price,B.D. (2002) A PCR-based amplification method retaining the quantitative difference between two complex genomes. *Nat. Biotechnol.*, **20**, 936–939.
23. Holland,P.M., Abramson,R.D., Watson,R. and Gelfand,D.H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5′-3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl Acad. Sci. USA*, **88**, 7276–7280.
24. Diatchenko,L., Lau,Y.F., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D. *et al.* (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA*, **93**, 6025–6030.
25. Forozan,F., Mahlamaki,E.H., Monni,O., Chen,Y., Veldman,R., Jiang,Y., Gooden,G.C., Ethier,S.P., Kallioniemi,A. and Kallioniemi,O.P. (2000) Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. *Cancer Res.*, **60**, 4519–4525.
26. Snijders,A.M., Nowak,N., Segraves,R., Blackwood,S., Brown,N., Conroy,J., Hamilton,G., Hindle,A.K., Huey,B., Kimura,K. *et al.* (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nature Genet.*, **29**, 263–264.
27. Bharathi,A., Ghosh,A., Whalen,W.A., Yoon,J.H., Pu,R., Dasso,M. and Dhar,R. (1997) The human RAE1 gene is a functional homologue of *Schizosaccharomyces pombe* rae1 gene involved in nuclear export of Poly(A)+ RNA. *Gene*, **198**, 251–258.
28. Riou,G., Mathieu,M.C., Barrois,M., Le Bihan,M.L., Ahomadegbe,J.C., Benard,J. and Le,M.G. (2001) c-erbB-2 (HER-2/neu) gene amplification is a better indicator of poor prognosis than protein over-expression in operable breast-cancer patients. *Int. J. Cancer*, **95**, 266–270.
29. Stoecklein,N.H., Erbersdobler,A., Schmidt-Kittler,O., Diebold,J., Schardt,J.A., Izbicki,J.R. and Klein,C.A. (2002) 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.*, **161**, 43–51.
30. Cheung,V.G. and Nelson,S.F. (1996) 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*, **93**, 14676–14679.
31. Zhang,L., Cui,X., Schmitt,K., Hubert,R., Navidi,W. and Arnheim,N. (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proc. Natl Acad. Sci. USA*, **89**, 5847–5851.
32. Lovmar,L., Fredriksson,M., Liljedahl,U., Sigurdsson,S. and Syvanen,A.-C. (2003) Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Res.*, **31**, e129.