

Award Number: DAMD17-03-1-0240

TITLE: Expression and Genomic Profiling of Minute Breast Cancer Samples

PRINCIPAL INVESTIGATOR: G. Mike Makrigiorgos, Ph.D.

CONTRACTING ORGANIZATION: Dana Farber Cancer Institute
Boston Massachusetts 02115

REPORT DATE: July 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 01-07-2006		2. REPORT TYPE Final		3. DATES COVERED (From - To) 1 Jul 2003 – 30 Jun 2006	
4. TITLE AND SUBTITLE Expression and Genomic Profiling of Minute Breast Cancer Samples				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER DAMD17-03-1-0240	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) G. Mike Makrigiorgos, Ph.D. E-Mail: mmakrigiorgos@yahoo.com				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Dana Farber Cancer Institute Boston Massachusetts 02115				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT To be able to apply the newest genetic analysis technologies in breast cancer research, enough DNA material must be available to perform this analysis. Often this is impossible with the minute amounts of tissue obtained via fine needle aspiration or laser capture microdissection – currently the best methods available for removal of small amounts of tissue. Therefore whole genome-based DNA amplification techniques (e.g. PCR) are essential. The aim of this proposal was to evaluate a newly developed method, balanced PCR, which overcomes the difficulty of non-linear PCR-amplification of complex genomes and faithfully retains the difference among corresponding genes or gene fragments. In the three years of work we demonstrated the application of balanced PCR in performing genomic profiling of breast CA cell lines and samples (part of Tasks 2 and 3). In the second year we demonstrated the application of the method for gene expression profiling of breast CA cell lines (Task 1). This report describes the work conducted over the three years.					
15. SUBJECT TERMS Molecular biomarkers; Genomics; whole genome amplification; Breast cancer					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	70	19b. TELEPHONE NUMBER (include area code)

Table of Contents

Introduction.....	4
Body.....	5
Key Research Accomplishments.....	12
Reportable Outcomes.....	12
Conclusions.....	13
References.....	13
Appendices.....	

1. INTRODUCTION

Genetic and gene expression profiling-based diagnosis promises to refine (1) and potentially revolutionize (2) the existing cancer staging system and the management of early disease. Microarray-based gene expression profiling and Array-based Comparative Genomic Hybridization (array-CGH) offers global views of cancer genomes and transcriptomes 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/gene expression changes in tumors requires ‘μgs’ of pure tumor DNA/cDNA (12,13). Routine tumor biopsies often consist of heterogeneous mixtures of stromal cells plus tumor cells with a wide range of genetic/gene expression profiles (14). Techniques such as Fine Needle Aspiration (FNA) 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/RNA extracted from such small number of cells has to be amplified to provide sufficient material for microarray screening. Whole genome/transcriptome 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). The aim of this proposal is to evaluate our newly developed method, *balanced* PCR, which overcomes the difficulty of non-linear PCR-amplification of complex genomes and faithfully retains the difference among corresponding genes or gene fragments.

The work conducted during the three years of research lead to the optimization of balanced PCR for (a) performing unbiased array-CGH profiling from fresh, as well as paraffin-embedded DNA and (b) performing unbiased gene expression profiling in cDNA obtained from breast cancer cells. The lowest amount of starting RNA/cDNA material for which the method is reliable was defined and a direct comparison of balanced PCR with two other methods for whole genome/transcriptome amplification was conducted. The results are summarized below.

2. BODY

The work during the three years of research focused towards realizing Tasks 1-3 in the approved Statement of Work.

METHODS.

(a) Balanced PCR on genomic DNA, followed by array-CGH:

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 per company's 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 (Oregon, USA) and are depicted in Table I. 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 Nla-III (10 units/µl stock, 37°C, 2 hours, New England Biolabs, Beverly, MA) in 1x buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, 25 µg/ml BSA). Nla-III was subsequently inactivated by incubation at 70°C for 1 hour. 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 one hour. After inactivation of ligase at 65°C for 40 minutes, 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, NJ). Thermocycling conditions were: 8 min at 72°C; 1 min at 95°C; 20 x (30s at 95°C and 60 s at 72°C); 5 min at 72°C. Following thorough DNA purification with 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 reaction was carried out using primers P2a (BT474 target genome) or P2b (HMEC control 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, NJ) with following thermocycling conditions: 1 min at 95⁰C; 10 x (30 s at 95⁰C and 60 s at 72⁰C); 5 min at 72⁰C. Alternatively, instead of BT474 DNA, the target DNA used for balanced-PCR amplification was DNA (10 µg) extracted from paraffin-embedded tissue.

Quantitation using real time (TaqMan) PCR: Real time PCR, TaqMan (33) 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, CA) in a Smart-Cycler™ (Cepheid, Sunnyvale, CA). Primers and probes for exonic regions of thirteen genes (Table I) 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 1xABI TaqMan master mix™, 4 µM each primer, and 2 µM probe. This reaction mix was split into three different 20 µl PCR reactions and thermo-cycled. The cycling program was 50⁰C 2 minutes 1 cycle, 95⁰C for 10 minutes 1 cycle, and 40 cycles at 95⁰ C for 15 seconds and 60⁰C for 1 minute. The relative genomic copy number was calculated using the comparative threshold (Ct) method (11).

Array-CGH using cDNA microarrays: Array-based comparative genomic hybridization (Array-CGH) was performed on Agilent Human 1 cDNA arrays using Nla-III 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. Further details on the experimental methods applied can be found in the accompanying paper, published in *Nucleic Acids Research* (19)-*copy Appended*.

(b) Balanced PCR on cDNA, followed by gene expression profiling on microarrays:

cDNA synthesis from total RNA: 10 µg of total RNA from the BT474 breast cancer cell line and StratRef RNA (Stratagene, La Jolla, CA), was reverse transcribed using Stratascript RT (Stratagene, La Jolla, CA) in the presence of 10 µg of random hexamer (Amersham Pharmacia) and oligo d(T)₂₄NN.

Microarrays. The 20,862 cDNAs used in these studies were from Research Genetics (Huntsville, AL). On the basis of Unigene build 166, these clones represent 19,740 independent loci. All clones corresponding to gold standard QPCR assays were sequence verified. Hybridization, washing, scanning and primary data analysis was performed as described ²²; www.microarrays.org).

Microarray Data analysis: Hierarchical clustering. Gene expression was analyzed with Cluster ²³ using the average linkage metric, and displayed using Treeview (<http://rana.lbl.gov/EisenSoftware.htm>). Genepix median of ratio values from the experiment were subjected to linear normalization in NOMAD (<http://derisilab.ucsf.edu>), log-transformed (base 2) and filtered for genes where data were present in 80% of experiments, and where the absolute value of at least one measurement was > 1.

Statistical analysis for microarrays (SAM) analysis. After linear normalization, log (base 2) transformation, and hierarchical clustering, the total RNA cluster dataset was imported into the SAM software package. One class analysis was performed to identify genes representative of StratRef and genes representative of BT474 (with 2-4 fold differences in expression). Data was censored if more than one data value was flagged in each group to eliminate poor quality array data. Delta was chosen to limit the output gene list so that less than 1% predicted false positives would be included.

Statistics: Pearson correlation coefficients comparing microarray and QPCR gene expression measurements were made in Excel (Microsoft, Redmond, WA). Global Pearson correlation coefficients for microarrays were calculated using the statistical software package R (<http://www.r-project.org/>). Further details on the experimental methods can be found in the accompanying *Manuscript In Preparation* (see Appendix).

3. RESULTS.

a. Array-CGH studies for fresh and paraffin samples, following balanced-PCR amplification of genomic DNA.

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 from 5 ng each 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 1. Concordance between the two sets of data was $R^2 = 0.51$, which increased substantially if nearest neighbor averaging was applied to the data ($R^2 = 0.71, 0.79, 0.87$ for averaging signals by 2, 5, 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 2-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 datasets is about 300 kb.

Screening of DNA from formalin-fixed, paraffin-embedded tissue, following balanced-PCR amplification.

DNA obtained from paraffin-embedded tissues 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 samples and amplified via balanced-PCR demonstrated amplification efficiency similar to that obtained from cell lines. The array-CGH profiling successfully revealed the main features obtained from direct screening of unamplified samples. A typical result obtained from FFPE samples is depicted in **Figure 2**. In **Frame A**, the DNA fragmentation associated with the formalin treatment is depicted. In **Frame B**, all 23

chromosomes are depicted and regions of amplification in chromosomes 4 and X are indicated. In **Frame C** the chromosomal region from chromosome 4 flanking the amplified region of interest (~7 Mb long) is depicted. Similarly, when examined via Taqman real time PCR, samples amplified via balanced-PCR demonstrated concordance with unamplified DNA for 8 out of 9 genes examined (**Figure 2, D**). 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.

In summary, we demonstrated a balanced-PCR procedure that allows unbiased amplification of genomic DNA from fresh or paraffin-embedded DNA samples. We demonstrated genome-wide retention of the differences among alleles following balanced-PCR amplification of DNA from breast cancer and normal human cells and genomic profiling by array-CGH (300kb 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 was used, only balanced PCR overcomes problems associated with formalin fixation and produces unbiased amplification. Balanced-PCR allowed amplification and recovery of partially degraded genomic DNA from formalin-fixed samples for subsequent retrospective analysis of human tumors with known outcomes.

b. Gene expression profiling studies.

To evaluate more accurately the value of balanced-PCR, we compared it with two established RNA amplification strategies, modified T7 linear amplification¹⁴⁻¹⁸, and Arcturus RiboAmp HS linear amplification (Arcturus, Mountain View, CA). We used a cDNA microarray platform containing 20,620 clones representing 19,700 distinct genes for hybridizations of Stratagene Universal Human Pooled Reference RNA (StratRef), a pool of 11 cells line RNAs, compared to BT474 breast cancer cell line RNA using each of the amplification methods. The results (**Figure 3**) demonstrate agreement between the data obtained from unamplified DNA, balanced-PCR and the two established methods, modified T7 and Arcturus. **Table 1** analyzes the cost associated with each method. Balanced-PCR is significantly less expensive than the other two amplification methods.

In summary, RNA amplification technologies serve translational clinical research well. Already, linear amplification has enabled examination of gene expression in clinical core needle biopsies (20,21) fine needle aspirates (21) and even single human cells (22). Our results demonstrate that balanced-PCR amplification is reproducible, and highly correlated with gold standard quantitative PCR measurements using picogram-range RNA samples. Balanced PCR displays similar accuracy as established RNA amplification methods while it is rapid, more convenient to use and of lower cost. We predict that balanced-PCR will be used widely by investigators studying fresh or fixed breast CA tissues or circulating tumor cells, and will allow answering important questions by enabling analysis of samples previously considered to be of insufficient quantity for expression array analysis.

5. KEY RESEARCH ACCOMPLISHMENTS

- a. It was verified that starting from 5-10 ng DNA obtained from breast cancer cell lines, whole genome amplification via balanced PCR allows successful screening via comparative genomic hybridization.
- b. Using 10 ng DNA extracted from breast cancer biopsies embedded in paraffin, it was demonstrated that balanced PCR is uniquely applied to perform unbiased whole genome amplification.
- c. It was verified that using total RNA obtained from breast cancer cell lines, whole transcriptome (cDNA) amplification via balanced-PCR allows successful screening via gene expression microarrays and via Taqman real time PCR assays.
- d. The minimum amount of input total RNA that is required for successful downstream analysis following balanced PCR is 500 pg.
- e. Balanced-PCR compares favorable in performance with 2 established, commercial RNA amplification methodologies (Arcturus and Modified T7) while it is more rapid, convenient and of lower cost.
- f. Ongoing work: A no-cost extension has been requested to allow us to finalize the balanced-PCR evaluation and optimization. In the next few months, the limits of the technology will be pushed even further to be able to amplify lower starting amount material and highly degraded FFPE samples (dating at least 15 years old for which the clinical outcome is already known) and perform genomic and gene expression profiling on them, which will be invaluable for identification of new cancer biomarkers.

6. LIST OF REPORTABLE OUTCOMES/BIBLIOGRAPHY

Wang G, Brennan C, Rook M, Wolfe J, Leo C, Chin L, Pan H, Liu W, Price B, and Makrigiorgos GM. Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples. *Nucleic Acids Research*, 2004; 32:e76 –Appended.

Wang G, Price B and Makrigiorgos GM. PCR-based amplification method of retaining the quantitative difference between two complex genomes. In *Cell Biology: A laboratory handbook*, 3rd Edition, Julio Celis, Editor, Elsevier Publishing, London, UK, (2005) (Review)-Appended.

Jin Li and G.M. Makrigiorgos, Whole genome amplification technologies for screening cancer biomarkers in fresh or paraffin tissue samples and in bodily fluids in breast CA. Era of Hope meeting, June 8-11 2005, Department of Defence Breast Cancer Research Meeting, Philadelphia, PA (Lecture).

Paper in preparation: A Comparison of RNA Amplification Techniques at Low-Input Concentration (Appended).

7. CONCLUSION

We have optimized the balanced-PCR whole genome amplification as well as the whole-transcriptome amplification methodology and shown its effectiveness in measuring array-comparative genomic hybridization, gene expression via microarrays and real time PCR. This method should allow effective amplification of cDNA from breast CA cell lines, fresh and paraffin-embedded tissues and the study of cancers when tissue is limited. Further applications in pre-implantation diagnosis and biotechnology can be envisioned.

REFERENCES

1. Lakhani, S.R. and Ashworth, A. (2001) Microarray and histopathological analysis of tumours: the future and the past? *Nat 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 U S A*, **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. *Nat Genet*, **23**, 41-46.
6. Pinkel, D., Se Graves, 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. *Nat Genet*, **20**, 207-211.
7. Albertson, D.G., Ylstra, B., Se Graves, 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. *Nat 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., 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 U S A*, **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 U S A*, **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 U S A*, **86**, 6686-6690.
19. Wang, G., Brennan, C., Rook, M., Wolfe, J.L., Leo, C., Chin, L., Pan, H., Liu, W.H., Price, B. and Makrigiorgos, G.M. (2004) Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples. *Nucleic Acids Res*, **32**, e76.
20. Fox, M.S., Ares, V.X., Turek, P.J., Haqq, C. and Reijo Pera, R.A. (2003) Feasibility of global gene expression analysis in testicular biopsies from infertile men. *Mol Reprod Dev*, **66**, 403-421.
21. Papadopoulou, E., Davilas, E., Sotiriou, V., Koliopoulos, A., Aggelakis, F., Dardoufas, K., Agnanti, N.J., Karydas, I. and Nasioulas, G. (2004) Cell-free DNA and RNA in plasma as a new molecular marker for prostate cancer. *Oncol Res*, **14**, 439-445.
22. Dobson, A.T., Raja, R., Abeyta, M.J., Taylor, T., Shen, S., Haqq, C. and Pera, R.A. (2004) The unique transcriptome through day 3 of human preimplantation development. *Hum Mol Genet*, **13**, 1461-1470.

LIST OF PERSONNEL SUPPORTED BY THIS GRANT

1. G. M. Makrigiorgos, Ph.D (Principal Investigator)
2. Gang Wang, Ph.D (postdoctoral fellow)
3. Jin Li, Ph.D (postdoctoral fellow)
4. Lilin Wang, MSc (research assistant)

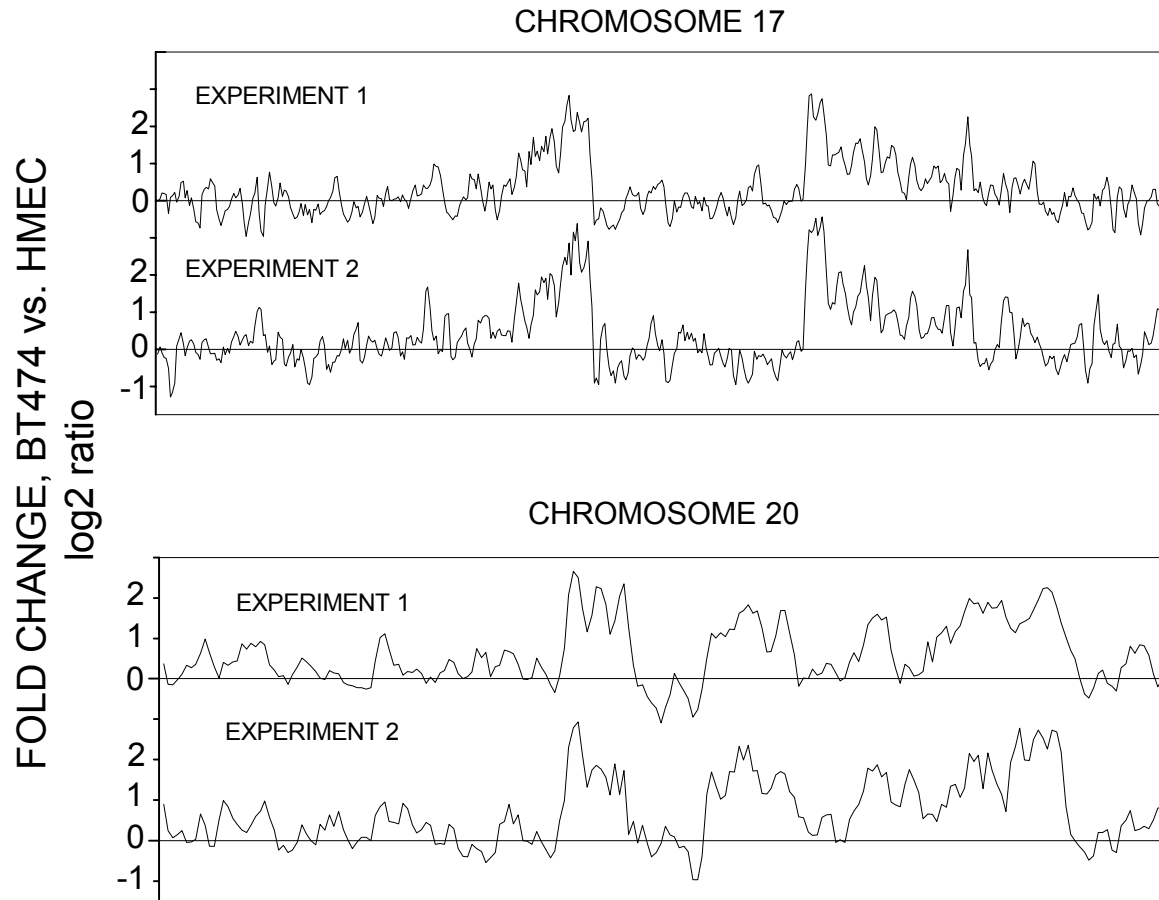


Figure 1

Figure 1. 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.

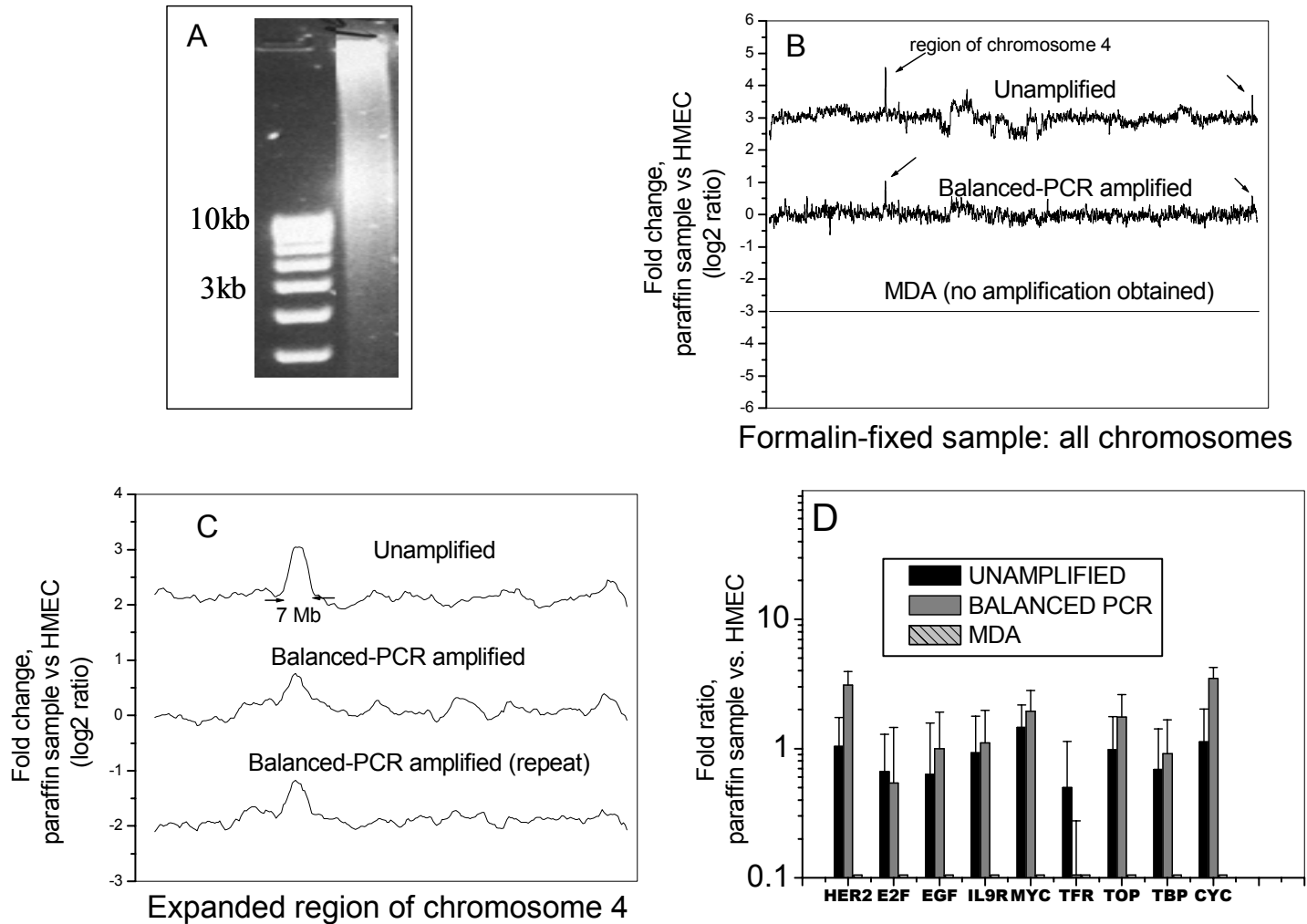


Figure 2

Figure 8. Screening of DNA from paraffin-embedded DNA. **A:** Agarose gel profiling of FFPE sample. **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. **D:** evaluation of single genes using unamplified DNA, balanced-PCR amplified DNA and MDA-amplified DNA using Taqman real time PCR.

Cluster Analysis of 17 Arrays for 2098 Genes* Comparing Amplification Techniques using BT474 vs StratRef

Legend

■ = under expressed
in sample vs
StratRef

■ = over expressed
in sample vs
StratRef

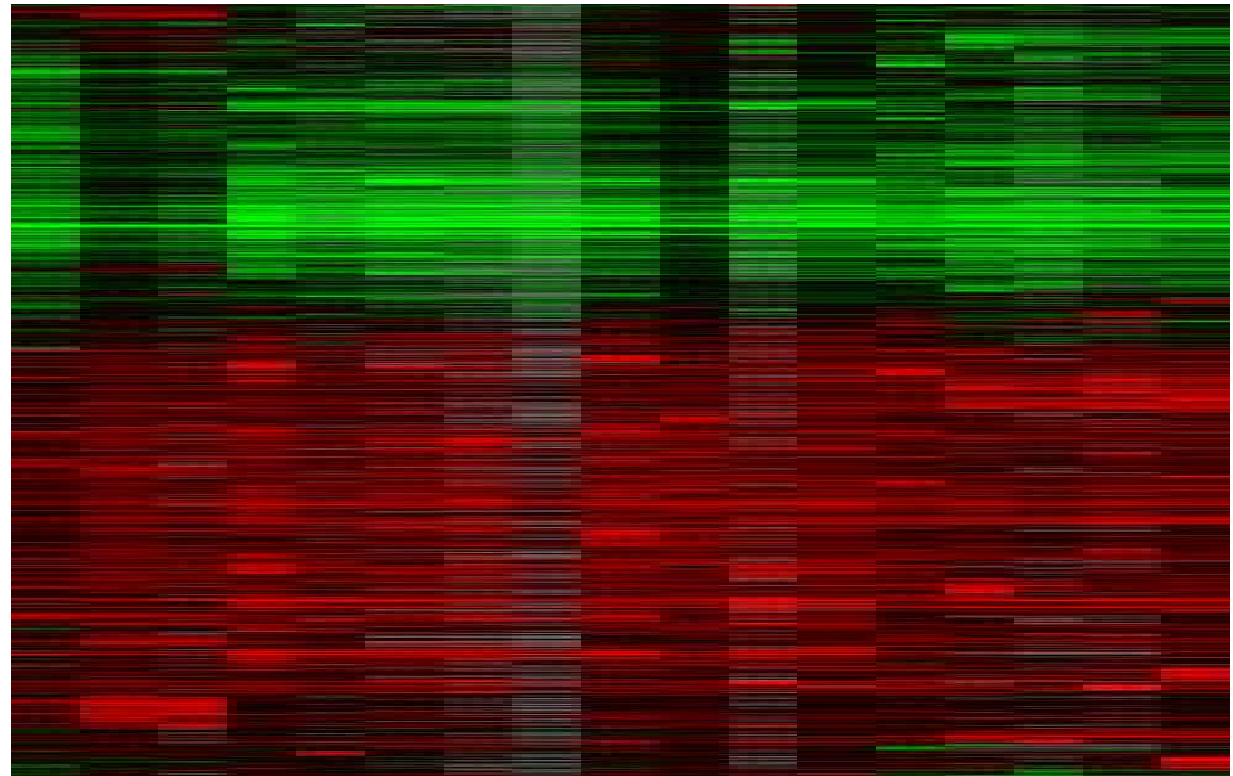
Un-amplified	Modified T7	Arcturus	Balanced PCR
--------------	-------------	----------	--------------

10 ug Un-amplified Total RNA primed with DTT7 + random hexamers

Modified T7 Amplified from 500 pg – 1ng total RNA

Arcturus Amplified RiboAmpHS from 250pg –1ng total RNA

Balanced PCR Amplified from 500pg-3.33 ng cDNA



*passed quality filters (>2fold expression, spot intensity>600)

Figure 3

Table 1 Comparison of Expenses for 3 Amplification Techniques

Balanced pcr expenses			
Reagent	Unit Price		Expense/sample
Advantagell	\$380/100x		\$1.90
Titanium	\$260/100x		\$15.60
Ligase	\$176/50ul		\$0.88
NLAIII	\$40/50ul		\$0.20
Qiagen	\$68/50x		\$2.72
Oligo+primers	\$168		\$0.50
			\$21.80
			x 2 (sample +reference)
Total cost/reaction for reagents			\$43.60
Tech time		1 day	\$131.15
Direct Labeling of balanced pcr product			
BioPrime	\$238/kit		\$13.60
Cy3dUTP	\$435		\$70.83
Cy5dUTP	\$435		\$70.83
dNTP's 10mM	\$178		\$0.01
Qiagen	\$68/50x		\$2.72
Cot	\$145/500uL		\$2.46
Quia quick	\$77		\$1.54
Tech time		2 hours	\$32.78
Total for labeling reagents			\$161.99
Total cost/reaction (amplification and labelling) reagents			\$205.59
Arcturus expenses			
			\$139.00
			x 2 (sample +reference)
RiboAmpHS kit reagents	\$695/5 samples		\$278.00
Tech time		2 days	\$262.30
RT			
N6 hexamers	\$134	makes 1ml	\$0.23
Stratascript RT	\$160	200 ul	\$2.40
aa dUTP	\$83/mg	makes 100ul	\$0.50
Microcon 30	\$214		\$2.14
Cot	\$145/500uL		\$2.46
Cy3	\$216		\$18.00
Cy5	\$216		\$18.00
Quia quick	\$77		\$1.54
Tech time		1 day	\$131.15
Total for labeling reagents			\$45.27

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

Gang Wang, Cameron Brennan¹, Martha Rook², Jia Liu Wolfe², Christopher Leo³, Lynda Chin¹, Hongjie Pan, Wei-Hua Liu, Brendan Price and G. Mike Makrigiorgos*

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

Received February 4, 2004; Revised April 13, 2004; Accepted April 21, 2004

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 7122; 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

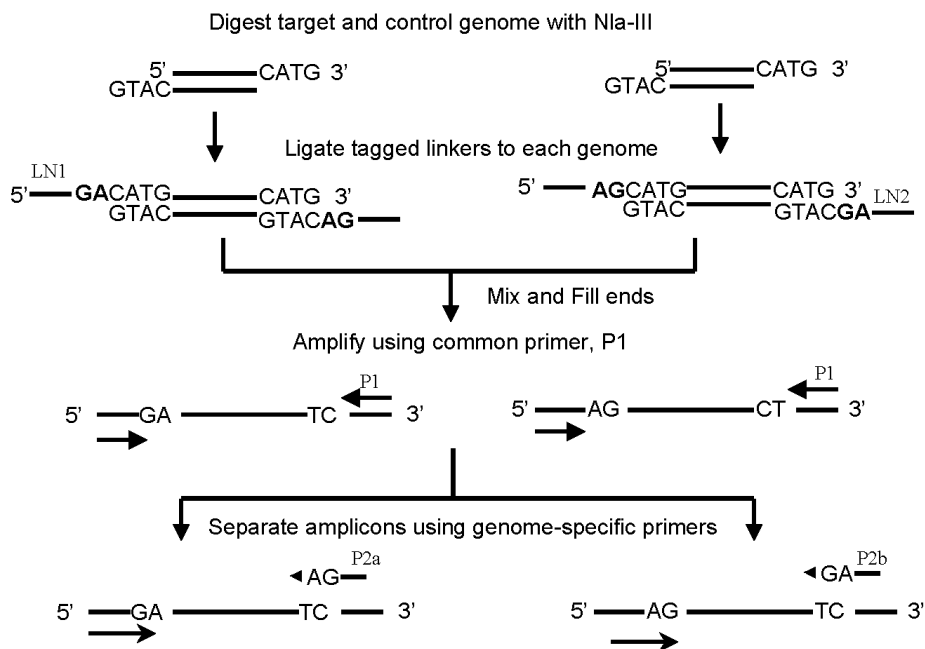


Figure 1. Protocol used for the unbiased amplification of two genomic DNAs via balanced-PCR.

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 MgCl₂, 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	CCTTGTGCTGTGTCCATCTTTC
	Reverse	CCTGGCGACAGCTTGAA
	Probe	CCTGCTGACTGGCCCGACCTACC
E2F1, 17458490	Forward	TGGCTGGGCGTGTAGGA
	Reverse	CGCTCCATTAAGCTTCAATCA
	Probe	AGAGCACTTCTGTCTTAAAGTTTTTT
TBP, 27484631	Forward	GGGCATTATTTGTGCACTGAGA
	Reverse	AGCAGCACGGTATGAGCAACTGTGAGA
	Probe	CACCGCGCAGCGTGACTGT
RAN, 34194620	Forward	TGGAGCCCAAGCGTCAGA
	Reverse	CGCTGCACCGCTGACAT
	Probe	TCTAGTTTTATAGGCAGCTGTCCTGT
TOP1, 17484369	Forward	GACAGCCCCGGATGAGAAC
	Reverse	AAGAATTGCAACAGCTCGATTG
	Probe	TCCCAGCGAAGATCCTTTCTTATAACCGTG
TFR, 29728873	Forward	GCCAATGAGGTCTGAAATGGA
	Reverse	GGCCTTATTCCTGCAATCAACA
	Probe	CTTCTGCTGGATAAAAATGAGGTTCAA
CYC, 29745697	Forward	GCCATGGAGCGCTTTGG
	Reverse	TCCACAGTCAGCAATGGTGAC
	Probe	TCCAGGAATGGCAAGACCAGCAAGA
GAPDH, 29744218	Forward	CGTCCTTGACTCCCTAGTGTC
	Reverse	CCGTAACCGCTAGTAGCC
	Probe	ATGGGAGGTGATCGGTGCTGGTT
HoxB5, 29738788	Forward	CCGAGAAGGAGTTTACAAAGT
	Reverse	CGCATACATAGCAAAACGAA
	Probe	CTTGATTTGTGGATCGTGGTCGTTA
PCK1, 17484369	Forward	CGAGAGAGAGATCCTTGCCTT
	Reverse	TTCAGATCTGCTCACGGTGT
	Probe	CAGTAGGAGCAAGAGAGGGCAAGTGTT
RAE1, 17484369	Forward	TATTTCTATGTTTGGGGTG
	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 *Nla*III 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× 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 C_Ts/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 $\Delta\Delta C_T$ (ΔC_T DNA of interest - ΔC_T 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× 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× SSC-SDS [100 ml of 20× SSC, 0.03% SDS (10%) (v/v)] at 65°C for 5 min, followed by additional 5 min wash steps in 1× SSC, then 0.2× 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

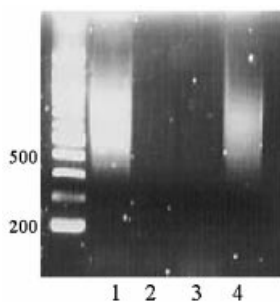


Figure 2. Evaluation of the specificity of primers P2a and P2b for amplifying target and control genomes ligated to LN1 and LN2, respectively. The protocol of Figure 1 was applied to co-amplify and, subsequently, to re-separate the two genomes. Lane 1, P2a-amplified genome; lane 4, P2b-amplified genome; 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 μ g (~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

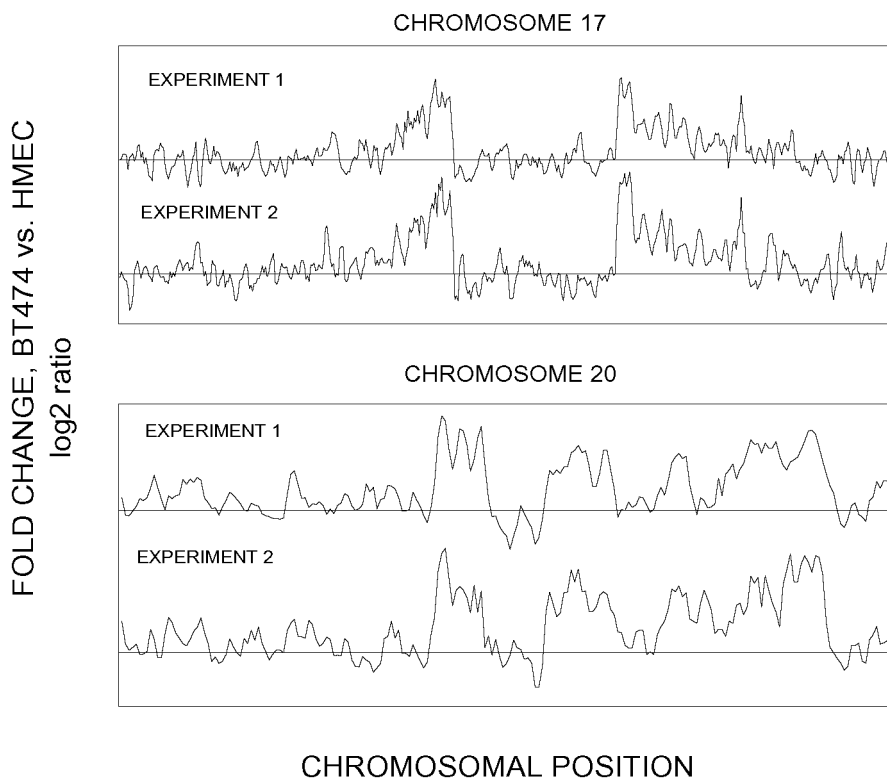


Figure 3. Reproducibility of array-CGH screening of samples amplified via balanced-PCR. In two independent experiments, genomic DNA from BT474 and HMEC cells was amplified via balanced-PCR and then screened on different human cDNA microarrays. Fold change versus chromosomal position for chromosomes 17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

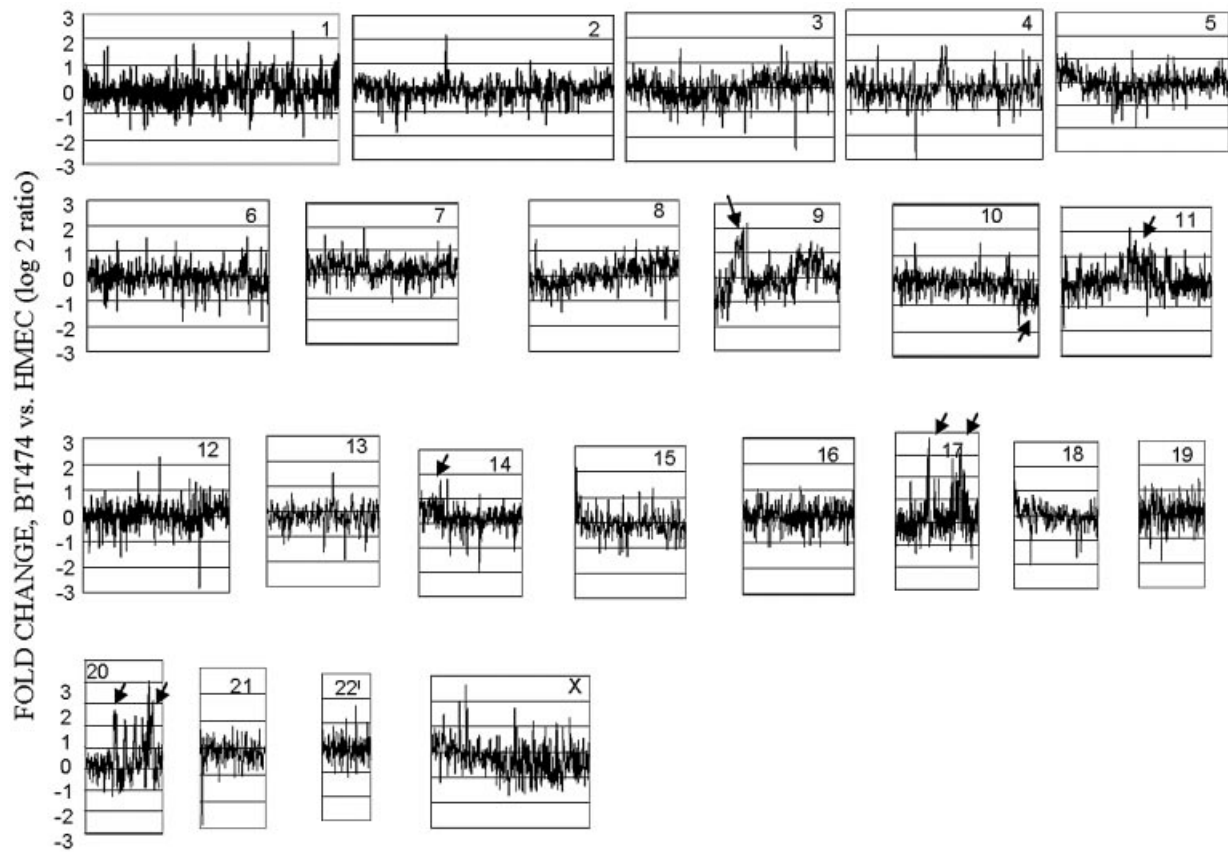


Figure 4. Array-CGH screening of genomic DNA from human female BT474 and HMEC cells, using unamplified DNA. Chromosomes 1–23 are depicted and arrows indicate regions of known amplifications and deletions for the BT474 cell line.

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

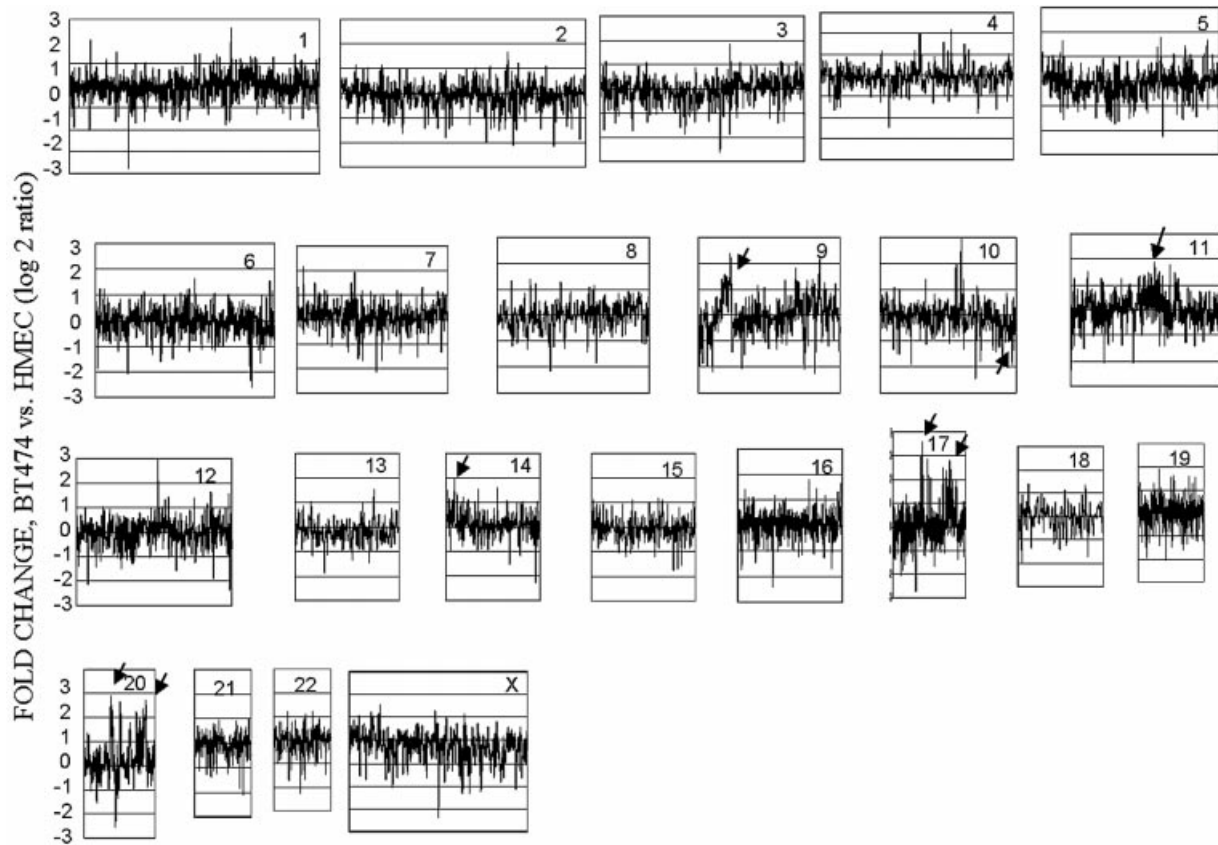


Figure 5. Array-CGH screening of genomic DNA from human female BT474 and HMEC cells, using balanced-PCR amplified DNA. Chromosomes 1–23 are depicted and arrows indicate highlighted regions of known amplifications and deletions for the BT474 cell line.

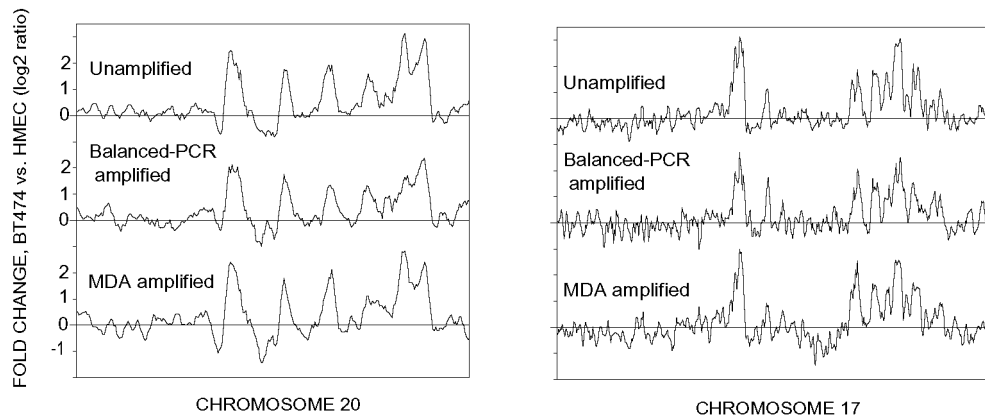


Figure 6. Array-CGH screening of chromosomes 17 and 20 from human female BT474 and HMEC cells: comparison of results using unamplified DNA (top curve), balanced-PCR amplified DNA (middle curve) and MDA amplified DNA (bottom curve).

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 *Nla*III 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

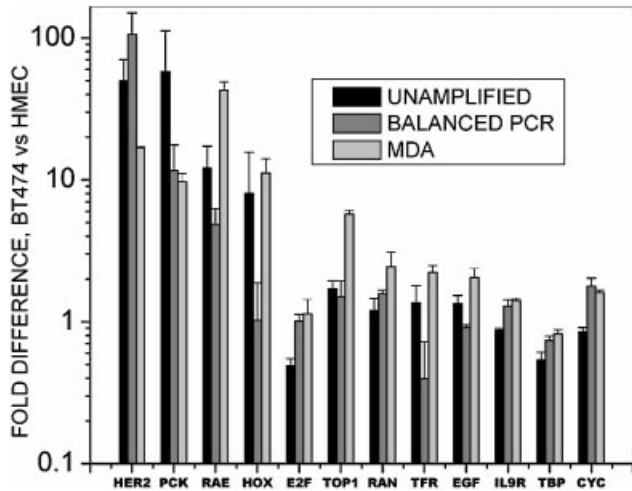


Figure 7. Real-time PCR screening (TaqMan assay) of relative gene copy numbers for breast cancer cells (BT747, 'target') versus HMEC cells ('control'). First column (black), amplification directly from unamplified genomic DNA. Second column (dark gray), amplification from balanced-PCR amplified genomic 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 successfully 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 LN₁, LN₂, LN₃, ..., 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-

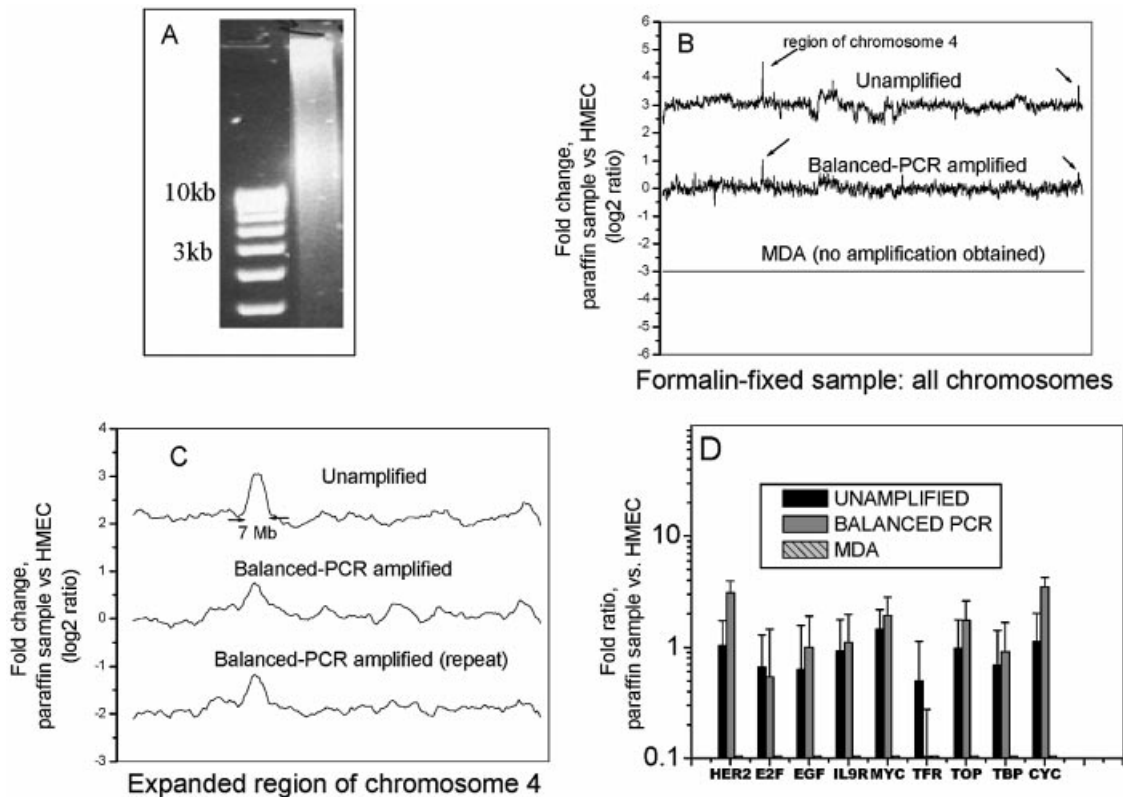


Figure 8. Screening of DNA from paraffin-embedded DNA. (A) Gel electrophoresis profile from a formalin-fixed, paraffin-embedded sample indicating DNA degradation. (B) Array-CGH screening of all 23 chromosomes using unamplified DNA (top curve), balanced-PCR-amplified DNA (middle curve) and MDA amplified 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

Funding for this work was provided in part by DOD grant BC020504, by NIH P50 CA 93683, by the Arthur and Rochelle Belfer Cancer Genomics Center, and by Variagenics, Inc. W.-H.L. was supported by training grant NIH 5 T32 CA09078.

REFERENCES

- Lakhani, S.R. and Ashworth, A. (2001) Microarray and histopathological analysis of tumours: the future and the past? *Nature Rev. Cancer*, **1**, 151–157.
- 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.
- 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.
- 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.
- 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.
- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, J., 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.
- Albertson, D.G., Ylstra, B., Seagraves, 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Rubin, M.A. (2002) Understanding disease cell by cell. *Science*, **296**, 1329–1330.
- 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.
- Becker, I., Becker, K.F., Rohrl, M.H. and Hofer, H. (1997) Laser-assisted preparation of single cells from stained histological slides for gene analysis. *Histochem. Cell Biol.*, **108**, 447–451.
- 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.
- 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.
- Hawkins, T.L., Detter, J.C. and Richardson, P.M. (2002) Whole genome amplification—applications and advances. *Curr. Opin. Biotechnol.*, **13**, 65–67.
- 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.
- Lage, J.M., Leamon, J.H., Pejovic, T., Hamann, S., Lacey, M., Dillon, D., Seagraves, 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.
- 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.
- 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.
- 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.
- 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.
- Snijders, A.M., Nowak, N., Seagraves, 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.
- 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.
- 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.
- 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 degenerate oligonucleotide primed-polymerase chain reaction for global amplification of minute amounts of DNA from microdissected archival tissue samples. *Am. J. Pathol.*, **161**, 43–51.
- 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.
- 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.
- 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.

IN PREPARATION FOR SUBMISSION

A Comparison of RNA Amplification Techniques at Low-Input Concentration

Julie E. Lang^{*}, Janet H. Scott[§], David G. Ginzinger[^], Gang Wang[¶], G. Mike Makrigiorgos[¶], Daisuke F. Nonaka[§], Julia Crothers[§], Scot Federman[§], Daniel Khodabakhsh[§], Laura J. Esserman^{*}, John Park[§], and Christopher M. Haqq^{§†}

^{*}Department of Surgery, UCSF Comprehensive Cancer Center, San Francisco, CA;

[§]Department of Medicine and Prostate Cancer Program, UCSF Comprehensive Cancer

Center, San Francisco, CA 94115; [¶]Dana Farber Cancer Institute, Harvard Medical

School, Boston, MA 02115. [^]Genome Analysis Core Laboratory, UCSF Comprehensive

Cancer Center, San Francisco, CA 94115.

[†]Reprint requests and correspondence should be addressed to C.H.

(chaqq@itsa.ucsf.edu). Address: 1600 Divisadero Street, Third Floor, San Francisco, CA

94115; Phone: (415) 514-2559; FAX: (415) 502-3179.

Key Words: T7 linear amplification, balanced polymerase chain reaction, quantitative real time polymerase chain reaction, cDNA microarray

Abstract

Characterization of gene expression from rare clinical specimens requires high-fidelity amplification techniques at the picogram level. Although there are many different amplification techniques utilized by different investigators no group has compared fidelity of these methods on a single microarray platform.

Aliquots of commercial reference and BT474 cell line RNA were independently amplified using two linear methods: 1) modified T7, 2) Arcturus RiboAmp HS and a logarithmic method 3) Balanced PCR. Spotted 20,621 cDNA microarrays were hybridized for each of the probe pairs. Data from each amplification method was compared to the gold standard of quantitative real-time PCR (QPCR) for 37 genes and Pearson correlations were calculated. Replicate amplifications were R^2 0.75 for modified T7 linear amplification, R^2 0.86 for Arcturus HS linear amplification and R^2 0.87 for Balanced PCR logarithmic amplification. The false expression rate (FER), defined as an inverse microarray expression ratio measurement compared to expression ratios from QPCR of total RNA, was measured. The mean FER for all methods were similar: modified T7 14.6% (5.4/37), Arcturus HS 13.5%% (5/37), and Balanced PCR 11.3% (4.2/37). On comparison of QPCR of amplified to QPCR of total RNA, Arcturus yielded an R^2 of 0.86 (FER 0/21 for 0%), modified T7 R^2 0.87 (FER 1/22 for 4.5%), and Balanced PCR R^2 0.75 (FER 3/19 for 15%). These results demonstrate feasibility of expression analysis starting with picogram level input of total RNA samples. Selection of an optimal method for each laboratory will require balancing local labor versus reagent costs.

Introduction

Quantitative analysis of circulating tumor cells (CTC's/micrometastases) has demonstrated prognostic significance equivalent to lymph node status in breast cancer¹. The relationship of CTC's to the cells that actually comprise solid organ metastases has been a subject of controversy and speculation, as no robust methodology to perform unbiased expression profiling of these rare cells has previously been available. Similarly, many other types of clinical specimens (for example, fine needle aspirates, cells isolated by laser capture microdissection, and other rare cell populations) have a limited quantity of RNA available for analysis. A major obstacle to the expression profiling of these rare specimens with microarrays is that only picograms to nanograms of RNA are available, while microarray assays require at least 2 μ g. The average amount of total RNA in a human epithelial cell is estimated to be between 10 and 40 pg² (1-6 pg of mRNA per single cell^{3,4}).

Our group developed a flow cytometry (FACS) strategy for the isolation of circulating tumor cells from the blood of patients with cancers of epithelial origin, such as breast and prostate cancer⁵. Typically, 10 mL peripheral blood samples from patients with advanced breast and prostate cancer yield no more than 10-300 cells. Therefore, RNA amplification techniques for CTC analysis must be able to faithfully represent the transcriptome starting with just 100 picograms to 3 nanograms of total RNA, assuming the total RNA content of CTC's to be 10pg/cell. As these cells are extremely rare but of great clinical significance, we determined which amplification technique is most faithfully able to measure gene expression when starting with picogram quantities of total RNA.

We compared three popular RNA amplification strategies, modified T7 linear amplification⁶⁻¹⁰, Arcturus High Sensitivity linear amplification (Arcturus, Mountain View, CA), and Balanced PCR, a recently developed exponential amplification method¹¹. We used a cDNA microarray platform containing 20,620 clones representing 19,700 distinct genes for hybridizations of Stratagene Universal Human Pooled Reference RNA (Stratref), a pool of 11 cells line RNAs, compared to BT474 breast cell line RNA using each of the amplification methods.

For a gold standard measurement to compare to array results, we utilized quantitative Taqman RT-PCR (QPCR) for a panel of 37 genes. Unlike prior studies where QPCR was used to validate expression of outliers - genes predominantly expressed in one RNA sample versus another - we selected QPCR primers to measure genes that are under-expressed, equivalent, or over-expressed in BT474 relative to StratRef total (un-amplified) RNA. Thus, our experiments were designed to determine whether fidelity of amplification was compromised without regard to the amplitude of the ratio of gene expression between two RNA samples.

Results showed that each method had a different lower limit of input RNA quantity, below which success in amplification was unreliable. However, when using sufficient input total RNA, each method resulted in array expression ratio measurements that were well correlated with QPCR results. Few differences in fidelity were seen comparing the techniques. Thus, we also assessed the time required to perform each technique and the reagent costs to compare the cost-effectiveness of each assay. Our analysis provides a useful guide to laboratories faced with the challenge of analyzing multiple samples containing picogram total RNA quantities.

Methods

Total RNA labeling without amplification: 10 micrograms of total RNA from the BT474 breast cancer cell line and StratRef RNA (Stratagene, La Jolla, CA), was reverse transcribed using Stratascript RT (Stratagene, La Jolla, CA) in the presence of 10 micrograms of random hexamer (Amersham Pharmacia) and oligod(T)₂₄NN.

Modified T7 RNA amplification: Total RNA from the BT474 breast cancer cell line and StratRef (Stratagene, La Jolla, CA) was linearly amplified through two rounds of modified *in vitro* transcription⁷.

Arcturus HS amplification: Total RNA from the BT474 breast cancer cell line and StratRef (Stratagene, La Jolla, CA), was linearly amplified through two rounds of *in vitro* transcription according to the manufacturer's instructions (Arcturus, Mountain View, CA).

Balanced PCR amplification: Total RNA from the BT474 breast cancer cell line and StratRef (Stratagene, La Jolla, CA), was reverse transcribed using separate oligo dTT7 primers, pooled and exponentially amplified in the same PCR tube^{11,12}. Although the balanced PCR reactions were carried out in a different laboratory than the linear amplifications, an aliquot of RNA from the same tube of StratRef RNA and an aliquot of the same preparation of BT474 RNA as was used to minimize input variability. Reverse transcription of this RNA was performed prior to shipment of the cDNA on dry ice for subsequent Balanced PCR.

Assessment of Transcript Integrity: The molecular weight profile and integrity of each amplified RNA/DNA species was evaluated using the Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA). All RNA was verified to be intact with well resolved 18 and 28S peaks and no evidence of RNase contamination prior to beginning all experiments. The size of the amplified products ranged from 100-4400 bases.

Fluorescent labeling: Amplified RNAs (aRNAs) produced with Modified T7 and Arcturus RiboAmp HS were converted to amino-allyl modified cDNA and coupled to N-hydroxysuccinimidyl esters of Cy3 or Cy5 (Amersham, Piscataway, NJ).¹³ The Balanced PCR amplified cDNA's were labeled with Klenow from BioPrime (Invitrogen, Carlsbad, CA) and Cy3/Cy5 dUTP (Amersham, Piscataway, NJ). All specimens were then hybridized to a microarray slide at 65°C for 12-16 hours. The slide was then washed and immediately scanned with Axon Imager 4000b (Axon Instruments, Union City, CA), utilizing GenePixPro 3.0 software.

Microarrays. The 20,862 cDNAs used in these studies were from Research Genetics (Huntsville, AL). On the basis of Unigene build 166, these clones represent 19,740 independent loci. All clones corresponding to gold standard QPCR assays were sequenced to verify their identity¹⁴. Hybridization, washing, scanning and primary data analysis was performed as described¹⁵; www.microarrays.org).

Microarray Data analysis: Hierarchical clustering. Gene expression was analyzed with Cluster¹⁶ using the average linkage metric, and displayed using Treeview (<http://rana.lbl.gov/EisenSoftware.htm>). Genepix median of ratio values from the experiment were subjected to linear normalization in NOMAD (<http://derisilab.ucsf.edu>),

log-transformed (base 2) and filtered for genes where data were present in 80% of experiments, and where the absolute value of at least one measurement was > 1 .

Statistical analysis for microarrays (SAM) analysis. After linear normalization, log (base 2) transformation, and hierarchical clustering, the total RNA arrays' cluster data table was imported into the SAM software package. One class analysis was performed to identify genes representative of StratRef and genes representative of BT474 (with 2-4 fold differences in expression). Data was censored if more than one data value was flagged in each group to eliminate poor quality array data. Delta was chosen to limit the output gene list so that less than 1% predicted false positives would be included.

Quantitative RT-PCR: cDNA was made from total RNA for both BT474 and StratRef, in 100- μ L reactions using M-MLV reverse transcriptase and random hexamers incubated at 25°C for 10 min then 48°C for 30 min. Expression of each gene was analyzed using the 5' nuclease assay (real-time TaqMan RT-PCR; ¹⁷) with the ABI PRISM 7700 instrument (Applied Biosystems (ABI), Foster City, CA). Probe sequences and cycle conditions are available upon request. Relative expression levels were calculated compared to beta-glucuronidase as detailed previously¹⁸. Six of the 38 (16%) genes for which we performed QPCR failed repeated attempts at sequence verification from the original E. coli library microarray source plate. However, these six genes only contributed 2.5% (2/79) of all the FER's for all microarrays and they are therefore included in this analysis.

Statistics: Pearson correlation coefficients comparing microarray and QPCR gene expression measurements were made in Excel (Microsoft, Redmond, WA).

Cost Analysis: UCSF institutional prices for each reagent used in each amplification technique were determined, and the fractional price per amplification reaction was

determined. For time analysis, only the time actually spent in the laboratory by the technician performing the assay (i.e. not the time needed for incubation of the PCR reactions or in vitro transcription reactions, which were typically run overnight) was used. The labor costs for a UCSF entry level technician including 10% fringe benefits were used as the basis for calculation.

Results

Determination of Amplification Linearity

Replicates of each amplification method and replicates of control total RNAs without amplification were assayed. It is important that replicate experiments provide a high overall correlation in ratios measured for each microarray target before amplification strategies are accepted as trustworthy methods. Replicate two-round amplifications were well correlated for our 38 gene panel in log Cy3 (StratRef):Cy5(BT474) ratios: R^2 0.75 for Modified T7 linear amplification, R^2 0.86 for Arcturus HS linear amplification and R^2 0.87 for Balanced PCR logarithmic amplification.

Determination of lowest input RNA concentrations for reproducible RNA amplification

Serial dilutions of the same tube of StratRef and BT474 RNA served as the substrate for all amplification reactions to minimize sources of variability. The lower limits of total RNA required for each method were defined as the lowest RNA input amount where amplification reactions consistently yielded sufficient product (10 micrograms) to permit analysis on cDNA microarrays. These were 500 pg for modified T7, 250 pg for Arcturus RiboAmp HS, and 500 pg for balanced PCR (Table 1).

Determination of false expression measurements occurring with each amplification method

When dealing with clinical samples, microarray results are often validated with QPCR, therefore techniques that demonstrate a low FER by this type of analysis are very desirable. We define a false expression result (FER) as measurement of an inverse ratio by microarray compared to gold standard QPCR for the same gene assayed using unamplified total RNA. Both array and QPCR measurements were normalized to levels of β -glucuronidase to facilitate comparison. Table 1 lists performance of each amplification method at differing input RNA concentrations with number false and percentage FER in comparison to QPCR of total RNA for 37 genes. Balanced PCR showed a mean percent FER of 11.3%, Arcturus RiboAmp HS showed 13.5%, and modified T7 showed 14.6%. It was interesting to observe that the FER rates for each method were independent of input RNA level by ANOVA ($p=0.39$).

Fig. 1 shows the FERs among the 37 QPCR assays (FER genes are boxed) for the modified T7 method starting with 1 nanogram of total RNA. FERs were calculated in the identical fashion for each of the amplifications (not shown). **Figure 2** shows the overall Pearson correlation for each of the three methods compared with QPCR of amplified and total RNA. Table 2 presents a comparison of QPCR of amplified to QPCR of total RNA for each method. By this analysis Arcturus had a 0% FER, modified T7 had a 4.5% FER, and Balanced PCR had a 15%FER. These methods of analysis give a platform independent measure, the FER, useful in comparing the amplification methods.

Incidence of FER is not correlated with presence of repeat elements in the microarray platform cDNA clones

We noted that FER was common to all amplification techniques for three sequence verified clones, DFF (Unigene ID AA487452), ELK1 (AA844141) and GRP(AA026118). Since this suggests that a significant contribution to false expression measurements was the microarray platform itself, we used bioinformatics tools to examine characteristics of the clones that contributed repeated FER results across methods. A bioinformatics query for repetitive elements in the sequence of clones that contributed FER's to this analysis found that only 41.1% (7/17) of the FER clones had regions of sequence repeats, suggesting that nonspecific hybridization to repetitive elements did not exclusively explain FER.

Hierarchical clustering analysis of Stratref and BT474 samples amplified by different methods

Fig. 3 presents gene expression for Stratref and BT474 after hierarchical clustering, as visualized by Treeview¹⁶. As expected, nodes highlighting such breast cancer specific genes as V-Erb-b2 (Her2/neu) were consistently detected in all amplifications of BT474 RNA. It is gratifying that all methodologies yielded globally similar profiles of gene expression. Each of the three amplification techniques yielded fairly consistent expression results within the constraints of each technique's input threshold of total RNA. Therefore, to further aid in selection of a standard methodology for RNA amplification, we performed a cost analysis.

Cost per Amplification

The cost per amplification and labeling reaction for each method were calculated as cost for reagents per individual sample. Technician labor costs were estimated based on an annual salary of \$31,000 plus 10% fringe benefits. Labor costs were determined based on the number or fraction of days of actual work time (not incubations) based on a 5 day work week. The full data required for calculation are provided as supplementary Table 1, and the final costs by method are presented in Table 3. Technician costs would increase with increasing numbers of samples beyond some reasonable threshold that an individual researcher could efficiently amplify in a given workday. Generally, technician costs would be comparable for amplification of 1-10 samples/day in our experience.

Discussion

RNA amplification technologies serve translational clinical research well. Already, linear amplification has enabled examination of gene expression in clinical core needle biopsies^{8,19} fine needle aspirates¹⁹ and even single human cells²⁰. Our results demonstrate that amplification technology is reproducible, and highly correlated with gold standard QPCR measurements using such picogram range RNA samples. We predict that these methods will be used by investigators studying circulating tumor cells, and will allow answering important questions by enabling analysis of samples previously considered to be of insufficient quantity for expression array analysis.

Many groups rely on these amplification techniques to provide data on gene expression, yet to our knowledge this is the first report comparing the fidelity of 3 amplification

methods at low-input range on a single microarray platform. While each method was able to provide data in the picogram range, certain methods are advantageous over others in terms of lower limit of RNA that can reliably be amplified, cost per reaction, and number of days required for processing of samples.

Below 1ng the modified T7 method could not reproducibly amplify such that insufficient RNA was typically generated for even a single microarray hybridization. While we were successful in hybridizing 3 arrays with this method at 500pg we do not recommend this method below 1ng of input total RNA as several technicians quite experienced with this method could not repeat these results. This technique may be optimized by HPLC purification of the oligodT₍₂₄₎T7 primer. One drawback of this technique is the greater length of time involved (3days) compared to other amplification reactions (2days) and the relative complexity of the protocol.

Arcturus RiboAmp HS was able to provide expression array data at a lower input concentration than any of the other tested methods, and we were able to use smaller amounts than the manufacturer's recommended minimum sample input of 500 pg total RNA. Below 250pg, even this method typically fails to amplify. This likely represents a theoretical limit of 10-25 cells total RNA content (for laser capture microdissection more would be required because of fractionation of cellular material), unless specialized tissues such as oocytes are examined. It is somewhat concerning, however, that the %FER was observed to increase from 10.8% at 500pg to 19% at 250pg in our study.

Balanced PCR is a promising technique for the amplification of low-input quantities of RNA. It maintains a high degree of accuracy with an input as low as 667pg of RNA (FER 10.8-13.5%). While theoretical concern exists regarding the accuracy of logarithmic amplification methods, this method overcomes the potential problem by stopping the PCR reaction before the logarithmic phase of the PCR curve. This method had the lowest cost per reaction and also required the least amount of technician time compared to the other methods. In addition, it has been recently demonstrated that the same balanced-PCR protocol used for cDNA amplification may also be used for the unbiased amplification of whole genomic DNA followed by array-CGH analysis¹². However, several iterations were required for new technicians to learn to successfully perform balanced PCR.

Each lab will have to weigh their decision on which amplification technique is most suitable based on factors including amount of starting input total RNA, cost per reaction, technician time, and experience/comfort level with the techniques. Labs that routinely work with samples in excess of 1ng starting material should focus on cost-savings as each of the methods tested proved to be reliable above this threshold. It is likely that balanced PCR could be further optimized to include amino-allyl-dUTP incorporation in the PCR reaction. This would facilitate indirect Cy dye labeling, which would dramatically reduce the labeling cost for this method.

It is important to ascertain the linearity of a chosen method at the low input range before going on to work with precious clinical specimens. Each of the 3 tested methods performed surprisingly accurately when amplifying from low inputs of total RNA based on microarray analysis validated with QPCR of 37 genes. We have demonstrated that it is feasible to reliably and accurately perform expression profiling from picogram quantities of total RNA. These methods will likely enable exciting new directions for molecular analysis of samples previously considered to be of insufficient quantity of total RNA for expression profiling.

Acknowledgments

C.H. was supported by National Cancer Institute Grants P50 CA89520, R01 CA101042-01 and a gift from Bank of the West. J. P was supported by a U54 grant. G.M.M and G. Wang were supported by DOD grant BC020504.

Figure Legends

Fig. 1. A Comparison of Microarray Expression Ratios for 37 Genes of Modified T7 Amplified RNA to Taqman of Un-amplified RNA

Fig. 2. Correlation in Expression Ratios Between Taqman of Amplified and Total RNA for 3 Different Methods

Fig. 3 Cluster Analysis of 17 Arrays for 2098 Genes Comparing Amplification Techniques Using BT474 Versus StratRef

Table Legend

Table 1 Performance of Amplification Method by Evaluation of Array Expression Ratios as Compared to Taqman of 37 Genes

Table 2 Taqman of Amplified RNA: Accuracy of 3 Methods

Table 3 Expenses for 3 Amplification Techniques

Table 3A Final Costs by Method

References

1. Braun S, V. F., Schlimok G, Diel IJ, Janni W, Gerber B, Gebauer G, Coombes RC, Pierga J-Y, Naume B, Pantel K, Collaborative Group Bone Marrow Micrometastasis Dept. Ob/Gyn, Innsbruck, Austria; Bozen, Italy; Dept. Hem/Oncol, Augsburg, Germany; Dept. Ob/Gyn, Heidelberg, Germany; Dept. Ob/Gyn, Munich, Germany; Dept. Ob/Gyn, Rostock, Germany; Dept. Ob/Gyn, Nuernberg-Erlangen, Germany; Div. Med., London, United Kingdom; Dept. Hem/Oncol, Paris, France; Dept. Hem/Oncol, Oslo, Norway; Inst. Tumorbiology, Hamburg, Germany. in *San Antonio Breast Cancer Symposium* (2003).
2. Brady, G. Expression profiling of single mammalian cells--small is beautiful. *Yeast* **17**, 211-7 (2000).
3. Dixon, A. K., Richardson, P. J., Pinnock, R. D. & Lee, K. Gene-expression analysis at the single-cell level. *Trends Pharmacol Sci* **21**, 65-70 (2000).
4. Klein, C. A. et al. Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol* **20**, 387-92 (2002).
5. J. Lang, J. S., D. Nonaka, D. Khodabakhsh, J. Crothers, D. Ginzinger, L. Esserman, C. Haqq, J. Park. in *Society of Surgical Oncology* (New York, New York, 2004).
6. Van Gelder, R. N. et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A* **87**, 1663-7 (1990).
7. Baugh, L. R., Hill, A. A., Brown, E. L. & Hunter, C. P. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* **29**, E29 (2001).
8. Fox, M. S., Ares, V. X., Turek, P. J., Haqq, C. & Reijo Pera, R. A. Feasibility of global gene expression analysis in testicular biopsies from infertile men. *Mol Reprod Dev* **66**, 403-21 (2003).
9. Ryder, M. I., Hyun, W., Loomer, P. & Haqq, C. Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. *Oral Microbiol Immunol* **19**, 39-49 (2004).
10. Rogatsky, I. et al. Target-specific utilization of transcriptional regulatory surfaces by the glucocorticoid receptor. *Proc Natl Acad Sci U S A* **100**, 13845-50 (2003).
11. Makrigiorgos, G. M., Chakrabarti, S., Zhang, Y., Kaur, M. & Price, B. D. A PCR-based amplification method retaining the quantitative difference between two complex genomes. *Nat Biotechnol* **20**, 936-9 (2002).
12. Wang, G. et al. Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples. *Nucleic Acids Res* **32**, e76 (2004).
13. Hughes, T. R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**, 342-7 (2001).
14. Taylor, E. et al. Sequence verification as quality-control step for production of cDNA microarrays. *Biotechniques* **31**, 62-5 (2001).
15. DeRisi, J. et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* **14**, 457-60 (1996).

16. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-8 (1998).
17. Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W. & Deetz, K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**, 357-62 (1995).
18. Ginzinger, D. G. et al. Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. *Cancer Res* **60**, 5405-9 (2000).
19. Sotiriou, C., Khanna, C., Jazaeri, A. A., Petersen, D. & Liu, E. T. Core biopsies can be used to distinguish differences in expression profiling by cDNA microarrays. *J Mol Diagn* **4**, 30-6 (2002).
20. Dobson, A. T. et al. The unique transcriptome through day 3 of human preimplantation development. *Hum Mol Genet* **13**, 1461-70 (2004).

Table 1

Performance of Amplification Method by Evaluation of Array Expression Ratios as Compared to Taqman of 37 Genes

Method	Quantity of input	#False	%FER
ModT7	500pg	5	13.5%
	500pg	5	13.5%
	500pg	6	19%
	1ng ¹	6	19%
	1ng ¹	5	13.5%
Mean		5.4	14.6%
Arcturus	250pg ¹	6	19%
	250pg ¹	6	19%
	500pg ³	4	10.8%
	1ng	4	10.8%
Mean		5	13.5%
Balanced PCR	500pg ⁴	1	2.7%
	667 pg ²	5	13.5%
	667 pg ²	4	10.8%
	3.3 ng	5	13.5
	3.3 ng	6	19%
Mean		4.2	11.3%

¹ Threshold level in which linear technique could repeatedly amplify

² Threshold level of input at logarithmic method could reliably amplify

³ Manufacturer's stated lower recommended threshold

⁴ Could not be repeatedly amplify at this input threshold

Figure 1

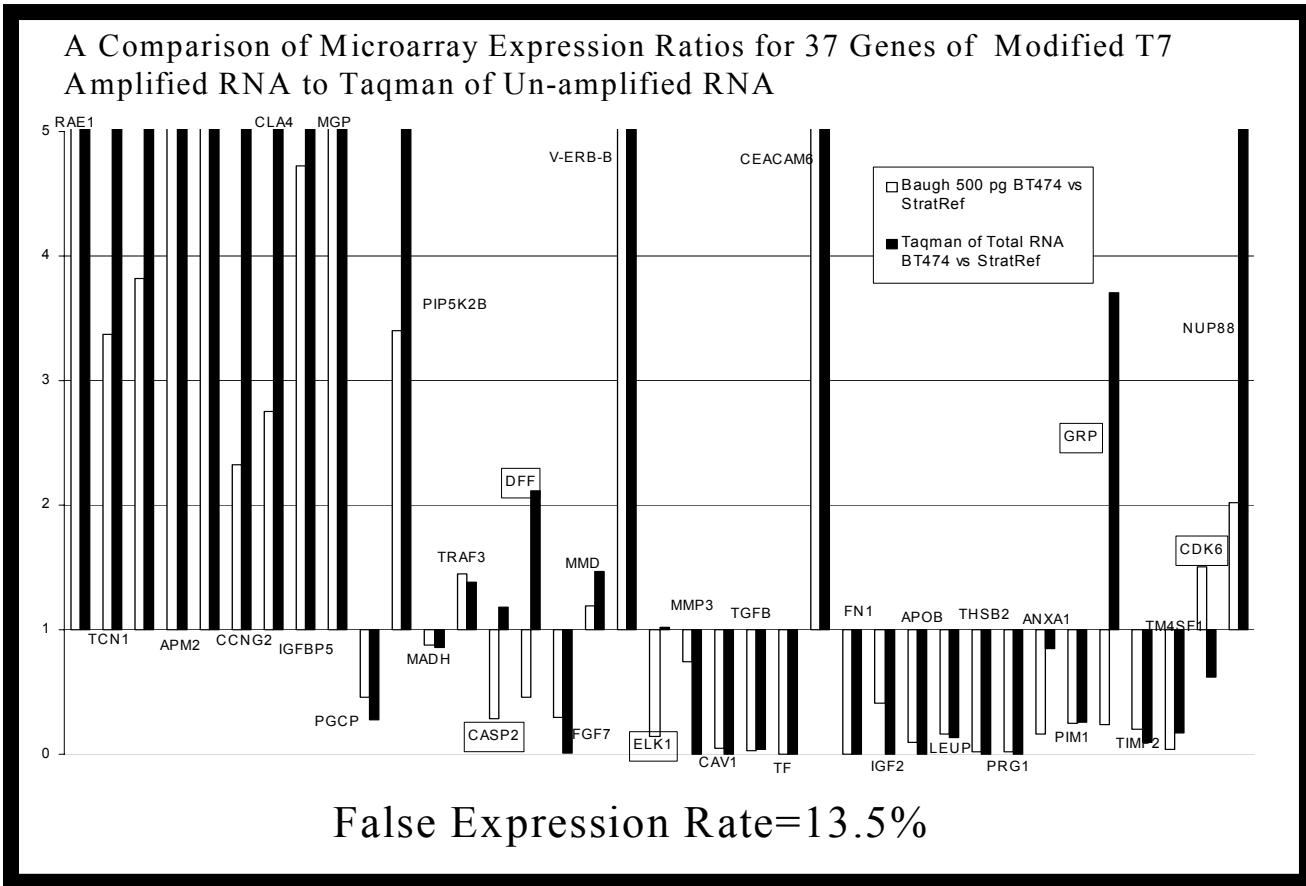


Figure 2

Correlation in Expression Ratios Between Taqman of Amplified and Total RNA for 3 Different Methods

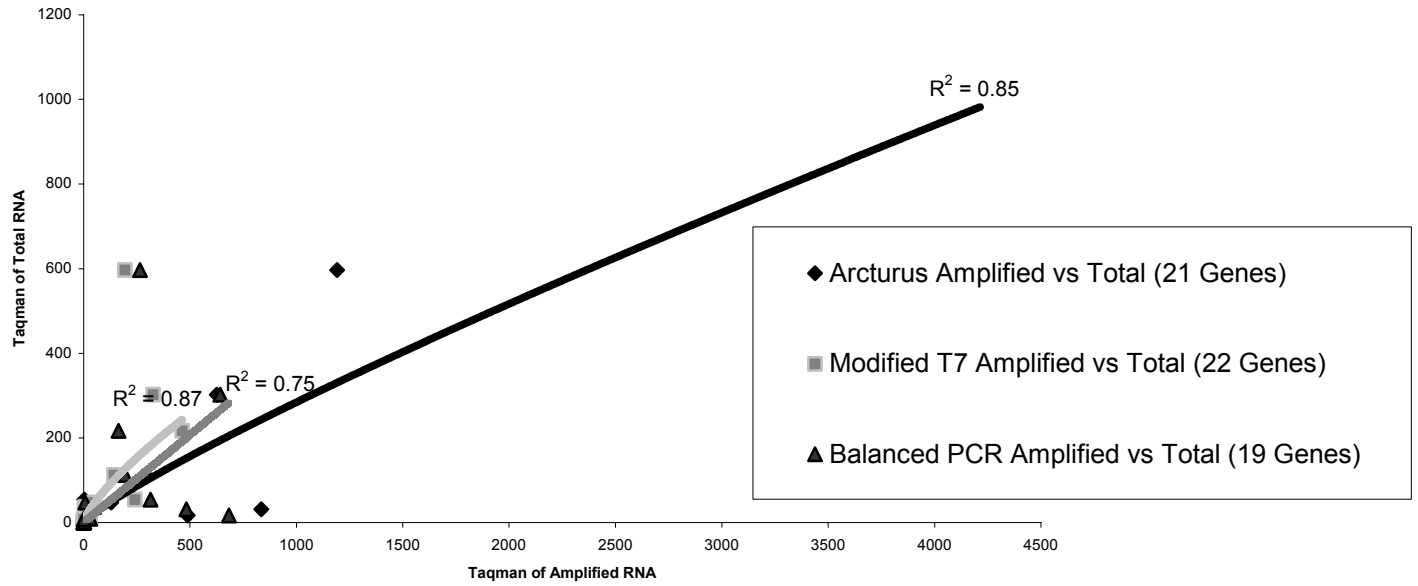


Table 2

Taqman of Amplified RNA: Accuracy of 3 Methods

Method Name	Number of Genes Analyzed	Number False	%FER
Arcturus	21	0	0
Modified T7	22	1	4.5
Balanced PCR	19	3	15

Table 3 Comparison of Expenses for 3 Amplification Techniques

Balanced pcr expenses			
Reagent	Unit Price		Expense/sample
Advantagell	\$380/100x		\$1.90
Titanium	\$260/100x		\$15.60
Ligase	\$176/50ul		\$0.88
NLAIII	\$40/50ul		\$0.20
Qiagen	\$68/50x		\$2.72
Oligo+primers	\$168		\$0.50
			\$21.80
			x 2 (sample +reference)
Total cost/reaction for reagents			\$43.60
Tech time		1 day	\$131.15
Direct Labeling of balanced pcr product			
BioPrime	\$238/kit		\$13.60
Cy3dUTP	\$435		\$70.83
Cy5dUTP	\$435		\$70.83
dNTP's 10mM	\$178		\$0.01
Qiagen	\$68/50x		\$2.72
Cot	\$145/500uL		\$2.46
Quia quick	\$77		\$1.54
Tech time		2 hours	\$32.78
Total for labeling reagents			\$161.99
Total cost/reaction (amplification and labelling) reagents			\$205.59
Arcturus expenses			
			\$139.00
			x 2 (sample +reference)
RiboAmpHS kit reagents	\$695/5 samples		\$278.00
Tech time		2 days	\$262.30
RT			
N6 hexamers	\$134	makes 1ml	\$0.23
Stratascript RT	\$160	200 ul	\$2.40
aa dUTP	\$83/mg	makes 100ul	\$0.50
Microcon 30	\$214		\$2.14
Cot	\$145/500uL		\$2.46
Cy3	\$216		\$18.00
Cy5	\$216		\$18.00
Quia quick	\$77		\$1.54
Tech time		1 day	\$131.15
Total for labeling reagents			\$45.27

Total cost/reaction (amplification and labelling) reagents			\$323.27
Baugh			
dTT7 oligo	\$75	makes 800ul	\$0.09
SuperScript II RT	\$171	50 ul	\$1.54
dNTP's 10mM	\$178	10000 ul	\$0.01
T4GP32	\$337	100 ul	\$2.29
Rnase inhibitor	\$227	250 ul	\$0.23
E. Coli DNA Pol I	\$50	100 ul	\$1.00
E. Coli DNA ligase	\$212/1000U	100 ul	\$1.06
E. Coli Rnase H	\$113/50U	33 ul	\$1.71
5X Second Strand Buffer	\$64	500 ul	\$1.92
dNTP's	\$178	10000 ul	\$0.03
T4 DNA Polymerase	\$45	50 ul	\$2.97
Zymo Clean and Concentrator	\$220/200rxn		\$1.10
Ampliscribe	\$205	200 ul	\$2.05
100mM ATP	\$72	400 ul	\$0.27
100mM CTP	\$72	400 ul	\$0.27
100mM UTP	\$72	400 ul	\$0.27
100mM GTP	\$72	400 ul	\$0.27
Rnase inhib.	\$227	250 ul	\$0.68
T7 RNA Polymerase	\$276	125 ul	\$2.21
Quiagen Rneasy kit	\$857/250rxn		\$3.43
random hexamers	\$134	makes 1ml	\$0.23
SuperScript II RT	\$171	50 ul	\$1.54
dNTP's 10mM			\$0.01
T4GP32	\$337	100 ul	\$2.29
Rnase inhibitor	\$84	250 ul	\$0.23
dTT7 oligo	\$70	makes 800ul	\$0.09
E. Coli DNA Pol I	\$50/50 ul	100 ul	\$2.00
E. Coli Rnase H	\$113/50U	33 ul	\$1.71
5X Second Strand Buffer	\$64	500 ul	\$1.92
dNTP's			\$0.03
T4 DNA Polymerase	\$45	50 ul	\$2.97
Zymo Clean and Concentrator	\$220/200rxn		\$1.10
Tech time		3 days	\$393.45
			\$37.52
			x 2 (sample +reference)
Total cost/reaction for amplification reagents			\$75.04
RT			
N6 hexamers	\$134	makes 1ml	\$0.23
Stratascript RT	\$160	200 ul	\$2.40
aa dUTP	\$83/mg	makes 100ul	\$0.50
Microcon 30	\$214		\$2.14
Cot	\$145/500uL		\$2.46

Cy3	\$216	\$18.00
Cy5	\$216	\$18.00
Quia quick	\$77	\$1.54
Tech time	1 day	\$131.15
Total for labeling reagents		\$45.27
Total for amplification and labeling reagents		\$120.31

Table 3A - Final Costs by Method

Amplification Technique	Total Cost (including coupling and labor)	Technician time (days)
Modified T7	\$644.91	4
Arcturus	\$716.72	3
Balanced PCR	\$369.52	1.25

Published In Cell Biology: A laboratory handbook, 3rd Edition, Julio Celis, Editor,
Elsevier Publishing, London, UK, (2005) (Review).

PCR-based amplification method of retaining the quantitative difference between two complex genomes

Gang Wang, Brendan D. Price and G. Mike Makrigiorgos⁺,

Department of Radiation Oncology, Dana Farber Cancer Institute, Harvard
Medical School, Boston, MA, USA.

⁺Address correspondence to: G. Mike Makrigiorgos, Ph.D, Brigham and
Women's Hospital, Level L2, Radiation Therapy, 75 Francis street, Boston, MA
02115, USA. Tel (617) 6326905, Fax (617) 6326900, Email
mmakrigiorgos@partners.org

I. INTRODUCTION

Technologies for analyzing gene expression and gene copy number changes are increasingly used in the detection, diagnosis and therapy of cancer. The clinical outcome of various breast cancer therapies correlates closely with distinct mRNA expression profiles detected using DNA microarrays (Alizadeh et al. 2001; Perou et al. 1999; Ross and Perou 2001; Sorlie et al. 2001; van 't Veer et al. 2002). Array-based Comparative Genomic Hybridization (array-CGH) can detect the amplification or deletion of candidate breast cancer genes as well as genomic instability within tumor cells (Albertson et al. 2000; Kallioniemi et al. 1994; Kallioniemi et al. 1992; Pinkel et al. 1998; Pollack et al. 1999). Subtractive hybridization methods, such as Differential Display or Representational Difference Analysis are also used for breast cancer gene discovery (Scheurle et al. 2000). Such genetic profiling-based diagnosis can potentially revolutionize the existing staging system and the management of early breast disease (Burki et al. 2000). However, analysis of genetic changes in tumors using these techniques requires ‘ μ gs’ of pure tumor DNA (Klein et al. 1999; Lucito et al. 1998). Routine tumor biopsies often consist of inhomogenous mixtures of stromal cells plus tumor cells with a wide range of genetic profiles (Rubin 2002). Newer techniques, such as Fine Needle Aspiration (FNA) and Laser Capture Microdissection (LCM), allow for the removal of minute amounts of tissue from tumors (Rubin 2002). LCM can isolate homogeneous populations of normal or tumor cells, potentially resolving tissue into single cells (Assersohn et al. 2002; Emmert-Buck et al. 1996). However, the yield of RNA/DNA from small cell numbers dictates that LCM must be coupled to a DNA amplification step, usually by use of the Polymerase Chain Reaction (PCR (Assersohn et al. 2002)).

A major problem with PCR is that amplification occurs in a non-linear manner and reproducibility is influenced by stray impurities (Heid 1996). The exponential mode of DNA amplification and the concentration-dependent PCR saturation are notorious for introduction of bias (Heid 1996). As a result, when amplifying two complex DNA populations, the quantitative relationship between two genes after amplification is generally not the same as their relation prior to amplification. Real time PCR strategies can retain the initial relation among alleles when a single gene is amplified from two sources (Celi et al. 1994). Further, methods exist to PCR-amplify whole genomic DNA from as little as a single cell (Klein et al. 1999; Nelson et al. 1989; Zhang et al. 1992a). However, the quantitative amplification of the entire population of DNA fragments (‘alleles’) from two different, complex genomes is not possible using conventional PCR. Multiple strand displacement isothermal amplification (MDA) is an alternative to

PCR that has shown promise in a number of investigations (Dean et al. 2002; Zhang et al. 1992b). On the other hand, MDA requires long DNA stretches to work effectively and therefore it is inefficient when formalin fixed, archival genomic DNA is to be amplified (Lage et al. 2003) or when cDNA amplification for gene expression profiling on microarrays is required.

We have recently described balanced-PCR (Makrigiorgos et al. 2002), a method which overcomes biases associated with PCR-amplification of complex genomes and faithfully retains the difference among corresponding genes, or gene fragments over the entire sample. This approach, which can be applied to the amplification of both genomic DNA and cDNA, utilizes a simple principle (**Figure 1**). Two distinct genomic DNA samples, a ‘target’ sample and a ‘control’ sample, are tagged with oligonucleotides (LN1, LN2) containing both a common (P1) and unique DNA sequence (P2a, P2b). The genomic DNA samples are pooled and amplified in a single PCR tube using the common DNA tag, P1. By mixing the 2 genomes, PCR ‘loses’ the ability to discriminate between the different alleles and the influence of impurities tends to cancel. The PCR-amplified pooled samples can subsequently be differentially labeled or separated using the DNA tag unique to each individual DNA sample. This balanced-PCR approach has been validated with amplification of cDNA for gene expression profiling (Makrigiorgos et al. 2002) and genomic DNA for array-CGH profiling (Wang et al, submitted for publication).

II. MATERIALS

Nla-III (Cat. No. R0125S), DpnII (Cat. No. R0543S), Sau3A (Cat. No. R0169S) and T4 DNA ligase (Cat. No. M0202T) were purchased from New England Biolabs. Advantage™ 2 PCR Kit (K1910-1) and TITANIUM™ Taq PCR Kit (K1915-1) were purchased from BD Biosciences. RNeasy Mini Kit (Cat. No. 74104) and QIAquick PCR Purification Kit (Cat. No. 28104) was purchased from QIAGEN. SuperScript Double-Stranded cDNA Synthesis Kit (Cat. No. 11917-020) was purchased from Invitrogen. Picogreen™ dsDNA Quantitation reagent (P-7581) was purchased from Molecular Probes. Linkers were synthesized from Oligos Etc. PCR reactions were performed with a TechGene™ thermocycler (TECHNE).

III. PROCEDURES

1. Double-strand cDNA synthesis. The protocols recommended by the manufacturers were used to extract total RNA from breast or prostate cells (RNeasy Mini Kit), to reverse transcribe to cDNA using Oligo(dT)₁₂₋₁₈ primers, and to synthesize double stranded cDNA (SuperScript Double-Stranded cDNA Synthesis Kit).
2. Balanced-PCR protocol. This procedure is a modification of the one originally reported (Makrigiorgos et al. 2002), and can be used for amplification of either cDNA or whole genomic DNA. The procedure has been tested with starting amounts of 1-10 ng total mRNA and with 1-10 ng of total genomic DNA extracted from target (e.g. tumor) and control (e.g. normal tissue) cells.

Steps:

1. Digestion. The protocol described here employs either NlaIII or DpnII/Sau3A for double stranded cDNA digestion. Mix 1 µl of 10 ng/µl cDNA from the target cells (e.g. tumor) or from the control cells (e.g. normal tissue) with 0.5 µl of 10x T4 DNA ligase buffer, 0.5µl of 10 U/µl NlaIII/DpnII/Sau3A, and 3 µl of H₂O. Incubate this mixture at 37°C for one hour.

2. Ligation. Add 0.5 µl of 10x ligase buffer, 0.3 µl of 2.8 µg/µl linker, and 3.7 µl H₂O into digestion solution. For digestion with NlaIII, linker LN1 is used for control and LN2 for target cDNA (**Table I**). For digestion with DpnII or Sau3A, linker LN1 and an equimolar amount LN1a are used for ligation to the control cDNA; and linker LN2 and an equimolar amount of LN2a are used for ligation to the target cDNA (**Table II**). Anneal the appropriate linkers to cDNA by serially decreasing temperature of the sample from 50 °C to 10 °C at 5 °C ramp in 5 minute steps. Then add 0.5 µl of 2,000U/µl T4 DNA ligase and incubate at room temperature for 1 hour.

3. Purification. Mix together cDNAs ligated to different linkers and purify the mixture with a QIAquick™ PCR Purification Kit. Purification is not needed if only a fraction of the ligation mixture (e.g. 10% of the total volume) is used in the subsequent co-amplification PCR reaction.

4. Co-amplification PCR. To 20 µl of purified-ligated DNA, add 5 µl of 10x Advantage™ 2 PCR Buffer, 1 µl of 50x Advantage™ 2 Polymerase Mix, 1 µl of 50x dNTP mix (10 mM ea.), 1 µl of 10 µM common primer P1 and 22 µl of H₂O. PCR is performed at 72 °C for 8 minutes; 95 °C for 1 minute; 20

cycles of 95 °C for 30 seconds and 72 °C for 1 minutes; then 72 °C for 5 minutes. Purify PCR product twice with QIAquick™ PCR Purification Kit and elute the DNA in 50 µl of H₂O. Quantify cDNA concentration with Picogreen™. This procedure usually yields 2-3 µg cDNA from an original material of ~5 ng cDNA.

5. Separation. Mix 1 µl of 3 ng/µl DNA with 5 µl of 10x TITANIUM™ Taq PCR Buffer, 1 µl of 50x TITANIUM™ Taq Polymerase, 1 µl of 50x dNTP Mix (10 mM ea.), 5 µl of 4 µM P2a for LN1-ligated cDNA or P2b for LN2-ligated cDNA, and 37 µl of H₂O. Separate and amplify cDNA at 95 °C for 1 minute; 10 cycles of 95 °C for 30 seconds and 72 °C for 1 minute; and 72 °C for 5 minutes. Each 10-cycle PCR reaction is expected to produce 1-1.5 µg cDNA. Scale the number of individual reactions as needed to produce the desired total amount of amplified cDNA.

IV. EXAMPLES

Microarray screening for prostate and lung cDNA, before and after *balanced* PCR. As an example of balanced PCR's ability to retain the difference among alleles between two cDNA populations, microarray studies of human prostate (representing the 'target') and lung-derived cDNA (representing the 'control') were employed. Digested cDNA was ligated to linkers, and directly screened on the Affymetrix Genechip^R Cancer microarrays following the procedure we described earlier (Zhang et al. 2001). Next, prostate and lung cDNA samples were 1:1 mixed, and amplified via balanced PCR for three consecutive PCR rounds of 20 cycles each. The samples were then separated using the procedure of **Figure 1** and screened on microarrays. The ratio of signal intensities after balanced PCR was plotted versus the same ratio prior to balanced PCR (**Figure 2, Frame A**). The ratio of expression levels for the majority of genes remained relatively unchanged after balanced PCR, as indicated by the distribution of data in **Frame A** ($R^2=0.92$). Next the experiment was repeated the 'traditional' way, i.e. by PCR-amplifying separately the prostate and lung cDNA samples and screening each on microarrays (**Figure 2, Frame B**). The data indicate that, for a substantial fraction of genes the ratio of expression levels is substantially different from the original one, presumably due to PCR-introduced changes in the original relative expression levels among prostate and lung ($R^2=0.38$).

In **Figure 3, A and B**, the comparison between balanced-PCR and conventional PCR is depicted for 30 genes that presented the highest up-regulation in prostate versus lung. Most are widely known prostate – specific genes, such as the prostate specific antigen (PSA), prostatic acid phosphatase, and prostatic kallikrein. **Figure 3 frame A** indicates a good retention of the relative expression levels before and after balanced-PCR for almost all these genes (correlation coefficient=0.800). In contrast, **Figure 3 frame B** demonstrates that distortions are introduced if the samples are amplified separately, using conventional PCR, presumably due to a PCR-introduced change in the original relative expression levels among prostate and lung (correlation coefficient=0.28). Genes important to prostate cancer development, such as prostate-specific antigen (PSA) and prostatic acid phosphatase are overestimated by more than a factor of 10 when amplified via traditional PCR but correctly quantitated when amplified via balanced-PCR prior to microarray screening. Of all 407 genes considered the percent of genes that had their relative signal change by more than two-fold or by more than 1.3-fold after performing PCR amplification is depicted in **Figure 3 frame C**. Since the deviations observed using balanced-PCR are less or equal to the microarray – related deviation (established by repeated application of a single sample on different arrays (Makrigiorgos et al. 2002)) it is concluded that balanced-PCR introduced minimal distortion in the relative expression among prostate and lung (i.e. balanced-PCR error < array error).

V. POTENTIAL PITFALLS USING BALANCED-PCR

- Efficiency of enzymatic treatments. A requirement for the success of balanced-PCR is that treatment of target and control DNA is identical at all stages prior to mixing the samples. We conducted control studies and we included internal standards for digestion using Sau3A and ligation to derive the efficiency of digestion and ligation steps (Makrigiorgos et al. 2002). Both were found to be more than 95% efficient. However, if the enzymatic efficiency is reduced due to degradation of the enzyme stocks, impurities, or due to methylation sensitivity bias may be introduced in the first step of the procedure. This can be avoided by using freshly obtained enzymes that are highly efficient and that are not sensitive to mammalian CpG methylation.
- Post-PCR separation. Another assumption is that the low-cycle PCR used for re-separation of the two genomes following the common PCR step does not produce distortions among DNA samples. It is in principle possible that this PCR might itself produce some bias among alleles in

the two populations. In practice however we have found that this 10 cycle separation PCR does not introduce significant distortion among alleles differing by at least 50-fold in initial concentration in any of the systems examined (plasmid, genomic DNA, cDNA, (Makrigiorgos et al. 2002)). However, it is not recommended to increase the separation-PCR cycles to beyond 10.

- The effect of mutations and polymorphisms: Balanced PCR uses templates from enzyme-digested fragments. If mutations occur within the restriction sequences in the target or control cDNAs then the enzyme will not digest at that position, but will act in the next available restriction sequence. As a result, certain gene fragments in the target genome will be different in size from their alleles in the control genome and PCR amplification may introduce bias if the fragment sizes are too different. Mutations that occur specifically at the restriction sites are not frequent. The most common form of mutations is single nucleotide polymorphisms (SNPs) which, between two given genomes, occur with a frequency of about 1:1000 bases. The chances that a 4-base cutter enzyme used in balanced PCR encounters a SNP is roughly $4/1000 = 0.4\%$, and therefore it would affect only a small fraction of the sequences amplified. Since several SNPs are already tabulated in databases and more will become known in the near future, one can use computational methods to predict which restriction sites will be altered due to a SNP in order to anticipate potential PCR-bias at these positions. If these sequences are vital, one may perform balanced PCR using a different restriction enzyme.

ACKNOWLEDGMENTS. Funding for this work was provided in part by DOD grant BC020504.

FIGURE LEGENDS

Figure 1: Outline of balanced-PCR amplification of cDNA or genomic DNA (reproduced with permission from Nature Publishing Group).

Figure 2: Comparison of relative expression of lung vs prostate tissue on microarrays, before and after PCR amplification. **Frame A**, amplification conducted using the current balanced-PCR method. **Frame B**, amplification conducted by performing conventional PCR, separately on lung and prostate cDNA samples.

Figure 3: Comparison of relative expression of lung vs. prostate specifically for the 30 genes highest-upregulated in prostate vs. lung. **Frame A**, amplification conducted using the current balanced PCR method. **Frame B**, amplification conducted by performing conventional PCR, separately on lung and prostate cDNA samples. **Frame C**. Fraction of genes whose relative expression among prostate and lung changes by more than 100% (**columns 1-3**) or 30% (**columns 4-6**) following PCR amplification. **Columns 1 and 4**, repeated application of the same sample on microarrays. **Columns 2 and 5**, amplification via balanced-PCR. **Columns 3 and 6**, amplification via conventional PCR (reproduced with permission from Nature Publishing Group).

Tables I and II: Linkers and primers used in conjunction with NlaIII DNA digestion (**Table I**) or DpnII/Sau3A digestion (**Table II**).

SUPPLIER ADDRESSES

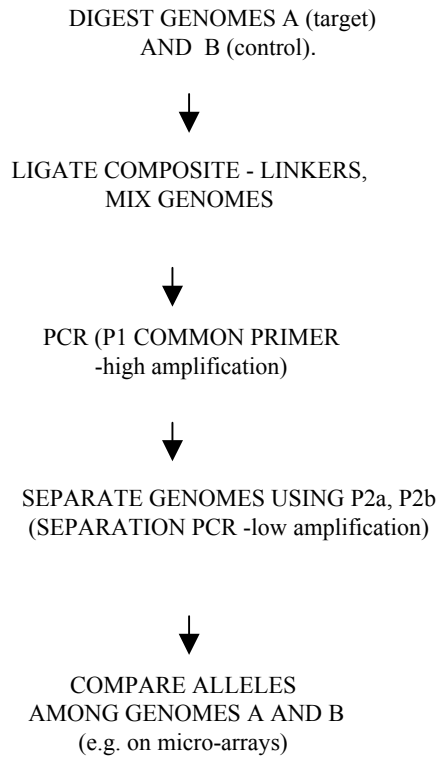
1. NEW ENGLAND BIOLABS
32 Tozer Road
Beverly, MA 01915-5599
(978) 927- 5054
2. BD BIOSCIENCES CLONTECH
1020 East Meadow Circle
Palo Alto, CA 94303-4230
3. QIAGEN Inc.
28159 Avenue Stanford
Valencia, CA 91355
4. Molecular Probes, Inc.
29851 Willow Creek Road
Eugene, OR 97402
5. Oligos Etc. Inc.
PO Box 727
9775 SW Commerce Circle C-6
Wilsonville, OR 97070
6. Techne
Duxford Cambridge
CB2 4PZ
England
7. Invitrogen Corporation
1600 Faraday Avenue
PO Box 6482
Carlsbad, California 92008

REFERENCES

- Albertson, D.G., B. Ylstra, R. Segraves, C. Collins, S.H. Dairkee, D. Kowbel, W.L. Kuo, J.W. Gray, and D. Pinkel. 2000. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat Genet* **25**: 144-146.
- Alizadeh, A.A., D.T. Ross, C.M. Perou, and M. van de Rijn. 2001. Towards a novel classification of human malignancies based on gene expression patterns. *J Pathol* **195**: 41-52.
- Assersohn, L., L. Gangi, Y. Zhao, M. Dowsett, R. Simon, T.J. Powles, and E.T. Liu. 2002. The feasibility of using fine needle aspiration from primary breast cancers for cDNA microarray analyses. *Clin Cancer Res* **8**: 794-801.
- Burki, N.G., R. Caduff, H. Walt, C. Moll, T. Pejovic, U. Haller, and D.C. Ward. 2000. Comparative genomic hybridization of fine needle aspirates from breast carcinomas. *Int J Cancer* **88**: 607-613.
- Celi, F.S., M.M. Cohen, S.E. Antonarakis, E. Wertheimer, J. Roth, and A.R. Shuldiner. 1994. Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (gd-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics* **21**: 304-310.
- Dean, F.B., S. Hosono, L. Fang, X. Wu, A.F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S.F. Kingsmore, M. Egholm, and R.S. Lasken. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* **99**: 5261-5266.
- Emmert-Buck, M.R., R.F. Bonner, P.D. Smith, R.F. Chuaqui, Z. Zhuang, S.R. Goldstein, R.A. Weiss, and L.A. Liotta. 1996. Laser capture microdissection. *Science* **274**: 998-1001.
- Heid, C., Stevens, J., Livak, K., Williams, P. 1996. *Real time quantitative PCR. Genome Methods: Genome Research*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Kallioniemi, A., O.P. Kallioniemi, J. Piper, M. Tanner, T. Stokke, L. Chen, H.S. Smith, D. Pinkel, J.W. Gray, and F.M. Waldman. 1994. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A* **91**: 2156-2160.
- Kallioniemi, A., O.P. Kallioniemi, D. Sudar, D. Rutovitz, J.W. Gray, F. Waldman, and D. Pinkel. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**: 818-821.
- Klein, C.A., O. Schmidt-Kittler, J.A. Schardt, K. Pantel, M.R. Speicher, and G. Riethmuller. 1999. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A* **96**: 4494-4499.
- Lage, J.M., J.H. Leamon, T. Pejovic, S. Hamann, M. Lacey, D. Dillon, R. Segraves, B. Vossbrinck, A. Gonzalez, D. Pinkel, D.G. Albertson, J. Costa, and P.M. Lizardi. 2003. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res* **13**: 294-307.

- Lucito, R., M. Nakimura, J.A. West, Y. Han, K. Chin, K. Jensen, R. McCombie, J.W. Gray, and M. Wigler. 1998. Genetic analysis using genomic representations. *Proc Natl Acad Sci U S A* **95**: 4487-4492.
- Makrigiorgos, G.M., S. Chakrabarti, Y. Zhang, M. Kaur, and B.D. Price. 2002. A PCR-based amplification method retaining the quantitative difference between two complex genomes. *Nat Biotechnol* **20**: 936-939.
- Nelson, D.L., S.A. Ledbetter, L. Corbo, M.F. Victoria, R. Ramirez-Solis, T.D. Webster, D.H. Ledbetter, and C.T. Caskey. 1989. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl Acad Sci U S A* **86**: 6686-6690.
- Perou, C.M., S.S. Jeffrey, M. van de Rijn, C.A. Rees, M.B. Eisen, D.T. Ross, A. Pergamenschikov, C.F. Williams, S.X. Zhu, J.C. Lee, D. Lashkari, D. Shalon, P.O. Brown, and D. Botstein. 1999. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* **96**: 9212-9217.
- Pinkel, D., R. Seagraves, D. Sudar, S. Clark, I. Poole, D. Kowbel, C. Collins, W.L. Kuo, C. Chen, Y. Zhai, S.H. Dairkee, B.M. Ljung, J.W. Gray, and D.G. Albertson. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**: 207-211.
- Pollack, J.R., C.M. Perou, A.A. Alizadeh, M.B. Eisen, A. Pergamenschikov, C.F. Williams, S.S. Jeffrey, D. Botstein, and P.O. Brown. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**: 41-46.
- Ross, D.T. and C.M. Perou. 2001. A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. *Dis Markers* **17**: 99-109.
- Rubin, M.A. 2002. Understanding disease cell by cell. *Science* **296**: 1329-1330.
- Scheurle, D., M.P. DeYoung, D.M. Binniger, H. Page, M. Jahanzeb, and R. Narayanan. 2000. Cancer gene discovery using digital differential display. *Cancer Res* **60**: 4037-4043.
- Sorlie, T., C.M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, T. Thorsen, H. Quist, J.C. Matese, P.O. Brown, D. Botstein, P. Eystein Lonning, and A.L. Borresen-Dale. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**: 10869-10874.
- Telenius, H., N.P. Carter, C.E. Bebb, M. Nordenskjold, B.A. Ponder, and A. Tunnacliffe. 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**: 718-725.
- van 't Veer, L.J., H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, and S.H. Friend. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530-536.
- Zhang, L., X. Cui, K. Schmitt, R. Hubert, W. Navidi, and N. Arnheim. 1992a. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A* **89**: 5847-5851.
- Zhang, L., X. Cui, K. Schmitt, R. Hubert, W. Navidi, and N. Arnheim. 1992b. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A* **89**: 5847-5851.

Zhang, Y., B.D. Price, S. Tetradis, S. Chakrabarti, G. Maulik, and G.M. Makrigiorgos.
2001. Reproducible and inexpensive probe preparation for oligonucleotide arrays.
Nucleic Acids Res **29**: E66-66, **APPENDED**.



EXAMPLE OF A 'COMPOSITE' LINKER

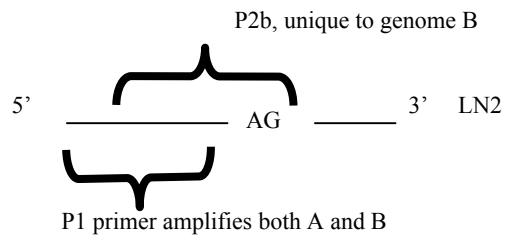
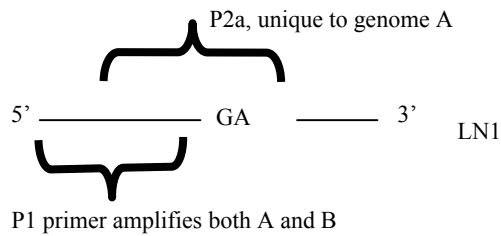


Figure 1

PROSTATE cDNA → Cut, ligate linker LN1

LUNG cDNA → Cut, ligate linker LN2

- Screen on microarrays
- Mix, apply balanced-PCR, screen on microarrays
- Apply conventional PCR, screen on microarrays

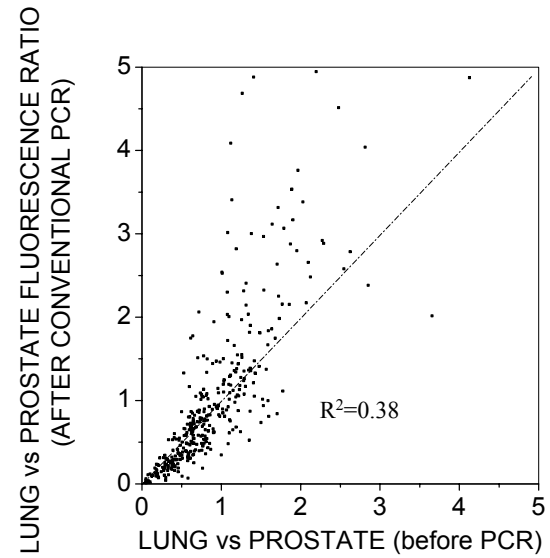
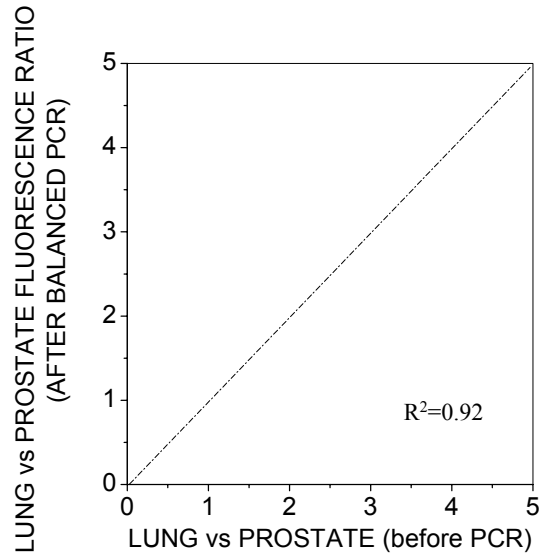


Figure 2

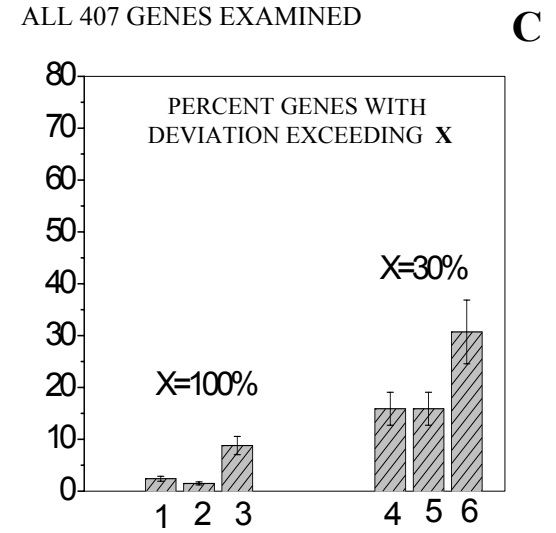
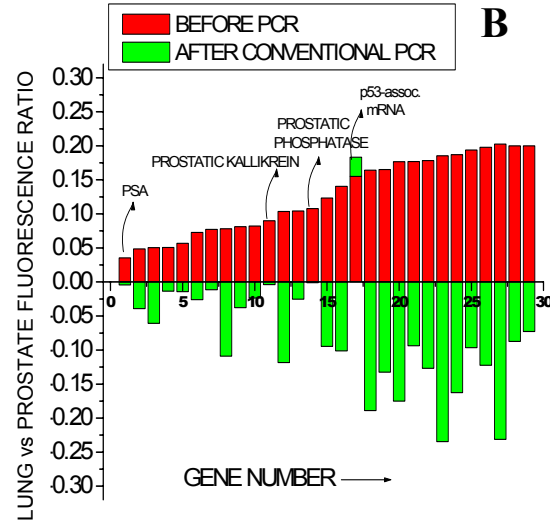
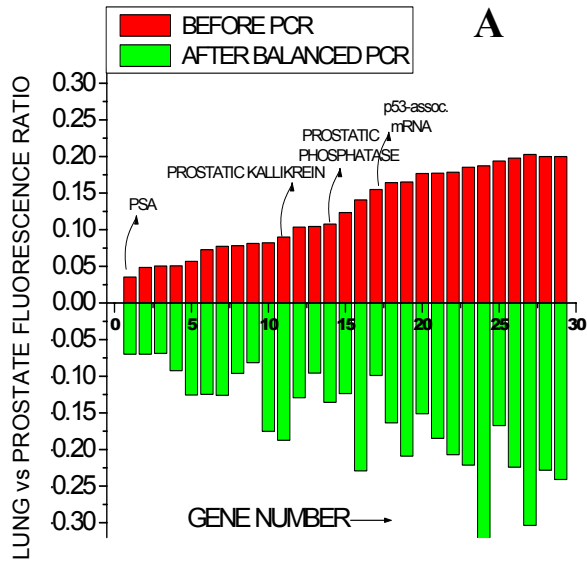


Figure 3

Table I. Sequences of Linkers and Primers used in conjunction with Nla-III digestion

		Sequences (5'-3')
Linkers & Primers	LN1	AACTGTGCTATCCGAGGGAAAGGACATG
	LN2	AACTGTGCTATCCGAGGGAAAGAGCATG
	P1	AGGCAACTGTGCTATCCGAGGGAA
	P2a	AACTGTGCTATCCGAGGGAAAGGA
	P2b	AACTGTGCTATCCGAGGGAAAGAG

Table II. Sequences of Linkers and Primers used in conjunction with Dpn-II or Sau3A digestion

		Sequences (5'-3')
Primers & Linkers	LN1	AACTGTGCTATCCGAGGGAAAGGACATG
	LN1a	GATCCATGTCCT
	LN2	AACTGTGCTATCCGAGGGAAAGAGCATG
	LN2b	GATCCATGCTCT
	P1	AGGCAACTGTGCTATCCGAGGGAA
	P2a	AACTGTGCTATCCGAGGGAAAGGA
	P2b	AACTGTGCTATCCGAGGGAAAGAG