AD_____

Award Number: DAMD17-94-J-4112

TITLE: Fellowship to Identify New Mechanisms of Tamoxifen Resistance in Breast Cancer Patients

PRINCIPAL INVESTIGATOR: Suzanne A.W. Fuqua, Ph.D.

CONTRACTING ORGANIZATION: The University of Texas Health Science Center at San Antonio
San Antonio, Texas 78284-7758

REPORT DATE: May 1999

TYPE OF REPORT: Final

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1999	Final (15 Sep 94 – 14 Apr 99)

4. TITLE AND SUBTITLE

Fellowship to Identify New Mechanisms of Tamoxifen Resistance in Breast Cancer Patients

5. FUNDING NUMBERS

DAMD17-94-J-4112

6. AUTHOR(S)

Suzanne A.W. Fuqua, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Texas Health Science
Center at San Antonio
San Antonio, Texas 78284-7758 email: sfuqua@bem.tmc.edu

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. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

The non-steroidal antiestrogen tamoxifen (TAM) has been used successfully in the treatment of tens of thousands of women with breast cancer, and has been shown to increase both disease-free and overall survival in treated patients. Unfortunately, virtually all patients treated with TAM eventually develop resistant disease. The research funded by this fellowship was directed at increasing our understanding of the mechanisms leading to the development of tamoxifen resistance. The two Specific Aims of this proposal have been completed and three altered estrogen receptors (ERs) have been identified from clinical samples, one of which exhibits a hormone-independent phenotype in transient assays (Aim 1 and 2). This work represents the work of the first postdoctoral fellow to work on the project, Dr. Douglas Wolf. We have also worked out the technical aspects of microdissection of archival and frozen clinical tumor samples to improve the use of differential display technologies to examine these samples for altered gene expression coincident with the phenotype of tamoxifen resistance (Aims 3 and 4). Dr. Wolf left the laboratory after completion of Aim 3, and Aim 4 of the fellowship project was completed by Dr. Rhonda Hansen.

14. SUBJECT TERMS

Tamoxifen resistance, estrogen receptor mutations

15. NUMBER OF PAGES 37

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

___ Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

___ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

NA In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

NA For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

X In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

X In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

___ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature 2/14/00 Date

TABLE OF CONTENTS

MAIN REPORT

ABSTRACT

The non-steroidal antiestrogen tamoxifen (TAM) has been used successfully in the treatment of tens of thousands of women with breast cancer, and has been shown to increase both disease-free and overall survival in treated patients. Unfortunately, virtually all patients treated with TAM eventually develop resistant disease. The research funded by this fellowship is directed at increasing our understanding of the mechanisms leading to the development of tamoxifen resistance.

The two Specific Aims of this proposal have been completed and three altered estrogen receptors (ERs) have been identified from clinical samples, one of which exhibits a hormone-independent phenotype in transient assays (Aim 1 and 2). This work was published in Cancer Research 57:1244-1249, and represents the work of the first postdoctoral fellow to work on the project, Dr. Douglas Wolf. We have also worked out the technical aspects of microdissection of archival and frozen clinical tumor samples to improve the use of differential display technologies to examine these samples for altered gene expression coincident with the phenotype of tamoxifen resistance (Aims 3 and 4). This work was published in a Laboratory Guide to RNA: Isolation, analysis and synthesis. John Wiley and Sons, New York, NY. Pages 237-249, 1996. Dr. Wolf left the laboratory after completion of Aim 3, and Aim4 of the fellowship project was completed by Dr. Rhonda Hansen. Her work was published in JNCI 91:453-459, 1999.

INTRODUCTION

Breast cancer is the most common cancer among women in America, and is second only to lung cancer as the cause of cancer death. Indeed, this year alone almost 180,000 American women will be diagnosed with the disease. Fortunately, many of these women will have tumors that are dependent upon estrogen for their growth, and

these patients will be successfully treated with endocrine therapies which are associated with much less severe side-effects that are most forms of chemotherapeutic treatment. The antiestrogen tamoxifen (Tam) is the most successful of these endocrine treatments, and has been shown to prolong disease-free and overall survival both in cases of advanced disease (1, 2), and when used as an adjuvant treatment (3). Unfortunately, however, most patients who are successfully treated with Tam eventually fail therapy, and their disease progresses to become Tam resistant.

Studies designed to increase our understanding of the mechanisms leading to Tam resistance are the subject of this postdoctoral fellowship proposal. In order to better understand the phenomena contributing to Tam resistance, we proposed the following **Specific Aims:**

1. To investigate the association between clinical Tam resistance and the presence of altered ERs.
2. To evaluate the functional activity of any receptor variants identified in Aim 1.
3. To identify new genes using differential display, the expression of which is associated with clinical Tam resistance.
4. To confirm the identity and evaluate the function of any genes identified in Aim 3.

BODY

Specific Aim 1

Thirty metastatic Tam-resistant breast cancer samples were screened for mutations in the ER gene by SSCP analysis. SSCP were identified in six of the metastatic tumors. The six positive samples were further studied by both direct sequencing analysis, and cloning of shifted bands followed by sequencing analysis.

Three of the SSCP shifted bands were found to constitute previously identified polymorphisms: TCT to TCC at nucleotide 30 in exon 1, TGC to TGT at nucleotide 720 in exon 3, and CGC to CGT at nucleotide 729 in exon 3, none of which results in amino acid substitutions. The remaining three shifted SSCP bands were found to constitute missense mutations: G140C (Ser47Thr) in exon 1, A1591G (Lys531Glu) in exon 8 , and T1609A (Tyr537Asn) in exon 8. While the former two sequence changes result in substitution of residues within hydrophilic amino acid groups, the later mutation changes a hydrophobic tyrosine to a hydrophilic asparagine.

It has been estimated that missense mutations are present in only about 1% (2/188) of primary tumors (4). In agreement with this, we have not detected any missense ER alterations in 60 primary breast cancers that we have examined using SSCP analysis (unpublished data). In this study of 30 cases of metastatic, Tam-resistant breast cancer, we detected three missense ER mutations. Thus ER mutations in metastatic breast tumors may be more frequent as compared to that seen in primary lesions (10% as compared to 1%, p=0.004 with Fisher's exact test).

Specific Aim 2

The functional transactivational status of the three missense ER mutations isolated from the metastatic breast tumors were investigated using transient transactivation assays, measuring the transcriptional activity of the mutant ERs with ERE-reporter gene constructs. Since the majority of studies examining the effect of specific alterations in the ER on function have utilized consensus EREs, most commonly the vitellogenin A2 ERE, we felt that it would be important to test ER function on constructs which might be more relevant to breast cancer biology, eg. those genes which are endogenously regulated by ER in breast cancer cells. We thus prepared ERE-CAT reporters to the estrogen-regulated pS2, Cathepsin D, and lactoferrin gene promoters. The activity of

Ser47Thr and Lys531Glu were not different from that of t wild-type ER using any of the four different ERE constructs. However, the Tyr537Asn mutant exhibited strong constitutive transactivation activity (15-20-fold over wild type ER activity on the vitellogenin ERE) in the absence of hormone. This elevated constitutive activity was also observed on the other ERE constructs. All of these results **(Specific Aims 1 and 2)** are shown in our paper in Cancer Research, which is included for review (Reprint #1).

Specific Aims 3 and 4

The second part of this research fellowship proposal (Specific Aims 3 and 4) involved the identification of new genes potentially involved in the development of Tam resistance. This was to be accomplished using the technique of differential display (DD) analysis of RNA isolated from Tam-sensitive vs. Tam-resistant tumors. We found in the first year of the grant that DD was not well suited for these tumors because they are composed of many different cell types, and thus we have spent the second year of the proposal addressing this technical issue. The results of DD from whole tumors are simply uninterpretable unless the RNA is prepared from a highly purified sample of one cell type. To this aim, we have now developed methodologies for microdissection from frozen and paraffin-embedded, archical breast tumor tissues. Microdissection is quite difficult on unstained slides, which motivated us to first do a study comparing various histochemical stains to find one that provided adequate visualization for microdissection without interfering with RNA isolation. Our results showed that nuclear-fast-red (NFR) satisfied these requirements, and we now use NFR routinely for all microdissections. We found that during microdissection, specific cells of interest could be visualized by staining briefly with NFR and quickly redrying the slide, allowing for better direct visualization and only slightly decreasing the yield of RNA. Accurately separating different types of cells obviously requires familiarity with the histopathological features of the tissue, and Dr. Craig Allred, a Breast Pathologist at this Institution and a member

of the Baylor Breast Cancer Research Team, collaborated with us on this aspect of the project. [These results summarized here for **Specific Aim 3** have been published in the Laboratory Guide to RNA Isolation book (Reprint #2 is included in the report for review).]

We originally started using DD to identify genes differentially expressed in tamoxifen-resistant patients. However, expression array techniques became available which are more sensitive and reliable than DD, and thus began to use this technique for **Specific Aim 4**. Again, we were faced with a technical problem, that of data analysis of expression array data, and we undertook a study to optimize the technique and develop methods for statistical analysis of the data. We were successful in both of these objectives and the data is published in our JNCI paper (Reprint #3) that is included in the Appendix for review. We will continue to study genes that were identified in this study for their relevance to clinical tamoxifen resistance.

KEY RESEARCH ACCOMPLISHMENTS

1. Identified ER mutation in a metatstatic breast tumor with constitutive activity
2. Developed methods for microdissection and RNA expression analysis of archival tumor samples
3. Identified gene expression patterns associated with tamoxifen resistance using expression array analysis

REPORTABLE OUTCOMES

1. (3) manuscripts included in the Appendix
2. Successful training of Drs. Doug Wolf and Rhonda Hansen. Both fellows are currently employed in academic institutions and are working on breast cancer research.

CONCLUSIONS:

We have successfully completed all of our Aims. Two postdoctoral fellows worked on this project and both have published their findings. Exciting results describing ER mutations were generated, and we have developed methods for examining clinical samples for expression array profiling to identify genes associated with tamoxifen resistance. The work and findings from this postdoctoral fellowship award will continue to be studied in the laboratory of the mentor, Suzanne A. W. Fuqua. Undoubtedly, the funds for this fellowship have been well spent with much success. It is also conceivable that this line of research will lead to direct translational benefit.

REFERENCES

1. Tanaka, M. et. al. Jpn. J. Clin. Oncol. 8:141, 1978.
2. Patterson, J.S., et. al. Revs. Endocrine-Related Cancer. 9:563, 1982.
3. Early Breast Cancer Trialists' Collaborative Group. Lancet 339:1, 1992.
4. Roodi, N., et. al. J. Natl Cancer Inst 87:446, 1995.

BIBLIOGRAPHY

Task 1 and 2: Zhang, Q - X., Borg, A., **Wolf, D**. M., Oesterreich, S., and Fuqua, S. A. W.: An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. Cancer Research, 57:1244-1249, 1997.

Task 3: Fuqua, S. A. W., Benedix, M. G., **Wolf, D**. M., Castles, C. G., and Allred, D. C.: Application of differential display to cancer research: Amplification of RNAs isolated from air-dried frozen and archival paraffin-embedded tissues. A Laboratory Guide to RNA: Isolation, analysis, and synthesis. John Wiley & Sons, Inc. New York, NY, 237-249, 1996

Task 4: Hilsenbeck, S. G., Friedrichs, W. G., Schiff, R., O' Connell, P., **Hansen, R. K.**, Osborne, C. K., and Fuqua, S. A. W.: Statistical analysis of microarray expression data as applied to the problem of tamoxifen resistance. JNCI 91:453-459, 1999.

PERSONNEL

Dr. **Douglas Wolf**, Postdoctoral Fellow, worked on the project from 1994-1996. No one was appointed on the grant from 1996 to1998 until we could recruit another fellow to work on the project. We were also delayed in this endeavor because we had to obtain an extension from Dr. Modrow to extend the funding of the project, and to authorize the appointment of another fellow to the grant. This was received in a letter dated September 18, 1997 from Dr. Modrow. The second individual recruited was Dr. Rhonda Hansen, a Postdoctoral Fellow who completed work on the grant in 1998.

APPENDICES

Reprint 1: Zhang, Q - X., Borg, A., **Wolf, D.** M., Oesterreich, S., and Fuqua, S. A. W.: An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. Cancer Research, 57:1244-1249, 1997.

Reprint 2: Fuqua, S. A. W., Benedix, M. G., **Wolf, D.** M., Castles, C. G., and Allred, D. C.: Application of differential display to cancer research: Amplification of RNAs isolated from air-dried frozen and archival paraffin-embedded tissues. A Laboratory Guide to RNA: Isolation, analysis, and synthesis. John Wiley & Sons, Inc. New York, NY., 237-249, 1996.

Reprint 3: Hilsenbeck, S. G., Friedrichs, W. G., Schiff, R., O' Connell, P., **Hansen, R. K.**, Osborne, C. K., and Fuqua, S. A. W.: Statistical analysis of microarray expression data as applied to the problem of tamoxifen resistance. JNCI 91:453-459, 1999.

[CANCER RESEARCH 57, 1244–1249, April 1, 1997]

Advances in Brief

An Estrogen Receptor Mutant with Strong Hormone-independent Activity from a Metastatic Breast Cancer[1]

Qiu-Xia Zhang, Åke Borg, Douglas M. Wolf, Steffi Oesterreich, and Suzanne A. W. Fuqua[2]

Department of Oncology, University Hospital, S-221 85 Lund, Sweden [Q-X. Z., Å. B.], and Department of Medical Oncology, University of Texas Health Science Center, San Antonio, Texas 78284-7884 [D. M. W., S. O., S. A. W. F.]

Abstract

Thirty tumors from metastatic breast cancer patients were screened for mutations in the estrogen receptor (*ER*) gene using single-strand conformation polymorphism and sequence analysis. Three missense mutations, Ser47Thr, Lys531Glu, and Tyr537Asn, were identified in these lesions. To investigate these mutated ERs or altered transcriptional activation function, expression vectors containing wild-type (wt) and mutant ERs were constructed and cotransfected with different estrogen response element reporter gene constructs into HeLa cells and MDA-MB-231 human breast cancer cells. The first two ER mutants were similar to wt ER. However, the Tyr537Asn ER mutant possessed a potent, estradiol-independent transcriptional activity, as compared to wt ER. Moreover, the constitutive activity of the Tyr537Asn ER mutant was virtually unaffected by estradiol, tamoxifen, or the pure antiestrogen ICI 164,384. Tyr537 is located at the beginning of exon 8 in the COOH-terminal portion of the hormone-binding domain of the ER, to which dimerization and transcription activation functions have also been ascribed. It has been identified as a phosphorylation site implicated in hormone binding, dimerization, and hormone-dependent transcriptional activity. Our results suggest that the Tyr537Asn substitution induces conformational changes in the ER that might mimic hormone binding, not affecting the ability of the receptor to dimerize, but conferring a constitutive transactivation function to the receptor. If present in other metastatic breast tumors, this naturally occurring ER mutant may contribute to breast cancer progression and/or hormone resistance.

Introduction

The human ER[3] belongs to the steroid/thyroid hormone receptor superfamily of ligand-activated transcription factors (1). Hormone binding is thought to induce conformational changes in the structure of the ER, homodimerization, and high-affinity binding of the hormone-ER dimer complex to well-defined palindromic DNA sequences, termed EREs, usually located upstream of estrogen-responsive genes (2). Transcription is then induced by two separate AFs of the ER, the NH_2-terminally located and constitutive AF-1 region and the ligand-inducible AF-2 region located within the hormone-binding domain (3), both of which are probably dependent on further interaction with specific sets of ER-associated proteins (4). Receptor activity is also thought to be modulated by phosphorylation and dephosphorylation on multiple serine and tyrosine residues through more or less defined signaling pathways (5, 6).

It is well accepted that the presence of ER identifies those breast cancer patients with a lower risk of disease recurrence and better survival, who may also have a better response to endocrine interventions (7). The discovery of ER variants with altered function, *e.g.*, transcriptionally active in the absence of estrogen (dominant-positive) or transcriptionally inactive but preventing the activity of wt ER (dominant-negative), may help elucidate the different responses of individual breast tumors to treatment (8). However, the ultimate clinical significance of the ER variants remains to be defined until their presence can be verified at the protein level in clinical studies. We and others have previously described the presence of several ER-splicing variants in human breast cancer cell lines and tumors (8–12). Single-bp changes within the ER, however, appear to be infrequent in primary breast tumors (13–15). To examine their occurrence in metastatic breast tumors, we used SSCP and sequence analysis to screen DNA from 30 tumor specimens for mutations in all eight coding exons of the *ER* gene. The transcriptional activity of the three missense mutations that we found was assessed by their transfection with different ERE-reporter gene constructs into HeLa cells and human MDA-MB-231 breast cancer cells. One of these naturally occurring ER mutants, Tyr537Asn, is at a previously described phosphorylation site located in the COOH-terminal end of the hormone-binding domain. This alteration conferred a potent constitutive transcriptional activity that was independent of estradiol binding and was essentially unaffected by Tam or the pure antiestrogen ICI 164,384.

Materials and Methods

Tumor Samples. Thirty breast tumor samples from residual tumor material remaining after routine steroid receptor measurements were utilized for the study. Tumors samples were frozen within 1 h from surgical removal and kept at −80°C for the present investigation. Data for these tumors and patients were collected from pathological examination and follow-up of clinical records. These cases were all from metastatic sites distant from the original breast lesion. Genomic DNA was isolated from 50–100 mg of frozen tumor tissues using phenol-chloroform extraction (16) and kept at 4°C until use. The patient exhibiting the Tyr537Asn substitution presented with stage IV disease, and the sample used for the analysis was a bone metastasis that recurred after treatment with diethylstilbestrol hormonal therapy. The metastatic bone sample was ER negative and progesterone receptor negative by ligand binding analysis. No additional clinical follow-up is available on this patient.

PCR Amplification and SSCP Analyses. Twelve primers were designed for analysis of genomic DNA according to the exon/intron locations defined in Ponglikitmongkol *et al.* (17). All coding exons were examined. The primer sequences, the expected sizes of the PCR products, and their location within the ER are shown in Table 1. The forward primers were 5′ biotinylated for direct sequencing analysis.

Eighty ng of genomic DNA were used as template in PCR, which was performed in a 30-μl volume, including 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.3–1.5 mM $MgCl_2$, 200 μM of each deoxynucleotide triphosphate, 0.13 μM of each primer, and 0.75 units of Taq polymerase (Boehringer Mannheim).

Received 11/5/96; accepted 2/14/97.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This study was supported by grants from the Swedish Cancer Society, the Medical Faculty of Lunds University, the Gunnar A. E. Nilsson Cancer Foundation, and the Mrs. Berta Kamprad Foundation (to Å. B.). This work was also supported by NIH Grants CA30195 and CA54174, and United States Army Medical Research Defense Contract Postdoctoral Fellowship DAMD17-94-J-4112 (to S. A. W. F.).

[2] To whom requests for reprints should be addressed, at Department of Medical Oncology, University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78284-7884. Fax: (210) 567-6687; E-mail: suzanne_fuqua@oncology.uthscsa.edu.

[3] The abbreviations used are: ER, estrogen receptor; ERE, estrogen-response element; AF, activation function; SSCP, single-strand conformation polymorphism; CAT, chloramphenicol acetyltransferase; β-gal, β-galactosidase; Tam, 4-hydroxytamoxifen; tk, thymidine kinase; SH-2, src homology.

Table 1 *ER primers used for PCR amplification of DNA*

Name	Sequence	Size	Location
1A	CCACGGACCATGACCATGAC	151 bp	−9, exon 1
1B	TGCTGCTGTCCAGGTACA		142, exon 1
2A	AGCTCAAGATCCCCCTGGA	140 bp	88, exon 1
2B	GGTCTGACCGTAGACCTG		228, exon 1
3A	CAGGTCTACGGTCAGACC	252 bp	211, exon 1
3B	GCGCGGGTACCTGTAGGAA		452 + intron 1
4A	TTTTTTCCCCCCAGGCCAA	224 bp	Intron 1–457, exon 2
4B	TCAACACACTATTACCTTGA		643 + intron 2
5A	CCTCTTGCTTTTAATAGGAC	152 bp	Intron 2–646, exon 3
5B	TGGGAGAGATGTACCTACCA		760 + intron 3
6A	TCTACCTGTGTTTTCAGGGATA	215 bp	Intron 3–765, exon 4
6B	ACAAGGCACTGACCATCTGG		958, exon 4
7A	CTCATGATCAAACGCTCGAAG	138 bp	886, exon 4
7B	CGAAGCTTCACTGAAGGGTC		1023, exon 4
8A	GACCCTTCAGTGAAGCTTCGAT	111 bp	1004, exon 4
8B	TGCGCTTCGCATTCTTACCTGG		1096 + intron 4
9A	CTTGCTTGTTTTCAGGCTTT	167 bp	Intron 4–1101
9B	CCAGGTCACTTACCTGTC		1235 + intron 5
10A	TGCTATGTTTTCATAGGAACC	168 bp	Intron 5–1240
10B	TTGTGTTATCAACTCACCAGA		1369 + intron 6
11A	TCTCTGCGCATTCAGGAGTG	214 bp	Intron 6–1474
11B	CCAGAGATGCCTCACCTCAT		1548 + intron 7
12A	GTCTTCCCACCTACAGTA	128 bp	Intron 7–1555
12B	ACGGCTAGTGGGCGCATGTA		1665, exon 8

DMSO (5%) was also added for primer pair number 3 (Table 1) to facilitate amplification. Negative control reactions (H_2O instead of DNA) were included in every experiment. Amplification was carried out under the following conditions: 94°C for 3 min, followed by 32–42 repetitive cycles of 94°C for 1 min, 52–63°C for 45 s, 72°C for 1 min, and an extension step at 72°C for 5 min. PCR products were first visualized on 7.8% polyacrylamide gels to verify the expected sizes and then stored at 4°C until used for direct sequence analysis.

PCR for SSCP analysis (18) was performed as above with the exception of including 5 μl of diluted (1:50 in dH_2O) [^{32}P]dCTP (10 μCi/μl, 3000 Ci/mmol; Amersham Corp.) and reducing the concentration of unlabeled dCTP to 100 μM. One μl of the PCR product was mixed with 9 μl of denaturing solution (95% deionized formamide, 10 mM NaOH, 0.25% xylene cyanol, and 0.25% bromphenol blue), denatured at 95°C for 3 min, and put on ice for 3 min. Four μl of the mixture were then run on a nondenaturing 0.5× mutation detection enhancement acrylamide gel (AT Biochem) without glycerol at 4°C and 40 W for 4–5 h or with 5% glycerol at room temperature and 15 W overnight. The gel was then transferred onto Whatman filter paper, dried under vacuum at 80°C for 2 h, and put on X-ray film at −70°C overnight or longer. Samples manifesting migration shifts in SSCP gels were analyzed further by DNA sequencing.

DNA Sequencing. DNA sequencing was performed both directly from PCR-amplified genomic DNA and from shifted bands cut from the SSCP gels. For direct sequence analysis, single-stranded DNA template was produced by binding the biotinylated strand from 20–30 μl of PCR product to Dynabead M280-Streptavidin solid supports (Dynal AS, Oslo, Norway). Sequence analysis was then performed, using the bound biotin-labeled strand as template, according to the Sanger dideoxynucleotide chain termination (19) using ^{35}S-labeled dCTP. In addition, 2 μl of the PCR product from samples displaying altered migration bands were cloned into the pCR-TA plasmid (Invitrogen), sequenced, and compared to the published sequence for ER (17, 20).

Expression Vector Construction. A wt ER expression vector was made by subcloning the *Bam*HI-*Eco*RI fragment from the yeast expression vector YEPE 10 (6) into the pcDNAI expression vector (Invitrogen, San Diego, CA). To study the function of the three identified ER mutants, we individually introduced these three substitutions into pcDNAI-wtER plasmid using the Transformer site-directed mutagenesis kit (Clontech, Palo Alto, CA) according to manufacturer's instructions. The following three mutagenic primers were used: Ser47Thr, 5′-CTGGACAGCACCAAGCCCGCC-3′; Lys531Glu, 5′-CAGCATGAAGTGCGAGAA CGTGG-3′; and Tyr537Asn, 5′-GTGCCCCT-CAATGACCTGCTG-3′. The selection primer employed to disrupt the unique *Nhe* site within the plasmid to facilitate cloning of the ER mutations was: 5′-GCAACGCAAGCATGCTTCTAGC-3′. The sequences of the three mutated ER expression plasmids were verified by sequence analysis.

ERE Reporter Construction. We have prepared CAT reporter constructs to the promoters of the four estrogen-inducible genes progesterone receptor (*PgR*), *pS2*, *lactoferrin*, and *cathepsin D*. All the ERE-CAT vectors were derived by modification of the vector pTZ-tk-CAT (21), provided generously by Dr. Benita Katzenellenbogen. This plasmid contains a multiple cloning site upstream of the herpes simplex virus type I tk promoter (bases −150 to +56) controlling the expression of the *CAT* gene. This plasmid was linearized by *Bam*HI digestion and dephosphorylated with calf intestine alkaline phosphatase (Promega, Madison, WI) to prevent recircularization. The ERE-CAT constructs were prepared by ligating this vector in the presence of a 100-fold molar excess of oligonucleotides containing a 13-bp ERE plus 9–11 bp of the requisite flanking gene sequence on either side. Oligonucleotides corresponding to the plus and minus strands of each ERE appended by a sequence generating a *Bam*HI compatible overhang were kinased and annealed prior to addition to the ligation reaction. ERE oligonucleotide sequences (consensus sequences appear in bold and underlined and non-consensus sequences are non-bold; underlining designates two half-ERE sites) were derived from published promoter sequences of the human *lactoferrin* (GATCCCAGTCTCA-CA**GGTCA**A-GG**CGATC**TTCAAGTAG; Ref. 22), *pS2* (GATCCCCTG-CAA**GGTCA**CGG**TGGCC**ACCCCG-TGAGCG; Ref. 23), and *cathepsin D* (GATCCCAGAAGCTG**GGCCG**GGC**TGACC**CCGCGG-GCGG; Ref. 24) genes. Subsequent analysis of the products of these ligation reactions showed that the vectors contained three (pS2) or four (cathepsin D and lactoferrin) inserted ERE sequences. A similar protocol was used to generate the plasmid pERE2-tk-CAT, which contains a tandem insertion of two vitellogenin A2 (consensus) EREs in the *Bam*HI site of pTZ-tk-CAT.

Cell Culture and Transient Transactivation Assays. HeLa cells were maintained in complete MEM supplemented with 10% fetal bovine serum (Sigma Chemical Co., St. Louis, MO). Cells were plated into six-well cluster dishes (Falcon, Franklin Lakes, NJ) at a density of 1.5×10^5 cells/well and grown in the above medium. MDA-MB-231 breast cancer cells (2×10^5) were plated and grown in the same medium supplemented with 2 mM L-glutamine and 6 ng/ml insulin and allowed to recover for 2 days at 37°C in 5% CO_2. Transient transfections were carried out in a total volume of 1 ml of OptiMEM reduced-serum medium with 8 μl of lipofectamine/well according to the manufacturer's instructions (Life Technologies, Inc., Gaithersburg, MD). Transfections of individual wells were performed using 1 μg of ERE-CAT reporter vector. Cytomegalovirus-driven β-gal plasmid (100 ng) was added into each well to monitor transfection efficiency, and 20 ng of either the wt ER or ER-mutant plasmid was transfected into HeLa cells or MDA-MB-231, respectively. The reporter was also transfected into cells without ER as a control for background level of CAT activity. Following an 8-h incubation at 37°C, the transfection medium was removed, and 2 ml of phenol red-free MEM supplemented with charcoal-stripped fetal bovine serum was added and incubated at 37°C for 18–20 h. Cells were then treated with 10^{-11} or 10^{-9} M estradiol, 10^{-7} M Tam, 10^{-7} M ICI 164,384 (a kind gift from Dr. Allen Wakeling of Zeneca, Macclesfield, England), or an ethanol vehicle for 20–22 h at 37°C. All the transfections were performed in duplicate. The transfected cells were then rinsed once with PBS, and 150 μl of cell lysis buffer (Promega) were added. Cell extracts were spun for 5 min, and 20 μl of the supernatant were used in the β-gal assay. The remaining supernatant was heated to 65°C for 15 min to remove any endogenous acetylases or nonlabeled acetyl-CoA and centrifuged again. Various volumes of the extracts, calculated to contain equivalent amounts of β-gal activity, were incubated with substrate mix [1 ml of 1 M Tris (pH 7.8), 80 μl of ^{14}C-labeled chloramphenicol, 800 μl of 4 mM acetyl-CoA, and 4.52 ml of dH_2O] at 37°C for 90 min (HeLa cells) or 120 min (MDA-MB-231 cells). The reaction was then terminated by adding 1 ml of cold ethylacetate vigorously mixed and centrifuged. The supernatant was dried under vacuum, and the residue was resuspended in 20 μl of ethyl acetate, spotted on precast TLC plates (Sigma Chemical Co.), chromatographed in 95% chloroform and 5% methanol at room temperature for 30–40 min, dried, and exposed to Biomax X-ray film (Kodak, Rochester, NY) overnight. All the assays were done in duplicate.

Results

ER Mutations Identified in Metastatic Breast Tumors. Thirty metastatic breast cancer samples were screened for mutations in the *ER* gene by SSCP analysis. SSCP shifts (evident by electrophoresis both at ambient and subambient temperatures) were identified in six of the metastatic tumors. The six positive samples were studied further

both by direct sequencing analysis and by cloning of shifted bands followed by sequencing analysis. Three of the SSCP-shifted bands were found to constitute previously identified polymorphisms (TCT to TCC at nucleotide 30 in exon 1, TGC to TGT at nucleotide 720 in exon 3, and CGC to CGT at nucleotide 729 in exon 3), none of which results in an amino acid substitution (data not shown). The remaining three shifted SSCP bands (Fig. 1*A*, *arrows*) were found to constitute missense mutations: G140C (Ser47Thr) in exon 1, A1591G (Lys531Glu) in exon 8, and T1609A (Tyr537Asn) in exon 8 (Fig. 1*B*). Whereas the first two sequence changes exchange hydrophilic amino acids, the last mutation changes a hydrophobic tyrosine to a hydrophilic asparagine.

Functional Studies. The functional transactivational status of the three missense ER mutations isolated from the metastatic breast tumors was investigated using transient transactivation assays, measuring the transcriptional activity of the mutant ERs with ERE-reporter gene constructs. Although the majority of studies examining the effect of specific alterations in the ER on function have utilized consensus EREs, most commonly the vitellogenin A2 ERE, we felt that it would be important to test ER function on constructs that might be more relevant to breast cancer biology, *e.g.*, those genes that are regulated endogenously by ER in breast cancer cells. We therefore prepared ERE-CAT reporters from the estrogen-regulated *pS2, cathepsin D,* and *lactoferrin* gene promoters. Each mutant ER was then cotransfected into HeLa cells (Fig. 2) with the different ERE reporter vectors and compared to wt ER activity. CAT activity was determined relative to the activity of the transfected reporter vector alone; activity was also corrected for transfection efficiency by cotransfection of a β-gal vector.

The activities of Ser47Thr and Lys531Glu were not different from that of wt ER using any of the four different ERE constructs in either HeLa or MDA-MB-231 cells (data not shown). However, the Tyr537Asn mutant exhibited strong constitutive transactivation activity (15–20-fold over wt ER activity on the vitellogenin ERE) in the absence of hormone (Fig. 2, *Vitellogenin, con*). This elevated constitutive activity was also observed on the other ERE constructs (5-fold on the pS2 control [*con*], 8-fold on the cathepsin D control, and 17-fold on the lactoferrin ERE control). Estradiol was required for the induction of wt ER activity on all four of the ERE reporter constructs (maximum inductions were 11-, 2-, 3-, and 4-fold on the vitellogenin, pS2, cathepsin D, and lactoferrin reporters, respectively, using 10^{-9} M estradiol; Fig. 2). As expected, tamoxifen alone had no effect on basal activity of the wt ER and completely inhibited the stimulatory effect of estradiol (Fig. 2, *E2+Tam*).

In contrast, the addition of estradiol had only minimal influence on the already high constitutive transcriptional activity of the Tyr537Asn mutant [Fig. 2; compare the control (*con*) with the estradiol-stimulated levels (*E2-9* and *E2-11*)]. Interestingly, tamoxifen appeared to slightly inhibit to varying degrees the basal activity of the Tyr537Asp mutant on three of the ERE promoters; for instance, 45% inhibition of basal activity was seen using the cathepsin D ERE reporter. The only notable difference among the four ERE reporters was in the insignificant tamoxifen inhibition of basal transcriptional activity on the lactoferrin promoter (Fig. 2, compare control with Tam levels). Similar results to those seen with tamoxifen were observed when these experiments were repeated using the pure antiestrogen ICI 164,384 in place of tamoxifen in these cells (data not shown).

Because we know that the transcriptional activity of the ER is highly dependent on the cell and the promoter context (25) in which the receptor is expressed, we also tested the transcriptional activity of the Tyr537Asn ER in MDA-MB-231 breast cancer cells. Similar results were seen in these cells, although the maximum inductions by wt ER (2–3-fold) over control in the absence of estradiol were all reduced due to the high basal activity of the tk-CAT reporter alone in these cells (compare the controls in Fig. 3 to the controls in Fig. 2). Again, the Try537Asn mutant displayed high constitutive activity on all four of the ERE reporters, ranging from 3- to 7-fold depending on the ERE reporter, and this activity was essentially unaffected by estrogen, tamoxifen (Fig. 3), and the ICI pure antiestrogen (data not shown).

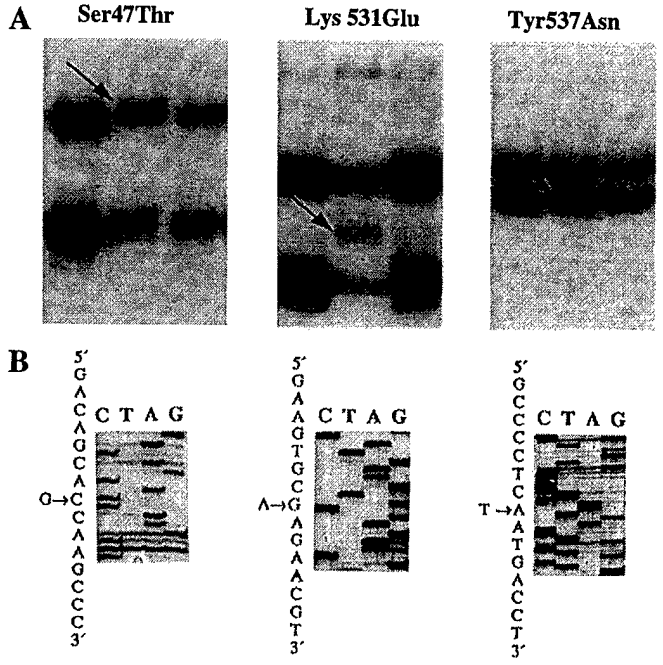

Fig. 1. Identification of three ER missense mutations in human breast cancer. *A*, SSCP analyses of normal (*outside lanes*) and tumor (*middle lane*) DNA revealed three shifted bands (*arrows*). The shifted bands were cut from the SSCP gel, sequenced (*B*), and represent missense mutations: G140C (*Ser47Thr*) in exon 1, A1591G (*Lys531Glu*) in exon 8, and T1609A (*Tyr537Asn*) in exon 8.

Discussion

Although ER-splicing variants have been shown to be ubiquitous in human breast cancer (12), the number of naturally occurring missense mutations identified in primary breast cancers to date is extremely low. It has been estimated that missense mutations are present in only about 1% (2 of 188) of primary tumors (15). In agreement with this, we have not detected any missense ER alterations in 60 primary breast cancers that we have examined using SSCP analysis.[4] Karnik *et al.* (14) have also recently examined five primary and metastatic breast tumor pairs for ER sequence alterations, again using SSCP analysis. They found that one of the five metastatic lesions, but not the primary tumor from the same patient, contained a single nucleotide deletion (1294delT) in the coding region of the ER. This deletion generates a frameshift in the hormone-binding domain of the receptor and is predicted to give rise to a premature translation termination with an ER protein maintaining an intact DNA-binding domain but with a defective ligand-binding domain. Functional analysis of this ER mutant, however, has not yet been reported. The fact that this ER mutation was present in the metastatic lesion but not in the corresponding primary tumor suggests that some ER mutations may be associated with tumor progression.

In the present study of 30 cases of metastatic breast cancer, we detected three missense ER mutations. Thus, ER mutations in meta-

[4] Q-X. Zhang, Å. Borg, and S. A. W. Fuqua, unpublished data.

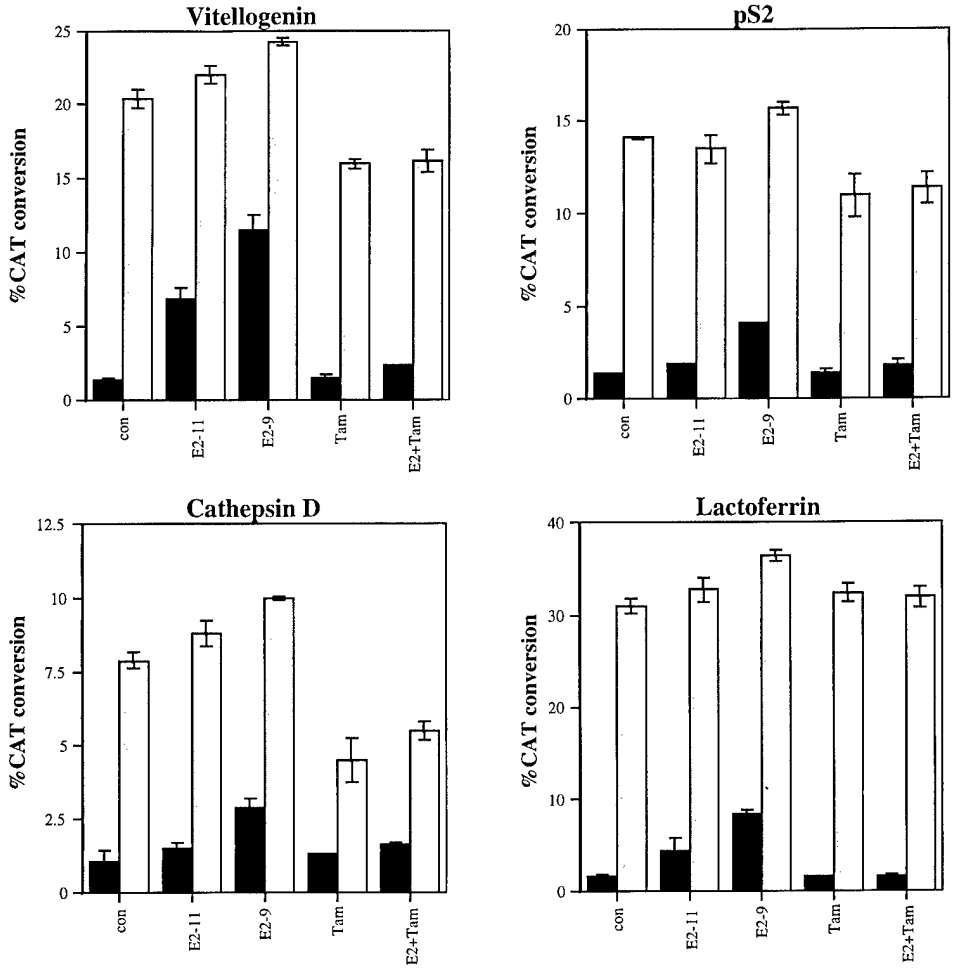

Fig. 2. The Tyr537Asn ER mutant displays strong hormone-independent transcriptional activity in HeLa cells. Transactivation assay comparing the Tyr537Asn ER mutant (□) with wt ER (■) in the absence and presence of estradiol (*con*, control group; *E2-11*, 10^{-11} M; *E2-9*, 10^{-9} M), Tam (*Tam*, 10^{-7} M), and a combination of both (*E2+ Tam*, 10^{-9} M estradiol and 10^{-7} M Tam). The results from four different ERE constructs are shown in separate panels. Data are shown as percentage of CAT conversion (corrected for β-gal activity) from duplicate wells; *bars*, SD.

static breast tumors may be more frequent than in primary lesions [3 of 30 as compared to 2 of 248 (see above); $P = 0.004$ with Fisher's exact test). Unfortunately, the corresponding primary tumors from these three patients were not available for ER mutational analysis, and future studies will be directed at addressing whether ER mutations arise during metastatic spread of the tumor in certain patients. Two of the identified ER mutations (Ser47Thr and Lys531Glu) did not alter ER transcriptional activity in transient transfection assays. The third mutation, however, resulted in an ER protein with a very high constitutive transcriptional activity. This mutation (Tyr537Asn) is located at the beginning of exon 8 of the *ER* gene encoding the COOH-terminal portion of the hormone-binding domain, a region of the ER to which dimerization and AF-2 functions have been ascribed (2). This is also a region that is evolutionarily conserved between species but is divergent from other members of the steroid and thyroid hormone receptor superfamily (17).

The Tyr537Asn mutation eliminates a tyrosine residue that is a potential phosphorylation site within the ER. We found that the Tyr537Asn ER mutant manifested strong transactivation activity in both HeLa and breast cancer cells in the absence of hormone binding and that this activity was only marginally affected by estradiol, tamoxifen, or the pure steroidal antiestrogen ICI 164,384. Furthermore, this constitutive activity was similarly evident whether using an idealized ERE (vitellogenin) or endogenous EREs such as the *cathepsin D, pS2,* and *lactoferrin* gene promoters. Several potential mechanisms could explain the high constitutive activity of the Tyr537Asp ER mutation. One explanation, and one that we favor (Fig. 4), is that the Tyr537Asn substitution may produce a conformational change in the receptor that mimics hormone binding. We know that this residue lies within the hormone-binding domain of the ER; functional analysis of ER deletion mutants has suggested that the COOH-terminal boundary of both the estrogen and tamoxifen-binding domains are similar, lying between residues 522 and 538 (26, 27). The role of ligand binding in the formation of AF-2 is unknown but is believed to involve conformational changes in the receptor that generate a productive association between the AF-1 and AF-2 domains (28). In our model (Fig. 4), we envision that similar conformational changes may be induced by the Tyr537Asn substitution. As a result of this conformational change, the mutant ER might only weakly bind estrogen and tamoxifen, explaining their limited effects on mutant ER activity. Although the metastatic bone lesion from which we isolated the Tyr537Asn mutant was indeed ER negative by ligand-binding analysis, this result may not be conclusive, because low tumor cellularity is often associated with bone lesions. Unfortunately, additional tumor material was not available to address the question of hormone binding ability of this variant using immunohistochemical techniques.

There is evidence to suggest that phosphorylation at this site is required for efficient estrogen binding. Arnold *et al.* (29) determined that Tyr 537 is a physiological phosphorylation site in ER isolated from MCF-7 human breast cancer cells. Using site-directed mutagenesis to separately replace all five tyrosine residues within the hormone-binding domain of the ER with phenylalanine, Castoria *et al.* (30) demonstrated that phosphorylation of *in vitro* synthesized ER at Tyr537 confers efficient estrogen binding ability. Thus, phosphoryl-

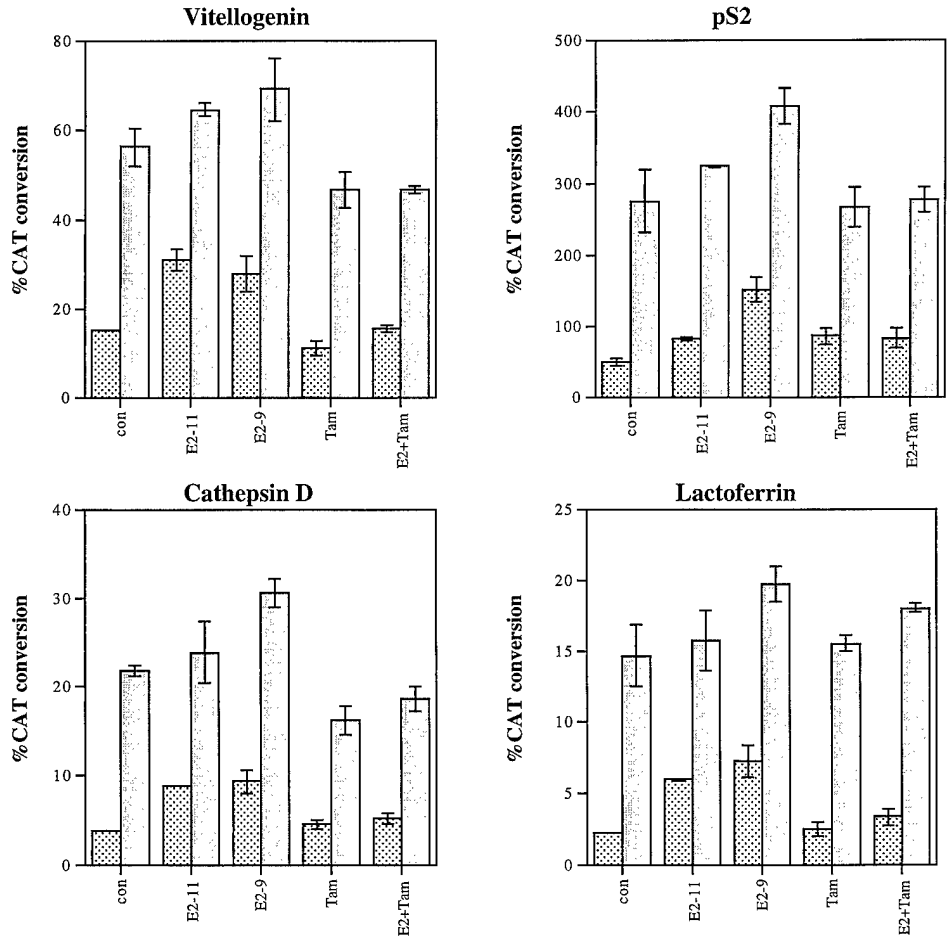

Fig. 3. The Tyr537Asn ER mutant displays strong hormone-independent transcriptional activity in MDA-MB-231 cells. Transactivation assay comparing the Tyr537Asn ER mutant (□) with wt ER (▨) in the absence and presence of estradiol (*E2–11*, 10^{-11} M; *E2–9*, 10^{-9} M), Tam (10^{-7} M), and a combination of both (*E2+Tam*, 10^{-9} M estradiol and 10^{-7} M Tam). The results from four different ERE constructs are shown in separate panels. Data are shown as percentage of CAT conversion (corrected for β-gal activity) from duplicate wells; *bars*, SD.

ation of Tyr537 may induce conformational changes that are necessary for ligand binding.

It has also been suggested that phosphorylation at Tyr537 is a necessary step for ER dimerization (31). These authors propose that ER dimerization occurs through specific interactions between phosphotyrosine residues and SH-2-like domains, similar to the activation of the STAT family of transcription factors (32). These data would therefore suggest that the Tyr537Asn ER mutant would be incapable of dimerization due to the absence of a target for the SH-2 like domain of its dimerization partner. However, as ER dimerization is necessary for ERE binding and transcriptional activity, this hypothesis is clearly inconsistent with our demonstration of strong transcriptional activity with the Tyr537Asn mutant and implies that other regions of the hormone-binding domain, such as the adjacent leucine zipper motif (32), are involved in dimerization as well.

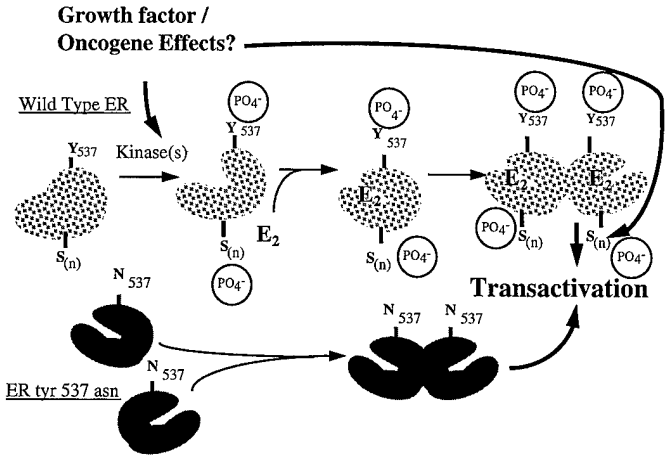

Fig. 4. Model of transactivation of wt and Try537Asp ERs. The activity of wt ER is influenced by the growth factor and oncogene signaling pathways via kinase activities. Specific residues within the ER are phosphorylated, such as several serines within the AF-1 domain and tyrosine 537 within the AF-2 domain. Phosphorylation at Tyr537 allows efficient estrogen (*E2*) binding, the ER undergoes a conformation change, and dimerization is affected. We propose that the ER tyr537asn mutation exerts a hormone-independent conformational change in the AF-2 domain, leading to dimerization and transactivation in the absence of estradiol.

Tyr537 may represent a basal phosphorylation site of the human ER, which is under strict control by both specific tyrosine kinases and phosphatases. This was suggested by studies demonstrating that two members of the *src* family of tyrosine kinases were capable of Tyr537 phosphorylation, as well as by the finding that protein tyrosine phosphatase-1B and the SH-2 protein tyrosine phosphatase-1 dephosphorylated Tyr537 (6). Thus, the Tyr537Asn ER mutant may have escaped from phosphorylation-mediated transcriptional regulation that is present *in vivo*, as was seen in the HeLa and MDA-MB-231 cells used in the present transfection studies. As reflected in the model diagrammed in Fig. 4, the activity of the ER can be modulated by the phosphorylation of a number of residues through growth factor and oncogene signaling pathways. There are multiple sites of phosphorylation in the ER; several serine residues in the NH_2-terminal AF-1 domain are targeted through the Ras mitogen-activated protein kinase (5) and the protein kinase A or C signaling pathways (33). Recently, it was demonstrated that the *HER-2* oncogene targets the ER leading to phosphorylation of the ER on tyrosine residues and ligand-inde-

pendent signaling through the receptor, resulting in ligand-independent signaling through the ER (34). Disruption or dysregulation of phosphorylation at specific sites within the ER may therefore be important in the clinical problem of hormone-independent tumor growth, as would be predicted for patients harboring the Tyr537Asn mutation detected in this study. It is of note that the patients from which this mutation was identified presented with advanced metastatic disease. Future studies will be directed at investigating the frequency of this specific alteration in patients with metastatic breast cancer to determine whether this constitutive mutation is common in patients with dissemination of their disease.

References

1. Evans, R. M. The steroid and thyroid hormone receptor superfamily. Science (Washington DC), *240:* 889–895, 1988.
2. Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. R., and Chambon, P. Functional domains of the human estrogen receptor. Cell, *51:* 941–951, 1987.
3. Berry, M., Metzger, D., and Chambon, P. Role of the two activating domains of the oestrogen receptor in the cell-type and promoter-context dependent agonistic activity of the anti-oestrogen 4-hydroxytamoxifen. EMBO J., *9:* 2811–2818, 1990.
4. Halachmi, S., Marden, E., Martin, G., MacKay, H., Abbondanza, C., and Brown, M. Estrogen receptor-associated proteins: possible mediators of hormone-induced transcription. Science (Washington DC), *264:* 1455–1458, 1994.
5. Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, Y., Uchiyama, S., Sasaki, H., Masushige, S., Gotoh, Y., Nishida, E., Kawashima, H., Metzger, D., and Chambon, P. Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. Science (Washington DC), *270:* 1491–1494, 1995.
6. Arnold, S. F., Obourn, J. D., Jaffe, H., and Notides, A. C. Phosphorylation of the human estrogen receptor on tyrosine 537 *in vivo* and by Src family tyrosine kinases *in vitro*. Mol. Endocrinol., *9:* 24–33, 1995.
7. Clark, G. M., and McGuire, W. L. Steroid receptors and other prognostic factors in primary breast cancer. Semin. Oncol., *15:* 20–25, 1988.
8. Fuqua, S. A. W., Chamness, G. C., and McGuire, W. L. Estrogen receptor mutations in breast cancer. J. Cell. Biochem., *51:* 135–139, 1993.
9. Fuqua, S. A. W., Fitzgerald, S. D., Allred, D. C., Elledge, R. M., Nawaz, Z., McDonnell, D. P., O'Malley, B. W., Greene, G. L., and McGuire, W. L. Inhibition of estrogen receptor action by a naturally occurring variant in human breast tumors. Cancer Res., *52:* 483–486, 1992.
10. Fuqua, S. A. W., Fitzgerald, S. D., Chamness, G. C., Tandon, A. K., McDonnell, D. P., Nawaz, Z., O'Malley, B. W., and McGuire, W. L. Variant human breast tumor estrogen receptor with constitutive transcriptional activity. Cancer Res., *51:* 105–109, 1991.
11. Zhang, Q-X, Borg, Å., and Fuqua, S. A. W. An exon 5 deletion variant of the estrogen receptor frequently coexpressed with wild-type estrogen receptor in breast cancer. Cancer Res., *53:* 5882–5884, 1993.
12. Zhang, Q-X., Hilsenbeck, S., Fuqua, S. A. W., and Borg, Å. Multiple splicing variants of the estrogen receptor are present in individual breast tumors. J. Steroid Biochem. Mol. Biol., *59:* 251–260, 1996.
13. Garcia, T., Sanchez, M., Cox, J. L., Shaw, P. A., Ross, J. B. A., Lehrer, S., and Schachter, B. Identification of a variant from of the human estrogen receptor with an amino acid replacement. Nucleic Acids Res., *17:* 8364, 1989.
14. Karnik, P. S., Kullkarni, S., Liu, X-P., Budd, G. T., and Bukowski, R. M. Estrogen receptor mutations in tamoxifen-resistant breast cancer. Cancer Res., *54:* 349–353, 1994.
15. Roodi, N., Bailey, L. R., Kao, W-Y., Verrier, C. S., Yee, C. J., Dupont, W. D., and Parl, F. F. *Estrogen receptor* gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. J. Natl. Cancer Inst., *87:* 446–451, 1995.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual, Ed. 2, pp. 9.16–9.23. New York: Cold Spring Harbor Laboratory Press, 1989.
17. Ponglikitmongkol, M., Green, S., and Chambon, P. Genomic organization of the human *oestrogen receptor* gene. EMBO J., *7:* 3385–3388, 1988.
18. Orita, M., Iwahara, H., Kanazawa, H., Hayashis, K., and Sekiya, T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformational polymorphosis. Proc. Natl. Acad. Sci. USA, *86:* 2766–2770, 1989.
19. Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, *74:* 5463–5467, 1977.
20. Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., and Shine, J. Sequence and expression of human estrogen receptor complementary DNA. Science (Washington DC), *231:* 1150–1154, 1986.
21. Kraus, W. L., Montano, M. M., and Katzenellenbogen, B. S. Cloning of the rat *progesterone receptor* gene 5′-region and identification of two functionally distinct promoters. Mol. Endocrinol., *7:* 1603–1616, 1993.
22. Teng, C. T., Liu, Y., Yang, N., Walmer, D., and Panella, T. Differential molecular mechanisms of the estrogen action that regulates *lactoferrin* gene in human and mouse. Mol. Endocrinol., *6:* 1969–1981, 1992.
23. Nunez, A-M., Berry, M., Imler, J-L., and Chambon, P. The 5′ flanking region of the *pS2* gene contains a complex enhancer region responsive to oestrogens, epidermal growth factor, a tumour promoter (TPA), the c-Ha-ras oncoprotein and the c-jun protein. EMBO J., *8:* 823–829, 1989.
24. Augereau, P., Miralles, F., Cavailles, V., Gaudelet, C., Parker, M., and Rochefort, H. Characterization of the proximal estrogen-responsive element of human *cathepsin D* gene. Mol. Endocrinol., *8:* 693–703, 1994.
25. Tzukerman, M. T., Esty, A., Santiso-Mere, D., Danielian, P., Parker, M. G., Stein, R. B., Pike, J. W., and McDonnell, D. P. Human estrogen receptor transactivational capacity is determined by both cellular and promoter context and mediated by two functionally distinct intramolecular regions. Mol. Endocrinol., *8:* 21–30, 1994.
26. Lees, J. A., Fawell, S. E., and Parker, M. G. Identification of two transactivation domains in the mouse oestrogen receptor. Nucleic Acids Res., *17:* 5477–5488, 1989.
27. Fawell, S. E., Lees, J. A., and Parker, M. G. A proposed consensus steroid-binding sequence: a reply. Mol. Endocrinol., *3:* 1002–1005, 1989.
28. Kraus, W. L., McInerney, E. M., and Katzenellenbogen, B. S. Ligand-dependent, transcriptionally productive association of the amino- and carboxyl-terminal regions of a steroid hormone nuclear receptor. Proc. Natl. Acad. Sci. USA, *92:* 12314–12318, 1995.
29. Arnold, S. F., Obourn, J. D., Jaffe, H., and Notides, A. C. Phosphorylation of the human estrogen receptor on tyrosine 537 *in vivo* and by src family tyrosine kinases *in vitro*. Mol. Endocrinol., *9:* 24–33, 1995.
30. Castoria, G., Migliaccio, A., Green, S., Di Domenico, M., Chambon, P., and Auricchio, F. Properties of a purified estradiol-dependent calf uterus tyrosine kinase. Biochemistry, *32:* 1740–1750, 1993.
31. Arnold, S. F., and Notides, A. C. An antiestrogen: a phosphotyrosyl peptide that blocks dimerization of the human estrogen receptor. Proc. Natl. Acad. Sci. USA, *92:* 7475–7479, 1995.
32. Fawell, S. E., Lees, J. A., White, R., and Parker, M. G. Characterization and colocalization of steroid binding and dimerization activities in the mouse estrogen receptor. Cell, *60:* 953–962, 1990.
33. Le Goff, P., Montano, M. M., Schodin, D. J., and Katzenellenbogen, B. S. Phorphorylation of the human estrogen receptor: identification of hormone-regulated sites and examination of their influence on transcriptional activity. J. Biol. Chem., *269:* 4458–4466, 1994.
34. Pietras, R. J., Arboleda, J., Reese, D. M., Wongvipat, N., Pegram, M. D., Ramos, L., Gorman, C. M., Parker, M. G., Sliwkowski, M. X., and Slamon, D. J. HER-2 tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth in human breast cancer cells. Oncogene, *10:* 2435–2446, 1995.

CHAPTER 14

APPLICATION OF DIFFERENTIAL DISPLAY TO CANCER RESEARCH: AMPLIFICATION OF RNA ISOLATED FROM AIR-DRIED FROZEN AND ARCHIVAL PARAFFIN-EMBEDDED TISSUES

SUZANNE A. W. FUQUA, MARGARET G. BENEDIX, DOUGLAS M. WOLF, CARL G. CASTLES, and D. CRAIG ALLRED

INTRODUCTION

Accumulating molecular evidence in the field of cancer research has demonstrated that the processes contributing to carcinogenesis are largely controlled by the activities of two opposing types of gene products (Cooper, 1995). Oncogenes are genes which produce proteins that have normal cellular functions regulating growth and passage through the cell cycle. However, when expressed inappropriately or when mutated to become more active stimulators of cell growth, oncogene products contribute to the conversion of a normal cell to a cancerous cell. Similarly, tumor suppressor gene (or antioncogene) products are proteins that serve to downregulate cell growth, and if these genes are mutated or lost, the effect is similar to that of removing a brake, allowing a cell to grow and replicate in an uncontrolled fashion. One of the major goals of cancer researchers has been the identification of oncogenes and tumor suppressor genes by a variety of techniques, both to provide an understanding of the mechanisms of the carcinogenic process and to conceivably provide targets for new therapeutic interventions.

The development of the differential display polymerase chain reaction (DD-PCR) method (Liang and Pardee, 1992) has added an important tool to the list of techniques available for the identification of potential oncogene and tumor suppressor gene products involved in carcinogenesis. Products detected in cancer cells but not corresponding normal tissues may be oncogenes, whereas products specifically expressed by normal tissues but not cancer cells are good candidates for tumor suppressor genes. Careful application of differ-

A Laboratory Guide to RNA: Isolation, Analysis, and Synthesis, Edited by Paul A. Krieg.
ISBN 0-471-12536-9

ential display, coupled with appropriate methods for confirmation of the results of DD-PCR analysis, allows the researcher to screen RNA prepared from diverse sources to identify these genes. RNA for DD-PCR analysis can be obtained from sources as disparate as cancer-derived cell lines, tumor biopsy specimens, or archival blocks of paraffin-embedded specimens.

There are several critical items to be considered for successful application of DD-PCR to cancer-derived materials. Steps must first be taken to enrich target cells as much as possible. This can be done by working with tumor cell lines, which are homogeneous populations of cancer-derived cells, so that differences in cDNAs are attributable to differences in specific manipulations. However, differential cDNA expression can be simply due to differences in cell confluency, accumulation of inhibitory/stimulatory factors in the culture media, or other factors. Thus, even DD-PCR results from a relatively simple system, such as cultured cells, must be interpreted with caution and confirmed with alternative techniques.

When working with animal tissues such as human tumors, the difficulties associated with interpreting differential display results increase dramatically. First, tissue for analysis may be limited since the majority of it may be needed for clinical diagnostic evaluation. However, this problem can be partially circumvented by methods that make use of archival material. Second, all DD-PCR results must be viewed with the caveat that tumor samples are not a homogeneous population of tumor cells, but rather an admixture of cancerous and non-cancerous cells such as normal epithelium, fibroblasts, endothelial cells due to neovascularization, as well as infiltrating inflammatory cells. Therefore, investigators run the risk of having to perform secondary screens on many apparently differentially expressed cDNAs, where in reality, the differences in gene expression are due to varying amounts of contaminating cell types in the original sample. Ideally, isolation of RNA for differential display should be performed only after the investigator has taken steps to isolate the cells of interest (in this case tumor cells) from surrounding noncancerous elements. One method that has proven successful in addressing both of these concerns is microdissection of cancer cells from histological frozen sections that have been air-dried after sectioning to facilitate histiological analysis and microdissection, or thick sections cut from archival paraffin-embedded tissue blocks. These methods form the core of the procedures discussed below.

Finally, it cannot be overemphasized that, in all cases, some additional method should be used to confirm the differences in cDNA expression seen with differential display. Ideally, analysis of expression levels of newly identified gene fragments should be accomplished by probing Northern mRNA blots with a probe prepared by reamplification of the isolated DD band in the presence of a ^{32}P-labeled nucleotide (or nucleotides). Unfortunately, there may be insufficient material available from frozen biopsy specimens to do multiple Northern blots. However, confirmation of DD-PCR bands in clinical materials is feasible with RT-PCR of RNA prepared from frozen sections, or RT-PCR coupled with primer-extension preamplification (PEP) of RNA prepared from paraffin-embedded samples (Zhang et al., 1992), and methods for these are described below.

MATERIALS FOR BASIC METHODS

Preparation of Air-Dried Frozen Histological Sections

Glass microscope slides, 25 × 75 mm, uncoated
Tissue-Tek cryomolds (Miles Inc., Elkhart, IN)
Tissue-Tek OCT compound (Miles Inc.)

Extraction of RNA from Histological Sections

Xylene
80% ethanol
Guanidinium resuspension buffer
Phenol:chloroform (70:30)
Mussel glycogen (5 Prime → 3 Prime, Inc. cat. #5306-851159)
Isopropanol
RQ1-DNase (Promega)
100% ethanol
Phenol/chloroform/isoamyl alcohol
Chloroform/isoamyl alcohol
DEPC treated H_2O
Nutator (Adams Co.)

Differential Display PCR (DD-PCR)

dNTP stock mix
10× RT buffer (↑ Mg)
10× PCR buffer (↓ Mg)
AMV (or M-MLV) reverse transcriptase[a]
$T_{11}XX$ primer at 1 mM
random sequence 5′ 10-mer oligonucleotide primer
^{35}S-dATP, 3000 Ci/mmol, 12.5 mCi/mL (DuPont-NEN cat: #NEG-034H)
Taq polymerase
Mineral oil
10× TBE
50% Long Ranger™ gel solution (AT Biochem)
Formamide loading buffer
pBR-322/*Msp*I radiolabeled marker
Thermal cycler
Sequencing gel apparatus and power supply

RT-PCR Amplification

dNTP stock mix
10× ↑ Mg PCR buffer
5′ and 3′ gene-specific 20-mer oligonucleotide primers
AMV or M-MLV reverse transcriptase[a]
5% acrylamide/TBE minigel

Taq polymerase
mineral oil
Thermal cycler

Primer-Extension Preamplification (PEP)

10× PEP buffer
dNTP stock mix
Random 15-mer primer mix
Pair of gene-specific primers that flank the target sequence (outer 5′ and 3′ primers)
Pair of nested gene specific internal primers (nested 5′ and 3′ primers)
AMV (or M-MLV) reverse transcriptase[a]
Taq polymerase
5% acrylamide/TBE minigel
Mineral oil
Thermal cycler

RECIPES

Guanidinium Resuspension Buffer

6 mg/mL proteinase K
1 M guanidinium thiocyanate
25 mM β-mercaptoethanol
0.5% *N*-lauryl-sarcosine
20 mM Tris-HCL pH 7.5

dNTP stock mix (25 mM each dNTP)

20 μL 0.1M dCTP
20 μL 0.1M dGTP
20 μL 0.1M dATP
20 μL 0.1M dTTP

10× PEP Buffer

100 mM Tris, pH 8.3
500 mM KCl
30 mM $MgCl_2$
10 mM spermidine

10× ↑ Mg PCR Buffer (also 10× RT Buffer)

100 mM Tris, pH 8.3
500 mM KCl
30 mM $MgCl_2$

10× PCR Buffer (↓ Mg)

100 mM Tris, pH 8.3
500 mM KCl
10 mM $MgCl_2$

10× TBE

0.89 M Tris base
0.89 M boric acid
0.02 M EDTA

Formamide Loading Buffer

4 mL formamide
500 μL 10× TBE
0.025 g bromophenol blue
0.025 g xylene cyanol
H_2O to 5.0 mL

STEPS FOR BASIC METHODS

Preparation of Air-Dried Frozen Histological Sections

1. Sample tissue is snap-frozen and immediately adhered to a cryostat chuck.[b]
2. Cut sections at 20 μm and thaw-mount onto glass slides.
3. Desiccate slides in a 37°C incubator for 5 min, then microdissect immediately.[c] Dried slides may be stored in an air-tight container (along with a desiccant) at < -70°C.

Extraction of RNA from Mounted Samples. RNA is extracted from histological sections using methods essentially as described by Stanta and Schneider (1991).

1. Carefully transfer tissue areas microdissected from histological sections into 1.5-mL microcentrifuge tubes.[d] Paraffin-embedded samples can be deparaffinized by adding 1.0 mL of xylene and mixing on a Nutator of 20 mins.
2. Pellet the samples in a microcentrifuge for 5 mins at room temperature, rinse the pellets with 0.5 mL of 80% ethanol (cold), and air-dry for 10 min.
3. Resuspend the dried pellets in 200 μL guanidinium resuspension buffer (6 mg/mL proteinase K, 1 M guanidinium thiocyanate, 25 mM β-mercaptoethanol, 0.5% *N*-lauryl-sarcosine, 20 mM Tris-HCl pH 7.5).[e]
4. Then incubate samples at 45°C for 6 h, followed by a 7-min incubation at 100°C to inactivate the proteinase K.
5. Extract with 200 μL of phenol:chloroform (70:30), mix, then microcentrifuge for 3 min at room temperature, and transfer the aqueous phase to a fresh tube.
6. Add 2 μg of mussel glycogen, an equal volume of room-temperature isopropanol, and precipitate the samples for 1 h at −20°C.[f]

7. Pellet the RNA in a microcentrifuge for 10 min at 4°C, wash the pellet with 500 μL of 80% ethanol (cold), and air-dry the pellet.
8. Resuspended the RNA in 100 μL of DEPC-treated H_2O.
9. Add 1U RQ1 DNase (Promega) to 100 μL of RNA from step 8. Incubate at 37°C for 15 min.
10. Extract once with phenol/chloroform/isoamyl alcohol and once with chloroform/isoamyl alcohol.
11. Add 0.1 vol of 3 M NaOAc and 2.5 vol of 100% ethanol. Precipitate for 15–30 min on ice or overnight at −20°C.
12. Repeat step 8.
13. Resuspend the RNA in either 10μL of DEPC-treated H_2O (for RT-PCR), or 100 μL of 1× ↑ Mg-PCR amplification buffer (for DD-PCR).

Differential-Display PCR (DD-PCR)

1. Prepare reverse transcription (RT) master mix, say, for 20 reactions:

Add	Final Concentration
100 μL 10× RT (↑ Mg) buffer	1×
5 μL 2.5 mM dNTP mix[g]	12.5 μM
1 μL 1 mM T_{11}XX primer	1 μM
H_2O to 940 μL	

2. Mix 47 μL of master mix with 1–2 μL (0.2–1.0 μg) of RNA solution from step 13.[h]
3. Heat samples at 65°C for 5 min.
   Cool samples to 40°C
   Add 5 units of AMV reverse transcriptase.[i]
   Incubate at 40°C for 60 min.
4. Heat samples to 94°C for 3–5 min to inactivate the RT. Store on ice until ready to perform the DD-PCR, below.
5. Prepare DD-PCR master mix, say, for 20 reactions:

Add	Final Concentration
36 μL 10× RT (↓ Mg) buffer	1×
2 μL 0.5 mM dNTP mix[j]	2.5 μM
1 μL 1 mM T_{11}XX primer	2.5 μM
0.5 μL 1 mM random 10-mer 5′ primer[k]	1.25 μM
15 μL ^{35}S-dATP (3000 Ci/mmol, 12.5 mCi/mL)	~0.5 mCi/mL
H_2O to 350 μL	

6. For each sample, combine in a PCR reaction tube:
   17.5 μL DD-PCR master mix

2 μL RT reaction product
0.5 μL (2.5 units) Taq DNA polymerase

7. Overlay tubes with 50 μL mineral oil.
8. Heat reactions to 94°C for 90 s to ensure complete denaturation.
9. Do 40 cycles of PCR at 94°C for 30 s, 42°C for 90 s, and 72°C for 30 s. Do a single final extension at 72°C for 5 min. To inactivate the remaining Taq enzyme, heat for 10 min at 98°C, and bring samples to 27°C. Samples should be stored at 4°C until ready to load onto the gel.
10. Prepare 5% Long Ranger™/1.0× TBE sequencing gel.
11. Remove PCR samples to fresh tubes[l] and add 5 μL formamide loading buffer.
12. Heat samples and pBR322/*Msp*I radiolabeled marker at 80–85°C for 3 min.
13. Load marker and 10 μL of each sample into 1.0-cm-wide wells on 5% Long Ranger™/1.0× TBE gel.
14. Run gel at 70 W constant power until the xylene cyanol (upper) marker reaches the bottom of the gel.[m]
15. Dry gel onto Whatman 3MM paper under vacuum at 80°C. Expose to film for 16–24 h at room temperature.
16. Bands of interest may be excised and reamplified as described by Liang and Pardee (1992) (see also Chapter 13, this volume).

In the next sections we describe methods to confirm differential expression of the sequences identified using the differential-display procedure. It will be necessary to obtain at least partial DNA sequence of the specific differential display product because the methods described below require sequence specific PCR primers.

DIFFERENTIAL-DISPLAY PROTOCOLS

RT-PCR Amplification of RNA Isolated from Frozen Specimens for Confirmation of Expression of cDNAs Identified by DD-PCR

RT-PCR is performed essentially as described by Fuqua et al. (1990):

1. Set up RT-PCR reactions:
   5–10 μL RNA from extraction procedure (≥ 200 ng RNA)
   10 μL 10× ↑ Mg PCR buffer
   3.2 μL dNTP stock mix
   1 μL each 0.1 nmol of 5′ and 3′ gene-specific 20-mer oligonucleotide primers
   H_2O to 98.5 μL

2. Heat to 94°C for 2 min. Cool to 42°C and add 2 units of AMV reverse transcriptase.
3. Heat to 42°C for 45 min. Heat to 94°C for 2 min.
4. Add 2.5 units of Taq polymerase and overlay with 100 μL of mineral oil.
5. Do 35 cycles of PCR at 94°C for 1 min, 55°C for 2 min, 72°C for 3 min. Do a single extension of 72°C for 7 min. Heat 98°C for 10 min. Bring samples to 27°C.
6. Store samples at 4°C.
7. Run 10 μL of each sample on a 5% acrylamide minigel, with standards, to check for expression of candidate cDNAs identified by DD-PCR.

Alternative Protocol for PEP (Primer-Extension Preamplification) of RNA isolated from Archival Specimens for Confirmation of Expression of cDNAs Identified by DD-PCR

1. Set up RT reaction:
   10 μL RNA (all the RNA isolated from a microdissected, 10 μM paraffin-embedded sample)
   5 μL 10× PEP buffer
   1.6 μL dNTP stock solution
   40 μM random 15-mer
   H_2O to final volume of 50 μL
2. Heat to 94°C for 2 min.
3. Add 2 units of AMV Reverse Transcriptase.
4. Heat to 42°C for 45 min and then 94°C for 2 min.
5. Add 2.5 units of Taq polymerase and overlay with 100 μL mineral oil for the initial PCR with random 15-mer oligonucleotide primers.
6. Do 25 cycles of PCR at 94°C for 1 min, 37°C for 2 min, 55°C for 2 min. Do a single extension of 55°C for 3 min. Heat to 98°C for 10 min. Bring samples to 27°C.
7. Set up the first round of gene-specific PCR:
   5 μL from the initial PCR from step 6
   10 μL 10× PEP PCR buffer
   3.2 μL dNTP stock solution
   1 μL 0.1 nmol outer 5′ primer
   1 μL 0.1 nmol outer 3′ primer
   H_2O to a final volume of 99.5 μL
8. Add 2.5 units of Taq polymerase and overlay with 100 μL mineral oil.
9. Do 35 cycles of PCR at 94°C for 1 min, 55°C for 2 min, 72°C for 3 min. Do a single extension of 72°C for 7 min. Heat to 98°C for 10 minutes. Bring samples to 27°C.

10. Set up the second round of gene-specific PCR:
    2 μL of the first round PCR
    10 μL 10× PEP PCR buffer
    3.2 μL dNTP stock solution
    1 μL 0.1 nmol nested 5′ primer
    1 μL 0.1 nmol nested 3′ primer
    H_2O to a final volume of 99.5 μL
11. Add 2.5 units of Taq polymerase and 100 μL mineral oil.
12. Repeat PCR cycles as in step 9.
13. Run 10 μL of each sample from the first- and second-round PCRs on a 5% acrylamide minigel with standards to check for expression of candidate cDNAs identified by DD-PCR.

Notes on the Methods

a. We routinely use AMV reverse transcriptase in our experiments; however, it is likely that M-MLV reverse transcriptase is equally effective. If you decide to use M-MLV reverse transcriptase, alter the reaction buffer to that recommended for M-MLV in order to ensure maximum enzyme activity (see for example, Chapter 13, above).
b. The tissue can also be stored at −70°C for later sectioning. Thawing should be avoided to prevent RNase activity.
c. Typical equipment for microdissection includes an inverted scope and a mechanical micromanipulator for holding and manipulating the cutting tool.
d. This is usually accomplished by scraping with a sterile instrument such as a scalpel blade.
e. Non-paraffin-embedded sections, such as frozen sections or air-dried frozen sections, and paraffin-embedded sections that were not deparaffinized are similarly resuspended in this buffer directly following step 1 of the extraction protocol.
f. For best results, an overnight precipitation is recommended.
g. Do a 1:10 dilution of the dNTP stock to prepare 2.5 mM dNTP mix. Dilute solutions of dNTPs are not stable for storage, and should be made fresh from concentrated stock immediately prior to use.
h. We have had success with as little as 50 ng of RNA per reaction, but for best results we recommend using an amount in the range shown, and using the same amount of RNA in each sample to provide consistency across lanes in the gel.
i. A working stock of RT can be made by diluting the enzyme to 2.5 IU/μL in the RT master mix prepared in step 2.
j. Do a 1:50 dilution of the dNTP stock mix to obtain the 0.5 mM dNTP mix.
k. Less 5′ primer is used to favor the formation of products containing a

$T_{11}XX$ 3′ end. We have found that if equal amounts of primer are used, products containing the random 10-mer at both ends predominate.

l. Sample recovery can be accomplished either by the addition of 75 μl chloroform, which causes the aqueous phase to float to the top, or by carefully pipetting the aqueous reaction mixture from underneath the mineral oil.

m. On a 5% gel, xylene cyanol migrates with an apparent size equivalent to an ~150-base single-stranded DNA. We and others have found that products below this size generally represent contaminating artifacts.

COMMENTARY

Examples of the method of the method are illustrated in Figures 14.1 and 14.2.

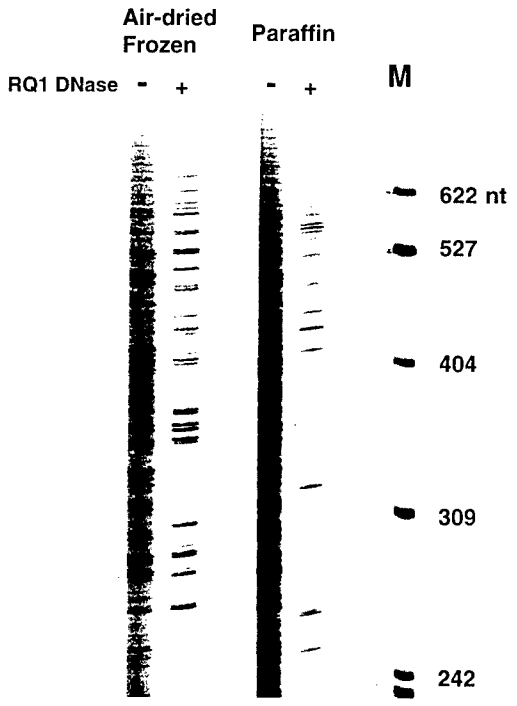

Figure 14.1. Differential display from air-dried, frozen, and paraffin-embedded samples. DD-PCR was performed on RNAs isolated from air-dried frozen sections, or paraffin-embedded permanent sections of normal human cervical tissue. The RNAs were incubated with or without DNase before DD-PCR. Although RNAs can be isolated from both frozen and fixed samples, the best results are obtained from frozen, air-dried, microdissected samples. These figures also emphasize the necessity of DNase-treatment of RNA prior to DD-PCR analysis.

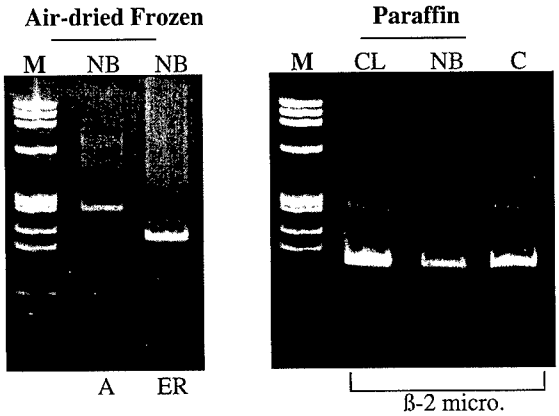

Figure 14.2. RT-PCR of RNA from air-dried, frozen sections or RT-PCR coupled with PEP of RNA isolated from paraffin-embedded permanent sections. RNAs were isolated from air-dried, frozen sections of normal human breast (NB) tissue, or from paraffin-embedded permanent sections of human breast cancer cell line (CL), NB, and normal human cervix (C). The air-dried RNAs were amplified by RT-PCR using either actin-specific (A) primers, and primers to the human estrogen receptor (ER), which is expressed at low levels. The RNAs isolated from the paraffin-embedded sections were amplified by RT-PCR coupled with PEP using primers specific for human β-2-microglobulin. These results demonstrate that confirmation of DD-PCR cDNAs can be performed by RT-PCR methodology using tissue samples microdissected from air-dried, frozen, or paraffin-embedded sections. Neither of these methods, RT-PCR and RT-PCR coupled with PEP, are appropriate methods for quantitation of cDNA expression, but rather are appropriate only for determining whether a candidate cDNA identified by DD-PCR is expressed in specific tissues and tumors.

Critical Parameters

Evaluating gene expression by differential display may be relatively simple if the source of RNA is specific and abundant, such as with cell lines. Mammalian tissues, such as human tumors, may contain sufficient RNA, but they are derived from many cell types that are present in varying proportions. Specific cells can be obtained from heterogeneous animal tissues by microdissecting them from histological sections viewed under a light microscope. Histological sections are typically prepared from formalin-fixed, paraffin-embedded tissue and stained with hematoxylin and eosin (H&E). Unfortunately, the RNA extracted from this type of tissue preparation is usually of insufficient quality for routine DD-PCR. However, adequate amounts of quality RNA can be obtained from air-dried, frozen histological sections. The relative absence of water in these sections is apparently sufficient to inhibit endogenous RNAse activity, making it possible to routinely amplify cDNA's larger than 600 base pairs from this source of RNA.

During microdissection, specific cells of interest can be visualized without staining if the microscope contrast is high (e.g., by unfocusing the condensor or using phase-contrast rings) and an adjacent H&E stained section is used as a guiding template. Alternatively, staining briefly with aqueous hematoxylin

and quickly redrying the slide allows for better direct visualization and only slightly decreases the yield of RNA. Accurately separating different types of cells obviously requires familiarity with the histopathological features of the tissue. About 10^4 cells are needed to obtain sufficient RNA to run one primer pair in DD-PCR or RT-PCR. This may require dissecting one to several slides depending on target cell distribution and density. Obtaining sufficient cells may be tedious and time-consuming if the target cells are rare and/or scattered in small groups. Thus, harvesting enough cells can be greatly facilitated by preselecting specimens containing relatively large areas of apposed target cells. It is possible to routinely prepare samples enriched to >95% target cellularity, although 100% purity is nearly impossible because of intermingled capillary endothelium, fibroblasts, lymphocytes, and other cell types present in tissues.

DD-PCR is best performed using RNA isolated from microdissected frozen, or air-dried frozen-section samples. DD-PCR performed on RNAs isolated from routine clinical paraffin-embedded sections appears to be unreliable, thus making interpretation difficult (although RNA from tissues rapidly fixed in formalin for short periods of time can be of high quality). Furthermore, it must be emphasized that cDNAs differentially displayed in DD-PCR must be confirmed for expression using alternative methods of RNA analysis. RT-PCR of RNA isolated from frozen or air-dried frozen sections is a feasible technique for confirmation using gene-specific primers to the cDNA of interest isolated from the DD-PCR gels. However, the amount of RNA isolated from paraffin-embedded samples precludes direct amplification using gene-specific primers. Instead, one must couple nonspecific amplification techniques, such as PEP, to obtain sufficient cDNA for subsequent PCR with gene-specific primers. Using either of these two techniques, RT-PCR or RT-PCR coupled with PEP, one can quickly confirm differential gene expression of candidate cDNAs in a relatively short timeframe in the specific cells of interest. Microdissection and direct confirmation of gene expression in the target cells of interest is critical for successful application of DD-PCR to clinical samples.

Troubleshooting

The most common problem associated with the recovery of RNA from archival, paraffin-embedded sections is RNA degradation during the tissue fixation and paraffin-embedding procedures. We have found that most of the archival specimens contain partially degraded RNA. Therefore, one must design gene-specific primers that flank the target cDNA sequence such that the target is no greater than 200 base pairs (i.e., $\leq$ 200 bp) in length, with shorter PCR target fragments preferable. Another potential source of degradation is improper storage of the air-dried, frozen mounted samples. The mounted sections, or any microdissected segments from these sections, must be kept dry and used as soon as possible.

Time Considerations

It may take from a few minutes to an hour or more to microdissect histological sections, depending on the amount and distribution of target cells on the slide.

Paraffin-embedded samples take about 50 min to deparaffinize. These samples, along with non-paraffin-embedded sections, are then incubated for 6 h. After approximately 15 min for purification steps, the samples can be precipitated from one hour to overnight. Twenty minutes is then required to prepare the samples to be used in subsequent procedures.

The DD-PCR procedure takes approximately 30 min to one hour to prepare the reverse transcriptase reactions, depending on the number of samples being set up. The RT reaction takes about 70 min, and the PCR cycling reactions an additional 3 h depending on the thermal cycling parameters. It takes about 30 min to prepare the Long Ranger™ gel and one or more hours for the gel to set while the thermal cycler is running. Allow about 30 min to set up the sequencing gel apparatus, to prepare, and then load the samples. The gel will require between 2 and 3 h to run, with approximately one hour to vacuum-dry. Exposure of the gel to film takes 16–24 h (usually done overnight).

Between 30 min and one hour is required to prepare the RT-PCR reactions. The RT reaction takes about 50 min, and after adding Taq polymerase, the PCR cycling requires over 5 hours depending on the model cycler used. For RT-PCR coupled with PEP, one again needs about 30 min to one hour to prepare the reaction mixes. The RT reaction takes about 50 min, and the PEP procedure requires a little over 3 h. The first and second rounds of PCR each require a little over 5 h. The first round can be run overnight and left at 4°C until the next morning, when the second round can be set up and run.

REFERENCES

Cooper GM (1995): *Oncogenes*, 2nd ed. Boston: Jones & Bartlett.

Fuqua SAW, Fitzgerald SD, McGuire WL (1990): A simple polymerase chain reaction method for detection and cloning of low-abundance transcripts. BioTechniques 9:206–211.

Liang P, Pardee AB (1992): Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967–971.

Stanta G, and Schneider C (1991): RNA extracted from paraffin-embedded human tissues is amenable to analysis by PCR amplification. *BioTechniques* 11:304–308.

Zhang L, Xiangfeng C, Schmitt K, Hubert R, Navidi, W, Arnheim, N. (1992): Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci* (USA) 89:5847–5851.

Work was supported by NIH CA 585183.

Statistical Analysis of Array Expression Data as Applied to the Problem of Tamoxifen Resistance

Susan G. Hilsenbeck, William E. Friedrichs, Rachel Schiff, Peter O'Connell, Rhonda K. Hansen, C. Kent Osborne, Suzanne A. W. Fuqua

Background: **Although the emerging complementary DNA (cDNA) array technology holds great promise to discern complex patterns of gene expression, its novelty means that there are no well-established standards to guide analysis and interpretation of the data that it produces. We have used preliminary data generated with the CLONTECH Atlas™ human cDNA array to develop a practical approach to the statistical analysis of these data by studying changes in gene expression during the development of acquired tamoxifen resistance in breast cancer.** ***Methods:*** **For hybridization to the array, we prepared RNA from MCF-7 human breast cell tumors, isolated from our athymic nude mouse xenograft model of acquired tamoxifen resistance during estrogen-stimulated, tamoxifen-sensitive, and tamoxifen-resistant growth. Principal components analysis was used to identify genes with altered expression.** ***Results and Conclusions:*** **Principal components analysis yielded three principal components that are interpreted as 1) the average level of gene expression, 2) the difference between estrogen-stimulated gene expression and the average of tamoxifen-sensitive and tamoxifen-resistant gene expression, and 3) the difference between tamoxifen-sensitive and tamoxifen-resistant gene expression. A bivariate (second and third principal components) 99% prediction region was used to identify outlier genes that exhibit altered expression. Two representative outlier genes, erk-2 and HSF-1 (heat shock transcription factor-1), were chosen for confirmatory study, and their predicted relative expression levels were confirmed in western blot analysis, suggesting that semiquantitative estimates are possible with array technology.** ***Implications:*** **Principal components analysis provides a useful and practical method to analyze gene expression data from a cDNA array. The method can identify broad patterns of expression alteration and, based on a small simulation study, will likely provide reasonable power to detect moderate-sized alterations in clinically relevant genes. [J Natl Cancer Inst 1999; 91:453–9]**

Tremendous effort in cancer research has been devoted to identifying biologically relevant, differentially expressed genes by comparing, for example, tumor cells with normal cells or primary cells with metastatic cells. Until recently, most studies have been limited to quantitation of expression of at most a few genes at a time. Complementary DNA (cDNA) arrays offer the potential to simultaneously quantify expression of many genes. Advances in cDNA array technology to address issues, such as array size, probe density, probe content, and readout, now make this technology sufficiently flexible, accessible, and practical for application in the laboratory *(1)*. The novelty of this technology means that there are no well-established and widely accepted standards to guide analysis and interpretation of the data that it produces. Thus far, cDNA arrays of one type or another have been most often used in paired comparisons (e.g., control versus cancer) to identify differentially expressed genes in only a few types of cancer, such as melanoma *(2)*, Ewing's sarcoma *(3)*, oral cancer *(4)*, glioblastoma multiforme tumors *(5)*, and gastrointestinal tumors *(6)*. After standardization, rules for gene selection were typically based on ratios of expression [for example, greater than twofold difference *(7)*, greater than three standard deviations of control genes ratio *(2)*, or an arbitrary percent]. Application of the technology to more complex experimental designs involving simultaneous analysis of multiple experimental conditions or sampling over several time points will require a more general approach.

Tamoxifen is the most frequently prescribed drug for the treatment of breast cancer. Its use in breast cancer treatment has expanded from first-line treatment for advanced metastatic disease *(8)*, to adjuvant therapy after surgery for primary disease *(9)*, and possibly to prevent breast cancer *(10)*. Acquired tamoxifen resistance is a clinically important problem because a majority of patients with breast cancer will be offered tamoxifen at some time during their treatment, and although tamoxifen is initially effective in many patients, resistance eventually develops. Clinical resistance is almost certainly heterogeneous and multifactorial. Changes may be at the level of the target estrogen receptor *(11–14)*, at a postreceptor point in the estrogen-receptor-response pathway *(15–18)*, and/or downstream of the response pathway *(19–21)*. With cDNA array technology *(6,22,23)*, we may be able to discern the potentially complex patterns of gene expression that are involved in the acquisition of resistance.

In this study, we have used principal components analysis as a practical, but statistically valid, approach to simultaneously examine array data from several time points in an *in vivo* model of acquired resistance. The model simulates the clinical tamoxifen-resistant phenotype by using estrogen receptor-positive MCF-7 breast cancer tumors growing in athymic nude mice *(24)*. We demonstrate that principal components analysis can reliably detect moderately sized alterations in gene expression that we have confirmed by western blot analysis.

Materials and Methods

Tumors and Microarray Hybridization

MCF-7 breast cancer cells were injected into the mammary fat pads of athymic nude mice supplemented with an estrogen pellet as described previously *(24)* until tumors grew. The estrogen pellets were removed and the animals were treated with

Affiliations of authors: S. G. Hilsenbeck, W. E. Friedrichs, R. Schiff, R. K. Hansen, C. K. Osborne, S. A. W. Fuqua (Departments of Medicine/Oncology), P. O'Connell (Department of Pathology), The University of Texas Health Science Center, San Antonio.

Correspondence to: Suzanne A. W. Fuqua, Ph.D., The University of Texas Health Science Center, Departments of Medicine/Oncology, 7703 Floyd Curl Dr., San Antonio, TX 78248–7884 (e-mail: suzanne_fuqua@oncology.uthscsa.edu).

See "Notes" following "References."

tamoxifen. Tumor volumes then declined and remained stable for several months. Invariably, however, after initial growth suppression, the tumors became resistant and growth resumed. Animals were killed at various times to obtain estrogen-stimulated tumors before tamoxifen treatment, tamoxifen-sensitive tumors during tamoxifen treatment but before acquired resistance, and tamoxifen-resistant tumors after tumor growth had resumed. We collected five tumors from each group. We then prepared total RNA with RNeasy kits (Qiagen Inc., Valencia, CA), and isolated messenger RNA on Dynabeads (Dyna, Oslo, Norway) according to manufacturer's instructions. For each group, the RNAs were pooled and used to synthesize ^{32}P-radiolabeled cDNAs for hybridization to the Atlas™ human cDNA expression array-1, according to the manufacturer's instructions *(25)* with SuperScriptII reverse transcriptase (Life Technologies, Inc. [Gibco BRL], Gaithersburg, MD). The CLONTECH Atlas™ human cDNA expression array is a positively charged nylon membrane (8 × 12 cm) that is spotted in duplicate with 200- to 600-base-pair cDNA fragments representing 588 genes and 21 housekeeping genes or control sequences *(25)*. Genes are arrayed in six quadrants with genes of like function (i.e., oncogenes, assorted receptors, etc.) grouped together geographically. The hybridization data were collected with a Molecular Dynamics PhosphoImager™ (Molecular Dyanmics, Sunnyvale, CA). This array was essentially the only one available when these experiments were done. Although the array does not include the estrogen receptor, it does include many other genes of potential interest in breast cancer, including two that we have studied previously, hsp27 and heregulin-α. We collected data from three arrays, one array for each tumor type.

Western Blot Analysis

Pulverized frozen tumors were manually homogenized in 5% sodium dodecyl sulfate. After boiling and microcentrifugation (10 minutes at 10 000 rpm, room temperature), clear supernatants were collected, and the protein concentration was determined by the bicinchoninic acid method (Pierce Chemical Co., Rockford, IL) as previously described *(26)*. Twenty-five micrograms of protein was separated on a denaturing polyacrylamide gel and transferred by electroblotting to nitrocellulose membranes (Schleicher and Schuell, Inc., Keene, NH). The blots were first stained with StainAll dye (Alpha Diagnostic Intl., Inc., San Antonio, TX), to confirm uniform transfer of all samples, and then incubated in blocking solution (5% nonfat dry milk in Tris–HCl buffered saline–Tween [TBST = 50 m*M* Tris–HCl at pH 7.5, 150 m*M* NaCl, and 0.05% Tween 20]). After brief washes with TBST, the filters then were reacted with primary antibodies to erk-2 (UBI, Lake Placid, NY) or heat shock transcription factor-1 (HSF-1) (Stressgen, Victoria, Canada) for 1 hour at room temperature followed by extensive washes with TBST. Blots were then incubated with horseradish peroxidase-conjugated secondary antibody (Amersham Life Science Inc., Arlington Heights, IL) for 1 hour, washed with TBST, and developed by the ECL procedure (Amersham Life Science Inc.). The autoradiograms from the western blots were scanned with a densitometer, and the data are presented as the area determined for each individual tumor sample.

Statistical Considerations

In this pilot study, each hybridization (m = three arrays) resulted in expression values for 588 genes and 21 control genes (putative housekeeping genes and negative control genes). The control genes, which were arrayed in a separate row at the bottom of the array and were more difficult to quantitate reliably in replicated experiments using the same RNA (data not shown), were not included in the statistical analyses. Expression of the highest and lowest expressed genes on the array varied by two to three orders of magnitude. Logarithmic transformation of the raw data reduced this range and helped equalize variability. This also means that additive effects on the log scale can be interpreted as fold changes in actual expression.

Because of the expense, limited amounts of RNA, and other considerations, array experiments usually have few replications and invariably have orders of magnitude more variables (genes and expressed sequence tags) than observations (hybridizations). In this study, we switch the roles of variables and observations, treating each tumor type as a variable (m = three arrays) and each expressed gene sequence as an observation (n = 588 genes).

Principal components analysis of mean-centered log-transformed data, based on the variance-covariance matrix *(27)*, was then used to standardize across the three hybridizations and to extract three new axes (components P1, P2, and P3), expressed as linear combinations of the original axes (variables ES [estrogen-stimulated], TS [tamoxifen-sensitive], and TR [tamoxifen-resistant]).

$$P1 = A_1 * ES + B_1 * TS + C_1 * TR$$

$$P2 = A_2 * ES + B_2 * TS + C_2 * TR$$

$$P3 = A_3 * ES + B_3 * TS + C_3 * TR$$

In principal components analysis, the coefficients (As, Bs, and Cs) are chosen so that the first component (P1) explains the maximal amount of variance in the data. The second component (P2) is perpendicular to the first and explains the maximal residual squared variation, and the third component (P3) is perpendicular to the first two. Meaning was ascribed to the new axes by visual examination of the coefficients. In these array experiments, P1 represents the average level of expression across the tumor types and P2 and P3 represent differences between tumor types. A bivariate analysis that results in two new axes (P1 and P2) was also performed to compare tamoxifen-sensitive gene expression with tamoxifen-resistant gene expression. The coefficents do not always have a nice biologically sensible interpretation, although the higher-order components can still be used to identify outlier genes, regardless of interpretation (*see below*).

We used P2 (and P3 in the higher-order analysis) to identify outlier genes that might represent true alterations in gene expression. In the bivariate principal components analysis of tamoxifen-sensitive gene expression versus tamoxifen-resistant gene expression, we used a normal approximation to construct a 99% prediction region for component P2 (i.e., $0 \pm 2.57 * SD_r$, where SD_r = interquartile range/1.35). A robust estimate of the standard deviation (SD_r) was used to reduce the variance-inflating effects of outliers *(28)*. Genes outside the region were identified for further study. Analogously, in a trivariate principal components analysis (estrogen-stimulated, tamoxifen-sensitive, and tamoxifen-resistant gene expression), we computed a 99% bivariate normal prediction ellipse *(27,29)* for components P2 versus P3, and genes outside the ellipse were selected for investigation.

This "robust prediction interval" approach seems justified on the following basis. Although the distribution of P1 is highly skewed, higher-order components are roughly symmetric. When there is no differential expression, as in a bivariate analysis of two array hybridizations using the same pool of RNA, the higher-order components are approximately normally distributed (data not shown). In experiments comparing different pools of RNA, where some genes may be differentially expressed, the observed distribution of each higher-order component (P2, P3, etc.) should be a mixture of central ($\mu = 0$) and noncentral ($\mu \neq 0$) distributions. By using a robust estimator that focuses on the middle of the observed distribution, which should represent primarily unaltered genes, we hope to increase sensitivity to identify truly altered genes. The prediction level (99%), which is analogous to the specificity of a diagnostic test, was chosen arbitrarily as representing a reasonable balance between identifying too many spuriously "significant" genes and missing true alterations. For display purposes, we have back-transformed the data by exponentiation of P2 and P3 so that the data are shown as approximate fold increases or decreases in expression.

The ability of this methodology to detect true alterations was examined in a small simulation study. Log-transformed values from a hypothetical bivariate array experiment with 588 genes were generated to have a common log-normally distributed component for level of expression [i.e., $\exp(X) + 8$, where $X \sim N(\mu = 0, \sigma = 0.6)$], and independent normally distributed errors [i.e., $\log_e$ (Control) = $\exp(X) + 8 + Y$ and $\log_e$ (Experimental) = $\exp(X) + 8 + Z$, where $Y,Z \sim N(\mu = 0, \sigma = 0.17)$].

The distributional parameters were chosen to mimic data seen in our real experiments. A small percentage of truly altered genes (2% or 4%) were created by shifting the error distribution for the experimental member of the pair up or down (with 50% probability) to represent an average 2- or 2.5-fold change from baseline [i.e., $\log_e$ (Experimental) = $\exp(X) + 8 + W$, where $W \sim N(\mu = \pm 0.7, \sigma = 0.17)$]. The generated data were then analyzed as described above, and the numbers of truly altered and spuriously altered genes falling outside the 99% prediction region were tabulated. Each scenario was replicated 100 times, and the results were summarized over all replications. All analyses were performed with the SAS program package (Version 6.11, SAS Institute, Cary, NC).

Results

Bivariate Analysis

Fig. 1 shows the three bivariate log-log scatter plots that arise from pairwise comparisons of the data from the three cDNA array hybridizations (one for estrogen-stimulated tumors, one for tamoxifen-sensitive tumors, and one for tamoxifen-resistant tumors). Each gene of the 588 genes on the array (excluding housekeep-

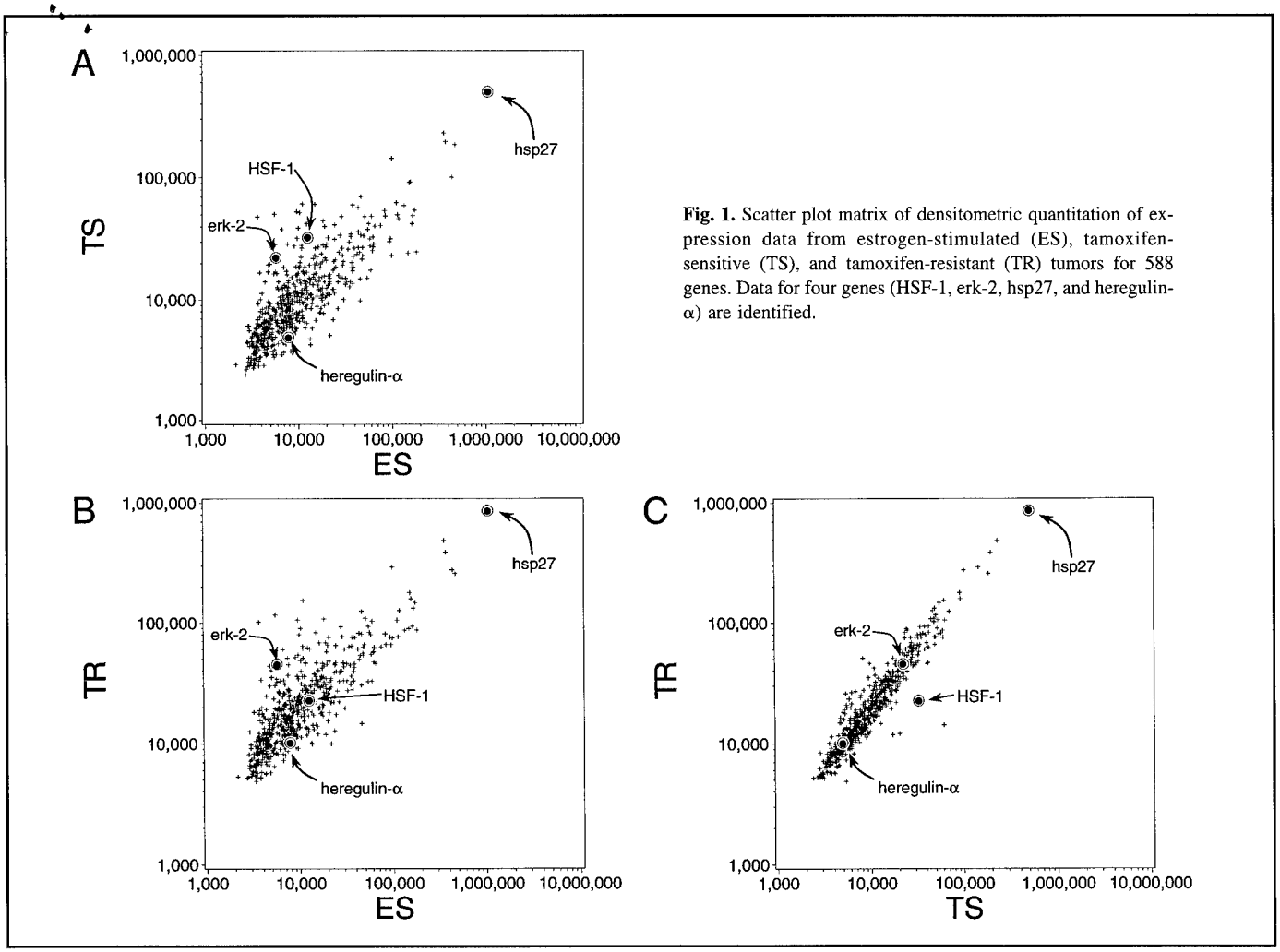

Fig. 1. Scatter plot matrix of densitometric quantitation of expression data from estrogen-stimulated (ES), tamoxifen-sensitive (TS), and tamoxifen-resistant (TR) tumors for 588 genes. Data for four genes (HSF-1, erk-2, hsp27, and heregulin-α) are identified.

ing and control genes) is represented by a point on the scatter plots. The individual values ranged over two to three orders of magnitude, indicating that the most highly expressed genes were expressed at 100- or 1000-fold higher levels than the lowest expressed genes. For example, the 27-kd heat shock protein (hsp27) was the most highly expressed gene on the array in all three tumor types. This finding is consistent with our previously published result that hsp27 is amplified and overexpressed in the late-passage MCF-7 cells used in this model *(30)*. Similarly, the array results are consistent with previous findings *(31)* that heregulin-α is expressed at relatively low levels in all three types of tumor cells.

In each scatter plot, most genes lie fairly close to a diagonal line of "identity." This line may not be centered on the graph if there are differences in the average level of radioactivity of probes used in each hybridization. The distance along this line denotes differences in the level of expression between genes, such as we see between hsp27 and heregulin-α. The perpendicular distance away from the line denotes differences in expression within the same gene between tumor types.

Principal components analysis of the log-transformed expression data was used to produce a new set of axes (Fig. 2). For tamoxifen-sensitive tumors versus tamoxifen-resistant tumors (Fig. 2, A), the new x axis or first principal component (P1) roughly corresponds to the line of "identity" and represents level of expression. The second principal component (P2) is perpendicular to the first and represents difference in expression between tumor types. In the bivariate analysis, more than 97% of the total variation in the log-transformed data was associated with P1, leaving about 3% for P2. The two components are, by definition, not correlated ($\rho = 0$). The distribution of P1 is skewed, because many genes on the array are expressed at low to moderate levels, but only a few are expressed at extremely high levels. The distribution of P2 is roughly symmetric, and a 99% robust prediction interval identified 35 outlier genes that may be over- or under-expressed in tamoxifen-resistant tumors relative to tamoxifen-sensitive tumors (Fig. 2, B).

Trivariate Analysis

Bivariate principal components analysis could be performed for each pair of tumor types; however, a more comprehensive three-way analysis is preferred and is more biologically relevant. Principal components analysis of the mean-centered log-transformed data (for estrogen-stimulated tumors, tamoxifen-sensitive tumors, and tamoxifen-resistant tumors) yields three new axes (P1, P2, and P3) that account for 90.5%, 8%, and 1.5% of the variation in the data, respectively. By inspection of the coefficients, the first principal component (P1) is again interpreted as the "average level of expression" because the coefficients were all positive and similar in value (0.63, 0.55, and 0.55, respectively). The second principal component (P2) clearly contrasts data from estrogen-stimulated tu-

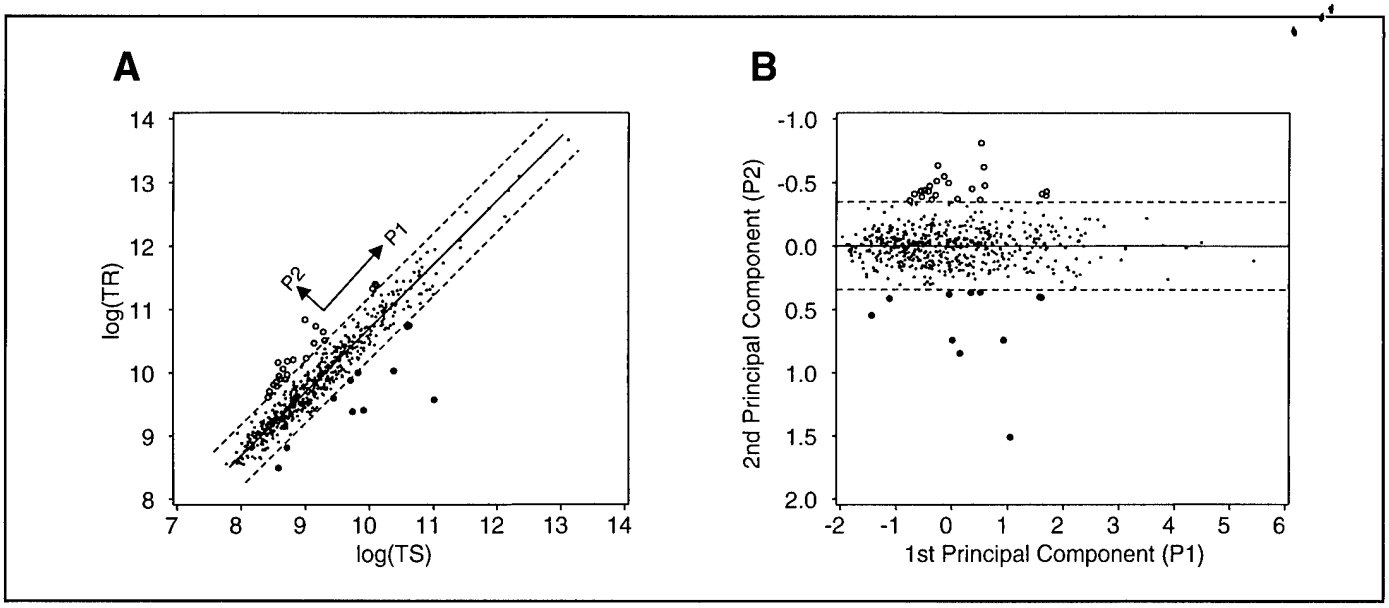

Fig. 2. A) Scatter plot of log-transformed expression data for tamoxifen-sensitive (TS) and tamoxifen-resistant (TR) tumors showing the line of identity (solid line) and 99% prediction region (dashed line). The open and solid circles indicate genes that are overexpressed or underexpressed, respectively, in tamoxifen-resistant tumors relative to tamoxifen-sensitive tumors. Dots indicate genes that are not affected by tamoxifen. **B)** Scatter plot of first and second principal components from the same data shown in A.

mors to the average of tamoxifen-sensitive and tamoxifen-resistant tumors because the P2 coefficient for the estrogen-stimulated data is negative (−0.78) and roughly equal to the sum of the tamoxifen-sensitive and tamoxifen-resistant coefficients (0.46 and 0.43, respectively). The third principal component (P3) primarily represents differences between the tamoxifen-sensitive and the tamoxifen-resistant tumors, because the P3 coefficient for the estrogen-stimulated tumors is small (0.02) and the tamoxifen-sensitive and tamoxifen-resistant coefficients are nearly equal but opposite in sign (0.69 and −0.72, respectively). Fig. 3 shows a scatter plot of P2 versus P3. Points near the center represent genes that were similarly expressed in all three tumor types, whereas points on the periphery exhibit alterations in expression. Data have been back-transformed to show the approximate fold changes in expression. We used a bivariate normal approximation with robust estimates of standard deviations to compute a 99% prediction ellipse. Genes lying outside the region may exhibit real alterations in the level of expression that are associated with the biologic effects during the transition from estrogen-stimulated to tamoxifen-sensitive status and tamoxifen-sensitive to tamoxifen-resistant status.

In addition, different regions of the P2 × P3 plane correspond to different temporal patterns of expression alteration. For example, expression of genes to the far right in Fig. 3 (i.e., near erk-2) is increased by tamoxifen relative to the expression of genes in estrogen-stimulated tumors but expression of genes in this area is unchanged in tamoxifen-resistant tumors relative to tamoxifen-sensitive tumors. In contrast, expression of genes to the lower right in Fig. 3 (i.e., near HSF-1) is increased in tamoxifen-sensitive tumors relative to estrogen-stimulated tumors but is decreased in tamoxifen-resistant tumors.

Confirmation of Gene Expression by Western Blot Analysis

We selected two genes just outside of the 99% prediction ellipse (erk-2 and HSF-1) for quantitation by western blot analysis. These two genes were chosen because of their relatively low expression (Fig. 1) and modest alteration, so that we could address sensitivity questions and the ready availability of specific antibodies. The erk-2 kinase is a known mediator of the growth factor signaling pathway, and it has been shown that the estrogen receptor can activate its activity in MCF-7 cells *(32)*. HSF-1 is involved in cellular stress responses *(33)* and is thus a potential marker of tamoxifen-induced stress. We found that the relative levels of erk-2 and HSF-1 predicted in the array experiment were indeed confirmed in an independent set of individual tumors (Fig. 3, B, lanes 1–15) from the athymic nude mouse model. As predicted by Figs. 1, A, and 3, A, western blot results for HSF-1 indicate a substantial increase in expression in tamoxifen-sensitive tumors relative to estrogen-stimulated tumors, which is followed by a decrease in tamoxifen-resistant tumors to approximately the levels in estrogen-stimulated tumors (Fig. 1, B). Similarly for erk-2, there is an increase in expression in tamoxifen-sensitive tumors relative to estrogen-stimulated tumors (Fig. 1, A), but there is relatively less change between tamoxifen-sensitive and tamoxifen-resistant tumors.

Power Considerations

Using distributional parameters from some of our pilot studies, we ran a series of simulations to investigate the likely sensitivity of these methods to detect real differences of moderate size (Table 1). With modest changes (twofold) in 2%–4% of genes, 99% of the unchanged genes were correctly classified as unchanged by the 99% prediction interval, and 59% of the altered genes were correctly identified as outliers. With larger differences (e.g., 2.5-fold), the proportion of correctly identified outliers goes up (85%). Although the outliers will always be contaminated by a few spuriously identified genes, these results suggest that the method has reasonable power to detect real differences.

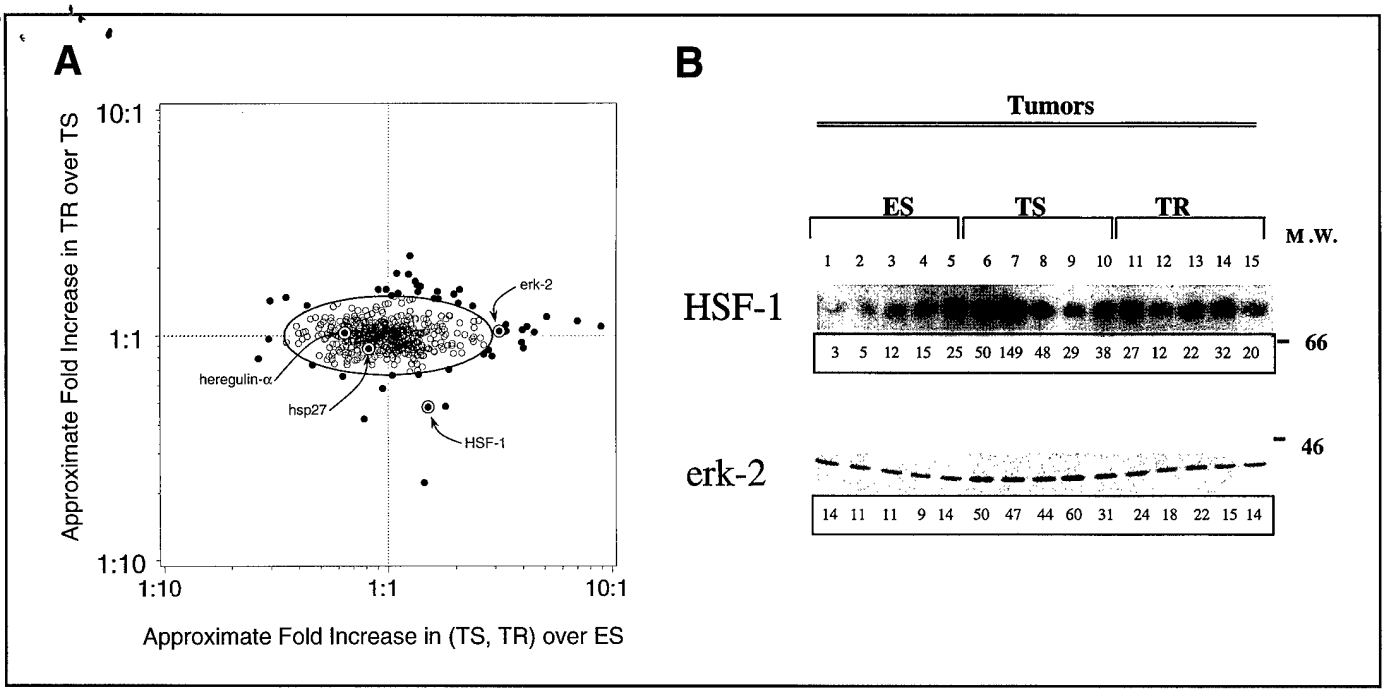

Fig. 3. A) Scatter plot of second and third principal components from principal components analysis of log-transformed gene expression data from estrogen-stimulated (ES), tamoxifen-sensitive (TS), and tamoxifen-resistant (TR) tumors, back-transformed to show approximate fold alterations. Axis labels describe the qualitative interpretation of principal components analysis coefficients. Genes inside or outside the 99% prediction ellipse (solid line) are shown as open or solid circles, respectively. Data for four genes (HSF-1, erk-2, hsp27, and heregulin-α) are identified. **B)** Western blot analysis with erk-2 and HSF-1 antibodies in estrogen-stimulated (ES, lanes 1–5), tamoxifen-sensitive (TS, lanes 6–10), and tamoxifen-resistant (TR, lanes 11–15) tumors (five tumors are in each group). The positions of molecular weight (M.W.) markers (in 10^{-3} kd) are shown to the right. Densitometric scan values (in relative units) for each lane are shown in the boxed area below each western blot lane. For HSF-1 protein expression, there was a fivefold increase in the tamoxifen-sensitive tumors and a 1.8-fold increase in the tamoxifen-resistant tumors relative to the estrogen-stimulated tumor group. For erk-2 expression, there was a fourfold increase in the tamoxifen-sensitive tumors and a 1.6-fold increase in the tamoxifen-resistant tumors relative to the estrogen-stimulated tumor group; the slight difference in protein levels compared with that predicted by the RNA array analysis may reflect posttranscriptional and/or translational control of erk-2 protein.

Table 1. Results of simulation study involving 588 genes in two tumor types and using a 99% prediction interval

Average fold change in altered gene expression	% of genes with altered expression	% of genes with unaltered expression inside interval*	% of genes with altered expression outside interval†
2.0	2	99	59
	4	99	60
2.5	2	99	86
	4	99	85

*This is the observed specificity and is analogous to prediction level ($1 - \alpha$).
†This is the observed sensitivity and is analogous to power.

Discussion

cDNA microarray expression profiling offers tremendous potential to simultaneously characterize the expression of large numbers of gene sequences. In theory, comparisons of hybridization data from pairs or a series of RNA pools, representing cells from various tumors or experimental conditions, should allow us to identify differentially expressed genes or sequences that may be involved in the biologic process under investigation. In practice, it is not so easy to distinguish true differences in expression from differences in expression due to experimental variability only. In a traditional study of one or a few genes, statistical analysis of experimental replicates would be used to estimate variability in expression for each gene to determine whether expression is altered. Variability between replicates is often large, and moderate-sized differences (two- to 10-fold) can require many experimental replications. Due to expense, limited amounts of RNA, and other considerations, array experiments usually have few replications and invariably have orders of magnitude more variables (genes and expressed sequence tags) than observations. In our study of acquired tamoxifen resistance, we have switched the roles of variables and observations and used principal components analysis, coupled with robust estimates of 99% prediction regions on higher-order components, as a practical approach to screening array data for likely candidates for further study. The method presumes that the vast majority of genes will be altered very little and uses information from all genes to obtain more stable estimates of variability. The method is not limited to pairwise comparisons but can be used to study several tumor types or experimental conditions simultaneously. In a small simulation study, we have shown that this approach is capable of reliably identifying 60%–85% of genes exhibiting moderate degrees of differential expression (2- to 2.5-fold), without increasing the number of spuriously identified outliers.

In this study, we used an *in vivo* athymic mouse model of acquired tamoxifen resistance *(24)* to explore the power of microarray expression profiling. In this tamoxifen-resistance model, we have previously shown that one potential resistance mechanism is stimulation of the tumor by tamoxifen, which acts as a partial agonist. As our first analysis, we used the

array technology to identify those genes that might be associated with this growth stimulation. We hypothesized that the tamoxifen-stimulated phenotype could result from the deregulated expression of downstream growth-regulatory pathways that liberate the cell cycle from normal steroid control. Indeed, it has been reported that overexpression of single growth regulatory genes such as cyclin D1 *(34)*, protein kinase A *(35)*, and transforming growth factor β *(21)* can influence a cell's response to tamoxifen treatment. However, there are probably multiple mechanisms that coexist in tumors and in conjunction contribute to the clinical tamoxifen-resistant phenotype. The microarray expression profiling technology is well-suited for this clinical problem. Principal components analysis of our preliminary data suggests that distinct patterns of temporal alteration in gene expression can be distinguished. Our future studies will be aimed at identifying which of the outlier genes are most contributory to the tamoxifen-stimulated phenotype and testing these genes in clinical samples on custom microarrays. From these studies, we expect to identify the gene expression patterns predictive of tamoxifen-resistant growth.

In summary, principal components analysis of log-transformed array data provides a practical approach to data reduction, visualization, and identification of "significant" outlier genes. As a result, analysis of cDNA expression arrays can identify genes and pathways that are altered during the process of resistance. We predict that principal components analysis or related methods of analysis of microarray expression data will lead to the identification of novel growth pathways that are important for the generation of tamoxifen resistance and thus will generate new predictive clinical paradigms.

REFERENCES

(1) Marshall A, Hodgson J. DNA chips: an array of possibilities. Nat Biotechnol 1998;16:27–31.

(2) DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, et al. Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nat Genet 1996;14:457–60.

(3) Welford SM, Gregg J, Chen E, Garrison D, Sorensen PH, Denny CT, et al. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. Nucleic Acids Res 1998;26:3059–65.

(4) Chang DD, Park NH, Denny CT, Nelson SF, Pe M. Characterization of transformation related genes in oral cancer cells. Oncogene 1998;16:1921–30.

(5) Sehgal A, Boynton AL, Young RF, Vermeulen SS, Yonemura KS, Kohler EP, et al. Application of the differential hybridization of Atlas Human expression arrays technique in the identification of differentially expressed genes in human glioblastoma multiforme tumor tissue. J Surg Oncol 1998;67:234–41.

(6) Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, et al. Gene expression profiles in normal and cancer cells. Science 1997;276:1268–72.

(7) Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci U S A 1996;93:10614–9.

(8) Bezwoda WR, Esser JD, Dansey R, Kessel I, Lange M. The value of estrogen and progesterone receptor determinations in advanced breast cancer. Estrogen receptor level but not progesterone receptor level correlates with response to tamoxifen. Cancer 1991;68:867–72.

(9) Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Early Breast Cancer Trialists' Collaborative Group. Lancet 1992;339:71–85.

(10) Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel P-1 Study. J Natl Cancer Inst 1998;90:1371–88.

(11) Fuqua SA, Fitzgerald SD, Chamness GC, Tandon AK, McDonnell DP, Nawaz Z, et al. Variant human breast tumor estrogen receptor with constitutive transcriptional activity. Cancer Res 1991;51:105–9.

(12) Fuqua SA, Wiltschke C, Castles C, Wolf D, Allred DC. A role for estrogen receptor variants in endocrine resistance. Endocrine-related cancer 1995;2:19–25.

(13) Daffada AA, Johnston SR, Smith IE, Detre S, King N, Dowsett M. Exon 5 deletion variant estrogen receptor messenger RNA expression in relation to tamoxifen resistance and progesterone receptor/pS2 status in human breast cancer. Cancer Res 1995;55:288–93.

(14) Gallacchi P, Schoumacher F, Eppenberger-Castori S, Von Landenberg EM, Kueng W, Eppenberger U, et al. Increased expression of estrogen-receptor exon-5-deletion variant in relapse tissues of human breast cancer. Int J Cancer 1998;79:44–8.

(15) Smith CL, Nawaz Z, O'Malley BW. Coactivator and corepressor regulation of the agonist/antagonist activity of the mixed antiestrogen, 4-hydroxytamoxifen. Mol Endocrinol 1997;11:657–66.

(16) Jackson TA, Richer JK, Bain DL, Takimoto GS, Tung L, Horwitz KB. The partial agonist activity of antagonist-occupied steroid receptors is controlled by a novel hinge domain-binding coactivator L7/SPA and the corepressors N/CoR or SMRT. Mol Endocrinol 1997;11:693–705.

(17) Lavinsky RM, Jepsen K, Heinzel T, Torchia J, Mullen TM, Schiff R, et al. Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. Proc Natl Acad Sci U S A 1998;95:2920–5.

(18) Berns EM, van Staveren IL, Klijn JG, Foekens JA. Predictive value of SRC-1 for tamoxifen response of recurrent breast cancer. Breast Cancer Res Treat 1998;48:87–92.

(19) Zwijsen RM, Wientjens E, Klompmaker R, van der Sman J, Bernards R, Michalides RJ. CDK-independent activation of estrogen receptor by cyclin D1. Cell 1997;88:405–15.

(20) Lonning E, Lien EA. Mechanisms of action of endocrine treatment in breast cancer. Crit Rev Oncol Hematol 1995;21:158–93.

(21) Thompson AM, Kerr DJ, Steel CM. Transforming growth factor beta 1 is implicated in the failure of tamoxifen therapy in human breast cancer. Br J Cancer 1991;63:609–14.

(22) Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467–70.

(23) Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, et al. Accessing genetic information with high-density DNA arrays, Science 1996;274:610–4.

(24) Osborne CK, Hobbs K, Clark GM. Effect of estrogens and antiestrogens on growth of human breast cancer cells in athymic nude mice. Cancer Res 1985;45:584–90.

(25) CLONTECH Laboratories Inc. Atlas™ Human cDNA expression arrays user manual. Palo Alto (CA): CLONTECH Laboratories, Inc; 1997.

(26) Tandon AK, Clark GM, Chamness GC, Ullrich A, McGuire WL. HER-2/neu oncogene protein and prognosis in breast cancer. J Clin Oncol 1989;7:1120–8.

(27) Tatsuoka MM, editor. Multivariate analysis: techniques for educational and psychological research. New York (NY): John Wiley & Sons, Inc; 1971. p. 94–149.

(28) Venables WN, Ripley BD. Modern applied statistics with S-plus. New York (NY): Springer-Verlag; 1994. p. 203–8.

(29) Anderson TW. An introduction to multivariate statistical analysis. New York (NY): John Wiley & Sons, Inc.; 1958. p. 123.

(30) Fuqua SA, Benedix MG, Krieg S, Weng CN, Chamness GC, Oesterreich S. Constitutive overexpression of the 27,000 dalton heat shock protein in late passage human breast cancer cells. Breast Cancer Res Treat 1994;32:177–86.

(31) Tang CK, Perez C, Grunt T, Waibel C, Cho C, Lupu R. Involvement of heregulin-beta2 in the acquisition of the hormone-independent phenotype of breast cancer cells. Cancer Res 1996;56:3350–8.

(32) Migliaccio A, Di Domenico M, Castoria G, de Falco A, Bontempo P, Nola E, et al. Tyrosine

kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. EMBO J 1996;15:1292–300.

(33) Rabindran SK, Giorgi G, Clos J, Wu C. Molecular cloning and expression of a human heat shock factor, HSF1. Proc Natl Acad Sci U S A 1991;88:6906–10.

(34) Neuman E, Ladha MH, Lin N, Upton TM, Miller SJ, DiRenzo J, et al. Cyclin D1 stimulation of estrogen receptor transcriptional activity independent of cdk4. Mol Cell Biol 1997;17:5338–47.

(35) Fujimoto N, Katzenellenbogen BS. Alteration in the agonist/antagonist balance of antiestrogens by activation of protein kinase A signaling pathways in breast cancer cells: antiestrogen selectivity and promoter dependence. Mol Endocrinol 1994;8:296–304.

NOTES

Supported by USAMRDC DAMD17–94-J-4112 and Public Health Service grants CA58183, CA30195, and CA54174, National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

We thank Julia Perkins for preparation of the manuscript.

Manuscript received August 3, 1998; revised December 18, 1998; accepted December 30, 1998.