We are working to develop a novel genomic subtraction method to rapidly identify all chromosomal regions that undergo Loss of Heterozygosity (LOH) in breast cancer. LOH is a major mechanism in the genomic alteration that transforms a normal cell into an unregulated tumor cell. The identification of all regions of LOH will identify causative genes and gene interactions and will yield new understanding of the etiology and pathogenesis of breast cancer and lead to new approaches to detection and treatment.

We designed several genomic subtraction methods to isolate enriched populations of DNA fragments representing LOH regions in individual breast cancer cell lines. Our approach exploits the minor degree of DNA sequence heterogeneity expected between the members of each chromosome pair. Genomic subtraction requires presence/absence differences, and we transform the DNAs to create such differences. Subtraction of tumor DNA from normal tissue DNA from the same individual should leave a remainder population of DNA fragments enriched for the different sequences missing from the tumor DNA due to LOH. We continue to refine and compare alternative methods in model systems to determine the superior approach.
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.

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).

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

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

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.

May 26, 1990

PI - Signature    Date
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRONT COVER</td>
<td>1</td>
</tr>
<tr>
<td>REPORT DOCUMENTATION PAGE (SF298)</td>
<td>2</td>
</tr>
<tr>
<td>FOREWORD</td>
<td>3</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>4</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>5</td>
</tr>
<tr>
<td>BODY</td>
<td>7</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>14</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>14</td>
</tr>
<tr>
<td>BIBLIOGRAPHY / PERSONNEL</td>
<td>15</td>
</tr>
</tbody>
</table>
INTRODUCTION

Our goal is to use genomic subtraction to identify chromosomal regions of Loss of Heterozygosity (LOH) in breast cancer cells. Patterns of LOH will reveal new information about progressive genetic changes in breast cancer, enhancing our understanding of this disease and offering new opportunities for therapeutic intervention.

LOH is a frequent and characteristic event in many or all tumor types, and it operates to unmask deleterious mutations that have occurred, or have been inherited, in particular genes that control growth and differentiation (1). When the normal allele is lost, only the altered function is left. Thus LOH is an integral part of the process of deregulating a cancer cell and may be crucial in the development of most tumors. LOH clearly is important in the progression of malignant disease. Losses at 1p and 11p15 have been associated with metastasis of breast cancer to regional lymph nodes, and LOH at 1q was associated with early disease recurrence (2, 3). Thus LOH patterns can yield information regarding clinical properties of tumors as well as pointing to important oncogenes involved in their genesis. Any LOH that is observed repeatedly probably points to a gene whose alteration is important in malignant progression. Identification of the altered gene and its function would offer new opportunity for diagnosis, intervention and therapy. A systematic study of LOH over the whole genome in many individuals at various stages has not been done for breast cancer or any other tumor. A feasible approach to this goal will bring promise of much valuable new information. We are developing a genomic subtraction method that derives informative probes from the individual's own DNA. We will use these as probes for Fluorescent in situ Hybridization (FISH, 4-6) to identify all areas of LOH over the entire genome in a single analysis.

To do this we must overcome the difficulty that tumor DNA sequences are virtually identical to normal sequences even though there may be chromosome rearrangements or amplifications. We want to exploit the subtle differences that LOH produces, the chromosome regions where allelic sequences are missing. The challenge for a genomic subtraction strategy is to select DNA specifically from these regions of difference and capture sequences that are nearly identical between the retained and the "lost" allele in tumor genomes undergoing LOH. Because only the alleles from one parent are missing while the very similar allelic sequences from the other parent are present, usually in two copies, straightforward subtraction cannot succeed. There is no presence/absence (+/-) difference that would leave a remainder after subtraction of one genome from the other. The system can be transformed, however, into one that has +/- differences suitable for genomic subtraction, and we use DOP-PCR to accomplish this.

We create sampled populations of sequences for subtraction by DOP-PCR, or Degenerate Oligonucleotide Primed PCR (7-9), which uses a short primer sequence corresponding to a commonly occurring hexamer to amplify intervening DNA. Allelic
differences in whether or not a DOP priming site is present will result in the presence or absence of sequences from the amplified populations. These differences are present at high frequency. Subtraction of the amplified tumor DNA from amplified normal DNA from the same individual yields specifically those sequences that differ between alleles in the region(s) of LOH. A diagram of example chromosomes is shown in Figure 1. As illustrated, allelic differences in the location and presence of PCR priming sites will yield similar but not identical sets of PCR products from the two chromosome homologs. If LOH occurs at this region in tumor tissue the PCR products from normal cell DNA will include species that are not produced from the tumor DNA. Subtraction will isolate those particular products which can then be used as FISH probes to identify this chromosomal region.

Figure 1. Diagrammatic representation of heterozygosity giving rise to PCR-produced presence/absence differences.

```
M: < < >aaaa< >bbbbbb< cccccc<  
DOP-PCR products: >aaaa< >bbbbbb<

P: < < >aaaa >bbbbbb< >ccccc<  
DOP-PCR products: >bbbbbb< >ccccc<
```

Diagrammed are segments of the Maternally derived (M) and Paternally derived (P) homologs of an arbitrary chromosome. The > and < characters represent DOP-PCR priming sites. Due to nucleotide sequence differences in the two homologs creating or destroying DOP sites the two homologs yield similar but not identical populations of amplification products. If this were a region of LOH in a tumor, and, e.g., the paternal allele were lost, the tumor PCR products would lack the product ">cccccc<". Note that a sequence identical to or very similar to "cccccc" is present on the Maternal chromosome but without both priming sites.

Advantages inherent in the subtractive approach are that it does not depend on previously defined Restriction Fragment Length Polymorphisms or Simple Sequence Repeat markers or Single Nucleotide Polymorphisms that may or may not be informative in each individual (10, 11); rather, it exploits whatever uncharacterized DNA sequence differences there are in each individual. The goal is to produce a single comprehensive analysis of LOH across the entire genome simultaneously, rather than requiring an intensive series of individual marker assays. FISH painting with subtracted DNAs from different individuals will identify consistent sites of LOH, and the staining patterns will define the "minimum region of overlap" of LOH in different tumors; this will guide positional cloning and identification of novel oncogenes and tumor suppressor genes.
Our objectives for this grant were the development of the methods and the demonstration of genomic subtraction using progressively more complex models. We proposed to begin by demonstrating genomic subtraction in a YAC-Yeast system and then in the more complex congenic mouse DNAs and tumor cell line DNAs. At the end of this two year IDEA grant we can report progress, but not completion. Our goal of a successful genomic subtraction technology is difficult and challenging; because we have to learn to walk before we can run, we are still working in the yeast system as our experimental approach evolves.

In developing the genomic subtraction technique we have investigated two critical variables, the genome sampling strategy and the subtraction method. In sampling the genome we aim for a simple and robust technique that produces a very broad representation of the genome. The subtraction method should also be simple and robust, but the primary aim is effectiveness, i.e., subtraction should be as near to 100% as can be achieved.

We compared three sampling strategies, degenerate-oligo-primed PCR (DOP-PCR, 7-9), inter-repeat-sequence PCR (IRS-PCR, 12), and restriction fragment length polymorphisms (RFLP). We also compared four methods of subtraction, physical subtraction through phenol extraction of avidin-bound biotin-derivatized subtractor DNA (13-15), solid phase subtraction with avidin-beads (16), suicide subtraction through enzymatic degradation of subtractor DNA and its hybrids (17), and most recently, the use of amplification-inert subtractor.

RFLP SAMPLING
Our initial experiments used mouse DNAs, RFLP sampling and phenol subtraction. The products were cloned as plasmid libraries and analyzed by hybridization. The RFLP sampling method -- size fractionating restricted DNAs by agar gel electrophoresis, slicing the gel and extracting the DNA fractions -- proved feasible but quite cumbersome in repeated use. When we analyzed the cloned subtraction products most carried repetitive DNA, and no difference was apparent between the starting DNA and the subtracted DNA. This strategy had no means to amplify the subtracted material so that it could be subjected to additional rounds of subtraction to improve the enrichment. It was very clear that we needed to incorporate PCR amplification of the recovered subtraction products to allow further subtraction and to increase the yield of the final product for use as probes.

PCR SAMPLING
We then compared the two PCR methods for genome sampling and found that the DOP-PCR yielded a much broader and more useful sampling, as well as being very
straightforward and reproducible. All subsequent development has employed this sampling method. The generation of the initial genome sample by PCR then allows additional amplification as needed using the same PCR primers.

Figure 2. **Sampling of genomic DNAs by Inter-Repeat-Sequence PCR (IRS-PCR) and Degenerate-Oligo-Primer PCR (DOP-PCR).** Using gel purified mouse YACs as small model genomes we compared the products of IRS-PCR sampling between repeat elements (B1 in mouse, Alu in human) in the genome with the products of DOP-PCR sampling. IRS-PCR yields small numbers of discrete products, but DOP-PCR produces a continuous spectrum (smear) of products that is a more complete sampling of the genomes.

**BIOTIN-AVIDIN SUBTRACTION**

We compared alternative methods of biotinylating the subtractor DNA and physically removing the avidin-bound biotin-DNA. We have used the photobiotinylation method (13) predominately, but have also used biotin-derivatized oligonucleotide primers or biotin-dUTP in the PCR amplification of the subtractor DNA. All methods appear to work without a clear advantage for any. We also compared the phenol-chloroform extraction of the avidin-biotin-DNA complex (13) to binding the biotin-DNA with avidin conjugated to magnetic beads (CPG, Inc.). Again, both were similarly effective. In these experiments we also incorporated tritium tracer in the subtractor PCR product, and this revealed that a very small (1-3%) fraction of the subtractor DNA was resistant to removal, perhaps due to incomplete biotinylation. Since the subtractor/target DNA ratio is at least 30/1, this is a serious problem. This is confirmed by analysis of the remainder after subtraction; even after three rounds most of the YAC-yeast products are still yeast sequences. Nonetheless, subtraction does remove major yeast components and alter the composition of the YAC DNA sample as shown in Figure 3.
Figure 3. Biotin-avidin mediated subtraction.
This figure shows duplicate polyacrylamide gels comparing DOP-PCR products of yeast, YAC (including the yeast host), and the products of 3 rounds of subtraction of yeast from YAC. With increasing rounds of subtraction the prominent discrete bands in the yeast and YAC lanes appear to be removed leaving a uniform smear of species in the final lane.

**SUICIDE SUBTRACTOR**

We have taken several approaches to address the problem of residual subtractor DNA besides the straightforward one of refining and improving the degree of biotinylation of the subtractor. We developed a "suicide subtractor" approach in which the subtractor and any target strands hybridized to it are degraded and destroyed. To do this we incorporate uridine into the subtractor PCR product in place of thymidine. This does not alter the base pairing specificity of the subtractor which hybridizes to the target strands to be subtracted. After hybridization, treatment with Uracil-DNA Glycosylase (UDG) removes the uracil bases from the phosphodiester backbone of the DNA strand (17). These abasic sites are sensitive to Nuclease S1 cleavage; this nuclease nicks the abasic strand and then cleaves the strand opposite the nick. The resulting fragments cannot be reamplified, thus both the subtractor strand and the target strand are removed from further participation.

Initial experiments with only 1 round of subtraction show evidence of subtraction but are not sufficient to quantify enrichment and will require additional rounds of subtraction. While promising, this approach has considerable complexity, and we have devoted more attention to simpler alternatives.

**AMPLIFICATION-INERT SUBTRACTOR**

We decided to develop a simpler strategy that would be similar to suicide subtraction but would use a subtractor that would be inert to PCR amplification. Our first tactic used a cleavable DOP primer to prepare both target and subtractor DNA. After PCR preparation of the subtractor the primer sequences on the ends are cleaved off, rendering them inert in future rounds of PCR. To also make the target strands that are hybridized to the subtractor strands incapable of being amplified, the primer ends
are removed from the target/subtractor heteroduplexes after PERT hybridization (15) by "polishing" the ends with Mung Bean Nuclease to remove single strand overhangs. In diagram form we have PCR amplified target molecules (double stranded):

```
pppppppttttttttttttttttttttttpppppp
pppppppttttttttttttttttttttttpppppp
```

and subtractor:

```
ppppppsssssssssssssssssssssppppppp
ppppppsssssssssssssssssssssssppppppp
```

which is digested to give:

```
psssssssssssssssssssssp
psssssssssssssssssssssp
```

When these are mixed, melted and reannealed in subtractor excess, most of the target sequences that have matches in the subtractor population are in heteroduplexes:

```
pppppppttttttttttttttttttttttpppppp
psssssssssssssssssssssp
```

After mung bean nuclease end polishing these are shortened:

```
pTTTTTTTTTTTTTTTTTTTTTTTTp
psssssssssssssssssssssp
```

Thus those target strands duplexed with subtractor strands will lose their priming sites for further amplification, and only the target homoduplexes will retain priming sites and reamplify. The cleavable DOP primer is shown:

```
TAATACGACTCACTATAGGGatccNNATGTGG
[------ T7 ------] [-DOP--]
[Bam ]
```

After amplification of the subtractor DNA the T7 portion of its primer is cleaved off with BamHI. Later rounds of PCR after the initial genome sampling use only the 20-mer T7 primer, ensuring that the subtractor and the subtracted and end-polished target molecules are inert to amplification. Figure 4 shows amplification of YAC DNAs with this primer and apparent complete BamHI cleavage and size reduction of the subtractor pYAC4.
Figure 4. Cleavable primer amplification and cleavage.
Lane 1: size marker, PhiX174/HincII digest. Lanes 2 and 3: primer T7-Bam-DOP amplification of yeast+YAC DNAs from pYAC4 (subtractor) and ADGC9 (target) respectively. There is a good spectrum of product with distinct bands. Products are similar, but clear differences are evident. Lanes 4-6: amplified pYAC4 digested 0, 1, 14 hours with BamHI.

FAILURE OF CLEAVEABLE PRIMER STRATEGY
Although PCR sampling by T7-Bam-DOP priming produces different products as seen in Figure 4, lanes 2 and 3, several attempts at subtraction and reamplification with only the T7 primer did not enrich the ADGC9-specific bands; instead, the subtractor pattern was recovered each time. Although Bam digestion appeared exhaustive, T7 reamplification of the subtractor alone gave robust yields, indicating that significant amounts of subtractor had not been cleaved but remained amplifiable. To quantify this we titrated full-length undigested target into BamHI digested subtractor, and assessed the reamplification products (Figure 5).

Figure 5. Estimation of PCR competent subtractor.
Serial 10 fold dilution of target (T) T7-Bam-DOP amplified ADGC9 into Bam digested subtractor (S), amplified pYAC4. Lanes 1 and 8: 20 bp ladder. Lane 3: T/S=1. Lane 4: T/S=0.1. Lane 5: T/S=0.01. Lane 6: T/S=0.001. Lane 7: T/S=0.0001. The product pattern shifts from ADGC9 to pYAC4 between lanes 4 and 5 (see, for example, shift in product in 80-90 bp range).

From Figure 5 we estimate that greater than 1% of the subtractor is PCR competent and is reamplified with the target after subtraction. Since we need to use subtractor in 10-100 fold excess, this negates any enrichment.

What is responsible for this disastrous obstacle? We surmise two contributing factors. First, the BamHI cleavage may not have gone to completion, even though we digested exhaustively. Second, since primer synthesis is not 100% efficient at each step, a very small proportion, but perhaps one or a few percent, of the synthesized primer can be expected to have variant sequences in the BamHI restriction site. (Another, less
frequent source of variant restriction sites is the known misincorporation rate of Taq polymerase.) These variant primers would participate in amplification but those products cannot be cleaved. Perhaps extreme efforts in both primer purification and BamHI digestion might overcome this problem, but we have chosen to attempt a simpler and more robust approach that eliminates restriction enzyme cleavage.

**SEGMENTAL PRIMER STRATEGY**

Here we reverse the cleavable primer strategy; rather than removing part of the primer sequence after PCR to render the subtractor inert to reamplification, we add a longer primer only to the target. The approach is as follows:

1) Both target and subtractor are sampled by PCR using a 10-mer primer, 10aDOP, shown below (bold). This primer contains the same frequently occurring 3' hexamer, ATGTGG, used in previous DOP primers.

2) The target population exclusively is reamplified using the 30-mer, JunTail-10aDOP, composed of a 5' 20-mer "tail" attached to the 3' 10-mer, 10aDOP.

3) The tailed targets are annealed to the shorter subtractors producing target/subtractor heteroduplexes that have single stranded overhanging tails. These are removed by mung bean nuclease leaving only the 10aDOP portion of the primer on the ends of heteroduplexed target strands.

4) The products at this point consist of 10aDOP-primed subtractor homoduplexes, target/subtractor heteroduplexes that also have only 10aDOP primer sequences on the ends, and JunTail-10aDOP-primed target homoduplexes. This mixture is diluted and reamplified using only the JunTail primer, now missing from the "subtracted" target strands. All the primers are shown below:

```
10aDOP         TGAGATGTGG
JunTail-10aDOP  CCCAAGCTTGCATGCGAATTTGAGATGTGG
JunTail        CCCAAGCTTGCATGCGAATT
```

Steps 3 and 4 can then be repeated multiple times to increase the degree of enrichment as needed. The physical subtraction is accomplished only by the dilution for each such cycle of reamplification, and reamplification with the JunTail primer can only amplify remaining target sequences.

This approach is the latest, but simplest, iteration in our long search for a robust subtractive methodology. With this approach we have obtained at this time good genomic sampling by PCR with the 10aDOP primer, and good extension of the full spectrum of these products using the JunTail-10aDOP primer. We are currently optimizing these steps. We believe we have now identified all the pitfalls in our previous strategies, and
we have designed an approach that avoids them. This approach is simple and direct; the sole enzymatic step is the removal of single stranded overhangs by mung bean nuclease, and, if this is not complete, it will reduce the efficiency of subtraction but should not completely negate it. We expect that this strategy will enable us to finally achieve our goal of simple and robust genomic subtraction. Although the term of this IDEA grant is finished we will continue to pursue this goal because of the overriding importance of the information it should provide about breast cancer and other cancers.

METHODS

GENOMIC SAMPLING BY DOP-PCR

The initial sampling of both target and subtractor genomes is performed using 50 ng of template DNA, 25 ng of primer 10aDOP, TGAGATGTGG, 2.5 millimolar Mg++ in a 50 microliter volume. The PCR regimen is 35 cycles of 94 degrees C for 1 minute, 40 degrees C annealing for 1 minute and 72 degrees C extension for 2 minutes.

The target is then tailed by reamplifying the 10aDOP-primed PCR product (1 microliter used as template in a 50 microliter reamplification) with the JunTail-10aDOP primer [CCCAAGCTTGCATGCGAATTTGAGATGTGG], 25 ng in a 50 microliter reaction, 35 cycles of 30 sec at 94 deg, 1 min at 55 deg, and 2 min at 72 deg.

Subsequent reamplifications after PERT hybridization and nuclease trimming of overhangs use the JunTail primer [CCCAAGCTTGCATGCGAATT] at 2 micromolar and 35 cycles of 30 sec at 94 deg, 30 sec at 55 deg, and 1 min at 72 deg.

SUBTRACTION

Our subtraction technique does not require physical removal of the subtractor DNA and the target strands hybridized to subtractor strands; rather the aim is to render them inert to reamplification by PCR so that they are simply diluted out in the multiple rounds of PERT hybridization, dilution, and reamplification.

Phenol Emulsion Reassociation (PERT, 15): Subtractor and target PCR sampled DNAs are mixed in 10:1 to 30:1 mass ratios in PERT buffer and placed in the thermal cycler. The mixtures are denatured for 5 minutes at 94 degrees, and then cooled rapidly to 37 degrees forming the phenol emulsion. The emulsion is maintained by reforming it every 15 minutes by cycling up to 68 degrees for 30 seconds and cooling back to 37 deg. Reassociation is driven to completion in 24 hours. The PERT buffer is 1.5 M sodium thiocyanate, 120 mM sodium phosphate, 10 mM EDTA, and 8 % phenol.

Removal of single strand overhangs: Mung bean nuclease has the ability to degrade both 5' and 3' single stranded overhangs without endonucleolytic activity and thus is ideal for our strategy to render inert to further PCR amplification any target DNA strands that are heteroduplexed with shorter subtractor strands.

The PERT mixture is diluted in TE and the DNA is precipitated with ethanol. The
precipitate is washed several times with 70% ethanol to remove residual salts and is redissolved in water. Mung bean nuclease and its reaction buffer are added and digestion is allowed to proceed for 30 minutes. The nuclease is removed by phenol/chloroform extraction and the DNA is ethanol precipitated and redissolved in TE. A small aliquot (1%) of the DNA is then reamplified for analysis and the next round of subtraction.

CONCLUSIONS

This is a challenging project that presents many problems. We have made progress in developing a successful genomic subtraction strategy and have described additional approaches that should bring this project to a successful conclusion and provide novel and important characterizations of the patterns of Loss of Heterozygosity that reflect the initiation and progression of breast cancer.

REFERENCES


BIBLIOGRAPHY

No publications or meeting abstracts have resulted from this grant to date.

PERSONNEL (1996-1998)

Christophe Chevillard Ph.D.
Brendan Brinkman B.S.
Roy J. Riblet Ph.D.