AD_____

GRANT NUMBER DAMD17-94-J-4484

TITLE: Cloning Human Chromosome 17 Genes: Candidate Genes for BRCA1

PRINCIPAL INVESTIGATOR: Cheng-Chi Lee, Ph.D.

CONTRACTING ORGANIZATION: Baylor College of Medicine
Houston, Texas 77030

REPORT DATE: October 1996

TYPE OF REPORT: Annual

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1996	Annual (15 Sep 95 - 14 Sep 96)

4. TITLE AND SUBTITLE
Cloning Human Chromosome 17 Genes: Candidate Genes for BRCA1

5. FUNDING NUMBERS
DAMD17-94-J-4484

6. AUTHOR(S)
Cheng-Chi Lee, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Baylor College of Medicine
Houston, Texas 77030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, 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*

From loss of heterozygosity (LOH) studies, it has been determined that three regions 17p13, 17q12-22, and 17q24-25 of human chromosome 17 are frequently deleted in breast cancer. The presence of LOH in these region suggest the location of putative tumor suppressor. Our research is focused on the isolation of chromosome 17 genes. A method for the isolation of chromosome specific cDNAs using high density arrayed cDNA and chromosome specific cosmid libraries was developed. To date we have isolated and mapped 105 unique cDNAs of chromosome 17. Of these we have mapped 72 of these cDNAs to the three regions that have been determined to be associated with LOH in breast cancer. These cDNAs are now potential candidate genes for the putative tumor suppressor associated with breast cancer on human chromosome 17. Potential function of these cDNAs will be determined by comparing sequence information to known protein motif in the genome data base. Genes encoding for proteins involved in cellular functions including signal transduction, transcription and cell cycle pathways are prime candidate for tumor suppressor. These genes will be further characterize to determine if they play any role in breast cancer etiology.

14. SUBJECT TERMS
Breast Cancer , Chromosome 17 genes

15. NUMBER OF PAGES
10

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 US 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.

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

lel ____ 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.

Lel Cheng 10/10/96

PI - Signature Date

Table of Contents

Introduction

The role of genetics in cancer is now well established with the identification of several genes where the presence of mutation has been associated with cancer formation (1). It has been shown that mutations of tumor suppressor genes, as negative regulators of cell division, contribute to oncogenesis by interference with mechanisms restraining cell multiplication. Thus genes encoding proteins involved in cellular functions including signal transduction, transcription and phosphorylation/dephosphorylation cell cycle pathways are prime candidates for tumor suppressor genes. In familial forms of cancer, a combination of germ-line and somatic mutations on each allele results in chromosome loss or deletion, meiotic recombination, or gene conversion. Similar events uncovering recessive somatic mutations also occur in sporadic forms of cancer. The loss of genetic material inherited from one parent can be detected by loss of heterozygosity (LOH) analysis using genetic markers. LOH studies on tumors and linkage analysis in inherited forms of cancers have resulted in identification of several tumor suppressor genes (2). The gene for hereditary breast cancer, namely BRCA 1 is also associated with hereditary cancer of the ovary (2,3). The gene for BRCA 1 was mapped to chromosome 17q21, a region that is also associated with allele losses (loss of heterozygosity, (LOH)) in sporadic breast and ovarian cancer (4,5,6,7,8,9). After intensive investigation by many research groups, the gene BRCA 1 was identified via positional cloning methods (10). Many mutations in BRCA 1 gene were found in patients with hereditary breast and ovarian cancer (11,12). Surprisingly, these studies also show that mutations in BRCA1 are rare in sporadic breast and ovarian cancers that are thought to be due to susceptibility to the disease at this locus (13,14,15). Together, the LOH studies and the lack of mutation in BRCA 1 have led to the proposal that there is another gene within this region of 17q12-q22 that is associated with sporadic breast and ovarian cancer in women (16). Many studies have demonstrated LOH in other regions of human chromosome 17 associated with breast cancer (17,18). These studies indicate that a region telomeric to the P53 gene at 17p13.3 (about 3 cM) is believed to harbor a separate tumor suppressor gene associated with breast cancer. In the meantime, another study also shows that regions 17q24-25 are associated with another tumor suppressor gene (8,16). The challenge to identify these potential sporadic breast cancer genes on the chromosome 17 is expected to be great. The strategies for cloning a disease-related gene included either functional or positional cloning approaches (19). With the effort of the Human Genome Initiative in cDNA and expressed sequence tag (EST) mapping, a candidate gene approach to finding human disease genes has been predicted to be the future trend. Our research interest is focused on the development of new strategies to identify genes from chromosome 17. The isolation of genes transcribed from chromosome 17 will provide candidates for the proposed sporadic breast and ovarian cancer genes. The next phase of this proposal is to find candidate tumor suppressor genes associated with sporadic breast cancer in the regions of 17p13, 17q12-22, and 17q24-25 by combining candidate, functional and positional cloning strategies.

Results

We have reported a method for the isolation of chromosome specific cDNAs using high density arrayed cDNA and chromosome specific cosmid libraries (20). The ability to isolate genes in a chromosome specific manner provides simultaneous identification of the expressed sequence and a chromosomal location. This new technology identifies expressed sequences by reciprocal probing of arrayed cDNA libraries and a chromosome specific cosmid library.

The isolated chromosome specific cDNA clones were sequenced through one pass sequencing from the 5' and 3'ends. The corresponding cosmids were used for *fluorescent in situ hybridization* (FISH) mapping to localize their chromosomal position. The sequence information

was used to generate sequence tag site (STS) primers for polymerase chain reaction (PCR) mapping on chromosome 17 somatic hybrid cell-lines to further confirm the cDNA and the corresponding cosmid map position.

During our last report we described our research on a human placental cDNA library where we have isolated and characterized 42 cDNAs to chromosome 17. In the past year we have completed our goal of arraying 40,000 clones from a placental and an ovarian cDNA libraries. The generation of PCR products for high density filters from these libraries have also been completed. This phase of the project is on target with our statement of work. Our preliminary analysis of the human ovarian library indicates that the level of gene diversity is not high. This is not surprising since the ovary is a highly specialized organ. To reduce redundancy of clones and to increase the level of gene diversity we elected to work with a human heart tissue library. In line with this we carried out the arraying of a 20,000 clone human heart cDNA library. Initial characterization of this library suggested a relatively high level of gene diversity and we elected to proceed with this library for our next level of study. Probes from 20,000 heart cDNAs were used to screen a human chromosome 17 cosmid library. A total of 732 clones from the chromosome 17 cosmid library were identified to contain expressed sequences. Keeping in mind that the cosmid libraries were generated with a 5-10 X coverage of the chromosome, this number of cosmids could be associated with about 60-80 potential genes. Indeed our analysis of these cosmids gave 63 unique cDNAs in addition to those found in the placenta studies. As can be seen from our progress we have exceeded our goals for the first 24 months. We are in the of process identifying cosmids associated with expressed sequence for probes from the ovarian library.

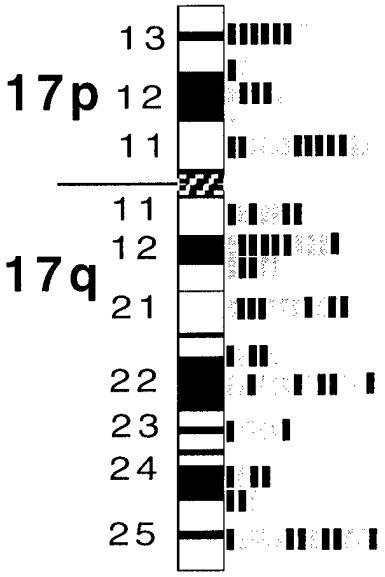
So far using our approach, 105 unique cDNAs of chromosome 17 have been identified. Each of the cDNAs identified in this study was sequenced from the 5' and 3' ends of the clone. Sequences were searched against the Genome data base to determine homologies to existing genes of known function. From these searches, we were able to identify human genes that have not been described. These include genes that are novel, i.e. no homology in the data base and sequences with homology to expressed sequence tags (EST). Homology to non-human DNA and protein sequence motifs were also included in the group of genes to be studied. cDNAs with sequence homology to characterized human genes were excluded from further studies. 43 of the 105 cDNAs we identified are either novel or have homology to ESTs in the data base; 16 cDNAs have previously been described and mapped to chromosome 17; the remaining 46 cDNAs are homologous to previously described but unmapped genes (see Table 1). These cDNAs can be further grouped into their respective mapped position on human chromosome 17. We mapped 8 cDNAs to the region 17p13, 44 cDNAs to the region of 17q12-22, and 20 cDNAs to the region of 17q24-25. The distribution of the 105 characterized cDNA is shown in Figure 1. Together, there are 72 cDNAs localized in the regions of 17p13, 17q12-22 and 17q24-25, three regions targeted in this study.

Table.1 . Identified cDNAs from Chromosome 17

# of cDNA previously mapped to chromosome 17	# of cDNAs with Known Function but not mapped	# of cDNAs with Unkown Function
16	46	43

Figure. 1

Mapped position of 105 Human Genes



▮ — A described gene

▮ — A novel gene or matched an EST.

Given that there are 72 cDNAs mapped to the targeted regions we have to devise selection criteria for taking these cDNAs for detailed characterization. This is essentially our plan of research from month 18-48. We propose to carry out the following strategies for analysis of these clones as candidates for tumor suppressor genes.

1). Isolate full-length cDNAs of the genes localized in the LOH region by screening specific primed cDNA libraries, and obtain the full-length DNA sequence by automatic fluorescence sequencing method. Based on a cDNA's sequence information we can determine its potential biological function by its homology to existing protein motifs. The encoded amino acid sequences will be searched against protein databases. The rapid development of computational methods and search tools have greatly facilitated the biologist to compare and determine the function of novel genes. Recently our department has developed a new search tool, namely BEAUTY, to compare the novel gene with genes of known function in databases (21). Unlike conventional search methods this search tool is focused on the functional domain similarity comparison, which gives us the valuable information about possible functions of novel genes. Genes with protein motifs that falls under the following cellular functions including signal transduction, transcription and phosphorylation/dephosphorylation cell cycle pathways will be chosen for further studies. Once identified, the construction of the gene structure and exon-intron boundary by primer directed sequencing of cosmids associated with their corresponding cDNAs will be carried out. Although we have mapped 72 cDNAs to the regions of 17p13, 17q12-22, and 17q24-25 by FISH, we still do not know if these cDNAs are localized in the narrower regions where sporadic breast cancer patients showed LOH. In order to answer this question, we need to find the polymorphic marker related to these cDNAs. The cosmids associated with the cDNAs will be digested with a combination of restriction enzymes into around 500 bp fragments. The digested fragments will be subcloned into M13 vector and screened by a (CA)13 repeat oligonucleotide probe. The positive clones will be sequenced. The level of heterozygosity for the microsattelite repeat is determine by PCR based typing on genomic DNA of twenty individuals with primers derived from the sequences flanking the (CA)n repeat. Heterozygosity larger than 0.6 will be considered as a polymorphic marker. In similar line we will use the existing LOH microsatellite

markers to determine if they are located in the cosmids associated with our genes. DNA from breast tumor cell lines will be used for typing the LOH microsatellite obtained from our genes. This will determine whether any of these genes are located in the LOH regions. Mutation detection will be PCR based single strand conformation polymorphism (SSCP), heteroduplex analysis, and chemical mismatch cleavage methodologies. Sequence the mutated region by PCR amplification of the corresponding genomic DNA for confirmation.

2). We would like to embark on an approach parallel with from the strategy described above. This involved the screening of genes for microsatellite (dinucleotide and trinucleotide) repeat sequences. Over the past six years, 8 human diseases have been discovered that are associated with trinucleotide repeat expansion (22). The trinucleotide repeat expansion mechanism has provided a solution to the non-mendelian genetics clinically termed anticipation. Anticipation is used in heritable diseases to describe the increasing incidence of a disease in a family from generation to generation as well as the earlier onset of the disease in individuals from generation to generation. Simple trinucleotide repeats are now proven to be the molecular origin for anticipation in seven of these heritable diseases. They include spinobulbar muscular atrophy (SBMA), fragile X syndrome (FX), Myotonic dystrophy (MD), Huntington's chorea disease (HD), Spinocerebellar ataxia type I (SCA), Dentatorubral-pallidolusian atrophy (DRPLA), and Machado-Joseph disease (MJD). The exception to this rule was the recently identified gene for Friedreich ataxia which is autosomal recessive without genetic anticipation. A novel trinucleotide repeat expansion of GAA within an intron sequence was discovered to cause Friedreich ataxia. This recent discovery suggest other variations of microsatellite sequences are involved in other human diseases. The majority of breast cancer are non-Mendelian, i.e sporadic form. Thus this mechanism based on microsatellite instability is an excellent idea for explaining sporadic human diseases including breast cancer. In fact, colon cancers due to mutation in DNA mismatch repair genes have a high level of microsatellite instability (23). Our goal is to identify genes with these microsatellite features and to type them on breast cell line DNA to determine the presence of aberrant alleles. The aberrant alleles can be identified from normal length polymorphism by comparing the alleles from disease state against non-disease state.

Conclusions

The proposal goals for the past year focused on the development of resources and the application of this technology so that genes specific to chromosome 17 could be identified rapidly. To date we have successfully generated the resources proposed for this project and have utilized them for the isolation of 105 unique cDNAs. In addition to the proposed libraries we have also added a human heart cDNA library to this study. To date we have kept up with the 5' and 3' sequencing of the cDNAs isolated. The information obtained from the 5' and 3' ends of these cDNAs has allowed for the search against the Genome data base for sequence homology. Of the characterized cDNAs, 43 of these cDNAs are either novel or have homology to ESTs in the data base; 16 cDNAs have previously been described and mapped to chromosome 17; the remaining 46 cDNAs are homologous to previously described but unmapped genes. We mapped 8 cDNAs to the region 17p13, 44 cDNAs to the region of 17q12-22, while 20 cDNAs to the region of 17q24-25, three regions frequently observed to have LOH in breast cancer tissue. We also proposed to isolate the full-length cDNA to these genes and the generation of genomic DNA material from disease material for LOH studies.

The present study demonstrates that this novel strategy for isolating chromosome specific gene is efficient. The reagents generated by the reciprocal probing strategy including cDNAs, map cosmids, and sequence tag site (STS) primers can provide a high level of transcript map characterization. The isolation and mapping of chromosome 17 cDNAs has provided candidates for the proposed sporadic breast cancer genes and other human diseases mapped to this chromosome.

References

1. Knudson, A. G., Hereditary cancer, cancer oncogenes, and antioncogenes. *Cancer Res.* **45;** 1437-1443, (1985).

2. Narod, S. A., Feuteun, J., Lynch, H. T., Watson, P., Conway, T., Lynch, J., and Lenior,G. M. Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet.* **338:** 82-83, (1991).

3. Easton, D. F., Bishop, D. T., Ford, D., and Crockford, G. P. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The breast Cancer Linkage Consortium. *Am. J. Hum. Genet.* **52:** 678-701, (1993).

4. Futreal, P. A., Sodervist, P., Marks, J. R., Iglehard, J. D., Cochran, C., Barrett, J. C., and Wiseman, R. W. Deletion of frequent allelic loss on proximal chromosome 17q in sporadic breast carcinoma using microsatellite length polymorphism. *Cancer Res.* **52:** 2624-2627, (1992).

5. Jacobs, I. J. A deletion unit on chromosome 17q in epithdial ovarian tumors distal to the familial breast/ovarian cancer locus. *Cancer Res.* **53:** 1218-1221, (1993).

6. Saito, H., Inazawa, J., Salto, S., Kasumi, F., Koi, S., Sagae, S., Kudo, R., Saito, J., Noda, K., and Nakamura, Y. Detailed deletion mapping of chromosome 17q in ovarian and breast cancer: 2 cM region on 17q21.3 often and commonly deleted in the tumors. *Cancer Res.* **53:** 3382-3385, (1993).

7. Cropp, C. S., Champeme, M. H., Lidereau, R., and Callahan, R. Identification of three regions on chromosome 17q in primary human breast carcinomas which are frequently deleted. *Cancer Res.* **53:** 5617-5619, (1993).

8. Nagai, M. A., Pacheco, M. M., Breutani, M. M., Marques, L. A., Breutani, R. R., Ponder, B. A. J. and Mulligen, L. M. Allelic loss on distal chromosome 17p is associated with poor prognosis in a group of Brazilian breast cancer patients. *Br. J. Cancer.* **69:** 754-758, (1994).

9. Claus, E. B., Thompsom, W. D., and Risch, N. Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. *Am. J. Hum. Genet.* **48:** 232-242, (1991).

10. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochron, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Stran, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bodgen, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P., Norris, F., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, C., Lewis, C., Neuhasen, S., Canonalbright, L., Golgar, D., Wiseman, R., Kamb, A., and Skolnick, M.H. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science.* **266** (5182): 66-71, (1994).

11. Futreal, P, A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L., Haugen-strano, A., Swensen, J., Miki, Y., Eddington, K., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Soderkvist, P., Terry, L., Jhanwar, S., Berchuck, A., Iglehart, J.D., Marks, J., Ballinger, D.G., Barret, J.C., Skolnick, M. H., Kamb, A., and Wiseman, R. BRCA1 mutations in primary breast and ovarian carcinomas. *Science.*

266 (5182): 120-122, (1994).

12. Inoue, R., Fukutomi, T., Ushijima, T., Matsmoto, Y., Sugimura, T., And Nagao, M. Germline mutation of BRCA1 in Japan breast cancer families. *Cancer Res.* **55:** 3521-3524, (1995).

13. Takahashi, H., Behbakht, K., McGover, P. E., Chiu, H., Couh, F. J., Weber, B. L., Friedman, L. S., King, M., Furusato, M., Livolsi, V., Menzin, A. W., Liu,P. C., Benjamin, I., Morgan, M. A., King, S. A., Rebane, B. A., Cardonick, A., Mikuta, J., Rubin, S. C., and Boyd, J. Mutation analysis of the BRCA 1 gene in ovarian cancer. *Cancer Res.* **55:** 2998-3002, (1995).

14. Merajver, S. D., Pham, T. M., Caduff, R. F., Chen, M., Poy, E. L., Cooney, K. A., Weber, B. L., Collins, F. S., Johnston, C., and Frank, T. S. Somatic mutations in the BRCA 1 gene in sporadic ovarian tumors. *Nature Genet.* **9:** 439-443, (1995).

15. Hosking, L., Trowsdal, J., Nicolai, H., Solomon, E., Foulks, W., Stamp, G., Siner, E., and Jeffreys, A. A somatic BRCA 1 mutation in an ovarian tumour. *Nature Genet.* **9:** 343-344, (1995).

16. Radford, D, M., Fair, K. L., Philips, N. J., Ritter, J. H., Steinbrueck, T., Holt, M. S., and Donis-Keller, H. Allelotyping of ductal carcinoma *in situ* of the breast: deletion of loci on 8p, 13q, 16q, 17p, and 17q. *Cancer Res.* **55:** 3399-3405, (1995).

17. Cornelis, R. S., van Vliet, M., Vos, C. B., Cleton-Jansen, A., van de Vijer, M., Peterse, J., Khan, P. M., Borrensen, A., Cornelisse, C., and Devilee, P. Evidence for a gene on 17p13.3, distal to p53, as a target for allele loss in breast tumors without p53 mutations. *Cancer Res.* **54:** 4200-4206, (1994).

18. Coles, C., Thompson, A. M., Elder, P. A., Cohen, B. B., Mackenzie, I. M., Cranston, G., Chetty, U., Mackay, J., Macdonald, M., Makamura, Y., Hoyheim,B., and Steel, C. M. Evidence implicating at least two genes on chromosome 17p in breast carcinogenesis. *Lancet.* **336:** 761-763, (1990).

19. Collins, F. S. Positional cloning: Let's not call it reverse anymore. *Nature Genet.* **1:** 3-6, (1992).

20. Lee, C. C., Yazdani, A., Wehnert, M., Zhao, Z. Y., Lindsay, E. A., Bailey, J., Coolbough, M. I., Couch, L., Xiong, M. Chinault, A. C., Baldini, A., and Caskey, C. T. Isolation of chromosome-specific genes by reciprocal probing of arrayed cDNA and cosmid libraries. *Hum. Mole. Genet.* **Vol 4** , 8: 1373-1380, (1995).

21. Brenner SE. BLAST, Blitz, BLOCKS and BEAUTY: sequence comparison on the net. *Trends in Genetics.* **11(8)**, 330-1, (1995).

22. Warren, S.T. The expanding world of trinucleotide repeats. Science, **271,** 1374-1375 (1996).

23. Kolodner RD. Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends in Biochemical Sciences.* **20(10)**, 397-401 (1995).