AD_____

MIPR NUMBER: 93MM3508

TITLE: Role of LINE-1 Retrotransposons in Human Breast Cancer

PRINCIPAL INVESTIGATOR: Thomas G. Fanning, Ph.D.

CONTRACTING ORGANIZATION: Armed Forces Institute of Pathology
Washington, DC 20306-6000

REPORT DATE: October 1995

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1995	Annual 27 Oct 92 - 30 Sep 95

4. TITLE AND SUBTITLE
Role of LINE-1 Retrotransposons in Human Breast Cancer

5. FUNDING NUMBERS
93MM3508

6. AUTHOR(S)
Thomas G. Fanning, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armed Forces Institute of Pathology
Washington, DC 20306-6000

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 human *LINE-1* retrotransposon (*L1Hs*) is known to cause mutations by inserting into genes and inactivating them. The element is expressed in many breast tumors and breast tumor cell lines, suggesting that *L1Hs*-induced mutations may play some role in this malignancy. Prior evidence suggested that unmethylated *L1Hs* elements are transcriptionally active. Therefore, we used the inverse polymerase chain reaction technique to clone 12 elements that are unmethylated in the breast tumor cell line, T47D. Characterization of these clones shows that some elements are unmethylated in other breast cancer cell lines, but are methylated in several normal breast cell lines that do not express *L1Hs*. In addition, *L1Hs* elements that are unmethylated in malignant breast cells are usually methylated in malignant germ cells that express *L1Hs*. Our results suggest the possibility that different subsets of active *L1Hs* elements are expressed in different cancers.

14. SUBJECT TERMS	15. NUMBER OF PAGES
LINE-1 retrotransposon; Breast cancer	12
	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

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2.
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- Leave blank.
DOE	- Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- Leave blank.
NTIS	- Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

*U.S.GPO:1993-0-358-779

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.

____ 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.

Thomas C. Fanning Oct. 18, 1995

PI - Signature Date

Table of Contents

Introduction

Human *LINE-1* (*L1Hs*) is a transposable element that encodes a reverse transcriptase and moves *via* an RNA intermediate [1]. It therefore seems possible that cells in which *L1Hs* is active may be subject to insertional mutagenesis. We have recently found that the element is expressed in a significant number of germ cell cancers [2,3], as well as in many breast tumors and breast tumor cell lines [4]. This last finding raises the possibility that the initiation or progression of some breast cancers is facilitated by *L1Hs*-induced insertional mutations.

In addition to insertional mutagenesis, there are several other characteristics of the *L1Hs* element that suggest its potential as an oncogenic agent. For example, *L1Hs* has an internal promoter which could potentially lead to readthrough transcription and activation of downstream genes. In addition, the *L1Hs*-encoded p40 protein has a leucine zipper motif, suggesting a possible interaction with other cellular proteins. Such interactions at inappropriate times might lead to the disruption of important cellular functions. Thus, *L1Hs* involvement in cancer could occur by several mechanisms, either singly or in combination.

Our long range goals focus on the identification and characterization of *L1Hs* elements that are expressed in breast cancers. *L1Hs* expression may identify a class of breast neoplasms in which *L1Hs*-encoded proteins initiate or maintain the neoplastic state. This suggests that down-regulating *L1Hs* expression could influence the course of breast cancer. To approach this problem we have proposed to study the expression of *L1Hs* in cell lines and in solid tumors by cloning and characterizing actively expressed *L1Hs* elements. Specifically we have proposed to:

(1) Clone active *L1Hs* elements from several breast cancer cell lines.
(2) Characterize the active *L1Hs* sequences and determine if the same *L1Hs* elements are active in solid tumors.

Experimental

Isolation of hypomethylated *L1Hs* elements.

Our original plan was to isolate unmethylated (potentially active) *L1Hs* elements by the method of inverse polymerase chain reaction (PCR) [5], and this approach has worked quite well. Our current protocol runs as follows: we partially cleave DNA extracted from the T47D breast tumor line with *Hpa*II, an enzyme that appears to cleave *L1Hs* sequences only in transcriptionally active cells [6]. We then isolate fragments from 0.3-1.3 kb in size on acrylamide gels and circularize the fragments by incubation with T4 DNA ligase under conditions of dilute DNA concentration (1-2 ng/μl). We collect the circularized DNAs and cleave with the restriction enzyme *Bgl*I (*Bgl*I has a consensus site located at position 10-20 on the *L1Hs* consensus sequence; the enzyme introduces cuts between nucleotides 13/14 and 16/17 giving 3' overhangs) and perform PCR under relaxed conditions with primers that are complementary to regions 3-16 and 14-27 on the *L1Hs* consensus sequence. This results in PCR products that contain 5' flanking DNA bracketed by *L1Hs* sequences. In most cases the *L1Hs* moiety of the clone stretches from bp 1 to a *Hpa*II site at bp 36.

Following inverse-PCR and cloning of the products into pBluescript KS or pDK101 [7], we isolated 12 unique clones deemed suitable for analysis (Tables I and II, pp. 10-12). We use the following nomenclature when describing the cloned sequences: when discussing the

clone, without regard to the *L1Hs* or the flanker moiety, no additional designation is given (*e.g.*, clone 593B). When referring only to the *L1Hs* moiety of the clone, an -L1 suffix is provided (*e.g.*, 593B-L1). When referring only to the sequence flanking the *L1Hs* moiety, an -Fl suffix is provided (*e.g.*, 593B-Fl).

Sequences flanking the hypomethylated *L1Hs* elements.

The sequences directly flanking the 5' end of the 12 hypomethylated *L1Hs* elements were of special interest to us since it is possible that they harbor regulatory signals that distinguish these *L1Hs* elements from the bulk of the elements in the genome. Therefore, we sequenced and analyzed about 220 bp of each flanker directly adjacent to the *L1Hs* element. The sequences were unremarkable. In general they were A+T-rich (average = 61%; range = 51-75%) with no large regions of strand asymmetry (*e.g.*, no polypurine runs). Although the occasional clone had a transcriptional regulatory motif (*e.g.*, CAAT), most did not. It is possible, of course, that uncommon variants of a common motif are present in the flanking sequences, or that novel regulatory signals are present and not recognizable. We have begun to analyze selected flankers in promoterless and enhancerless expression plasmids to test these possibilities.

A GenBank search provided matches with only three of the flanker sequences. 893J-Fl consists entirely of an *Alu* element. 893I-Fl contains multiple copies of the four bp sequence, TGGA. 593F-Fl consists of a novel repetitive sequence (CASP repeat) that is being characterized by us. Currently we know that the repeat is present in about 250-300 copies in Old World monkeys and is absent in New World monkeys. In humans, copies of the sequence are located on all chromosomes, including the Y chromosome.

Chromosomal location of hypomethylated *L1Hs* elements.

We prepared primers specific for each flanker and used them together with an *L1Hs*-specific primer to amplify *L1Hs*-flanker fragments in somatic cell hybrid DNAs, each containing a specific human chromosome in a rodent background (Coriell Mapping Panel #2). Eleven of the twelve flankers were assigned this way to specific chromosomes (Table I). The 11 hypomethylated *L1Hs* elements are found on 7 different chromosomes, with two each on chromosomes 5 and 8 and three on chromosome 15. This non-random distribution could reflect a non-randomness in the hypomethylation of various regions of the human genome during tumorigenesis [8,9].

One locus, 893N, yielded no PCR product with any of the chromosome-specific hybrid DNAs, although DNA isolated from the human cell line used to make the hybrids did give a signal. Thus, the locus containing 893N must have been deleted sometime subsequent to the fusion of the human and rodent cells. Other workers have previously reported that some chromosomal loci are deleted in the Mapping Panel #2 lines [10].

Methylation status of hypomethylated *L1Hs* elements in different cell lines.

We were able to analyze the methylation status of 9 loci in a number of cell lines (Table II). The methylation status of three loci (593F, 893I, 893J) could not be ascertained since the flankers contained sequences that are multicopy in the human genome. In Table II a minus sign indicates the sequence is primarily unmethylated while a minus/plus indicates that the

locus was partially unmethylated and partially methylated. This could be due to differential methylation of the two alleles, or to differential methylation of the locus in different cells, or both. Note also that many of these cells are aneuploid and the copy number of the locus being probed is not necessarily two/cell. Pluses in Table II indicate that the locus is methylated: a single plus indicates that a defined, high molecular weight band was present, while triple plus indicates that only a smear of hybridizing material was seen at the very top of the lane. A double plus indicates an intermediate condition.

The first conclusion from Table II is that certain loci are hypomethylated in a variety of malignant cell lines. For example, locus 593E appears to be hypomethylated in all cell lines except those derived from germ cell tumors. Thus, 593E may represent a locus that is demethylated late in embryogenesis. 593B has a pattern similar to that of 593E, but it is partially hypomethylated in one germ cell line and is methylated in several non-germ cell lines. 893Y is methylated in all lines except T47D (breast) and DAOY (medulloblastoma). 593C is not methylated in cell lines originating from breast tumors, but is methylated in almost every other cell line.

The one non-breast cell line where 593C is also hypomethylated is the medulloblastoma line, DAOY. An examination of Table II demonstrates that the T47D and DAOY lines have virtually identical methylation patterns for all nine hypomethylated loci. In a separate series of experiments, using hypomethylated *L1Hs* flankers isolated from NTera2D1 cells, we found a clone that demonstrated an RFLP between the T47D and DAOY cell line DNAs with the restriction enzyme *Msp*I. Thus, the DAOY DNA is clearly not T47D DNA that had been mislabeled. We have no explanation for the very similar hypomethylation patterns in these two cell lines derived from quite different tumors.

The second conclusion from Table II is that two subsets of hypomethylated loci are present: one in somatic cells and one in germ cells. With only a few exceptions, all of the loci that are hypomethylated in various malignant and non-malignant somatic cells are fully methylated in the germ cells. This indicates that some regions of the genome that are demethylated in tumors derived from somatic cells remain highly methylated in germ cell tumors. In this respect it is interesting to note that all of the loci that are hypomethylated in T47D cells are highly methylated in the NTera2D1 cell line. Two loci (893O and 893V) are hypomethylated in both 2102Ep and PA-1, and one (593B) is hypomethylated only in PA-1. The reason for the complete methylation of the loci in NTera2D1, but not in the other two germ cell lines, is presently unknown. One possible explanation stems from the fact that the NTera2D1 line is capable of retinoic acid induced differentiation and might reflect the real embryonic cell more faithfully that either of the other two germ cell lines.

Originally, we expected to find loci that were hypomethylated because they had been involved in amplification events. However, we found only one such case: 593C exhibits an approximate 10-15 fold amplification in the SK-BR-3 line. 593C is located on chromosome 8 (Table I). The long arm of chromosome 8 is known to be amplified in a number of primary breast tumors and the 8q21-8q24 region is amplified in the SK-BR-3 cell line [11]. Therefore, it seems likely that 583C is within the 8q21-8q24 region.

Cloning of full-length hypomethylated *L1Hs* elements.

We have screened a human genomic library in phage lambda with some of the flanker

sequences described above. To date we have lambda clones for the following hypomethylated elements: 593C (11 clones of which 10 have been restriction mapped), 593D (2 clones; 1 mapped), 593E (8 clones; 4 mapped), 593F (2 clones; 0 mapped). These clones will be used for coupled transcription/translation assays to determine if the associated *L1Hs* elements are capable of producing intact, functional proteins.

In addition, we have isolated 40 lambda phage that contain CASP sequences (see above) selected randomly from the genome, and 33 lambda phage with CASP sequences located on chromosome 9. We are currently attempting to determine if the CASP repeat is involved in any aspect of cellular function.

Conclusions

We have partially characterized a number of *L1Hs* elements that may be active in several breast cancer cell lines. Our results suggest that *L1Hs* elements that are unmethylated in breast tumors are methylated in germ cell tumors. This suggests that specific, non-overlapping sets of elements are active in different cancers. We are currently testing this by investigating unmethylated *L1Hs* elements from an embryonal carcinoma cell line and from a medulloblastoma cell line. If there are cancer-specific patterns of *L1Hs* expression, the probes we are developing today may be useful as clinical tools in the future.

The intent of the present work was to characterize the *L1Hs* elements that are differentially hypomethylated in malignant cell lines. Hypomethylation of *L1Hs* elements or surrounding sequences may be important in malignant transformation, especially since demethylation of genomic DNA appears to be an early event in tumor progression [8,9]. Of an estimated 120 hypomethylated *L1Hs* elements per T47D cell, we have isolated and partially characterized twelve. If these twelve elements are representative, then they should provide a picture of the hypomethylated, potentially active, *L1Hs* elements in malignant cells.

Our results suggest that there are many *L1Hs* elements that can be demethylated during tumorigenesis, and some of these elements are demethylated in a variety of cell lines from independent tumors. For example, the locus 593C is hypomethylated in three breast cancer cell lines and one brain cancer cell line, but is methylated in all other lines (Table II). Locus 893V is hypomethylated in one of three non-malignant cell lines and two of three breast cell lines, two of three germ cell lines and two of three brain cell lines (Table II). Thus, our findings eliminate many models that might have described hypomethylated *L1Hs* elements in malignant cells: for example, that each cell line has a different subset of hypomethylated elements.

Why certain loci are hypomethylated in numerous malignant cell lines is presently unknown. These regions of the genome may be demethylated randomly, then held in a demethylated state by selection. Selection could result from the presence of a gene or *L1Hs* element in the hypomethylated region whose transcription is necessary to maintain the malignant state. We note that *L1Hs* need not necessarily act as an insertion mutagen in order to modify the cellular phenotype. The element encodes at least two proteins whose functions (with the exception of the reverse transcriptase encoded by ORF2) are largely unknown. The p40 protein, the reverse transcriptase, or possibly other polypeptides originating from ORF2, could be involved in the initiation or maintenance of the malignant state, either alone or in combination with other gene products.

References

[1] Fanning, T.G. and M.F. Singer (1987) *LINE-1*: a mammalian transposable element. *Biochim. Biophys. Acta* 910:203-212.

[2] Bratthauer, G.L. and T.G. Fanning (1992) Active *LINE-1* retrotransposons in human testicular cancer. *Oncogene* 7:507-510.

[3] Bratthauer, G.L. and T.G. Fanning (1993) *LINE-1* retrotransposon expression in pediatric germ cell tumors. *Cancer* 71:2383-2386.

[4] Bratthauer, G.L., R.D. Cardiff and T.G. Fanning (1994) Expression of *LINE-1* retrotransposons in human breast cancer. *Cancer* 73:2333-2336.

[5] Ochman, H., A.S. Gerber and D.L. Hartl (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120:621-623.

[6] Thayer, R.E., M.F. Singer and T.G. Fanning (1993) Undermethylation of specific *LINE-1* sequences in human cells producing a *LINE-1*-encoded protein. *Gene* 133:273-277.

[7] Kovalic, D., J-K. Kwak and B. Weisblum (1991) General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nuc. Acids Res.* 19:4560.

[8] Feinberg, A.P. and B. Vogelstein (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89-92.

[9] Goelz, S.E., B. Vogelstein, S.R. Hamilton and A.P. Feinberg (1985) Hypomethylation of DNA from benign and malignant human colon neoplasias. *Science* 228:187-190.

[10] Dubois, B.L. and S.L. Naylor (1993) Characterization of NIGMS human/rodent somatic cell hybrid mapping panel 2 by PCR. *Genomics* 16:315-319.

[11] Kallioniemi, A., O-P. Kallioniemi, J. Piper, M. Tanner, T. Stokke, L. Chen, H.S. Smith, D. Pinkel, J.W. Gray and F.M. Waldman (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl. Acad. Sci. USA* 91: 2156-2160.

Notes to Table I.

L1Hs (bp): size of the L1Hs moiety in clone.
Flank (bp): size of the DNA flanking the 5' end of the L1Hs element.
Copy no.: estimated copy number of flank DNA.
*: HpaII site at position 36 is CCGG, but not cut by enzyme.
?: could not be determined.

Notes to Table II.

Type: normal is non-malignant breast epithelial (HBL-100, MCF10A) or fibroblast (GM0327); breast is adenocarcinoma (MCF7, SK-BR-3) or infiltrating ductal carcinoma (T-47D); brain is medulloblastoma (DAOY, TE671) or glioblastoma (A172); germ cell is testicular embryonal carcinoma (2102Ep, NTera2D1) or ovarian embryonal carcinoma (PA-1).
Scoring system (-, -/+, *etc.*) is given in text.
?: could not be determined.
xx: locus rearranged in this cell line.

Table I. Location and Size of Clones

clone	chromosome	L1Hs (bp)	flank (bp)	copy no.
593B	5	103	1200	single
593C	8	36	500	single
593D	4	103*	500	single
593E	8	36	700	single
593F	9	36	400	250-300
893I	15	103*	350	multicopy
893J	15	103	350	multicopy
893N	?	36	200	single
893O	7	36	400	single
893V	15	36	450	single
893W	5	36	350	single
893Y	3	103	150	single

Table II. Methylation Status of Clones

cell line	type	593B	593C	593D	593E	893N	893O	893V	893W	893Y
HBL100	normal	-	+ + +	+ +	-	+ + +	?	+ + +	+	+ + +
MCF10A	"	-	+ + +	-/+	-	+ + +	-/+	-/+	+	+ + +
GM0327	"	-/+	+ + +	+ +	-/+	+ + +	-/+	+ + +	+	+ + +
MCF7	breast	+ +	-/+	-/+	-	?	?	?	+ + +	?
T-47D	"	-	-	-/+	-	-/+	-/+	-	-/+	-
SK-BR-3	"	-/+	-	-/+	-	+ + +	+ +	-/+	+ +	+ + +
DAOY	brain	-	-/+	-/+	-	-/+	-	-	-/+	-
TE671	"	+ + +	+ + +	+ + +	-	-/+	+ +	+ + +	+ +	+ + +
A172	"	-	+ + +	-/+	-	?	-/+	-	+ + +	+ + +
2102Ep	germ cell	+ + +	+ + +	+ + +	+ + +	xx	-/+	-/+	+ + +	+ + +
NTera2D1	"	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +	+ + +
PA-1	"	-/+	+ + +	+ +	+ +	+ + +	-/+	-	+ + +	+ + +