AD_____

Award Number: DAMD17-99-1-9267

TITLE: Long Interspersed Nuclear Element (LINE) Mediated Genomic Instability

PRINCIPAL INVESTIGATOR: Lalitha Nagarajan, Ph.D.

CONTRACTING ORGANIZATION: The University of Texas
MD Anderson Cancer Center
Houston, Texas 77030

REPORT DATE: July 2000

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. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2000	Annual (1 Jul 99 - 30 Jun 00)

4. TITLE AND SUBTITLE
Long Interspersed Nuclear Element (LINE) Mediated Genomic Instability

5. FUNDING NUMBERS
DAMD17-99-1-9267

6. AUTHOR(S)
Lalitha Nagarajan, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Texas MD Anderson Cancer Center
Houston, Texas 77030

E-MAIL:
lnagarajan@notes.mdacc.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)***

We have isolated and characterized a novel human homeobox gene designated Mix.1 homeobox (*Xenopus laevis*)-like gene 2 (*MIXL2*), as a gene that contains a repetitive element in its 3'UTR (MER22). The *MIXL2* gene encodes a predicted 232 amino acid open reading frame, which contains a highly conserved mix.1-like homeobox and an acidic motif. Radiation hybrid mapping and Fluorescence in situ hybridization localize the *MIXL2* gene to human chromosome 1q32-41, a region of gain in breast tumors. Consistent with these cytogenetic findings we detected allelic imbalance for loci flanking MIXL2 in 14/25 primary breast tumors. Additionally, a search of the human dbEST database suggested a highly restricted expression as we detected only three EST clones (one each from normal germinal center B cell, undifferentiated germ cell tumor, and infiltrating ductal carcinoma of the breast). These findings suggest *MIXL2* may play a role in development and mammary carcinogenesis.

14. SUBJECT TERMS
Breast Cancer, geneomic insatbility, chromosomal gain, 1q41 and homeobox gene

15. NUMBER OF PAGES
16

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

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

FOREWORD

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

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

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

N/A 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.

N/A 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).

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

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

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

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

Lalitha nagaraja 7/19/00

PI - Signature Date

Table of Contents

Introduction:

The overall goals of this proposal are to develop a rapid PCR assay to detect the genome wide Long Interspersed Sequence (L1Hs) mediated instability and to isolate target oncogenes, tumor suppressor genes and regions of chromosomal instability by the use of L1Hs specific PCR. In order to accomplish these goals we modified our approach and took advantage of the recent progress in the human genome intiative and cancer genome anatomy projects (CGAP).

DNA sequence analysis of a novel L1Hs element that we characterized in the laboratory revealed a segment of homology to MER22 repetitive elements. These findings coupled with the rapid progress in the human genome initiative led us to redesign our approach to detect target genes. Thus we screened the cDNA database dBEST for novel genes which contain repetitive elements in them. About the same time, other studies in our laboratory identified a partial cDNA clone for a key regulatory homeobox MIX like gene in the dBEST database. Interestingly, the 3' UTR of this gene contains MER repetitive elements. Furthermore the expression of this novel gene appeared to be highly restricted with the identification only 3 cDNA clones in the entire dBEST database with more than 3 million entries. Our interest in this regulatory gene was further heightened as one of the 3 cDNA clones was from a laser captured tumor tissue from infiltrating mammary adenocarcinoma *in situ*. All of the subsequent studies detailed in this report stemmed from these key observations made in early stages of the first year of this award.

Year 1 of the proposal was to focus on specific Aim 1. The following report summarizes what we have accomplished towards specific tasks in the approved statement of work.

Specific Aim I (i-iii)

(i) Development of primers and conditions for IPCR and L1Hs PCR
(ii) Analysis of hybridization patterns
(iii) Subcloning of unique fragments

In accomplishing the tasks i-iii of Specific Aim I, we have taken advantage of the rapid progress made in the human genome sequencing initiative, the cancer genome anatomy project and more importantly characterization of a novel L1Hs element in our laboratory. The novel element designated LREHL showed segments of homology to MER repetitive elements.

We initiated a screen of the expressed sequence tagged database (dBEST) electronically for transcripts with repetitive sequences. About the same time, other investigations in our laboratory were searching for the human homolog of the homeobox gene *Mix.1*, which regulates mesoderm and endoderm differentiation in Xenopus embryos. A single candidate EST clone (AA84781) with a putative *Mix* like homeodomain was identified in a library from germinal center B lymphoid cells. When the EST clone AA84781 was sequenced, we identified a partial open reading frame. A recursive search of the dBEST database identified only two other cDNA clones for this novel gene, suggesting an extremely restricted pattern of expression. Interestingly, one of these ESTs (AA911377) was from a cDNA library made from laser captured infiltrating ductal carcinoma of the breast (Fig.1). The significance of this finding was further enhanced when we detected alu and MER like repetitive elements in the 3' (> 1.0kbp) UTR of this gene.

(iv) Sequence analysis and development of STSs

We developed unique STSs from the 3'untranslated region of the MIX like homeodomain gene. Using these primers we screened a human genomic library in Bacterial Artificial Chromosome vector to isolate a large segment of the genome containing the gene. The BAC was used to characterize the full length open reading frame of this novel homeogene that we designated *MIXL2*. The gene encodes an open reading frame of 232amino acids. Similar genes whose expression is restricted to the primitive streak have been identified in mice and chicken (1-3). MIXL2 shows 94% identity in mix-like domain with an overall identity of 69% to the murine gene MML (Fig.2). With the chicken gene CMIX, we find a 79% identity in the homeodomain and an overall identity of 41%. Identification of the three distinct and yet highly related genes suggests the existence of a novel mix-like homeobox gene family.

Additionally, we characterized the genomic organization and expression pattern of MIXL2. The gene consists of two exons (Fig. 3) and shows a highly restricted expression pattern. None of the breast cancer cell lines tested showed abundant transcripts. However, reverse transcription coupled PCR detected *MIXL2* transcripts in the cell line MDA-MB 453 and not in the immortalized breast cell line MCF10A.

Simultaneously, we fine mapped the novel gene to human genome using DNA from a panel of well-characterized radiation hybrids. Interestingly, the MIX like gene localized to human chromosome 1q32-41 locus, between the microsatellite loci *D1S479* and, a region of gain in breast cancer (Fig.4).

(v) Development of the PCR for minimal material

In collaboration with Dr. Nour Sneige in the Dept. of Pathology, we dissected 25 breast tumors and isolated normal and tumor DNAs. The status of D1S479 was examined in these paired samples by polymerase chain reaction (PCR) with end labeled primers. Interestingly, 12 out of 25 tumors showed allelic imbalance for *D1S479* suggesting that the MIXL gene was a candidate target gene of chromosomal gain in 1q (Fig. 5).

(vi) Correlation of data from tasks (i-v) with stage and grade

All of the 25 breast tumors used to screen for allelic imbalance of *D1S479* showed lymph node involvement. Studies are underway to determine whether there is intra-tumoral heterogeneity within for gain of 1q32-41 loci. We will utilize both FISH and MIXL specific antibody based immunofluorescence staining approaches to examine such a possibility.

Key Research Accomplishments:

- Identification of a novel regulatory homeobox gene with repetitive elements in the 3'UTR

- Isolation and characterization of the full length open reading frame for the novel gene designated *MIXL2*

- Localization of *MIXL2* to 1q41, a region of gain in breast cancer

- Detection of allelic imbalance for *D1S479*, a marker physically linked to *MIXL2*, in 12/25 breast tumors with lymph node involvement

Reportable outcomes:

Manuscripts in preparation:

1. Hejlik DP and Nagarajan L. Retrotransposition and recombination of LINE-1 Elements as a mechanism of chromosomal deletion and genomic instability

2. Guo W., Ma, J., Liang H., and Nagarajan L. A novel MIX like paired type homeobox gene localizes to a region of in breast cancer

3. Guo W., Chan A., Ma J., Liang H., Etkin L and Nagarajan L. MIXL2 shows evolutionary and functional conservation: a role in hematopoiesis

Development of FISH and immunohistochemical staining probes:

We have characterized a BAC probe that can be used in Fluorescence in situ hybridization (FISH).

We have developed antibodies to MIXL2, which can be used in immuno-histochemical staining.

Informatics:

The cDNA sequence for MIXL2 has been submitted to human genome organization. The full length sequence will be submitted to Genbank.

Funding applied for based on work supported by this award:

Application by Wei Guo entitled "Functional characterization of a novel embryonic homeobox gene MIXL2 in breast cancer" for Graduate training fellowship, from the Breast Cancer Research Program, Department of Defense Submitted on June 7, 2000.

Conclusions:

A genome wide-electronic screening approach yielded us clues on an important regulatory homeobox gene with repetitive elements in the 3' UTR.

The *MIXL2* gene is highly restricted in its expression. There are only 3 cDNA clones for this gene in the entire human cDNA expressed sequence tag (EST) database that contains more than 3 million entries. Interestingly, one of the three ESTs for MIXL2 is from a laser captured infiltrating ductal carcinoma in situ of the breast.

The *MIXL2* gene consists of two exons and encodes an open reading frame of 232 amino acids.

We localized it to a region of chromosomal gain (1q41) in breast cancer.

Interestingly *D1S479*, a marker that is physically linked to *MIXL2* is gained in 12 out of 25 breast tumors with lymph node involvement.

Homeobox genes encode a family of regulatory transcription factors that primarily play a crucial role in development. Other evidences suggest that certain homeobox genes provide an important link between the processes of cell cycle control, development and cancer.

Future experiments will identify the functional consequence of gain of *MIXL2* gene in breast cancer as well as characterization of other novel regulatory genes with L1 repetitive elements.

References:

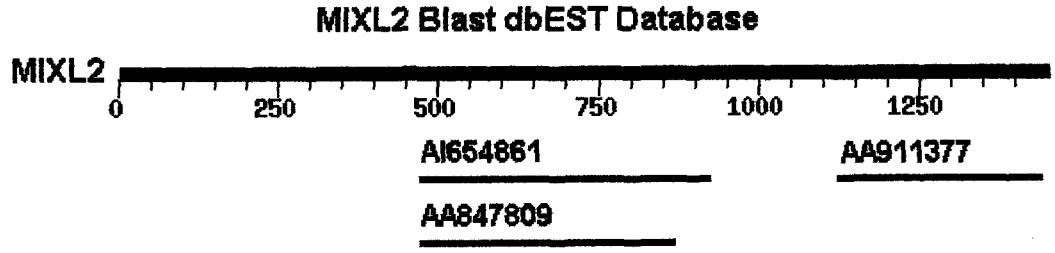
1. Pearce JJ, Evans MJ. Mml, a mouse Mix-like gene expressed in the primitive streak. *Mech. Dev.* **87**:189-192,1999.

2. Stein S, Roeser T, Kessel M. CMIX, a paired-type homeobox gene expressed before and during formation of the avian primitive streak. *Mech. Dev.* **75**:163-165,1998.

3. Peale FV Jr, Sugden L, Bothwell M. Characterization of CMIX, a chicken homeobox gene related to the Xenopus gene mix.1. *Mech. Dev.* **75**: 167-170,1998.

dbEST Database Search

A.



B.

EST	Library	Source
AA847809	NCI_CGAP_GCB1	germinal center B cell
AI654861	NCI_CGAP_GC6	pooled germ cell tumors
AA911377	NCI_CGAP_Br5	infiltrating ductal carcinoma of the breast

Figure 1. A search of dbEST database. AA847809 was found in the dbEST database because the partial sequence of the EST clone shows a mix-like homeodomain. The complete sequencing of the EST clone allows us to retrieve the other two EST clones from different cDNA libraries. The detection of only three EST clones suggests a highly restricted expression of MIXL2.

Homology between human MIXL2 and mouse MML

```
MIXL2           MATAESRALQFAEGAAFPAYRAPHAGGALLPPPSPAAALLPAPPAGPGPATFAGFLGRDP 60
MML             MAAAGSQQLQFAEGAAFPIFPAAHPGGQLLPAMRPASGLPAAPHDSRAPAATQCFPNRDS 60
                **:* *: ********** : *.*.** ***.  **:.* .**  . .**:   * .**.

MIXL2           GPAPPPPASLGSPAPPKGAAAPSASQRRKRTSFSAEQLQLLELVFRRTRYPDIHLRERLA 120
MML             SPTAQTPAGLDPPGPSKGSAAPSAPQRRKRTSFSSEQLQLLELVFRQTMYPDIHLRERLA 120
                .*:. .**.*..*.*.**:*****.*********:***********:* **********

MIXL2           ALTLLPESRIQVWFQNRRAKSRRQSGKSFQPLARPEIILNHC-APGTETKCLKPQLPLEV 179
MML             ALTLLPESRIQVWFQNRRAKSRRQSGKSFQPLSSRRGVFLHCPAPGTEARCLKPQLPLEA 180
                ********************************:  . :: ** *****::*********.

MIXL2           DVNCLPEPNGVGGGISDSSSQGQNFETCSPLSEDIGSKLDSWEEHIFSAFGNF 232
MML             DVNHVPDPSMTGGGVCTSGS--QSFETYSSLSEDIGSKLDSWEEHIFSALGNF 231
                *** :*:*. .***:. *.*  *.*** *.*******************:***
```

Figure 2. Homology between human MIXL2 and mouse MML. MML is a mix-like gene in mice. MIXL2 shows 94% identity on the mix-like domain and overall identity of 69% with mouse MML. The carboxy terminal acidic motif is also conserved.

Genomic Organization of MIXL2

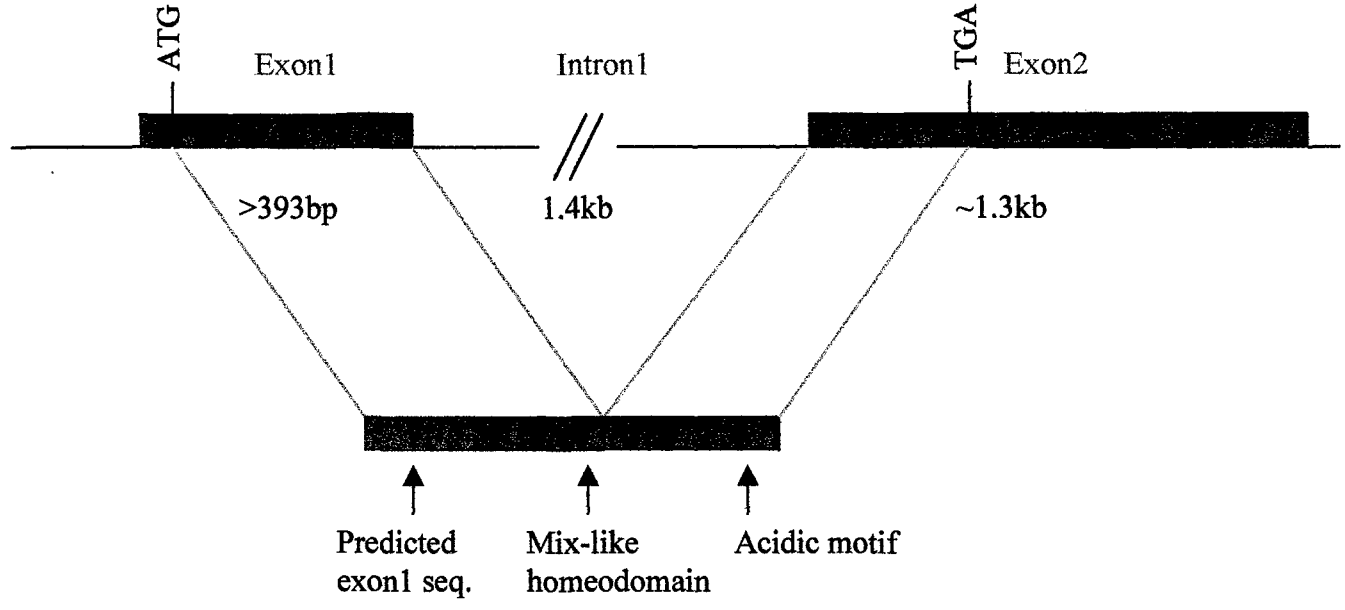


Figure 3. Genomic organization of MIXL2. The predicted MIXL2 consists of at least two exons. The exon1 is predicted to have more than 393bp. 3'-RACE shows that the exon2 has about 1.3kb. The long PCR amplified a 1.4kb intron. The exon-intron junction in the mix-like domain perfectly matches the consensus GT...AG. The part of exon1 sequence is assembled from BAC216I11 sequence. The context around the predicted ATG start codon shows G-3 and G+4 (GGAGCGatgG), which matches the consensus motif of start codon. The EST clone AA847809 covers part of exon1 and exon2. Notes: 3'-RACE uses a Qt-adapted single strand cDNA pool made from KG1 cells (Methods in Enzym., vol 218, 340-357) to amplify MIXL2 exon2 and 3' UTR. The long PCR uses Expand High Fidelity PCR System from Boehringer, and PCR conditions follow instruction in the kit.

Genomic Localization of *MIXL2*

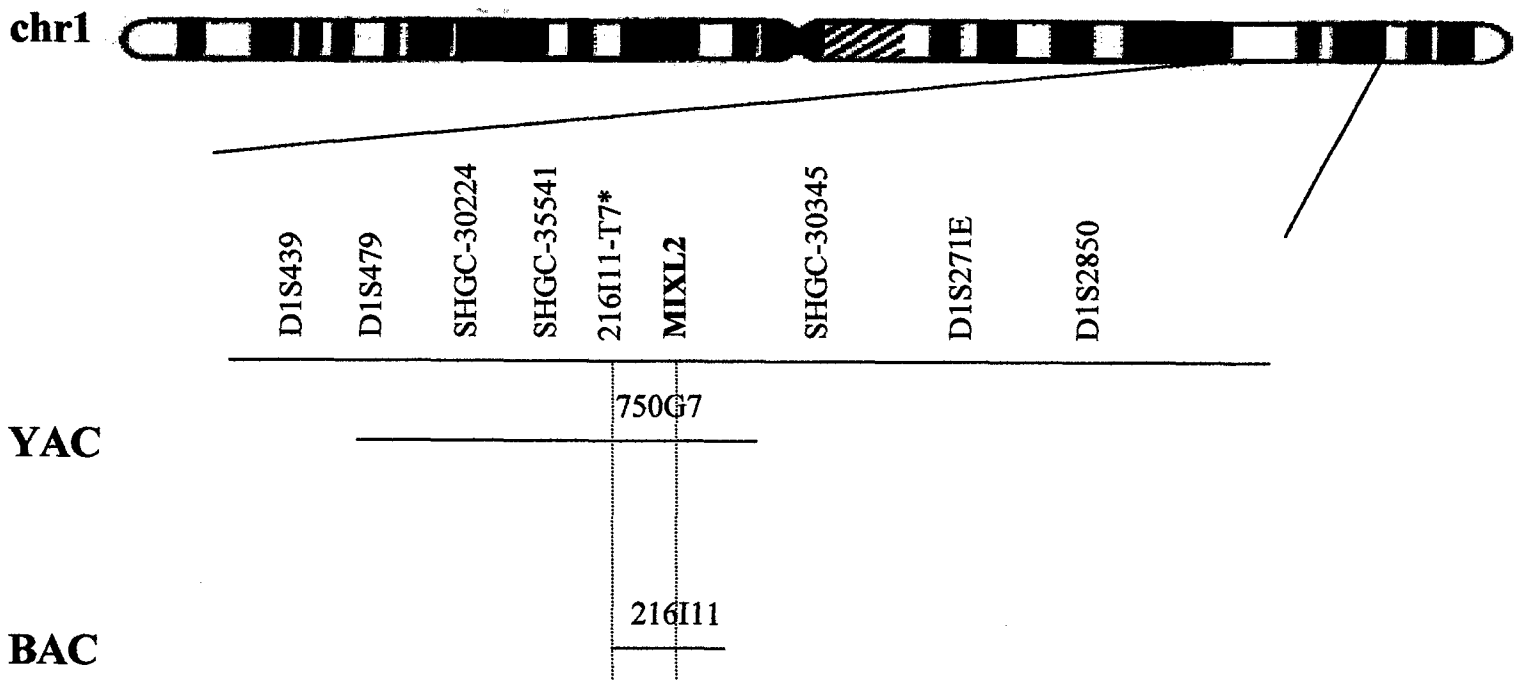


Figure 4. Genomic localization of *MIXL2*.MIXL2 is localized to human chromosome 1q32-41 by radiation mapping and FISH (data not shown). PCR confirms that MIXL2 can hit YAC750G7 in the contig Whitehead Institute Contig WC-516 of Chr1q. The BAC216I11 was retrieved from human BAC library (Research Genetics, Inc) by PCR screening. One end of BAC216I11 also hit YAC750G7

Allelic Imbalance of Chromosome 1q32-41in Breast Cancer

A. Allelic imbalance in 12/25 cases

Case #	1	2	3	4	5	6	7	8	9	10	11	12	13
Accession#	SS97-17729	SS97-18252	SS97-18659	SS97-19607	SS97-22197	SS97-22332	SS97-25357	SS97-25363	SS97-25832	SS97-30517	SS97-31380	SS97-31390	SS97-35459
Tumor	+/+	N	+++/+	N	N	++/+	NI	NI	NI	+/+++	+/+++	++/+	+/+++
Normal	N	N	N	N	N	N	NI	NI	NI	N	N	N	N

Case #	14	15	16	17	18	19	20	21	22	23	24	25
Accession#	SS97-38032	SS97-38312	SS97-38569	SS97-39449	SS97-40702	SS97-42153	SS97-42385	SS97-43544	SS97-44939	SS97-45814	SS97-46375	SS97-51787
Tumor	+++/+	+/+++	+/+++	+/+++	N?	N?	N	N?	N	NI	+/+++	+/+++
Normal	N	N	N	N	N	N	N	N	N	NI	N	N

B.